Unique secondary and tertiary structural features of the eucaryotic selenocysteine tRNA$^{\text{Sec}}$

Christine Sturchler, Eric Westhof, Philippe Carbon and Alain Krol*

Unité CNRS 'Structure des Macromolécules Biologiques et Mécanismes de Reconnaissance', Institut de Biologie Moléculaire et Cellulaire, 15 Rue Descartes, 67084 Strasbourg Cedex, France

Received December 24, 1992; Revised and Accepted February 8, 1993

**ABSTRACT**

Cotranslational insertion of selenocysteine into selenoenzymes is mediated by a specialized transfer RNA, the tRNA$^{\text{Sec}}$. We have carried out the determination of the solution structure of the eucaryotic tRNA$^{\text{Sec}}$. Based on the enzymatic and chemical probing approach, we show that the secondary structure bears a few unprecedented features like a 9 bp amino-acid-, a 4 bp thymine- and a 6 bp dihydrouridine-stems. Surprisingly, the eighth nucleotide, although being a uridine, is base-paired and cannot therefore correspond to the single-stranded invariant U8 found in all tRNAs. Rather, experimental evidence led us to propose that the role of the invariant U8 is actually played by the tenth nucleotide which is an A, numbered A8 to indicate this fact. The experimental data therefore demonstrate that the cloverleaf structure we derived experimentally resembles the hand-folded model proposed by Bock et al (ref. 3). Using the solution data and computer modelling, we derived a three-dimensional structure model which shows some unique aspects. Basically, A8, A14, U21 form a novel type of tertiary interaction in which A8 interacts with the Hoogsteen sites of A14 which itself forms a Watson–Crick pair with U21. No coherent model containing the canonical 15–48 interaction could be derived. Thus, the number of tertiary interactions appears to be limited, leading to an uncoupling of the variable stem from the rest of the molecule.

**INTRODUCTION**

Incorporation of the aminoacid selenocysteine into proteins is directed by specifically programmed UGA stop codons, both in procaryotes and eucaryotes (1, 2). Distinctive molecules in the selenocysteine insertion machinery have been shown to function in procaryotes (reviewed in ref. 3). Briefly, a special tRNA (tRNA$^{\text{Sec}}$) is charged by serine, which is then converted in situ to selenocysteyl by selenocysteine synthase. The resulting selenocysteyl-tRNA$^{\text{Sec}}$ is specifically recognized by an alternate elongation factor, SELB, which is implicated in the delivery of this tRNA to the ribosome. The *E. coli* tRNA$^{\text{Sec}}$ deviates in several positions considered to be invariant in elongator tRNAs. Most interestingly, its longer acceptor stem (8 bp) and extra arm are crucial determinants which enable this tRNA to decode reprogrammed UGA codons (4).

In eucaryotes, the existence of a particular tRNA that accepts serine was reported more than twenty years ago but was originally implicated in suppression of opal UGA codons (5). More recently, it has been shown that this minor tRNA exists in the form of selenocysteyl-tRNA$^{\text{Sec}}$ in mammals as well, suggesting that it serves as the direct donor of selenocysteine in the synthesis of selenoenzymes (6, 7). Very little is known about the eucaryotic machinery that reprograms UGA codons to insert selenocysteine (8). Remarkably, and in contrast to the procaryotic situation, the mRNA determinants required for the incorporation of selenocysteine are not located in the vicinity of the UGA codon (9) but rather far downstream in the 3′ untranslated region of the mRNA (10).

A secondary structure model has been proposed for the eucaryotic tRNA$^{\text{Sec}}$ which exhibited some variation with respect to the canonical cloverleaf structure of tRNAs (11). An alternative folding of this tRNA proposed by another group (3) renders it closer to its *E. coli* counterpart in that it possesses an acceptor stem longer than in the majority of elongator tRNAs. Discriminating between both models is obviously a prerequisite before undertaking any structure–function study of this tRNA. We therefore carried out a study to determine its solution structure. In this report, we propose both a secondary structure model based on enzymatic and chemical probing data and a tertiary structure model derived by computer modelling using the experimental data. The model sheds some light on salient features of the tRNA$^{\text{Sec}}$ secondary and tertiary structures which differ to some extent from those of the other elongator tRNAs.

