Abstract

Nuclear Magnetic Resonance Studies of RNA Structure:
An RNase III Processing Signal and the
Anticodon Stem-Loops of Initiator and Elongator Methionine tRNA

David C. Schweisguth

May 1997

The structures of several RNAs, an RNase III processing signal and the anticodon stem-loops of initiator and elongator tRNA, were investigated by NMR spectroscopy and other techniques.

1. The structure of a 41 nucleotide variant of the T7 R1.1 RNase III processing signal from bacteriophage T7 was examined by NMR spectroscopy, optical melting, and chemical and enzymatic modification. It has two Watson-Crick base-paired helices separated by an internal loop, consistent with its predicted secondary structure. The internal loop is neither rigidly structured nor completely exposed to solvent, and seems to be helical. The secondary structure of R1.1 RNA is largely insensitive to the monovalent cation concentration, which suggests that the monovalent cation sensitivity of secondary site cleavage by RNase III is not due to a low salt-induced RNA conformational change. However, spectroscopic data show that Mg^{2+} affects the conformation of the internal loop, suggesting a divalent cation binding site(s) within this region. The Mg^{2+}-dependence of RNase III processing of some substrates may reflect not only a requirement for a divalent cation as a catalytic cofactor, but also a requirement for a local RNA conformation which is divalent cation-stabilized.

2. The anticodon domains of initiator and elongator methionine tRNA have the same loop sequence but different stem sequences. Their loops are known to differ in structure as shown by their susceptibility to cleavage by S1

nuclease. Oligonucleotide analogs were examined by NMR spectroscopy. The structure of the initiator analog was solved by distance geometry and simulated annealing. Its loop is qualitatively the same as the anticodon loops of elongator tRNAs. The nearly identical chemical shifts of the two molecules show that their average structures are nearly identical as well. They do differ in their dynamic behavior: the initiator analog is more rigid, while the elongator analog is more flexible. This explains the difference in nuclease susceptibility, and it argues that the initiation apparatus recognizes the sequences of the stems rather than a structural difference between the loops.

Nuclear Magnetic Resonance Studies of RNA Structure: An RNase III Processing Signal and the Anticodon Stem-Loops of Initiator and Elongator Methionine tRNA

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
David C. Schweisguth

Dissertation Director: Professor Peter B. Moore

May 1997

For Penny

"`... And now, which of these finger-posts ought I to follow, I wonder?'

It was not a very difficult question to answer, as there was only one road through the wood, and the two finger-posts both pointed along it. `I'll settle it,' Alice said to herself, `when the road divides and they point different ways.'

But this did not seem likely to happen. She went on and on, a long way, but wherever the road divided, there were sure to be two finger-posts pointing the same way, one marked `To Tweedledum's House' and the other `To the House of Tweedledee'.

`I do believe,' said Alice at last, `that they live in the *same* house!'"

– Lewis Carroll, Through the Looking-Glass

Foreword

The application of high-resolution nuclear magnetic resonance to RNA was a novelty around 1990 when I began graduate school. It is now an established specialty. I am happy to report that RNA structures solved by NMR are now too numerous for me to list. A new structure is interesting not because it has been solved at all, but because it has something to say about function. This has required spectroscopists to choose projects with increasing care; I hope it has also increased the impact of the field on those outside it.

This dissertation is written in that spirit – I have concentrated on the biological message and described methods only where I differed from standard references and practice. Now, to specifics: The reader is expected to be familiar with Saenger (Saenger 1983), Wüthrich (Wüthrich 1986), and Varani-and-Tinoco (Varani and Tinoco 1991b). I have not re-documented nucleic acid nomenclature or diagrammed the anomeric-aromatic walk. The spectroscopic techniques used herein have been described elsewhere, and it suffices to refer to them. The reader who needs further explication of these basics will find it where I did, in Alex Szewczak's Ph.D. thesis (Szewczak 1994).

Part 1 reproduces a paper (Schweisguth et al. 1994) co-authored with Bhadrani Chelladurai and Prof. Allen Nicholson of Wayne State University. I thank them for an enjoyable collaboration. The material in part 2 has been submitted for publication. This dissertation (in PostScript form), the programs discussed in appendix 2, and the X-PLOR files used to calculate the structure of ITALY2 have been placed in ftp://proton.chem.yale.edu/dcs-dissertation/.

Scores of others provided essential support. I thank my advisor, Peter Moore, my research committee, Profs. Don Crothers and Joan Steitz, my outside reader, Prof. Uttam RajBhandary, the past and present members of the Moore group, and each and every inhabitant of the first floor of Kline Chemistry, which has been a fine place to work. Two people require special mention: Joe Kim pointed out to me the initiator tRNA anticodon stem-loop problem, and Alex Szewczak taught me nearly every technique I used herein. Finally, I thank the Howard Hughes Medical Institute for a predoctoral fellowship.

Dave Schweisguth
September 1996
New Haven, CT

Table of contents

Figures

Tables

Abbreviations

ac^4	N^4-acetyl-
CMCT	cyclohexyl-N'-[2-(N-methylmorpholino)ethyl] carbodiimide *p*-toluene sulfonate
DEPC	diethylpyrocarbonate
DG	distance geometry
DMS	dimethylsulfate
dsRNA	double-stranded RNA
MetRS	methionyl-tRNA synthetase
mnm^5s^2	5-methylaminomethyl-2-thio-
ms^2i^6	2-methylthio-N^6-isopentenyl-
nt	nucleotides
PRT	phosphoribosyl transferase
PTH	peptidyl-tRNA hydrolase
RMS	root mean square
RMSD	RMS deviation
RNase V_1	cobra venom ribonuclease
SA	simulated annealing
t^6	N^6-threonylcarbamoyl-
$tRNA_f^{Met}$	formylmethionine initiator tRNA (prokaryotic)
$tRNA_i^{Met}$	methionine initiator tRNA (eukaryotic)
$tRNA_m^{Met}$	methionine elongator tRNA (prokaryotic or eukaryotic)

Nomenclature

"Prokaryote" as used herein means eubacterial and not archaebacterial. When describing the translational machinery, "eukaryotic" refers to that in the eukaryotic cytoplasm, not to that in mitochondria or chloroplasts.

A phosphorus atom in a polynucleotide, as in the monomer, belongs to the nucleotide on its 3' side.

The names of spins have been pluralized by adding an "s", not an "'s". For example, "H2's" means more than one H2', not more than one H2.

1. An RNase III Processing Signal

1.1 Introduction

Protein recognition of double-stranded RNA is important in many gene regulatory mechanisms (St. Johnston et al. 1992; Haines et al. 1991; Libonati et al. 1980). *Escherichia coli* ribonuclease III (RNase III, E.C.3.1.24) (Robertson et al. 1968) is an endonuclease that specifically hydrolyzes dsRNA (reviewed in (Court 1993)). It cleaves primary rRNA transcripts, releasing the immediate precursors to the 16S and 23S rRNAs. A number of cellular pre-mRNAs are also substrates for RNase III. RNA cleavage occurs at specific sites called RNase III processing signals, which invariably contain approximately two turns of more or less regular dsRNA. Additional sequence and structural elements are present in RNase III processing signals which determine the way they are cleaved, and often have important functional consequences for the metabolic stabilities and translational activities of the transcripts (Court 1993).

The life cycles of a number of coliphages depend on RNase III. The early region of the bacteriophage T7 genome, for example, encodes five RNase III processing signals that occur in the intercistronic regions, and several cleavage sites have been found in transcripts from the T7 late region (Dunn and Studier 1983). RNase III cleavage of the early primary transcript creates mature mono- and dicistronic mRNAs, and also determines their translational activities and stabilities (Dunn and Studier 1983; Dunn 1976; Saito and Richardson 1981). The T7 R1.1 processing signal (figure 1.1.1A), immediately upstream of gene 1.1, is an approximately 60 nt, irregular RNA hairpin, consisting of two dsRNA segments which are separated by an asymmetric internal loop. It contains the sequence information which is necessary and sufficient for accurate in vitro RNase III cleavage (Nicholson et al. 1988; Chelladurai et al. 1991). The R1.1 element is cleaved at a single point

within the internal loop, leaving a short hairpin which protects the upstream RNA segment from 3'-5' exonucleolytic degradation (Dunn and Studier 1983; Panayotatos and Truong 1985). The R1.1 processing signal has been placed in plasmid expression vectors in order to produce transcripts with precisely defined 5' ends, or having prolonged metabolic stabilities (Linn and St. Pierre 1990; Studier et al. 1990; Ziemke and McCarthy 1992).

Although over 30 different RNase III processing signals have been identified, and several studied biochemically and genetically, little is known about their solution conformations. The T7 R1.1 RNA has been studied as a model substrate with which to determine the RNA sequence and structural elements that establish RNase III processing reactivity (Chelladurai et al. 1991; Chelladurai et al. 1993). Determination of its structure may provide information as to why some substrates are cleaved once, while others undergo coordinate double cleavage, and afford insight into the reason for secondary site cleavage and its dependence on salt concentration.

In this report we describe imino proton NMR and optical melting studies of the R1.1 processing signal, and the results of enzymatic and chemical probing experiments using reagents specific for structured or unstructured RNA regions. The data indicate that the R1.1 processing signal is extensively base-paired, and that its internal loop is helical in character but is less stable than a canonical dsRNA element. The conformation of the R1.1 processing signal is largely insensitive to the monovalent salt concentration, suggesting that the influence of salt on secondary site cleavage by RNase III reflects altered interactions between enzyme and substrate rather than changes in RNA structure. In contrast, Mg^{2+} does influence the conformation of the internal loop. Divalent cations may direct the pattern RNase III processing by altering substrate structure, as well as serving as essential catalytic cofactors.

1.2 Methods

1.2.1 Materials

Water was deionized and distilled. Chemicals were reagent grade. Ribonucleoside triphosphates were purchased from Boehringer-Mannheim. [γ-^{32}P]ATP (3000 Ci/mmol) and [α-^{32}P]UTP (3000 Ci/mmol) were purchased from Dupont-NEN. DEPC and DMS were obtained from Aldrich, while RNases U2, T2, and T1 were from GIBCO-BRL, and RNase V1 from Pharmacia. T7 RNA polymerase and RNase III were purified as described from overexpressing bacterial strains (Grodberg and Dunn 1988; Li et al. 1993). The oligodeoxynucleotide transcription templates were synthesized at the Wayne State Macromolecular Analysis Facility, and the fully deprotected forms purified by denaturing gel electrophoresis (Chelladurai et al. 1991). The sequence of the R1.1[LSΔ6] RNA template and the 18 nt promoter oligonucleotide are as follows: 5'-AGGATCATAAAGGCCACTCTTGCGAAT-GACCTTGAGTTCCCTATAGTGAGTCGTATTA-3'; 5'-TAATACGACTCACT ATAG-3'.

1.2.2 RNA synthesis and purification

Preparative-scale synthesis of R1.1[LSΔ6] RNA was carried out by the procedure of Uhlenbeck and coworkers (Milligan et al. 1987) with the modifications described by Wyatt et al. (Wyatt et al. 1991). Transcription reactions (5-10 ml) were prepared in standard buffer (Milligan et al. 1987), which included 4 mM of each ribonucleoside triphosphate, 20 mM Mg^{2+}, T7 RNA polymerase (2500 units/ml) and 55 nM of transcription template annealed to the promoter oligonucleotide. Following incubation for 4 hours at 37°C, the reactions were terminated by adding EDTA (50 mM final concentration), dialyzed against water, then phenol extracted and ethanol precipitated. The crude RNA was resuspended in water and electrophoresed (25 V/cm) in 1.5 mm thick, 15% polyacrylamide gels containing 7 M urea in TBE buffer. The gel region containing the full-length transcript was located by UV-shadowing. Minor amounts of longer RNA species were also observed, representing transcripts with one or two additional nucleotides at the 3' terminus (Milligan et al. 1987). The RNA was extracted from crushed gel bands using 0.5 M ammonium acetate, 1 mM EDTA (pH 7). The aqueous volume was reduced by repeated extraction with 1-butanol, and the RNA ethanol-precipitated and resuspended in H_2O. The yield was determined by UV absorbance (calculated extinction coefficient: 4.45×10^5 M^{-1} cm^{-1}). 5'-^{32}P-labeled R1.1[LSΔ6] RNA was prepared by dephosphorylation of unlabeled transcript with calf alkaline phosphatase, followed by reaction (approximately 50 pmol; 30 minutes at 37°C with T4 polynucleotide kinase (20 units) and [γ-^{32}P] ATP (0.89 µM; 1870 Ci/mmol) in kinase buffer. The ^{32}P-labeled RNA was purified by gel electrophoresis as described above. Base-specific cleavage analysis of ^{32}P-labeled R1.1[LSΔ6] RNA using RNases T1 and U2 was in full agreement with the predicted sequence.

1.2.3 NMR spectroscopy

All NMR experiments were conducted on a sample containing approximately 0.6 µmol of R1.1[LSΔ6] RNA. Buffers contained 4 mM sodium cacodylate and 0.2 mM EDTA, pH 7.5, with 200 mM KCl and/or 5 mM $MgCl_2$ added as described. The RNA was dialyzed extensively against each of the buffers in which it was to be examined and concentrated to approximately 400 µl in a Centricon-3 concentration device (Amicon). 20 µl D_2O (MSD Isotopes) was added to the sample as a lock reference and 1 µl 2.5% dioxane as a chemical shift standard. The chemical shift of dioxane was taken to be 3.741 ppm at all temperatures.

One-dimensional spectra were water-suppressed using the twin-pulse method (Kime and Moore 1983b). Difference NOE experiments were collected in interleaved sets of 64 scans, irradiating each frequency of interest in turn with a 300 ms low-power pulse (Kime and Moore 1983a). Two-dimensional NOESY spectra were collected in phase-sensitive mode with States phase cycling and a 300 ms mixing time. Water was suppressed using a self-refocused 1331 pulse before acquisition (Takegoshi et al. 1990). 2048 complex points were collected in t2 and approximately 300 points were collected in t1. All data were apodized with 30° to 70°-shifted sine bells before Fourier transformation.

One-dimensional NMR experiments were done using a homebuilt 490 MHz spectrometer and NOESYs were collected on a Bruker AM-500, both in Yale's Chemical Instrumentation Center. Data were processed on a Silicon Graphics workstation using Felix (Hare Research).

1.2.4 Optical melting

Aliquots of R1.1[LSΔ6] RNA were dialyzed extensively against each buffer, diluted to concentrations of 0.1, 0.3, 1, 3 and 10 OD, degassed under vacuum, and placed in stoppered quartz cuvettes with path lengths of 1 cm (0.1, 0.3 and 1 OD samples) or 0.1 cm (3 and 10 OD samples). Melting was done in a Varian Cary 1 UV-visible double-beam spectrophotometer. Five samples, one at each concentration, were melted simultaneously. Cuvettes filled with degassed buffer were used as references. The temperature was raised from 5 to 95°C at 0.5°C per minute and the absorbance at 260 nm of each sample recorded every 0.5°C. The temperature was returned to 20°C at the end of each melt and the absorbance checked to ensure that solvent had not been lost by evaporation during the melt. Each melt was done in duplicate.

Melting data were normalized to 1 OD at 20°C and duplicates were averaged. Observed absorbance was modeled as the sum of a baseline (the absorbance of fully stacked RNA) and two sigmoids (the hypochromicities of two independently melting regions), each varying linearly with temperature, using the following equation:

$$A = a_0 + b_0 T + \frac{a_1 + b_1 T}{1 + e^{\frac{\Delta H^\circ_1}{R}\left(\frac{1}{T} - \frac{1}{T_{m,1}}\right)}} + \frac{a_2 + b_2 T}{1 + e^{\frac{\Delta H^\circ_2}{R}\left(\frac{1}{T} - \frac{1}{T_{m,2}}\right)}}$$

A is absorbance, T is temperature, ΔH°_1 and ΔH°_2 are the enthalpies of the two transitions, T_{m1} and T_{m2} are their melting temperatures and a_n and b_n are the zero and first-order coefficients of variation of each term with temperature. This equation was fit to the averaged data using an implementation of the simplex algorithm (Tirado-Rives 1993). Nearly all fits converged to the data within a RMS difference of 10^{-6}; those which did not, due to experimental error, were rejected. Restarting converged fits did not significantly change either the values of the parameters or the goodness of fit. The enthalpies and T_m values were averaged, and the standard deviations calculated. Entropies were calculated as $\Delta H^\circ/T_m$ and free energies at 37°C as $\Delta H^\circ - (310.15\ K \cdot \Delta S^\circ)$. Enthalpies, entropies and energies for the upper and lower stems were summed to obtain the parameters for the entire molecule.

1.2.5 Enzymatic and chemical structure probing

The protocols followed for probing of 5'-^{32}P-labeled R1.1[LSΔ6] RNA using RNase T2 or RNase V1 were based on those described previously (Ehresmann et al. 1987). Chemical probing using DMS or DEPC was carried out essentially as described elsewhere (Peattie and Gilbert 1980). The reaction buffer was 4 mM sodium cacodylate (pH 7.5), 0.2 mM EDTA, supplemented with 200 mM KCl and/or 5 mM $MgCl_2$, as indicated. Enzymatic RNA sequencing reactions were performed using RNase U2 or RNase T1, following the protocol provided by GIBCO-BRL. Reactions were analyzed by electrophoresis in 0.4 mm thick, 15% polyacrylamide, 1.5% bisacrylamide sequencing gels containing 7 M urea and TBE buffer. Autoradiography was carried out using Fuji RX film and intensifying screens.

1.3 Results

1.3.1 Choice of experimental target

R1.1[LSΔ6] RNA , shown in figure 1.1.1B, is a 41-nt derivative of the T7 R1.1 processing signal. The RNA incorporates the wild-type upper stem and asymmetric internal loop, and possesses a lower stem of sufficient size for the RNA to be faithfully processed by RNase III in vitro, albeit with a reduced reactivity compared with the wild type sequence (Chelladurai et al. 1993). Unlike its parent molecule, R1.1[LSΔ6] RNA has helices of different lengths, avoiding ambiguity in assigning the results of NMR spectroscopy and optical melting to the correct parts of the molecule. Its lower stem lacks 5' or 3' single-stranded extensions, reducing the likelihood of intermolecular aggregation (Szewczak et al. 1990).

The RNase III reactivity of R1.1[LSΔ6] RNA was tested in the buffer used in the NMR experiments (4 mM cacodylate, 5 mM Mg^{2+}, 200 mM KCl). The RNA was cleaved singly at the canonical site, at a rate comparable to that observed in the physiologically relevant potassium glutamate-based buffer (Li et al. 1993) (data not shown). Moreover, when KCl was omitted from the reaction buffer cleavage occurred at the expected secondary site (data not shown). We conclude that the cacodylate-based buffer does not perturb R1.1[LSΔ6] RNA structure in a manner that inhibits the proper pattern of primary and secondary site enzymatic processing.

1.3.2 Spectroscopic analysis

Downfield spectra of R1.1[LSΔ6] RNA at 5°C in three different ionic conditions are shown in figure 1.3.1. Each spectrum shows thirteen peaks of roughly stoichiometric intensity, representing G or U imino protons protected from solvent by hydrogen-bonding, and six to ten peaks of lesser intensity, representing imino protons in faster exchange with solvent or with each other. Thus, all or nearly all of the 22 imino protons in R1.1[LSΔ6] RNA are visible and are at least somewhat protected from solvent exchange at 5°C in each ionic condition. There are also several low-intensity but narrow peaks, due either to minor conformations or small amounts (less than 1%) of transcripts with 3' extensions (see Materials and Methods). The chemical shifts of the stoichiometric resonances change relatively little with ionic conditions. Most of the differences between these spectra reflect changes in the weaker resonances.

Spectra collected at temperatures up to 60°C, or 30°C when the solvent contained Mg^{2+} (data not shown), demonstrated that although the imino proton exchange rates increase with temperature, no resonances disappear (except through increased exchange with solvent) and no new ones appear. Moreover, since the chemical shifts change only slightly, the RNA undergoes no gross structural rearrangements as a function of temperature. This result was obtained under all ionic conditions tested.

The imino-imino region of a NOESY of R1.1[LSΔ6] RNA in 200 mM KCl at 5°C is shown in figure 1.3.2. Nearly all of the stoichiometric resonances and some of the weak resonances correlate with one or two others, indicating spatial proximity. Sequences of adjacent imino resonances can thus be identified. NOESYs were also collected at 30°C, where small changes in

chemical shift at the higher temperature helped to resolve overlapping peaks (data not shown). One-dimensional difference NOE experiments (data not shown) were also carried out to confirm these results. Difference experiments were often more sensitive, but two-dimensional experiments provided higher resolution. Comparison between the two was important for confident assignment of NOEs.

The uracil imino proton in an A:U base pair typically shows a strong NOE to the adenine H2 proton. These NOEs were used to identify A:U pairs and align the sequence of adjacent imino protons determined by NOE experiments with that expected from the predicted secondary structure of R1.1[LSΔ6] RNA. The alignment is unambiguous in the upper stem-loop. All thirteen of the stoichiometric imino resonances originate in that part of the molecule. The assignments are summarized in figure 1.3.3. It was not possible to distinguish between partners in the two G:U pairs in the upper stem by their NOEs, as each shows NOEs to its partner and to the imino protons in the base pairs above and below, as well as several NOEs to protons in the amino region. The guanine proton in a G:U pair is typically far upfield of the uracil proton, so peak O is probably a G and peak I a U, but peaks M and N cannot be thus assigned. Each resonance assigned to a uracil imino proton in an A:U base pair shows the expected U imino-AH2 NOE (data not shown). These assignments are supported by data obtained from a related R1.1 RNA derivative having the same upper stem, but with a lower stem containing seven different Watson-Crick base pairs. All of the assigned protons are present in the spectrum of this R1.1 variant, and exhibit the same NOEs (data not shown).

Peak R is assigned to the imino proton of G_{19} in the tetraloop, which is expected to be protected from solvent exchange as discussed below. Two pairs

of resonances, L and Q and N_2 and S, show detectable NOEs to each other but to no other protons. (The N and N_2 peaks overlap completely, but are distinguishable by the chemical shifts of their crosspeaks to other resonances in 0 mM KCl.) They are all broadened and shifted upfield by exchange, and thus most likely located in the loop or lower stem. The imino spectrum of the R1.1 variant mentioned above, in which the lower stem sequence is different, lacks the L and Q resonances, and the L-Q and N_2-S NOEs are not seen (data not shown). One of these pairs might arise from the G:U pair at the base of the lower stem. This would account for their strong mutual NOE, which is detectable despite the broadening of the individual resonances, and the absence of that NOE in the variant molecule, which does not have a G:U pair in the lower stem.

Similar experiments in buffers lacking K^+ or containing only 5 mM Mg^{2+} gave similar results. Base-pairing in the upper stem is insensitive to Mg^{2+} or the monovalent cation concentration, except the base pair represented by the unassigned resonance P, which is evident only in low salt, and the A_9:U_{32} pair (peak D) adjacent to the internal loop, whose imino proton resonance shifts upon the addition of Mg^{2+}.

We conclude that the upper stem-loop of the R1.1 processing signal is stable in all solvent conditions tested. Of the six to ten unassigned resonances, no more than five can come from the four base pairs (one a G:U) in the lower stem. This indicates that at least one of the four imino protons in the internal loop is weakly protected from solvent exchange. Monovalent salt significantly affects no more than one base-pairing interaction in either the internal loop or lower stem. The addition of Mg^{2+} affects the chemical shift of several imino proton resonances in the internal loop and lower stem, but neither creates nor destroys structural features which protect imino protons from solvent exchange.

1.3.3 Optical melting

Figure 1.3.4 shows the absorbance of R1.1[LSΔ6] RNA at 260 nm as a function of temperature in each buffer, differentiated with respect to temperature to highlight the melting transitions. Two transitions are evident in each case, presumably corresponding to the melting of the upper and lower stems. That the transitions are distinct indicates a discontinuity in stacking in the internal loop, and the absence of a third transition suggests that internucleotide interactions in the internal loop either contribute little to meltable hypochromicity, or melt in concert with the upper or lower stem.

Melts were done at concentrations of RNA from 0.1 to 10 OD (0.25 to 25 μM). Melting behavior was independent of concentration in each buffer, showing that R1.1[LSΔ6] RNA is monomeric over this concentration range, and therefore almost certainly monomeric at the concentration used in the NMR experiments, which was less than 100-fold larger than the highest concentration melted.

Table 1.3.1 presents the thermodynamic parameters derived from these melting experiments, calculated assuming the two transitions to be independent. The transition with the larger enthalpies and entropies of melting is assigned to the upper stem, on the assumption that the magnitude of these parameters should increase with the size of the base-paired region (Searle and Williams 1993). In support of this, the related R1.1 molecule (see above) has two melting transitions, each having a much higher T_m than that

Lower stem

Solvent	T_m (°C)	ΔH° (kcal/mol)	ΔS° (cal/mol K)	$\Delta G°_{37}$ (kcal/mol)
0 mM KCl	39.4 ± 6.5	48.3 ± 0.7	154 ± 2	0.5 ± 1.0
200 mM KCl	55.1 ± 3.9	33.7 ± 2.3	103 ± 7	1.8 ± 3.2
200 mM KCl + 5 mM $MgCl_2$	69.1 ± 0.2	29.6 ± 1.5	87 ± 4	2.0 ± 2.0

Upper stem

Solvent	T_m (°C)	ΔH° (kcal/mol)	ΔS° (cal/mol K)	$\Delta G°_{37}$ (kcal/mol)
0 mM KCl	51.4 ± 0.6	70.9 ± 5.8	218 ± 18	3.3 ± 8.0
200 mM KCl	76.5 ± 0.6	104.3 ± 14.7	298 ± 42	11.9 ± 19.7
200 mM KCl + 5 mM $MgCl_2$	80.4 ± 0.4	125.3 ± 6.5	317 ± 16	27.0 ± 6.5

Total

Solvent	T_m (°C)	ΔH° (kcal/mol)	ΔS° (cal/mol K)	$\Delta G°_{37}$ (kcal/mol)
0 mM KCl		119.2 ± 5.8	372 ± 17	3.8 ± 8.1
200 mM KCl		138.0 ± 14.9	401 ± 43	13.7 ± 20.0
200 mM KCl + 5 mM $MgCl_2$		154.9 ± 6.6	404 ± 17	29.0 ± 6.8

Table 1.3.1: Thermodynamic parameters of R1.1[LSΔ6] RNA melting in the indicated solvents. Errors are standard deviations.

for the transition which we attribute to the lower stem of R1.1[LSΔ6] RNA (data not shown).

The upper stem melts well above 37°C in all conditions, but the lower stem is mostly dissociated at 37°C in 0 mM KCl. Melting temperatures and free energies of melting are higher in the presence of potassium and magnesium, which counteract charge repulsion between backbone phosphates. Interestingly, the enthalpies and entropies of melting decrease with increasing ionic strength in the lower stem, but increase in the upper stem. This may indicate that the bases in the internal loop stack on the lower stem at low ionic strength, but stack on the upper stem at high ionic strength.

1.3.4 Chemical and enzymatic structure probing

To further define conformational features of the R1.1 processing signal, chemical and enzymatic structure probing was performed on 5' ^{32}P-labeled R1.1[LSΔ6] RNA. To correlate the results with the NMR data, the structure probing experiments were carried out at 30°C in the presence or absence of 200 mM KCl or 5 mM $MgCl_2$ in pH 7.5 cacodylate buffer. The enzymes were selected as complementary probes, as RNase T2 cleaves unstructured RNA regions (Ehresmann et al. 1987), while RNase V1 catalyzes the hydrolysis of helical or double-helical RNA (Lowman and Draper 1986). The chemical reactivity of R1.1[LSΔ6] RNA was examined using DMS and DEPC. DMS probes the steric accessibility and electronic environment of guanine N7 atoms (Ehresmann et al. 1987; Peattie and Gilbert 1980), while DEPC examines adenine N7 atoms by the same criteria, with an additional difference being that DEPC cannot modify adenines in a regular double helix (Ehresmann et al. 1987).

The results of representative experiments are shown in figure 1.3.5 and summarized in figure 1.3.6. At 200 mM KCl, essentially the only RNase T2-sensitive region in R1.1[LSΔ6] RNA is the GCAA tetraloop (G_{19},C_{20},A_{21}>C_{18}>U_{17}) (figure 1.3.5A, lanes 3 and 4; and figure 1.3.5B, lanes 10 and 11). This result supports the spectroscopic evidence for base-paired upper and lower stems. A minor amount of RNase T2 cleavage occurs in the internal loop at the 3' end-proximal side (U_{33}>U_{32},A_{34}) (figure 1.3.5A, lanes 3 and 4), suggesting some deviation from a stable structure in this region. Omitting KCl lowers the overall RNase T2 reactivity of the RNA, but increases the relative reactivity of the internal loop compared to the tetraloop (data not shown), suggesting that low salt destabilizes the internal loop structure (see also below). The presence of Mg^{2+} has no effect on the RNase T2 reactivity pattern in high salt (data not shown), and omitting Mg^{2+} in low salt buffer inhibits all RNase T2 cleavage. However, RNase T2 is able to cleave the top of the internal loop with low efficiency in low-salt with Mg^{2+}, and under the same conditions RNase III is able to cleave a second time in the internal loop. If these differences in enzymatic reactivity result solely from monovalent ion-induced changes in RNA structure, those changes must be quite subtle and largely undetectable using these techniques.

As anticipated, R1.1[LSΔ6] RNA is sensitive to RNase V1, but the cleavage pattern is quite selective. In 200 mM KCl, 5 mM $MgCl_2$ the most reactive region is the 5' side of the upper stem, with A_{15} as the most reactive residue. A lesser amount of cleavage occurs at the flanking nucleotides U_{13}, C_{14} and U_{16} (figure 1.3.5A, lanes 6 and 7). The reactivity of these sites is consistent with the double-helical nature of the upper stem. Some reactivity is also observed at several nucleotides centered at A_{10}, which is also located in the upper stem. Interestingly, the 3' end-proximal segment of the internal

loop is also cleaved by RNase V1. Specifically, a significant amount of cleavage occurs at U_{35}, which is also the primary RNase III cleavage site (figure 1.3.5A, lanes 6 and 7). Additional cleavages are observed at flanking residues A_{34} and G_{36}, as well as C_{30}, U_{31} and U_{32}, indicating that this portion of the internal loop possesses (or has access to) a helical structure. There is no significant RNase V1 reactivity of the sequence between A_{37} and the 3' end (figure 1.3.5A), or the sequence between the 5' end and U_8 (data not shown). The lack of RNase V1 reactivity of these segments does not rule out the double-stranded nature of the lower stem, as nonuniform reactivities of dsRNA towards RNase V1 have been observed elsewhere (Lockard and Kumar 1981; Auron et al. 1982), and reflect the specifics of RNase V1 interaction with helical RNA. Moreover, both of these segments are resistant to RNase T2 (see above), supporting the double-helical nature of this region. Finally, RNase V1 cleavage of R1.1[LSΔ6] RNA is greatly inhibited at low salt (data not shown), and omitting Mg^{2+} abolishes all cleavage, reflecting the divalent cation requirement of RNase V1 (Lowman and Draper 1986).

The DMS reactivity pattern of R1.1[LSΔ6] RNA at 30°C in 200 mM KCl indicates that all guanine N7 atoms are reactive, with no guanine residue notably more reactive than the others (the methylation sensitivity of G_1, G_2 and G_3 could not be definitively assessed, due to the compressed gel electrophoretic mobilities of the short RNA products). Moreover, the DMS reactivity pattern is insensitive to KCl or Mg^{2+} (figure 1.3.5A, lanes 9-12). Specifically, we observe that the N7 of G_{36} in the internal loop is no more reactive than the other guanine N7 atoms, indicating that G_{36} is not looped out or otherwise solvent exposed. Interestingly, G_{19} in the tetraloop is significantly less reactive than the other guanine residues, and the relative amount of DMS methylation of this residue does not increase at 90°C (data not shown). The reduced reactivity may reflect the local electronic and steric environment of G19 within the highly structured context of the GCAA tetraloop (Heus and Pardi 1991b; SantaLucia et al. 1992). The DEPC pattern indicates that A_{21} and A_{22} in the tetraloop are the most reactive residues (figure 1.3.5B, lanes 2-5). A smaller yet significant amount of reaction is observed at A_{34} and A_{37} (A_{34}>A_{37}), whose DEPC sensitivities are enhanced in low salt. The low salt reactivity enhancement is not seen with the other adenines in the upper stem-loop. The DEPC sensitivities of A_3 and A_4 could not be definitively assessed, due to the small size of the reaction products and the differing mobilities of enzyme and DEPC-cleaved fragments (see also the legend to figure 1.3.5). The overall DEPC reactivity pattern is insensitive to Mg^{2+} in either the absence or presence of 200 mM KCl (figure 1.3.5B, lanes 2-5). The lack of reactivity of the adenine residues in the upper stem is evidence for their existence within an A-form double helix (Ehresmann et al. 1987). Minor DEPC reactivity was observed with non-adenine residues in the tetraloop and internal loop. The DEPC sensitivity of non-adenine nucleotides in non-double-helical regions has been noted elsewhere (Ehresmann et al. 1987). This particular aspect of the DEPC reactivity pattern provides further evidence for the lack of a canonical double helical structure in the internal loop.

1.4 Discussion

1.4.1 The secondary structure of R1.1[LSΔ6] RNA

The continuous chain of NOE connectivities and susceptibility to RNase V1 cleavage clearly show that the upper stem of R1.1[LSΔ6] RNA is an A form double helix, as predicted by sequence comparison (see (Chelladurai et al. 1991) and references therein). The data for the lower stem are less clear, as the imino resonances are too broadened by exchange to give detectable NOEs. However, given (i) the detectability (if not complete assignability) of all but one to three of the twenty-two imino protons and (ii) the RNase T2 resistance of nucleotides near the 5' and 3' ends, we argue that the R1.1[LSΔ6] RNA possesses a double-helical lower stem. Also, it is possible that either the L-Q or N_2-S NOE arises from the terminal G:U pair in the lower stem.

