Université de Montréal

Role of tertiary interactions in determining RNA architecture

par
Anatoli Ioudovitch

Département de Biochimie
Faculté de Médecine

Thèse présentée à la Faculté des études supérieures
En vue de l'obtention du grade de
Philosophiæ Doctor (Ph.D.)

Avril 1999

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée:

Role of tertiary interactions in determining RNA architecture

présenté par

Anatoli Ioudovitch

a été évaluée par un jury composé des persones suivantes:

Président-rapporteur: Franz Bernd Lang
Directeur de recherche: Robert J. Cedergren
Co-directeur; Stephen Michnick
Membre du jury: François Major
Examinateur externe: Stephen Harvey
Représentant du doyen: Serguei Chteinberg

Thèse acceptée le: 22 juillet 1999

SUMMARY

An important step in understanding the mechanisms of any biochemical process is the realization of the relationships between the structure and function of its components. These relationships can be interpreted in terms of correlation between particular details of the architecture of the functional site and the performed function. In this work transfer RNAs have been chosen as a model for systematic investigation of such correlations. Among different intra-molecular interactions stabilizing the architecture of RNA molecules, the tertiary interactions play a very significant role. These interactions are built mainly via formation of hydrogen bonds and base-base stacking. In spite of the three-dimensional character of tertiary interactions, their formation requires a particular sequence pattern, which can be recognized by comparative analysis of related sequences. The first step of this analysis consists of the compilation of all available tRNA nucleotide sequences, their alignment, annotation and correction. The published Compilation of tRNA sequences and sequences of tRNA genes (Chapter I) is the result of collaborative efforts in this direction. The aligned sequences of cytosolic tRNAs are very alike except for the selenocysteine tRNAs. The analysis of the structure-function relationships in these unusual tRNAs is presented in Chapters II and III. This analysis shows that despite the notable deviation of the secondary structure of the selenocysteine tRNAs from the standard one, their three-dimensional architectures satisfy the general tRNA structural constrains. In the case of the eukaryotic $tRNA^{Sec}$, the available information has allowed to model the tertiary structure of this tRNA (Chapter IV). The core region of the model has a structural motif similar to that seen in all other known Class II tRNA structures. Another interesting aspect of the tRNA structure, which was revealed during the analysis of the cytosolic tRNAs, has dealt with relationship between the nucleotides not directly involved in any contacts and the formation of tertiary interactions. For nucleotides involved in tertiary interactions and concentrated in a relatively small region of the sequence, the maintenance of their interactions may be sterically impossible without any intervening nucleotides. This proved to be the case for two nucleotides, 46 and 48, involved in the formation of the core tertiary interactions 21-46 and 15-48 in the tRNA

structure. The presence of nucleotide 47 allows the formation of both these interactions without restrictions, while the absence is compensated by a non-canonical base pair U13-G21 (Chapter V). The presented results show that the theoretical approach connecting the primary structure and the function via modeling the elements of the tertiary structure can be fruitful for understanding different types of structure-function relationships.

RÉSUMÉ

La détermination des relations entre la structure moléculaire et le rôle de ses composants constitue le premier pas en vue de la compréhension de n'importe quel processus dans le domaine de la biochimie. Ces relations peuvent être considérées en tant que corrélations existant entre des détails particuliers de l'architecture du site fonctionnel et de la fonction à remplir. L'architecture des molécules d'ARN est créée par différents types d'interactions intramoléculaires parmi lesquelles les interactions tertiaires jouent un rôle significatif. Généralement, toutes les interactions nucléotide-nucléotide, hormis les hélices doubles de type Watson-Crick, sont nommées des structures tertiaires. Ces interactions incluent la formation de divers types de ponts hydrogène et d'interactions de superposition. Dans le cadre de ce travail, différents aspects des relations existant entre la structure tertiaire et la fonction chez les ARN de transfert (ARNt) ont été étudiés.

La capacité de renaturation des ARNt suggère que les éléments nécessaires à un repliement adéquat soient présents dans la séquence. Par conséquent, une analyse systématique de la séquence des ARNt peut fournir une excellente source d'information quant aux interactions tertiaires, de leur variabilité chez différentes espèces d'ARNt ainsi que de leur rôle dans le repliement. Évidemment, la première étape de cette analyse est la compilation de toutes les séquences disponibles d'ARNt, de leur alignement, de leur annotation et, dans certains cas, des corrections s'y rattachant. La compilation des séquences d'ARNt et des séquences de gènes d'ARNt "The Compilation of tRNA sequences and sequences of tRNA genes" (Chapitre I), constitue le fruit d'efforts collectifs en vue d'atteindre ce but.

De façon générale, tous les ARNt peuvent être séparés en deux groupes tout dépendant de leur origine. Le premier groupe comprend tous les ARNt cytoplasmiques comportant des éléments de séquence très bien conservés. Le second groupe est constitué d'ARNt provenant de différents organites, de certains virus et de bactéries symbiotiques, et où les éléments conservés présents dans le premier groupe disparaissent en tout ou en partie. Tous les ARNt cytoplasmiques ainsi que plusieurs organites peuvent être repliés uniformément en un diagramme "en feuille de trèfle" représentant leur structure

secondaire. Dans ce diagramme, les éléments de séquence conservés occupent toujours la même position. De plus, la longueur de tous les domaines hélicoïdaux, hormis un, est déterminée très strictement. Les caractéristiques universelles de la structure secondaire "en feuille de trèfle" comprennent cinq paires de bases dans la tige T, sept paires de bases dans la tige acceptrice, trois ou quatre paires de bases dans la tige D et six paires de bases dans la tige de l'anticodon. Seule la région du bras supplémentaire, dont la longueur peut varier de seulement quatre nucléotides à quelques douzaines, fait exception. Habituellement, les ARNt dans lesquels le bras supplémentaire est suffisamment long pour former une structure tige-boucle sont classifiés en ARNt de Classe II tandis que les ARNt possédant un bras supplémentaire court sont des ARNt de Classe I.

Quel que soit le critère considéré, les ARNt sélénocystéine ($ARNt^{Sec}$) représentent un type exceptionnel d'ARNt. En effet, leur structure secondaire diffère de façon significative de celle de tous les autres ARNt cytoplasmiques. Deux structures secondaires d'$ARNt^{Sec}$ eucaryotes se distinguent, toutes deux satisfaisant aux caractéristiques des séquences apparentées de phylogénie. Elles présentent respectivement sept et cinq paires de bases dans la tige acceptrice et la tige T (structure 7/5) ou encore neuf et quatre paires de bases (structure 9/4). Bien que la structure 7/5 soit la seule capable de maintenir la juxtaposition normale des domaines T et D telle que présente chez les autres ARNt cytoplasmiques, la fonction unique des $ARNt^{Sec}$ laisse toujours une possibilité qu'ils ne correspondent pas au squelette standard des ARNt. Afin d'établir une distinction entre les structures secondaires 7/5 et 9/4 des $ARNt^{Sec}$ eucaryotes, l'analyse des résultats expérimentaux disponibles sur la sérylation, la sélénylation et la phosphorylation de différents mutants des $ARNt^{Sec}$ eucaryotes a été effectuée (Chapitre II). Il a été démontré que plusieurs de ces mutants, incapables de se replier en une structure 9/4, étaient actifs dans les différents processus enzymatiques tandis que la perte de leur capacité à se replier en une structure 7/5 était dommageable pour la fonctionnalité. Ainsi, les résultats de l'analyse corroborent bien le fait que les $ARNt^{Sec}$ eucaryotes possèdent une structure secondaire 7/5. En se basant sur les résultats de cette analyse ainsi que sur la comparaison des séquences de nucléotides disponibles, un nouveau modèle tridimensionnel des interactions tertiaires de la région centrale des $ARNt^{Sec}$ eucaryotes a

été proposé (Chapitre IV). Le modèle suggère un système unique d'interactions tertiaires dans la région entre le grand sillon de la tige D et la première paire de bases du bras supplémentaire, lequel ne jouira d'aucune flexibilité quant à son orientation. L'importante similarité entre le modèle proposé et la structure connue d'un ARNt de Classe II, l'ARNt^{Ser}, est illustrée.

La tige T de l'ARNt^{Sec} de l'archéobactérie *Methanococcus jannascii* contient seulement quatre paires de bases, soit une paire de bases de moins que dans tous les autres ARNt cytoplasmiques. Notre analyse de la structure moléculaire (Chapitre III) indique qu'une telle tige T ne peut permettre qu'une interaction normale entre les boucles D et T ait lieu. Elle affecte donc la juxtaposition de ces deux domaines en hélice altérant par le fait même la fonction de l'ARNt. De plus, cet ARNt possède une autre caractéristique inhabituelle, soit une tige D particulièrement longue constituée de sept paires de bases qui pourrait aussi rompre l'interaction normale des boucles D et T. Cependant, grâce à des techniques de modélisation moléculaire, il a été prouvé que l'effet compensatoire de la petite tige T et de la grande tige D produit une juxtaposition normale des domaines. Dans le cas des nucléotides impliqués dans les interactions tertiaires et qui sont concentrés dans une région relativement petite de la séquence, le maintien des interactions en question s'avère parfois impossible, en raison de considérations reliées à la stéréochimie, en l'absence de nucléotides additionnels. Le rôle structural d'un nucléotide qui relie deux nucléotides impliqués dans une interaction tertiaire importante a été analysé dans le cas du nucléotide 47 des ARNt (Chapitre V). La présence de ce nucléotide dans la structure de l'ARNt^{Phe} de la levure permet la formation des interactions tertiaires canoniques 15-48 et 22-46 dans le domaine D. Par contre, la formation de l'une de ces interactions tertiaires s'avère impossible en l'absence du même nucléotide. Toutefois, cette situation peut être compensée par la présence d'un flottement (wobble base pair) U13-G22. L'analyse de la banque de données des ARNt démontre que la grande majorité des ARNt cytoplasmiques possèdent soit un nucléotide à la position 47, soit une paire U13-G22.

TABLE OF CONTENTS

List of tables

List of figures

Introduction

Chapter I

Chapter II

Chapter III

Chapter IV

Chapter V

Abbreviations

bp	base pair
euk	eukaryotic
HDV	hepatitis delta virus
HIV	human immunodeficiency virus
MMTV	mouse mammary tumor virus
NMR	nuclear magnetic resonance
nt	nucleotide
prok	prokaryotic
R	purine
RNA	ribonucleic acid
RRE	Rev response element
rRNA	ribosomal RNA
Sec	selenocysteine
tRNA	transfer ribonucleic acid
TYMV	turnip yellow mosaic virus
WC	Watson-Crick
wt	wild type
Y	pyrimidine

To my family

and also

to the memory of Robert Cedergren

Introduction

1. Structural motifs

Early studies of protein and nucleic acid structure showed that different molecules often contained similar structural elements (Rao & Rossmann, 1973). By now, such elements have been identified at different levels of structural organization. In both proteins and nucleic acids, they may be a part of the secondary structure, a particular tertiary arrangement or even a single interaction. It is generally assumed that the existence of such similarities, usually described as motifs, reflects resemblance either in the function or in folding of the molecule. In RNA, motifs can be found almost at any level of their organization, from sequence patterns to intricate tertiary arrangements.

Based on crystal and NMR RNA structures, a large number of motifs have been described so far, although most of them have been seen only in few structures. The best known RNA motif is the U-turn, which was first observed in the crystal structure of the yeast $tRNA^{Phe}$ (Kim *et al.*, 1974; Robertus *et al.*, 1974). This motif refers to the nucleotide conformation and the system of nucleotide-nucleotide interactions in a sharp turn of the RNA polynucleotide chain (Fig 1). Although the fine details of the conformation within the same motif can vary from molecule to molecule, the common structural organization makes motif description a very powerful tool in the studies of structure-function relationships. For example, the tRNA L-shape that describes the orientation of the tRNA helical domains and the set of interactions necessary to achieve it is a structural motif common for all tRNAs. However, a conformation of the polynucleotide chain in each of tRNA species can be different.

The difficulties associated with the determining the biomolecular structure and the rapid accumulation of sequence information have pushed forward the development of approaches to identify structural motifs in gene sequences. Generally, it is assumed that the sequence *per se* contains sufficient information to guarantee the proper folding. The problem is to decipher this information and distinguish it from that information which is "unimportant" for the structure but is also encoded in the gene. A possible solution is to identify sequence patterns that correspond to the known structural motifs. However, many

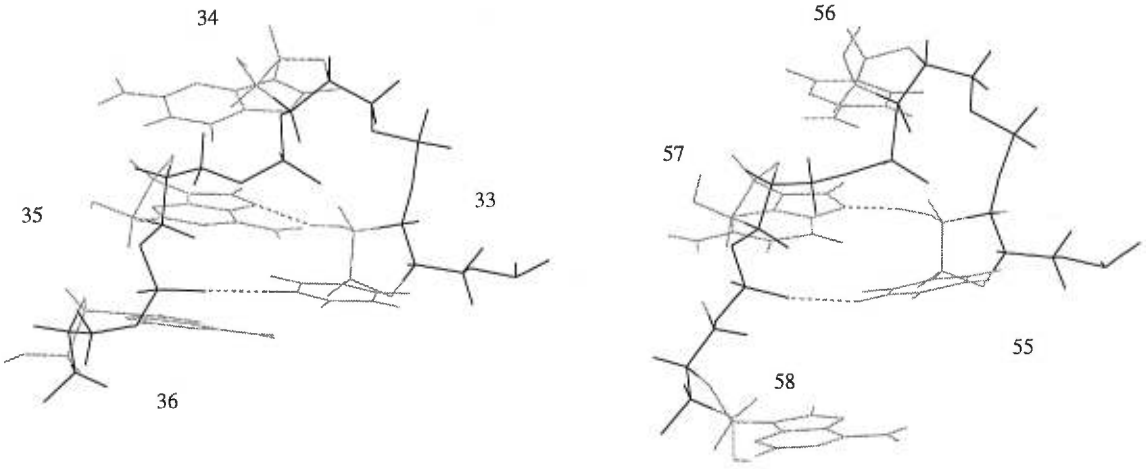


Figure 1. The U-turn motif in the anticodon loop (left) and in the T-loop (right) of the yeast $tRNA^{Phe}$ (Robertus *et al.*, 1974; Kim *et al.*, 1974). Important hydrogen bonds are shown as broken lines. The nomenclature of nucleotides is taken from Chapter I.

structural motifs found in RNA, like the above mentioned U-turn, are not sequence-specific. Another strategy to predict structure from sequence is to use comparative sequence analysis of all phylogenetically related sequences, assuming that in most cases similar structural elements are expected to have similar sequences. This approach has been more or less successfully used in the RNA secondary structure prediction, when a sufficient number of homologous nucleotide sequences was taken for analysis (Woese *et al.*, 1983; Michel, *et al.*, 1989). It is obvious that the major problem of this approach is that, on one hand, very conserved regions do not provide any useful information, while on the other hand, very dissimilar regions are very difficult to align. Phylogenetic comparison of the structures aligned by their secondary structure has been used in a number of cases to predict base-base tertiary interactions or motifs (Levitt, 1969).

A special class of structural motifs observed in different molecules is characterized by the presence of compensatory effects. A potentially disrupting change in a motif observed in one or several homologous molecules can be compensated by another change in a different part of the same molecule. An example of this effect can be found in the coaxial arrangement of RNA double helices. If for whatever reason only the total length of the domain made of two coaxial helices is important, the shortening of the one helix will be compensated by the extension of the other. The presence of such a compensation has been used as an indicator of coaxiality between helices in the ribosomal RNAs (Woese *et al.*, 1983).

2. RNA structure

Folded RNA molecules are stabilized by a variety of interactions, the most prevalent of which are base stacking and hydrogen bonding between bases. Generally, the interactions found in a three-dimensional RNA structure can be divided into two categories: secondary interactions and tertiary interactions. RNA secondary interactions are Watson-Crick interactions between the bases in the anti-parallel double helix. They are represented on a scheme of base paring (secondary structure) by a nonintersecting

line, which connects the paired bases. Tertiary interactions occur when elements of the secondary structure interact with each other.

2.1 Secondary structure

The secondary structures of real RNA molecules contain a significant number of unpaired regions. According to their place in the secondary structure, unpaired regions can be hairpin loops, internal loops, bulges or connector regions (bifurcation loops) in junctions (Fig 2).

The secondary structure *per se* does not provide any information regarding spatial arrangements of its elements. Structural information accumulated so far can help clarify how these elements are arranged and what are the motifs of their general folding. Taken together, the single stranded regions and their conformations can be considered as blocks from which the overall three-dimensional structure is built.

2.1.1 Double helix

A Watson-Crick type RNA duplex forms a right-handed helix (so-called "A-form") with two strands being in the antiparallel orientation (Fig. 3). Nucleotides in this helix have a C3'-*endo* sugar pucker with the distance between the neighboring phosphates of about 5.9 Å. As a result of the base pair displacement of approximately 4.4 Å from the helical axis and of the positive base pair tilt angle of about 16-19°, the RNA double helix has a very deep major groove and a rather shallow minor groove (Saenger 1984).