**MATERIALS AND METHODS**

Cloning of the *X.laevis* tRNA$^{\text{Sec}}$ gene downstream of a T7 promoter

The parental *X.laevis* tRNA$^{\text{Sec}}$ gene was construct C286 from (12). In order to insert a T7 promoter, a PCR reaction was

* To whom correspondence should be addressed
performed with C286 as the template and primers CTGGAT-
CCTAATCAGACTCATAAGGCGGATGACCCCTCAGTCTCTG
GGG (52-mer) and GAATTGCCTGGCGCCTGCTCATCC
GGTGGAATGGAACCTCGCCGGAAAGTGGTAAGAAATC
(34-mer). The 52-mer contains the T7 promoter and a BamHI site and the 34-mer an EcoRI and BstNI site. The amplified fragment was digested with BamHI and EcoRI and then ligated to EcoRI/BamHI pUC119 vector to yield tRNA(5'( +1, +86). The 3' extended version of the tRNA was made using the same 52-mer primer and primer GAATTCAAAAGGAGAAGCCCTCGCCGGAA instead of the 34-mer.

In vitro transcription with T7 RNA polymerase
T7 RNA polymerase was purified with the help of Dr Giege's lab. The transcription reaction was performed in 250 μl of a mixture containing 25 μg of pUC119 DNA linearized with BstNI (to generate the mature CCA 3' end of the tRNA after transcription), 40 mM Tris- HCl pH 8, 22 mM MgCl2, 1 mM spermidine, 5 mM DTE, 4 mM each NTP, 0.01% Triton, 60 units RNasin and 150 units T7 RNA polymerase. The transcription medium was incubated for 4 hours at 37°C. The transcripts were purified on 12% polyacrylamide gels and electroeluted.

Enzymatic cleavages
RNase V1 cleavage occurred at 0.02 unit/μg, RNase T2 at 0.05 unit/μg and RNase T1 at 0.01 unit/μg (total tRNA added as carrier) in 10 mM Tris - HCl pH 7.5, 10 mM MgCl2, 50 mM KCl. Incubation was performed at 20°C. More details are given in (15, 16, 22).

Chemical modifications
Chemical reactions with dimethylsulfate (N7G, N1A and N3C), diethylpyrocarbonate (N7A) and carbodiimide (NIG and N3U) were used as described in (13—15) and in legends to Figures. Modifications of N1 and N2G with kethoxal were performed in 50 μl of cacodylate buffer pH 7 in which 2 to 5 μl of 20 mg/ml kethoxal were added. Reactions were stopped by adding 20 μl 50 mM borate buffer pH 7. Salt and temperature conditions were varied to provide native (10 mM MgCl2, 50 mM KCl, 37°C), semi-denaturing (1 mM EDTA, 37°C) and denaturing (1 mM EDTA, 90°C) conditions.

Detection of cleavage positions or modified bases
Positions of enzymatically-cleaved phosphodiester bonds were mapped using either molecules labeled at their 3' ends with (5'32P) pCp or by extension with reverse transcriptase of 5' labeled primers complementary to positions 47 to 57 or to positions 88 to 103 (21) of the elongated tRNA (primers TGAACCACT-CGTCGCTAGAC and AATTCAAAAAGGAGAAGCCCT, respectively).

Alkylation at N7G or N7A with DMS or DEPC, respectively, was detected after aniline-induced strand scission of 3' labeled molecules (13). Methylation at N3C was observed after strand scission (13) or by a pause of reverse transcription. The other Watson - Crick positions were analyzed by primer extension. Chemical and enzymatic sequencing with reverse transcriptase were used to map the relevant positions (13, 15, 16).

Computer modelling
The programs NAHELIX and FRAGMENT (17) were used to build standard RNA helices (the 9 bp of the acceptor stem, the 6 bp of the dihydrouridine loop and the 5 bp helix of the variable stem) and to obtain the anticodon and thymine stem-loop structures based on the crystal structure of the yeast tRNA (18), respectively. The four-base loop of the variable stem was modelled as a GNRA loop of the type first built for the X. laevis 5S RNA (19). The RNA fragments as well as the whole tRNA structure were subjected to geometrical and stereochemical least-squares refinement (18). The accessibilities of the base positions, which were chemically probed, were calculated using the program ACCESS (20). Drawings were obtained with the programs DRAWNA (Westhof, unpublished) and PLUTO (Evans and Motherwell, MRC, Cambridge).
RESULTS