The monovalent salt concentration has a minor effect on the structure of R1.1[LSΔ6] RNA, as only one imino resonance (P) undergoes a significant shift. The relative chemical shifts of resonances J and K also change slightly as a function of salt concentration, but as the chain of NOEs in the upper stem persists this is likely to represent only a minor structural perturbation. The salt-dependent, opposing trends in enthalpy and entropy of upper and lower stem melting may indicate a conformational change in the internal loop (see also below). If this is the case, it must reflect a relatively small quantitative structural change, rather than a qualitative one. The increase in T_m values and free energies of melting with increasing ionic strength may influence the energetics of interaction of RNase III with its processing substrates. At physiological salt concentrations RNase III cleaves only primary sites, while at low salt secondary sites are also recognized, and otherwise unreactive RNAs are cleaved (Dunn 1976; Chelladurai et al. 1991; Chelladurai et al. 1993). Given the salt-insensitivity of R1.1 RNA secondary structure, our results suggest that secondary site cleavage by RNase III results from an altered enzyme-substrate interaction, rather than from a salt-induced change in RNA conformation.

The R1.1 tetraloop, which is distant from the RNase III cleavage site and irrelevant to recognition (Chelladurai et al. 1993), appears similar in conformation to a tetraloop with the same loop sequence and closing base pair (Heus and Pardi 1991b). The G_{19} imino proton of the R1.1 tetraloop exchanges slowly with solvent, as was observed with the corresponding guanine of the model tetraloop. The GCAA tetraloop guanine imino proton is not predicted to hydrogen-bond to another nucleotide, but it may be protected from solvent exchange by hydrogen binding to a water molecule (SantaLucia et al. 1992). The R1.1 tetraloop is cleaved by RNase T2 at G_{19}, C_{20}, $A_{21} > C_{18} > U_{17}$, and A_{21} and A_{22} are the most DEPC-reactive residues. Although the NMR analysis suggests that these adenines (and in particular the N6 protons) are part of a network of hydrogen bonds, these bonds have small free energies of formation (SantaLucia et al. 1992), and would not be expected to significantly affect the chemical or enzymatic modification reactions.

1.4.2 The R1.1 internal loop

The R1.1 internal loop is asymmetric, and cannot be a canonical double helix. A lack of a rigid helical structure for this local structure is suggested by (i) the RNase T2 sensitivity of the 3' end-proximal portion of the internal loop, (ii) the DEPC-sensitivity of nucleotides within the internal loop, (iii) the absence of measurable protection of several of the four imino protons in the internal loop from exchange with solvent and (iv) the separate melting transitions of the upper and lower stems. The exchange-broadening of the imino protons in the internal loop and the lower stem is also evidence of flexibility. This flexibility may be responsible for the UV light-induced covalent crosslinking of the two RNase T1-resistant oligonucleotides which encompass the R1.1 internal loop (Robertson 1990). The UV reactivity pattern may suggest the presence of interstrand base-stacking in the R1.1 internal loop, which occurs in 5S RNA loop E, and is responsible for the UV sensitivity of this structure (Branch et al. 1985; Wimberly et al. 1993).

However, RNase III efficiently cleaves Watson-Crick base-paired dsRNA (Court 1993), so one might expect the internal loop to possess dsRNA character. Indeed, RNase V1 cleavage of the U_{35}-G_{36} phosphodiester bond (which is also the primary RNase III processing site) indicates that this portion of the R1.1 internal loop is engaged in (or has access to) a helical conformation. A similar coincidence of RNase V1 and RNase III cleavage sites has also been observed in an RNase III processing signal of phage λ (Steege et al. 1987). The unexceptional methylation sensitivity of the N7 atom of G_{36} suggests that this residue is not solvent exposed. This can be contrasted with the strong chemical reactivity of the conserved guanine in the 5S rRNA internal loop E (Romaniuk et al. 1988).

Similar internal loops are present in a number of other T7 and T3 RNase III processing signals (Court 1993; Dunn and Studier 1983). Apparently this motif was selected to optimize efficient single cleavage of the phage transcripts in vivo. Other types of asymmetric internal loops in unrelated phage and cellular substrates also enforce single enzymatic cleavage (Court 1993; Dunn and Studier 1983). It remains to be determined how processing signals with internal loops of different sequences and structures interact with RNase III to yield a similar outcome.

1.4.3 The effect of magnesium

Magnesium is absolutely required for RNase III processing, so its influence on substrate structure is of particular interest. The chemical and enzymatic reactivity of R1.1 RNA is largely insensitive to magnesium, but magnesium does shift the resonances of a number of unassigned imino protons, several of which are within the internal loop, and it also affects the imino proton of the A_9:U_{32} pair. It does not so affect the imino protons in the upper stem or tetraloop, so its effect has some specificity. The spectroscopic changes are unlikely to reflect a major conformational change in the internal loop, however, as magnesium only perturbs these resonances, instead of abolishing them or creating new ones (as seen, for example, in the binding of argininamide to the bulge loop of TAR RNA (Puglisi et al. 1992)). Unfortunately, no further structural information could be gained, as no NOEs between any of these resonances were detected.

Although perhaps not relevant to every processing signal, magnesium may stabilize RNA tertiary structure important for RNase III reactivity, in addition to being an essential cofactor in the chemical step. In this regard, we have found that the interaction of R1.1 RNA with a catalytically inactive mutant of RNase III is significantly stabilized in the presence of magnesium (H. Li and A.W.N., unpublished experiments). Magnesium may therefore play an important role in the determining the reactivity of other RNase III substrates containing higher-order structural features.

1.4.4 Comparison with the thermodynamically predicted secondary structure

The RNA structure prediction program MFOLD (Zuker 1989) predicts a lowest-energy secondary structure for R1.1[LSΔ6] RNA similar to that presented here, but with U_7-A_9 base-paired to U_{35}-A_{37}. This structure is incompatible with our assignment of imino resonance D to the A_9:U_{32} base pair. While U_{35} rather than U_{32} might pair with A_9, we do not observe the other computer-predicted pairs. Moreover, the computer-predicted conformation has bases U_{32}-A_{34} in a bulge loop, which would coincide with several RNase V1 cleavage sites. The lowest-energy structure which contains the full-length upper stem has a similar bulge loop, but the next higher-energy structure pairs C_6 with G_{36}, and has no completely unopposed bulge loops. This internal loop pairing scheme is supported by the unexceptional DMS reactivity of G_{36}. This R1.1[LSΔ6] RNA structure differs from the lowest-energy conformation by less than 3 kcal/mol, which is insignificant since MFOLD neglects most non Watson-Crick interactions other than G:U base pairs, and uses thermodynamic data collected in ionic conditions (1 M NaCl) different from those used here.

For the same reason, we can only compare in a qualitative fashion our thermodynamic parameters with those calculated by MFOLD. The parameters calculated for what we argue is the most likely secondary structure of R1.1 RNA are very similar to those we observed in either salt condition (data not shown). We would expect the melting temperature to rise at higher ionic strength, but fall if additional stabilizing non Watson-Crick interactions were neglected, and perhaps the two effects cancel.

1.4.5 The dsRNA mimicry model

Finally, the NMR and structure probing data presented here, as well as a previous mutational analysis (Chelladurai et al. 1991), argue against the proposal of a more complicated internal loop folding motif (Robertson and Barany 1978). Specifically, if the "dsRNA mimicry" tertiary folding were present, we would expect an imino proton resonance from the proposed A_5:U_{33} pair, with accompanying NOEs to adjacent base-pairs in the upper and lower stem, and these were not observed.

2. The Anticodon Stem-Loops of Initiator and Elongator Methionine tRNA

2.1 Introduction

2.1.1 The problem of initiator-elongator discrimination

The first amino acid in every polypeptide chain is a methionine. This methionine is delivered to the translation apparatus by a special initiator tRNA. Methionines other than the first are carried by a separate methionine elongator tRNA. Initial and internal methionines are coded for by the same codon, AUG, and initiator and elongator methionine tRNAs recognize that codon with the same anticodon sequence, CAU. (In prokaryotes, the initiation codon may differ from AUG by one nucleotide in special cases, but it always specifies methionine and is recognized by the same initiator tRNA (Gualerzi and Pon 1990).) Both methionine tRNAs are aminoacylated by the same methionyl-tRNA synthetase, which recognizes the same nucleotides in each tRNA (see below). What one would think are the two most salient features of these tRNAs are thus identical, and yet they have completely non-overlapping functions. How the translation apparatus recognizes them as different is the problem of initiator-elongator discrimination. Although it is an old problem by the standards of molecular biology, many of its aspects remain unresolved. This dissertation discusses one facet of initiator-elongator discrimination, the role of the sequence of the anticodon stem.

2.1.2 Translation initiation

Translation initiation is the process by which initiator tRNA and mRNA are assembled onto the ribosome (Gualerzi and Pon 1990; Voorma et al. 1994). The first stage of translation initiation is the formation of the

30S/40S complex (30S in prokaryotes, 40S in eukaryotes), shown in figure 2.1.1. In prokaryotes, initiation factors bind to the 30S subunit first, fMet-$tRNA_f^{Met}$ and mRNA bind in random order, and a conformational change occurs in which the codon recognizes the anticodon. In eukaryotes, initiation factors, an eIF-2·GTP·Met-$tRNA_i^{Met}$ complex and mRNA bind to the 40S subunit in that order, the 40S subunit moves from the 5' end of the mRNA to the initiation codon, and codon-anticodon recognition takes place.

By this point all of the recognition events which distinguish initiator from elongator tRNA have occurred and the initiator tRNA is in the P site of the ribosome. From the ribosome's point of view this is the fundamental purpose of initiation: it places the initiator tRNA in the P site directly, without it first having to bind to the A site and undergo peptidyl transfer and translocation. In subsequent steps, not shown, initiation factors dissociate and the large ribosomal subunit binds to form the 70S/80S (70S in prokaryotes, 80S in eukaryotes) initiation complex. Having a peptidyl-tRNA in the P site, the 70S/80S complex is equivalent to a ribosome which has just completed a round of elongation, and it moves into the elongation cycle in the same way (and out of our present scope of interest).

Initiation factors enforce the specificity of translational initiation. (They have many other functions as well, described in the reviews referenced above.) In the absence of initiation factors any tRNA can initiate translation if a cognate codon is near the Shine-Dalgarno sequence. The prokaryotic initiation factors IF1, IF2 and IF3 restrict initiation to $tRNA_f^{Met}$ (Hartz et al. 1989) at an AUG codon (Sussman et al. 1996). IF2 binds to the acceptor stem of fMet-$tRNA_f^{Met}$ (Petersen et al. 1979; Petersen et al. 1981; Wakao et al. 1989), and rejects tRNAs without a formyl group (Sundari et al. 1976). IF3 causes the initiation complex to reject tRNAs whose anticodon domains do not have

the unique features of $tRNA_f^{Met}$ (Hartz et al. 1990). IF3 does not bind tRNA, although it does bind to 16S ribosomal RNA. IF1 stimulates the activities of IF2 and IF3 through an unknown mechanism. Initiation in eukaryotes is much more complicated. It suffices to mention here that it is the initiation factor eIF-2 (which binds Met-$tRNA_i^{Met}$ to form a complex which then binds to the 40S subunit) which discriminates between initiator and elongator tRNAs, and which supervises the selection of the AUG initiation codon.

2.1.3 The purpose of initiator-elongator discrimination

It is interesting to speculate on why the cell discriminates between molecules which would otherwise function identically. There are two sides to this coin: elongator tRNAs are excluded from initiation, and initiator tRNAs are excluded from elongation. There must be benefits to both.

Why might the cell want to exclude elongator tRNAs from initiation? Doing so incidentally excludes abortive peptidyl-tRNAs from the ribosome, but the elongation apparatus is able to solve the same problem in another way simply by requiring a free α-amino group (Schulman et al. 1974). Alternatively, initiator identity might be an adjunct to methionine identity: given that the initiation codon always codes for methionine, excluding all elongator tRNAs from initiation ensures that the only tRNAs which can enter initiation (that is, initiator methionine tRNAs) will have a methionine anticodon. Perhaps the difficulty of evicting a noncognate anticodon from the initiation complex justifies the cost of excluding it at the door.[1]

---

[1]Another topic of speculation is the choice of methionine as the universal initiating amino acid. It is removed from most proteins after translation (Waller 1963), so cannot have a general structural role. The initiation apparatus (RajBhandary and Chow 1995) and the N-end rule degradation pathway (Meinnel et al. 1993a) probably adapted to this choice rather than making it. Use as an indicator of cellular energy balance (Meinnel et al. 1993a) does not seem to

Why exclude initiator tRNAs from elongation? In most modern prokaryotes the aminoacyl-initiator tRNA is formylated; if it were able to enter the elongation cycle its blocked α-amino group would prevent peptidyl transfer. As eukaryotes are able to initiate perfectly well without the formyl group, this only begs the question of why prokaryotic initiator tRNA is formylated – the formyl group is probably a marker for initiator tRNA which arose after the segregation of initiators and elongators. Note also that the elongation apparatus is able to exclude peptidyl-elongator tRNAs, which, like fMet-$tRNA_f^{Met}$, have a blocked α-amino. Similarly, if Met-$tRNA_f^{Met}$ were able to elongate, abortive translation would result in peptidyl-$tRNA_f^{Met}$ which would not be cleaved by peptidyl-tRNA hydrolase (Schulman and Pelka 1975; Dutka et al. 1993), but if there were no formyl group peptidyl-tRNA hydrolase would not need to exclude $tRNA_f^{Met}$. $tRNA_f^{Met}$ can recognize alternate codons by wobble in the first codon position, which would be undesirable in an elongator methionine tRNA, but since eukaryotic initiator tRNAs do not undergo first-position wobble it cannot have spurred the divergence of initiator and elongator tRNAs.

Again, initiator identity might really be methionine identity; perhaps ur-proteins used methionine only for an initial amino acid. Since modern proteins contain internal methionines, this seems unlikely. The amount of $tRNA_f^{Met}$ needed for initiation is comparable to the amount of $tRNA_m^{Met}$ needed for elongation (Bremer and Dennis 1987), so other elongator tRNAs would be able to compete for EF-Tu with a $tRNA_f^{Met}$ that could bind it. Some feature of initiator tRNAs might cause them to be inefficient in elongation,

---

mandate such strict conservation. Perhaps proteins begin with a methionine *codon* simply to reduce possible start sites 64-fold, and the methionine side chain is frozen chance. If so, this choice of initiation codon must have preceded the development of the Shine-Dalgarno sequence and the divergence of initiator from elongator methionine tRNA. The latter seems more likely than the divergence of elongator methionine tRNA from initiator methionine tRNA.

but the only feature shared by prokaryotic and eukaryotic initiator tRNAs that is not also found in their elongator methionine tRNAs is the GGG:CCC sequence in the anticodon stem (see below), and it is present in other elongator tRNAs in some species. In *E. coli*, a $tRNA_f^{Met}$ mutant which can elongate does so only 1.5 times more efficiently when its GGG:CCC sequence is ablated (Seong and RajBhandary 1987b), a $tRNA^{Trp}$ mutant with a GGG:CCC sequence is as efficient an amber suppressor as its parent (Yarus et al. 1986) and a elongating, tryptophan-inserting mutant of $tRNA_f^{Met}$ is fully 50% as efficient as $tRNA^{Trp}$ itself (Pak et al. 1992). Further, it seems unlikely that the cell would (or could) evolve a marker for excluding elongators from initiation that was so detrimental to elongation that initiators would then need to be excluded from elongation.

The most likely reason for excluding initiator tRNA from elongation lies in the sizes of the pools of factors with which it must interact. There are roughly as many molecules of each initiation factor as of $tRNA_f^{Met}$ in *E. coli*, but about twenty times as many molecules of EF-Tu (Jakubowski and Goldman 1984; Bremer and Dennis 1987). If initiator tRNA could bind to EF-Tu with the same affinity as it does to initiation complexes, it would be sequestered by the much larger pool of EF-Tu and unavailable for initiation. A yeast strain without $tRNA_m^{Met}$ or phosphoribosyl transferase (Åström et al. 1996) is viable, showing that $tRNA_i^{Met}$ without the A64 phosphoribosyl modification (see below) is competent in both initiation and elongation, but it grows slowly. This could be because the supply of $tRNA_i^{Met}$ is sequestered by EF-Tu, because $tRNA_i^{Met}$ without the phosphoribosyl modification functions less well in initiation, or both.

2.1.4 Recognition elements in initiator tRNAs

Initiator tRNAs may be the most information-dense molecules in nature. In the course of their biosynthesis and of initiation they interact with a host of RNAs and proteins, and they must be excluded from another host of interactions in which elongator tRNAs take part. Figure 2.1.2 shows the sequence elements which admit them to or exclude them from these interactions and the molecules which recognize or reject them.

Initiator tRNA must be aminoacylated and, in prokaryotes, formylated, before it can enter the initiation complex. Methionyl-tRNA synthetase in both prokaryotes and eukaryotes, like many tRNA synthetases, chiefly recognizes the CAU methionine anticodon (Meinnel et al. 1990; Meinnel et al. 1992). However, acceptor stem nucleotides are known to contribute as well in prokaryotes (Meinnel et al. 1993b; Senger et al. 1992); the equivalent experiments have not been done in eukaryotes. While the prokaryotic synthetase can aminoacylate a "minihelix" (the acceptor-TψC domain) (Martinis and Schimmel 1993), the eukaryotic synthetase requires most of the tRNA structure to be present (Senger et al. 1995). A comparison of *E. coli* and rabbit synthetases (Senger et al. 1995) indicated that the former requires a 7-nt anticodon loop and the latter 9 (what would be the last pair of the anticodon stem is a ψ-ψ mismatch in eukaryotic $tRNA_m^{Met}$s), but the latter requirement cannot be absolute since rabbit MetRS must also aminoacylate $tRNA_i^{Met}$. In any case, the same synthetase aminoacylates both initiator and elongator tRNAs in a given organism, so methionine identity is a separate problem from initiator-elongator discrimination. However, in prokaryotes, Met-$tRNA_f^{Met}$, but not Met-$tRNA_m^{Met}$, is formylated to become fMet-$tRNA_f^{Met}$ (thus the "f" designation of the tRNA) by methionyl-tRNA transformylase;

this enzyme recognizes sequences at the end of the acceptor stem and an A11-U24 pair in the D stem which are not present in elongator tRNAs (Lee et al. 1991; Guillon et al. 1992).

The prokaryotic initiation factor IF2 prefers, but does not require, initiator tRNA to be aminoacylated with methionine (Varshney and RajBhandary 1992). However, it does require the α-amino group of the amino acid to be formylated (Sundari et al. 1976); recognition elements for transformylase are thus indirectly recognition elements for initiation. IF2 does not recognize tRNA sequence directly. IF3 recognizes a stack of three G:C pairs in the anticodon stem (Hartz et al. 1989; Hartz et al. 1990). The prokaryotic initiation complex can form without initiation factors, but it does not then require these features of the initiator tRNA. Initiation factors may recognize these elements directly, or may affect other parts of the initiation complex so that these elements are recognized – without detailed structural information, the two possibilities are indistinguishable.

The eukaryotic initiation factor eIF-2 does require a methionine (Wagner et al. 1984). Initiation in eukaryotes requires nucleotides in the acceptor stem, the TψC loop, and the anticodon stem (Drabkin et al. 1993), the latter being the same three G:C pairs found in prokaryotes. These might be recognized by eIF-2 or the ribosomal P site. The A1:U72 base pair found in eukaryotic $tRNA_i^{Met}$ but not in eukaryotic elongator tRNAs has been shown to be required for binding to eIF-2 (Farrugio et al. 1996).

Initiator tRNAs have sequence elements which are rejected by molecules which interact with elongator tRNAs. EF-Tu requires its substrate tRNA to have a free α-amino group, so rejects the formyl group of fMet-$tRNA_f^{Met}$ (Schulman et al. 1974). EF-Tu also requires bases 1 and 72 to be a Watson-Crick base pair, as they are in all elongator tRNAs; prokaryotic

$tRNA_f^{Met}$ has a C-A mismatch at the end of the acceptor stem, so it is rejected by EF-Tu (Schulman et al. 1974). Economically, the C1-A72 mismatch also prevents fMet-$tRNA_f^{Met}$ from interacting with peptidyl-tRNA hydrolase, an enzyme which hydrolyzes the aminoacyl linkage of tRNAs with blocked amino termini (Schulman and Pelka 1975; Dutka et al. 1993). Its usual function is to dispose of peptidyl-tRNAs which arise from abortive translation, but it would degrade fMet-$tRNA_f^{Met}$ as well if not prevented from doing so by this rejection element.

eEF-1α, the eukaryotic equivalent of EF-Tu, is prevented from binding $tRNA_i^{Met}$ by a completely different rejection signal in plants and fungi, a phosphoribosyl modification of A64 (Kiesewetter et al. 1990). It seems reasonable that this modification excludes $tRNA_i^{Met}$ from elongation by interfering sterically with binding of elongation factors. Phosphoribosyl transferase, which modifies A64, itself recognizes the U50:A64 base pair (not found in elongator tRNAs) and nucleotides at the base of the TψC loop (Åström and Byström 1994).[2] There is a unique three-dimensional structure in the elbow region of eukaryotic initiator tRNAs (Wagner et al. 1989) which is not present in eukaryotic elongator tRNAs (or in any prokaryotic tRNAs). Although it was proposed to exclude initiator tRNAs from elongation, it was shown not do so (Kiesewetter et al. 1990), and its function is otherwise unknown. The specificity of eukaryotic peptidyl-tRNA hydrolase is not well understood, but since eukaryotic initiator tRNAs are not formylated it need not specially exclude them. Eukaryotic PTH is able to hydrolyze prokaryotic fMet-tRNAfMet (Wagner et al. 1989), so does not reject the GGG:CCC

[2] Interestingly, nucleotide 64 in prokaryotic selenocysteine tRNA is part of an element which prevents it from binding to EF-Tu (Rudinger et al. 1996). The similar positions of these antideterminants may reflect similarities in prokaryotic and eukaryotic elongation factors, or perhaps just the exposed position of nucleotide 64 in the universally conserved global fold of tRNA.

sequence which is found in the anticodon stem of both prokaryotic and eukaryotic initiator tRNAs.

A distinctive feature whose purpose is not known is the replacement of the otherwise universal U33 by a C in vertebrate, insect and plant initiator tRNAs. Wheat germ $tRNA_i^{Met}$ can elongate if the phosphoribosyl modification is removed from its A64 (Kiesewetter et al. 1990), so the purpose of C33 is not to exclude $tRNA_i^{Met}$ from elongation. Yeast $tRNA_i^{Met}$ can initiate in a rabbit in vitro translation system (Wagner et al. 1989), so C33 is not required even in species where it is present. Human $tRNA_i^{Met}$ functions in yeast (Francis and RajBhandary 1990), so C33 has no properties which interfere with translational machinery which expects U33. While the conservation of U33 or C33 within a given phylum (Sprinzl et al. 1996) suggests that there is some functional difference between them, it must be very small.

2.1.5 The G:C stack and anticodon loop conformation

The stack of three G:C base pairs (hereafter the "G:C stack") at the base of the anticodon stem is the only initiator identity element present in both prokaryotic and eukaryotic initiator tRNAs. It is found in most archaeal (Kuchino et al. 1982) and organellar initiator tRNAs as well; virtually all of those species which do not have three G:C pairs have two (Sprinzl et al. 1996). Some organellar initiator tRNAs also function as elongators, even though they have the G:C stack. It is present in only a few elongator tRNAs as such, 21 of 2699 surveyed, about half what would be expected in a random distribution of Watson-Crick base pairs. It is more widely distributed, and thus evolutionarily older, than any other initiator identity element; if initiator and elongator methionine tRNAs diverged from a single ancestor, it was the first point of divergence. (This is true whether initiator tRNAs arose by gaining the G:C stack, or whether elongator methionine tRNAs arose, from an ancestor which both initiated and elongated, by losing the G:C stack.) It is also old in terms of human study, having been identified as an initiator identity element in 1969 (Dube et al. 1969).

The G:C stack has not been noted to have any special function in other RNAs. Interestingly, GGG:CCC is the single most common (651 of 2699) sequence in the three loop-closing base pairs of the TψC stems of tRNAs; CCC:GGG is unknown (Sprinzl et al. 1996). GGG:CCC occurs with about the expected frequency in the base pairs which close 7-nucleotide loops in small-subunit ribosomal RNAs; interestingly, again, CCC:GGG is almost never seen (Neefs et al. 1993). There is a notable instance in the L4 binding region of the large-subunit ribosomal RNA of *E. coli*: a 7-nucleotide loop closed by GGG:CCC. The last two nucleotides of this loop are predicted by phylogenetic comparison to interact with a nearby single-stranded region (Gutell and Woese 1990); doing so in a Watson-Crick manner would preclude a tRNA-like loop structure.

A recognition element in a nucleic acid or protein can distinguish itself from non-elements in two ways: by presenting a unique set of side chains in the context of an unremarkable structure, or by taking on (or causing residues with which it interacts to take on) an unusual structure specific to its sequence. In this case there is evidence that the unique sequence of the stem confers a unique conformation on the loop. The Rich group found that the anticodon loops of initiator tRNAs from several species are cleaved by S1 nuclease in a different way from those of methionine and other elongator tRNAs (Wrede and Rich 1979; Wrede et al. 1979). As shown in figure 2.1.3,

initiator tRNAs are cleaved after positions 34 and 35, while elongator tRNAs are cleaved after positions 33-36. Because the anticodon loops of initiator and elongator methionine tRNAs have the same sequence, the cleavage patterns must arise from different conformations of the same nucleotides which are dictated by the adjacent stem sequences. Because S1 nuclease preferentially cleaves single-stranded nucleic acid (Harada and Dahlberg 1975), the reduced amount of S1 cleavage (hereafter "S1 resistance") in the initiator loop was thought to reflect a more compact or stable structure. S1 resistance was present in a nucleolytic fragment of yeast $tRNA_i^{Met}$ lacking the TψC arm, supporting the idea that the anticodon stem, rather than some other part of the tRNA or its global fold, was its source. In an interesting sidelight, the Rich group noted that the D loops of *S. cerevisiae* and *S. pombe* $tRNA^{Phe}$, which have the same loop sequences but different stem sequences, were cleaved differently by T1 nuclease, although the interpretation of this result is complicated by the interaction of the D and TψC loops.

Later work by the RajBhandary group (Seong and RajBhandary 1987a) proved that the unusual conformation of *E. coli* fMet-$tRNA_f^{Met}$ arises from the G:C stack by mutating each G:C to another Watson-Crick base pair. Mutants with two, one or no G:C pairs were successively less S1-resistant, with the triple mutant (UCA:UGA, the sequence found in *E. coli* $tRNA_m^{Met}$) having S1 susceptibility similar to that of elongator tRNAs. In the course of a study of tRNA requirements for initiation factor selection (Hartz et al. 1990), the Gold group showed that the anticodon stem alone suffices to confer S1 resistance on the anticodon loop: oligonucleotides consisting solely of the sequence of the anticodon arm of *E. coli* $tRNA_f^{Met}$, even a 15-mer encompassing only the G:C stack and the anticodon loop (figure 2.1.3), were S1-resistant in a way qualitatively similar to their parent tRNA. Because these oligonucleotides were unmodified T7 transcripts, this also showed that modified nucleotides are not required for S1 resistance. Interestingly, a molecule with the G:C stack reversed (CCC:GGG) was not S1-resistant, showing that S1 resistance is not solely a function of the thermodynamic stability of the stem, nor were several molecules with varying anticodon loop sequences, showing that the S1-resistant conformation cannot be conferred on every loop.

Throughout these studies, S1 resistance was correlated with initiator function. The RajBhandary group's mutants were able to bind to the P site of the ribosome (measured both by a filter-binding assay and by puromycin reactivity of the initiation complex) in proportion to their S1 resistance. The Gold group measured the ability of their oligonucleotides to assemble initiation complexes by measuring the stopping of reverse transcription of the mRNA at the initiation codon, or "toeprinting". In the absence of initiation factors, all RNAs tested could form initiation complexes, but in the presence of IF3 only those RNAs which were S1-resistant could do so. This showed both that S1 resistance seems to be a marker for initiator tRNA anticodon stem-loop identity, and that it is IF3 which recognizes this aspect of initiator identity, or causes it to be recognized by some other component of the initiation complex.

Only recently has the RajBhandary group put forth an example of an initiation-competent tRNA which is not S1-resistant in the same way as initiator tRNAs as such (Mandal et al. 1996): $tRNA_f^{Met}$ with the mutation U35A36 (giving it an amber-suppressing CUA anticodon) is cleaved by S1 nuclease at positions 34, 35 and (slightly less strongly) 36. Mutations in the anticodon stem which severely reduce initiator function do not affect this cleavage pattern. Additionally, S1 cleavage of the U35A36 loop is the same in the presence or absence of the ms^2i^6 modification which was already known to be required for it to function as an initiator. Because neither the anticodon mutant (which can initiate) nor the double mutants (which cannot) have the same S1 cleavage pattern as natural tRNAs, these results must be interpreted carefully. It is possible that initiator and elongator loops differ structurally in some way which is reflected in S1 nuclease cleavage patterns in the context of the native anticodon loop sequence, but not in the context of the amber-suppressor sequence. Nonetheless, these results show that S1 cleavage does not unambiguously reflect initiator identity.

Finally, a puzzling observation in yeast (von Pawel-Rammingen et al. 1992): mutation of the G29:C41 and/or G31:C39 to the corresponding nucleotides in $tRNA_m^{Met}$ has virtually no effect on growth rate. This may mean, in contradiction to the universal conservation of the G:C stack, that it is not important in initiation in yeast, or perhaps just that initiation is not severely growth-limiting in yeast.

Other biochemical experiments also point to structural differences between initiator and elongator methionine tRNAs. An analysis of magnesium, europium and lead cleavage sites in yeast and lupin tRNAs showed that initiator methionine tRNAs were cleaved more readily than elongator methionine tRNAs, and both more readily than $tRNA^{Phe}$ (Ciesiolka et al. 1989). The authors speculated that metal ions might bind more tightly to a rigid structure, explaining why the initiator loop (thought to be more rigid from its resistance to S1 nuclease) was cleaved most easily. While residue 30, when a G, is always reactive to reagents which attack the N7 atom of purines, whatever the surrounding bases, it is unreactive in $tRNA_f^{Met}$, suggesting that not just the anticodon loop but the stem itself is more rigid than that of elongator tRNAs (Wakao et al. 1989).

2.1.6 Structures of initiator tRNAs and related molecules

The biological interest in and the evidence for the unusual structure of initiator tRNAs made them early targets of X-ray crystallography. An early structure of yeast $tRNA_i^{Met}$ (Schevitz et al. 1979) was of insufficient resolution (4.5 Å) to furnish a detailed model, but indicated that it had an architecture very similar to that of yeast $tRNA^{Phe}$, excepting a small difference in the position of the anticodon arm. A 3.5 Å structure of *E. coli* $tRNA_f^{Met}$ (Woo et al. 1980) was also very similar to yeast $tRNA^{Phe}$ but with an important difference: the nearly universally conserved base U33, which is part of (and gives its name to) the U-turn in yeast $tRNA^{Phe}$, is flipped out of the anticodon loop and makes a lattice contact. The "uridine swivel" hypothesis took this to be biologically significant, suggesting that U33 might be flipped out in a tRNA in the P site of the ribosome and part of the U-turn in a tRNA in the A site of the ribosome. Unfortunately, a later re-refinement of yeast $tRNA_i^{Met}$ to 3.0 Å (Basavappa and Sigler 1991) failed to find this feature: while it did reveal the unique eukaryotic $tRNA_i^{Met}$ D-TψC structure and the phosphoribosyl modification of A64 discussed above, the anticodon stem-loop was poorly defined and involved in lattice contacts, preventing the authors from discussing its features with confidence.[3] The Westhof group reexamined the yeast $tRNA_i^{Met}$ data (Dirheimer et al. 1995) and proposed an anticodon loop structure in which U33 was stacked into the loop but t^6A37 was flipped out! They noted the resemblance of this structure to that of the TψC loop, in which the equivalent bases 59 and 60 are flipped out to pair with bases in the D loop, and suggested that the initiator tRNA anticodon might adopt a TψC-like conformation.

---

[3] At this writing there are no publically available coordinates for any initiator tRNA.

Structural data collected in solution do not support the various conformations seen by crystallography. In a chemical and enzymatic probing study of $tRNA_f^{Met}$ (Wakao et al. 1989), U33 was observed to be unreactive with CMCT, indicating that its imino proton, which would be hydrogen-bonded to the phosphate oxygen of U36 in a U-turn but exposed to solvent in the $tRNA_f^{Met}$ structure, was in some way protected in solution. A37's N7 atom was strongly reactive with DEPC, but no more so than was predicted for the equivalent base in the crystal structure of $tRNA^{Phe}$. A38's N7 was unreactive with DEPC under native conditions and was probably stacked into the loop as is the equivalent base in $tRNA^{Phe}$.

There are no other obvious hints at the nature of the special structure of the anticodon loop of initiator tRNAs. While the repetitive nature of the G:C stack seems suggestive at first, poly(G):poly(C) is not expected to be structurally unusual on thermodynamic grounds (GC:GC is somewhat more stable than GG:CC (Freier et al. 1986), and GGG:CCC is of course exactly as stable as CCC:GGG), and crystallographic studies of molecules containing G:C runs have not revealed any unusual features (Betzel et al. 1994; Valegård et al. 1994; Portmann et al. 1995). Although G:C is the most stable closing base pair for GA:AG tandem mismatches (Walter et al. 1994), it is no better than C:G at closing AUAAUA hairpin loops (Serra et al. 1993) (recall that CCC:GGG does not substitute for GGG:CCC in initiator tRNAs).