At a low ionic strength, the A-RNA double helical conformation with 11 base pairs per turn predominates. Increasing the ionic strength triggers transformation of the A-RNA to the A'-RNA characterized by 12 base pairs per turn. These two conformations differ mainly in the pitch parameter, which is about 30 Å in the A-RNA, while 36 Å in the A'-RNA (Arnott *et al.*, 1973). The helical repeat in solution for the double stranded

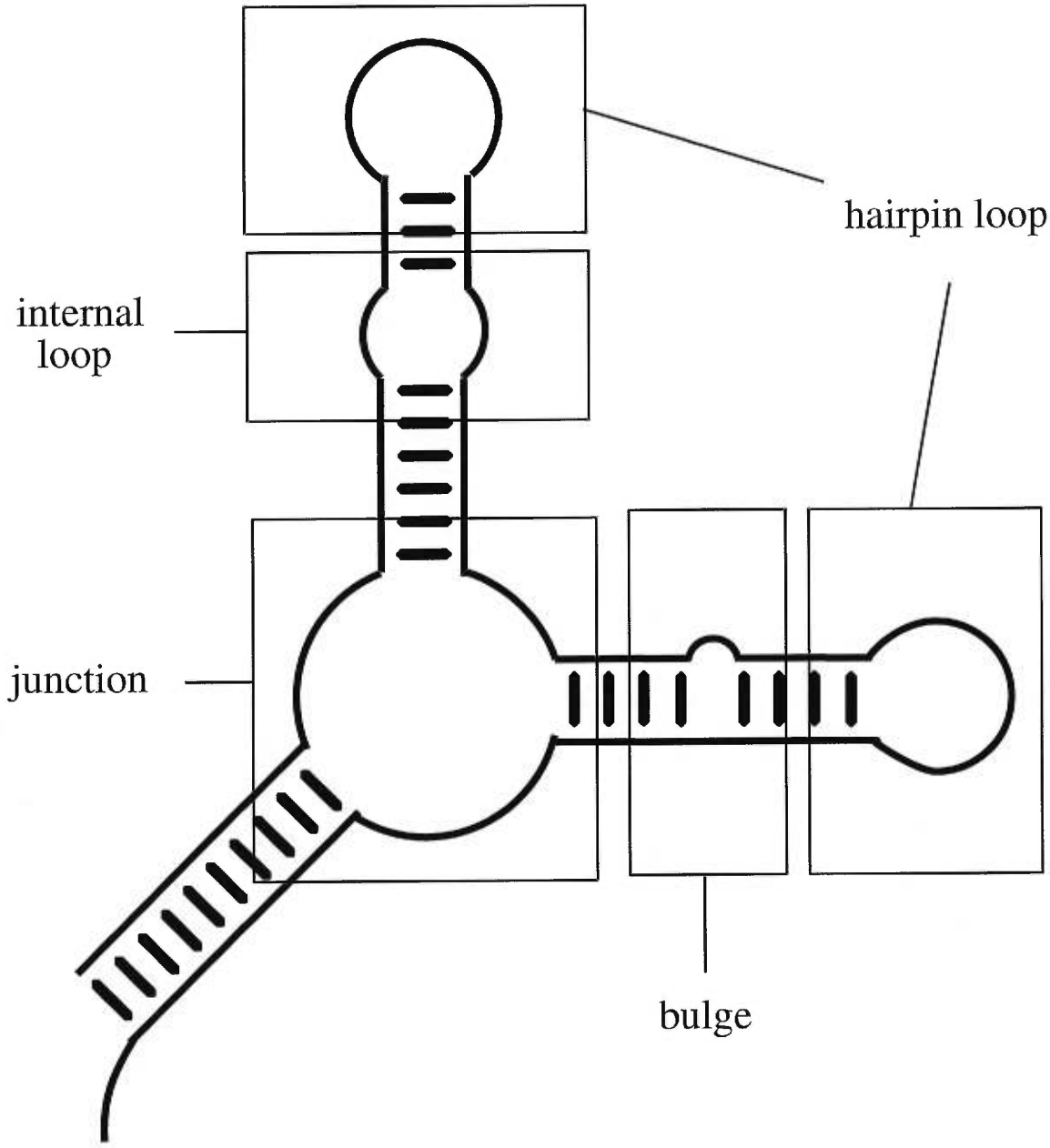


Figure 2. RNA secondary structure elements.

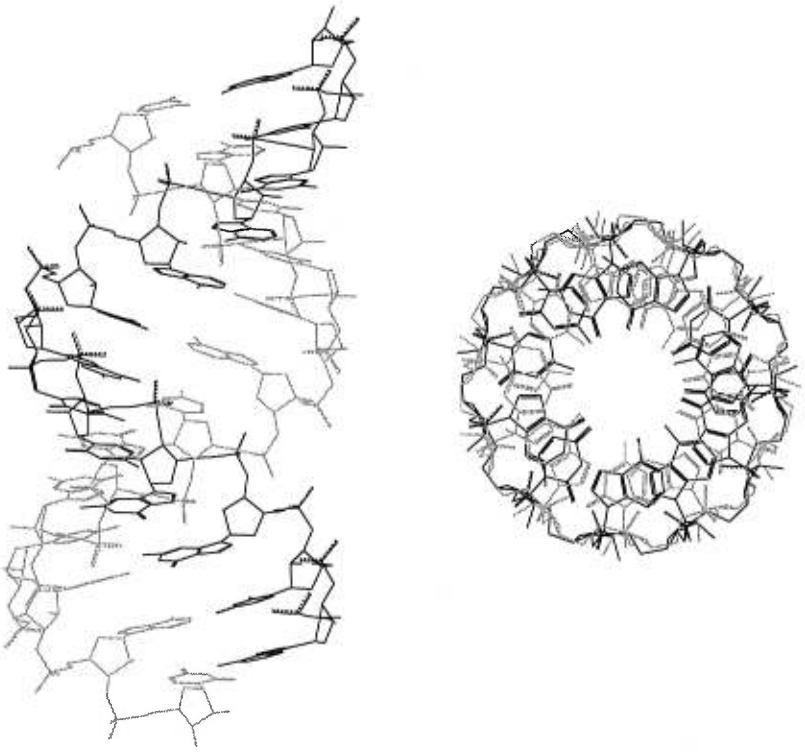


Figure 3. RNA double helix. Two projections are presented: perpendicularly to the axis of the helix (left) and along the helix (right). Two strands are shown in black and gray.

RNA has been shown to be between 11.3 and 11.6 base pairs per turn (Tang & Draper, 1990).

Under unusual conditions some other double helical structures have been observed. This includes the Z-RNA in which the alternating G-C base pairs are arranged in the left-handed helix (Hall *et al.*, 1984). Another example is the parallel double helix of poly(2-methylthio-A)-poly(U) with Hoogsteen type of base pairing (Hakoshima *et al.*, 1981). The biological relevance of these structures, if any, is not known.

2.1.2 Hairpin loops

As seen in Figure 2, a hairpin loop is formed when RNA folds back on itself. Hairpin loops are probably the most abundant elements of the secondary structure. They can contain as few as two nucleotides, but there is virtually no upper limit. Still, after a certain length, large hairpin loops do not exist in a self-sustained conformation and tend to be involved in inter or intra-molecular interactions. Recent advances in determination of the structure of relatively short RNA molecules by NMR have shed the light on their conformation.

The known three nucleotide loops are usually considered unstructured. NMR studies of the oligo-rCGC(UUU)GCG showed that the final model of the loop determined by restrained molecular dynamics lacks any stacking interactions within the loop with all nucleotides having adopted the C2'-*endo* conformation, despite the fact that the NOE-connectivity data suggested some stacking within tri-uridine loop (Davis *et al.*, 1993). A somewhat similar situation with the tri-uridine loop without any internal contacts was observed in the crystal structure of the 5abc region of the Group I intron (loop 6, Cate *et al.*, 1996).

Tetraloops are probably the most studied RNA hairpin loops. Loops with sequences GNRA and UNCG have been shown to predominate in the bacterial ribosomal RNAs (Woese *et al.*, 1990). Several solution structures for both loop types have revealed

common features of nucleotide interactions within the loops (Heus & Pardi, 1991; Varani *et al.*, 1991; Szewczak *et al.*, 1993; Allain & Varani, 1995). These features include the formation of a non-Watson-Crick base pair on top of the stem and the stacking of a third nucleotide to this pair. In the case of the GAAA and UUCG tetraloops, the second nucleotide is also stacked to the rest of the loop, forming a structure very similar to the U-turn found in the anticodon and T-loops of the tRNA (Fig. 4). The two middle nucleotides have the C2'-*endo* conformation, which helps to reverse the direction of the chain. On the other hand, the C3'-*endo* conformation of the nucleotides at both ends of the loop provides a decent stacking to the adjacent stem. Based on the stereochemical analysis of these tetraloops, Kajava & Rüterjans (1993) suggested that all stable conformations of different tetraloops depend on the type of the base pair formed by the first and the last nucleotides of the loop. However, a solved later structure of the CUUG tetraloop (Jucker & Pardi, 1995) had a conformation different from that suggested by the theoretical analysis. Strictly speaking, it was not a tetraloop at all, since the flanking C and G formed a normal Watson-Crick base pair, while the third nucleotide of the loop, uridine, stacked to the guanine. Another unusual structure has been observed in the loop AGUU of SL1 RNA from *Caenorhabditis elegans* where the adenine does not form even a single H-bond with the opposite uridine, while both stacked to the stem (Greenbaum *et al.*, 1996). Under unusual conditions and/or with help of unusual nucleotides, tetraloops can acquire new alternative conformations. A GNRA-like tetraloop containing N2-methylguanosine and two N6, N6-dimethyladenosines has quite a flexible conformation in which only m^2G stacks to the stem (Rife & Moore, 1998). Influence of metal ions on RNA conformation is very significant, as it has been highlighted by the solution structure of UGAA tetraloop (Butcher *et al.,* 1997a). In the absence of Mg^{2+} this tetraloop does not have a U-turn-like conformation, contrary to what one might expect from its sequence. Instead, it forms a turn between the second guanine and the third adenine, thus making the 3' and 5' sides of this loop equal in the number of stacked nucleotides.

The current knowledge of pentaloop structures is rather scarce. Only two examples of these loops with similar sequences GUUUC and GUCUC are known in which the loop is free of intra or inter molecular interactions (Sich *et al.*, 1997; Dallas &

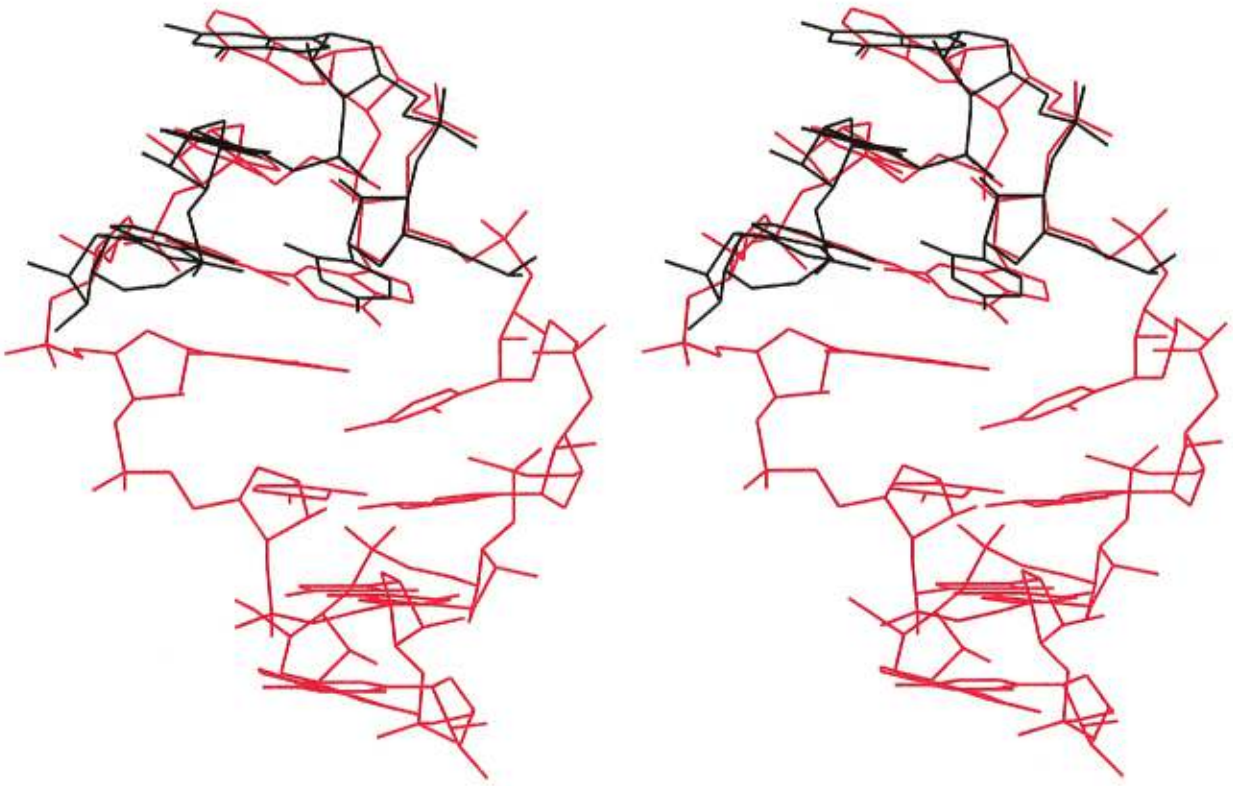


Figure 4. Stereo-drawing of the U-turn motif in the GAAA tetraloop (red; Jucker *et al.*, 1996) and in the anticodon loop of the yeast $tRNA^{Phe}$ (black; Robertus *et al.*, 1974). Despite the differences in the identities of the nucleotides involved in the U-turn in both molecules, the overall structures are very similar.

Moore, 1997). Both structures are characterized by the absence of the Watson-Crick GC-base pair as the closing pair of the loop. In the first structure, the terminal guanine and cytidine do not even stack on top of the stem, which leaves the loop completely unstructured. In the second structure, these nucleotides form a somewhat disturbed Watson-Crick pair with only one hydrogen bond, while neither of the three intermediate pyrimidines is involved in any particular interactions.

Three out of the four known hexaloop structures, the GUAAAA loop from HIV-1 (Puglisi & Puglisi, 1998), the GUAACA loop in the U2 snRNA from *Saccharomyces cerevisiae* (Stallings & Moore, 1998) and the GUAAUA loop from the prokaryotic large subunit rRNA (Huang *et al.*, 1996) display very similar conformations, as one could expect from their sequence similarity. The closing G-A base pair is formed in a manner similar to that in the tetraloops. The second, the third and the forth nucleotides of the loop form the U-turn. The only difference between the structures of these loops deals with the conformation of the fifth nucleotide of the loop. The adenine in the GUAAAA loop and the cytidine in the GUAACA loop stay within the stacked part of the loop, making this structure very similar to that of the tRNA anticodon loop. Uridine in the GUAAUA loop, on the contrary, is excluded from the stack on the 3' side of the loop. The hexanucleotide loop CUCGGA from TAT RNA appears to be disordered (Aboul-Ela *et al.*, 1996).

The only example of a loop structure not involved in any inter- and intra-molecular interactions has been for a long time the anticodon loop of the yeast $tRNA^{Phe}$ (Kim *et al.*, 1974; Robertus *et al.*, 1974). This is a heptaloop in which five nucleotides stack on the 3' side and two nucleotides stack on the 5' side. The sharp bend between the two stacks has a U-turn conformation (Fig 1, 4 and 9). The crystal structures of other tRNAs, which were determined later, displayed a conformation almost identical to that in the yeast $tRNA^{Phe}$ (Moras *et al.*, 1980; Rould *et al.*, 1991). The only other known structure of a heptaloop not involved in intra-molecular interactions is the UCCUCGC loop from the fragment of the HDV antigenomic ribozyme. This loop has rather a disordered structure with a weak two-pyrimidine stack on the 5'side of the loop (Kolk *et al.*, 1997). A subsequent crystal structure of the larger fragment of the HDV showed this loop participating in intra-molecular interactions (Ferré-D'Amaré *et al.*, 1998).

The only known structure of a loop containing more then seven nucleotides in a self-sustaining conformation is a nominally nine-membered loop AUUUCUGAC. NMR studies have shown that the structure of this loop resembles that of the loops with three nucleotides. The disordered terminal loop UCU with all nucleotides in the C2'-*endo* conformation is closed by three base pairs U-G, U-A and A^+-C of which only the U-A base pair is of the Watson-Crick type (Puglisi *et al.*, 1990).

It should be noted that hairpin loops involved in inter or intra-molecular interactions often adopt conformations quite different from those in the free state. Their conformations can change depending on the ionic strength and/or pH of solution, the presence of different ligands, etc. An extreme example of such changes is the structure of oligonucleotide r(GGACUUCGGUCC), which forms in solution a hairpin with a UUCG tetraloop, while a non-canonical double helix in crystals (Kanyo *et al.*, 1996).

2.1.3 Bulge loops

A bulge loop (or simply a bulge; Fig 2) is an irregular region of a double helix where one of two strands has an unpaired nucleotide or nucleotides. Depending on their identity (purine or pyrimidine) and on the surrounding nucleotide context, single nucleotide bulges can either be a part of the helix or be exempt from the helical stack (Chastain & Tinoco, 1991). Not much is known about general behavior of the bulges consisting of more than one nucleotide, although two NMR structures are known to contain this element. These structures, the A-rich internal bulge from the 5abc region of the group I intron (Luebke *et al.*, 1997) and the TAR *cys*-acting RNA regulatory element in HIV-1 (Aboul-Ela *et al.*, 1996) display a similar bending of the RNA double helix of about 90° at the place of the bulge. Only the 5'-uridine of the UCU TAR bulge stacks on the 5' neighboring helix, while the other two pyrimidines are excluded from stacking. The situation is, however, quite different in the group I intron bulge, where two consecutive adenines on both sides of the bulge stack on their neighboring helices without any interaction between the two stacks. The uridine is excluded from both stacks and serves as a connector between them (Fig 5). Interestingly, the structures of both bulges display

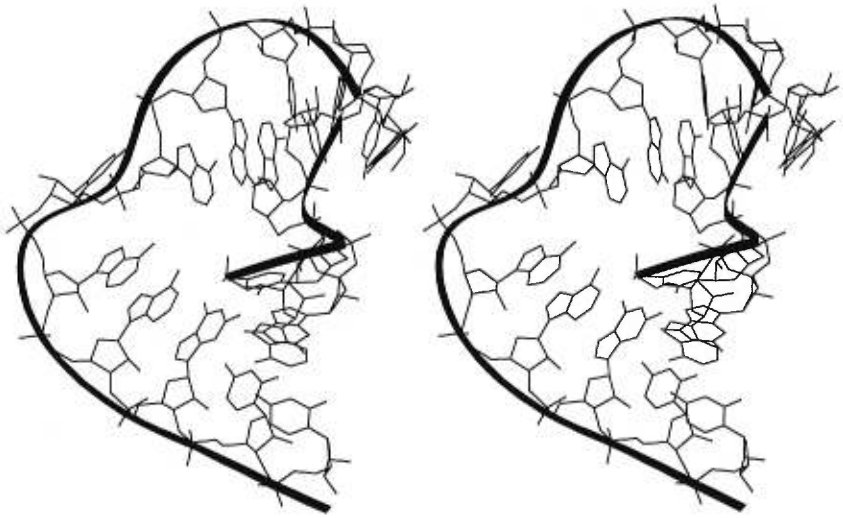


Figure 5. Stereo-drawing of the structure of the A-rich internal bulge from the 5abc region of the group I intron (Luebke *et al.*, 1997).

different conformations in the presence of ligands and while participating in some intra-molecular interactions. In the structure of the TAR-arginine complex (Puglisi *et al.*, 1992; Aboul-Ela *et al.*, 1995) the bend in the helix is smaller than in the uncomplexed RNA and all nucleotides of the bulge are excluded from the stacking interaction with the adjacent parts of the helix, while the 5' uridine forms a base triple with the A-U base pair of the stem. In the crystal structure of the 5abc region of the group I intron (Cate *et al.*, 1996) the secondary structure of the A-rich bulge is different from that observed in solution, mostly because of an extensive network of tertiary interactions around that region.

2.1.4 Internal loops

Internal loops occur when the corresponding nucleotides in both strands of a double stranded region do not constitute a Watson-Crick combination (Fig. 2). Thus, the smallest internal loop is a base-base mismatch. The secondary structure schemes are usually based on the Watson-Crick base pairing, which gives a misleading impression that internal loops are simply big floppy "bubbles" flanked by helical stems. Structural studies have, however, shown that internal loops are often highly structured and are actively involved in different types of base pairing and stacking. Usually, short internal loops adopt conformations relatively close to that of the RNA double helix, while long internal loops may not follow the behavior of short ones.