A 9 bp acceptor stem characterizes the eucaryotic tRNA<sup>sec</sup>

Two hand-folded models for eucaryotic tRNA<sup>sec</sup> were originally proposed (11). One of these was discarded later by the authors themselves, presumably because it could not accommodate the nucleotide changes required by phylogenetically related new sequences (21). The first admitted model was that displayed in Figure 1A. A second one was proposed more recently (Figure 1B) which rendered the secondary structure of the eucaryotic tRNA<sup>sec</sup> much closer to that proposed for its prokaryotic counterpart (3). Note that in the present work we renumbered the nucleotides in order to have a numbering system consistent with the invariant nucleotide positions in tRNAs. Also, note in this context that positions 17, 49 and 65 (the latter two corresponding to the 49–65 base-pair of elongator tRNAs) are missing but that 67A and 67B were introduced. We now have an easier way to compare nucleotide identity and tertiary interactions between this and other tRNAs. To distinguish between the two models, the strategy was to enquire about the status of nucleotides U6 to C9, on the one hand, and C64, C66 and U67, on the other. This choice was governed by the fact that resolving the base-pairing scheme at these positions will unambiguously discriminate one model from the other.

We first looked at the enzymatic accessibilities of the Xenopus tRNA<sup>sec</sup> transcribed in vitro under the control of a T7 promoter. RNase V1 (cleaves RNA helical and stacked regions), RNase T1 (G specific) and RNase T2 (single-strand specific) were employed in this study (gel not shown). We expected such an experiment to give information about the local conformation encompassing positions G7–C9 and C64–U67A. In particular, we thought that either C64–C66 or C66–U67 phosphodiester bonds would be cleaved by RNase T2 if model A (Figure 1A) appeared to be valid. Figure 2 summarizes the cleavage data obtained for the almost entire molecule from several experiments. In addition to cleavages provoked by RNase V1 and RNases T1 or T2 in stems and loops, respectively, the surprising consequence of this experiment was that the two regions of interest, U6–G7 and C66–U67, are not RNase T2 sensitive but, instead, accessible to RNase V1. This finding was a clue that model B (Figure 1B) was more likely to be valid than model A since it is difficult to interpret the occurrence of these RNase V1 cuts unless one assumes the existence of base-pairings between U6–U67, G6–C66 and G50–C64.

To confirm the validity of model 1B, we then examined the reactivities of the bases towards various chemical probes. Reactivities of the Watson–Crick positions N1A and N3C to dimethylsulfate (DMS), N1G and N3U to carbodiimide (CMCT), N1G and N2G to kethoxal are shown in Figure 3A and B. A summary is displayed in Figure 3D. Reactivities of the N7 atoms of adenines to diethylpyrocarbonate (Figure 4A) and guanines to DMS (gel not shown) were also monitored and are summarized in Figure 4B. More than 70% of the molecule was probed in this way. The autoradiographs shown are representative examples of chemical probing. It must be pointed out that the data compiled in Figures 3D and 4B result in fact from several experiments in which the nucleotide mapping arises from either 3'-end labeled molecules or primer extension. Again, the area we will focus on, because of its strategical importance, is composed of U6–C9 and C64–U67. For unknown reasons, we repeatedly failed in obtaining information about the reactivities of N1–A8 and N7–A8 as well as N3–C9. In contrast, informative results were obtained for the remainder of the region of interest. N3–U6 is reactive to CMCT under native conditions (Figure 3B and D). N1–G7 and N2–G7 become reactive towards kethoxal under semi-denaturing conditions only (Figure 3A and D). N3–C64 is protected even under semi-denaturing conditions and becomes available to DMS alkylation under denaturing conditions only (Figure 3C and D). This reactivity provides the strongest evidence against C64 being bulged, as proposed in model A (Figure 1A). N3–C66 and N3–U67 are protected under native conditions and need semi-denaturing conditions to become reactive (Figure 3C, B and D).

The unique pairing scheme which can be proposed to fit both the enzymatic accessibilities and chemical protection data is that shown in Figure 3D. It differs from model A in that U6–U67, G7–C66 and C64–G50 are base-paired but is closer to model B. The RNase T1 cleavage between G5 and G5A is likely to be due to the local deformation of the helix caused by the G5A–U67B pair. The fact that the N3 atoms of the U6–U66 pair exhibit a differential reactivity towards CMCT is not unprecedented in RNAs (22) and does not argue against their actual involvement in base-pairing. We have no straightforward explanation for the reactivities of G52 and G53 towards kethoxal under native conditions but the reactivities of their pairing-partners C61 and C62 strongly argue for the existence of the G52–C62 and G53–C61 base-pairs in our experimental conditions. Also deserving interest is the lack of reactivity under native conditions of N1–A14, N3–U21 and N1–G15 (Figure 3D). Protection of the first two indicates that they interact to form an additional A14–U21 pair in the D-stem. The behavior of N1 and N2 of G15 as well as those of N7-purines (shown in Figure 4B) will be approached below in connection with the establishment of the tertiary structure model.