Our understanding of RNA hairpin loop structure is limited to a few special cases at this point. While we might expect initiator tRNA anticodon loops to have something like a U-turn on functional grounds, we cannot predict the nature of the difference between initiator and elongator loops solely on the basis of their sequences, and the biochemical data are not sufficiently precise to dictate a particular three-dimensional structure. Existing structures of tRNA anticodon loops which differ from that in elongator tRNA crystal structures are probably not relevant here. The unusual structure of the anticodon loop of *E. coli* $tRNA^{Lys}$ is not well understood, but certainly results from the mnm^5s^2 modification of U34 (Watanabe et al. 1994). Unmodified DNA analogs of the yeast $tRNA^{Phe}$ anticodon loop can adopt an alternate secondary structure (Chen et al. 1993), but this is incompatible with the anticodon loop sequence of methionine tRNAs.

2.1.7 Thesis objective

The structure of the anticodon stem-loop of initiator tRNAs is, then, of extreme biological interest. Crystallographic methods have so far failed to reveal it convincingly. We know from biochemical studies that the S1-resistant structure of initiator tRNA anticodon loops is preserved in model oligonucleotides containing only the anticodon stem-loop, and that it does not require nucleotide modifications. Such a model oligonucleotide could easily be synthesized and would have a molecular weight well within the range accessible to solution nuclear magnetic resonance spectroscopy. The elucidation, by solution NMR, of the structures of the anticodon stem-loops of initiator and elongator methionine tRNAs constitutes the remainder of this work.

2.2 Methods

2.2.1 Design of constructs

Figure 2.2.1 shows the sequences of the anticodon stem-loops of yeast initiator and *E. coli* elongator methionine tRNA, and of the oligonucleotides that were designed to mimic them. The oligonucleotides begin with two Gs to promote T7 transcription, which are opposed by two Cs to prevent aggregation (Szewczak 1994). Except for the absence of modifications, the remainder of their sequences is that of their parent tRNAs, and they are numbered accordingly. The identities of the two most loop-distal base pairs of the anticodon stem (27:43 and 28:42) are not believed to be important for initiator/elongator discrimination or other tRNA function. Sequences from these particular species were chosen because those two base pairs happen to be the same. The two constructs thus differ from each other only in the three base pairs adjoining the anticodon loop. They are designated ITALY2 (Initiator TRNA Anticodon Loop from Yeast) and METAL2 (Methionine Elongator TRNA Anticodon Loop).

2.2.2 Synthesis

RNAs were transcribed from chemically synthesized DNA templates with T7 RNA polymerase (Beckett and Uhlenbeck 1984; Milligan et al. 1987; Seggerson 1994). Table 2.2.1 shows the components of the reaction. Chemicals were from commercial sources. RNA polymerase was prepared in-house and was a kind gift of other members of the laboratory. Table 2.2.2 shows the sequences of the templates. ITALY2T71 and METAL2T71 have the

Component	Final concentration	Stock concentration	μl for 50 μl transcription	ml for 10 ml transcription
Tris	40 mM	400 mM	5	1
Triton X-100	0.01%	0.1%	5	1
spermidine	1 mM	20 mM	2.5	0.5
DTT	80 mM	500 mM	8	1.6
$MgCl_2$	20 mM	200 mM	5	1
NTPs	2 mM each	20 mM each	5	1
DNAs	0.5 μM each	5 μM each	5	1
Polymerase	40 μg/ml	5 mg/ml	0.4	80 μl
H_2O			14.1	2.82

Table 2.2.1: T7 RNA polymerase transcription reaction components.

Template	Sequence
T7TOP1	`5'-TAATACGACTCACTATAG-3'`
ITALY2T7	`5'-GGCAGGGTTATGAGCCCTGCCTATAGTGAGTCGTATTA-3'`
METAL2T7	`5'-GGCATCATTATGAGTGATGCCTATAGTGAGTCGTATTA-3'`

Table 2.2.2: Templates for T7 RNA polymerase transcription of ITALY2 and METAL2. The residues which differ in ITALY2T7 and METAL2T7 are underlined.

complement to the T7 promoter at their 3' ends and the complement to the sequence to be transcribed at their 5' ends. T7TOP1 is the T7 promoter plus the first G of the sequence to be transcribed, which is displaced by the polymerase.

Test reactions were done at 50 µl or 1 ml scale; preparative reactions were done at 10 ml scale. Reactions at all scales were done in disposable centrifuge tubes of sufficient size so that the reaction, extraction and precipitation could be carried out in the same tube. All components except polymerase were mixed and warmed to 37°C, then polymerase was added and mixed gently. The reaction was incubated for 4-6 hours at 37°C. 1/10 volume of 0.5 M EDTA was added to stop the reaction and redissolve the magnesium pyrophosphate precipitate generated during the reaction. The reaction was extracted once with 25:24:1 phenol:$CHCl_3$:isoamyl alcohol buffered with Tris-HCl at pH 8. The RNA was precipitated with 2.5 volumes of ice-cold ethanol, washed with 70% ethanol and dried under vacuum.

Formamide loading buffer contained 80% formamide, 0.5x TBE (Maniatis et al. 1982) and 0.25% bromphenol blue; xylene cyanol was omitted since it comigrates with 21-nt RNA. Test transcriptions were resuspended in 50% formamide loading buffer in TE (Maniatis et al. 1982) and separated on denaturing 20% polyacrylamide gels containing 6 M urea, 0.5x TBE, 0.8 mm thick x 15 cm long, prerun at 5 W constant power until current dropped to a stable minimum, then run at 15 W constant power until the bromphenol blue had migrated most of the length of the gel (about 30'). Gels were fixed in 0.1 M HOAc, stained with methylene blue (0.4% methylene blue, 0.2 M NaOAc, pH 4.7), destained with warm tap water and photographed. Preparative transcriptions were resuspended in a few ml TE and dialyzed overnight against deionized, distilled water to remove nucleotides and short abortive transcripts. An equal volume of formamide loading dye was added and the transcriptions were heated to 80°C for 5-10'. The RNA was separated on denaturing 20% polyacrylamide gels, 0.5x TBE, 1.6 mm thick x 45 cm long, prerun at 20 W constant power and run at 50 W constant power until the bromphenol blue began to exit the gel (about 3 hours). Enough RNA was loaded onto each gel to yield 20 ODs of RNA after elution. It was essential to 1) prerun the gel thoroughly, 2) dialyze transcriptions before loading, 3) not overload the gel and 4) load in the minimum possible volume (lyophilizing again if necessary) to maximize resolution. Bands were visualized by UV shadowing and cut from the gel. RNA was electroeluted from the gel slices using an Elutrap apparatus (Schleicher and Schuell). The eluate was lyophilized and resuspended in TE.

2.2.3 S1 nuclease cleavage

RNAs were 5'-end-labeled, sequenced and cleaved with S1 nuclease as described in appendix 1.

2.2.4 Nondenaturing gel electrophoresis

0.5 or 1 μl of RNA solution was removed directly from an NMR sample (see below), mixed with 1.5 or 1 μl nondenaturing loading buffer, and analyzed on a 20% polyacrylamide gel prepared in S1 NMR buffer (see below), 0.8 mm thick x 15 cm long, run at 70 W until the bromphenol blue had migrated most of the length of the gel (about 24 hr). The gel buffer was recirculated with a peristaltic pump to maintain a constant ionic composition. RNA was visualized with methylene blue as above.

2.2.5 UV melting spectroscopy

Samples were prepared and melted as described in section 1.2.4 using S1 NMR buffer (see below).

2.2.6 NMR sample preparation

RNAs were dialyzed extensively (three changes of at least four hours each) against S1 NMR buffer (100 mM KCl, 50 mM NaCl, 4 mM Na cacodylate, 0.2 mM EDTA) in 1K MWCO tubing (Spectrum). 11.5 mm diameter 1K MWCO tubing, which is quite stiff, rather than a microdialysis apparatus, was found to be the most convenient way to dialyze small samples. RNAs were concentrated to less than 400 µl in a Centricon-3 concentrator (Amicon). The concentrator membrane contains glycerol, which is visible in proton NMR spectra around 3 ppm. Soaking the concentrator in NMR buffer overnight removed the glycerol and the unwanted signals. Samples were placed in a 5 mm NMR tube (Wilmad 528-PP) and brought up to approximately 400 µl with buffer. 20 µl D_2O and 0.5 µl 2.5% dioxane were added to samples destined for exchangeable-proton spectroscopy for lock and chemical shift references. Samples for non-exchangeable proton spectroscopy were frozen, lyophilized and resuspended in 99.96% D_2O twice, then frozen, lyophilized and resuspended in 99.996% D_2O plus 0.5 µl 2.5% dioxane in D_2O. D_2O samples were always sealed with Parafilm and refrigerated or frozen between experiments to reduce contamination with atmospheric water vapor.

2.2.7 NMR spectroscopy

Spectra were collected on instruments in the Yale Chemical Instrumentation Center and the Bass NMR facility: the Yale 490 (originally controlled by an Aspect 2000 and later by a Tecmag Libra spectrometer), a Bruker AM500, a GE Omega 500, a Varian Unity 500 and a Varian Unity+ 600.[4] Homonuclear experiments were collected on whatever instrument was available using standard pulse programs. Natural-abundance ^{13}C-^{1}H HSQC spectra were collected on the 600 MHz spectrometer to take advantage of its high field and thus its superior sensitivity. Phosphorus and phosphorus-proton experiments were collected on the AM500. ^{31}P-^{1}H COSY spectra were collected as described in (Sklenár et al. 1986) and ^{31}P-^{1}H heteroTOCSY and heteroTOCSY-NOESY spectra as in (Kellogg et al. 1992; Kellogg 1992). The carrier was set at 4.7 ppm (H_2O or HDO) for proton, -3.5 ppm for ^{31}P and 148 or 79 ppm, depending on the range of ^{13}C chemical shifts which was to be observed, for ^{13}C. ^{1}H chemical shifts were referenced to an 0.2 mM dioxane (3.741 ppm) internal standard, ^{31}P chemical shifts to a trimethyl phosphate (0 ppm) external standard, and ^{13}C chemical shifts to a sodium formate (26.1 ppm) external standard. Table 2.2.3 shows typical acquisition parameters for each experiment. All spectra were collected at 30°C unless otherwise specified.

Data were transferred to Silicon Graphics workstations and processed using Felix (MSI). Table 2.2.4 shows typical processing parameters for each experiment. The residual water peak in homonuclear experiments was removed by convolution before apodizing. The sweep width in the phosphorus dimension of phosphorus-proton experiments is very narrow, so

---

[4]The author has petitioned to have "Comparative NMR Instrumentation" appended to the title of his degree.

Experiment	SW_2 (ppm)	SW_1 (ppm)	Mix (ms)	Recycle time (s)	Scans	t_1 points
NOESY	8	8	50-300	2	64	500-600
COSY	8	8		2	64	"
HSQC (aromatic)	8	33		2	64	"
HSQC (5, ribose)	8	33		2	64	"
TOCSY	8	8	200	4	32	"
GE JRSE NOESY	20	20	300	1	64	900
^{31}P-^{1}H COSY	8	2.5		2.5	64	250
heteroTOCSY	8	2.5	40	2.5	128	160-180
heteroTOCSY-NOESY	8	2.5	40 (TOCSY) 500 (NOESY)	2.5	384	"

Table 2.2.3: Parameters for NMR data collection. "SW_2" and "SW_1" are the sweep widths in the direct and indirect dimensions respectively. "Mix" is the NOESY or TOCSY mixing time, where appropriate. "Recycle time" is the total time including acquisition for a single scan.

Experiment	t_2 apodization	t_1 apodization	t_2 points	t_1 points
NOESY	ss 768 50	ss 50	2048	2048
COSY	ss 768 0	ss 0	1024	1024
HSQC (aromatic)	ss 768 50	ss 90	1024	1024
HSQC (5, ribose)	ss 768 50	ss 90	1024	1024
TOCSY	ss 768 50	ss 50	2048	2048
GE JRSE NOESY	ss 768 50	ss 50	2048	2048
^{31}P-^{1}H COSY	ss 1024 70	qss 0 0.5	1024	256
heteroTOCSY	ss 768 50	ss 50	2048	256
heteroTOCSY-NOESY	ss 768 90	ss 90	2048	256

Table 2.2.4: Parameters for NMR data processing. Apodization is expressed as the appropriate Felix command: "ss X Y" is a squared sinebell X points long shifted by Y degrees and "qss X Y Z" is the same sinebell with a skew of Z. The length of the apodization function in t1 is the number of data points in t_1. "t_2 points" and "t_1 points" are the size to which the data were zero-filled in the appropriate dimension before Fourier transformation.

they can be processed into matrices with very few points in that dimension. Smaller matrices (1024 instead of 2048 points) were also used for scalar coupling-based experiments which produced less crosspeaks and required less detail.

2.2.8 Constraints for structure calculation

2.2.8.1 Base-pairing constraints from observations of exchangeable protons

Bases which were judged to be Watson-Crick base-paired were constrained in two ways: 1) Distance constraints were defined between each of the two (for the A:U base pair) or three (for the G:C base pairs) hydrogen bond donor-acceptor pairs. 2) A planar pseudoenergy term was defined for each base pair, requiring six atoms (three in each base) to be coplanar. Conveniently, all four RNA bases contain atoms designated N1, N3 and C5, so planarity terms could use these atoms and be independent of base type.

2.2.8.2 A-form constraints

Regions which were judged to be A-form helix were constrained to be so in two ways. 1) An ideal A-form helix with the appropriate sequence was built using Insight II (MSI), and a list of all proton-proton distances under 5 Å was generated from it and converted to distance constraints in X-PLOR format using the programs dm (A2.1) and dm2xplor (A2.2). 2) Backbone and ribose dihedral angles were constrained to A-form values with an allowed deviation of only 15°.

2.2.8.3 Distance constraints from NOEs

Distance constraints were extracted from NOESY spectra with 50, 100 and 300 ms mixing times. Each NOESY spectrum was displayed using Felix' "fire ramp" color scheme, which gives increasingly higher contours increasingly brighter colors. Peaks were picked on both sides of the diagonal, since there were cases in which peaks which could not be resolved in the upper left could be resolved in the lower right and vice versa. Peak intensities were scored as strong, medium or weak, using contour levels to maintain consistency. Peak lists were exported from Felix by saving peaks entities as ASCII files. The CHURN suite of programs (A2.3) was used to convert each peak list to a crosspeak database (peaks2cpd), check it against the sequence, covalent structure and known chemical shifts of the molecule (check-cpd) and combine the three peak lists into a list of distance ranges readable by X-PLOR (cpd2xplor).

Peaks which were strong or medium at 50 ms were placed in the short distance range (2.4 ± 0.6 Å). Peaks which were absent or weak at 100 ms but strong or medium at 100 ms were placed in the medium distance range (3 ± 1 Å). Peaks which were absent or weak at 50 and 100 ms but which appeared (regardless of intensity) at 300 ms were placed in the long distance range (4 ± 1 Å). If a peak had different intensity scores on opposite sides of the diagonal, the higher of the two scores was used. There were no peaks whose intensity score decreased with increasing mixing time. Peaks involving an H5' or H5'' were assigned arbitrarily to one or the other and their uncertainties were increased by 1.8 Å (the distance between a H5' and H5'').

2.2.8.4 χ dihedral angle constraints from NOEs

χ, the dihedral angle around the base-ribose glycosyl bond (see (Saenger 1983) for dihedral angle definitions), is always found in one of two conformations in RNAs: anti (200°), where the base is pointed away from the ribose, or syn (80°), where the base is over the ribose. Anti is preferred for steric reasons except in unusual circumstances. The distance between the 6 or 8 proton of a base and its ribose H1' is 3.8 Å when the base is anti but 2.5 Å when it is syn, so a syn base would be distinguished by a short-range anomeric-aromatic NOE (Varani and Tinoco 1991b). χ was constrained to the anti value ± 30° for each nucleotide without such an NOE.

2.2.8.5 Ribose dihedral angle constraints from $^3J_{H1'-H2'}$s

In C3'-endo ribose rings, the H1'-H2' dihedral angle is close to 90° and $^3J_{H1'-H2'}$ is small (0.5 Hz). The H1'-H2' dihedral angle increases to nearly 180° and $^3J_{H1-H2'}$ increases to about 10 Hz in a C2'-endo ribose. Intermediate coupling constants reflect a weighted average of states. The fraction of C3'-endo character, here designated $f_{C3'\text{-endo}}$, can be calculated from $^3J_{H1-H2'}$ with the equation (van den Hoogen 1988)

$$f_{C3'\text{-endo}} = 1.149 - 0.145 \cdot {}^3J_{H1'-H2'}$$

Absorptive and dispersive peak splittings were measured from the COSY spectrum and the program ext_j (A2.4) was used to calculate $^3J_{H1'-H2'}$ as described in (Kim and Prestegard 1989). $f_{C3'\text{-endo}}$ was calculated from $^3J_{H1'-H2'}$ with the program j2nu (A2.5). Ribose dihedral angles ν_0-ν_4 were constrained to C3'-endo values when ${}^3J_{H1'-H2'}$ was unmeasurably small, constrained to C2'-endo values when $f_{C3'\text{-endo}}$ was less than 25%, and left unconstrained when $f_{C3'\text{-endo}}$ had an intermediate value.

The dihedral angles for a ribose ring fixed at a conformation intermediate between C3'- and C2'-endo can be calculated as follows (Altona 1982):

$$P = f_{C3'\text{-endo}} \cdot P_{C3'\text{-endo}} + (1 - f_{C3'\text{-endo}}) \cdot P_{C2'\text{-endo}}$$

$$\nu_i = \tau_m \cos(P\pi/180 + 4\pi(i - 2)/5),\ i = 0\text{-}4$$

P is the pseudorotation phase angle, which varies linearly with $f_{C3'\text{-endo}}$, $P_{C3'\text{-endo}}$ and $P_{C2'\text{-endo}}$ are the observed values of P when $f_{C3'\text{-endo}}$ is 1 and 0 respectively and τ_m is the pucker amplitude. $P_{C2'\text{-endo}}$, $P_{C3'\text{-endo}}$ and τ_m were fixed at 9°, 162° and 37° respectively. P and ν_0-ν_4 were also output by j2nu.

2.2.8.6 α and ζ dihedral angle constraints from ${}^{31}P$ chemical shift

Phosphorus chemical shifts in A-form RNA are normally between -3.5 and -4.5 ppm. Downfield shifts reflect distortion of the bond angles around the phosphorus by non-A-form values of the dihedral angles α or ζ (Gorenstein 1984). α and ζ angles involving phosphorus atoms with A-like chemical shifts were constrained to the A-form values ± 30°. Since it is not possible to tell whether a downfield shift arises from non-A-form α, ζ or both, α and ζ angles involving downfield-shifted phosphorus atoms were left unconstrained.

2.2.8.7 β and γ dihedral angle constraints from ^{31}P-^{1}H coupling constants

$^{3}J_{P-H5'}$ and $^{3}J_{P-H5''}$ are both small when the β dihedral angle is trans, as it is in A-form RNA (Varani and Tinoco 1991b). Simple visual inspection of a ^{31}P-^{1}H COSY spectrum indicated that there were no intense crosspeaks between phosphorus and 5' or 5'' protons, and thus that all β were trans. (Phosphorus-H5'/'' crosspeaks would be distinguished from phosphorus-H3' crosspeaks by being split by the -12.8 Hz scalar coupling between H5' and H5''.) All β angles were constrained to the A-form value ± 30°.

$^{4}J_{P-H4'}$ is never large, but weak phosphorus-H4' crosspeaks are detected in a ^{31}P-^{1}H COSY spectrum when the two spins are in the optimal W conformation, i.e. β and γ are both trans (Varani and Tinoco 1991b). γ was constrained to the A-form value ± 30° for each nucleotide in which a phosphorus-H4' crosspeak, however weak, was observed.

2.2.8.8 ε dihedral angle constraints from steric considerations

Constraints on ε were derived from steric considerations rather than from NMR data. For technical reasons they were incorporated into X-PLOR calculations in the same way as were dihedral angle constraints derived from NMR data, so they are discussed here. $^{3}J_{P-H3'}$ is not very sensitive to ε, and there are no commonly observed four-bond couplings which involve ε. The only way to determine ε is to measure $^{3}J_{C-P}$ couplings, which would require a ^{13}C-labeled sample. However, ε is prevented by steric interactions from being gauche^{+} (0° to 120°) (Varani and Tinoco 1991b). ε in every nucleotide was constrained to -120 ± 120°.

2.2.9 Structure calculation

Structures were calculated with X-PLOR version 3 using the distance geometry/simulated annealing approach described in the X-PLOR manual (Brünger 1992). The scripts used (appendix 3) were derived from scripts given in the manual. Space does not permit a detailed description of the modifications made to each of the scripts, but they may be easily discovered by comparing derivative and original, for which reason the derivatives, although for the most part not conceptually different from the originals, are reproduced in full.

X-PLOR's 'knowledge' of molecular structure and energetics comes from two files, a topology file which defines atoms, bonds and angles and a parameter file which defines force constants for bonds, angles and nonbonded interactions (electrostatics and van der Waals). The parameter file uses the definitions in the topology file, so the two must be written and used together. Given the fundamental importance of this information to a successful structure calculation, it is surprising that none of the topology/parameter sets distributed with X-PLOR 3.1 are free of significant errors. One such error is that ribose ring bond lengths and angles are based on those of tetrahydrofuran, so are inconsistent with the dihedral angles observed in ribose rings. One or the other cannot be refined to low energy.

Fortunately, an adequate topology/parameter set became available just before the present calculations began. The Berman group published a heavy-atom only topology/parameter set based on a survey of the Cambridge small-molecule crystallographic database (Parkinson et al. 1996), and Jason Rife in the Moore laboratory (Rife and Moore 1996) modified it to include the protons necessary for NMR calculations. The protonated topology/parameter set is available at ftp://proton.chem.yale.edu/pub/rna-structure/.

A further minor error was found in the course of the present work. There is no improper angle term to preserve the chirality of the non-backbone phosphate oxygens O1P and O2P, so they can switch places during simulation. This has no effect on structure calculations, since these atoms are never involved in experimental constraints, but it inflates RMSDs. Since this error was found after calculations were complete, calculated structures were repaired after the fact by manually editing their PDB files.

Structures were calculated in three stages: preliminaries (generation of a molecular structure and template), distance geometry/simulated annealing and refinement. A simple protocol (make-rsf.inp, A3.3) combined the general information in the topology file with the sequence of ITALY2 into a "molecular structure file" used by all subsequent protocols.[5] A second preliminary protocol (make-template.inp, A3.4) generated a template, an extended structure with correct local geometry which is used as a reference by the DG/SA protocol. This protocol distributes all of the atoms in the molecule randomly in x and y but sequentially in z, and then uses minimization and dynamics to restore correct local geometry. This process may (about one in ten times with the current protocol) fail to find a structurally reasonable minimum. Failure manifests as crumpled purine rings, which are easily detected by their high energy and multiple bond and angle violations or by visual inspection using a molecular graphics program. Later protocols do not correct such flaws, so they must be eliminated before proceeding. Ten templates were generated and one was selected.

[5]This is usually referred to as a "protein structure file", or PSF. Appendices 3 and 4 use "RNA structure file", or RSF.

Initial structures were generated by distance geometry (dg.inp, A3.6). This algorithm transforms distance constraints (including distance constraints as such and those derived from covalent geometry and dihedral angle constraints) directly to coordinates. Unfortunately, incomplete (as is inevitably the case with NMR data) and or inconsistent input produces solutions which do not precisely satisfy the input distances. Distance geometry structures are globally correct, but appear to have had each of their atoms shifted by an Ångström or two in a random direction. The initial portion of the post-distance geometry simulated annealing protocol (dgsa.inp, A3.7) corrects this 'noise' by replacing each nucleotide in turn with the corresponding nucleotide of the template structure. While this does correct most faults in local geometry, it distorts the internucleotide (O3'-P) bond and erases the effect of experimental information on the intranucleotide dihedral angles. This procedure would be much more appropriate for proteins, which have many more but much simpler monomers than nucleic acids.

Two fine points of distance geometry deserve some comment. The first is metrization, a smoothing procedure which improves consistency among the initially undefined distances. Unfortunately, test calculations showed that metrization was detrimental, producing DG structures in which atoms were forced into a thin spherical shell. The system under study was sufficiently underconstrained that the random values assigned to initially undefined distances, which metrization tends to preserve, overwhelmed the actual data. Metrization was not used. The second is that nucleic acids contain large rigid groups (each base with all of its immediately exocyclic atoms) within which all interatomic distances are known. Incorporating these distances into distance geometry ought to improve its results. X-PLOR has a facility for doing so, but it does not work. The authors have been notified.

The protocol continues with simulated annealing. To speed calculations, only repulsive nonbonded energies are used; attractive electrostatics and van der Waals terms are omitted. Atoms are assigned velocities equivalent to 3000 K and dynamics are carried out for 12 ps while dihedral and improper angle constraints are slowly introduced. Coordinates and velocities are stored, the coordinates are reversed along the y axis and the entire protocol up to this point is carried out again. The enantiomer with lower energy is selected to be refined further. The system is then cooled to 300 K over 15 ps.

The final protocol (refine.inp, A3.9) applies the complete force field, including all nonbonded interactions, and carries out dynamics for 20 ps. The last 10 ps are averaged to produce the output structure. The effect on the structure is minor; most constraint violations present at the end of the DG/SA protocol remain after refinement.

While the procedure just described worked well enough, as will be seen in the results, it is needlessly complicated. Distance geometry is theoretically appealing, in that it produces a result directly, without the intervention of molecular mechanics, but it treats the input distances roughly and much effort must be expended to undo the damage. One could calculate structures by simulated annealing starting from templates with random dihedral angles or even random atomic coordinates, eliminating the need to build a template (Brünger 1992), or using torsion angle dynamics (Rice and Brünger 1994). Even a substantial increase in computational time would be offset in practice by the reduced complexity. Readers about to embark on solving a structure are advised to consider their options.

The DG/SA and refinement calculations were rerun repeatedly (usually in groups of 1, 10 or 20 structures) as bugs were fixed and new data were incorporated. Careful attention was paid to constraints which were repeatedly violated. Some NOEs were clearly affected by secondary transfer pathways and were increased or reduced by a distance class. A few, probably arising from unassigned NOEs with the same chemical shifts as assigned NOEs, were consistently violated by several Å or more and were eliminated.

Although the results of one set of calculations often informed the choices made when running the next set, structures were always calculated de novo, never recycled through multiple rounds of the same protocol as has been done in the past (Szewczak 1994; Rife and Moore 1996). It was more straightforward simply to increase the length of energy minimization and the temperature and length of molecular dynamics until the yield of converged structures no longer increased. Note that simply recycling a structure through the DG/SA protocol would subject it to template fitting each time, resetting nucleotide conformations to those in the template and undoing the good work of the simulated annealing component.

In the final calculation, the DG/SA and refinement protocols were run 30 times, producing 30 output structures. Structures were considered to have converged if they violated no constraints. The thresholds for violation of each constraint are listed near the end of refine.inp (A3.9). Dihedral angles in each converged structure were calculated with the program cdih_measure (A2.9) and averaged with the program average-dihedrals (A2.10). The protocol average.inp (A3.10) averaged the coordinates of selected structures, minimized the averaged structure to restore correct local geometry and calculated the RMSD between each converged structure and the average and the RMS RMSD over all structures for the stem, loop and entire molecule.

2.3 Results

2.3.1 Synthesis

Conditions for T7 transcription of ITALY2 and METAL2 were chosen primarily to maximize yield of RNA per unit of NTPs, in anticipation of synthesizing ^{13}C,^{15}N-labeled RNAs, and secondarily to maximize yield of RNA per unit of polymerase. The concentration of NTPs was fixed at 2 mM each, a factor of two less than is usually used for transcriptions with unlabeled nucleotides but typical of transcriptions with labeled nucleotides. 20 or 40 μg of polymerase were used per ml of transcription depending on availability. Doubling polymerase concentration approximately doubled yield, suggesting that polymerase was limiting, but yield also varied widely between batches of polymerase. The concentration of $MgCl_2$ was the only other parameter which strongly affected yield; a 2.5:1 ratio of $MgCl_2$ to total NTPs was optimal, as other workers have found (Puglisi 1989). Reaction conditions were optimized extensively at 50 μl and less extensively at larger scale; no change in conditions was necessary for large-scale transcription.

ITALY2 and METAL2 both yielded 0.5 to 1 OD of purified RNA per ml of transcription, and 1 to 2 OD when polymerase concentration was doubled. Yields were consistently slightly higher for METAL2, although the variation between templates was small compared to that among transcription batches. Transcriptions contained no high-molecular weight contaminants ("RNA X"), but did contain 1-10 nt failure sequences and substantial amounts of 22-nt runoff sequences ("N + 1"). The first round of gel purification did not completely remove runoff transcripts, which remained as a few percent of the "purified" RNA. This problem was particularly severe for METAL2, whose runoff band had a mobility unusually close to that of the full-length transcript. A second round of purification completely removed runoffs at a cost of 30-50% in yield.

As judged by the quality of its NMR spectra, the quality of the purified RNA was extremely high, much better than that of earlier samples which had been chemically synthesized and purified by HPLC (Webster et al. 1991). In fact, spectra of the chemically synthesized samples had been so poor that it appeared that heteronuclear methods would be needed to assign them. T7 transcription was begun so that isotopically labeled samples could be prepared. Spectra of the transcribed samples were so much better that the original plan of using only natural-abundance NMR methods was resumed. The poor quality of the earlier samples was due in part to the inadequacy of the second step of the HPLC purification: single or multiple rounds of ion-exchange HPLC did not completely remove failure sequences. Problems with the synthetic chemistry, which was executed outside the laboratory, may also have contributed, and it is possible that the synthetic molecules, which lacked the first two and last two nucleotides of the transcribed molecules, were inherently less suited to NMR spectroscopy. If the second step of the purification of the chemically synthesized molecules were replaced by gel electrophoresis or a different size-selecting HPLC column, failure sequences could be completely removed and the quality of the chemically synthesized molecules better assessed. Note also that several practical advances in deprotection and HPLC purification of RNA have been made recently (Sproat et al. 1995; Wincott et al. 1995; Anderson et al. 1996).

2.3.2 S1 nuclease cleavage

Figure 2.3.1 shows the results of sequencing and S1 nuclease cleavage experiments done on ITALY2 and METAL2. Several factors complicate their interpretation: (1) Cleavage of RNA by hydroxide ion and by T1 and U2 nucleases produces a 3' cyclic phosphate, while cleavage by S1 nuclease produces a 3' hydroxyl. S1 fragments thus run almost one "nucleotide" slower than the corresponding hydroxyl, T1 and U2 fragments on a sequencing gel. (2) Even in the absence of base or nucleases, ITALY2 and METAL2 are cleaved by an unknown agent, most strongly between C34 and A35 but also between U33 and C34 and between U36 and A37. This cleavage could not be eliminated despite considerable effort. It has also been seen in the intact tRNA by others (Wakao et al. 1989). Fortunately, the spontaneous cleavage products comigrate with hydroxide cleavage products, so they can be distinguished from S1 cleavage products. In addition, most spontaneous cleavage products are shorter than, and so cannot give rise to, the S1 cleavage products. Cleavage of RNA by an unknown agent in YpA sequences has been attributed to structurally promoted chemical hydrolysis (Dock-Bregeon and Moras 1987) or to environmental nucleases (Plaipa and Juodka 1996). (3) When equal amounts of labeled ITALY2 and METAL2 are cleaved and visualized under the same conditions (as in figure 2.3.1), the S1 cleavage products of METAL2 are much fainter than those of ITALY2, because METAL2 fragments are much more susceptible to secondary cleavage than ITALY2 fragments. Note also that the non-enzymatic cleavage products are present in all S1 lanes for ITALY2, but not for METAL2. This is reasonable since the stem of METAL2 can dissociate more easily to allow S1 to degrade

the non-enzymatic cleavage products, and was taken into account when comparing cleavage intensities.

ITALY2 is cleaved by S1 nuclease after C34 and A35, and less strongly after U36. METAL2 is cleaved by S1 nuclease after C34, A35, U36 and less strongly after A37. The strong cleavages are the same as those reported for tRNA anticodon loops by the Rich group (Wrede et al. 1979; Wrede and Rich 1979), except for the cleavage they observed between U33 and C34, which is very weak here.

The patterns of cleavage of both molecules are unaffected by magnesium (compare lanes 5 and 7). They are also unaffected by pH over the range from 4.5, which is the optimum for S1 nuclease, to 6.5, at which Zn^{++} is still soluble but the activity of S1 is reduced (compare lanes 5, 7, 13 and 15 to 6, 8, 14 and 16). ITALY2 is cleaved at pH 6.5 much as it is at pH 4.5. Fragments of METAL2 do not undergo as efficient secondary cleavage at high pH, and cleavage can be seen continuously from U33/C34 to A37/A38. (The U33/C34 cleavage band is very close to a spontaneous cleavage band.) Although S1 requires Zn^{++}, which precipitates at alkaline pH, the Rich group reported that the patterns of cleavage of initiator and elongator tRNAs were maintained at pHs as high as 7.4 (Wrede and Rich 1979), consistent with what is observed here.

Results obtained using a modified S1 cleavage buffer with twice the usual salt (KCl and NaCl) concentration (not shown) were indistinguishable from what is shown here, consistent with an earlier report that S1 activity is insensitive to ionic strength (Wrede and Rich 1979).

