High-Resolution Nuclear Magnetic Resonance Determination of Transfer RNA Tertiary Base Pairs in Solution. 2. Species Containing a Large Variable Loop

Ralph E. Hurd, George T. Robillard, and Brian R. Reid

ABSTRACT: The number of base pairs in the solution structure of several class III D3VN tRNA species from E. coli has been determined by analyzing the number of low-field (~15 to ~11 ppm) proton resonances in their nuclear magnetic resonance spectra at 360 MHz. Contrary to previous reports indicating the absence of tertiary resonances, all the spectra exhibit the expected number of secondary base pair resonances plus approximately ten extra resonances derived from tertiary base pairs in the three-dimensional folding of these molecules. The possible origins of some of these tertiary resonances are discussed; none of the spectra exhibits the characteristic resonance of the 8-14 tertiary base pair seen in class I D4V5 tRNA spectra.

High-resolution NMR spectroscopy has proven to be an extremely valuable tool in monitoring the solution base pairing of small nucleic acid molecules, especially tRNA. This derives from the fact that the ring NH hydrogen bond, if adequately long-lived (ca. 5 ms or longer), generates a resonance in the low-field region of the NMR spectrum (~11 to ~15 ppm); furthermore there is only one ring NH hydrogen bond for each base pair. During the past 5 years, analysis of exchangeable proton resonances had been used to study nucleic acid structure in solution and the application of high-resolution NMR in the study of tRNA folding has been reviewed by Kearns and Shulman (1974) and by Kearns (1976).

The majority of transfer RNA molecules can be divided into two major groups, namely, those with small variable loops (typically 5 nucleotides) and those with large variable loops (typically 13 to 15 nucleotides or more). The former species typically contain four base pairs in their DHU stem and are designated class I tRNAs or D4V5 tRNAs (Kim et al., 1974). The latter species typically contain three base pairs in their DHU stem and contain internal base pairs in their large variable loops; they are designated class III tRNAs or D3VN tRNAs where N is usually 13 or 15 nucleotides and can be as large as 21 nucleotides (Kim et al., 1974). The ability of all tRNAs sequenced to date to be arranged in a two-dimensional cloverleaf representation facilitates the prediction of the expected number of secondary structure base pairs in any given tRNA. Hence an extremely valuable application of the NMR approach has been to ask if additional base pairs from tertiary folding exist in the tRNA structure in solution. D4V5 tRNAs, such as yeast tRNA^Phe, typically contain 20 secondary base pairs and the low-field spectrum of this tRNA has been interpreted to contain 19-20 resonances (Jones and Kearns, 1975; Wong et al., 1975b); i.e., tertiary resonances were claimed to be absent.

More recently Kearns and his co-workers have revised their method of integration and extended their studies to include E. coli tRNA^Asp (20.2 resonances), E. coli tRNA^Trp (21.6 resonances), E. coli tRNA^Glu (21.5 resonances), E. coli tRNA^Val (23 resonances), and E. coli tRNA^Met (23 resonances); i.e., the revised estimates suggest 1-3 tertiary resonances in class I tRNAs (Kearns, 1976; Bolton et al., 1976). Our own data on 14 class I tRNAs indicate that they all contain approximately seven tertiary base pair resonances in their NMR spectra (see Reid et al., 1977) and prompted the present study on class III tRNA species. There are only a few amino acids for which the corresponding tRNAs are always class III D3VN species regardless of the biological source. For instance, bacterial tRNA^Tyr is a class III D3V13 species, whereas yeast tRNA^Tyr contains only five nucleotides in its variable loop (Barrell and Clark, 1974). However, all serine-specific tRNAs and all leucine-specific tRNAs isolated to date from viral, bacterial, yeast, and mammalian sources are class III D3VN species (Barrell and Clark, 1974).

Relatively few studies have been carried out on the solution base pairing of class III tRNAs by NMR methods compared with studies on class I tRNAs. Perhaps the most studied class III tRNA is yeast tRNA^Leu (UUG). This is a D3V13 species containing 22 secondary base pairs in the cloverleaf structure. NMR studies on this tRNA by Kearns and co-workers reported the presence of either 21.3 or 22 resonances in the low-field spectrum (Wong et al., 1973; Kearns et al., 1974a,b). A more recent analysis by this group reported an intensity of 21 ± 2 for the low-field spectrum of yeast tRNA^Leu^U^3 (UUG) and extended the study to another class III tRNA, namely, yeast tRNA^Leu^C (CUA). This latter D3V13 species contains 21 secondary base pairs; its low-field NMR spectrum was reported to contain 20 resonances (Rordorf et al., 1976). The conclusion from these studies is that the low-field NMR spectrum contains no resonances from tertiary base pairs.
We have uncovered several errors in previous NMR analyses of tertiary base pairing in class I tRNAs (see Reid et al., 1977). Consequently we decided to extend our studies to include several class III tRNAs. We have purified to homogeneity four class III leucine-accepting tRNAs from E. coli and studied their low-field NMR spectra at 360 MHz. Contrary to previous claims that the spectra contain no tertiary resonances, we find that the low-field spectra contain approximately ten resonances from tertiary base pairs.