Among the known internal loops one can distinguish at least three groups of structures. The first group includes single base pair mismatches, mostly purine-pyrimidine or pyrimidine-pyrimidine. Depending on the nucleotide context, these mismatches can be stacked within the helix in conformations close to that of the Watson-Crick base pairs. The most studied example is the G-U base pair, originally seen in the structure of the yeast $tRNA^{Phe}$ (Robertus *et al.*, 1974; Kim *et al.*, 1974). The second group includes the GA/AG tandem mismatches. The so-called sheared pair G-A, being introduced into the A-RNA helix, will over-wind it. However, an A-G pair adjacent to the

first G-A will have an opposite effect, under-winding the helix and thus restoring the A-type helix conformation. There are many examples of RNA structures with this type of internal loop, including oligo-$(GGCGAGCC)_2$ (SantaLucia & Turner, 1993) and oligo-$(GGGCUGAAGCCU)_2$ (Heus *et al.*, 1997). An interesting yet distinct conformation has been described for the symmetrical internal loop GAAA, where the sheared G-A and A-G base pairs are separated by two reverse Hoogsteen type A-A base pairs (Baeyens *et al,* 1996). The winding and unwinding of the helix in the presence of the A-A base pairs is so strong that the major groove of the helix almost disappears while the minor groove becomes extremely wide and almost "flat". The sheared A-A base pair has an overall geometry close to that of the sheared G-A pair, although kept only by one H-bond. It is, therefore, possible for the tandem A-A mismatches to have a structure close to that found in the tandem of the sheared G-A base pairs. At least one example of such a structure is found in the internal loop J4/5 of the group I intron (Cate *et al.*, 1996).

An internal loop structure similar to that found in the E-loop of the eukaryotic 5S ribosomal RNA constitutes the third group (sometimes called E family; Shen *et al.* 1995). All internal loops in this group share a common sequence motif 5'-ANUA-3'/5'-AAG-3' which can exist alone or be a part of a larger internal loop. The latter case occurs in the E-loop itself, which consists of five nucleotides in one strand and four in the other (UAGUA/UAAG; Wimberly *et al.*, 1993). The structure of the racin/sarcin loop from the 28S rRNA shown in Fig. 6 includes a sheared G-A base pair followed by the Hoogsteen A-U base pair and by a Hoogsteen-like A-A base pair with one nucleotide excluded from stacking interactions (Szewczak *et al.*, 1993). However, crystal structures of the same internal loops have a conformation somewhat different from that observed in the NMR structures. The nucleotide excluded from stacking in the NMR structure becomes a part of the stack in the crystal structure (Correll *et al.*, 1997).

Involvement of internal loops in different inter or intra-molecular interactions can affect their conformation. The AA-platform motif found in the J6a/6b internal loop of the group I intron participates in the interaction with the tetraloop L5b (Cate *et al.*, 1996). However, in the solution structure, the same internal loop, taken separately, does not form the same motif. (Butcher *et al.*, 1997b). In the RRE RNA the stacking pattern and the

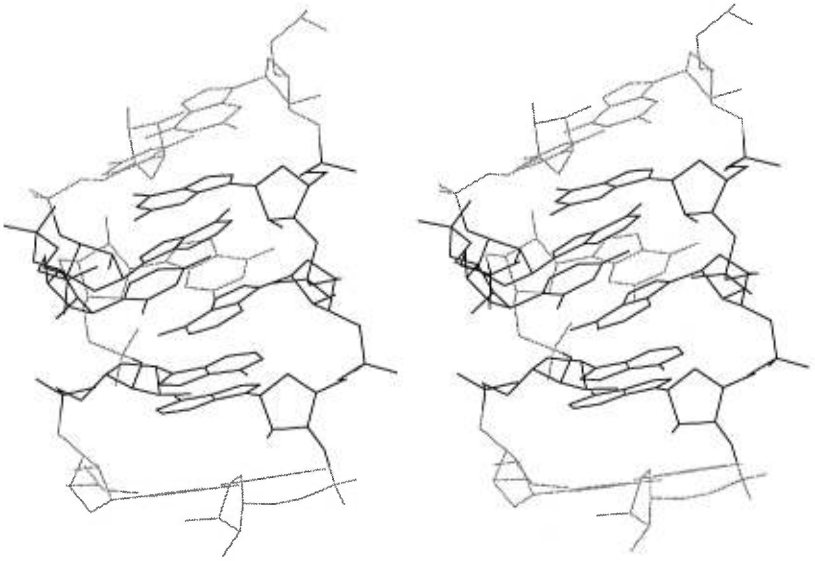


Figure 6. Stereo-drawing of the structure of the racin/sarcin loop from the 28S rRNA (Szewczak *et al.*, 1993). The nucleotides important for the E-family motif are shown in black.

nucleotide interactions are different for the free and protein-bound states (Battiste *et al.*, 1996). Among RNA aptamers one can find more interesting examples of internal loops with different lengths. Structures of their complexes with corresponding ligands displayed sometimes quite peculiar conformations (Dieckmann *et al.*, 1996; Fan *et al.*, 1996; Jiang *et al.*, 1996). In most cases the ligands are "buried" inside the internal loop structure and participate in various types of interactions that stabilize the loop.

2.1.5 Junctions

Junctions are the places in the RNA structure where three or more double stranded regions adjoin (Fig. 2). There can be none or several unpaired nucleotides between each paired region in a junction. These nucleotides usually participate in different tertiary interactions, which will be discussed later.

In general, bifurcation loops play an important role in the structures of large RNAs, providing necessary links between relatively rigid double helical domains. Unlike in DNA, most RNA junctions contain unpaired nucleotides (Altona, 1996). It is generally assumed that helices in the tight (with no nucleotides between helices) junctions having an even number of branches are mutually coaxial. In the 3-way or 5-way tight junctions, the helices expect to form quasicontinuous stacked structures, in which two of the helices stack together to the third one (so called Y-shape). However, no structures of this type have been observed yet.

So far, structures of only three types of RNA junctions have been solved at atomic resolution. A three-way junction was observed in the hammerhead ribozyme (Pley *et al.*, 1994; Scott *et al.*, 1995) and in the 5abc region of the group I intron (Cate *et al.*, 1996). Four-way junctions exist in several known tRNA structures (Robertus *et al.*, 1974; Kim *et al.*, 1974; Moras *et al.*, 1980; Rould *et al.*, 1989; Basavappa & Sigler, 1991). The structure of the $tRNA^{Ser}$ from *T. thermophylus* can be considered as a five-way junction (Biou *et al.*, 1994).

In both the ribozyme and group I intron structures two of three helices of the 3-way junction are coaxial. Also, in both cases, non-Watson-Crick base pairs adjust the stacking between two coaxial stems, while the third helix leans toward one of them. In the hammerhead ribozyme, the sharp turn needed for the third helix to get its position is provided by the U-turn-like conformation in the longer connector, which makes the major grooves of the two helices facing each other. It has been found that in the group I intron, helix P5c faces helix P5a by its minor groove using a special "purine-pinch" motif (Steinberg, unpublished).

The tRNA four-way junction consists of two pairs of coaxial helices whose perpendicular arrangement resembles letter "L". Nucleotides of the bifurcation loop are involved in different types of tertiary interactions, which further stabilize the structure. In the $tRNA^{Ser}$ the existence of one more helix attached perpendicularly to one of the helical domains makes this structure a 5-way junction.

2.2 Tertiary structure

Tertiary interactions are usually referred to as contacts observed in the three-dimensional structure between elements of the secondary structure. Thus, tertiary interactions occur via contacts involving two helices, two unpaired regions or one unpaired region and a double stranded helix. Tentatively, one can distinguish Watson-Crick interactions occurring between single stranded regions from other interactions, since they will result in the formation of a normal double helix. For the purpose of clarity, structures with predominantly Watson-Crick interactions are considered first, while all other nucleotide-nucleotide interactions will follow.

2.2.1 Watson-Crick tertiary base pairing

Tertiary base pairing in a Watson-Crick manner between single stranded regions can result in several different types of structure. The structure of so-called "kissing loops", in which two hairpin loops interact with each other thus forming a somewhat distorted double helix, is one example. In fact, the interaction between a loop (hairpin, internal or bulge) and another single stranded region adjacent to a helix in the same polynucleotide chain produces knots or pseudoknots (Studnicka *et al.*, 1978). Although one can theoretically propose many different types of such base pairing, only few have been observed so far (Kang *et al.*, 1996; Du *et al.*, 1996; Ferré-D'Amaré *et al.*, 1998).

A six member "kissing loop" of HIV TAR element with sequence UCCCAG interacts with its complementary sequence, making a double helix which is quasi-coaxial to the stems of the hairpins (Chang & Tinoco, 1997). The extension of the loop for one nucleotide, as it occurs in the structure of the inverted sequence of ColE1 (Lee & Crothers, 1998), increases the bend between the helices and makes the conformation of the Watson-Crick pairs and the stacking in the loop less distorted than in TAR.

The most studied type of pseudoknot is that formed by a hairpin loop with a single stranded region adjacent to the hairpin stem. The two double helices of this pseudoknot are coaxial, with connector regions of several nucleotides crossing the major groove of one helix and the minor groove of the other (Fig. 7). Such structures were observed in pseudoknots from MMTV (Kang *et al.*, 1996), gene 32 mRNA of bacteriophage T6 (Du *et al.*, 1996), TYMV (Kolk *et al.*, 1998) and the aptamer inhibiting the HIV reverse transcriptase (Jaeger *et al.*, 1998). A more complicated system of base pairing was observed in the case of the HDV ribozyme (Ferré-D'Amaré *et al.*, 1998) where a polynucleotide chain forms a nested double pseudoknot with five double helical segments.

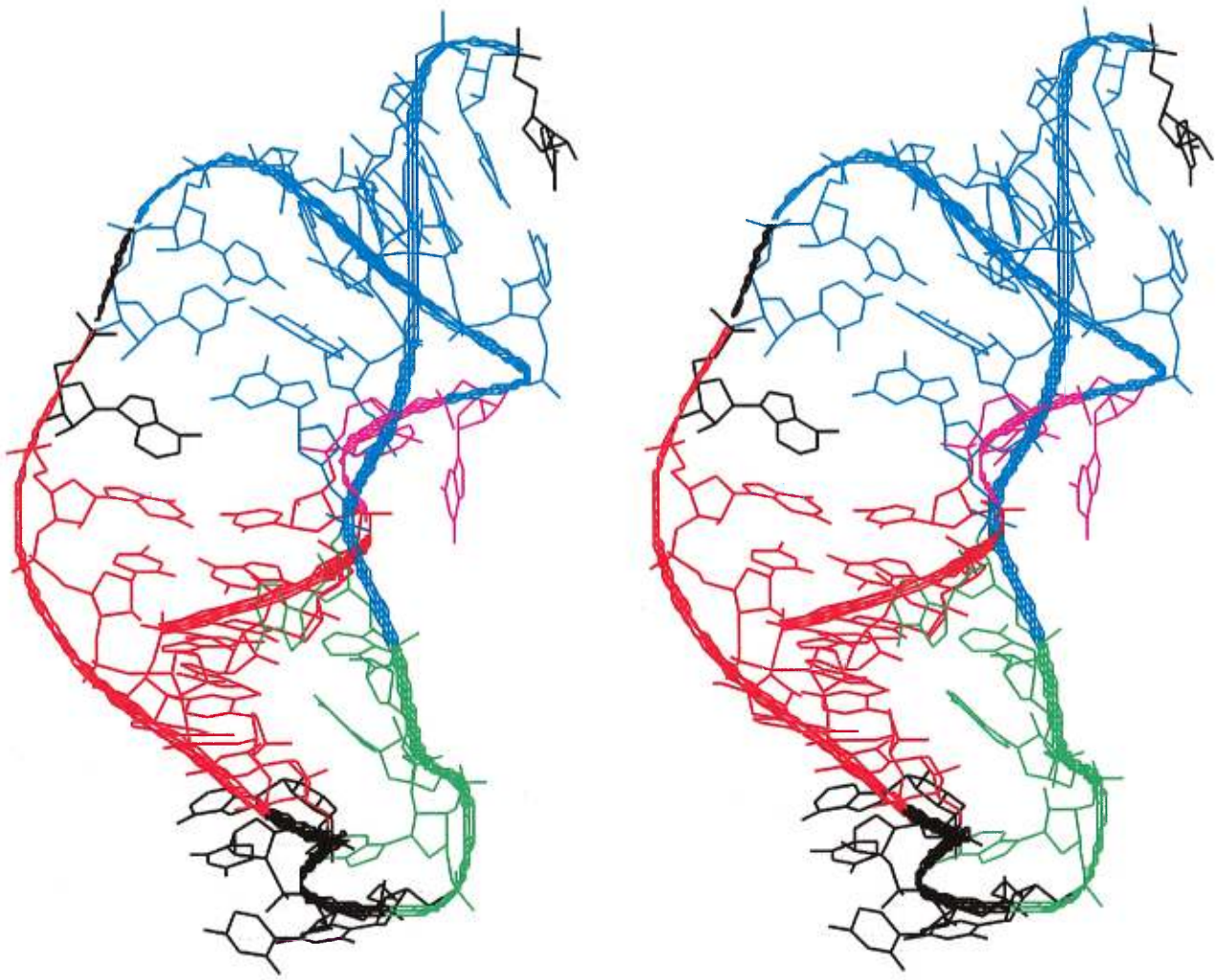


Figure 7. Stereo-drawing of the RNA pseudoknot from MMTV (Shen & Tinoco, 1995). The two helical domains are shown in red and blue, while the connector regions are colored in green and magenta.

2.2.2 Tertiary base pairs, triples and quadruples

The potentials for base pairing in RNA are not limited to the Watson-Crick type. At least 28 different schemes of base pairing can be suggested for the uncharged tautomeric forms of nucleotides (Saenger, 1984). The variety of possible base pairs increases even more if one takes into account a possibility for ionization and participation of water molecules in the formation of hydrogen bonds. However, not all theoretically possible combinations have been observed in either crystal or NMR structures so far.

Tertiary base pairing can occur between any two single-stranded elements of the secondary structure. The first known examples of such interactions were hairpin loop - bifurcation loop (reverse Watson-Crick base pair G15-C48) and hairpin loop - hairpin loop (G18 and Ψ55 of the D and T-loop, respectively) interactions in the $tRNA^{Phe}$ (Robertus *et al.*, 1974; Kim *et al.*, 1974).

Most of the single-stranded regions, however, are characterized by their own quasi-independent structure with internal base-base interactions. Thus, tertiary base pairs often become part of base triples. Such a tertiary base triple between a hairpin loop and an internal loop can be seen in the structure of the group I intron, where adenine of tetraloop L5b interacts with the Hoogsteen A-U base pair of the "tetraloop receptor" in the internal loop J6a/6b (Cate *et al.*, 1996). Triples can also involve Watson-Crick base pairs within double helices, like the base triple A9-U12-A23 in the $tRNA^{Phe}$ or can even become a regular structure, making a triple helix (Broitman *et al.*, 1987).

Structures with a base quadruple have not been observed in biologically relevant molecules yet. However, the possibility of their existence has been suggested based on the solution structure of the oligo-(UGGGGU) (Cheong & Moore, 1992). This oligonucleotide forms a tetraplex of four parallel strands with four stacked layers of guanines uniformly interacting with their neighbors in a non-Watson-Crick manner.

2.2.3 Intercalation, stacking and base – backbone interactions

Tertiary interactions are not limited to base-base hydrogen bonds. A wide range of different interactions between riboses, backbone and bases can also be found in the RNA tertiary structure. Most of these interactions are not sequence specific and there are no known general similarities among them. This, however, may reflect the fact that only a handful of RNA structures are known at atomic resolution. Another problem is that interactions between the biomolecule and the molecules of a solvent cannot be determined by NMR methods while the resolution of many crystal structures is not high enough to see them.

Base-base stacking interactions include the interaction between neighboring nucleotides in the sequence as well as intercalation. Intercalation is an insertion of a nucleotide into the stack between two neighboring nucleotides belonging to another region of the molecule. In the tRNA structure intercalation occurs at two different places (Robertus *et al.*, 1974; Kim *et al.*, 1974). The first is an insertion of nucleotide 18 of the D-loop between nucleotides 57 and 58 of the T-loop. The second is the intercalation of nucleotide 21 from the D-loop into the nucleotide stack of junction 46-48. Interdomain stacking can be also seen in the tRNA at the place of the D/T-loop contact, where the tertiary base pair 15-48 from the D-domain stacks to nucleotide 59 of the T-loop. A somewhat similar situation occurs in the core of the $tRNA^{Ser}$ between nucleotide 20b of the D-loop and the helix of the variable arm (Biou *et al.*, 1994). Interestingly, in both cases the stacking interaction occurs between the helices and the nucleotides not involved in stacking interactions within the regions adjacent to them. Thus, these interactions contribute to the perpendicular orientation of the helical domains. However, an example of another orientation of helical domains is also known. The stacking interaction between the tetraloop and the tetraloop receptor in the structure of the group I intron provides a quasicontinuous stacking between the corresponding domains (Cate *et al.*, 1996). Such a structure becomes possible due to a special conformation of the internal loop J6a/J6b known as the AA-platform. This conformation consists of a special A-A dinucleotide, which presents one of its adenines for stacking with another adenine from tetraloop L5b.

Hydrogen bonds between bases and phosphates have been observed in many structures. The U-turn motif includes a hydrogen bond between H1 of uridine and the oxygen from the phosphate group of the second nucleotide after the uridine. Another interesting example was seen in the conformation of the T-domain of the tRNA, in which C61 forms a hydrogen bond with the phosphate group of nucleotide 59. This hydrogen bond plays an important role in maintenance of the T-loop conformation (Romby *et al.*, 1987).

Hydrogen bonds involving the O2'-hydroxyl groups of the riboses are common for RNA. Although these interactions occur in many different structures, there are only few motifs that specifically include them. For example, a "ribose zipper" found in the structure of the group I intron is formed by the riboses of two stacked regions in the minor groove (Cate *et al.*, 1996). This interaction is characterized by hydrogen bonding between the 2'-hydroxyl and pyrimidine O2 (or purine N3) of one base the 2'-hydroxyl of its partner. It is difficult, however, to generalize based solely on this structure, because a crystal structure of two RNA helices packed via their minor grooves displayed quite a different H-bonding pattern (Schindelin *et al.*, 1995).