2.3.3 Nondenaturing gel electrophoresis

Figure 2.3.2 shows a nondenaturing polyacrylamide gel run in S1 NMR buffer. ITALY2 and METAL2 migrate between ITALY1 and METAL1, which are 17-nt stem-loops, and ABISS1, a 30-nt stem-loop. ITALY2 and METAL2 must be stem-loops rather than dimers under these conditions, which are nearly the same as those used for NMR spectroscopy. Although this gel was run at 5°C, at which temperature NMR spectra of ITALY2 are indicative of multiple conformations (see section 2.3.5), only one species is seen. This suggests that dimerization is not responsible for the multiple conformations seen. Gels run at 30°C give identical results (not shown). Note that METAL1 and METAL2 run slightly faster than ITALY1 and ITALY2 respectively.

2.3.4 UV melting spectroscopy

Figure 2.3.3 shows first-derivative plots of UV melts of ITALY2 and METAL2 at concentrations from 0.1 to 10 OD (0.6 to 60 μM) in S1 NMR buffer. Each displays a single transition well above room temperature, ITALY2 at 85°C and METAL2 at 72°C. Neither set of melts shows multiple transitions which might indicate sequential melting of multiple stacked structures or aggregation. Since melting does not depend on concentration, the molecules are stem-loops, not bimolecular duplexes, consistent with the nondenaturing gel electrophoresis results.

Melting temperatures can be predicted using a nearest-neighbor algorithm (Freier et al. 1986). This algorithm predicts melting temperatures of 99.5°C and 83.7°C for ITALY2 and METAL2 respectively. Since the data on which those predictions are based were collected in 1 M NaCl, the predicted

melting temperatures cannot be compared directly to the melting temperatures observed here. However, the differences between observed and predicted melting temperatures are comparable for the two molecules. This suggests that the difference in the sequence of the stems of these molecules determines their differences in stability. They do not support the idea that ITALY2 contains a stable substructure which is not present in METAL2.

2.3.5 Spectroscopic characterization

The effect of temperature on the imino proton spectrum of ITALY2 is shown in figure 2.3.4A. Six resonances are observed at room temperature, corresponding to the base pairs G26:C44 through G31:C39. (Their assignments are described in section 2.3.6.) As is usual with RNAs beginning with pppGG and ending with CC, the imino proton of the terminal base pair is not observed at room temperature. At 5°C, G25's imino is observed (at 12.9 ppm), as are two additional resonances which must arise from U33 and U36 in the loop.

While it was tempting to suppose that the non-exchanging loop imino protons observed at 5°C reflect a structure unique to initiator tRNAs, further observations suggest that this is not the case. The resonances of U33 and U36 are substoichiometric, and they appear in concert with a multitude of lesser substoichiometric peaks. In addition, ITALY2's nonexchangeable proton spectrum, shown in figure 2.3.5A, changes drastically with temperature. The widespread appearance and disappearance of peaks as the temperature is lowered makes it clear that the predominant structure at low temperature is different from that at room temperature. Another change at low temperature is a marked increase in linewidths, which is consistent with exchange

between multiple conformations and with dimerization or higher-order multimerization, but also with a simple increase in correlation time caused by increased viscosity and/or transient aggregation. Both multiple conformations and increased linewidths are observed in two-dimensional spectra of ITALY2 at low temperature, which consequently have very low signal-to-noise. Its natural abundance ^{13}C-^{1}H HSQC was completely featureless. Taken together, these observations suggest that ITALY2 is inhomogeneous in conformation, aggregation state, or both.

Interestingly, METAL2's imino proton spectrum (figure 2.3.4B) does not change at low temperature aside from the appearance of an imino proton resonance from the terminal base pair. METAL2's nonexchangeable spectrum does not change drastically either, although linewidths increase (figure 2.3.5B) and two-dimensional experiments degrade much as with ITALY2.[6] The appearance and disappearance of peaks in the spectrum ITALY2 as the temperature is lowered must, then, depend on the sequence of its stem. Since the differences between the patterns of S1 nuclease cleavage of ITALY2 and METAL2 are evident at 30°C, it was not necessary to characterize the low-temperature form. All subsequent spectroscopy was done at 30°C.

2.3.6 Assignment of ITALY2

The strategy used to assign the NMR spectra of ITALY2 and METAL2 is outlined in table 2.3.1. It was for the most part the traditional strategy used to assign RNAs at natural abundance in that it began with a sequential

[6]To the extent that this is a general phenomenon (if it involves transient aggregation, it may depend to some extent on sequence features common to the two molecules) it suggests that the upper limit on the size of RNAs which are accessible to solution NMR is much lower at low temperature.

Experiments	Assignments made
Aromatic-anomeric region of NOESY (with COSY and ^{13}C-^{1}H HSQC)	Aromatics, 1's
NOESY in water	Iminos and aminos, confirmation of above
1'-ribose regions of COSY and 50 ms NOESY	2'
heteroTOCSY-NOESY	^{31}P
^{31}P-^{1}H COSY	3'
100 ms NOESY	Some 4'
TOCSY, heteroTOCSY, aromatic-ribose regions of NOESYs	4' and 5'/'', confirmation of above

Table 2.3.1: Assignment strategy. Experiments are listed in the order in which they were analyzed.

assignment of aromatic (6 and 8) and anomeric (1') protons (Wüthrich 1986). However, phosphorus-proton correlations, and in particular ^{31}P-^{1}H heteroTOCSY-based experiments, played an important part in confirming these assignments and extending them beyond 2' sugar protons.

The aromatic-anomeric region of a 300 ms NOESY spectrum of ITALY2 is shown in figure 2.3.6. It contains crosspeaks between pyrimidine 5 and 6 protons, between 1' protons and aromatic (6 and 8) protons on the same and following nucleotides, and between adenine H2 protons and 1' protons on the following nucleotides on the same and opposite strands. While the presence of pyrimidine 5-6 crosspeaks are guaranteed by the covalent structure of the bases, the internucleotide crosspeaks on which sequential assignment depends are present only in A-form RNA or similarly structured regions. Fortunately, almost all of this molecule can be assigned as if it were A-form. There is continuous A-like connectivity from G25H1' through C45H6, with only a single break between U33H1' and C34H6, and there is only one non-A-like crosspeak, which correlates U33H1' and A35H8.

Aromatic-anomeric assignments were aided and confirmed by two ancillary experiments (data not shown): a COSY and a ^{13}C-^{1}H HSQC. The only crosspeaks in the aromatic-anomeric region of a COSY spectrum are those between H5 and H6 protons on the same pyrimidine base. Comparison of this region of the COSY and NOESY spectra identified the H5-H6 crosspeaks in the NOESY spectrum. In A-form RNA there are two crosspeaks between 1' protons and the H6 of the H5-H6 crosspeak, which identify the H1' of the ribose to which the pyrimidine base is directly bonded and that of the preceding nucleotide. There is sometimes also a weak crosspeak between the H5 and the H6 or H8 of the preceding nucleotide.

The ^{13}C-^{1}H HSQC experiment detects proton-carbon pairs joined by a single bond, and is thus relatively insensitive to conformation. However, carbon chemical shifts reflect chemical type in ways which proton chemical shifts do not: adenine C2s are separated from C6s and C8s; cytosine and uracil C5s are separated from each other and from C1's; and C2's, C3's, C4's and C5'/C5''s each resonate in a distinct region (Varani and Tinoco 1991a). By correlating crosspeaks in the ^{13}C-^{1}H HSQC and NOESY spectra, crosspeaks in the NOESY were sorted into these chemical classes. It remained only to distinguish adenine and guanine H8s, which was done by the conformational strategy discussed above.

A NOESY spectrum collected in water allowed assignment of the exchangeable protons in the molecule. Peaks in the imino-imino region arise from spatially adjacent imino protons. There are three such peaks, between G26H1 and G43H1, U42H3 and G29H1, and G30H1 and G31H1. As mentioned above, the imino protons of the terminal G (G25H1) and of the Us in the loop (U33H3 and U36H3) are not observed at 30°C. In the imino-aromatic region, each G imino has a crosspeak with the two amino protons of the C to which it is paired, and U42H3 has a crosspeak to A28H2. In the imino-anomeric region, secondary transfer from the G imino proton through the G and C amino protons gives rise to two anomeric and one CH5 crosspeak for each G imino (Heus and Pardi 1991a). Imino protons were assigned using their correlations with known H2, H5 and H1' chemical shifts.

Imino spectra collected at higher temperatures (not shown) were consistent with these imino assignments: imino proton resonances in base pairs near the ends of the helix disappeared most quickly with increasing temperature. Note also that the lone U imino is further downfield than the G imino, as expected. All this confirmed the earlier aromatic and anomeric assignments and demonstrated that ITALY2's secondary structure is what it was anticipated to be. Unfortunately, no exchangeable protons were observed in the loop.

Assignments of H1's were extended to H2's using the COSY spectrum and a NOESY spectrum collected at short mixing time. The coupling constant between 1' and 2' protons is conformationally sensitive; it is small (0.5 Hz) for the C3'-endo ribose rings found in A-form RNA and large (8 Hz) for the C2'-endo ribose rings found in B-form nucleic acids and other extended and/or dynamic conformations. The 1'-2' region of a COSY spectrum has a crosspeak for each ribose which is at least partially C2'-endo, of which there are six in ITALY2 (figure 2.3.7A). Two of these belong to the 5' and 3'-terminal nucleotides, which are at the end of the helix and relatively unconstrained. Their ribose protons, particularly the 2's, have chemical shifts which are distinctive because they are in unusual environments, and which are familiar because so many NMR constructs begin with two G:C base pairs. The terminal H2's were identified and the assignments of the terminal H1's were confirmed in this way.

1' and 2' protons are separated by about 3 Å regardless of ribose conformation, so a crosspeak is always detected in a NOESY spectrum for each ribose in an RNA. They are the only crosspeaks that can be detected in the 1'-ribose region at a short (about 50 ms) mixing time. At longer mixing times intranucleotide H1'-H4' crosspeaks appear, as do crosspeaks between H2's and the H1's on the following nucleotide. H1'-H3' and H1'-H5'/H5'' crosspeaks appear only at very long mixing times through secondary NOEs.

All of the foregoing is standard practice in natural-abundance NMR of RNA. Less commonly applied are the heteroTOCSY-based experiments developed by Kellogg (Kellogg 1992; Kellogg et al. 1992). The heteroTOCSY

pulse sequence simultaneously transfers magnetization between scalar-coupled heteronuclei and between scalar-coupled homonuclei. In RNA, a ^{31}P-^{1}H heteroTOCSY transfers magnetization primarily from phosphorus to the 3' proton of the N-1 ribose and thence among ribose protons. Displayed in two dimensions, this experiment groups all of the coupled protons in each ribose by the chemical shift of the phosphorus on its 3' side. The heteroTOCSY-NOESY experiment uses a NOESY step to transfer magnetization from protons excited by a heteroTOCSY sequence to additional protons which are not scalar coupled but are spatially adjacent, thus correlating phosphorus atoms with nearby anomeric and aromatic protons. Several NOESY pathways can contribute to the detected magnetization, so this experiment is less dependent on conformation than is a simple NOESY experiment.

The aromatic and anomeric regions of a heteroTOCSY-NOESY spectrum of ITALY2 are shown in figure 2.3.8. In the aromatic region (figure 2.3.8A), the strongest correlations are between phosphorus atoms and the aromatic protons of their own nucleotides, i.e. the aromatic proton of the base to the 3' side of each phosphorus. These arise by NOESY transfer from preceding ribose protons to the following aromatic protons, e.g. H2' (n-1) to H6 or H8 (n). There are also many crosspeaks between phosphorus atoms and the aromatic protons of the nucleotides to their 5' side, which come from NOESY transfer between ribose and aromatic protons of each nucleotide. Together these form an aromatic-phosphorus "walk" similar to the aromatic-anomeric walk discussed above, but with anomeric proton chemical shifts replaced by the corresponding phosphorus chemical shifts.

In the anomeric region (figure 2.3.8B), most phosphorus atoms have only one crosspeak, to the anomeric proton of the preceding nucleotide. This

comes from NOESY transfer between H2' and H1' of the each ribose. There are a few weaker crosspeaks between phosphorus atoms and the H1's of their own riboses, which come primarily from sequential H2' to H1' NOEs. As with the aromatic region, we see familiar proton-proton NOEs in a phosphorus guise. Although in this region one can "walk" through only a few short runs of sequence, one can correlate known H1' chemical shifts with unknown ^{31}P chemical shifts and assign the latter.

While the heteroTOCSY-NOESY experiment is not completely independent of conformation, and its signal-to-noise is not particularly high (the experiment shown took 48 hours to collect), it has the enormous advantage of correlating two sets of easily assignable protons with the same phosphorus atoms. Where crosspeaks are missing, weak or overlapped in one region, they are invariably present in the other.

This experiment allowed ITALY2's phosphorus spectrum to be assigned with a high degree of confidence, and it confirmed the prior aromatic and anomeric assignments. Even the non-A-form features of ITALY2's NOESY spectrum appear in its heteroTOCSY-NOESY spectrum: there is a weak "backwards" C34P to A35H8 crosspeak arising from NOEs between U33 ribose protons and A35H8, and since there is little or no NOE transfer from U33 ribose protons to C10H6 there is no C34P to C34H6 crosspeak.

With the phosphorus spectrum assigned, it was trivial to assign H3' chemical shifts from a ^{31}P-^{1}H COSY spectrum, which contained 20 distinct, strong crosspeaks. C45H3' is not adjacent to a phosphorus so did not give a crosspeak in this spectrum, but could be assigned by elimination from the C3' region of the ^{13}C-^{1}H HSQC spectrum. Since it is in the often-assigned terminal GG:CC sequence, its chemical shift was expected to be close to 4.16, which it is.

A number of experiments were used to assign the remaining protons (H5'/H5''s and the remaining H4's) and to further confirm prior assignments. As mentioned, the ^{13}C-^{1}H HSQC identified ribose protons by type. Additionally, many H5'/H5'' pairs could be correlated with one another via their shared carbon chemical shifts. A proton-proton TOCSY experiment correlated H1's to other ribose sugars. A TOCSY mixing time of 200 ms served to transfer substantial magnetization through even the 0.5 Hz $^{3}J_{H1'-H2'}$ couplings in the C3'-endo riboses in the rigidly A-form parts of the molecule. A heteroTOCSY experiment correlated phosphorus to protons in the ribose ring to the 5' side, as well as to 5'/5'' and sometimes 4' protons to the 3' side. Protons on the same ribose appeared in the same spin system in both the TOCSY spectrum and the heteroTOCSY spectrum, but H4's and H5'/5''s appearing in that spin system in the heteroTOCSY spectrum but not in the TOCSY spectrum were assigned to the ribose to the 3' side of the phosphorus. Finally, the aromatic-ribose and anomeric-ribose regions of NOESY spectra at various mixing times correlated spatially adjacent protons. H5'/5''s often gave NOESY crosspeaks to the aromatic proton of the same nucleotide. In an A-form conformation the aromatic-5' and aromatic-5'' distances would be similar. Stereospecific assignments of 5' and 5'' protons were not attempted.

Proton and phosphorus assignments for ITALY2 are tabulated in table 2.3.2, and its ^{13}C assignments are shown in table 2.3.3. Virtually every proton assignment is supported by correlations in at least three spectra or regions thereof.

	H1/ H3	NH1	NH2	H6/ H8	H2/ H5	H1'	H2'	H3'	H4'	H5'/''	H5'/''	P
G25	NO	NO	NO	8.13	NA	5.81	4.93	4.71	4.57	4.26	4.41	NO
G26	13.25	NO	NO	7.60	NA	5.90	4.56	4.56		4.27	4.51	-3.68
C27	NA	6.79	8.41	7.66	5.25	5.51	4.55	4.54	4.45	4.14		-4.26
A28	NA	NO	NO	7.96	6.98	5.94	4.66	4.68	4.50	4.15		-3.77
G29	12.69	NO	NO	7.13	NA	5.60	4.54	4.43		4.06		-3.91
G30	12.73	NO	NO	7.04	NA	5.71	4.60	4.39		4.03		-3.82
G31	12.95	NO	NO	6.99	NA	5.75	4.59	4.38	4.41	4.03	4.27	-3.78
C32	NA	NO	NO	7.59	5.33	5.55	4.31	4.42	4.37	4.09		-4.40
U33	NO	NA	NA	7.63	5.59	5.71	4.73	4.55	4.35	4.27	4.45	-3.71
C34	NA	NO	NO	7.70	5.83	5.48	4.07	4.36	4.19	4.00	4.02	-2.30
A35	NA	NO	NO	8.23	7.72	5.90	4.62	4.98	4.36	4.04	4.18	-3.35
U36	NO	NA	NA	7.74	5.73	5.89	4.48	4.71	4.51	4.20	4.29	-4.14
A37	NA	NO	NO	8.21	7.75	5.91	4.95	4.60	4.44	4.26		-3.07
A38	NA	NO	NO	7.70	8.08	5.64	4.14	4.44	4.53	4.20		-3.98
C39	NA	6.96	8.58	7.49	5.40	5.06	4.23	4.39		4.09	4.70	-4.60
C40	NA	6.91	8.46	7.85	5.44	5.50	4.36	4.52	4.40	4.06		-4.52
C41	NA	6.85	8.34	7.81	5.47	5.48	4.38	4.50	4.41	4.15	4.53	-4.33
U42	13.36	NA	NA	7.87	5.39	5.50	4.62	4.60	4.45	4.11		-4.28
G43	12.50	NO	NO	7.78	NA	5.78	4.48	4.59	4.62	4.13	4.54	-4.10
C44	NA	6.81	8.51	7.67	5.19	5.46	4.21	4.43	4.38	4.04		-4.52
C45	NA	NO	NO	7.66	5.48	5.74	3.99	4.16	4.45			-4.22

Table 2.3.2: ITALY2 proton and phosphorus chemical shifts. NA = "Not Applicable" (that nucleotide does not have a spin for that column). NO = "Not Observed". H5' s and H5''s are not stereospecifically assigned.

	C6/C8	C2/C5	C1'	C2'	C3'	C4'	C5'
G25	139.2	N A	91.3	75.1		83.6	67.4
G26	136.9	N A	93.1				66.3
C27	140.8	97.6	93.5				
A28	139.5	152.7	93.0	75.9	73.1	82.1	
G29	135.8	N A	92.5				
G30	135.8	N A	93.6				
G31	135.7	N A	93.4				
C32	140.5	97.3	93.8	76.2			
U33	142.1	104.4	93.0			83.0	67.2
C34	143.3	98.6	94.3	76.0	73.8	83.5	
A35	141.9	154.7	90.8		75.9	84.7	66.2
U36	143.4	105.5	91.1	75.8		85.1	67.6
A37	141.0	154.3	92.8	75.7			
A38		155.1	92.8			82.6	
C39	140.2	98.0	94.0	75.3			
C40	141.6	97.5	93.8	75.5			
C41	141.7	97.6	94.3	75.5			65.4
U42	141.9	103.7	94.2				
G43	136.5	N A	92.9	75.4			64.6
C44	141.2	97.1	94.3	75.7			
C45	141.9	98.3	93.0	77.7	69.8		

Table 2.3.3: ITALY2 carbon chemical shifts.

2.3.6 Qualitative features of the structure of ITALY2

Figure 2.3.9 summarizes the features of the spectra of ITALY2 which provide qualitative structural information. (1) Exchangeable spectra confirm that what was expected to be the stem is in fact base-paired. (2) There are no interactions in the loop which protect exchangeable protons. (3) There are sequential A-like anomeric-aromatic NOEs throughout the molecule, except for the a break between U33 and C34, and there is a nonsequential NOE between U33 and A35. (4) The absence of strong intranucleotide H1'-H6/H8 NOEs indicates that all bases are in the anti conformation. (5) H1'-H2' coupling constants are small (A-like) throughout the molecule, except in and just after the anticodon where they have values between those typical of A and B form nucleic acids, indicating flexibility. (6) Phosphorus chemical shifts, and thus α and ζ backbone angles, have values like those in A-form RNA except at C34, A35 and A37.

2.3.7 Constraints for calculation of the structure of ITALY2

Table 2.3.4 summarizes the constraints used to calculate the structure of ITALY2.

All relevant data indicated that the fourteen nucleotides in what was expected to be the stem were A-form and Watson-Crick base-paired. Distances between hydrogen bond donors and acceptors were constrained to ideal Watson-Crick values (A4.1) and paired bases were constrained to be planar (A4.2). The fourteen stem nucleotides were constrained to A-form distances and dihedral angles. Distances between protons closer than 5 Å apart were

Observation	Constraint	#
Stem		1736
Non-exchanging imino protons	Distances between Watson-Crick hydrogen-bonding partners	20
	Planarity of paired bases	7
A-form helical RNA	A-form distances	1560
	A-form dihedral angles	149
Loop		182
Nuclear Overhauser effect	Proton-proton distances	
	short (2.4 ± 0.6 Å)	27
	medium (3 ± 1 Å)	33
	long (4 ± 1 Å)	68
	total	128
	χ glycosidic dihedral angle	7
H1'-H2' coupling constants	ν_0-ν_4 ribose dihedral angles	20
^{31}P chemical shifts in A-form range	α and ζ backbone dihedral angles	8
^{31}P-^{1}H coupling constants	β and γ backbone dihedral angles	12
Steric exclusion	ε backbone dihedral angle	7

Table 2.3.4: Experimental observations and the resulting constraints on the structure of ITALY2.

constrained to those measured from an ideal A-form helix. Experimental information about the conformation of the stem was otherwise ignored.

Experimental data were considered in detail only for the seven loop nucleotides. 128 distance constraints, 71 of them intraresidue and 57 of them interresidue, were determined from NOEs involving these nucleotides (A4.3). Since there were no strong aromatic-anomeric NOEs which would indicate syn bases, all χ dihedral angles were constrained to anti values. H1'-H2' coupling constants, $f_{C3'\text{-endo}}$ and the conformational parameters derived therefrom are shown in table 2.3.5A. U36's ribose was determined to be only 23% C3'-endo and was constrained to have pure C2'-endo values of ν_0-ν_4. The riboses of C34, A35 and A37 had around 50% C3'-endo character and were left unconstrained.

The backbone dihedral angle constraints used are summarized in table 2.3.6. The phosphates of C32, U33, U36 and A38 had chemical shifts similar to those found in A-form RNA; their α and ζ dihedral angles were constrained to A-form values (Gorenstein 1984). Since there were no intense crosspeaks between phosphorus and 5' or 5'' protons, all β dihedral angles were constrained to be trans (Varani and Tinoco 1991b). Phosphorus-H4' crosspeaks were observed, confirming that β was trans and establishing that γ was trans (Varani and Tinoco 1991b), for all loop nucleotides excepting A35 and U36. Because the gauche$^+$ conformation is sterically forbidden for ε, ε was constrained not to be gauche$^+$ for all loop nucleotides. The complete set of dihedral angle constraints is shown in A4.4.

A: ITALY2

	$^3J_{H1'-H2'}$	$f_{C3'\text{-endo}}$	P	ν_0	ν_1	ν_2	ν_3	ν_4
G25	3.60	62.7	66.1	-27.5	7.7	15.0	-32.0	36.8
C34	4.90	43.8	94.9	-36.0	24.2	-3.2	-19.1	34.1
A35	4.50	49.7	86.0	-34.3	19.6	2.6	-23.8	35.9
U36	6.30	23.6	126.0	-35.2	35.2	-21.7	0.0	21.8
A37	3.80	59.8	70.5	-29.4	10.5	12.3	-30.5	37.0
C45	4.20	54.0	79.4	-32.5	15.9	6.8	-26.9	36.7

B: METAL2

	$^3J_{H1'-H2'}$	$f_{C3'\text{-endo}}$	P	ν_0	ν_1	ν_2	ν_3	ν_4
G25	3.60	62.7	66.1	-27.5	7.7	15.0	-32.0	36.8
A31	2.45	79.4	40.6	-14.2	-8.6	28.1	-36.9	31.6
C32	2.55	77.9	42.8	-15.5	-7.2	27.2	-36.7	32.3
U33	3.15	69.2	56.1	-22.8	1.3	20.6	-34.7	35.6
C34	3.70	61.3	68.3	-28.5	9.1	13.7	-31.3	36.9
A35	4.40	51.1	83.8	-33.8	18.4	4.0	-24.8	36.2
U36	5.25	38.8	102.7	-36.8	27.8	-8.1	-14.7	31.8
A37	3.20	68.5	57.2	-23.4	2.1	20.0	-34.5	35.8
A38	2.40	80.1	39.4	-13.5	-9.3	28.6	-36.9	31.2
C39	2.15	83.7	33.9	-10.1	-12.7	30.7	-37.0	29.1
C45	2.75	75.0	47.2	-18.1	-4.4	25.1	-36.3	33.6

Table 2.3.5: H1'-H2' coupling constants and derived quantities.

Residue	α	β	γ	ε	ζ
G25	none	none	none	A	A
G26	A	A	A	A	A
C27	A	A	A	A	A
A28	A	A	A	A	A
G29	A	A	A	A	A
G30	A	A	A	A	A
G31	A	A	A	A	A
C32	A	A	A	not g^+	A
U33	A	A	A	not g^+	none
C34	none	A	A	not g^+	none
A35	none	A	none	not g^+	A
U36	A	A	none	not g^+	none
A37	none	A	A	not g^+	A
A38	A	A	A	not g^+	A
C39	A	A	A	A	A
C40	A	A	A	A	A
C41	A	A	A	A	A
U42	A	A	A	A	A
G43	A	A	A	A	A
C44	A	A	A	A	A
C45	A	A	A	none	none

Table 2.3.6: Structural constraints on backbone dihedral angles in ITALY2. "A" means that the angle was constrained to the A-form value (± 15° in residues 25-31 and 39-45, ± 30° in residues 32-38), "not g^+" that the angle was constrained to not be gauche$^+$, and "none" that the angle was not constrained.

2.3.8 Calculation of the structure of ITALY2

Twelve of the thirty structures calculated violated no constraints and had low total energy. This success rate, 40%, was lower than that reported for proteins but as good as or better than usual for nucleic acids (Moore 1995). Note, however, that it was sensitive to the number, accuracy and precision of the constraints that were used and, in the SA protocol used here, to the temperature and length of simulation. Rates of convergence in test calculations varied between 10% to 80% as the protocols and input data were developed. The current success rate could unquestionably be raised, but probably at the cost of an offsetting increase in calculation time.

Many parameters were examined and modified in preliminary calculations. Minimization and dynamics times and temperatures were increased in several places, the weighting of dihedral angle terms was quadrupled and the weighting of NOE terms was doubled, all of which reduced the number of violations of one kind or another in the final structures.

Of particular interest is the choice of cutoff for NOE violations. It was originally set, following precedent, to 0.5 Å, allowing proton pairs evincing weak NOEs to be separated by up to 5.5 Å. 80% of structures converged without violations using this criterion, some to the structure which will be described presently and others to more or less related structures in which many pairs of protons constrained by weak NOEs were separated by 5.0 to 5.5 Å. Allowing only a 0.2 Å cutoff, and thus a maximum of 5.2 Å between pairs of protons evincing weak NOEs, reduced the success rate by half and narrowed down the converged structures to a single class.

All the dihedral angles in ITALY2 had similar values in all converged structures, except the α, β and γ dihedral angles of A35 which had one set of values in nine structures and another in the other three structures. Table 2.3.7 lists the average values of the dihedral angles, calculated over all twelve structures except for A35's α, β and γ which were calculated separately over the sets of nine and three structures. The coordinates of the set of nine structures were averaged and energy-minimized to produce the final structure.

All twelve converged structures are shown superimposed on the averaged structure in figure 2.3.10. The RMS RMSD between the individual structures and the average structure was 0.5 Å for the entire molecule, 0.3 Å for the stem alone (residues 25-31 and 39-45, which were artificially constrained) and 0.6 Å for the loop (residues 32-38). Individual RMSDs ranged from 0.3 Å to 0.8 Å for the entire molecule, 0.2 Å to 0.4 Å for the stem and 0.3 Å to 0.9 Å for the loop. Table 2.3.8 lists the dihedral angles of the averaged structure.

2.3.9 Description of the structure of ITALY2

Figure 2.3.11 shows the loop nucleotides of the averaged structure. On the 5' side of the loop, C32 stacks on G31 and U33 stacks to some degree on C32. There is a sharp turn between U33 and C34, due almost entirely to C34's α dihedral angle which is gauche$^+$ rather than the gauche$^-$ usual for A-form RNA. On the 3' side, C34 through A38 stack on C39. The C2'-endo ribose ring of U36 creates a kink between U36 and A37. C34's ribose is also relatively extended, while those of A35 and A37 are close to C3'-endo.

	α	β	γ	ε	ζ	χ
G25	0 ± 0	-110 ± 0	41 ± 2	-148 ± 0	-74 ± 0	-157 ± 0
G26	-58 ± 1	173 ± 0	44 ± 1	-154 ± 0	-73 ± 1	-156 ± 0
C27	-62 ± 1	177 ± 1	47 ± 0	-151 ± 0	-75 ± 0	-161 ± 0
A28	-58 ± 0	175 ± 0	45 ± 0	-152 ± 0	-73 ± 0	-155 ± 0
G29	-57 ± 0	174 ± 0	45 ± 0	-153 ± 0	-73 ± 0	-156 ± 0
G30	-59 ± 0	174 ± 0	47 ± 0	-155 ± 1	-72 ± 1	-155 ± 0
G31	-60 ± 1	176 ± 1	51 ± 0	-150 ± 4	-63 ± 3	-157 ± 0
C32	-67 ± 3	-165 ± 2	44 ± 2	-122 ± 8	-84 ± 8	-162 ± 2
U33	-43 ± 3	164 ± 5	39 ± 4	-156 ± 9	-52 ± 9	-163 ± 4
C34	121 ± 19	163 ± 7	70 ± 6	-110 ± 11	-83 ± 11	-130 ± 6
A35	-104 ± 14	158 ± 5	56 ± 11	-169 ± 3	-67 ± 3	-126 ± 3
	86 ± 2	-151 ± 0	175 ± 1			
U36	-66 ± 5	171 ± 2	69 ± 2	-98 ± 1	-94 ± 9	-127 ± 0
A37	-90 ± 19	153 ± 1	53 ± 12	-159 ± 0	-70 ± 2	-149 ± 2
A38	-64 ± 1	149 ± 0	83 ± 0	-117 ± 0	-67 ± 0	174 ± 2
C39	-52 ± 0	162 ± 0	44 ± 1	-151 ± 0	-80 ± 0	-162 ± 0
C40	-52 ± 0	174 ± 0	41 ± 0	-149 ± 0	-76 ± 0	-160 ± 0
C41	-58 ± 0	173 ± 0	45 ± 0	-154 ± 0	-76 ± 0	-162 ± 0
U42	-55 ± 0	176 ± 0	43 ± 0	-149 ± 0	-77 ± 0	-161 ± 0
G43	-52 ± 0	172 ± 0	42 ± 0	-153 ± 0	-72 ± 0	-156 ± 0
C44	-60 ± 0	177 ± 0	45 ± 0	-154 ± 0	-74 ± 0	-161 ± 0
C45	-61 ± 0	178 ± 0	48 ± 0	0 ± 0	0 ± 0	-160 ± 0

	ν_0	ν_1	ν_2	ν_3	ν_4
G25	2 ± 0	-25 ± 0	37 ± 0	-36 ± 0	21 ± 0
G26	0 ± 0	-23 ± 0	38 ± 0	-39 ± 0	25 ± 0
C27	1 ± 0	-25 ± 0	37 ± 0	-37 ± 0	22 ± 0
A28	0 ± 0	-24 ± 0	37 ± 0	-38 ± 0	23 ± 0
G29	0 ± 0	-23 ± 0	37 ± 0	-39 ± 0	25 ± 0
G30	0 ± 0	-24 ± 0	37 ± 0	-38 ± 0	24 ± 0
G31	-1 ± 0	-20 ± 0	34 ± 0	-35 ± 0	23 ± 0
C32	14 ± 1	-29 ± 1	33 ± 0	-25 ± 0	7 ± 1
U33	-1 ± 3	-19 ± 2	32 ± 1	-34 ± 1	22 ± 2
C34	-14 ± 4	21 ± 16	-20 ± 21	12 ± 20	0 ± 10
A35	-2 ± 5	-20 ± 3	34 ± 0	-36 ± 3	24 ± 5
U36	-29 ± 1	23 ± 2	-10 ± 2	-6 ± 2	22 ± 2
A37	12 ± 0	-34 ± 0	42 ± 0	-36 ± 0	14 ± 1
A38	4 ± 2	-27 ± 1	39 ± 0	-37 ± 1	20 ± 2
C39	0 ± 0	-24 ± 0	38 ± 0	-39 ± 0	25 ± 0
C40	3 ± 0	-25 ± 0	37 ± 0	-36 ± 0	20 ± 0
C41	1 ± 0	-25 ± 0	38 ± 0	-38 ± 0	23 ± 0
U42	4 ± 0	-26 ± 0	37 ± 0	-36 ± 0	20 ± 0
G43	1 ± 0	-25 ± 0	38 ± 0	-38 ± 0	23 ± 0
C44	2 ± 0	-25 ± 0	37 ± 0	-37 ± 0	21 ± 0
C45	1 ± 0	-23 ± 0	35 ± 0	-35 ± 0	21 ± 0

Table 2.3.7: Average values of dihedral angles in the calculated structures of ITALY2, ± standard deviations. A35's α, β and γ are calculated over the sets of nine and three structures in which they each take on two distinct values.