![Figure 2](image-url)
Modelling the tertiary structure revealed original structural features

During the assembly of the helices, several tertiary interactions were tested and built. We first tried to maintain tertiary interactions as similar as possible to those found in elongator tRNAs. Such an approach did not lead to difficulties with the anticodon and T-loops and for the standard 26–44 base-pair, based on the N3–C32, N3–U54, N7–A58; N3–U26 and N1–A44 protections (Figures 3D and 4B). A subset of the tertiary interactions found in tRNAs involves pairing between residues 8–14 and 15–48. In yeast tRNA^Ap and tRNA^Phe, residue 45 interacts with base-pair 10–25, while this interaction does not exist in yeast tRNA^Sec (23). In the Xenopus tRNA^Sec, several interactions are possible between positions A8 and A14, as well as between G15 and A48, or G45 and the D-helix major groove. Position 8 is so critical in the folding at the junction between the acceptor and D helices that we settled for the most conservative solution, i. e. a standard Watson–Crick pair for A14–U21, with A8 binding via its Watson–Crick sites to the A14 Hoogsteen sites. This pairing scheme accounts for the N7–A14 protection (Figure 4B). Thus, A8–A14 forms a trans Watson–Crick: Hoogsteen pair (Figure 5). We then moved to the 15–48 interaction. All pairings we could form led to excessive calculated accessibilities for either N7–G45 or N7–A48, which does not fit with the experimental data indicating they are protected (Figure 4B). Likewise, the N7–G15 protection could not be accounted for. The protection of N1–G15 and N2–G15 towards kethoxal (Figure 3D) finally led us to build a six base-pair D-helix, G15 pairing with C20A, thereby precluding a possible tertiary interaction with A48. In order to protect N7–G45 and N7–A48, which as well as N1–G45 (shown in Figure 3D), we constructed on top of the five base-pair helix of the variable region a cis Watson–Crick G45–A48 base-pair, similar to the G26–A44 pair in yeast tRNA^Phe or tRNA^Ap.
This arrangement gives the best agreement between chemical probing data and calculated accessibilities (not shown). The variable arm is thus uncoupled from the D-arm (Figure 6A). In the T-loop, the U54–A58 tertiary interaction was detected. In this loop also, N3–C56 is protected (Figure 3D) and might therefore form a Watson–Crick pair with G19. G18 could intercalate between A57 and A58. In the hairpin loop of the variable arm, N7–A47E is reactive whereas N7–G47F is protected under native conditions (Figure 4B). The loop was therefore modelled as a GNRA loop (19). There is also a possible hydrogen bond between O2–U47D and N4–C47G. Figure 6B displays a stereoview of the three-dimensional structure model for the Xenopus tRNA\textsuperscript{Sec}. The only existing long-range tertiary interactions consist in A8–A14 and the assumed interactions between the invariant G8, G19 and the T-loop.
DISCUSSION

We have carried out the solution structure determination of an in vitro transcribed Xenopus tRNA sec. We are aware that the modified bases contained in an authentic tRNA might influence the overall conformation of the molecule. In the case of tRNA sec, however, this possibility seems unlikely since the number of base modifications carried by this molecule is extremely limited compared to regular tRNAs (11). Of the two possible secondary structures originally proposed for the eucaryotic tRNA sec (3, 21) and which are displayed in Figure 1A and B, only the model we show in Figures 2, 3 and 4 is consistent with the solution structure data and compatible with the phylogeny (21). It is different from the hand-folded model A but brings experimental and additional evidence in favor to model B proposed earlier by others (3). Derivation of a 3D structure model was obtained by computer modelling. A number of interesting unusual features stem from the model, both at the secondary and tertiary structure levels. For example, the lengths of the acceptor, D- and T-stems are not canonical, being 9, 6 and 4 bp long, respectively. The length of the 4 nucleotide D-loop is not usual either. Lastly, the role of the invariant U8 is actually played by the tenth nucleotide which is an adenine in this tRNA. It has been known for some time that the E.coli tRNA sec displays a few unprecedented secondary structure features as well, like an 8 bp long acceptor stem (24). More recently, this molecule was submitted to enzymatic and chemical probing and those data used to propose a 3D model by computer modelling as well (C. Baron, E. Westhof, A. Böck and R. Giegé, submitted). These authors arrived to the conclusion that the D-stem is also 6 bp long. It must be pointed out here that in addition to the use of experimental data and structural constraints, establishment of the 3D models for both molecules was made possible thanks to the fact they illuminate each other.