3. Biogenesis and structure of tRNA

3.1 Function and lifecycle of tRNA

Transfer RNA plays a central role in the process of transformation of genetic information in the cell, being an *adapter* molecule in translating mRNA nucleotide sequence into the protein sequence of amino acids. In addition, tRNA has been found to play many other roles. The aminoacylated tRNA can be a donor of an amino acid not only in the ribosome-dependent protein synthesis, but also in the biosynthesis of aminoacyl-phosphatidylglycerol and glycyl-lipopolysaccharides (Littauer & Inouye, 1973), as well as in the transfer of terminal aminoacids to some proteins (Leibowitz & Soffer, 1969). tRNAs can be also involved in the transcriptional regulation of messenger RNAs for

enzymes associated with aminoacid synthesis (Henkin, 1994) and in the synthetic pathway of porphyrin derivatives (Shön *et al.*, 1986). Uncharged tRNA can serve as a primer for the reverse transcriptase in some retroviruses (Harada *et al.*, 1979) and as a transcription factor (TFIIIR) for the Pol III RNA polymerase (Dunstan *et al.*, 1994). Although for some transfer RNAs such a functional diversityhas been reported, it should be noted that in some cases tRNAs whose primary function is other then delivery of the amino acids to the ribosome do not participate in the translation at all. The tRNA "nature" of these molecules is recognized mainly on the basis of the conventional tRNA secondary structure.

tRNA biosynthesis proceeds differently in prokaryotes and eukaryotes; still, the resulting molecules are very much alike. Although some steps of tRNA transcript processing, including removal of extra 5' and 3' sequences, excision of introns and/or addition of the CCA terminus, have been studied for both eukaryotic and prokaryotic systems, much still remains to be elucidated (Deutscher, 1995). Even less is known about eukaryotic tRNAs that are transcribed as separate genes by the Pol III RNA polymerase. Another step of the tRNA maturation is nucleotide modification and RNA editing. The latter is known only for few cases (Beier *et al.*, 1992) and consists of replacing of one or several standard nucleotides in the RNA sequence by other, mostly unusual nucleotides. Nucleotide modifications occur in all known tRNAs, and predominantly touch anticodon and T-loops. The raison d'être of many modifications remains a mystery, although for some of the reasonable suggestions has been made. For example, modifications of nucleotide 37 are generally thought to affect the tRNA-mRNA interactions on the ribosome. The formation of the particular water-mediated interactions between the backbone and pseudouridine are suggested to be important for the tRNA structure (Arnez & Steitz, 1994). Formation of the N2,N2-dimethylguanine as well as 1-methyladenine have been shown to prevent alternative folding of the tRNA secondary structure (Helm *et al.*, 1998; Steinberg & Cedergren, 1995) by restricting the H-bond formation capabilities of the bases.

Although there are several different modes of tRNA recognition by the cognate aminoacyl-tRNA synthetase, all of them are thought to occur via interactions with so-

called identity elements in the tRNA sequence and/or structure. At least one universal element, the acceptor terminus 5'-CCA-3', is necessary for aminoacylation of all tRNAs. In the simplest case of the eubacterial $tRNA^{Ala}$ the major identity element for aminoacylation is the G3-U70 base pair in the acceptor stem. In other cases, tRNAs have several identity elements, located in the different parts of the molecule including the anticodon loop, anticodon and D-stems. For the tRNAs with a long extra arm the orientation and some particular nucleotides of the extra arm can also serve as identity elements (Achsel & Gross, 1993; Breitschopf *et al.*, 1995).

3.2 Sequence and secondary structure

Since the determination of the nucleotide sequence of yeast $tRNA^{Ala}$ (Holley *et al.*, 1965), about 3000 different tRNAs and tRNA genes from various organisms have been sequenced (Chapter I). Despite similarities observed in most of them, there are quite a few sequences that do not fit the general primary and secondary structure pattern. Roughly, all tRNAs can be divided into two groups depending on their origin. The first group includes cytosolic tRNAs which are characterized by a conserved sequence pattern. All organelle tRNAs and also tRNAs from some viruses and symbiotic bacteria belong to the second group. Sequence patterns observed in the first group are either distorted or simply absent in the second one. For the sake of clarity, this chapter deals exclusively with cytosolic tRNAs, their sequences and structures, while the next chapter will describe some rules of the structural organization of organelle tRNAs, mostly mitochondrial.

With only few exceptions, all sequences of cytosolic tRNAs and tRNA genes share three general features. The first and probably the most fascinating feature is that all tRNA sequences can be folded into the cloverleaf secondary structure (Holley *et al.*, 1968; Fig. 8). Second, within this secondary structure all tRNAs have the same feature, which can be summarized as follows (parenthesis contain alternative names and abbreviations):

a) seven base pairs in the amino acid acceptor stem (AA stem);

b) three or four base pairs in the dihydrouracil stem (D-stem);

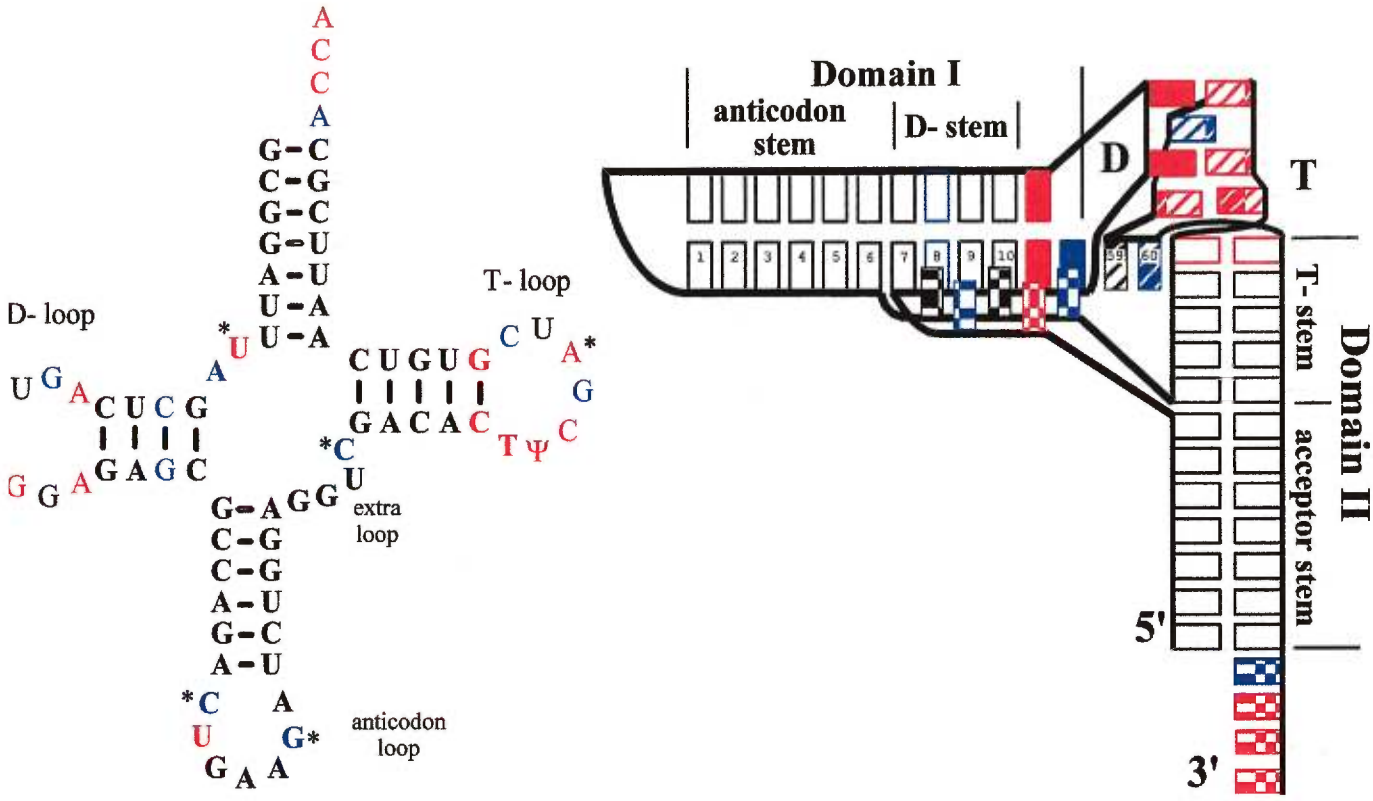


Figure 8. The cloverleaf secondary structure of the yeast $tRNA^{Phe}$ (left) and the standard tRNA L-form (right). The invariant and semiinvariant nucleotides are shown in red and blue, respectively. Asterisks on the cloverleaf structure denote positions in which nucleotide modifications occur very frequently. The L-form: open rectangles represent base paired nucleotides; filled and crosshatched rectangles stand for nucleotides of the D and T-loops, respectively. Checkered rectangles represent the unpaired nucleotides between helical domains and at the amino acid terminus. The small figures 1 to 12 refer to the layers of stacked nucleotides starting from the base pair closest to the anticodon loop. Numbers 59 and 60 refer to the T-loop nucleotides in the standard tRNA nomenclature (see Chapter I). Nucleotide 59 stacks to the last, twelfth layer of Domain I. The unstacked nucleotides in the D-loop are not shown.

c) five base pairs in the anticodon stem (AC stem);

d) five base pairs in the T-Ψ-C stem (T-stem);

e) two nucleotides between the acceptor and D-stems (Connector 1);

f) no nucleotides between the acceptor and T-stems;

g) one nucleotide between the D and anticodon stems;

h) seven nucleotides in the anticodon and T-loops;

i) from 7 to 10 nucleotides in the D-loop;

j) from 4 to 21 nucleotides in the variable arm (extra arm, extra loop, V loop, V arm, E arm; Connector 2);

Exceptions occur mainly in the selenocysteine tRNAs. The last general feature is that certain positions in the cloverleaf representation of the tRNA secondary structure are always occupied by invariant (conserved) or semi-invariant (semi-conserved) nucleotides (Fig. 8), except for the initiator tRNAs and few other special cases (Sprinzl *et al,* 1998). Sometimes a high number of modified nucleotides in the tRNA sequence is also considered as a general feature of tRNAs (Kim, 1978), although the type and number of these modifications depends on the particular tRNA species (Crain & McCloskey, 1997).

Based on the number of base pairs in the D-stem and the number of nucleotides in the extra loop, all cytosolic tRNAs with only few exceptions can be divided into two classes. Class I includes those tRNAs whose sequences in the cloverleaf type secondary structure have either three or four base pairs in the D-stem and either four or five nucleotides in the extra loop ($D_{3\text{-}4}V_{4\text{-}5}$). tRNAs with three base pairs in the D-stem and a long extra arm that has a stem-loop structure belong to Class II (D_3V_n).

3.3 General tRNA architecture

The structure of the polynucleotide chain in the Class I tRNAs is characterized by a so-called "L" shape (Robertus *et al.*, 1974; Kim *et al.*, 1974). Double helical regions suggested by the secondary structure are maintained in the three-dimensional structure. Two pairs of stems in the cloverleaf structure form coaxial double helical "arms", where

Domains I and II are arranged at approximately 90° to each other. Domain I consists of the anticodon and D-stems, while the acceptor and T-stems make up Domain II. The "corner" of the molecule is formed by interactions between the D and T-loops. The two major functional centers, the anticodon loop and the amino acceptor terminus, are positioned at the opposite ends of the "L", being at approximately 75 Å one from the other. If all tRNA molecules maintain the same shape, both helical domains should be conservative in their length (Steinberg *et al.*, 1997). This means that each of Domains I and II should always have twelve layers of stacked nucleotides. Both Connector I and the extra loop (Connector II) interact with the major groove of the D-domain, forming tertiary contacts with the D-stem and the D-loop, except for nucleotide 44, which stacks to the anticodon stem and forms a non-Watson-Crick base pair with nucleotide 26.

The detailed tertiary structure of the yeast $tRNA^{Phe}$ (Kim *et al.*, 1974) is shown on Fig 9. Interestingly, most of invariant or semiinvariant nucleotides are involved in tertiary interactions within the molecule. Although this structure contains the system of tertiary interactions that can be found in all cytosolic Class I tRNAs, some variations occur. One such variation is the *E.coli* $tRNA^{Gln}$ complexed with its aminoacyl tRNA-synthetase (Rould *et al.*, 1989). Nucleotide 46 is excluded from the base triple with pair 13-22 and from the nucleotide stacking, while its place in the triple is occupied by nucleotide 45. A possible reason for this rearrangement is that nucleotide 46, being a uridine, is unable to form the same type of interactions as G46 or protonated A46 in the $tRNA^{Phe}$ and in the $tRNA^{Asp}$, respectively.

Class II tRNAs differ from Class I tRNAs in some essential aspects of the tertiary structure. As seen in the crystal structure of $tRNA^{Ser}$ from *T. thermophilus* (Biou *et al.*, 1994) the accommodation of the extra arm to the rest of the molecule requires a special nucleotide arrangement between the extra arm and the major groove of the D-stem. Three tertiary nucleotides A21, C48 and G20b form a "shed", in which A21 and C48 play the role of "walls" supporting a "roof" (G20b) stacked to the extra arm. The stacking structure of this region is supported by nucleotide 9, which forms a base triple with pair 13-22 but not with 12-24 as in Class I tRNA structure. In other aspects, including D/T loop interactions, the Class II tRNAs are expected to be similar to the Class I tRNAs.

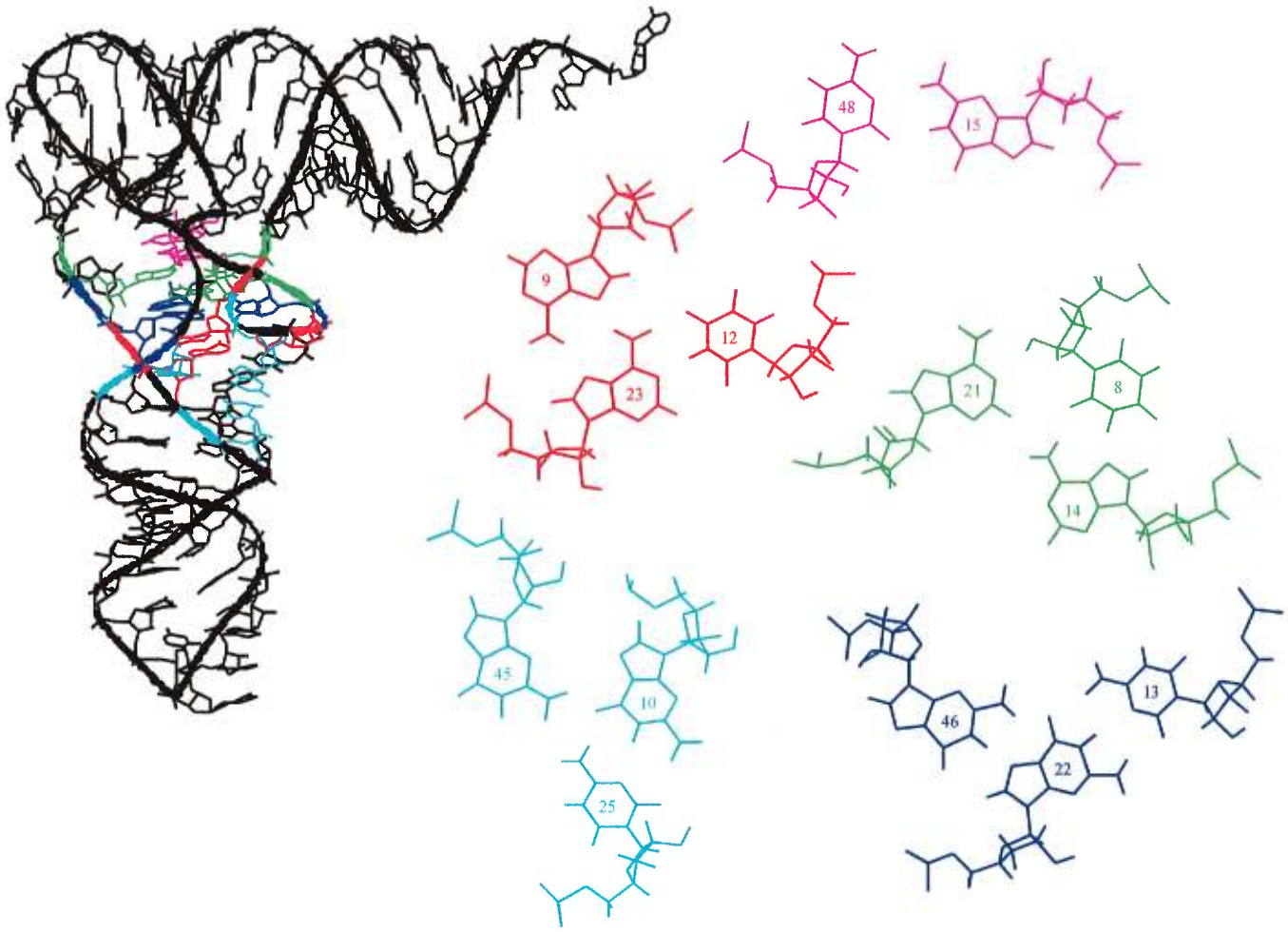


Figure 9. The tertiary structure of the yeast $tRNA^{Phe}$. The tertiary interactions in the D-domain are shown on the right and are marked with the corresponding color in the three-dimensional structure on the left. The nucleotide numeration corresponds to that in the standard tRNA nomenclature (see Chapter I).

3.4 Mitochondrial tRNAs within the standard L-shape

The sequences of many mitochondrial tRNAs are perplexing in a sense that they display both primary and presumed secondary structural patterns deviating significantly from those defined by the cloverleaf structure and by the presence of several highly conserved nucleotides in the cytosolic tRNAs (Wolstenholme, 1987). However, coexistence of these tRNAs together with the normal ones in the same mitochondria points to the possibility that despite these differences, their L-shaped spatial structures remain very similar to the standard (Steinberg & Cedergren, 1994). Moreover, there are experimental data suggesting that in some organisms normal cytosolic tRNAs are exported into the mitochondria, replacing those tRNAs that are missing in a mitochondrial genome, thus again indicating that three-dimensional structures of cytosolic and organelle tRNAs should be similar (Dietrich *et al.,* 1992).

Abnormalities in the mitochondrial tRNAs can occur in almost any part of the cloverleaf secondary structure. Many of these changes can be described by the "double-zipper" covariation (Steinberg & Cederegren, 1994) or fit into more general L-shape compensatory rules (Steinberg *et al.*, 1997). The "double-zipper" covariation is a correlation between the lengths of the anticodon stem and of both connectors in Class I mitochondrial tRNAs. Usually, for N base pairs in the anticodon stem, the minimal length for Connector 1 is 8-N nucleotides and for Connector 2 it is 9-N nucleotides. In terms of the tRNA L-shape, this means that the anticodon stem, in the presence of a shortened Domain I, can be extended at the expense of the connector regions. On the level of the tertiary structure this covariation represents a way to maintain the normal length of Domain I and the normal D/T loop interactions necessary for the formation of the L-shape conformation. In other words, whatever secondary structure the tRNA molecule has, Domain I and the T-stem within the L-shape tertiary structure *must* have twelve stacked layers and five stacked base pairs, respectively. However, if the T-stem has only four base pairs, the extension of Domain I for one more stacking layer can compensate this deficiency (Steinberg *et al.*, 1997). The extension can be achieved by introduction of additional stacking layer formed by nucleotides from connector regions. One should note

that only a part of the cloverleaf variations found in sequences of mitochondrial tRNAs results from changes in Domain I. What happens with a tRNA structure lacking the normal T-loop sequence or shortening of the D-loop remains unclear.

4. Problems addressed by the author

Analysis of the RNA architecture requires extensive comparative sequence analysis.