	α	β	γ	ε	ζ	χ	ν_0	ν_1	ν_2	ν_3	ν_4
G25	0	-169	41	-148	-75	-158	2	-25	37	-37	22
G26	-59	174	44	-156	-72	-155	0	-24	38	-39	25
C27	-63	179	48	-152	-76	-161	2	-25	37	-37	22
A28	-58	175	46	-153	-74	-155	1	-24	37	-38	24
G29	-55	174	45	-152	-74	-156	-1	-24	38	-39	25
G30	-60	174	47	-158	-71	-156	0	-25	39	-39	25
G31	-62	178	52	-152	-61	-157	0	-22	34	-35	22
C32	-68	-163	44	-126	-85	-162	14	-30	34	-26	8
U33	-38	162	39	-150	-51	-162	-2	-20	33	-35	23
C34	119	167	68	-116	-81	-128	-13	23	-23	16	-2
A35	-98	157	51	-170	-71	-124	-8	-17	34	-39	29
U36	-64	170	70	-99	-94	-128	-30	24	-10	-7	24
A37	-92	155	53	-160	-71	-150	13	-35	42	-36	15
A38	-65	149	84	-117	-68	176	6	-28	39	-37	20
C39	-53	163	44	-152	-80	-163	0	-25	39	-40	25
C40	-53	175	42	-149	-77	-161	4	-26	37	-36	20
C41	-58	172	45	-154	-78	-163	2	-26	39	-39	23
U42	-53	176	43	-149	-77	-161	4	-26	38	-36	20
G43	-53	172	43	-154	-73	-156	1	-25	38	-39	24
C44	-61	178	46	-155	-74	-162	2	-25	38	-38	22
C45	-62	178	50	0	0	-161	2	-24	36	-36	21

Table 2.3.8: Dihedral angles in the averaged structure of ITALY2.

Qualitatively, this structure contains the same "U-turn" as do the anticodon loops in crystal structures of elongator tRNAs. There are no looped-out bases, as there are in the crystal structure of *E. coli* $tRNA_f^{Met}$ (Woo et al. 1980). The superposition of the ITALY2 loop on the anticodon loop of $tRNA^{Phe}$ (Hingerty et al. 1978) is shown in figure 2.3.12. The RMSD between non-hydrogen backbone atoms in the two loops is 1.2 Å. (By way of comparison, the RMSD between the same atoms in two independently determined structures of $tRNA^{Phe}$ (Hingerty et al. 1978; Holbrook et al. 1978) is 0.8 Å.) While this is the best quantitative comparison that can be made, because the two loops differ in sequence, it understates the difference in the orientation of the bases. Although the bases of C32, U33, A37 and A38 superimpose well on the equivalent bases in $tRNA^{Phe}$ (figure 2.3.12A), the bases in the anticodon of the NMR structure are tilted about 40° towards the tip of the loop (figure 2.3.12B). This results from the extended C2'-endo conformation of U36's ribose.

Interatomic distances are consistent with two of the three hydrogen bonds reported to be characteristic of a U-turn (Jack et al. 1976): U33N3 is 2.9 Å from U36O2P and U33O2' is 2.9 Å from A35N7, just as in $tRNA^{Phe}$. However, A35N6 is 4.3 Å from U33O2', versus 3.1 Å in $tRNA^{Phe}$. A potential non-Watson-Crick base-pairing interaction at the base of $tRNA^{Phe}$'s anticodon loop may also be present in ITALY2: A38N6 is 3.5 Å from C32O2, closer than it is in $tRNA^{Phe}$ (4.2 Å).

2.3.10 Assignment of METAL2

The spectra of METAL2 were assigned in exactly the same way as were those of ITALY2. Although the data required were collected over a period of

several months, assignment was done in a single weekend. This was possible chiefly because METAL2 has essentially the same chemical shifts as ITALY2, except where their sequences actually differ (see below). Assignments are given in tables 2.3.9 and 2.3.10, and $^3J_{H1'-H2'}$ are given in table 2.3.5B. It was possible to assign the spectra of METAL2 somewhat more thoroughly than the spectra ITALY2, due to the superior quality of the sample which was used, the wider chemical shift dispersion in its stem, and its larger H1'-H2' coupling constants, which resulted in more efficient transfer in the proton-proton TOCSY experiment.

2.3.11 Qualitative features of the structure of METAL2 and comparisons with ITALY2

Figure 2.3.13 presents the structurally informative features of METAL2's spectra in the same format as figure 2.3.9. The two molecules are similar in most respects. They are both stem-loops. They have the same pattern of sequential and non-sequential NOEs around the loop (figure 2.3.14). Unusual phosphorus chemical shifts are found at the same positions in both. Comparison of proton chemical shifts in ITALY2 and METAL2 (figure 2.3.15) underscores their similarity: chemical shifts in the distal part of the stem and, importantly, in the loop, where the sequences are the same, are nearly identical. Chemical shifts differ only in and around the three base pairs proximal to the loop, where the sequences differ and there is no reason to expect similar shifts. These observations indicate that the structures of ITALY2 and METAL2 must be essentially the same.

However, returning to figure 2.3.4, we see two distinct and significant differences. The first is that U39H3, which we would expect to be protected by

	H1/ H3	NH1	NH2	H6/ H8	H2/ H5	H1'	H2'	H3'	H4'	H5'/''	H5'/''	P
G25	NO	NO	NO	8.14	NA	5.81	4.93	4.73	4.58	4.27	4.41	NO
G26	13.30	NO	NO	7.65	NA	5.91	4.56	4.60	4.54	4.28	4.52	-3.70
C27	NA	6.83	8.48	7.71	5.29	5.53	4.58	4.57	4.46	4.14		-4.27
A28	NA	NO	NO	8.04	7.33	5.96	4.51	4.64	4.69	4.05	4.16	-3.91
U29	14.04	NA	NA	7.64	5.03	5.46	4.37	4.43	4.42	4.12		-4.51
C30	NA	6.81	8.22	7.76	5.58	5.51	4.43	4.50	4.38			-4.14
A31	NA	NO	NO	7.88	7.35	5.88	4.50	4.58	4.44	4.12		-3.97
C32	NA	NO	NO	7.45	5.39	5.36	4.15	4.35	4.32	4.05		-4.17
U33	NO	NA	NA	7.62	5.57	5.66	4.49	4.50	4.27	4.03		-3.97
C34	NA	NO	NO	7.65	5.81	5.56	4.08	4.37	4.14	3.96	4.10	-2.99
A35	NA	NO	NO	8.20	7.85	5.89	4.65	4.88	4.36	4.03	4.17	-3.57
U36	NO	NA	NA	7.71	5.72	5.84	4.41	4.65	4.45	4.15	4.25	-4.09
A37	NA	NO	NO	8.15	7.76	5.83	4.81	4.65	4.52	4.22	4.36	-3.36
A38	NA	NO	NO	7.81	8.06	5.69	4.60	4.48	4.52	4.19	4.43	-3.83
U39	NO	NA	NA	7.47	5.42	5.24	4.47	4.51	4.42	4.10		-4.18
G40	11.82	NO	NO	7.75	NA	5.78	4.64	4.65	4.50	4.18		-3.87
A41	NA	NO	NO	7.84	7.65	5.91	4.46	4.62	4.50	4.14	4.39	-4.10
U42	13.47	NA	NA	7.56	5.01	5.46	4.47	4.50	4.40	4.06		-4.38
G43	12.55	NO	NO	7.68	NA	5.79	4.48	4.56	4.14	4.12	4.17	-4.04
C44	NA	6.81	8.54	7.67	5.20	5.48	4.22	4.43	4.39	4.04		-4.52
C45	NA	NO	NO	7.66	5.50	5.75	3.99	4.16	4.47	4.01	4.16	-4.22

Table 2.3.9: METAL2 proton and phosphorus chemical shifts.

	C6/C8	C2/C5	C1'	C2'	C3'	C4'	C5'
G25	139.3	N A	91.3	75.0	75.1	83.6	67.4
G26	137.0	N A	93.1	75.4		82.6	66.1
C27	143.5	97.5	93.9	75.5			
A28	139.8	153.3	93.3	75.7			
U29	143.4	102.7	93.4	75.5			
C30	141.3	98.1	94.0				
A31	139.5	154.0	92.9				
C32	141.1	97.5	93.3	76.0	73.0	82.5	
U33	142.5	104.6	92.8			83.5	
C34	141.7	98.7	92.7	75.9	74.7	83.7	66.0
A35	141.9	155.1	90.5		76.1	84.7	66.6
U36	143.5	105.3	91.2	75.7		84.8	67.5
A37	141.2	154.6	92.0	75.8		83.8	67.0
A38		155.1	92.1	75.8		83.0	66.1
U39	141.1	104.4	92.9				
G40	137.0	N A	92.8				
A41	139.8	154.1	93.1				
U42	141.0	103.2	93.1				
G43	136.3	N A	93.0			83.5	
C44	141.3	97.1	94.2	75.7			
C45	141.9	98.3	93.0	77.6	69.7		65.8

Table 2.3.10: METAL2 carbon chemical shifts.

base pairing between A31 and U39, is not observed. METAL2 really has a nine-nucleotide loop. The second is that while in ITALY2 there were measurable H1'-H2' coupling constants only for the riboses of C34 through A37, those in METAL2 are large throughout the entire loop and in the riboses adjacent to the loop. Taken together these indicate that METAL2's loop is more flexible than that of ITALY2, and that this flexibility is a result of structural differences at the base of the loop and/or the loop-proximal base pairs of the stem.

The effect of the relative flexibilities of the loops of ITALY2 and METAL2 on the conformations of their individual riboses is shown in figure 2.3.16. Although many more riboses have measurably large $f_{C2'\text{-endo}}$ in METAL2 than in ITALY2, those few in ITALY2 are larger. ITALY2 has a fixed structure in which only a few riboses sample non-A-like conformations, while in METAL2's structure many riboses are non-A-like at least part of the time.

A final interesting comparison is that between phosphorus chemical shifts, shown in figure 2.3.17. While the downfield-shifted phosphorus atoms are the same in each molecule, again indicating that their backbones are similarly structured, the magnitudes of the downfield shifts are uniformly smaller in METAL2. One interpretation of this is that while ITALY2 and METAL2 have the same structure, that of METAL2 is less stable. The chemical shift of a spin in fast exchange among conformational states is a time-weighted average of the chemical shifts in each state. The phosphorus chemical shifts of METAL2 reflect a time-weighted average between the relatively rigid conformation like that of ITALY2 with its distinctive phosphorus shifts and one or more extended, mobile conformations which have more A-like phosphorus shifts.

This interpretation is supported by ^{31}P spectra at low temperature (figure 2.3.18). Recall that ITALY2 undergoes a significant structural change when cooled from room temperature to 5°C (see section 2.3.5), but METAL2 does not. At 5°C METAL2's downfield phosphorus peaks move much further downfield, suggesting that low temperature freezes out the conformation which gives rise to these downfield shifts. In fact, the downfield parts of the phosphorus spectra of the two molecules are much more similar at low temperature than at 30°C. Incidentally, this also suggests that ITALY2's low-temperature conformation, although different by some measures, is still similar to the structure discussed above.

2.4 Discussion

2.4.1 The structures of ITALY2 and $tRNA_f^{Met}$ and of METAL2 and $tRNA_m^{Met}$

Technical limitations prevent us from studying the structure of tRNAs directly. ITALY2 and METAL2 differ from $tRNA_f^{Met}$ and $tRNA_m^{Met}$ in several respects: they are only fragments of the natural molecules, they are unmodified and they are studied in a simple magnesium-free buffer rather than in the cell.

Several lines of evidence give us confidence that stem-loops accurately reproduce the properties of the anticodon arm in intact tRNA. Most importantly, oligonucleotide stem-loops similar to ITALY2 are able to form initiation complexes (Hartz et al. 1990). Stem-loops also reproduce, at least qualitatively, the S1 nuclease cleavage pattern seen in initiator tRNAs ((Hartz et al. 1990) and the present work). Furthermore, the global structure of tRNA isolates the anticodon loop from other parts of the molecule. If the rest of the tRNA affects anticodon loop structure, it can only do so through a subtle perturbation of the structure of the stem. The anticodon stem in crystal structures of tRNA is unremarkable A-form RNA, and every indication is that this is true of ITALY2 as well.

Initiator tRNAs are modified to a lesser extent than other tRNAs. While the special nature of the anticodon stem-loop is universally conserved, the few modifications that it has are not. The only modification in the anticodon arm of most eukaryotic initiator tRNAs is a t^6A modification of A37. This has been shown to strengthen the adjacent anticodon-codon base pair and to reduce wobble at that position (Björk et al. 1987). It is also present in prokaryotic and eukaryotic elongator methionine tRNAs. It is not found in prokaryotic initiator tRNAs, which can recognize alternate initiation codons through wobble base pairs to the first position of the codon. Since prokaryotic $tRNA_f^{Met}$ can initiate translation in a eukaryotic system (Wagner et al. 1989), t^6A cannot be required for initiation in either kingdom.

The only modification of the anticodon arm of prokaryotic $tRNA_f^{Met}$ is a 2'-O-methylation of C32. This may protect the C32-U33 phosphodiester bond from spontaneous hydrolysis (Dock-Bregeon and Moras 1987). The 2' hydroxyl of C32 is not involved in any hydrogen-bonding interactions in the structure of ITALY2. The Cm at the same position of yeast $tRNA^{Phe}$ promotes a magnesium-induced conformational transition (Guenther et al. 1992), but the base-pairing interactions involved in that transition are not compatible with the sequence of the anticodon loop of methionine tRNAs. Since unmodified oligonucleotides can form initiation complexes in a prokaryotic system, neither t^6A nor Cm can be required for initiation.

Elongator methionine tRNAs are more extensively modified in both prokaryotes and eukaryotes. In addition to t^6A37, *E. coli* $tRNA_m^{Met}$ contains an ac^4C34 and a $\psi39$. Yeast $tRNA_m^{Met}$ has a $\psi31$:$\psi39$ mismatch. Like t^6A, all of these modifications improve stacking of the modified nucleotide on adjacent bases. ac^4C34 is known to reduce misreading of the AUA codon for isoleucine (Stern and Schulman 1978). Most tRNAs have ψ rather than U at the base of the anticodon loop in positions 37, 38 and 39 (Grosjean et al. 1995). As seen by comparison of the structures of natural and unmodified $tRNA^{Gln}$, which also has ψ in position 39, ψ stabilizes RNA structure by water-mediated contacts to the N and N-1 phosphates (Arnez and Steitz 1994). ψ is observed to increase stacking in small oligonucleotides (Davis 1995). The modifications in $tRNA_m^{Met}$ might actually make its structure more similar to $tRNA_f^{Met}$ than the structure of METAL2 is to that of ITALY2.

Finally, the structure of ITALY2 was determined under conditions rather different from those in the cell. In particular, the NMR buffer contained no magnesium. Magnesium is known to bind stably to the anticodon loop of $tRNA^{Phe}$, where it induces a conformational change (Salemink et al. 1981; Labuda and Porschke 1982; Striker et al. 1989). However, no magnesium ions were located in the crystal structures of $tRNA_f^{Met}$ (Woo et al. 1980) and $tRNA_i^{Met}$ (Basavappa and Sigler 1991), although this could be due to the insufficient resolution of the data. The ^{31}P spectrum of $tRNA_f^{Met}$ was almost unchanged by the removal of magnesium (Gorenstein and Goldfield 1982). A single peak attributed to the anticodon loop of $tRNA_f^{Met}$ was shifted less far downfield in the absence of magnesium, but still had the downfield shift indicative of a conformation other than gauche$^-$/gauche$^-$. Furthermore, the NMR structure of the $tRNA^{Phe}$ anticodon domain (Clore et al. 1984), determined in the absence of magnesium, was the same as that in the crystal structure including magnesium ions. Recall also that the S1 cleavage patterns of $tRNA_f^{Met}$ (Wrede and Rich 1979) and ITALY2 (present work) are the same in the presence and absence of magnesium.

2.4.2 The structure of ITALY2

ITALY2 is composed of familiar structural elements. The A-form helical stem is of course the fundamental unit of RNA secondary structure. The U-turn in its loop is found not only in the anticodon and TψC loops of tRNAs but in other hairpin loops and in the hammerhead ribozyme (see below). The entire loop is essentially that of elongator tRNAs.

The loop of ITALY2 is not stabilized by non-Watson-Crick base pairing, as are many hairpin and internal loops in RNA. The only potential hydrogen bonds are the two which stabilize the U-turn and one which could form between C32 and A38. The lack of hydrogen bonds involving the Watson-Crick faces of the bases of the anticodon is explained by the functional constraint that they must be able to base-pair with cognate codons.

The structure of the loop is maintained, however, by extensive base-base stacking interactions. C32 and U33 stack on the 5' side of the stem, and C34 through A38 stack on the 3' side. Only between U33 and C34 is there a break. Although the face of U33 that would stack on C34 in A-form RNA is relatively exposed, it is close to the phosphate of A35, as it is in the structure of $tRNA^{Phe}$ (Quigley and Rich 1976). Stacking is also poor between C34 and A35 and between U36 and A37, where the backbone is kinked by the C2'-endo riboses of C34 and U36 respectively.

In the crystal structure of $tRNA_f^{Met}$, U33 was flipped out into solution, an observation that was the basis of the "uridine swivel hypothesis" (Woo et al. 1980). Here it is stacked on C32 and hydrogen-bonded to other loop residues. A37, which was flipped out into solution in the crystal structure of yeast $tRNA_i^{Met}$ (Basavappa and Sigler 1991) as reported in (Dirheimer et al. 1995), is likewise stacked inside the loop. Nor are A37 and A38 flipped out like the equivalent bases of the TψC loop, as suggested by (Dirheimer et al. 1995). As discussed in detail below, the conformation of the loop of ITALY2, and by extension the anticodon loop of initiator tRNAs, is not greatly different from that of the anticodon loop of elongator tRNAs.

2.4.3 The structures of ITALY2 and METAL2

The close similarity of their chemical shifts and NOE connectivities indicates that the time-averaged structures of ITALY2 and METAL2 are nearly identical. They differ, however, in their dynamic behavior. The structure of ITALY2 is a rigid one, the non-A-like properties of which are concentrated in the U-turn nucleotides and the C2'-endo ribose of U36. METAL2 samples more frequently than does ITALY2 conformational states in which the backbone in the loop is extended, riboses throughout the loop have C2'-endo character and phosphate dihedral angles are closer to A-form. Because the difference between ITALY2 and METAL2 is one of dynamics, not of average structure, it would not be evident from a comparison of their NMR structures. For that reason, the structure of METAL2 was not determined.

What properties of the two molecules determine the balance between rigid and relaxed states? Nearest-neighbor parameters (Freier et al. 1986) predict that the A:U loop-closing base pair of METAL2 should stack less well on the underlying C:G pair (-1.8 kcal/mol) than should the loop-closing G:C of ITALY2 on its underlying G:C (-2.9 kcal/mol). The C32-A38 mismatch may also stack differently on the loop-closing G:C and A:U pairs. If the bottom of the loop and the loop-closing base pair can unstack and open more easily in METAL2 than in ITALY2, the structure of the rest of the loop will be more flexible.

This conformational change is too fast to be seen by NMR. The chemical shifts, scalar couplings and NOEs observed are averaged over a sum of conformational states. At most a few hydrogen bonds (the C32:A38 mismatch and, in METAL2, the A31:U39 pair) are broken and reformed. This is very different from the conformational change seen in the DNA analog of the $tRNA^{Phe}$ anticodon loop, which involves formation and disruption of two Watson-Crick base pairs across the center of the loop. The open and closed states of that loop depend on the absence and presence of magnesium, so can be separately observed by NMR.

2.4.4 The structure and NMR properties of ITALY2

The three-dimensional structure of ITALY2 explains many of the unusual features of its NMR spectra: its exchangeable proton spectrum, its NOE connectivity, its H1'-H2' coupling constants and its ^{31}P chemical shifts.

The imino protons of U33 and U36 are not observed at room temperature. U36 is in the anticodon and pointed into solution, so there can be no hydrogen-bonding interactions to protect its imino proton from exchange. Although U33H3 is within hydrogen-bonding distance of U36PO2 in the calculated structure, it too is not observed. This is not surprising, since the C2'-endo character of the anticodon riboses indicates that the structure of the anticodon is flexible. The U33H3-U36PO2 hydrogen bond must not be sufficiently stable to protect U33H3 from exchange.

The nonsequential U33H1'-A35H8 NOE, and the absence of the sequential U33H1'-C34H6 NOE, result from the sharp turn between U33 and C34, which moves C34's base away from and A35's base closer to U33's ribose. The relative orientation of the riboses of U33 and C34 also changes, so that the usual sequential H2' to H1' is not observed but there is an unusual "backwards" U33H1'-C34H2' NOE.

Based on their observed $^3J_{H1'-H2'}$s, U36's ribose was constrained to be C2'-endo and the riboses of C34, A35 and A37 were not constrained. In the calculated structures, A35 and A37 are C3'-endo, which allows A35 to stack on U36 and A37 to stack on A38. However, in a dynamic structure these bases might easily unstack and their riboses take on C2'-endo character. In most of the calculated structures, C34's ribose is C2'-endo. This may reflect a structural requirement for C34 to span a longer distance in the U-turn (6.8 Å from U33P to C34P in the present structure) than does a nucleotide in an A-form helix

(5.8 Å). On the other hand, it may reflect the reduced structural constraints of its position at the end of the anticodon stack. The C2'-endo character of the terminal riboses, of course, results from their exposed position at the end of the stem.

The ^{31}P chemical shift which is the farthest downfield in the molecule is that of C34, whose α dihedral angle is gauche$^+$ rather than the A-form gauche$^-$. The other two phosphates with downfield shifts, those of A35 and A37, are gauche$^-$/gauche$^-$. However, both α dihedral angles are unusually far towards trans, and, while C34's ζ has the A-form value, U36's ζ is the farthest towards trans in the molecule. It is also worth nothing that while the ε dihedral angle is trans in A-form RNA, the ε dihedral angles of C34 and U36 are the two most gauche$^-$ in the molecule. This may affect the chemical shifts of their associated phosphates.

The correlation between structure and ^{1}H chemical shifts is less straightforward. As usual in NMR constructs synthesized by T7 RNA polymerase, the 5' terminal ribose protons are among the most downfield in the molecule, possibly due to the absence of ring current shifts that would be produced by a base on the 5' side. The 3' terminal ribose protons are the most upfield, due to the absence of a 3' phosphate. However, neither of these factors account for the upfield chemical shift of C34H2', which is in a structural environment much like that in a 5' terminal nucleotide. A37H2' is shifted downfield, perhaps because the base of U36 is poorly stacked on A37 so exerts a lesser ring current shift on its H2'. A35H3' is also shifted downfield, but since A35H2' is not, poor stacking of C34 on A35 cannot be the explanation. In fact, A35H3' is in the plane of the base of U33, so should be shifted upfield. Finally, the cause of the upfield shift of the H1' of C39, whose conformation is by every indication A-form, is a mystery.

2.4.5 The structures of ITALY2 and $tRNA^{Phe}$

Is the loop of ITALY2, and by extension the anticodon loop of initiator tRNA, the same as the anticodon loop of elongator tRNAs? Superposition (figure 2.3.12) shows that the loops of ITALY2 and $tRNA^{Phe}$ (Hingerty et al. 1978) are identical except for nucleotides C34-U36. Nonetheless, those nucleotides of ITALY2 form a U-turn which is qualitatively similar to that in $tRNA^{Phe}$. Note that while many backbone dihedral angles in ITALY2 differ from those in $tRNA^{Phe}$, backbone dihedral angles in different structures of $tRNA^{Phe}$ differ from each other.

The only large difference between the loops of ITALY2 and $tRNA^{Phe}$ is the kink in ITALY2's backbone caused by the C2'-endo conformation of U36's ribose and the resulting reorientation of the bottom of the loop. None of the crystal structures of $tRNA^{Phe}$ (Hingerty et al. 1978; Holbrook et al. 1978; Westhof and Sundaralingam 1986) have any C2'-endo riboses in the anticodon loop. The NMR structure of the $tRNA^{Phe}$ anticodon domain (Gronenborn and Clore 1995) likewise has only C3'-endo riboses, as observed by the absence of H1'-H2' couplings.

While these differences between the structure of ITALY2 and the structures of $tRNA^{Phe}$ may reflect the difference in their sequences, there are two confounding factors. The first is that while all of the $tRNA^{Phe}$ structures studied were fully modified, ITALY2 is not. As discussed above, the modifications in the initiator tRNA anticodon loop are expected to increase stacking and decrease mobility. The second is the differences in experimental conditions. The $tRNA^{Phe}$ anticodon domain NMR structure was solved at a much higher ionic strength (0.5 M NaCl) than that used here (150 mM KCl + NaCl). In the crystal structures, C34s in the anticodon loops of symmetrically related tRNAs stack on one another (Holbrook et al. 1978). All of these factors should favor stacked, compact conformations. The anticodon loop of initiator tRNA may thus be even more similar to that of elongator tRNA than their crystal and NMR structures indicate. This is underscored by the fact that the riboses in METAL2's loop have even more C2'-endo character than do those in ITALY2's loop, so must not arise from the unique stem sequence of initiator tRNAs.

2.4.6 The structures of ITALY2 and other RNA hairpin loops

U-turns and similar sharp turns are found in many RNAs other than tRNA anticodon loops. Table 2.4.1 lists RNA hairpin loops which contain sharp turns and summarizes some of their properties. Some have U-turns which are nearly identical to that in tRNA anticodon loops, including the TψC loop, also found in tRNAs (Jack et al. 1976), and the hexaloop from the L11 binding site in 23S rRNA (Huang et al. 1996; Fountain et al. 1996). A bona fide U-turn occurs at the active site of the hammerhead ribozyme (Pley et al. 1994b; Scott et al. 1995). An octaloop from the T4 gp43 mRNA (Mirmira and Tinoco 1996a) and a tetraloop mutant of it (Mirmira and Tinoco 1996b) have surprisingly similar structures, with a sharp turn after the third and first loop nucleotides respectively. In contrast to these loops is the TAR hairpin (Jaeger and Tinoco 1993), whose six-nucleotide loop has a smoothly turning backbone and no extreme phosphorus chemical shifts.

Even smaller loops can – or must – have sharp turns similar to the U-turn. It has been observed (Jucker and Pardi 1995a) that the GNRA tetraloop has a sharp turn at the α dihedral angle of the phosphate after the first G. Furthermore, that G interacts with the phosphate of nucleotide N+3 just as

RNA molecule	Loop sequence	Location of turn	Downfield ^{31}P shift	Non-A-like NOEs	C2'-endo riboses	Exposed bases
Loop D in 5S RNA[1]	$U_{88}CU$	?	C89P[2]	no U88→C89	U88-U90	?
UCU model[3]	U_7CU	between U8 and U9	yes; not assigned	no U8→C9	U7-U9	U7-U9
UUU model[4]	U_4UU	between U5 and U6	U5P	no U5→U6	U4-U6	U4-U6
UUCG tetraloop[5]	U_5UCG	U6 α, 180°	U6P	Note 6	U6, C7	U6
GNRA tetraloop[7]	G_5CAA, G_5AAA	A6 α, 148°[8]	N6P	no G5→N6 G5→A7	N6, A7	N6, A7
L11 BS hexaloop[9]	G_5UAAUA	A7 α, 180°	A7P	no U6→A7 U6→A8	U9	A7-A9
ITALY2[10]	$C_{32}UCAUAA$	C34 α, 119°	C34P	no U33→C34 U33→A35	C34-A37	C34-U36
TψC loop[11]	T_{54}ψCGmAUC	C56 α, 165°	?	no ψ55→C56 ψ55→G57[12]	none	C34-mA36
T4 gp43 octaloop[13]	$A_5AUAACUC$	A8 α, 141°	A8P	no U7→A8 U7→A9	none	A8-C10
T4 gp43 mutant[13]	$A_5GCAACCU$	A8 α, 169°	A8P	C7→A9	none	A8-C10

Table 2.4.1: RNA hairpin loops which contain sharp turns. The sequence number of the first nucleotide in each loop is indicated by a subscript. Only those downfield ^{31}P chemical shifts and non-A-like NOEs which bear on the similarity of each loop to a U-turn are indicated.

1 (Dallas et al. 1995).
2 (Dallas 1996). Observed, but not assigned, in (Zhang et al. 1989).
3 (Puglisi et al. 1990).
4 (Davis et al. 1993).
5 (Cheong et al. 1990; Varani et al. 1991; Allain and Varani 1995).
6 Although it has a sharp turn, the UUCG tetraloop is quite different from a U-turn, and has many non-A-like NOEs which reflect its unique structure.
7 (Heus and Pardi 1991; Legault and Pardi 1994).
8 Stated in (Jucker and Pardi 1995); measured from (Pley et al. 1994).
9 (Huang et al. 1996; Fountain et al. 1996). Note that the conformation of U9 is different in the two structures.
10 This work. Other tRNA structures containing U-turns are in (Hingerty et al. 1978; Holbrook et al. 1978; Westhof and Sundaralingam 1986; Clore et al. 1984; Westhof et al. 1988).
11 (Hingerty et al. 1978; Holbrook et al. 1978; Westhof and Sundaralingam 1986; Westhof et al. 1988).
12 Predicted by the crystal structure.
13 (Mirmira and Tinoco 1996a; Mirmira and Tinoco 1996b).

does the U in a U-turn. The downfield chemical shift of the turning phosphate was identified as a marker of unusual conformation (Legault and Pardi 1994). The same work pointed out unassigned downfield phosphates in the region of the hammerhead ribozyme which we now know contains a U-turn. The UUCG tetraloop also has a sharp turn after its first loop nucleotide, although it is not otherwise similar to a U-turn (Cheong et al. 1990). The third known tetraloop, CUUG, has a unique conformation involving sharp turns at the beginning and end of its loop, only one of which is signaled by a downfield phosphorus chemical shift (Jucker and Pardi 1995b). Sharp turns have been observed in several three-nucleotide loops as well.

All of the loops listed in table 2.4.1 whose structures are known in detail involve a trans α dihedral angle, known as a π turn (Saenger 1983). The turning phosphate always has a downfield chemical shift. The turn moves the base to the 3' side of the turn away from the ribose to its 5' side, so the sequential H1' to H6 or H8 NOE which would be seen in A-form RNA is absent. In the larger loops there is an NOE from the H1' of the ribose to the 5' side of the turn to the H6 or H8 of the second base to the 3' side. In the smaller loops this nonsequential NOE is absent. The π-turn is not necessarily associated with C2'-endo riboses. Smaller loops require the extended C2'-endo conformation to span the helix, while larger loops do not. Recall also that C2'-endo character depends on environment (section 2.4.1).

Many different loops can be based on a π-turn. tRNA anticodon loops have two unpaired bases on the 5' side and five on the 3' side. The TψC loop has one and two, the hammerhead two and two, the L11 binding site hexaloop two and four, the GNRA tetraloop one and three and the three-nucleotide loops two and one. Not all of these loops have a U before the turn; the hydrogen bond from the U imino to the N+3 phosphate appears to be dispensable. The U imino is not protected from exchange with solvent at room temperature in ITALY2 or METAL2 or in other loops containing U-turns (Clore et al. 1984; Huang et al. 1996; Fountain et al. 1996). Furthermore, as mentioned above, some eukaryotic initiator tRNAs have a C at position 33 rather than the "universal" U33. The N O2' to N+2 purine H6 and N HO2' to N+2 purine N7 hydrogen bonds seen in the $tRNA^{Phe}$ and ITALY2 anticodon loops and in the hammerhead ribozyme must also not be required, as the second position of the tRNA anticodon must accommodate any base. The anticodon loop of $tRNA^{Asp}$, which has a U in position 35, has an otherwise canonical U-turn (Westhof et al. 1988).

An interesting consequence of the π-turn is that the Watson-Crick edges of one or more bases following it are pointed out into solution. In the tRNA anticodon loop, these bases are the anticodon, poised to bind with a cognate codon. The bases following the U-turn in the hammerhead ribozyme are its active site. The exposed bases in the L11 binding site hexaloop and the T4 gp43 loops are thought to be important for interactions with the translation apparatus and T4 DNA polymerase respectively. The second and third bases of the GNRA tetraloop have been predicted (Michel and Westhof 1990; Costa and Michel 1995) and observed (Jaeger et al. 1994; Pley et al. 1994a) to engage in tertiary interactions.

2.4.7 Functional implications

The difference in dynamics between ITALY2 and METAL2 accounts for the difference in the ways in which they are cleaved by S1 nuclease. ITALY2 is rigid and highly structured, so is for the most part inaccessible to a nuclease which prefers to cleave single-stranded RNA. METAL2 is more flexible, so more of its loop is in a single-stranded-like state at least some of the time and accessible to the nuclease. The structural difference detected by S1 nuclease is a difference in dynamics, not in average conformation.

We already know that MetRS recognizes the sequence of the anticodon loop, and that it does not discriminate between $tRNA_f^{Met}$ and $tRNA_m^{Met}$. The K_ds for binding of *E. coli* MetRS to *E. coli* $tRNA_f^{Met}$ and $tRNA_m^{Met}$ are 1.5 and 1.3 μM respectively (Meinnel et al. 1993b). This is of course consistent with the finding that the structures of the anticodon loops of the two molecules are the same. Structure-based alignment indicates that MetRS is structurally very similar to, and probably binds tRNA in the same way as, glutamine tRNA synthetase (GlnRS) (Perona et al. 1991). GlnRS disrupts the anticodon loop of its substrate tRNA on binding (Rould et al. 1989). If MetRS does the same, the similarity of the K_ds with which it binds initiator and elongator tRNAs indicates that the stability of the two anticodon loops must be very similar.