Materials and Methods

Isolation of tRNA. There are four, and perhaps a fifth, species of E. coli leucine tRNA which can be resolved on RPCS columns (Natale and Elilat, 1976; Kelmers and Heatherly, 1971) and which can be equally well-resolved by BD-cellulose chromatography using the conditions of Gillam et al. (1967). However, there is some confusion in the literature concerning the ambiguous numbering of these isoaccepting leucine tRNAs.

First leucine species eluted during both BD-cellulose chromatography and RPCS chromatography. E. coli tRNA.LeU3 is the second most predominant leucine tRNA: it is the third leucine species eluted in the NaCl gradient and has been called tRNA.LeU3 by Kelmers and Heatherly (1971) and by Natale and Elilat (1976). E. coli tRNA.LeU5 and tRNA.LeU6 are minor leucine species and both require ethanol to be eluted from BD-cellulose. E. coli tRNA.LeU4 is the last leucine species eluted from Sepharose 4B columns and the fourth species eluting from RPCS columns just before tRNA.LeU5. E. coli tRNA.LeU5 is the first leucine species eluting from Sepharose 4B columns and the last species eluting from RPCS columns.

As mentioned earlier, the BD-cellulose ethanol fraction also contained tRNA.Tyr1 and tRNA.Tyr2 which resolved from each other on DEAE-Sephadex chromatography. The later-eluting tRNA.Tyr2 was further purified to homogeneity by Sepharose 4B chromatography and RPCS chromatography. The final material accepted 1530 pmol of tyrosine per A260 unit.

Lastly, we often observed a variable peak of leucine tRNA eluting from BD-cellulose between tRNA.LeU, and tRNA.LeU4. Further purification revealed that most of this material was tRNA.LeU1: covalently nicked in the variable loop. However, it may also contain a fifth minor leucine tRNA, i.e., tRNA.LeU5 (it would be designated tRNA.LeU5 in the Kelmers RPCS nomenclature). We have not studied this material further. The nicked tRNA.LeU5 is apparently identical with the material produced upon T2 phage infection (Kano-Sueoka and Sueoka, 1968). We note that Natale and Elilat (1976) observed only four leucine tRNA species upon RPCS fractionation of uninfected E. coli C-3000.

NMR Spectra. The purified tRNA species were dialyzed against distilled water containing 0.1 mM sodium thiosulfate at pH 7.0 and lyophilized. Aliquots of 5 mg were dissolved in 10 mM sodium cacodylate-100 mM NaCl-15 mM MgCl2 (pH 7.0) to give a final volume of 0.18 mL which was transferred to a 5 mm x 8 mm NMR microtube (Wilmad Glass Co., Buena, N.J.). Spectra were obtained at 360 MHz using either correlation spectroscopy (2400 Hz per 2 s; signal averaged for 20-30 min) or continuous wave spectroscopy (2400 Hz per 12 s; signal averaged for 4-7 h). Chemical shifts are reported in ppm from DSS (2,2-dimethylsilapentane-5-sulfonate). They were experimentally determined as chemical shifts from the water solvent to which the known chemical shift of water from DSS at the appropriate temperature was added.

Integration and Simulation of Spectra. The intensity of each peak was determined by measuring its area in cm² on the original full scale spectrum. These values were corroborated by computer simulation of the spectra using a series of Lorentzian lines having the observed line width of single resolved resonances. These Lorentzian peaks were convoluted to generate an approximation of the spectrum and the line positions adjusted until the simulated spectrum matched the experimental spectrum. The resulting computer listing then revealed the number of resonances in each peak and the chemical shifts of the component resonances.