The multiple sequence alignment is routinely used for analysis of the sequences in homologous molecules. However, the optimal sequence alignment for full-length molecules may not correspond to the optimal alignment between their structural domains, and thus will provide wrong information concerning the similarity between the regions. The development of the databases that incorporate elements of the structural information is an important step toward resolution of this problem. For the protein structure analysis, databases usually group protein sequences either by common tertiary folding motifs or by the function similarity. Sufficient sequence variations within the same type of protein structural motifs and a relatively large number of available structures make this approach fruitful. In the structural analysis of RNA, on the other hand, the small number of known three-dimensional structures of similar molecules hampers creation of databases based on tertiary motifs. Thus, such databases are usually built of molecules with a similar function that are aligned by their secondary structure. The latter is usually predicted with help of available algorithms or comes from experiments on representative molecules. This kind of databases includes those of the ribosomal RNAs (Maidak *et al.*, 1997), of the RNA part of the RNAse P (Brown, 1998) etc. and can be used for analysis and prediction of their tertiary structure (Michel & Westhof, 1990). They can also be useful for verification and correction of the sequences and of the proposed secondary structures.

Relationship between deviations in secondary structure and in tertiary fold within the same RNA architecture

In terms of structure-function paradigm, every structural element essential for the function should be preserved in all molecules performing this function. It is not clear, however, to which extent the structural elements should be conserved in order to gurantee the function.

For tRNA, the ability of sequences of cytosolic tRNAs to fit into the cloverleaf pattern of secondary structure and to form the L-shaped spatial structure is thought to be the most general criteria determining its functionality in translation. There are however, examples even among cytosolic tRNAs, when the secondary structure deviates significantly from the standard. In particular, the selenocysteine tRNAs have unusually long D- and acceptor stems. For some organelle tRNAs the cloverleaf secondary structure pattern can hardly be recognized. If the cloverleaf structure is important for the tRNA function, there should be rules describing how to cope with the situations when it is not maintained any more. Studies on mitochondrial tRNAs revealed that many of the most unusual tRNAs still have the ability to maintain the L-shape structure as in the normal cytosolic tRNAs if the standard D/T loop interactions are also maintained (Steinberg *et al.*, 1997).

Here, evidence is presented that despite the obvious deviation in the secondary structure of eukaryotic and archaeal selenocysteine tRNAs from the standard cloverleaf pattern of cytosolic tRNAs, these tRNAs are able to adopt a conformation satisfying the general constraints on the tRNA L-form.

Fitting the abnormal eukaryotic $tRNA^{Sec}$ to the normal tRNA architecture

Elucidation of tertiary interactions even within the framework of the established secondary structure is still a difficult task. For the cytosolic Class I tRNAs, this problem can be simplified by consideration of the isosteric replacements in the interactions observed in the known crystal structures. However, as the example of the $tRNA^{Gln}$ (Rould

et al., 1989) and our analysis of the role of nucleotide 47 (Chapter V) show, steric factors play a significant role. In the Class II tRNAs the introduction of an additional helical domain makes the situation even more complicated. When the crystal structure of the $tRNA^{Ser}$ was published (Biou *et al.*, 1994), it became clear that in the model of this tRNA (Dock-Bregeon *et al.*, 1989) only one tertiary interaction had been predicted correctly. Detailed analysis of the Class II tRNA sequences indicated that the system of tertiary interactions found in the X-ray structure of the $tRNA^{Ser}$ is only one of several possible (Chapter IV).

The fitting of the eukaryotic $tRNA^{Sec}$ into the 7/5 secondary structure has helped understand the role played by the D/T loop interactions in the arrangement of the helical domains. The established secondary structure of the eukaryotic $tRNA^{Sec}$, in turn, has allowed to propose a special arrangement of the nucleotides of Connector 1 that dock the long extra arm to the major groove of the D-stem. This arrangement has revealed unexpected similarities to the corresponding arrangements in the eubacterial and eukaryotic $tRNAs^{Ser}$.

Can bulged nucleotides not involved directly in either secondary or tertiary interactions affect RNA architecture?

The role of unstructured elements in the RNA structure is usually underestimated. Nucleotides not involved in either secondary or tertiary interactions are often considered dispensable. In tRNA, nucleotide 47 in Connector 2 is the most notorious example of such unstructured elements. This nucleotide is not involved in any interactions in those known tRNA structures where it is present. In the crystal structure of the $tRNA^{Asp}$, on the other hand, this nucleotide does not exist; still, all tertiary base-base contacts appear to be the same as in the $tRNA^{Phe}$. Interestingly, when the first crystal structure of a tRNA at atomic resolution had been obtained, the initial division of the tRNA cloverleaf secondary structures into three classes (Class I – D_4V_5, Class II – D_3V_n, Class III – D_3V_4) was replaced by a binary classification, in which Class I encapsulated Class III (Levitt, 1969; Quigley & Rich, 1976). The presence of nucleotide 47, which made the difference

between these two classes, was thus considered unimportant for the tRNA structure and function.

Based on the molecular modeling study, the reputation of nucleotide 47 as an important aspect of the formation of the canonical system of tertiary interactions in the tRNA has been restored. The absence of this nucleotide leads to steric collision between nucleotide 22 and 46, which in turn, will cause a disruption of either 22-46 or 15-48 base pair. Here we show that both of these pairs can be preserved via introduction of a non-canonical U13-G22 pair in the D-stem. The case of nucleotide 47 demonstrates that even nucleotides not involved directly in any interactions can strongly affect the general architecture of the molecule.

Chapter I

The Compilation of tRNA sequences and sequences of tRNA genes

Mathias Sprinzl, Carsten Horn, Melissa Brown, Anatoli Ioudovitch and Sergey Steinberg[1]

Nucleic Acids Res., 1998, **26**, 148-153.

[1] Contributions of each author:
Mathias Sprinzl – completeness of the database, general supervision
Carsten Horn, Melissa Brown – data entry
Anatoli Ioudovitch – correction of mistakes in sequences and alignments
Sergey Steinbreg – quality control

48–153 *Nucleic Acids Research, 1998, Vol. 26, No. 1*

Compilation of tRNA sequences and sequences of RNA genes

Mathias Sprinzl*, Carsten Horn, Melissa Brown, Anatoli Ioudovitch[1] and Sergey Steinberg[1]

Laboratorium für Biochemie, Universität Bayreuth, 95440 Bayreuth, Germany and [1]Université de Montréal, Faculté de Médicine, Départment de Biochimie, C.P. 6128, Succursale Centre-Ville, Montréal, Québec H3C 3J7, Canada

Received September 16, 1997; Revised and Accepted October 6, 1997

ABSTRACT

Sequences of 3279 sequences of tRNA genes and tRNAs published up to December 1996 are included in the compilation. Alignment of the sequences, which is most compatible with the tRNA phylogeny and known three-dimensional structures of tRNA, is used. Sequences and references are available under http://www.uni-bayreuth.de/departments/biochemie/trna/

INTRODUCTION

The 1997 compilation contains 3279 sequences of tRNAs and tRNA genes. The last edition which appeared two years ago (1) was supplemented by 579 new sequences covering the literature up to December 1995. The sequences of tRNA mutants and of tRNAs originating from transformed or differentiated cells were not considered.

The tRNAs included in the compilation are listed in Table 1. Each tRNA or tRNA gene is specified by the (abbreviated) name of the organism from which it was isolated and a four digit code: the first three digits identify the organism, the last digit specifies the particular isoacceptor. The amino acid specificity of the tRNA is indicated by a one-letter amino acid code. The tRNAs coding for selenocysteine were annotated with the letter Z. Initiator tRNAs are annotated with the letter X.

The references are restricted to the first publication of the complete sequence unless additional information (e.g., base modification, corrections, etc.) was later obtained. In such cases additional references were added.

In order to facilitate a computer analysis an alignment is used which is most compatible with the tRNA phylogeny and known three-dimensional structures of tRNA. The corresponding numbering system is shown in Figure 1.

As was the case in the previous edition (1), this publication does not contain a sequence printout. Instead, the sequences, references and footnotes of tRNAs and tRNA genes listed in Table 1 are deposited in the European Bioinformatics Institute (EBI) Data Library. In addition, a World Wide Web page has been established and is available under http://www.uni-bayreuth.de/departments/biochemie/trna/ . The present publication should be quoted as a reference for the electronically accessible data.

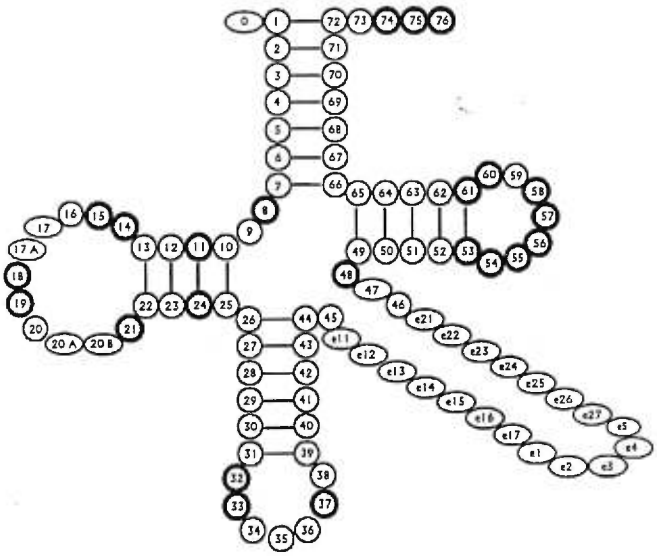


Figure 1. Numbering of nucleotides in tRNAs. Circles represent nucleotides which are always present; the ovals, nucleotides which are not present in each structure: these are nucleotides before the position 1 on the 5′-end, before and after the two invariant GMP residues 18 and 19 in the D-loop, and the nucleotides in the variable loop. The nucleotide to be added at a given site is indicated by the number of the preceeding nucleotide followed by a colon and a letter in alphabetical order. The nucleotides in the variable stem have the prefix 'e' and are located between position 45 and 46 obeying the base-pairing rules. The nucleotides in the 5′-strand and the 3′-strand are numbered by e11, e12, e13, ... and e21, e22, e23, ..., respectively; the second digit identifies the base-pair. In the case of a long variable region, the loop can be formed by up to 5 nt: e1, e2, e3, e4 and e5. Positions, in which invariant nucleotides usually occur are indicated by a thick line.

Researchers who wish to perform an advanced search for tRNA sequences according to several criteria, e.g., anticodon, amino acid specificity, modified nucleoside, or wish to print the requested sequence in the form of Table 2 or cloverleaf format (Fig. 1) can obtain appropriate software on diskette. Please contact M. Sprinzl, Laboratorium für Biochemie, Universität Bayreuth, D-95440 Bayreuth, Germany, Fax: +49 921 552432, Email: Mathias.Sprinzl@uni-bayreuth.de.

*To whom correspondence should be addressed. Tel: +49 921 552 420; Fax: +49 921 552 432; Email: mathias.sprinzl@uni-bayreuth.de

Table 1. List of tRNA sequences and sequences of tRNA genes included in the compilation

PART ONE: Sequences of tRNA genes

Source	Code	tRNA genes
VIRUSES	000-029	
MYCOBACTERIOPH. L5	020	NQW
PHAGE PHI C31	031	
PHAGE T4	022	GILPQRST
PHAGE T5	026	ACDEFGHIKLMNPQSSTVWXY
ARCHAEBACTERIA	030-109	
ARCHAEGLOBUS FULG.	034	A
HALOBACTERIUM CUT.	038	AC
HALOBACTERIUM HAL.	042	A
HALOBACTERIUM MAR.	044	LS
HALOBACTERIUM MED.	046	W
HALOFERAX VOLCANII	050	CW
METHANOBAC.FORMI.	058	A
METHANOBAC.THERM.	062	A
METHANOCOCCUS JAN.	065	AACDEFGGHIILLLKMMNQPRRRSSSTTVVVWXY
METHANOCOC.VANI.	066	ADEFHIKLNPQRTTVY
METHANOTHRIX SOEH.	067	A
METHANOTHERM. FER.	068	ADEHIKLMNPST
RUMINOBACTER AMYLO	070	E
METHANOCOC.VOLTAE	074	DKPTY
METHANOPYRUS KAND.	076	KLQS
METHANOSPIR. HUNG.	078	A
SULFOLOBUS SOLFA.	086	FGLSVX
THERMOPLASMA ACID.	090	M
THERMOCOCCUS CELER.	094	APT
THERMOFIL. PENDENS	096	GM
THERMOPROT. TENAX	098	AALLX
EUBACTERIA	110-239	
BARTONELLA BACIL.	110	IX
BARTONELLA ELIZAB.	111	AI
BARTONELLA HENSELA.	112	I
BARTONELLA QUINT.	113	AI
MYCOPLASMA CAPRIC.	114	ACDEFGHIIKKLLMNPQRRSSTTVWWXY
MYCOPLASMA GEN.	115	ACDEFGGHIIKKLLMMNPQRRSSSSTTTWWY
MYCOPLASMA MYCOID.	118	ADEFGIMNPRRSTVX
MYCOPLASMA PNEU.	120	ACDEEGGHIIKKLLMNPQRRRSSSSTTTVWWXYY
MYCOPLASMA PG50	122	KL
ACHOLEPLASMA LAID.	123	AACDEFGHIIKKLLLMMNQRSSTVW
SPIROPLASMA CITRI.	125	SWW
SPIROPLASMA MELIF.	126	ACDFIMPRSX
BORRELIA BURGDORF.	128	AI
STREPTOMYCES GRIS.	130	S
STREPTOMYCES COEL.	131	L
STREPTOMYCES RIM.	134	EQQXX
STREPTOMYCES LIV.	135	CDEEEGGKNNQQRSVVY
STREPTOMYCES AMBO.	136	P
CHLOSTRIDIUM PERFR.	139	S
MYCOBACT. TUBERC.	140	PV
KLEBSIELLA AEROGE.	141	N
AGROBACTER. TUME.	142	R
CLOSTRIDIUM THERM.	143	Z
DESULFOMICR. BACU.	144	Z
CLOSTRIDIUM ACETO.	145	T
PLESIOMONAS SHIGE.	146	E
ENTEROCOCCUS HIRAE	147	A
STAPHYLOCOC. AURE.	148	ACDDFGGGGHIKLLLMPQRSSTTVVWXY
LACTOBAC. BULG.	150	DEGNPRSV
LACTOBAC.DELBRUEC.	152	S
LACTOCOCCUS LACTIS	153	AAAEFGINSX
BACILLUS SUBTILIS	154	AAAACDEFFGGGHHIIIKKLLLLLLMMNNPQ RSSSTTTVWXXY
BACILLUS CIRCULANS	156	P
BACILLUS SP. PS3	157	DENSV
THERMUS THERMOPHI.	158	GGTTY
THERMOTOGA MARIT.	159	MMTWY
RHODOTHERMUS MAR.	160	AI
THIOBACILLUS FERRO	162	AI
STIGMATELLA AURANT.	163	GTTY
E.COLI	166	AACDEFGGGHIIKLLLLLMNPPPQQRRRRR SSSSTTTTTVVVWXXYYZ
SALMONELLA TYPHI.	170	HLPRR
AZOSPIRILLUM LIPO.	172	KTV
TRICHODESMIUM SPEC	173	AI
PHOTOBACT. PHOSPH.	174	HP
PHOTOBAC. LEIOGNA.	175	LM
AEROMONAS HYDROPH.	178	AEHILPR
PSEUDOMONAS AER.	182	AGITTY
PSEUDOMONAS FLUOR.	184	AI
CAMPYLOBAC.JEJUNI	186	AI
RICKETTSIA PROW.	187	GWY
CAULOBACTER CRES.	189	AI
BRUCELLA SUIS	190	AI
BRUCELLA MELLITENS.	191	AI
BRUCELLA ABORTUS	192	AAII
AZORHIZOBIUM CAUL.	193	G
RHIZOBIUM MELILOTI	194	L
AZOARCUS SP.BH72	195	IL
OCHROBACTRUM ANTH.	196	AI
BORDETELLA PERTUS.	198	L
HAEMOPHILUS INFLU.	200	AAAACDDDEFGGGHIIIKKKKLLLLLMMM NNPQQRRRRSSSSTTVVWY
ANACYSTIS NIDULANS	210	AI
SYNECHOCYSTIS SP.	214	AACDFGGGIIIHKKLLLLLNPPP QRRRRSSSSTTTVVWWXXY
SYNECHOCOCCUS SP.	215	L
CYANOPHORA PARAD.	218	AEGHILRS
PYLAIELLA LITTORA.	222	AI
STREPTOCOCCUS PN.	224	A
STREPTOCOCCUS SAL.	225	A
ORGANELLES		
CHLOROPLASTS	240-359	
CYANOPHORA PARAD.	240	AI
PYLAIELLA LITTORA.	241	AI
CHLAMYDOMONAS REIN	244	ACDEGIMRRTW
CHLAMYDOMO. MOEWU.	246	T
CHLORELLA ELLIPSO.	248	AIRS
LYCOPERSICON ESCU.	249	DLY
CUCUMIS SATIVUS	250	E
ASTASIA LONGA	251	ACDGIKLMPQRSSTV
EUGLENA GRACILIS	252	AACDEFGGHIIKLLLMNPQRRSSTVWXXY
CRYPTOMONAS SPEC.	254	AIR
SPIROGYRA MAXIMA	255	I
ANTITHAMNION SP.	257	AI
CYANIDIUM CALDAR.	258	AIK
OLISTHODISCUS LUT.	259	AI
MARCHANTIA POLYM.	260	ACDEFGGHIIKLLLMNPPQRRRSSSTTVVWXY
CUSCUTA REFLEXA	261	AHILMV
COLEOCHAETE ORBIC.	262	AI
HORDEUM VULGARE	264	GGMSTVX
TRITICUM AESTIVUM	268	CDEGGMPRSTWXY
ORYZA SATIVA	270	ACCDEFGGHIILLLMMNPQRRSSSTTVVWY
ZEA MAYS	272	AACDEFGGHHIIKLLLLMN PQRRSSSSSTTTVVVWXXY
EPIFAGUS VIRGINIA.	274	LNR
ARABIDOPSIS THAL.	276	IMP
ALLIUM PORRUM	278	R
BRASSICA OLERACEA	280	L
GLYCINE MAX	284	AIMV
MEDICAGO SATIVA	288	H
NICOTIANA TABACUM	292	ACDEFGGHIIIKLLLMNPQRRSSSTTVVWXY
NICOTIANA DEBNEYI	296	H
OENOTHERA SP.	300	PW
DAUCUS CAROTA	301	V
GOSSYPIUM HIRSUTUM	302	H
PELARGONIUM ZONALE	304	R
PENNISETUM AMERICA	308	I
PETUNIA HYBRIDA	312	H
PHASEOLUS VULGARIS	316	H
HELIANTHUS ANNUUS	317	HNY
PISUM SATIVUM	320	DEGHKLNPRRSTVWXY
PINUS THUNBERGII	322	ACDEFGGHIIKLLLMNPPQRRRSSSTTVVWXY
PINUS CONTORTA	323	HK
SINAPIS ALBA	324	HKQSV
SPINACIA OLERACEA	328	ACDEIIIILMRSSTTVY
SPIRODELA OLIGORH.	332	NRR
VICIA FABA	336	EFHLLTY
SORGHUM BICOLOR	340	L
MITOCHONDRIA	360-599	
SINGLE CELL ORGANISMS AND FUNGI	360-419	
PROTOTHECA WICKER.	360	ACDEFGGHIIKLLMNPQRRSSTVWXY
PYLAIELLA LITTOR.	361	KPY
CHONDRUS CRISPUS	362	ACEGGHIKLLMNPQRRSVWXY
PLATYMONAS SUBCORD.	363	KNPVY
CHLAMYDOMO. REINH.	364	MQW
ODONTELLA SINENSIS	365	AACDEFGGHIIIKLLNPPQRRSSTVWXY
PLASMODIUM FALCIP.	366	CDEGGHKLPQSSWXY
TRYPANOSOMA BRUCEI	368	AA
LEPTOMONAS COLLO.	369	H