A more important question is how the initiation complex discriminates between the anticodon domains of $tRNA_f^{Met}$ and $tRNA_m^{Met}$. Since there is no static structural difference between the anticodon loops of the two tRNAs, the initiation complex must detect the difference in loop flexibility or the sequence of the stem itself. The similarity of the affinities of MetRS for initiator and elongator tRNAs indicates that the difference between the stabilities of their anticodon loops is very small, so cannot be used to discriminate between them. The existence of mutations in $tRNA_f^{Met}$ which reduce initiation activity but which do not affect the pattern of S1 nuclease cleavage of the anticodon loop suggests that the initiation complex requires the unique anticodon stem sequence itself (Mandal et al. 1996).

If no special structure is required for the initiator tRNA to enter the P site directly, there is no reason to expect elongator tRNAs to adopt a different structure in the P site than they do in the A site. While the conformations of the anticodon loops of initiator or elongator tRNAs bound to the ribosome are not necessarily the same as that seen in crystal and solution structures, there is no reason to believe that they differ.

Appendices

Appendix 1: RNA sequencing and S1 cleavage

A1.1 General comments

Reagents are listed in the order in which they are added to the reaction. Mix well before adding enzyme. I always add enzymes last. Adding labeled RNA last would be slightly more convenient, but would increase the risk of contaminating one's stock of labeled RNA.

I did two sets of experiments, with 5' labeled synthetic oligonucleotides and with 3' labeled tRNAs. For 5' end labeling, I simply diluted my RNA to 100 pmol/µl and used it directly. For 3' end labeling, I guessed at a 50% yield, dephosphorylated 200 pmol of tRNA and resuspended the pellet after ethanol precipitation directly into the labeling buffer. There is no reason why one couldn't do things the other way around.

A1.2 Dephosphorylation

Materials (and amounts required)

dH$_2$O (a few µl)

10x NEBuffer #2 (a few µl)
  From NEB, included with enzyme

RNA, 100 pmol/µl (a few µl)

Calf intestine alkaline phosphatase (CIAP) (5 µl, 10 U/µl)
  From NEB. NEB recommends 0.1 U/pmol ends for 5' overhangs and 1 U/pmol ends for blunt ends or 3' overhangs, so one should use 100 U/100 pmol tRNA, or 10 µl. That's a lot. I used half that and saw no dimer bands in my subsequent ligations; the thrifty experimentalist may want to reduce the amount further.

25:24:1 phenol:chloroform:isoamyl alcohol equilibrated with TE (a few µl)
  See (Maniatis et al. 1982).

Chloroform (a few µl)

20 mg/ml glycogen (Boehringer Mannheim), 5 M NaCl, ice-cold ethanol and 70% ethanol for ethanol precipitation

Reaction

38 µl dH2O

5 µl 10x CIAP buffer

2 µl RNA

2.5 µl CIAP

☞ Incubate 30' at 37 C, add another 2.5 ul CIAP and incubate 60' at 37 C.

☞ Extract twice with PCA and twice with chloroform.

☞ Add 1 µl 20 µg/µl glycogen as a carrier and 5 µl 5 M NaCl as a salt. (Glycogen worked noticeably better for me than tRNA, and it does not interfere with any subsequent reactions.) Precipitate the RNA with 3 volumes of ice-cold ethanol, stopping the centrifuge as necessary to avoid heating. Wash the pellet with 70% ethanol and dry it for 15' in a Speed-vac on low heat. If the RNA is to be labeled next, just resuspend the pellet in the labeling reaction.

A1.3 5' end-labeling

Materials

dH$_2$O (a few µl)

10x polynucleotide kinase buffer (a few µl)
From NEB, included with enzyme

RNA (1 µl, 100 pmol/µl or in a pellet after dephosphorylating)
Synthetic RNA has no 5' phosphate and can be used directly. RNA purified from a natural source or made by T7 transcription probably needs to be dephosphorylated first; see below.

γ-^{32}P ATP (3 µl, 10 µCi/µl)

Polynucleotide kinase (1 µl, 10 U/µl)
From NEB

Formamide loading dye (a few µl)
See (Maniatis et al. 1982).

20% (for 17mers) or 10% (for tRNAs) polyacrylamide gel, 0.5x TBE, 0.8 mm thick (about 100 ml polyacrylamide stock for our hot lab rig)
See (Maniatis et al. 1982). Prerun overnight at 20 W constant power if possible and at least 1 hour at 50 W constant power just before loading in any case. An 0.4 mm gel will do if you happen to have one poured.

0.5x TBE
See (Maniatis et al. 1982).

Gel elution buffer (100 ml)

500 mM NH_4OAc	3.85 g
1 mM EDTA	200 µl 0.5 M stock
0.1% SDS	100 mg
pH 7	

Store at room temperature, as the NH_4OAc precipitates at 4°C.

Materials for ethanol precipitation as in A1.2, without 0.5M NaCl
TE, pH 7
See (Maniatis et al. 1982).

Reaction

13 µl dH2O

2 µl 10x kinase buffer

1 µl RNA

3 µl γ-^{32}P ATP

1 µl kinase

☞ Incubate 30' @ 37°C.

☞ Add 10 µl formamide loading dye & vortex. Heat 5' at 80°C or so if you think it is necessary to denature your RNA. It isn't for 17mers or 21mers.

☞ Load entire volume onto gel and run at 50 W constant power.

☞ Stop the gel when the bromphenol blue is either 1) 1/3 of the way down the gel, keeping labeled nucleotides in the gel and making cleanup easy, or 2) about to run off the gel, making cleanup difficult but maximizing resolution, whichever you think is more helpful.

☞ Take down gel, remove from plates and cover with plastic wrap. Expose gel to film for 1' with intensifying screen. Use Fuji film so you can use the bright safelight. Mark the gel and film and/or cut out part of the gel where the labeled bands are not so that the gel can be oriented on the film. and mark the position of the dye bands through the film. Develop the film. Realign the gel and film and cut out the bands.

☞ Cover the gel pieces with gel elution buffer (about 500 µl) and incubate overnight at room temperature.

☞ Remove the gel elution buffer from the gel pieces. Ethanol precipitate as above, but do not add salt.

☞ Resuspend the pellet in TE, pH 7, to a concentration of 40000 cpm/µl, measured by holding 0.5 µl of labeled material in a pipet tip next to the radiation counter. I got around 100 µl of labeled material at that concentration per labeling, enough for 100 to 200 gel lanes.

A1.4 3'-end labeling

Materials

dH_2O (a few µl)

10x T4 RNA ligase buffer (a few µl)
From NEB; included with enzyme

DMSO (a few µl)

100 mM ATP, pH 7 (a few µl)

0.1 µg/µl acetylated BSA (a few µl)

RNA (1 µl, 100 pmol/µl or in a pellet after dephosphorylating)
Synthetic RNA has no 5' phosphate and can be used directly. RNA purified from a natural source or made by T7 transcription probably needs to be dephosphorylated first; see above.

α-^{32}P pCp (12.5 µl, 10 µCi/µl)

T4 RNA ligase (4 µl, 10 U/µl)
From NEB

Materials for gel purification and ethanol precipitation as in A1.3

Reaction

6.5 µl dH_2O

4 µl 10x ligase buffer

4 µl DMSO

5 µl 100 mM ATP, pH 7

4 µl 0.1 µg/µl BSA

1 µl RNA

12.5 µl α-^{32}P pCp

4 µl ligase

☞ Incubate overnight at 4°C.

☞ Add 20 µl formamide loading dye and vortex.

☞ Gel-purify as above under A1.3.

A1.5 Sequencing and S1 cleavage

Materials

Labeled RNA (0.5 µl/lane, 40000 cpm/µl)
10000 cpm is adequate to visualize a sequencing lane exposed to film overnight (12-18 h) with an intensifying screen.

OH^- buffer (100 ml)

50 mM $NaCO_3$	0.53 g
pH 9.0 (for tRNAs) or 9.5 (for 17mers)	

T1 and U2 buffer (50 ml each)

33 mM Na citrate	0.97 g
1.7 mM EDTA	340 µl 0.5 M stock
7 M urea	42 g

Divide into two 50 ml portions.
T1 buffer: pH 5.0
U2 buffer: pH 3.5

T1 or U2 buffer + tRNA (a few µl)

475 µl T1 or U2 buffer
25 µl 20 µg/µl tRNA

T1 nuclease (a few µl)
Calbiochem's T1 comes as lyophilized powder containing 5000 units. Dissolve in 500 µl dH_2O to make a 10 U/µl stock. Dilute this with dH_2O to make a 0.1 U/µl (= "1/100x") working solution.

U2 nuclease (a few µl)
Calbiochem's U2 comes as lyophilized powder containing 100 units. Dissolve in 30 µl dH2O to make a 3 U/µl stock. Dilute this with dH_2O to make a 0.03 U/µl (= "1/100x") working solution.

Formamide loading dye (10 µl/lane)
See (Maniatis et al. 1982).

dH_2O (a few µl)

1 µg/µl tRNA (a few µl)

5x S1 cleavage buffer (100 ml)

125 mM NaOAc	1.7 g
250 mM KCl	1.86 g
25 mM $MgCl_2$	500 µl 5 M stock
5 mM $ZnCl_2$	500 µl 1 M stock
pH 4.5	

For experiments at pH 6.5 or 7.5, replace NaOAc with 125 mM Na cacodylate (1.73 g/100 ml). $ZnCl_2$ precipitates at pH 7.
For experiments without Mg^{++}, just omit the $MgCl_2$.
For experiments without Zn^{++}, omit the $ZnCl_2$ and use 100 mM $MgCl_2$.
For experiments without Mg^{++} or Zn^{++}, omit both $MgCl_2$ and $ZnCl_2$.
We can't use EDTA in the Mg^{++}-less or Zn^{++}-less buffers because

the other divalent metal is present, so I didn't bother to add EDTA to the Mg^{++}-less/Zn^{++}-less buffer.

S1 storage buffer (100 ml)

250 mM NaCl	1.46 g
10 mM NaOAc	0.14 g
50 mM $ZnCl_2$	5 ml 1 M stock
pH 4.5	
50% w/v glycerol	40 ml

For experiments without Zn^{++}, just omit the $ZnCl_2$.

S1 nuclease (a few µl)

Sigma's S1 comes at 450 U/µl in S1 storage buffer. Dilute with S1 storage buffer to approximately 1 U/0.5 µl for tRNAs or 1/16 U/0.5 µl (for 17mers), but plan on doing serial dilutions. For experiments without Zn^{++}, dilute similarly into S1 storage buffer without $ZnCl_2$.

S1 loading dye (10 µl/lane)

475 µl formamide loading dye
20 µl 100 mM ATP
5 µl 20 µg/µl tRNA

20% (for 17mers) or 15% (for tRNAs) polyacrylamide gel, 0.5x TBE, 0.4 mm thick, silanized plates (about 60 ml polyacrylamide stock for our hot lab rig)

See (Maniatis et al. 1982) Prerun overnight at 20 W constant power if possible and at least 1 hour at 50 W constant power just before loading in any case.

0.5x TBE

See (Maniatis et al. 1982).

Gel fixer (1 l/gel)

750 ml dH_2O
250 ml MeOH
100 ml HOAc

Reactions

OH^- ladder

4.5 µl OH- buffer pH 9.0 (for tRNAs) or 9.5 (for 17mers)
0.5 µl labeled RNA

☞ Incubate 12' @ 85°C (for tRNAs) or 92°C (for 17mers)
☞ Add 10 µl formamide loading dye to each reaction, vortex and keep on ice until loading.

T1 and U2 reactions

4 µl T1 or U2 buffer + tRNA
0.5 µl labeled RNA
0.5 µl T1 or U2

☞ Incubate 12' @ 55°C
☞ Add 10 µl formamide loading dye to each reaction, vortex and keep on ice until loading.

S1 reaction

2 µl water
1 µl 1 mg/ml tRNA
1 µl 5x S1 buffer
0.5 µl labeled RNA
0.5 µl enzyme (or S1 storage buffer for control lane)

☞ Incubate 5' (for 17mers) or 60' (for tRNAs) at 37°C.
☞ Add 10 µl S1 loading dye to each reaction, vortex and keep on ice until loading.

Gel

☞ Load 7 µl of each reaction per lane.
☞ Run gel until bottom dye is 2/3 of the way down the gel (17mers) or at the bottom (tRNAs).
☞ Fix and dry as usual.
☞ Expose to film overnight with intensifier screen.

Notes

☞ If you need to use more or less RNA (e.g. if your labeled material is a week or two old and you need to use 1 µl) adjust the buffer or water accordingly.
☞ All reactions are 5 µl total for convenience.
☞ RNases may be vortexed gently.

☞ RNases are not very temperature sensitive, but I keep them in a 4°C refrigerator or on ice.

A1.6 Sources

The dephosphorylation and labeling protocols are derived from Paul Richardson's protocols. The sequencing protocol is derived from an old protocol from PL Biochemicals, given to me by Paul Richardson. (This protocol gives conditions for *Phy. M.* and *B. cereus* enzymes, which I did not use.) The S1 cleavage protocol is derived from that in (Hartz et al. 1990) and references therein.

Appendix 2: Programs

The following programs were written by the author unless otherwise specified. All are available from the author. Most are written in the Perl language, the interpreter for which may be obtained without cost from (among other places) http://www.perl.com/.

A2.1: dm

Calculates the distances between pairs of protons in a PDB file. By default, only distances above 5 Å are printed; this limit may be changed with the -d option. By default, distances are printed regardless of residue number; a limit may be set with the -r option. 'dm -u' describes usage.

A2.2: dm2xplor

Takes as input output from dm and produces input to X-PLOR's "noe" application statement. The upper and lower bounds are set to 0.1 Å; this may be changed by editing the program.

A2.3: CHURN

CHURN ("Crosspeak Handling Utilities for RNA NMR") is an interdependent set of programs for managing NMR data. The core programs and data structures in CHURN were designed to support a variety of tasks including automated (or assisted) assignment. Time limitations have so far allowed the development to a useful level of only one planned aspect of CHURN, the conversion of NOE peak lists into X-PLOR-formatted distance constraints. Figure A2.3.1 shows the data structures with which CHURN presently concerns itself and the programs which transfer information between them.

The first of the two major data structures in CHURN is the spin-system database, or "ssd", which contains all the information known about a molecule. Molecular topology is encoded in a hierarchy of segments, residues and spins. Each level of the hierarchy has several properties (e.g. residues have types and spins have chemical shifts), one of which is a list of its constitutive elements at the next level of detail. The ssd also contains the secondary structure as a list of base pairs. The whole is stored in memory during use and dumped to disk as a binary file for storage.

The ssd structure is defined fully in the source code to the program seq2ssd, which is used to create an ssd. It reads a simple input file which describes a molecule's sequence and secondary structure and writes out an "empty" ssd, in which the molecular topology is fully defined but no chemical shifts are known. One can add chemical shift data by converting the ssd to a chemical shift table using the program ssd2table, editing the table with

a text editor, and incorporating the table back into the ssd with the inverse program, table2ssd.

The second major data structure is the crosspeak database, or "cpd", which describes a NMR dataset. It is simply a more structured version of the Felix peaks entity from which it is created by the program peaks2cpd. The source code of peaks2cpd fully describes the cpd structure. It contains a list of peaks, each of which contains a list of dimensions, each of which has a chemical shift and an assignment. Each peak also has an intensity score of s (strong), m (medium), w (weak), e (exchangeable) or a (absent). One indicates an intensity score in Felix by appending it to a crosspeak's assignment in either dimension. A cpd also contains spectral parameters (spectrometer frequency, sweep width, etc.) which are specified on the command line when running peaks2cpd. The program cpd2xplor reads one or more cpds and converts the intensity scores of peaks at multiple mixing times to distance constraints as described in section 2.2.7.1.

By way of example, here is the script used to generate distance constraints for ITALY2.

```
seq2ssd italy2.seq | table2ssd - italy2.table > italy2.ssd
peaks2cpd 499.9150269 4000 1271.06  3.741 2048 desolacion.peaks > desolacion.cpd
peaks2cpd 599.9494943 5000 1255.813 3.741 2048 corfu.peaks      > corfu.cpd
peaks2cpd 500.139032  4000 1267.92  3.741 2048 engenho.peaks    > engenho.cpd
check-cpd italy2.ssd desolacion.cpd corfu.cpd engenho.cpd
cpd2xplor italy2.ssd desolacion.cpd corfu.cpd engenho.cpd > noes.d2o.inp
```

desolacion, corfu and engenho are the names of 50, 100 and 300 ms NOESYs in D_2O. The first line constructs an ssd from the sequence and 2° structure description of ITALY2 (italy2.seq), incorporates into it the table of known chemical shifts (italy2.table) and saves it as italy2.ssd. The next three lines convert Felix peaks entities to cpds, including referencing information. The next checks the cpds for errors (see below), and the last combines the three cpds into distance constraints formatted for X-PLOR (noes.d2o.inp). It only remained to remove the NOEs involving only protons in the stem.

The fundamental purpose of CHURN is to move information between the molecule and the spectrum, that is, between the ssd and the cpd. At present there is only a single program which does so. check-cpd reads an ssd and a cpd and examines the cpd for consistency with the ssd. It checks for assignments whose residue type or number does not match that in the ssd, whose assignment is inconsistent with its residue type (e.g. g1h5) and so on. Where a chemical shift is known in the ssd, check-cpd flags assignments whose chemical shift does not match. The program prints a list of what it finds, which can be used to correct the assignments in Felix. The corrected peaks entity can be saved, reconverted to a cpd, re-checked, and so on until no errors are found. Crude though this procedure is, during the solution of the structure of ITALY2 it found scores of errors which might not have been found otherwise.

Further documentation is in the README included with CHURN, the usage messages which each program prints when run with the -u option and the source code to each program.

A2.4: ext_j

Written by Yangmee Kim in the laboratory of Dr. James Prestegard and published in (Kim and Prestegard 1989). Calculates $^3J_{H1'-H2'}$ from in-phase and antiphase peak separations, correcting for the effects of finite linewidth. A copy of the Fortran source code is available from the present author.

A2.5: j2nu

Takes as input a list of $^3J_{H1'-H2'}$ in Hz, one per line, and produces as output a table of $f_{C3'-endo}$, P, and ν_0-ν_4. $P_{C3'-endo}$, $P_{C2'-endo}$ and τ_m are set to the values given in section 2.2.8.5, but may be changed by editing the program.

A2.6: dihedrals2xplor

Takes as input a succinct table of dihedral angles and produces input to X-PLOR's "constraints dihedral" application statement. Angles may be omitted to use default A-form values. The wide array of input formats which are accepted (ranging from a simple list of residue types to the table shown in A4.4) is described in the source code, as are some fine points of program function. 'dihedrals2xplor -u' describes usage.

A2.7: ip

Runs a command "in place"; that is, runs a command on an input file and moves the output in place of the input file. 'ip foo bar' is equivalent to 'foo bar > baz; mv baz foo' with some error checking. It is mentioned here only because it is used in several scripts in appendix 3.

A2.8: repdb

Interconverts flavors of the PDB format. 'repdb -u' describes usage. This program was written with the intention of being upgraded to convert to and from every version (or perversion) of the PDB format the author encountered. As it happened, there was a regular need only to convert from Insight to X-PLOR (the default conversion) and from X-PLOR to Midas ('repdb -im -ox').

A2.9: cdih_measure

Written by Jon Lapham <lapham@tecate.chem.yale.edu> in the laboratory of Dr. Donald Crothers. Measures the dihedral angles of an RNA molecule in a PDB file. Minor modifications which correct an error in PDB file format parsing are available from the present author.

A2.10: average-dihedrals

Takes as input multiple lists of dihedral angles calculated by cdih_measure and produces as output a list of averaged dihedral angles and their standard deviations.

Appendix 3: X-PLOR scripts

A3.1: xplor

This csh script, intended to be named "xplor", sets up for and runs the actual X-PLOR program. It is used by the scripts below whose names end in ".run".

```
#!/bin/csh -f

# xplor (wrapper) v. 21 Apr 1995, Dave Schweisguth <dcs@proton.chem.yale.edu>
# Moore lab version

### Setup

# Set priority

nice +19

# Initialize variables

set whatami         = `basename $0`
set arch            = `uname -m`
set version         = ''
set versiond        = 3.820
set lib             = /usr/local/lib/xplor
unset seed
set seedfile        = xplor.seed        #.$$

# Process switches

while ("$1" =~ -* || "$1" =~ +*)        # Can't use [-+]. Grrr.
    switch ("$1")
    case +s:
        unset seed
        breaksw
    case -s:
        set seed
        breaksw
    case -u:
        set exit=0
        goto usage
        breaksw
    case -v:
        if ("$2" == "") then
            echo "${whatami}: $1 requires a version number for an argument."
            exit 1
        else
            set version = "$2"
            shift
            breaksw
        endif
    default:
        echo "${whatami}: $1 is not an option."
        set exit=1
        goto usage
        breaksw
    endsw
    shift
end

# Set version to default if not already set
```

```
if ("$version" == '') set version = $versiond
if (! -d $lib/$version) then
    echo "${whatami}: Xplor version $version is not available."
    exit 1
endif

### Source ulogin.com

set ulogin = $lib/$version/sgi/ulogin.com
if (-r $ulogin) then
    source $ulogin
else
    echo "${whatami}: Can't source $ulogin\!"
    exit 1
endif

### Decide on architecture

switch ("$arch")
case '':
    echo "${whatami}: Can't determine machine architecture\!"
    exit 1
case IP6:
    set executable = xplor.exe.R3000
    breaksw
case IP20:
    set executable = xplor.exe.R4000
    breaksw
default:
    echo "${whatami}: Don't know which executable to use for $arch architecture."
    exit 1
    breaksw
endsw

### Run Xplor

if ($?seed) then
    perl -e 'srand(time ^ $$); print rand(127773);' > $seedfile
        # Tell me why I chose 127773 and win a prize
#    setenv SEEDFILE $seedfile
    $OBJ/$executable $*
    rm -f $seedfile
else
    exec $OBJ/$executable $*
endif

exit

### Subroutines

# Usage

usage:
    cat <<EOE
Usage: $whatami [-s] [-v version] [-u]
-s          Generate temporary seed file, to be accessed from within Xplor
                with "set seed=@xplor.seed end"
+s          Don't generate temporary seed file (default)
-v version  Use that version of Xplor (default: $versiond)
-u          This message
EOE
    exit $exit
```

A3.2: make-template.run

This csh script runs the X-PLOR script make-rsf.inp once and make-template.inp ten times and summarizes the results. It can be run directly or used as a DQS job. The programs ip and repdb are described in A2.7 and A2.8.

```
#!/bin/csh -f
#$ -cwd -o make-template.run.o -e make-template.run.e -m bae -M
dcs@proton.chem.yale.edu

xplor < make-rsf.inp > make-rsf.out

set i=1
while ($i <= 10)
    set ii=`nawk 'BEGIN { printf "%02d", '$i' }'`

    xplor -s < make-template.inp > make-template.$ii.out
    ip 'repdb -ix -om' template.pdb
    mv template.pdb template.$ii.pdb

    @ i++
end

perl -ne 'print if /^Ener/ .. /^ X-PLOR/' make-template.??.out > make-template.summary
```

A3.3: make-rsf.inp

```
! make-rsf.inp -- Build RNA structure file
! Dave Schweisguth <dcs@proton.chem.yale.edu>, 11 Apr 1996
! Derived from generate/generatedna.inp

set echo=off message=off end            ! Normal use
!set echo=on message=all end            ! Auxiliary file debugging

topology @XPLORLIB:toppar/dna-rna-allatom.top end

set echo=on message=all end

segment
    name=a
    chain
        link nuc head - * tail + * end
        first 5pho tail + * end         ! Patches must precede sequence
        last 3ter head - * end
        sequence
            gua gua cyt ade gua gua gua
            cyt uri cyt ade uri ade ade
            cyt cyt cyt uri gua cyt cyt
        end
    end
end

write structure output=rsf.inp end

stop
```

A3.4: make-template.inp

```
! make-template.inp -- Build extended structure for local geometry reference
! Dave Schweisguth <dcs@proton.chem.yale.edu>, 27 Apr 1996
! Derived from nmr/generate_template.inp

set seed=@xplor.seed end                ! Use xplor -s

set echo=off message=off end            ! Normal use
!set echo=on message=all end            ! Auxiliary file debugging

structure @rsf.inp end
parameter @XPLORLIB:toppar/dna-rna-allatom.par end

set echo=on message=all end

vector ident (x) (all)                  ! Set X coordinate to atom number(!)
vector do (x=x/10) (all)
vector do (y=random(0.5)) (all)
vector do (z=random(0.5)) (all)
vector do (fbeta=50) (all)              ! Friction coefficient, in 1/ps
vector do (mass=100) (all)              ! Heavy masses, in amu

parameter
    nbonds
        cutnb=5.5
        rconst=20
        nbxmod=-2
        repel=0.9       ! Repulsive non-bonded energy only
        wmin=0.1
        tolerance=1
        rexp=2          ! default
        irexp=2         ! default
        inhibit=0.25
    end
end

! Parameter set may or may not contain dihedral energies

flags exclude * include bond angle dihe vdw end
minimize powell nstep=200 nprint=10 end

flags include impr end
minimize powell nstep=200 nprint=10 end

dynamics verlet
    nstep=200
    timestep=0.001      ! default
    iasvel=maxwell
    firsttemp=300
    tcoupling=true
    tbath=300
    nprint=50
    iprfrq=0
end

parameter
    nbonds
        rconst=2
        nbxmod=-3
        repel=0.75      ! Repulsive non-bonded energy only
    end
```

```
end

minimize powell nstep=400 nprint=25 end

dynamics verlet
    nstep=1000
    timestep=0.005
    iasvel=maxwell
    firsttemp=300
    tcoupling=true
    tbath=300
    nprint=100
    iprfrq=0
end

! H-building won't actually place protons, since we're using a fully protonated
!   topology/parameter set, but it may improve their positions
! We're using "nbonds repel", so the "elec" flag is irrelevant

flags exclude vdw end
vector do (mass=1) (name h*)
hbuild selection=(name h*) phistep=360 end

flags include vdw end
minimize powell nstep=800 nprint=50 end

! Analyze and write out the final structure.
! Total energy should be no more than about 100. If it is, it is probably
! because one or more purines are trapped in a crumpled conformation.

flags exclude * include bond angle dihe impr vdw end

print threshold=0.05 bonds
evaluate ($rms_bond=$result)
evaluate ($v_bond=$violations)
print threshold=5 angles
evaluate ($rms_angl=$result)
evaluate ($v_angl=$violations)
print threshold=15 dihedrals
evaluate ($rms_dihe=$result)
evaluate ($v_dihe=$violations)
print threshold=5 impropers
evaluate ($rms_impr=$result)
evaluate ($v_impr=$violations)

set echo=off message=off end

display Energy:      bond $bond, angle $angl, dihedral $dihe,
display              improper $impr, VdW $vdw, total $ener
display RMSD:        bond $rms_bond, angle $rms_angl, dihedral $rms_dihe,
display              improper $rms_impr
display Violations:  bond $v_bond, angle $v_angl, dihedral $v_dihe,
display              improper $v_impr

set echo=on message=all end

write coordinates output=template.pdb end

stop
```

A3.5: dgsa.run

This csh script runs the X-PLOR scripts dg.inp and dgsa.inp 30 times each and summarizes the results. See also A3.2.

```
#!/bin/csh -f
#$ -cwd -o dgsa.run.o -e dgsa.run.e -m bae -M dcs@proton.chem.yale.edu

set i=1
while ($i <= 30)
    set ii=`nawk 'BEGIN { printf "%02d", '$i' }'`

    xplor -s < dg.inp > dg.$ii.out
    xplor -s < dgsa.inp > dgsa.$ii.out
    ip 'repdb -ix -om' dg.pdb dgsa.pdb
    mv dg.pdb dg.$ii.pdb
    mv dgsa.pdb dgsa.$ii.pdb

    @ i++
end

perl -ne 'print if /^Ener/ .. /^ X-PLOR/' dgsa.??.out > dgsa.summary
perl -ne 'print if /0.2 noe/ .. /NOEPRI/' dgsa.??.out > dgsa.summary.noes
perl -ne 'print if /5 cdih/ .. /RMS deviation/' dgsa.??.out > dgsa.summary.cdih
qsub3 -q medium-phe refine.run
```

A3.6: dg.inp

```
! dg.inp -- Full metric matrix distance geometry
! Dave Schweisguth <dcs@proton.chem.yale.edu>, 7 May 1996
! Derived from nmr/dg_full_embed.inp

set seed=@xplor.seed end            ! Use 'xplor -s'

set echo=off message=off end        ! Normal use
!set echo=on message=all end        ! Auxiliary file debugging

structure @rsf.inp end
parameter @XPLORLIB:toppar/dna-rna-allatom.par end
! Uncomment the following to define rigid groups (which seem to do nothing)
!coordinates disposition=main @template.pdb         ! END in PDB file

noe nres=1000 @noes.stem.inp @noes.loop.inp @base-pairs.inp end
restraints dihedral nassign=300 @dihedrals.inp end

set echo=on message=all end

! Include NOE and dihedral restraints in mmdg

flags exclude * include bond angle dihe impr vdw noe cdih end
parameter nbonds repel=0.8 end end  ! Scale van der Waals radii

mmdg
!    @groups.inp                    ! Uncomment to define rigid groups
                                    ! (which seem to do nothing)
    selection=(all)
    storebounds
end

! Loop until embedding is successful. Once usually suffices.

evaluate ($embedded="false")
while ($embedded="false") loop embed
    mmdg recallbounds end
end loop embed

! Scale coordinates by $dgscale (defined by mmdg)

vector do (x = x * $dgscale) (known)
vector do (y = y * $dgscale) (known)
vector do (z = z * $dgscale) (known)

! Regularize

flags exclude * include dg end
minimize powell nstep=100 drop=10 nprint=25 end

write coordinates output=dg.pdb end

stop
```