Results

The cloverleaf sequences of E. coli tRNA.LeU1 (Dube et al., 1970), tRNA.LeU2 (Blank and Soll, 1971), tRNA.Tyr1, and tRNA.Tyr2 (Goodman et al., 1968) are shown in Figure 1. They are all class III D3VN tRNAs where N is either 15 or 13 and contain a total of 87 or 85 nucleotides. E. coli tRNA.LeU2 and tRNA.Tyr2 contain 23 normal Watson-Crick base pairs in their
secondary structure and \( E. \ coli \) tRNA\(^{Leu}_{2} \) contains 22 secondary base pairs. In addition they all contain nonstandard interactions such as GU, AC, G\( \Psi \), and A\( \Psi \) "pairing". \( E. \ coli \) tRNA\(^{Leu}_{4} \) and tRNA\(^{Leu}_{5} \) have not yet been sequenced; however, we have shown that they are both class III D3VN species by several methods. Both species have a larger chain length than the 76-nucleotide tRNA\(^{Val}_{1} \) when chromatographed on Sephadex G-200 at high temperature under the size-calibration conditions reported previously (Reid et al., 1972). Gel electrophoresis in 7 M urea at pH 7.2 easily resolved tRNA\(^{Leu}_{1} \) (87 nucleotides) from tRNA\(^{Val}_{1} \) (76 nucleotides) or any other class I tRNA. Similarly tRNA\(^{Leu}_{4} \) and tRNA\(^{Leu}_{5} \) were easily resolved from tRNA\(^{Val}_{1} \) on the basis of size in urea gel electrophoresis; yet neither could be resolved from tRNA\(^{Val}_{2} \) (Goodman et al., 1968). The cloverleaf sequences of representative class III D3VN tRNA species from \( E. \ coli \). The sequences are taken from the literature references cited in Materials and Methods.

Figure 2 shows the UV spectra of the isoaccepting leucine tRNA species from \( E. \ coli \) in the 300–380-nm range. The samples were adjusted to an absorbance of 100 at 260 nm so that the observed absorbance in the longer wavelength UV region directly reflects the percentage of the 260-nm extinction. The presence of 4-thiouridine was not detected during the sequencing of \( E. \ coli \) tRNA\(^{Leu}_{2} \) (Dube et al., 1970), and we observe no 340-nm peak for this tRNA. Although Blank and Soll (1971) did not detect 4-thiouridine in tRNA\(^{Leu}_{2} \), they point out the need to confirm this with unlabeled tRNA due to the difficulty of identifying this residue with only very small amounts of labeled tRNA. As shown in Figure 2, we observe the characteristic s4U absorption peak at 340 nm in tRNA\(^{Leu}_{2} \), tRNA\(^{Leu}_{4} \), and tRNA\(^{Leu}_{5} \); our fingerprint data and partial fragmentation of these tRNAs are consistent with this residue being in position 8 as expected. Quantitatively the 340-nm absorbance amounts to only 1.5 to 1.8% of the 260-nm extinction and this presumably reflects the 87-nucleotide chain length of these class III tRNAs. We conclude that \( E. \ coli \) tRNA\(^{Leu}_{2} \), tRNA\(^{Leu}_{4} \), and tRNA\(^{Leu}_{5} \) contain a s4U8 residue, whereas tRNA\(^{Leu}_{1} \) does not. The UV spectrum of our tRNA\(^{Tyr}_{2} \) sample contained a discrete 340-nm peak with an intensity of 3.9% of the 260-nm extinction; this reflects the presence of s4U at position 9 as well as position 8 in this tRNA (Goodman et al., 1968).

Figure 3 shows the low-field NMR correlation spectrum of \( E. \ coli \) tRNA\(^{Leu}_{2} \) at 360 MHz. Most of the intensity is located between -13.6 and -12.2 ppm in three large, poorly resolved complex peaks. If we assume that the resolved peaks at -14.2, -13.9, -11.7, and 11.4 ppm contain a single proton, then the total intensity between -15 and -11 ppm reflects the presence of approximately 30 base pairs.

The 360-MHz spectrum of \( E. \ coli \) tRNA\(^{Leu}_{2} \) is shown in Figure 4. The spectrum shows somewhat better resolution in the -14 to -12 ppm region and prompted us to attempt to integrate the complex peaks. Integration with respect to the single resonances at -14.6, -14.2, -11.9, -11.8, and -11.4 ppm gave the values indicated on the spectrum. This led to a total intensity estimate of approximately 30 protons in the -15 to -11 ppm region; the difficulty of integrating incompletely resolved peaks causes an uncertainty of at least 10% and we...
estimate that the spectrum contains 30 ± 3 low-field resonances.