Table 1. *continued*

PARAMECIUM PRIM.	372	XY
PARAMECIUM TETRA.	376	WY
PARAMECIUM AURELIA	377	FWY
TETRAHYMENA PYRIF.	380	EFHLWX
TETRAHYMENA THERM.	384	LXY
ASPERGILLUS FUMI.	387	EMMTV
ASPERGILLUS NIDUL.	388	ACCDEFGGHIKLLMMNPQRSSTVWXY
NEUROSPORA CRASSA	392	ACMR
PODOSPORA ANSERINA	396	DMNSVW
PODOSPORA CURVICOL	397	N
SACCHAROMYCES CER.	400	AACDEFGHIKLMNPQRRSSTTVVWXYY
SACCHAROMYCES EXI.	401	MP
PICHIA PIJPERI	402	LMM
WILLIOPSIS MRAKII	403	KLMPQSV
SCHIZOSACCHA.POM.	404	GHLPQ
KLUYVEROMYCES LAC.	405	CKLQ
CANDIDA PARAPSILO.	406	CEFGHIKLNPRRTVWY
HANSENULA WINGEI	407	ACCDEFGHIKLLLMMMMNPQRRRSSTVVWY
TORULOPSIS GLAB.	408	ACDEFGHIKLMNPQRSSTTVWXY
WILLIOPSIS SUAVE	409	M
PICHIA JADINII	410	M
TRICHOPHYTON MENT.	409	AFLMMTV
TRICHOPHYTON RUBR.	412	DGIKMQRWY
PENICILLIUM CHRYS.	413	NRY
ASCOBOLUS IMMERSUS	415	NNQ
PLANTS	420-459	
ARABIDOPSIS THAL.	424	EMQSSY
GLYCINE MAX	428	EMX
SOLANUM LYCOPERS.	430	C
SOLANUM TUBEROSUM	431	X
LUPINUS LUTEUS	432	GINX
BRASSICA NAPUS	434	K
OENOTHERA SP.	436	CFGHILNPSSSWXY
PHASEOLUS VULGARIS	440	NSY
HELIANTHUS ANNUUS	441	CEGHIKMNPQVWX
TRITICUM AESTIVUM	444	CDEFKNPQQSSSWXYYY
ORYZA SATIVA	446	FHNPRSW
ZEA MAYS	448	CDEHKMMPSSWXY
MARCHANTIA POLYM.	450	ACDEFGGHIKLLLMMNPQRRRSSTVWY
LARIX	452	HH
ANIMALS	460-599	
FASCIOLA HEPATICA	462	ADIKNPSW
ASCARIS SUUM	464	ACDEFGHIKLLNPQRSSTVWXY
CAENORHABDI.ELEG.	468	ACDEFGHIKLLNPQRSSTVWXY
MYTILUS EDULIS	470	ACDEFGHIKLLMMNPQRSSTVWY
ARTEMIA SP.	472	EFS
LOCUSTA MIGRATORIA	476	ACDDEFGGHIKKLLLLNPPQRSSTTVWXY
PSEUDOREGMA BAMBU.	477	L
METRIDIUM SENILE	478	X
NEPHILA CLAVIPES	479	AAAA
AEDES ALBOPICTUS	480	AEFGLNRSV
LOLIGO BLEEKERI	481	KKKKK
APIS MELLIFERA	482	ACDDEFGHIKLLMPQRSSVWY
DAPHNIA PULEX	483	IQVWXY
DROSOPHILA MELANO.	484	ACCDDEFGGHIKKLLLQRSSTVWWXYY
DROSOPHILA YAKUBA	488	ACDEFGHIKLLNPQRSSTVWXY
DROSOPHILA VIRILIS	496	IQX
CHORISTONEURA FUM.	497	L
PISASTER OCHRACEUS	498	ACDEGLLNPQTVWXY
PROTOPTERUS DOLLOI	499	ACDEFGHIKLLMNPQRSTVWY
ASTERINA PECTINI.	500	ACDGHLLMNPQSSVWY
CERATITIS CAPITATA	501	AEFNRS
ASTERIAS FORBESII	502	ACDGLLNVWXY
CYPRINUS CARPIO	503	ACDEFGHIKLLNPQRSSTVWXY
PARACENTROTUS LIV.	504	ACDEFFGHIKLLNPQRSSTVWXY
ANOPHELES QUADRIM.	505	ACDEFGHIKLLNPQRSSTVWXY
RAINBOW TROUT	506	FPT
ANAS PLATYRHYNCOS	507	ACDEFKLNSWY
STRONGYLOCEN.PURP.	508	ACDEFGHIKLLNPQRSSTVWXY
ACIPENSER TRANSM.	509	PT
GADUS MORHUA	510	ACDEFGHIKLLMNQRSSTVWY
ACANTHAMOEBA CAST.	511	ADEIKLMPQX
XENOPUS LAEVIS	512	ACDEFFGHIKLLNPQRSSTVWXY
ALLIGATOR MISSIS.	513	ACNWY
CROCODYLUS POROSUS	514	ACNWY
CARETTA CARETTA	515	ACNWY
RANA CATESBEIANA	516	ACFILNPQTWXY
MALACLEMYS TERRA.	517	ACNWY
SPHENODON PUNCTAT.	518	ACNY
EPICRATES SUBFLA.	519	ACEN
CEPHALORHYN.COM.	520	FPT
CROSSOSTOMA LACUS.	521	ACDEFGHIKLLNPQRSSTVWXY
CHICKEN	522	ACDEFGHIKLLMNPQRSSTVWY
DIDELPHIS VIRGINI.	523	DPT
ODOCOILEUS HEMIO.	524	PT
DICEROS BICORNIS	525	FP
MARMOSA SP.	526	DPT
PHILANDER SP.	527	D
RAT	528	ACCDDEFGHIKKLLNNNPPQQRSSSTTVWWXXY
METACHIRUS SP.	529	D
PHALANGER SP.	530	D
CNEDIMOPHORUS UNI.	531	EPT
MOUSE	532	ACDEFGHIKLLNPQRSSTVWXY
CERVUS NIPPON	533	PT
BALAENOPTERA PHYS.	534	ACDEFGHIKLLPQRTVWXY
BALAENOPTERA MUSC.	535	ACDEFGHIKLLNPQRSTVWXY
BOVINE	536	ACDEFGHIKLLNPQRSSTVWXY
HALICHOERUS GRYPUS	537	ACDEFGLNPQRRSTVWXY
PHOCA VITULINA	538	ACDEFGHIKLLNPQRSTVWXY
GADUS MORHUA	539	DEIPQST
LEPIDOSIREN PARAD.	542	V
RHINOCEROS UNICORN	544	ACDEFGHIKLLMNPQRTVWY
SCELOPORUS OCCID.	545	HILLMMVW
STRUTHIO CAMELUS	550	HILMRW
ERINACEUS EUROP.	555	ACDEFGHIKLMNPQRTVY
MACACA ASSAMENSES	556	HLS
MACACA NIGRA	557	HL
MACACA SILENUS	558	HL
MACACA THIBETANA	559	HL
GREEN MONKEY	560	F
SIAMANG	561	ACDEIKNWXY
MACACA FUSCATA	562	HLS
MACACA MULATTA	563	HLS
MACACA FASCICULA.	564	HLS
MACACA SYLVANUS	565	HLS
SAIMIRI SCIUREUS	566	HLS
PAPIO HAMADRYAS	567	HL
TARSIUS SYRICHTA	568	HLS
LEMUR CATTA	570	HLS
CHIMPANZEE	572	ACDEEFGHIKLLMNPQRSTVWXY
PYGMY CHIMPANZEE	573	ACDEIKNWXY
GIBBON	576	HLS
GORILLA	580	ACDEEGHIIKLMNPQRSTVWXY
ORANG UTAN	584	ACDEEFGHIKLLLMNPRSSTVWXY
HUMAN	588	AACCDEEFGHIKLLLMMNPQRRSSTVWX
AEPYCEROS MELAMPUS	590	FV
BOSELAPHUS TRAGOC.	591	FV
CEPHALOPHUS MAXW.	592	FV
DAMALISCUS DORCAS	593	FV
GAZELLA THOMSONI	594	FV
KOBUS ELLIPSIPRYM.	595	FV
MADOQUA KIRKI	596	FV
ORYX GAZELLA	597	FV
TRAGELAPHUS IMBER.	598	FV
EUKARYOTIC CYTOPLASM	600-999	
SINGLE CELL ORGANISMS AND FUNGI	600-669	
PLASMODIUM FALSI.	603	AILMNRRTV
TRYPANOSOMA BRUCEI	605	KKKNNQQRRRTVVY
TETRAHYMENA PYRIF.	606	NQS
LEISHMANIA TARENT.	609	GIKLQRRTVW
DICTYOSTELIUM DIS.	616	AEEHKKLMNQRRSSSTTVVWWY
PHYSARUM POLYCEPH.	618	X
NEUROSPORA CRASSA	620	FLR
CANDIDA ALBICANS	621	LSS
PHYTOPHTHORA PAR.	622	D
PODOSPORA ANSERINA	624	SS
SACCHAROMYCES CER.	628	AAACDDEEEFFFFGGHIIKKKLLLMMNP QQQRRRSSSSSTTTVVVWWWXXYY
SCHIZOSACCHA.POM	632	ADEEFHIKRRSSSVXX
CANDIDA CYLINDRA.	637	S
PLANTS	670-749	
CHLAMYDIA TRACHOM.	672	TW
ARABIDOPSIS THAL.	674	AFSSSSSSVWWXYYYY
GLYCINE MAX	690	DMX
PHASEOLUS VULGARIS	698	LPP
NICOTIANA RUSTICA	706	SSSSSSSYY
PETUNIA SP.	710	N
HELIANTHUS ANNUUS	712	L
SORGHUM BICOLOR	714	G
ORYZA SATIVA	718	G
TRITICUM AESTIVUM	720	YYYYY
TRITICUM VULGARE	724	S
SOYBEAN	730	C
ANIMALS	750-999	
CAENORHABDI. ELEG.	756	AAADEEGGHIKKKLLLLLLNPPP QQRRRRRSSTTVVVWXYZ
BOMBYX MORI	768	AAEGK
DROSOPHILA MELANO.	774	ADEEEFGGHIKKLLMNPRRSSTVVXYZ
DROSOPHILA SIMUL.	780	S
SQUID	785	K
XENOPUS LAEVIS	792	AFKLNVXXYYYZ
PODOCORYNE CARNEA	793	CFGSS
CHICKEN	804	AADDFKPPWZ
MOUSE	810	ACCDEGHIKKLPPXZ

Table 1. *continued*

RAT	916	DDEEEEEEEEEFGGKLLLPPQQQQQQ
BOVINE	928	SZ
HUMAN	999	AEEGGGKKLLMNNPPQQQRR SSSSSTTTVVVVVVVXXYY

PART TWO: tRNA Sequences

Source	Code	tRNA
VIRUSES	000-029	
AVIAN ONCO.-VIRUS	010	M
CHICKEN ASV/AMV/RS	014	W
MOUSE M-MULV	018	PP
PHAGE T4	022	GILPQRST
PHAGE T5	026	DLNPQ
ARCHAEBACTERIA	030-109	
HALOBACTERIUM CUT.	038	AGHNQRSTVVVX
HALOFERAX VOLCANII	050	AAACDEEFGGGGHIIKKLLLLLMNPPP QRRRSSSTTVVWXY
HALOCOCCUS MORRHUA	054	X
METHANOBAC. THERM.	062	GN
SULFOLOBUS ACIDO.	082	X
THERMOPLASMA ACIDO	090	MX
EUBACTERIA	110-239	
MYCOPLASMA CAPRIC.	114	ACDEFGHIIKKLLLMNPQRRSSTTVWWXY
MYCOPLASMA MYCOID.	118	AGIPSTVX
SPIROPLASMA CITRI	125	WW
STREPTOMYCES GRIS.	130	X
STREPTOMYCES COEL.	131	G
STAPHYLOCOC. EPID.	138	GG
MYCOBAC. SMEG.	142	X
BACILLUS STEARO.	146	FLVY
BACILLUS SUBTILIS	154	AFGIKKLMPRSSSTVWXYY
THERMUS THERMOPHI.	158	DFIMXX
E.COLI	166	AAACDEEEFGGGHIIIKLLLMNQQ RRRRRSSSSSTTVVVWXXYYZ
SALMONELLA TYPHI.	170	GGHLPPP
AZOSPIRILLUM LIPO.	172	N
RHODOSPIRIL. RUB.	202	FL
AGMENELLUM QUADR.	206	F
ANACYSTIS NIDULANS	210	LLX
SYNECHOCYSTIS SP.	214	E
ORGANELLES		
CHLOROPLASTS	240-359	
CHLAMYDOMONAS REIN	244	E
EUGLENA GRACILIS	252	F
CODIUM FRAGILE	253	GKMR
SCENEDESMUS OBLIQ.	256	MXY
LUPINUS ALBUS	263	Y
HORDEUM VULGARE	264	DDEQ
TRITICUM AESTIVUM	268	E
ZEA MAYS	272	I
GLYCINE MAX	284	LLL
NICOTIANA TABACUM	292	W
PHASEOLUS VULGARIS	316	FLLLWX
SPINACIA OLERACEA	328	FIILMPTVWX
MITOCHONDRIA	360-599	
SINGLE CELL ORGANISMS AND FUNGI	360-419	
TETRAHYMENA PYRIF.	380	FY
TETRAHYMENA THERM.	384	W
NEUROSPORA CRASSA	392	ALLTVWXY
SACCHAROMYCES CER.	400	FGHIKLMPRRSSSTWXY
PLANTS	420-459	
SOLANUM TUBEROSUM	431	IIL
OENOTHERA SP.	436	F
PHASEOLUS VULGARIS	440	FLLLLMPWXY
ANIMALS	460-599	
ASCARIS SUUM	464	FMS
AEDES ALBOPICTUS	480	DEGIKQRSVX
LOLIGO BLEEKERI	481	KKK
HAMSTER	524	DKRS
RAT LIVER	528	DDFKLLLRVVW
BOVINE LIVER	536	EGIKLLRSSSTVWXX
HUMAN	588	S
MARSUPIAL	599	D
EUKARYOTIC CYTOPLASM	600-999	
SINGLE CELL ORGANISMS AND FUNGI	600-669	
EUGLENA GRACILIS	604	DF
TETRAHYMENA THERM.	608	QQQX
SCENEDESMUS OBLIQ.	612	FXY
NEUROSPORA CRASSA	620	FX
SACCHAROMYCES CER.	628	ACDEFFGGHHIIKKLLLMNPPRRRSSSTTVVVWXY
SCHIZOSACCHA. POM.	632	EFXY
TORULOPSIS UTILIS	636	AILPVXY
CANDIDA CYLINDRA.	637	LLLSSSSS
PLANTS	670-749	
HORDEUM VULGARE	678	EEF
WHEAT GERM	682	FGKMRWXYY
BRASSICA NAPUS	686	F
LUPINUS LUTEUS	694	EFGHIMNPSVXY
PHASEOLUS VULGARIS	698	LLLLX
PISUM SATIVUM	702	F
SPINACIA OLERACEA	704	S
NICOTIANA RUSTICA	706	SSSSSYY
SOLANUM TUBEROSUM	707	LW
CUCUMIS SATIVUS	708	L
ANIMALS	750-999	
CAENORHABDL ELEG.	756	L
ASTERINA AMURENSIS	762	X
BOMBYX MORI	768	AAFFGGI
DROSOPHILA MELANO.	774	EFHKKSSSVVVXY
EUPHAUSIA SPERBA	786	X
XENOPUS LAEVIS	792	DFX
SALMON LIVER	798	X
CHICKEN	804	W
MOUSE	810	EFFFIKKMQQRRVXZ
RAT	916	DDEKKKLLNNQSSSVVX
RABBIT LIVER	922	DFKKKMV
BOVINE LIVER	928	DFFLLNQRRRTWYZ
CALF LIVER	934	F
COW MAMMARY GLAND	940	LL
SHEEP LIVER	946	HX
HUMAN	999	AAEFGGHLMNNQQSVXYYZ

PART THREE: tRNA and tRNA gene sequences that differ from the conventional alignment

Source	Code	tRNA/tRNA gene
ARCHAEBACTERIA	030-109	
METHANOCOCCUS JAN.	065	Z
MITOCHONDRIA	360-599	
SINGLE CELL ORGANISMS AND FUNGI	360-419	
PHYTOMONAS SP.	367	Q
TRICHOPHYTON MENT.	409	E
ANIMALS	460-599	
LOCUSTA MIGRATORIA	476	S
APIS MELLIFERA	482	T
DAPHNIA PULEX	483	C
DROSOPHILA MELANO.	484	P
PROTOPTERUS DOLLOI	499	S
BALAENOPTERA PHYS.	534	NSS
BALAENOPTERA MUSC.	535	S
HALICHOERUS GRYPUS	537	K
PHOCA VITULINA	538	S
SIAMANG	542	S
RHINOCEROS UNICORN	544	SS
SCELOPORUS OCCID.	545	ACS
STRUTHIO CAMELUS	550	AST
ERINACEUS EUROP.	555	SS
MACACA THIBETANA	559	S
PAPIO HAMADRYAS	567	S
CHIMPANZEE	572	S
PYGMY CHIMPANZEE	573	S
GORILLA	580	S
ORANG UTAN	584	S
HUMAN	588	N