A3.7: dgsa.inp

```
! dgsa.inp -- Clean up the output of dg.inp using simulated annealing
! Dave Schweisguth <dcs@proton.chem.yale.edu>, 22 Jul 1996
! Derived from nmr/dgsa.inp

evaluate ($init_t=3000)         ! Temperature for constant-temperature MD
evaluate ($high_steps=6000)     ! Number of steps at high temp
evaluate ($high_timestep=0.002) ! Time of each MD step at high temp
evaluate ($final_t=300)         ! Final temperature
evaluate ($cool_steps=7500)     ! Number of steps for cooling
evaluate ($cool_timestep=0.002) ! Time of each MD step when cooling
evaluate ($tempstep=50)         ! Degree increment for cooling

set seed=@xplor.seed end        ! Use 'xplor -s'

set echo=off message=off end    ! Normal use
!set echo=on message=all end    ! Auxiliary file debugging

structure @rsf.inp end
parameter @XPLORLIB:toppar/dna-rna-allatom.par end

noe nres=1000 @noes.stem.inp @noes.loop.inp @base-pairs.inp end
restraints dihedral nassign=300 @dihedrals.inp end
evaluate($planarweight=50)      ! AAFS' choice
restraints planar @planars.inp end

set echo=on message=all end

vector do (fbeta=10) (all)      ! Friction coeff. for MD heatbath, in 1/ps.
vector do (mass=100) (all)      ! Uniform heavy masses to speed MD.

noe                             ! Parameters for NOE effective energy term.
    ceiling=1000
    averaging  * cent
    potential  * square
    sqconstant * 1
    sqexponent * 2
    scale      * 100            ! Constant NOE scale throughout the protocol.
end

parameter                       ! Parameters for the repulsive energy term.
    nbonds
        repel=0.5               ! Initial value for repel--modified later.
        rexp=2
        irexp=2
        rcon=1
        nbxmod=-2               ! Initial value for nbxmod--modified later.
        wmin=0.01
        cutnb=4.5
        ctonnb=2.99
        ctofnb=3
        tolerance=0.5
    end
end

! Test for the correct enantiomer; if you want to bypass this test because
! the substructures were tested previously, simply remove the -1 from the
! next statement.

for $image in (1 -1) loop imag
```

```
coordinates initialize end
coordinates @dg.pdb                     ! END in PDB file
vector do (x=x * $image) (known)
vector identity (store1) (not known)    ! Set store1 to unknowns.

! Create local ideal geometry by template fitting;
! this also takes care of unknown atoms.

coordinates copy end        ! Store current coordinates in comparison set.
coordinates initialize end
coordinates @template.pdb               ! END in PDB file

set echo=off message=off end
for $id in id (tag) loop fit            ! Loop over residue tags.

    ! LSQ fitting using known coordinates.
    coordinates fit selection=(byresidue (id $id) and not store1) end

    ! Store fitted template coordinates for this residue.
    coordinates copy selection=(byresidue (id $id)) end

end loop fit
set echo=on message=all end

coordinates swap end

! Minimization of bonds, VDWs, and NOEs.

flags exclude * include bond vdw noe cdih plan end
constraints interaction (all) (all) weights * 1 vdw 20 end end
parameter nbonds nbxmod=-2 repel=0.5 end end
restraints dihedral scale=20 end
minimize powell drop=10 nstep=300 nprint=10 end

! Include angles.

flags include angl end
minimize powell drop=10 nstep=300 nprint=10 end

! Dynamics, slowly introducing dihedral and improper angles

flags include dihe impr end

evaluate ($nstep1=int($high_steps/8))   ! Dynamics runs 1, 2, 3 and 5
evaluate ($nstep2=int($high_steps/2))   ! Dynamics run 4

constraints inter (all) (all) weights * 0.1 dihe 0.05 impr 0.05 vdw 20 end end
dynamics verlet
    nstep=$nstep1 time=$high_timestep iasvel=maxwell firstt=$init_t
    tcoup=true tbath=$init_t nprint=100 iprfrq=0
end

constraints inter (all) (all) weights * 0.2 dihe 0.1 impr 0.1 vdw 20 end end
dynamics verlet
    nstep=$nstep1 time=$high_timestep iasvel=current firstt=$init_t
    tcoup=true tbath=$init_t nprint=100 iprfrq=0
end

constraints inter (all) (all) weights * 0.2 dihe 0.2 impr 0.2 vdw 0.01 end end
parameter nbonds repel=0.9 end end
dynamics verlet
    nstep=$nstep1 time=$high_timestep iasvel=current firstt=$init_t
    tcoup=true tbath=$init_t nprint=100 iprfrq=0
end
```

```
    constraints inter (all) (all) weights * 0.4 dihe 0.4 impr 0.4 vdw 0.003 end end
    parameter nbonds nbxmod=-3 end end
    dynamics verlet
        nstep=$nstep2 time=$high_timestep iasvel=current firstt=$init_t
        tcoup=true tbath=$init_t nprint=100 iprfrq=0
    end

    constraints inter (all) (all) weights * 1 dihe 1 impr 1 vdw 0.003 end end
    dynamics verlet
        nstep=$nstep1 time=$high_timestep iasvel=current firstt=$init_t
        tcoup=true tbath=$init_t nprint=100 iprfrq=0
    end

    if ($image=1) then
        vector do (store7=x) (all)          ! Store first image in stores.
        vector do (store8=y) (all)
        vector do (store9=z) (all)
        vector do (store4=vx) (all)
        vector do (store5=vy) (all)
        vector do (store6=vz) (all)
    end if

end loop imag

! Establish the correct handedness of the structure.

energy end
evaluate ($e_minus=$ener)
coordinates copy end
vector do (x=store7) (all)
vector do (y=store8) (all)
vector do (z=store9) (all)
energy end
evaluate ($e_plus=$ener)
if ($e_plus > $e_minus) then
    evaluate ($hand=-1)
    coordinates swap end
else
    evaluate ($hand=1)
    vector do (vx=store4) (all)
    vector do (vy=store5) (all)
    vector do (vz=store6) (all)
end if

! Increase VDW interaction and cool.

restraints dihedral scale=800 end

evaluate ($bath=$init_t)
evaluate ($ncycle=($init_t-$final_t)/$tempstep)
evaluate ($nstep=int($cool_steps/$ncycle))

evaluate ($ini_rad=0.9)
evaluate ($fin_rad=0.75)
evaluate ($radius=$ini_rad)
evaluate ($radfact=($fin_rad/$ini_rad)^(1/$ncycle))
evaluate ($ini_con=0.003)
evaluate ($fin_con=4.0)
evaluate ($k_vdw=$ini_con)
evaluate ($k_vdwfact=($fin_con/$ini_con)^(1/$ncycle))

evaluate ($i_cool=0)
while ($i_cool < $ncycle) loop cool
```

```
    evaluate ($i_cool=$i_cool+1)
    evaluate ($bath=$bath - $tempstep)
    evaluate ($radius=max($fin_rad,$radius*$radfact))
    evaluate ($k_vdw=min($fin_con,$k_vdw*$k_vdwfact))

    parameter nbonds repel=$radius end end
    constraints interaction (all) (all) weights * 1 vdw $k_vdw end end
    dynamics verlet
        nstep=$nstep time=$cool_timestep iasvel=current firstt=$bath
        tcoup=true tbath=$bath nprint=$nstep iprfrq=0
    end

    ! Abort condition.

    evaluate ($critical=$temp/$bath)
    if ($critical > 10) then
        display Rerun job with smaller timestep (i.e., 0.003)
        stop
    end if

end loop cool

! Final minimization.

minimize powell drop=10 nstep=600 nprint=25 end

! Analyze and write out the final structure.

print threshold=0.05 bonds
evaluate ($rms_bond=$result)
evaluate ($v_bond=$violations)
print threshold=5 angles
evaluate ($rms_angl=$result)
evaluate ($v_angl=$violations)
print threshold=15 dihedrals
evaluate ($rms_dihe=$result)
evaluate ($v_dihe=$violations)
print threshold=5 impropers
evaluate ($rms_impr=$result)
evaluate ($v_impr=$violations)
print threshold=0.2 noe
evaluate ($rms_noe=$result)
evaluate ($v_noe=$violations)
print threshold=5 cdih
evaluate ($rms_cdih=$result)
evaluate ($v_cdih=$violations)

set echo=off message=off end

display Energy:     bond $bond, angle $angl, dihedral $dihe,
display             improper $impr, NOE $noe, c-dihedral $cdih,
display             planar $plan, VdW $vdw, total $ener
display RMSD:       bond $rms_bond, angle $rms_angl, dihedral $rms_dihe,
display             improper $rms_impr, NOE $rms_noe, c-dihedral $rms_cdih
display Violations: bond $v_bond, angle $v_angl, dihedral $v_dihe,
display             improper $v_impr, NOE $v_noe, c-dihedral $v_cdih
display Handedness $hand, enantiomer discrimination $e_plus:$e_minus

set echo=on message=all end

write coordinates output=dgsa.pdb end

stop
```

A3.8: refine.run

This csh script runs the X-PLOR script refine.inp 30 times and summarizes the results. See also A3.2.

```
#!/bin/csh -f
#$ -cwd -o refine.run.o -e refine.run.e -m bae -M dcs@proton.chem.yale.edu

set i=21
while ($i <= 30)
    set ii=`nawk 'BEGIN { printf "%02d", '$i' }'`

    ln -s dgsa.$ii.pdb dgsa.pdb
    xplor -s < refine.inp > refine.$ii.out
    ip 'repdb -ix -om' refine.pdb
    mv refine.pdb refine.$ii.pdb
    rm -f dgsa.pdb refine.tra

    @ i++
end

perl -ne 'print if /^Ener/ .. /^ X-PLOR/' refine.??.out > refine.summary
perl -ne 'print if /0.2 noe/ .. /NOEPRI/' refine.??.out > refine.summary.noes
perl -ne 'print if /5 cdih/ .. /RMS deviation/' refine.??.out > refine.summary.cdih
```

A3.9: refine.inp

```
! refine.inp -- Clean up the output of dgsa.inp using simulated annealing,
!     including Van der Waals and electrostatic forces
! Dave Schweisguth <dcs@proton.chem.yale.edu>, 21 Jul 1996
! Derived from nmr/refine_gentle.inp

evaluate ($t=300)               ! Temperature for constant-temperature MD
evaluate ($steps=40000)         ! Number of steps of constant-temperature MD
evaluate ($timestep=0.0005)     ! Time of each MD step

set seed=@xplor.seed end        ! Use 'xplor -s'

set echo=off message=off end    ! Normal use
!set echo=on message=all end    ! Auxiliary file debugging

structure @rsf.inp end
parameter @XPLORLIB:toppar/dna-rna-allatom.par end

noe nres=1000 @noes.stem.inp @noes.loop.inp @base-pairs.inp end
restraints dihedral nassign=300 @dihedrals.inp end
evaluate($planarweight=50)      ! AAFS' choice
restraints planar @planars.inp end

set echo=on message=all end

noe                             ! Parameters for NOE effective energy term.
    ceiling=1000
    averaging  * cent
    potential  * square
    sqconstant * 1
    sqexponent * 2
    scale      * 100            ! Constant NOE scale throughout the protocol.
end

parameter
    nbonds
        ctonnb=9.5
        ctofnb=10.5
        cutnb=11.5
        tolerance=0.5
        rdie
        switch
        vswitch
    end
end

restraints dihedral scale=800 end

coordinates @dgsa.pdb           ! END in PDB file

flags exclude * include bond angl dihe impr vdw elec noe cdih plan end

! Intial minimization.

minimize powell nstep=200 nprint=25 end

! Constant temperature SA

vector do (fbeta=100) (all)     ! Coupling to heat bath.
dynamics verlet
    iasvel=maxwell
```

```
    firsttemperature=$t
    tbath=$t
    tcoup=true
    nstep=$steps
    timestep=$timestep
    nprint=250
    iprfrq=2500
    trajectory=refine.tra
    nsavc=200                   ! Save trajectory to file.
end
close refine.tra end

! Compute dynamics average over second half of trajectory.

evaluate ($steps2=$steps/2)
dynamics analysis average
    input=refine.tra
    begin=$steps2
    skip=200
    stop=$steps
end
close refine.tra end            ! disposition=delete doesn't work

! Final minimization.

minimize powell nstep=200 nprint=25 end

! Analyze and write out the final structure.

print threshold=0.05 bonds
evaluate ($rms_bond=$result)
evaluate ($v_bond=$violations)
print threshold=5 angles
evaluate ($rms_angl=$result)
evaluate ($v_angl=$violations)
print threshold=15 dihedrals
evaluate ($rms_dihe=$result)
evaluate ($v_dihe=$violations)
print threshold=5 impropers
evaluate ($rms_impr=$result)
evaluate ($v_impr=$violations)
print threshold=0.2 noe
evaluate ($rms_noe=$result)
evaluate ($v_noe=$violations)
print threshold=5 cdih
evaluate ($rms_cdih=$result)
evaluate ($v_cdih=$violations)

set echo=off message=off end

display Energy:      bond $bond, angle $angl, dihedral $dihe,
display              improper $impr, NOE $noe, c-dihedral $cdih,
display              planar $plan, VdW $vdw, elec $elec, total $ener
display RMSD:        bond $rms_bond, angle $rms_angl, dihedral $rms_dihe,
display              improper $rms_impr, NOE $rms_noe, c-dihedral $rms_cdih
display Violations:  bond $v_bond, angle $v_angl, dihedral $v_dihe,
display              improper $v_impr, NOE $v_noe, c-dihedral $v_cdih

set echo=on message=all end

write coordinates output=refine.pdb end

stop
```

A3.10: average.inp

Expects its input files to be named 'average.1.pdb' ... 'average.12.pdb'. Run directly with 'xplor -s < average.inp >& average.out &'.

```
! average.inp -- Computes the average structure, RMSDs from the average
! for each original structure and the rms RMSD.
! Dave Schweisguth <dcs@proton.chem.yale.edu>, 1 Sep 1996
! Derived from nmr/average.inp

set seed=@xplor.seed end     ! Use 'xplor -s'

set echo=off message=off end        ! Normal use
!set echo=on message=all end        ! Auxiliary file debugging

structure @rsf.inp end              ! Read the structure file.
parameter @XPLORLIB:toppar/dna-rna-allatom.par end

noe nres=1000 @noes.stem.inp @noes.loop.inp @base-pairs.inp end
restraints dihedral nassign=300 @dihedrals.inp end
evaluate($planarweight=50)   ! AAFS' choice
restraints planar @planars.inp end

set echo=on message=all end

noe                          ! Parameters for NOE effective energy term.
    ceiling=1000
    averaging  * cent
    potential  * square
    sqconstant * 1
    sqexponent * 2
    scale      * 100         ! Constant NOE scale throughout the protocol.
end

parameter
    nbonds
       ctonnb=9.5
       ctofnb=10.5
       cutnb=11.5
       tolerance=0.5
       rdie
       switch
       vswitch
    end
end

restraints dihedral scale=800 end

flags exclude * include bond angl dihe impr vdw elec noe cdih plan end

! Compute the mean structure.

evaluate ($nfile=9)
vector do (store1=0) (all)
vector do (store2=0) (all)
vector do (store3=0) (all)
vector do (store4=0) (all)

evaluate ($count=0)
while ($count < $nfile) loop main
    evaluate ($count=$count+1)
```

```
    evaluate ($filename="average."+encode($count)+".pdb")
    coordinates initialize end
    coordinates @@$filename
    if ($count=1) then
       coordinates copy end  ! Store first structure in comparison set.
    end if
    coordinates fit end
    vector do (store1=store1+x) (all)
    vector do (store2=store2+y) (all)
    vector do (store3=store3+z) (all)
    vector do (store4=store4+x*x+y*y+z*z) (all)
end loop main

vector do (x=store1 / $nfile) (all)
vector do (y=store2 / $nfile) (all)
vector do (z=store3 / $nfile) (all)
vector do (b=sqrt(max(0,store4/$nfile-(x**2+y**2+z**2)))) (all)

! Minimize the averaged structure

minimize powell nstep=200 nprint=25 end

! Analyze and write out the averaged structure

print threshold=0.05 bonds
evaluate ($rms_bond=$result)
evaluate ($v_bond=$violations)
print threshold=5 angles
evaluate ($rms_angl=$result)
evaluate ($v_angl=$violations)
print threshold=15 dihedrals
evaluate ($rms_dihe=$result)
evaluate ($v_dihe=$violations)
print threshold=5 impropers
evaluate ($rms_impr=$result)
evaluate ($v_impr=$violations)
print threshold=0.2 noe
evaluate ($rms_noe=$result)
evaluate ($v_noe=$violations)
print threshold=5 cdih
evaluate ($rms_cdih=$result)
evaluate ($v_cdih=$violations)

set echo=off message=off end

display Energy:      bond $bond, angle $angl, dihedral $dihe,
display              improper $impr, NOE $noe, c-dihedral $cdih,
display              planar $plan, VdW $vdw, elec $elec, total $ener
display RMSD:        bond $rms_bond, angle $rms_angl, dihedral $rms_dihe,
display              improper $rms_impr, NOE $rms_noe, c-dihedral $rms_cdih
display Violations:  bond $v_bond, angle $v_angl, dihedral $v_dihe,
display              improper $v_impr, NOE $v_noe, c-dihedral $v_cdih

set echo=on message=all end

! b array (last column) is the rms difference from the mean

write coordinates output=average.pdb end

! Compute an overall rms difference.

coordinates copy end
evaluate ($rmsd_rmsd=0)
evaluate ($rmsd_rmsd_stem=0)
```

```
evaluate ($rmsd_rmsd_loop=0)

evaluate ($count=0)
while ($count < $nfile) loop main
    evaluate ($count=$count+1)
    evaluate ($filename="average."+encode($count)+".pdb")
    coordinates initialize end
    coordinates @@$filename
    coordinates fit end
    coordinates rms end
    evaluate ($rmsd_rmsd=$rmsd_rmsd + $result**2)
    coordinates fit selection=(residue 1:7 or residue 15:21) end
    coordinates rms selection=(residue 1:7 or residue 15:21) end
    evaluate ($rmsd_rmsd_stem=$rmsd_rmsd_stem + $result**2)
    coordinates fit selection=(residue 8:14) end
    coordinates rms selection=(residue 8:14) end
    evaluate ($rmsd_rmsd_loop=$rmsd_rmsd_loop + $result**2)
end loop main

! RMS RMSD to the average structure

evaluate ($rmsd_rmsd=sqrt($rmsd_rmsd / $nfile))
evaluate ($rmsd_rmsd_stem=sqrt($rmsd_rmsd_stem / $nfile))
evaluate ($rmsd_rmsd_loop=sqrt($rmsd_rmsd_loop / $nfile))

stop
```

Appendix 4: Constraints for calculation of the structure of ITALY2

A4.1: base-pairs.inp – hydrogen-bonding constraints for Watson-Crick base-paired stem nucleotides

```
! Base pair restraints
!
! Heavy-atom to heavy-atom distances from Saenger p. 123
! Imino and amino to bonded nitrogen distances (all 1.01) from JPR's
!    parameter file
!
! AU pair
!
! AN6-UO4 (2.95) - AN6-AH62 (1.01) = AH62-UO4 (1.94)
! AN1-UN3 (2.82) - UH3-UN3 (1.01) = AN1-UH3 (1.81)
!
! GC pair
!
! GO6-CN4 (2.91) - CH42-CN4 (1.01) = GO6-CH42 (1.90)
! GN1-CN3 (2.95) - GN1-GH1 (1.01) = GH1-CN3 (1.94)
! GN2-CO2 (2.86) - GN2-GH22 (1.01) = GH22-CO2 (1.85)

! GC

assign (resid 1 and name O6) (resid 21 and name H42) 1.90 0.1 0.1
assign (resid 1 and name H1) (resid 21 and name N3) 1.94 0.1 0.1
assign (resid 1 and name H22) (resid 21 and name O2) 1.90 0.1 0.1

! GC

assign (resid 2 and name O6) (resid 20 and name H42) 1.90 0.1 0.1
assign (resid 2 and name H1) (resid 20 and name N3) 1.94 0.1 0.1
assign (resid 2 and name H22) (resid 20 and name O2) 1.90 0.1 0.1

! CG

assign (resid 3 and name H42) (resid 19 and name O6) 1.90 0.1 0.1
assign (resid 3 and name N3) (resid 19 and name H1) 1.94 0.1 0.1
assign (resid 3 and name O2) (resid 19 and name H22) 1.90 0.1 0.1

! AU

assign (resid 4 and name H62) (resid 18 and name O4) 1.94 0.1 0.1
assign (resid 4 and name N1) (resid 18 and name H3) 1.81 0.1 0.1

! GC

assign (resid 5 and name O6) (resid 17 and name H42) 1.90 0.1 0.1
assign (resid 5 and name H1) (resid 17 and name N3) 1.94 0.1 0.1
assign (resid 5 and name H22) (resid 17 and name O2) 1.90 0.1 0.1

! GC

assign (resid 6 and name O6) (resid 16 and name H42) 1.90 0.1 0.1
assign (resid 6 and name H1) (resid 16 and name N3) 1.94 0.1 0.1
assign (resid 6 and name H22) (resid 16 and name O2) 1.90 0.1 0.1
```

```
! GC

assign (resid 7 and name O6) (resid 15 and name H42) 1.90 0.1 0.1
assign (resid 7 and name H1) (resid 15 and name N3) 1.94 0.1 0.1
assign (resid 7 and name H22) (resid 15 and name O2) 1.90 0.1 0.1
```

A4.2: planars.inp – base-pair planarity constraints for Watson-Crick base-paired stem nucleotides

```
group
    selection=(
        (resid 1 and name N1) or
        (resid 1 and name N3) or
        (resid 1 and name C5) or
        (resid 21 and name N1) or
        (resid 21 and name N3) or
        (resid 21 and name C5)
    )
    weight=$planarweight
end

group
    selection=(
        (resid 2 and name N1) or
        (resid 2 and name N3) or
        (resid 2 and name C5) or
        (resid 20 and name N1) or
        (resid 20 and name N3) or
        (resid 20 and name C5)
    )
    weight=$planarweight
end

group
    selection=(
        (resid 3 and name N1) or
        (resid 3 and name N3) or
        (resid 3 and name C5) or
        (resid 19 and name N1) or
        (resid 19 and name N3) or
        (resid 19 and name C5)
    )
    weight=$planarweight
end

group
    selection=(
        (resid 4 and name N1) or
        (resid 4 and name N3) or
        (resid 4 and name C5) or
        (resid 18 and name N1) or
        (resid 18 and name N3) or
        (resid 18 and name C5)
    )
    weight=$planarweight
end

group
    selection=(
        (resid 5 and name N1) or
        (resid 5 and name N3) or
        (resid 5 and name C5) or
        (resid 17 and name N1) or
        (resid 17 and name N3) or
        (resid 17 and name C5)
    )
    weight=$planarweight
end

group
    selection=(
        (resid 6 and name N1) or
        (resid 6 and name N3) or
        (resid 6 and name C5) or
        (resid 16 and name N1) or
        (resid 16 and name N3) or
        (resid 16 and name C5)
    )
    weight=$planarweight
end

group
    selection=(
        (resid 7 and name N1) or
        (resid 7 and name N3) or
        (resid 7 and name C5) or
        (resid 15 and name N1) or
        (resid 15 and name N3) or
        (resid 15 and name C5)
    )
    weight=$planarweight
end
```

A4.3: noes.loop.inp – NOE-based distance constraints for loop nucleotides

The comment "increased or "reduced" indicates that a constraint was changed to a shorter or longer distance class respectively.

```
assign (resid 7 and name h8) (resid 8 and name h6) 4   1   1
assign (resid 7 and name h8) (resid 8 and name h5) 3   1   1
assign (resid 7 and name h1') (resid 8 and name h6) 4   1   1
assign (resid 7 and name h2') (resid 8 and name h6) 2.4 0.6 0.6
assign (resid 7 and name h2') (resid 8 and name h1') 4   1   1
assign (resid 7 and name h2') (resid 8 and name h5) 2.4 0.6 0.6
assign (resid 7 and name h3') (resid 8 and name h6) 3   1   1
assign (resid 7 and name h3') (resid 8 and name h5) 2.4 0.6 0.6
assign (resid 8 and name h6) (resid 8 and name h1') 3   1   1
assign (resid 8 and name h6) (resid 8 and name h5) 2.4 0.6 0.6
assign (resid 8 and name h6) (resid 8 and name h2') 3   1   1
assign (resid 8 and name h6) (resid 8 and name h3') 3   1   1        ! reduced
assign (resid 8 and name h6) (resid 8 and name h4') 3   1   1        ! reduced
assign (resid 8 and name h6) (resid 8 and name h5') 4 2.8 2.8
assign (resid 8 and name h1') (resid 8 and name h2') 2.4 0.6 0.6
assign (resid 8 and name h1') (resid 8 and name h3') 4   1   1
assign (resid 8 and name h1') (resid 8 and name h4') 4   1   1
assign (resid 8 and name h1') (resid 8 and name h5') 4 2.8 2.8
assign (resid 8 and name h1') (resid 9 and name h6) 4   1   1
assign (resid 8 and name h5) (resid 8 and name h3') 4   1   1
assign (resid 8 and name h5) (resid 9 and name h5) 4   1   1
assign (resid 8 and name h2') (resid 9 and name h6) 2.4 0.6 0.6
assign (resid 8 and name h2') (resid 9 and name h1') 4   1   1
assign (resid 8 and name h2') (resid 9 and name h5) 4   1   1
assign (resid 8 and name h3') (resid 9 and name h6) 3   1   1
assign (resid 8 and name h3') (resid 9 and name h5) 4   1   1
assign (resid 8 and name h5') (resid 9 and name h6) 4 2.8 2.8
assign (resid 9 and name h6) (resid 9 and name h1') 3   1   1
assign (resid 9 and name h6) (resid 9 and name h5) 2.4 0.6 0.6
assign (resid 9 and name h6) (resid 9 and name h2') 3   1   1
assign (resid 9 and name h6) (resid 9 and name h3') 4   1   1
assign (resid 9 and name h6) (resid 9 and name h4') 4   1   1
assign (resid 9 and name h1') (resid 9 and name h2') 2.4 0.6 0.6
assign (resid 9 and name h1') (resid 9 and name h3') 4   1   1
assign (resid 9 and name h1') (resid 9 and name h4') 3   1   1
assign (resid 9 and name h1') (resid 9 and name h5'') 3 2.8 2.8
assign (resid 9 and name h1') (resid 10 and name h2') 4   1   1
assign (resid 9 and name h1') (resid 11 and name h8) 3   1   1
assign (resid 9 and name h2') (resid 10 and name h6) 4   1   1
assign (resid 9 and name h2') (resid 11 and name h8) 3   1   1       ! reduced
assign (resid 10 and name h6) (resid 10 and name h1') 3   1   1
assign (resid 10 and name h6) (resid 10 and name h5) 2.4 0.6 0.6
assign (resid 10 and name h6) (resid 10 and name h2') 3   1   1
assign (resid 10 and name h6) (resid 10 and name h3') 3   1   1
assign (resid 10 and name h6) (resid 10 and name h4') 4   1   1
assign (resid 10 and name h6) (resid 10 and name h5') 4 2.8 2.8
assign (resid 10 and name h6) (resid 11 and name h8) 4   1   1
assign (resid 10 and name h1') (resid 10 and name h2') 2.4 0.6 0.6
assign (resid 10 and name h1') (resid 10 and name h4') 3   1   1
assign (resid 10 and name h1') (resid 10 and name h5') 4 2.8 2.8
assign (resid 10 and name h1') (resid 11 and name h8) 4   1   1
assign (resid 10 and name h2') (resid 11 and name h8) 2.4 0.6 0.6
assign (resid 10 and name h3') (resid 11 and name h8) 2.4 0.6 0.6
assign (resid 10 and name h4') (resid 11 and name h8) 4   1   1      ! reduced
assign (resid 10 and name h5') (resid 11 and name h8) 4 2.8 2.8
assign (resid 11 and name h2) (resid 12 and name h1') 4   1   1
assign (resid 11 and name h8) (resid 11 and name h1') 3   1   1
assign (resid 11 and name h8) (resid 11 and name h2') 3   1   1      ! reduced
assign (resid 11 and name h8) (resid 11 and name h3') 2.4 0.6 0.6
assign (resid 11 and name h8) (resid 11 and name h5') 4 2.8 2.8
assign (resid 11 and name h8) (resid 12 and name h6) 4   1   1
assign (resid 11 and name h1') (resid 11 and name h2') 3   1   1     ! increased
assign (resid 11 and name h1') (resid 11 and name h3') 4   1   1
assign (resid 11 and name h1') (resid 11 and name h4') 3   1   1
assign (resid 11 and name h2') (resid 11 and name h3') 3   1   1     ! increased
assign (resid 11 and name h2') (resid 12 and name h6) 3   1   1
assign (resid 11 and name h3') (resid 11 and name h4') 4   1   1
assign (resid 11 and name h3') (resid 11 and name h5') 4 2.8 2.8
assign (resid 11 and name h3') (resid 11 and name h5'') 4 2.8 2.8
assign (resid 11 and name h3') (resid 12 and name h6) 4   1   1
```

```
assign (resid 11 and name h3') (resid 12 and name h5'') 4 2.8 2.8
assign (resid 12 and name h6) (resid 12 and name h1') 3   1   1
assign (resid 12 and name h6) (resid 12 and name h5) 2.4 0.6 0.6
assign (resid 12 and name h6) (resid 12 and name h2') 2.4 0.6 0.6
assign (resid 12 and name h6) (resid 12 and name h3') 3   1   1
assign (resid 12 and name h6) (resid 12 and name h5') 4 2.8 2.8
assign (resid 12 and name h6) (resid 12 and name h5'') 4 2.8 2.8
assign (resid 12 and name h1') (resid 12 and name h2') 2.4 0.6 0.6
assign (resid 12 and name h1') (resid 12 and name h3') 4   1   1
assign (resid 12 and name h1') (resid 12 and name h4') 4   1   1
assign (resid 12 and name h1') (resid 12 and name h5') 4 2.8 2.8
assign (resid 12 and name h1') (resid 12 and name h5'') 4 2.8 2.8
assign (resid 12 and name h2') (resid 13 and name h8) 4   1   1
assign (resid 12 and name h2') (resid 13 and name h1') 4   1   1
assign (resid 12 and name h3') (resid 13 and name h8) 2.4 0.6 0.6
assign (resid 12 and name h4') (resid 13 and name h8) 4   1   1
assign (resid 13 and name h2) (resid 13 and name h2') 4   1   1
assign (resid 13 and name h2) (resid 14 and name h1') 2.4 0.6 0.6
assign (resid 13 and name h8) (resid 13 and name h1') 3   1   1
assign (resid 13 and name h8) (resid 13 and name h2') 3   1   1
assign (resid 13 and name h8) (resid 13 and name h3') 2.4 0.6 0.6
assign (resid 13 and name h8) (resid 13 and name h4') 3   1   1
assign (resid 13 and name h8) (resid 13 and name h5') 2.4 2.4 2.4
assign (resid 13 and name h1') (resid 13 and name h2') 2.4 0.6 0.6
assign (resid 13 and name h1') (resid 13 and name h3') 3   1   1     ! reduced
assign (resid 13 and name h1') (resid 13 and name h4') 4   1   1
assign (resid 13 and name h1') (resid 13 and name h5') 4 2.8 2.8
assign (resid 13 and name h1') (resid 14 and name h8) 4   1   1      ! reduced
assign (resid 13 and name h1') (resid 14 and name h1') 3   1   1
assign (resid 13 and name h1') (resid 14 and name h5') 4 2.8 2.8
assign (resid 13 and name h2') (resid 13 and name h3') 3   1   1     ! increased
assign (resid 13 and name h2') (resid 13 and name h4') 4   1   1
assign (resid 13 and name h2') (resid 13 and name h5') 4 2.8 2.8
assign (resid 13 and name h2') (resid 14 and name h8) 2.4 0.6 0.6
assign (resid 13 and name h2') (resid 14 and name h1') 2.4 0.6 0.6
assign (resid 13 and name h2') (resid 14 and name h4') 4   1   1
assign (resid 13 and name h2') (resid 14 and name h5') 4 2.8 2.8
assign (resid 13 and name h3') (resid 14 and name h8) 2.4 0.6 0.6
assign (resid 14 and name h2) (resid 15 and name h1') 2.4 0.6 0.6
assign (resid 14 and name h8) (resid 14 and name h1') 3   1   1
assign (resid 14 and name h8) (resid 14 and name h3') 4   1   1
assign (resid 14 and name h8) (resid 14 and name h5') 3 2.8 2.8
assign (resid 14 and name h8) (resid 15 and name h6) 4   1   1       ! reduced
assign (resid 14 and name h8) (resid 15 and name h5) 4   1   1
assign (resid 14 and name h1') (resid 14 and name h2') 2.4 0.6 0.6
assign (resid 14 and name h1') (resid 14 and name h3') 4   1   1
assign (resid 14 and name h1') (resid 14 and name h4') 4   1   1
assign (resid 14 and name h1') (resid 14 and name h5') 4 2.8 2.8
assign (resid 14 and name h1') (resid 15 and name h6) 4   1   1
assign (resid 14 and name h1') (resid 15 and name h5'') 4 2.8 2.8     ! reduced
assign (resid 14 and name h3') (resid 15 and name h6) 2.4 0.6 0.6
assign (resid 14 and name h3') (resid 15 and name h1') 4   1   1
assign (resid 14 and name h3') (resid 15 and name h5) 4   1   1
assign (resid 14 and name h4') (resid 15 and name h6) 4   1   1
! From 20°C NOESY, added by hand
assign (resid 11 and name h1') (resid 12 and name h6) 4   1   1
assign (resid 12 and name h1') (resid 13 and name h8) 4   1   1
! From 600 Mhz NOESY, added by hand
assign (resid 8 and name h6) (resid 9 and name h5) 4   1   1
assign (resid 12 and name h6) (resid 13 and name h8) 4   1   1
```

A4.4: dihedrals.d2x – dihedral angle constraints

The program dihedrals2xplor (A2.6) was used to convert the following table to X-PLOR format (dihedrals.inp).

```
# Dihedral angle constraints for ITALY2
# A-form values from Arnott and Chandrasekaran (1982) as reported by Sanger (1984)

#  alpha   beta    gamma  epsilon   zeta    chi      nu0       nu1       nu2       nu3       nu4

g  none    none    none   -153  15  -71 15  -158 15    3   15  -25   15   37   15  -36   15   21   15
g  -68 15  178 15  54 15  -153  15  -71 15  -158 15    3   15  -25   15   37   15  -36   15   21   15
c  -68 15  178 15  54 15  -153  15  -71 15  -158 15    3   15  -25   15   37   15  -36   15   21   15
a  -68 15  178 15  54 15  -153  15  -71 15  -158 15    3   15  -25   15   37   15  -36   15   21   15
g  -68 15  178 15  54 15  -153  15  -71 15  -158 15    3   15  -25   15   37   15  -36   15   21   15
g  -68 15  178 15  54 15  -153  15  -71 15  -158 15    3   15  -25   15   37   15  -36   15   21   15
g  -68 15  178 15  54 15  -153  15  -71 15  -158 15    3   15  -25   15   37   15  -36   15   21   15
c  -68 30  178 30  54 30  -120 120  -71 30  -158 30    3   30  -25   30   37   30  -36   30   21   30
u  -68 30  178 30  54 30  -120 120  none    -158 30    3   30  -25   30   37   30  -36   30   21   30
c  none    178 30  54 30  -120 120  none    -158 30  none      none      none      none      none
a  none    178 30  none   -120 120  -71 30  -158 30  none      none      none      none      none
u  -68 30  178 30  none   -120 120  none    -158 30  -21.7 30   35.2 30  -35.2 30   21.8 30    0   30
a  none    178 30  54 30  -120 120  -71 30  -158 30  none      none      none      none      none
a  -68 30  178 30  54 30  -120 120  -71 30  -158 30    3   30  -25   30   37   30  -36   30   21   30
c  -68 15  178 15  54 15  -153  15  -71 15  -158 15    3   15  -25   15   37   15  -36   15   21   15
c  -68 15  178 15  54 15  -153  15  -71 15  -158 15    3   15  -25   15   37   15  -36   15   21   15
c  -68 15  178 15  54 15  -153  15  -71 15  -158 15    3   15  -25   15   37   15  -36   15   21   15
u  -68 15  178 15  54 15  -153  15  -71 15  -158 15    3   15  -25   15   37   15  -36   15   21   15
g  -68 15  178 15  54 15  -153  15  -71 15  -158 15    3   15  -25   15   37   15  -36   15   21   15
c  -68 15  178 15  54 15  -153  15  -71 15  -158 15    3   15  -25   15   37   15  -36   15   21   15
c  -68 15  178 15  54 15  none      none    -158 15    3   15  -25   15   37   15  -36   15   21   15
```

Bibliography

Altona, C. (1982). "Conformational analysis of nucleic acids: Determination of backbone geometry of single-helical RNA and DNA in aqueous solution." J. Royal Neth. Chem. Soc. 101(12): 413-433.

Anderson, A.C., S.A. Scaringe, B.E. Earp and C.A. Frederick (1996). "HPLC purification of RNA for crystallography and NMR." RNA 2(2): 110-117.

Arnez, J.G. and T.A. Steitz (1994). "Crystal structure of unmodified $tRNA^{Gln}$ complexed with glutaminyl-tRNA synthetase and ATP suggests a possible role for pseudo-uridines in stabilization of RNA structure." Biochem. 33(24): 7560-7567.

Åström, S.U. and A.S. Byström (1994). "Rit1, a tRNA backbone-modifying enzyme that mediates initiator and elongator tRNA discrimination." Cell 79(3): 535-546.