Also deserving interest is the finding that a purine occurs in place of the invariant U8. Although a uridine is the eighth base from the 5' end, both the experimental data and the extra length of the tRNA sec led us to conclude that the uridine, numbered U6, cannot correspond to the invariant U8. Conspicuously, all tRNA sec lack the invariant U8 which is replaced by a purine at this position: an adenine is found in P. vulgaris and eucaryotes, a guanine in E.coli (25, 21, 26). In tRNAs, the U8-A14 interaction is universal. The observation that a purine nucleotide, numbered 8, was found instead of U8 in tRNA sec raised the issue of how this interaction can be preserved. This was solved in the eucaryotic tRNA sec by the finding of a novel type of tertiary interaction, forming the triple A8—A14—U21. In the prokaryotic tRNA sec, the triple G8—U14—A21 exists also but with different pairings (C. Baron et al., submitted). In the Xenopus tRNA sec, a C is found at position 9. In tRNAs, position 9 is mostly occupied by a purine which forms a triple with base-pair 12—23 (26). Other eucaryotic tRNA sec also contain a purine at position 9 (21). We could not put forward an interaction involving C9 in the Xenopus tRNA sec due to the lack of chemical data at this site, although a contact between N4—C9 and the Hoogsteen sites of G23 cannot be excluded. No coherent model could accommodate the canonical 15—48 interaction. Instead, G15 is involved in a Watson—Crick pair with C20A, thereby extending the D-stem by one base-pair. The variable stem is closed by a G45—A48 Watson—Crick pair and connected to the D-anticodon coaxial helix by the Watson—Crick U26—A44 tertiary pair only. This results in an uncoupling of the variable stem from the rest of the molecule. It appears that the number of tertiary interactions is limited in this tRNA, resulting in loose D-loop/T-loop interactions. This is reflected, for example, by the slight accessibility of the D-stem to RNase T2 (Figure 2).

One last point is worth discussing. The L-shape of tRNAs is formed by the coaxial stacking of the acceptor and T-stems, on the one hand, and of the D- and anticodon stems, on the other. In the E.coli tRNA sec, the acceptor and T-stems are 8 bp and 5 bp long, respectively, providing a 13 bp long coaxial helix. We have shown that in the eucaryotic tRNA sec, the acceptor stem is 9 bp long. Remarkably, the length of this coaxial helix has been conserved to 13 bp through evolution by reducing to 4 bp, in the eucaryotic tRNA sec, the otherwise 5 bp long T-stem. Work done by others (27) showed that an 8 bp long acceptor stem in the E.coli tRNA sec is the determinant for binding to the elongation factor SELB. Our observation that the 13 bp coaxiality is conserved both in procaryotes and eucaryotes leads to functional implications in line with those reported in (4, 27). It reinforces the fact that the tRNA sec has evolved unique structural features in order to respond to the selenocysteine insertion machinery.

ACKNOWLEDGEMENTS

F. William Studier, the members of the Giegé's lab, G. Keith and A. Przykorska are thanked for their gifts of T7 gene 1, T7 RNA polymerase and help at various stages of the work, respectively. A. Hoeft is thanked for oligonucleotides synthesis. We are grateful to R. Giegé, J. Rudinger, M. Frugier and P. Romby for stimulating discussions and advice. C. Baron and A. Böck are especially thanked for making available their data prior to publication. This work was partly supported by a grant from the Institut National de la Santé et de la Recherche Médicale (Contrat de Recherche Externe 920107).

REFERENCES

19. Westhof, E., Romby, P., Romaniuk, P. J., Ebel, J. P., Ehresmann, C. and
21. Lee, B. J., Rajagopalan, M., Kim, Y. S., You, K. H., Jacobson, K. B.,
   Res. 17, 7159–7165.
   2529–2540.
   387–397.