Table 2. Format of tRNA sequences in the databank

PART ONE: SEQUENCES OF tRNA-GENES

```
                                        accept  D-domain          anticodon domain   variable region          T-domain          accep
                                        stem                                         extra loop                                 stem
Number                                  012345678911111111111222222222222333333333344444eeeeeeeeeeeeeeeeeee4444555555555566666666667777777
      Anticodon                               01234567890001234567890123456789012345111111112345222222226789012345678901234567890123456
      Organism          Kingdom
DA0260 TGC PHAGE T5          VIRUS  -GGGCGAATAGTGTCAGC-GGG--AGCACACCAGACTTGCAATCTGGTA----------------G-GGAGGGTTCGAGTCCCTCTTTGTCCACCA
+                                   ==*=*==  ===*           *=========      ======                  =====      ========*=*==
DA0340 TGC ARCHAEGLOBUS FULG. ARCHAE -GGGCTCGTAGCTCAGC--GGG--AGAGCGCCGCCTTTGCGAGGCGGAG----------------GCCGCGGGTTCAAATCCCGCCGAGTCCA---
+                                   ==*====  ====           ====  =====      =====                   =====      ==========*==
DA0380 TGC HALOBACTERIUM CUT. ARCHAE -GGGCCCATAGCTCAGT--GGT--AGAGTGCCTCCTTTGCAAGGAGGAT----------------GCCCTGGGTTCGAATCCCAGTGGGTCCA---
+                                   ==*====  *===           ====* =====      =====                   =====      ==========*==
DA0420 TGC HALOBACTERIUM HAL. ARCHAE -GGGCCCATAGCTCAGT--GGT--AGAGTGCCTCCTTTGCAAGGAGGAT----------------GCCCTGGGTTGGAATCCCAGTGGGTCCA---
+                                   ==*====  *===           ====* =====      =====                   =====      ==========*==
DA0580 TGC METHANOBAC.FORMI.  ARCHAE -GGGCCCGTAGCTCAGACTGGG--AGAGCGCCGCCCTTGCAAGGCGGAG----------------GCCCCGGGTTCAAATCCCGGTGGGTCCA---
+                                   ==*====*  ====           ====  =====      =====                   =====      =====*===*==
DA0620 TGC METHANOBAC.THERM.  ARCHAE -GGGCCCGTAGCTCAGACTGGG--AGAGCGCCGCCCTTGCAAGGCGGAG----------------GCCCCGGGTTCAAATCCCGGTGGGTCCA---
+                                   ==*====*  ====           ====  =====      =====                   =====      =====*===*==
DA0650 TGC METHANOCOCCUS JAN. ARCHAE -GGGCCCGTAGCTCAGCT-GGG--AGAGCGCCGGCCTTGCAAGCCGGAG----------------GCCGTGGGTTCAAATCCCACCGGGTCCA---
+                                   ==*====  ====           ====  =====      =====                   =====      ==========*==
DA0651 GGC METHANOCOCCUS JAN. ARCHAE -GGGCTGGTAGCTCAGACTGGG--AGAGCGCCGCATTGGCTGTGCGGAG----------------GCCGCGGGTTCAAATCCCGCCCAGTCCA---
+                                   ==*====  ====           ====  =====      =====                   =====      ==========*==
DA0660 TGC METHANOCOC.VANI.   ARCHAE -GGGCCCGTAGCTCAGTT-GGG--AGAGCGCTGCCCTTGCAAGGCAGAG----------------GCCGTGGGTTCAAATCCCGCCGGGTCCA---
+                                   ==*====  ====           ====  =====      =====                   =*===      ===*======*==
DA0670 TGC METHANOTHRIX SOEH. ARCHAE -GGGCTTGTAGCTCAGCT-GGT--AGAGCGCCGCCTTTGCAAGGCGGAG----------------GCCCTGGGTCCGAATCCCAGCAAGTCCA---
+                                   ==*====  ====           ====  =====      =====                   =====      ==========*==
DA0680 TGC METHANOTHERM. FER. ARCHAE -GGGCCCATAGCTCAGCCTGGG--AGAGCGCCGCCCTTGCAAGGCGGAG----------------GCCCCGGGTTCAAATCCCGGTGGGTCCA---
+                                   ==*====  ====           ====  =====      =====                   =====      ==========*==
DA0780 TGC METHANOSPIR. HUNG. ARCHAE -GGGCTCGTAGCTCAGCT-GGA--AGAGCGCGGCGTTTGCAACGCCGAG----------------GCCTGGGGTTCAAATCCCCACGGGTCCA---
+                                   ==*=*==  ====           ====  =====      =====                   =====      ========*=*==
DA0940 TGC THERMOCOCCUS CELER ARCHAE -GGGCCGGTAGCTCAGCCTGGG--AGAGCGTCGGCTTTGCAAGCCGAAG----------------GCCCCGGGTTCGAATCCCGGCCGGTCCACCA
+                                   ==*====  ====           ====  =====      =====                   =====      ==========*==
DA0980 TGC THERMOPROT. TENAX  ARCHAE -GGGCCGGTAGTCTAGC--GGA--AGGACGCCCGCCTTGCGCGCGGGAG----------------ATCCCGGGTTCGAATCCCGGCCGGTCCA---
+                                   ==*====  ===*           *=== =====      =====                   =====      ==========*==
DA0981 CGC THERMOPROT. TENAX  ARCHAE -GGGCCGGTAGTCTAGC--GGA--AGGACGCCCGCCTCGCGCGCGGGAG----------------ATCCCGGGTTCGAATCCCGGCCGGTCCA---
+                                   ==*====  ===*           *=== =====      =====                   =====      ==========*==
DA1110 TGC BARTONELLA ELIZAB. EUBACT -GGGGCCGTAGCTCAGCT-GGG--AGAGCACCTGCTTTGCAAGCAGGGG----------------GTCGTCGGTTCGATCCCGTCCGGCTCCACCA
+                                   ==*====  ====           ====  =====      =====                   =  ===      === =====*==
DA1130 TGC BARTONELLA QUINT.  EUBACT -GGGGCCGTAGCTCAGCT-GGG--AGAGCACCTGCTTTGCAAGCAGGGG----------------GTCGTCGGTTCGATCCCGTCCGGCTCCACCA
+                                   ==*====  ====           ====  =====      =====                   =  ===      === =====*==
```

(Continued in the databank. See text for instructions.)

RESULTS

Presentation of sequences

The sequences in the database are divided into three parts. The first two parts contain the sequences of the tRNA genes and tRNAs, respectively, which can be fitted into the canonical tRNA alignment. The third part lists tRNA and tRNA gene sequences, mainly of animal mitochondria, whose secondary structures differ from most tRNAs and could not be aligned according to Figure 1.

An example for sequence presentation in the database is given in Table 2. Each sequence in the compilation occupies two consecutive lines. The first line begins with the letter 'D' or 'R' and contains the six-position identification code of the sequence ('D' or 'R' for DNA or RNA, respectively; a one-letter code for the amino acid, X for methionine-initiator, Z for selenocysteine; and the four-digit code specifying the organism and isoacceptor. After this, the sequence of the anticodon (in the case of tRNA sequences in its modified form) is given, followed by the name and the kingdom of organism (Table 1), and the sequence (99 standard positions). The second line begins with the sign '+' and contains the information about base-pairing (double helical regions only, tertiary interactions are not annotated). All other lines in the compilation begin with signs other than 'D,' 'R' or '+' (usually '*') and contain comments.

Nucleotides involved in Watson–Crick pairs are marked with '=', the GU pairs are indicated with the sign '*'. Nucleotides 26 and 44 are considered to form a base-pair included in the anticodon stem (Fig. 1).

The sequences in orginal publications denoted as 'yeast' are assigned to *Saccharomyces cerevisiae*. The user should be aware, however, that some of these organisms have possibly been misclassified and that the original literature should be consulted.

This compilation uses a one-letter code for all nucleotides including modified ones. For standard nucleotides, adenosine, cytidine, guanosine, thymidine and uridine the usual abbreviations, A, C, G, T and U, respectively, are used. To designate modified nucleotides, the other ASCII signs are employed as defined in Table 3. Terminology and structure of the modified nucleosides occurring in tRNAs were used according to refs 2 and 3. Positions in particular sequence which are not filled (gaps in the generalised structure, Fig. 1) are indicated by a dash. All nucleotide insertions are denoted by underlining at the place of insertion.

Numbering and alignment of the variable region

The alignment of the variable region has been done in accordance with Steinberg and Kisselev (4). The extra arm is placed between nucleotides 45 and 46. It includes two double helical strands forming a stem and a loop. The annotations of the nucleotides in the extra arm positions begin with the letter 'e' (extra) followed by a one- or two-digit number. We have reserved a space for 7 bp in the stem and 5 nt in the loop. The nucleotides in the loop are numbered from 1 to 5, whereas the nucleotides in the stem are numbered from 11 to 17 (5′-branch) and from 27 to 21, in the reverse order, (3′-branch), to indicate base-pair formation between nucleotides 11–21, 12–22, etc. (Fig. 1). In the tRNAs where the extra arm position 45 is empty but where the nucleotides 46–48 between the extra arm and T-domain are present, the positions will be filled in the order 48, 46,

Table 3. Modified nucleosides in tRNA and their abbreviations

One-letter code of nucleotides V	Symbol [2,3] V	Name [2,3] V
U	U	uridine
C	C	cytidine
A	A	adenosine
G	G	guanosine
T	T	thymine (for sequences of tRNA genes only)
-		empty position
_ (underline)		insertion (see footnote for further information)
.		unknown nucleotide
H	?A	unknown modified adenosine
*	m1A	1-methyladenosine
/	m2A	2-methyladenosine
+	i6A	N^6-isopentenyladenosine
*	ms2i6A	2-methylthio-N^6-isopentenyladenosine
=	m6A	N^6-methyladenosine
6	t6A	N^6-threonylcarbamoyladenosine
E	m6t6A	N^6-methyl-N^6-threonylcarbamoyladenosine
[	ms2t6A	2-methylthio-N^6-threonylcarbamoyladenosine
:	Am	2'-O-methyladenosine
I	I	inosine
O	m1I	1-methylinosine
^	Ar(p)	2'-O-(5-phospho)ribosyladenosine
'	io6A	N^6-(cis-hydroxyisopentenyl)adenosine
<	?C	unknown modified cytidine
%	s2C	2-thiocytidine
B	Cm	2'-O-methylcytidine
M	ac4C	N^4-acetylcytidine
?	m5C	5-methylcytidine
'	m3C	3-methylcytidine
}	k2C	lysidine
>	f5C	5-formylcytidin
°	f5Cm	2'-O-methyl-5-formylcytidin
;	?G	unknown modified guanosine
S	Gr(p)	2'-O-(5-phospho)ribosylguanosine
K	m1G	1-methylguanosine
L	m2G	N^2-methylguanosine
#	Gm	2'-O-methylguanosine
R	m22G	N^2,N^2-dimethylguanosine
|	m22Gm	N^2,N^2,2'-O-trimethylguanosine
7	m7G	7-methylguanosine
(	fa7d7G	archaeosine
Q	Q	queuosine
8	manQ	mannosyl-queuosine
9	galQ	galactosyl-queuosine
Y	yW	wybutosine
W	o2yW	peroxywybutosine
N	?U	unknown modified uridine
{	mnm5U	5-methylaminomethyluridine
2	s2U	2-thiouridine
J	Um	2'-O-methyluridine
4	s4U	4-thiouridine
&	ncm5U	5-carbamoylmethyluridine
1	mcm5U	5-methoxycarbonylmethyluridine
S	mnm5s2U	5-methylaminomethyl-2-thiouridine
3	mcm5s2U	5-methoxycarbonylmethyl-2-thiouridine
V	cmo5U	uridine 5-oxyacetic acid
5	mo5U	5-methoxyuridine
l	cmnm5U	5-carboxymethylaminomethyluridine
$	cmnm5s2U	5-carboxymethylaminomethyl-2-thiouridine
X	acp3U	3-(3-amino-3-carboxypropyl)uridine
,	mchm5U	5-(carboxyhydroxymethyl)uridinemethyl ester
)	cmnm5Um	5-carboxymethylaminomethyl-2'-O-methyluridine
~	ncm5Um	5-carbamoylmethyl-2'-O-methyluridine
D	D	dihydrouridine
P	Ψ	pseudouridine
]	m1Ψ	1-methylpseudouridine
Z	Ψm	2'-O-methylpseudouridine
T	m5U	ribosylthymine
F	m5s2U	5-methyl-2-thiouridine
\	m5Um	5, 2'-O-dimethyluridine

7, i.e., tRNAs use position 48, 46 and 47 for the first, second and hird nucleotide, respectively, depending on the length of the equence in this region. A similar situation occurs in tRNAs without . long extra arm, where the most variable position 47 is deleted in nany sequences.

Alignment of animal mitochondrial tRNAs

n properly aligned tRNA sequences, nucleotides occupying the ame position in different tRNA sequences should play a comparable structural or functional role. Most animal mitochondrial RNAs cannot be easily aligned with other tRNAs mainly because of the absence of information on their three-dimensional structure. Experimental data, however, point to the existence of tertiary nteractions in these tRNAs. In this compilation, we use an alignment which accounts for these interactions as much as possible. Where we could do so, the animal mitochondrial tRNAs were ncluded in Parts I and II. The alignment of animal mitochondrial RNA is, however, not yet unambiguous.

Some animal mitochondrial tRNAs have completely unusual secondary structure and cannot be fitted in the tRNA alignment used here (Parts I and II). We treated these sequences separately including them into Part III. Here, each particular sequence has its own alignment. To this group belong the tRNAs from: (i) mitochondria of a parasitic worm lacking the T- or D-domain, (ii) mitochondria of mollusks, insects and echinoderm, with extended anticodon and T-stems and (iii) mammalian mitochondria, lacking the D-domain.

For some tRNA genes the secondary structure pattern cannot be clearly established. We have also included these sequences in Part III. It is possible that posttranscriptional modifications of these tRNAs will result in improvement of the secondary structure.

ACKNOWLEDGEMENT

This project was supported by Fonds der Chemischen Industrie, Deutsche Forschungsgemeinschaft, Project Sp 243/5-1.

REFERENCES

1 Sprinzl,M., Steegborn,S., Hübel,F. and Steinberg,S. (1996) *Nucleic Acids Res.*, **24**, 68–72.
2 Limbach,P.A., Crain,P.F. and McCloskey,J.A. (1994) *Nucleic Acids Res.*, **22**, 2183–2196.
3 Crain,P.F. and McCloskey,J.A. (1997) *Nucleic Acids Res.*, **25**, 126–127. [See also this issue *Nucleic Acids Res.* (1998) **26**, 196–197.].
4 Steinberg,S.V. and Kisselev,L.L. (1992) *Biochimie*, **74**, 337–351.

Chapter II

The secondary structure of eukaryotic selenocysteine tRNA: 7/5 versus 9/4

Sergey V. Steinberg, Anatoli Ioudovitch and Robert Cedergren[2]

RNA 1998, **4**, 241-245

[2] Contributions of each author:
Sergey Steinberg – analysis of experimental data
Anatoli Ioudovitch – structural analysis
Robert Cedergren – general supervision

/A (1998). 4:241–245. Cambridge University Press. Printed in the USA.

ETTER TO THE EDITOR

he secondary structure of eukaryotic elenocysteine tRNA: 7/5 versus 9/4

ERGEY V. STEINBERG,[1] ANATOLI IOUDOVITCH, and ROBERT CEDERGREN
partement de Biochimie, Université de Montréal, Montréal, Québec H3C 3J7, Canada

eywords: selenocysteine tRNA; tRNA; tRNA identity; tRNA structure

ITRODUCTION

sertion of selenocysteine into a growing peptide requires the unusual tRNA^{Sec} (Zinoni et al., 1987; Stadtman, 1990; Böck et al., 1991). This tRNA has an extended -stem containing six base pairs, which, in the case of ukaryotic tRNA^{Sec} (euk-tRNA^{Sec}), is the key identity ement for selenylation and phosphorylation (Wu & ross, 1994; Amberg et al., 1996). Two secondary structures have been proposed for the euk-tRNA^{Sec}, which ffer in the base pairing of the acceptor/T helical domain (Diamond et al., 1981; Böck et al., 1991; Sturchler al., 1993). One structure has the normal seven base airs in the acceptor stem and five base pairs in the stem (7/5 structure, Fig. 1, left), and is characterized / an unusually long four-nucleotide unpaired region between the acceptor and D-stems (Connector 1) and an npaired nucleotide, C64a, in the T-stem. The alternate ructure features the normal two nucleotides in Connector 1 and a 13-base pair acceptor/T domain comised of nine base pairs in the acceptor stem and four in e T-stem (9/4 structure, Fig. 1, right). This 9/4 structre was initially proposed by analogy with the prokaryic tRNA^{Sec} (prok-tRNA^{Sec}), which also contains 13 ase pairs in the acceptor/T helical domain. However, in is case, there are eight and five base pairs in the acptor and T-stems, respectively. The acceptor/T helal domain having 13 base pairs is thought to be a key ructural element determining the functionalities patrn of tRNA^{Sec} in both prokaryotes and eukaryotes öck et al., 1991).

Using enzymatic and chemical probing, Sturchler al. (1993) favored the 9/4 structure, for which a threemensional model was proposed. Since then, new experimental data have been collected on serylation, selenylation, and phosphorylation of the euk-tRNA^{Sec} and mutants thereof (Wu & Gross, 1993, 1994; Ohama et al., 1994; Sturchler-Pierrat et al., 1995; Amberg et al., 1996). The point by point analysis presented here shows that the activities of the euk-tRNA^{Sec} and its mutants in serylation, selenylation, and phosphorylation are better explained by the 7/5 structure.

GENERAL CRITERIA

Recently, criteria for the juxtaposition of the acceptor/T and anticodon/D helical domains have been proposed based on the lengths of paired and unpaired regions in the tRNA secondary structure (Steinberg et al., 1997). One criterion requires a minimum of two nucleotides in Connector 1 to facilitate the connection between the acceptor and D-stems. Another states that the T-stem should consist of five or six layers of stacked nucleotides to allow for the normal D/T loop interaction. Violation of either criterion, if not compensated (Steinberg et al., 1997), leads to deformations in the arrangement of the helical domains, which may render the tRNA nonfunctional. Compensations include extension of the anticodon stem to more than the normal six base pairs for a shorter Connector 1 (Steinberg & Cedergren, 1994) and extension of the anticodon/D helical domain to more than the normal 12 layers for a shorter T-stem (Steinberg et al., 1997). In the following analysis, we have assumed that tRNA in serylation, selenylation, and phosphorylation must have the normal juxtaposition of the acceptor/T and anticodon/D helical domains and thus must fulfil the above criteria.

Analysis of the wt euk-tRNA^{Sec}

1. The "7/5" structure could have either five or six nucleotide layers in the T-stem, depending on whether the unpaired nt C64a is bulged or stacked into the helical domain. However, either way, the criteria for a

Reprint requests to: Sergey V. Steinberg, Département de Biomie, Université de Montréal, Montréal, Québec H3C 3J7, Canada; mail: serguei.chteinberg@umontreal.ca.
[1]On the leave from Engelhardt Institute of Molecular Biology, vilova 32, Moscow, Russia.