Åström, S.U., M. Nordlund, F.L. Ericson, E.M. Hannig and A.S. Byström (1996). "Modification of adenosine 64 of the yeast initiator tRNA has two different effects on translational initiation." Poster presentation, RNA Society meeting, Madison, WI, June 1996

Auron, P.E., L.D. Weber and A. Rich (1982). "Comparison of transfer ribonucleic acid structures using cobra venom and S1 endonucleases." Biochem. 21(19): 4700-6.

Basavappa, R. and P.B. Sigler (1991). "The 3 Å crystal structure of yeast initiator tRNA: Functional implications in initiator/elongator discrimination." EMBO J. 10(10): 3105-3111.

Beckett, D. and O.C. Uhlenbeck (1984). Enzymatic synthesis of oligoribonucleotides. In Oligonucleotide Synthesis: A Practical Approach, ed. M.J. Gait. Oxford: IRL Press.

Betzel, C., S. Lorenz, J.P. Furste, R. Bald, M. Zhang, T.R. Schneider, K.S. Wilson and V.A. Erdmann (1994). "Crystal structure of domain A of *Thermus flavus* 5S rRNA and the contribution of water molecules to its structure." FEBS Lett. 351(2): 159-164.

Björk, G.R., J.U. Ericson, C.E.D. Gustafsson, T.G. Hagervall, Y.H. Jönsson and P.M. Wikström (1987). "Transfer RNA modification." Ann. Rev. Biochem. 56: 263-287.

Branch, A.D., B.J. Benenfeld and H.D. Robertson (1985). "Ultraviolet light-induced crosslinking reveals a unique region of local tertiary structure in potato spindle tuber viroid and HeLa 5S RNA." Proc. Natl. Acad. Sci. USA 82(19): 6590-4.

Bremer, H. and P.P. Dennis (1987). 96. Modulation of chemical composition and other parameters of the cell by growth rate. In *Escherichia coli* and *Salmonella typhinurium*: Cellular and Molecular Biology, ed. F.C. Neidhardt. Washington, DC: American Society for Microbiology.

Brünger, A.T. (1992). X-PLOR Version 3.1: A System For X-Ray Crystallography and NMR. New Haven: Yale University Press.

Chelladurai, B., H. Li, K. Zhang and A.W. Nicholson (1993). "Mutational analysis of a ribonuclease III processing signal." Biochem. 32(29): 7549-58.

Chelladurai, B.S., H. Li and A.W. Nicholson (1991). "A conserved sequence element in ribonuclease III processing signals is not required for accurate in vitro enzymatic cleavage." Nucleic Acids Res. 19(8): 1759-66.

Chen, Y., H. Sierzputowska-Gracz, R. Guenther, K. Everett and P.F. Agris (1993). "5-methylcytidine is required for cooperative binding of Mg^{2+} and a conformational transition at the anticodon stem-loop of yeast phenylalanine tRNA." Biochem. 32(38): 10249-10253.

Cheong, C., G. Varani and I. Tinoco Jr. (1990). "Solution structure of an unusually stable RNA hairpin, 5'GGAC(UUCG)GUCC." Nature 346(6285): 680-682.

Ciesiolka, J., J. Wrzesinski, P. Gornicki, J. Podkowinski and W.J. Krzyzosiak (1989). "Analysis of magnesium, europium and lead binding sites in methionine initiator and elongator tRNAs by specific metal-ion-induced cleavages." Eur. J. Biochem. 186(1-2): 71-77.

Clore, G., A. Gronenborn, E. Piper, L. McLaughlin, E. Graeser and J. Van Boom (1984). "The solution structure of a RNA pentadecamer comprising the anticodon loop and stem of yeast $tRNA^{Phe}$. A 500 MHz ^{1}H-N.M.R. study." Biochem. J. 221(3): 737-751.

Costa, M. and F. Michel (1995). "Frequent use of the same tertiary motif by self-folding RNAs." EMBO J. 14(6): 1276-1285.

Court, D. (1993). In Control of Messenger RNA Stability, eds. J.G. Belasco and G. Brawerman. NY: Academic Press.

Davis, D.R. (1995). “Stabilization of RNA stacking by pseudouridine.” Nucleic Acids Res. 23(24): 5020-5026.

Dirheimer, G., G. Keith, P. Dumas and E. Westhof (1995). Primary, secondary and tertiary structures of tRNAs. In tRNA: Structure, Biosynthesis and Function, eds. D. Söll and U. RajBhandary. Washington, DC: American Society of Microbiology. 93-126.

Dock-Bregeon, A.C. and D. Moras (1987). “Conformational changes and dynamics of tRNAs: evidence from hydrolysis patterns.” C.S.H. Symp. Quant. Biol. 52: 113-121.

Drabkin, H.J., B. Helk and U.L. RajBhandary (1993). “The role of nucleotides conserved in eukaryotic initiator methionine transfer RNAs in initiation of protein synthesis.” J. Biol. Chem. 268(33): 25221-25228.

Dube, S.K., P.S. Rudland, B.F.C. Clark and K.A. Marcker (1969). “A structural requirement for codon-anticodon interaction on the ribosome.” C.S.H. Symp. Quant. Biol. 34: 161-166.

Dunn, J.J. (1976). “RNase III cleavage of single-stranded RNA. Effect of ionic strength on the fidelity of cleavage.” J. Biol. Chem. 251(12): 3807-14.

Dunn, J.J. and F.W. Studier (1983). “Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements.” J. Mol. Biol. 166(4): 477-535.

Dutka, S., T. Meinnel, C. Lazennec, Y. Mechulam and S. Blanquet (1993). “Role of the 1-72 base pair in transfer RNAs for the activity of Escherichia coli peptidyl-transfer RNA hydrolase.” Nucleic Acids Res. 21(17): 4025-4030.

Ehresmann, C., F. Baudin, M. Mougel, P. Romby, J.P. Ebel and B. Ehresmann (1987). “Probing the structure of RNAs in solution.” Nucleic Acids Res. 15(22): 9109-9128.

Farrugio, D., J. Chaudhuri, U. Maitra and U. RajBhandary (1996). “The A1-U72 base pair conserved in eukaryotic initiator tRNAs is important specifically for binding to the eukaryotic translation initiation factor eIF-2.” Mol. Cell. Biol. 16(8): 4248-4256.

Fountain, M.A., M.J. Serra, T.R. Krugh and D.H. Turner (1996). “Structural features of a six-nucleotide RNA hairpin loop found in ribosomal RNA.” Biochem. 35(21): 6539-6548.

Francis, M. and U.L. RajBhandary (1990). "Expression and function of a human initiator tRNA gene in the yeast *Saccharomyces cerevisiae*." Mol. Cell. Biol. 10: 4486-4494.

Freier, S.M., R. Kierzek, J.A. Jaeger, N. Sugimoto, M. Caruthers, T. Neilson and D.H. Turner (1986). "Improved free-energy parameters for predictions of RNA duplex stability." Proc. Natl. Acad. Sci. USA 83: 9373-9377.

Gorenstein, D. and E. Goldfield (1982). "High-resolution phosphorus nuclear magnetic resonance spectroscopy of transfer ribonucleic acids: Multiple conformations in the anticodon loop." Biochem. 21(23): 5839-5849.

Gorenstein, D.G. (1984). Phosphorus-31 chemical shifts: Principles and empirical observations. In Phosphorus-31 NMR: Principles and Applications, ed. D.G. Gorenstein. Orlando, FL: Academic Press. 7-36.

Grodberg, J. and J.J. Dunn (1988). "ompT encodes the Escherichia coli outer membrane protease that cleaves T7 RNA polymerase during purification." J. Bacteriol. 170(3): 1245-53.

Gronenborn, A.M. and G.M. Clore (1995). "Structures of protein complexes by multidimensional heteronuclear magnetic resonance spectroscopy." Crit. Rev. Biochem. Mol. Biol. 30(5): 35167.

Grosjean, H., M. Sprinzl and S. Steinberg (1995). "Posttranscriptionally modified nucleosides in transfer RNA: their locations and frequencies." Biochimie 77: 139-141.

Gualerzi, C. and C. Pon (1990). "Initiation of mRNA translation in prokaryotes." Biochem. 29(25): 5881-5889.

Guenther, R.H., C.C. Hardin, H. Sierzputowska-Gracz, V. Dao and P.F. Agris (1992). "A magnesium-induced conformational transition in the loop of a DNA analog of the yeast $tRNA^{Phe}$ anticodon is dependent on RNA-like modifications of the bases of the stem." Biochem. 31(45): 11004-11011.

Guillon, J.-M., T. Meinnel, Y. Mechulam, C. Lazennec, S. Blanquet and G. Fayat (1992). "Nucleotides of tRNA governing the specificity of Escherichia coli methionyl-$tRNA_f^{Met}$ formyltransferase." J. Mol. Biol. 224: 359-367.

Gutell, R.R. and C.R. Woese (1990). "Higher order structural elements in ribosomal RNAs: Pseudo-knots and the use of noncanonical pairs." Proc. Natl. Acad. Sci. USA 87: 663-667.

Haines, D.S., K.I. Strauss and D.H. Gillespie (1991). "Cellular response to double-stranded RNA." J. Cell. Biochem. 46(1): 9-20.

Harada, F. and J. Dahlberg (1975). "Specific cleavage of tRNA by nuclease S1." Nucleic Acids Res. 2(6): 865-871.

Hartz, D., J. Binkley, T. Hollingsworth and L. Gold (1990). "Domains of initiator tRNA and initiation codon crucial for initiator tRNA selection by Escherichia coli IF3." Genes Dev. 4(10): 1790-1800.

Hartz, D., D.S. McPheeters and L. Gold (1989). "Selection of the initiator tRNA by *Escherichia coli* initiation factors." Genes Dev. 3(12A): 1899-912.

Heus, H.A. and A. Pardi (1991a). "Novel ^{1}H nuclear magnetic resonance assignment procedure for RNA duplexes." J. Amer. Chem. Soc. 113: 4360-4361.

Heus, H.A. and A. Pardi (1991b). "Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops." Science 253(5016): 191-194.

Hingerty, B.E., R.S. Brown and A. Jack (1978). "Further refinement of the structure of yeast $tRNA^{Phe}$." J. Mol. Biol. 124: 523.

Holbrook, S.R., J.L. Sussman, R.W. Warrant and S.-H. Kim (1978). "Crystal structure of yeast phenylalanine tRNA. II. Structural features and functional implications." J. Mol. Biol. 123: 631.

Huang, S.G., Y.X. Wang and D.E. Draper (1996). "Structure of a hexanucleotide RNA hairpin loop conserved in ribosomal RNAs." J. Mol. Biol. 258(2): 308-321.

Jack, A., J.E. Ladner and A. Klug (1976). "Crystallographic refinement of yeast phenylalanine tRNA at 2.5 Å resolution." J. Mol. Biol. 108: 619-649.

Jaeger, J.A. and I. Tinoco (1993). "An NMR study of the HIV-1 TAR element hairpin." Biochem. 32(46): 12522-12530.

Jaeger, L., F. Michel and E. Westhof (1994). "Involvement of a GNRA tetraloop in long-range tertiary interactions." J. Mol. Biol. 236(5): 1271-1276.

Jakubowski, H. and E. Goldman (1984). "Quantities of individual aminoacyl-tRNA families and their turnover in *Escherichia coli*." J. Bacteriol. 158(3): 769-776.

Jucker, F.M. and A. Pardi (1995a). "GNRA tetraloops make a U-turn." RNA 1(2): 219-222.

Jucker, F.M. and A. Pardi (1995b). "Solution structure of the CUUG hairpin loop - a novel RNA tetraloop motif." Biochem. 34(44): 14416-14427.

Kellogg, G.W. (1992). "Proton-detected hetero-TOCSY experiments with application in nucleic acids." J. Mag. Res. 98: 176-182.

Kellogg, G.W., A.A.F. Szewczak and P.B. Moore (1992). "Two-dimensional hetero-TOCSY-NOESY: correlation of ^{31}P resonances with anomeric and aromatic ^{1}H resonances in RNA." J. Amer. Chem. Soc. 114: 2727-2728.

Kiesewetter, S., G. Ott and M. Sprinzl (1990). "The role of modified purine 64 in initiator/elongator discrimination of $tRNA_i^{Met}$ from yeast and wheat germ." Nucleic Acids Res. 18(16): 4677-4682.

Kim, Y. and J.H. Prestegard (1989). "Measurement of vicinal couplings from cross peaks in COSY spectra." J. Mag. Res. 84: 9-13.

Kime, M.J. and P.B. Moore (1983a). "Nuclear Overhauser experiments at 500 MHz on the downfield proton spectra of 5S ribonucleic acid and its complex with ribosomal protein L25." Biochem. 22(11): 2622-2629.

Kime, M.J. and P.B. Moore (1983b). "Physical evidence for a domain structure in *Escherichia coli* 5S RNA." FEBS Lett. 153(1): 199-203.

Kuchino, Y., M. Ihara, Y. Yabusaki and S. Nishimura (1982). "Initiator tRNAs from archaebacteria show common unique sequence characteristics." Nature 298(5874): 684-685.

Labuda, D. and D. Porschke (1982). "Magnesium ion inner sphere complex in the anticodon loop of phenylalanine transfer ribonucleic acid." Biochem. 21: 49-53.

Lee, C., B. Seong and U. RajBhandary (1991). "Structural and sequence elements important for recognition of Escherichia coli formylmethionine tRNA by methionyl-tRNA transformylase are clustered in the acceptor stem." J. Biol. Chem. 266(27): 18012-017.

Legault, P. and A. Pardi (1994). "^{31}P chemical shift as a probe of structural motifs in RNA." J. Mag. Res. B 103: 82-86.

Li, H.L., B.S. Chelladurai, K.J. Zhang and A.W. Nicholson (1993). "Ribonuclease III cleavage of a bacteriophage T7 processing signal - divalent cation specificity and specific anion effects." Nucleic Acids Res. 21(8): 1919-1925.

Libonati, M., A. Carsana and A. Furia (1980). "Double-stranded RNA." Mol. Cell. Bioch. 31(3): 147-164.

Linn, T. and R. St. Pierre (1990). "Improved vector system for constructing transcriptional fusions that ensures independent translation of lacZ." J. Bacteriol. 172(2): 1077-84.

Lockard, R.E. and A. Kumar (1981). "Mapping tRNA structure in solution using double-strand-specific ribonuclease V1 from cobra venom." Nucleic Acids Res. 9(19): 5125-40.

Lowman, H.B. and D.E. Draper (1986). "On the recognition of helical RNA by cobra venom V1 nuclease." J. Biol. Chem. 261(12): 5396-403.

Mandal, N., D. Mangroo, J.J. Dalluge, J.A. McCloskey and U.L. RajBhandary (1996). "Role of the three consecutive G-C base pairs conserved in the anticodon stem of initiator tRNAs in initiation of protein synthesis in *Escherichia coli*." RNA 2(5): 473-482.

Maniatis, T., E.F. Fritsch and J. Sambrook (1982). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Martinis, S.A. and P. Schimmel (1993). "Microhelix aminoacylation by a class I tRNA synthetase: Non-conserved base pairs required for specificity." J. Biol. Chem. 268(9): 6069-72.

Meinnel, T., Y. Mechulam and S. Blanquet (1993a). "Methionine as translation start signal: a review of the enzymes of the pathway in *Escherichia coli*." Biochimie 75: 1061-1075.

Meinnel, T., Y. Mechulam, F. Dardel, J.M. Schmitter, C. Hountondji, S. Brunie, P. Dessen, G. Fayat and S. Blanquet (1990). "Methionyl-tRNA synthetase from E. coli -- A review." Biochimie 72: 625-632.

Meinnel, T., Y. Mechulam, G. Fayat and S. Blanquet (1992). "Involvement of the size and sequence of the anticodon loop in tRNA recognition by mammalian and E. coli methionyl-tRNA synthetases." Nucleic Acids Res. 20(18): 4741-4746.

Meinnel, T., Y. Mechulam, C. Lazennec, S. Blanquet and G. Fayat (1993b). "Critical role of the acceptor stem of $tRNAs^{Met}$ in their aminoacylation by *Escherichia coli* methionyl-transfer RNA synthetase." J. Mol. Biol. 229(1): 26-36.

Michel, F. and E. Westhof (1990). "Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis." J. Mol. Biol. 216: 581-606.

Milligan, J.F., D. Groebe, G. Witherell and O.C. Uhlenbeck (1987). "Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates." Nucleic Acids Res. 15(21): 8783-8798.

Mirmira, S.R. and I. Tinoco (1996a). "NMR structure of a bacteriophage T4 RNA hairpin involved in translational repression." Biochem. 35(24): 7664-7674.

Mirmira, S.R. and I. Tinoco (1996b). "A quadruple mutant T4 RNA hairpin with the same structure as the wild-type translational repressor." Biochem. 35(24): 7675-7683.

Moore, P.B. (1995). "Determination of RNA conformation by nuclear magnetic resonance." Accts. Chem. Res. 28(6): 251-256.

Neefs, J.-M., Y. Van de Peer, P. De Rijk, S. Chapelle and R. De Wachter (1993). "Compilation of small ribosomal subunit RNA structures." Nucleic Acids Res. 21(13): 3025-3049.

Nicholson, A.W., K.R. Niebling, P.L. McOsker and H.D. Robertson (1988). "Accurate in vitro cleavage by RNase III of phosphorothioate-substituted RNA processing signals in bacteriophage T7 early mRNA." Nucleic Acids Res. 16(4): 1577-91.

Pak, M., L. Pallanck and L.H. Schulman (1992). "Conversion of a methionine initiator tRNA into a tryptophan-inserting elongator tRNA in vivo." Biochem. 31(13): 3304-3309.

Panayotatos, N. and K. Truong (1985). "Cleavage within an RNase III site can control mRNA stability and protein synthesis in vivo." Nucleic Acids Res. 13(7): 2227-40.

Parkinson, G., J. Voitechovsky, L. Clowney, A.T. Brünger and H. Berman (1996). "New parameters for the refinement of nucleic acid-containing structures." Acta Cryst. D 52: 57-64.

Peattie, D.A. and W. Gilbert (1980). "Chemical probes for higher-order structure in RNA." Proc. Natl. Acad. Sci. USA 77(8): 4679-82.

Perona, J.J., M.A. Rould, T.A. Steitz, J.-L. Risler, C. Zelwer and S. Brunie (1991). "Structural similarities in glutaminyl- and methionyl-tRNA synthetases suggest a common overall orientation of tRNA binding." Proc. Natl. Acad. Sci. USA 88: 2903-2907.

Petersen, H.U., T.A. Kruse, H. Worm-Leonhard, G.E. Siboska, B.F.C. Clark, A.S. Boutorin, P. Remy, J.P. Ebel, J. Dondon and M. Grunberg-Manago (1981). "A study of the interaction of *Escherichia coli* initiation factor IF2 with formylmethionyl-$tRNA_f^{Met}$ by partial digestion with cobra venom ribonuclease." FEBS Lett. 128(1): 161-165.

Petersen, H.U., T. Røll, M. Grunberg-Manago and B.F.C. Clark (1979). "Specific interaction of initiation factor IF2 of *E. coli* with formylmethionyl-$tRNA_f^{Met}$." Biochem. Biophys. Res. Comm. 91(3): 1068-1074.

Plaipa, R. and B. Juodka (1996). "The possible involvement of RNases in the so-called "RNA self-splitting" at Y-A bonds." Poster presentation, RNA Society meeting, Madison, WI, June 1996

Pley, H.W., K.M. Flaherty and D.B. Mckay (1994a). "Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix." Nature 372(6501): 111-113.

Pley, H.W., K.M. Flaherty and D.B. Mckay (1994b). "Three-dimensional structure of a hammerhead ribozyme." Nature 372(6501): 68-74.

Portmann, S., N. Usman and M. Egli (1995). "The crystal structure of r(CCCCGGGG) in two distinct lattices." Biochem. 34(23): 7569-7575.

Puglisi, J.D. (1989) RNA folding: Structure and conformational equilibria of RNA pseudoknots. Ph.D. thesis, U.C. Berkeley.

Puglisi, J.D., R. Tan, B.J. Calnan, A.D. Frankel and J.R. Williamson (1992). "Conformation of the TAR RNA-arginine complex by NMR spectroscopy." Science 257(5066): 76-80.

Quigley, G.J. and A. Rich (1976). "Structural domains of transfer RNA molecules." Science 194(4267): 796-806.

RajBhandary, U.L. and C.M. Chow (1995). Initiator tRNAs and initiation of protein synthesis. In tRNA: Structure, Biosynthesis and Function, eds. D. Söll and U.L. RajBhandary. Washington, D.C.: American Society for Microbiology. 511-528.

Rice, L.M. and A.T. Brünger (1994). "Torsion angle dynamics: Reduced variable conformational sampling enhances crystallographic structure refinement." Proteins: Structure, Function, and Genetics 19: 277-290.

Rife, J.P. and P.B. Moore (1996). "Solution structure of r(CAGGGGAACCUG)$_2$, an RNA that contains linear tetraloops." Nucleic Acids Res. submitted:

Robertson, H.D. (1990). "*Escherichia coli* ribonuclease III." Meth. Enzymol. 181: 189-202.

Robertson, H.D. and F. Barany (1978). Enzymes and mechanisms of RNA processing. In Gene Function / Proceedings of the 12th FEBS Congress, eds. S. Rosenthal, H. Bielka, C. Coutelle and C. Zimmer. NY: Pergamon Press. 285-295.

Robertson, H.D., R.E. Webster and N.D. Zinder (1968). "Purification and properties of ribonuclease III from *Escherichia coli*." J. Biol. Chem. 243(1): 82-91.

Romaniuk, P., I. de Stevenson, C. Ehresmann, P. Romby and B. Ehresmann (1988). "A comparison of the solution structures and conformational properties of the somatic oocyte 5S rRNAs of *Xenopus laevis*." Nucleic Acids Res. 16(5B): 2295-2312.

Rould, M., J. Perona, D. Soell and T. Steitz (1989). "Structure of E. coli glutaminyl-tRNA synthetase complexed with tRNA^{Gln} and ATP at 2.8 Å resolution." Science 246(4934): 1135-1142.

Rudinger, J., R. Hillenbrandt, M. Sprinzl and R. Giege (1996). "Antideterminants present in minihelix^{Sec} hinder its recognition by prokaryotic elongation factor Tu." EMBO J. 15(3): 650-657.

Saenger, W. (1983). Principles of Nucleic Acid Structure. New York: Springer-Verlag.

Saito, H. and C.C. Richardson (1981). "Processing of mRNA by ribonuclease III regulates expression of gene 1.2 of bacteriophage T7." Cell 27(3 Pt 2): 533-42.

Salemink, P.J.M., E.J. Reijerse, L.C.P.J. Mollevanger and C.W. Hilbers (1981). "Conformational changes of yeast tRNA^{Phe} as monitored by ^{31}P NMR." Eur. J. Biochem. 115: 635-641.

SantaLucia, J., Jr., R. Kierzek and D.H. Turner (1992). "Context dependence of hydrogen bond free energy revealed by substitutions in an RNA hairpin." Science 256(5054): 217-9.

Schevitz, R.W., A.D. Podjarny, N. Krishnamachari, J.J. Hughes and P.B. Sigler (1979). "Crystal structure of a eukaryotic initiator tRNA." Nature 278: 188-190.

Schulman, L.H. and H. Pelka (1975). "The structural basis for the resistance of *Escherichia coli* formylmethionyl transfer ribonucleic acid to cleavage by *Escherichia coli* peptidyl transfer ribonucleic acid hydrolase." J. Biol. Chem. 250(2): 542-7.

Schulman, L.H., H. Pelka and R.M. Sundari (1974). "Structural requirements for recognition of *Escherichia coli* initiator and non-initiator transfer ribonucleic acids by bacterial T factor." J. Biol. Chem. 249(22): 7102-10.

Schweisguth, D.C., B.S. Chelladurai, A.W. Nicholson and P.B. Moore (1994). "Structural characterization of a ribonuclease III processing signal." Nucleic Acids Res. 22(4): 604-612.

Scott, W.G., J.T. Finch and A. Klug (1995). "The crystal structure of an all-RNA hammerhead ribozyme - a proposed mechanism for RNA catalytic cleavage." Cell 81(7): 991-1002.

Searle, M.S. and D.H. Williams (1993). "On the stability of nucleic acid structures in solution - enthalpy-entropy compensations, internal rotations and reversibility." Nucleic Acids Res. 21(9): 2051-2056.

Seggerson, K. (1994). Personal communication.

Senger, B., R. Aphasizhev, P. Walter and F. Fasiolo (1995). "The presence of a D-stem but not a T-stem is essential for triggering aminoacylation upon anticodon binding in yeast methionine tRNA." J. Mol. Biol. 249(1): 45-58.

Senger, B., L. Despons, P. Walter and F. Fasiolo (1992). "The anticodon triplet is not sufficient to confer methionine acceptance to a transfer RNA." Proc. Natl. Acad. Sci. USA 89(22): 10768-10771.

Seong, B. and U. RajBhandary (1987a). "*Escherichia coli* formylmethionine tRNA: Mutations of GGG:CCC sequence conserved in anticodon stem of initiator tRNAs affect initiation of protein synthesis and conformation of anticodon loop." Proc. Natl. Acad. Sci. USA 84(2): 334-338.

Seong, B.L. and U.L. RajBhandary (1987b). "Mutants of Escherichia coli formylmethionine tRNA: A single base change enables initiator tRNA to act as an elongator in vitro." Proc. Natl. Acad. Sci. USA 84: 8859-8863.

Serra, M.J., M.H. Lyttle, T.J. Axenson, C.A. Schadt and D.H. Turner (1993). "RNA hairpin loop stability depends on closing base pair." Nucleic Acids Res. 21(16): 3845-3849.

Sklenár, V., H. Miyashiro, G. Zon, H.T. Miles and A. Bax (1986). "Assignment of the ^{31}P and ^{1}H resonances in oligonucleotides by two-dimensional NMR spectroscopy." FEBS Lett. 208(1): 94-98.

Sprinzl, M., C. Steegborn, F. Hubel and S. Steinberg (1996). "Compilation of tRNA sequences and sequences of tRNA genes." Nucleic Acids Res. 24(1): 68-72.

Sproat, B., F. Colonna, B. Mullah, D. Tsou, A. Andrus, A. Hampel and R. Vinyak (1995). "An efficient method for the isolation and purification of oligoribonucleotides." Nucleosides and Nucleotides 14(1 & 2): 255-273.

St. Johnston, D., N.H. Brown, J.G. Gall and M. Jantsch (1992). "A conserved double-stranded RNA-binding domain." Proc. Natl. Acad. Sci. USA 89(22): 10979-10983.

Steege, D.A., K.C. Cone, C. Queen and M. Rosenberg (1987). "Bacteriophage lambda N gene leader RNA: RNA processing and translational initiation signals." J. Biol. Chem. 262(36): 17651-8.

Stern, L. and L.H. Schulman (1978). "The role of the minor base N4-acetylcytidine in the function of the *Escherichia coli* noninitiator methionine transfer RNA." J. Biol. Chem. 253(17): 6132-6139.

Striker, G., D. Labuda and M. del Carmen Vega-Martin (1989). "The three conformations of the anticodon loop of yeast $tRNA^{Phe}$." J. Biomolec. Struct. Dynam. 7(2): 235-256.

Studier, F.W., A.H. Rosenberg, J.J. Dunn and J.W. Dubendorff (1990). "Use of T7 RNA polymerase to direct expression of cloned genes." Meth. Enzymol. 185: 60-89.

Sundari, R.M., E.A. Stringer, L.H. Schulman and U. Maitra (1976). "Interaction of bacterial initiation factor 2 with initiator tRNA." J. Biol. Chem. 251: 3338-3345.

Sussman, J.K., E.L. Simons and R. Simons (1996). "*Escherichia coli* translation initiation factor 3 discriminates the initiation codon in vivo." Mol. Microbiol. 21(2): 347-360.

Szewczak, A.A., S.A. White, D.T. Gewirth and P.B. Moore (1990). "On the use of T7 RNA polymerase transcripts for physical investigation." Nucleic Acids Res. 18(14): 4139-4142.

Szewczak, A.A.F. (1994) Natural Abundance NMR Spectroscopy and Molecular Modelling of RNA: The Solution Structure of the Sarcin/Ricin Loop from 28S Ribosomal RNA. Ph.D. thesis, Yale.

Takegoshi, K., S. Tsuda and K. Hikichi (1990). "Practical implementation of the self-refocusing 1331 solvent suppression sequence." J. Mag. Res. 89: 399-405.

Tirado-Rives, J. (1993). fit (program). New Haven, CT: Dept. of Chemistry, Yale Univ.

Valegård, K., J.B. Murray, P.G. Stockley, N.J. Stonehouse and L. Liljas (1994). "Crystal structure of an RNA bacteriophage coat protein-operator complex." Nature 371(6498): 623-626.

van den Hoogen, F. (1988) NMR Studies Concerning Base-Base Interactions in Oligonucleotides. Ph.D. thesis, Leiden.

Varani, G. and I. Tinoco (1991a). "Carbon assignments and heteronuclear coupling constants for an RNA oligonucleotide from natural abundance ^{13}C-^{1}H correlated experiments." J. Amer. Chem. Soc. 113: 9349-9354.

Varani, G. and I. Tinoco (1991b). "RNA structure and NMR spectroscopy." Quart. Rev. Biophys. 24(4): 479-532.

Varshney, U. and U.L. RajBhandary (1992). "Role of methionine and formylation of initiator transfer RNA in initiation of protein synthesis in *Escherichia coli*." J. Bacteriol. 174(23): 7819-7826.

von Pawel-Rammingen, U., S. Åström and A.S. Byström (1992). "Mutational analysis of conserved positions potentially important for initiator tRNA function in *Saccharomyces cerevisiae*." Mol. Cell. Biol. 12(4): 1432-1442.

Voorma, H.O., A.A.M. Thomas and H.A.A. Vanheugten (1994). "Initiation of protein synthesis in eukaryotes." Mol. Biol. Rep. 19(3): 139-145.

Wagner, T., M. Gross and P.B. Sigler (1984). "Isoleucyl initiator tRNA does not initiate eucaryotic protein synthesis." J. Biol. Chem. 259: 4706-4709.

Wagner, T., C. Rundquist, M. Gross and P.B. Sigler (1989). "Structural features that underlie the use of bacterial Met-$tRNA_i^{Met}$ primarily as an elongator in eukaryotic protein synthesis." J. Biol. Chem. 264(31): 18506-18511.

Wakao, H., P. Romby, E. Westhof, S. Laalami, M. Grunberg-Manago, J.-P. Ebel, C. Ehresmann and B. Ehresmann (1989). "The solution structure of the

Escherichia coli initiator tRNA and its interactions with initiation factor 2 and the ribosomal 30S subunit." J. Biol. Chem. 264(34): 20363-0371.

Waller, J.P. (1963). "The NH_2-terminal residue of the proteins from cell-free extracts of *E. coli*." J. Mol. Biol. 7: 483-496.

Walter, A.E., M. Wu and D.H. Turner (1994). "The stability and structure of tandem GA mismatches in RNA depend on closing base pairs." Biochem. 33(37): 11349-11354.

Watanabe, K., N. Hayashi, A. Oyama, K. Nishikawa, T. Ueda and K. Miura (1994). "Unusual anticodon loop structure found in E. coli lysine tRNA." Nucleic Acids Res. 22(1): 79-87.

Webster, K.R., Y. Shamoo, W. Konigsberg and E.K. Spicer (1991). "A rapid method for purification of synthetic oligonucleotides." BioTechniques 11(5): 658-661.

Westhof, E., P. Dumas and D. Moras (1988). "Restrained refinement of two crystalline forms of yeast aspartic acid and phenylalanine transfer RNA crystals." Acta Cryst. A 44: 112.

Westhof, E. and M. Sundaralingam (1986). "Restrained refinement of the monoclinic form of yeast phenylalanine transfer RNA: Temperature factors and dynamics, coordinated waters, and base-pair propeller twist angles." Biochem. 25(17): 4868-4878.

Wimberly, B., G. Varani and I. Tinoco (1993). "The conformation of loop E of eukaryotic 5S ribosomal RNA." Biochem. 32(4): 1078-1087.

Wincott, F., A. Direnzo, C. Shaffer, S. Grimm, D. Tracz, C. Workman, D. Sweedler, C. Gonzalez, S. Scaringe and N. Usman (1995). "Synthesis, deprotection, analysis and purification of RNA and ribozymes." Nucleic Acids Res. 23(14): 2677-2684.

Woo, N.H., B.A. Roe and A. Rich (1980). "Three-dimensional structure of *Escherichia coli* initiator $tRNA_f^{Met}$." Nature 286: 346-351.

Wrede, P. and A. Rich (1979). "Stability of the unique anticodon loop conformation of E. coli $tRNA_f^{Met}$." Nucleic Acids Res. 7(6): 1457-1467.

Wrede, P., N.H. Woo and A. Rich (1979). "Initiator tRNAs have a unique anticodon loop conformation." Proc. Natl. Acad. Sci. USA 76(7): 3289-3293.

Wüthrich, K. (1986). NMR of Proteins and Nucleic Acids. New York: Wiley-Interscience.

Wyatt, J.R., M. Chastain and J.D. Puglisi (1991). "Synthesis and purification of large amounts of RNA oligonucleotides." Biotechniques 11(6): 764-9.

Yarus, M., S. Cline, P. Wier, L. Breeden and R. Thompson (1986). "Actions of the anticodon arm in translation on the phenotypes of RNA mutants." J. Mol. Biol. 192(2): 235-255.

Ziemke, P. and J.E. McCarthy (1992). "The control of mRNA stability in *Escherichia coli*: manipulation of the degradation pathway of the polycistronic atp mRNA." Biochim. Biophys. Acta 1130(3): 297-306.

Zuker, M. (1989). "On finding all suboptimal foldings of an RNA molecule." Science 244(4900): 48-52.