The low-field NMR spectrum of E. coli tRNA\textsubscript{Leu}\textsubscript{4} is reasonably well-resolved and is shown in Figure 5. There are resolved single proton peaks at −14.5, −13.9, −13.7, −11.8, and −11.4 ppm. Based on these intensities, the peaks at −13.5, −12.3, and −11.9 ppm each contain 2 protons; the spectrum integrates to a total intensity of 33 ± 2 resonances between −15 and −11 ppm. The improved resolution in this class III tRNA spectrum encouraged us to simulate the spectrum with a series of convoluted Lorentzian lines having the experimentally observed 28 Hz line width. The lower trace in Figure 5 is the computer-simulated tRNA\textsubscript{Leu}\textsubscript{4} spectrum; it required 30 lines, not including the two protons at −11.1 ppm, and thus corroborates the independent estimate of 33 ± 2 low-field ring NH hydrogen bonds by direct integration. The computer simulation result of 32 low-field protons indicated that there were in fact 13 protons in the −12.8 to −12.4 ppm region but also suggested that the −13.5 to −13.1 ppm region contained 7 protons rather than 8 protons.

Figure 6 shows the 360-MHz NMR spectrum of E. coli tRNA\textsubscript{Leu}\textsubscript{5}. The resolution is quite good in that the seven complex peaks between −14 and −12 ppm are reasonably separated from each other. The spectrum was integrated as shown and led to a value of 33 ± 1 protons in the −15 to −11.3 ppm region. The computer simulation in the lower trace contains 32 lines of 30-Hz line width and a 20-Hz line of unit intensity at −11.4 ppm.

The last class III tRNA to be analyzed in this series was E. coli tRNA\textsubscript{Try}\textsubscript{2} and its 360-MHz low-field spectrum is shown in Figure 7. While not as well-resolved as the spectra of tRNA\textsubscript{Leu}\textsubscript{4} and tRNA\textsubscript{Leu}\textsubscript{5}, the complex peaks are nevertheless reasonably well distributed throughout the −14 to −12 ppm region. The intensity value for the various complex peaks are indicated on the spectrum; during the signal averaging of this spectrum, we experienced serious 3.3-KHz water interference at −13.9 ppm and, despite efforts to correct for it, the intensity of this peak is somewhat uncertain. In spite of these uncertainties, the tRNA\textsubscript{Try}\textsubscript{2} spectrum, like the other class III tRNA NMR spectra, apparently contains 33 ± 3 low-field protons of which approximately 10 must be derived from tertiary structure. Rordorf and Kearns (1976) have reported 25–26 low-field resonances (2–3 tertiary) in the low-field spectrum of E. coli tRNA\textsubscript{Try}.

Discussion

Since each base pair can generate only one ring NH hydrogen bond resonance in the low-field NMR spectrum, the most important question to be answered from these studies is whether or not extra base pairs from tertiary folding are detectable in solution. The class III D3VN tRNA species we have studied contain 22 or 23 base pairs; the number of stable base pairs detected in solution from the corresponding NMR spectra are listed in Table I. It is apparent that class III tRNAs show extensive tertiary interactions involving approximately 10 extra base pairs in their three-dimensional folding. This result does not agree with the class III leucine tRNA NMR studies of Kearns and co-workers (Wong et al., 1973; Kearns et al., 1974a,b; Rordorf et al., 1976). In these studies they claim that the low-field (11–15 ppm) region of the yeast tRNA\textsubscript{Leu}, (UUG) spectrum contains 21 or 22 resonances and the spectrum of tRNA\textsubscript{Leu} (CUA) contains only 20 resonances. In no cases was there even one extra resonance that could be attributed to tertiary base pairing. We are now in a position to explain the errors which led to this incorrect conclusion. The
and A14 (Kim et al., 1974). Most tRNAs instead of U at position 8 (Barrell and Clark, 1974) exhibit the s4U8-A14 tertiary base pair seen in class I11 tRNAs. Thus we conclude that class I11 tRNAs do not contain the 8-14 interaction.

Other Tertiary Interactions. The NMR data we have presented indicate that class III tRNAs utilize more tertiary base pairing in solution than class I or II tRNAs. A further possibility is that the 8-14 interaction may exist in solution but with too short a lifetime to be detected by magnetic resonance methods. Unequivocal answers to these questions must await crystallographic structure determination of these tRNAs.

The first error is that "the integrated intensity in the appropriate spectral region of the tRNALeu sample was compared with the intensity of a standard sample of yeast tRNA^Phe which exhibits 19 resonances in the 11-15-ppm region" (Kearns et al., 1974b). We have recently shown that yeast tRNA^Phe exhibits 26 ± 1 resonances in the 11-15-ppm region (see Reid et al., 1977). The second error in the case of the native tRNALeu spectrum was "the peak at 14.3 ppm was assumed to correspond to two protons and the rest of the spectrum was integrated on this basis" (Kearns et al., 1974b). It is not possible to simulate the shape of their peak at 14.3 ppm with two Lorentzian lines; the line shape of this peak can only be duplicated by three resonances. Hence the integration methods used lead to values of 66 to 75% of the correct values. Correction of these errors by the appropriate factor converts their estimate of approximately 22 base pairs to values approaching our own estimates for class III tRNAs which we find to contain even more tertiary base pairing in solution than class I tRNAs.

The 8-14 Tertiary Base Pair. The crystal structure of yeast tRNA^Phe reveals a reversed Hoogsteen base pair between U8 and A14 (Kim et al., 1974). Most E. coli class I tRNAs contain s4U instead of U at position 8 (Barrell and Clark, 1974) and also exhibit a single resonance at the extreme low-field end of their -15 to -11 ppm spectrum at approximately -14.8 ppm. The observation that this extreme low-field resonance moves upfield by ca. 0.6 ppm upon conversion of s4U8 to U8 has led us, and others, to assign the -14.8 ppm resonance to the s4U8-A14 base pair (Reid et al., 1975; Wong et al., 1975a).

The estimated intensities of the various peaks are shown on the spectrum.

The crystal structure of yeast tRNAPhe exhibits 19 resonances in the 11-15-ppm region (see Reid et al., 1975; Wong et al., 1975a). The class I E. coli tRNAs containing s4U8 which we have analyzed include tRNAVal^A1, tRNAVal^A2, tRNAVal^B1, tRNAHs, tRNA^cys, tRNA^lys, tRNA^thy, tRNA^gly, tRNA^met, tRNA^me2, and tRNA^his; we have not found a single exception to the rule that class I tRNAs containing a s4U8-A14 tertiary base pair contain a low-field resonance at -14.8 ± 0.1 ppm. However, this correlation is totally absent in class III tRNAs. E. coli tRNALeu^A1 contains no s4U but contains a resonance at -14.55 ppm; this same resonance is present in tRNALeu^A2 which does contain s4U. E. coli tRNALeu^A2 contains s4U (and also A14) and tRNALeu^A4 also contains s4U (and probably A14; Azhdarian and Reid, unpublished preliminary sequence data); neither contains a resonance lower than -14.2 ppm. There are no resonances lower than -13.9 ppm in the low-field spectra of E. coli tRNAVal^A1 and tRNAVal^B1, yet both s4U8 (and s4U9) as well as A14. Thus class III tRNAs with s4U8 and A14 in their sequence do not exhibit the -14.8-ppm resonance characteristic of the s4U8-A14 tertiary base pair seen in class I tRNA spectra. From this we conclude that class III tRNAs do not contain the 8-14 interaction and may well fold differently in this region of the molecule. The reason we do not observe the 8-14 interaction may be due to the absence of a secondary base pair involving G13 in the DHU stem in class III tRNAs. A further possibility is that the 8-14 interaction may exist in solution but with too short a lifetime to be detected by magnetic resonance methods. However, we feel this to be unlikely based on the long lifetimes of the other tertiary interactions. Unequivocal answers to these questions must await crystallographic structure determination of a D3VN tRNA.
pairs in their three-dimensional folding than do class I tRNAs. When their sequences are compared with yeast tRNA\textsuperscript{Phe}, we note the common sequences T\kCNA in the rT loop and two adjacent G residues at position 18 in the DHU loop. Hence the potential is certainly present to form the "T54-A58" base pair (65-69 in tRNA\textsuperscript{Leu} and 63-67 in tRNA\textsuperscript{Tyr}) as well as the "G18-\k55" and "G19-C56" interactions. There is obviously no class III interaction corresponding to G46-G22 in yeast tRNA\textsuperscript{Phe}; however, the complementarity between Pu15 and the nucleotide preceding the internal terminus of the rT helix discussed by Klug et al. (1974) and Kim et al. (1974) is maintained and may form a tertiary base pair analogous to the reversed Watson–Crick G15-C48 interaction in yeast tRNA\textsuperscript{Phe}. The identity of the remaining tertiary interactions in class III tRNAs remains unknown. Preliminary comparative studies on class III tRNA spectra have led us to tentatively suggest hydrogen-bonding interactions involving two of the four nucleotides in the loop at the end of the variable stem.

Acknowledgments

The authors thank Susan Ribeiro and Lillian McCollum for excellent technical assistance in purifying the various tRNA species. The use of the 360-MHz NMR facilities at the University of Groningen and at the Stanford Magnetic Resonance Laboratory is gratefully acknowledged as is the advice of Dr. W. W. Conover and Dr. S. L. Patt at SMRL.

References

Kearns, D. R. (1976), Prog. Nucleic Acid Res. Mol. Biol. 18, 91–149.