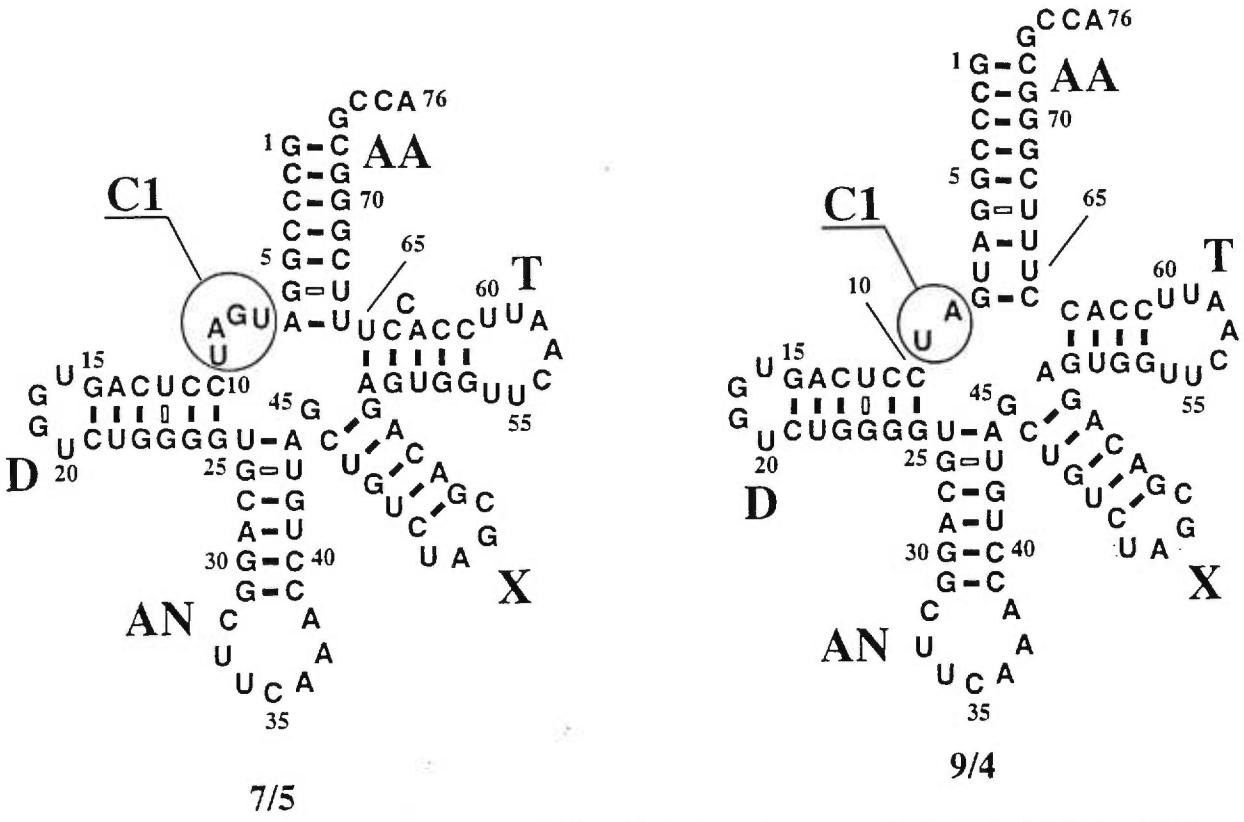


FIGURE 1. Nucleotide sequence of the human tRNA^{Sec} folded into alternate secondary structures: the 7/5 structure to the left and the 9/4 structure to the right. Numbering of nucleotides is taken from Sprinzl et al. (1996) and is different from that used in Sturchler et al. (1993). Nucleotides G9, U20, and C64 are followed by A9a and U9b, by C20b, and by C64a, respectively. AA, D, AN, T, X, and C1 represent the acceptor, D-, anticodon, and T-stems, the extra arm, and Connector 1, respectively. Structure 7/5 has a longer Connector 1 and an unpaired nucleotide in the T-stem.

normal D/T-loop interaction is satisfied (Fig. 2). The 9/4 structure, due to a T-stem of only four base pairs (Steinberg et al., 1997), does not provide for a normal D/T-loop interaction.

2. The 9/4 structure predicts two base pairing combinations, 8–65 and 9–64a. Nucleotide variations at these positions, however, do not support these pairs. Pair 8–65 is U-U in all euk-tRNAs^{Sec} and its conversion into a Watson–Crick or G-U combination has no major effect on either serylation or selenylation (Ohama et al., 1994; Sturchler-Pierrat et al., 1995). The nature of pair 9–64a does not have a Watson–Crick requirement either, because the mutant harboring the G9 → A replacement was effectively serylated and phosphorylated (Wu & Gross, 1994). In contrast, nt 8–65 and 9–64a in the 7/5 model belong to different domains and therefore would not be expected to have Watson–Crick relationships.

The bulged nucleotide in the T-stem

3. A deletion of nt C64a accompanied by replacement G9 → A does not affect either serylation or phosphorylation (mutant X12, Wu & Gross, 1994). The inability of the 9/4 structure to accommodate this mutant was recognized by Wu and Gross (1994, Fig. 1), because no more than seven base pairs could be formed in the acceptor stem. To the contrary, the 7/5 structure is not affected by this deletion (Fig. 2).

4. The replacement of the acceptor/T domain in the euk-tRNA^{Sec} by the corresponding region from the tRNA^{Ser} preserves both serylation and phosphorylation (mutant X9, Wu & Gross, 1993, 1994). This mutant folds exclusively in the 7/5 structure (Fig. 2).

5. The deletion of U65, together with the replacement G9 → A, does not seriously affect either selenylation or phosphorylation (mutant X12H, Amberg et al., 1996). The A49–C64a pair in this mutant can be accommodated in the 7/5 structure (Fig. 2), and, as described in #2 above, the G9 → A replacement does not affect selenylation. The 9/4 structure (see Fig. 5 in Amberg et al., 1996) is an unlikely form for this mutant because, in addition to the formation of pair A9–C64a, the intercalation of the unpaired U8 into the acceptor stem is required. The combination of both irregularities would damage the stability of the acceptor stem.

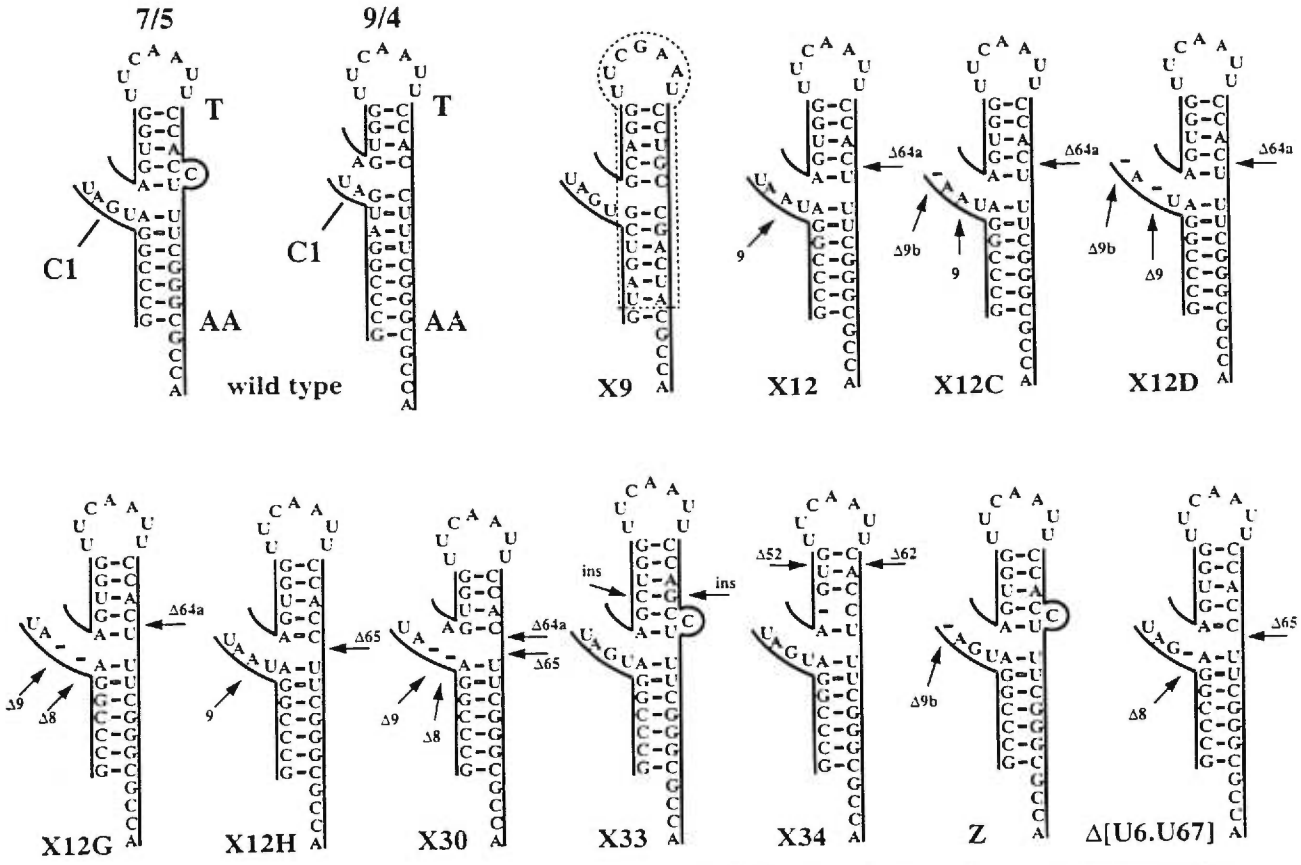


FIGURE 2. Structure of the acceptor/T helical domain in human $tRNA^{Sec}$ and mutants thereof discussed in this paper. For the wt tRNA, both the 7/5 and the 9/4 structures are presented, whereas, for the mutant tRNAs, only the 7/5 structures are shown. AA, T, and C1 in the wt tRNA structures stand for the acceptor stem, the T-stem, and Connector 1, respectively. Arrows indicate the nucleotides in mutant tRNAs that differ from those in the wt euk-$tRNA^{Sec}$. Numbers correspond to the nucleotide positions in Figure 1. Δ and "ins" stand for deletions and insertions, respectively. The region in mutant X9 surrounded by a dashed line, including the D-stem and loop and a part of the acceptor stem, was taken from the $tRNA^{Ser}$ (Wu & Gross, 1994). In mutant Z, nucleotide U9b is not deleted, but rather a part of the D-stem (see #7 in the text and Fig. 3).

The length of Connector 1

3. Deletion of U9b and C64a accompanied by the replacement G9 → A does not seriously affect either selenylation or phosphorylation (mutant X12C, Amberg et al., 1996). However, deletion of C64 deprives mutant X12C of the ability to be folded into the 9/4 structure. Moreover, the intercalation of A9 needed to form a nine-base pair acceptor stem (see Fig. 5, Amberg et al., 1996) leaves only one nucleotide in Connector 1, rendering the normal connection between the acceptor and D-stems sterically impossible. On the other hand, in the 7/5 structure, three nucleotides in Connector 1 would be retained (Fig. 2).

7. Shortening of Connector 1 by one nucleotide does not affect serylation. Ohama et al. (1994) reported that the mutant having two replacements C11 → G and G23 → C in the D-stem (Fig. 3, left) fully preserved the serylation capacity, even though these mutations result in two mismatches, G11–G24 and U12–C23, in the D-stem. A more probable structure of this region involves bulging U12 and forming three new pairs, G11–C23, C10–G24, and U9b-G25 (Fig. 3, right; Fig. 2Z). Because U9b comes from Connector 1 in this structure, Connector 1 must have more than two nucleotides, as in the 7/5 but not in the 9/4 structure.

8. Deletion of two nucleotides from Connector 1 and nt C64a in mutants X12D and X12G does not abolish either selenylation or phosphorylation (Amberg et al., 1996). Only the 7/5 structure is possible for these mutants (Fig. 2): a deletion of two nucleotides from Connector 1 would not affect this secondary structure, because two connector nucleotides remain. However, the attempt to restore the nine-base pair acceptor stem leaves no nucleotides for Connector 1 in the 9/4 structure (see Fig. 5 in Amberg et al., 1996).

The lengths of the acceptor and T-stems

9. Deletion of nt U8–U65 (mutant [U6.U67], Sturchler-Pierrat et al., 1995) is less detrimental for selenylation

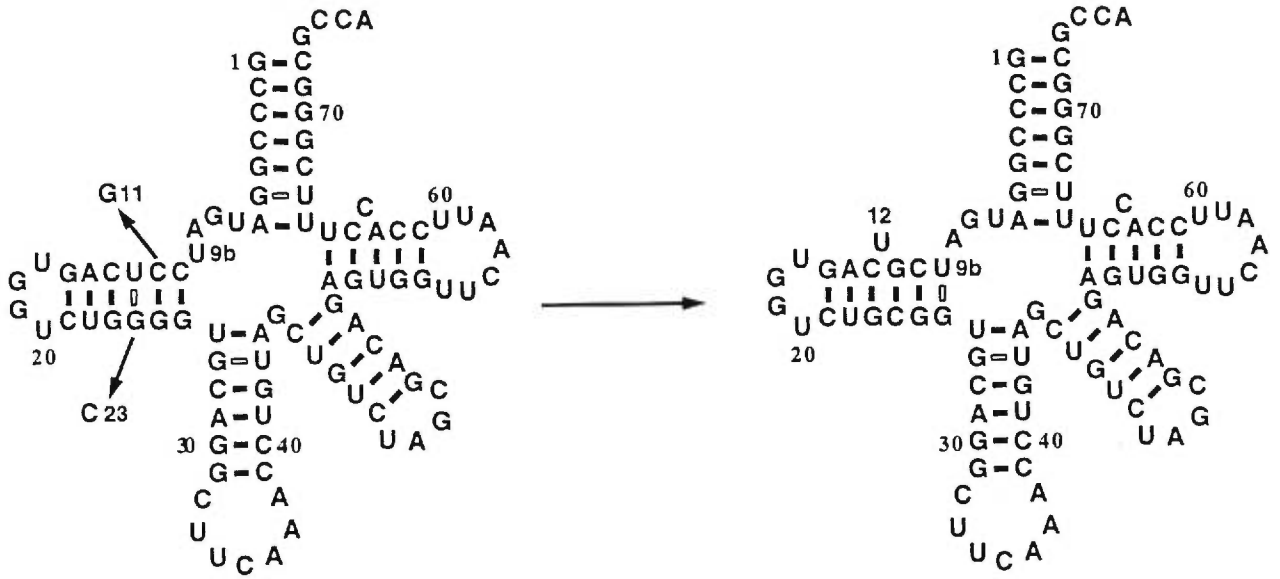


FIGURE 3. Nucleotide sequence of the wt and mutant tRNA^{Sec} from *Homo sapiens* (Ohama et al., 1994, see the acceptor/T domain representation in Fig. 2Z). Replacement of the C11 and G23 by G and C, respectively, results in two mismatches in the D-stem. The normal base pairing can, however, be restored, if U12 is bulged out and U9b is involved in the base pairing. Only 7/5 structure can accommodate this rearrangement. The 9/4 structure leaves only one nucleotide in Connector 1.

han deletion of base pairs C3-G70, G6–U67, or A7–J66 (respectively, [C3-G70], [G5a-U67b], and [A5b-J67a]). None of these deletions can be accommodated n the 9/4 structure, because they result in no more han eight base pairs in the acceptor stem. In the 7/5 structure, however, the U8–U65 combination, unlike the hree other combinations, does not form a base pair (Fig. 2), whereas deletion of U65 or a nucleotide from Connector 1, as mutants X12H and X12C have shown, has only a minor effect on selenylation.

10. A deletion of base pair G52–C62 from the T-stem improves serylation and only slightly diminishes selenylation and phosphorylation (mutant X34, Amberg et al., 1996). The 9/4 model cannot explain this fact because a deletion of a base pair from an already shortened T-stem would make it even more difficult to create the proper D/T-loop interaction. Although the 7/5 model is also affected by this deletion, intercalation of nt C64a could compensate for the deletion and restore the normal D/T-loop interaction (Fig. 2).

11. Deletion of nt U8, G9, C64a, and U65 abolishes both serylation and selenylation (mutant X30, Amberg et al., 1996). This mutant differs from X12G by the additional deletion of U65. In the 7/5 model, this deletion deprives A49 of its Watson–Crick partner in the T-stem, which would leave the latter with only four base pairs, thus preventing the normal D/T-loop interaction (Fig. 2).

12. Insertion of a base pair in the T-stem abolishes serylation (mutant X33, Amberg et al., 1996). Both the 9/4 and 7/5 structures are able to accommodate this mutation: in the 9/4 structure, the addition of a base pair in the T-stem provides the optimal five base pairs, whereas, in the 7/5 structure, it increases the length of the T-stem to the maximally allowable six base pairs (Fig. 2). The situation with the 7/5 structure is different, however, because the unpaired nt C64a (or C64), would have to be bulged, unlike in the wt sequence, to avoid extending of the T-stem to more than six layers. If this nucleotide was bulged, it could prevent the normal interaction with the seryl-tRNA synthetase and abolish the serylation.

This suggestion is compatible with the experimental data indicating that the eukaryotic seryl-tRNA synthetase probably interacts directly with the T-stem. It was recently shown by Acshel and Gross (1993) and by Ohama et al. (1994), that even minor modifications, such as changing of Watson–Crick pairs in this region of the T-stem, decreased the efficiency of serylation. We note that, of all mutants presented here, only those able to fold into a 7/5-type structure without requiring a bulged nucleotide in the T-stem are active in serylation. A bulge in the T-stem abolishing serylation is used in a further analysis (Ioudovitch & Steinberg, 1998) to explain the behavior of euk-tRNA^{Ser} mutants.

CONCLUSION

The above analysis strongly supports the 7/5 structure for the euk-tRNA^{Sec}. It also predicts that the acceptor/T helical domain does not contain any major identity elements for the enzymes involved in selenylation and phosphorylation. The existence of the unpaired nucle-

de in the T-stem of the wt euk-tRNA^{Sec} (nt C64 or i4a) is neither necessary nor harmful for the serylan, selenylation, or phosphorylation. Whether either i4 or C64a is bulged in the solution euk-tRNA^{Sec} ucture is not known, although the fact that the backne between C64a and U65 is sensitive to ribonuclee V1 (specific for stacked and helical regions) while ensitive to ribonuclease T2 (cleaving single-stranded gions) points to the possible insertion of C64a into e double helix (Sturchler et al., 1993). Whether C64 lges or not is less clear, because the linkage beeen C64 and C64a was not cleaved by either of V1 T2. The interpretation of these results may be comomised, however, by the inconsistent behavior of enmes V1 and T2: ribonuclease V1 cleaved between o unstacked nt U60 and C61, whereas ribonuclease : cleaved efficiently in the middle of the D-stem turchler et al., 1993).

Chemical protection experiments (Sturchler et al., 193) show a higher reactivity of N^3-U8 than N^3-35, which is consistent with the fact that U8 beigs to the connector region in the 7/5 structure, ereas U65 pairs to A49. On the other hand, the mplete accessibility observed for nt U12, G50, G52, 53, and A63, known to form base pairs in the D-id T-stems, raises questions about the applicability this approach. It seems that the probing experients do not distinguish well between the two alterte secondary structures, whereas the activity data rongly support the 7/5 model.

CKNOWLEDGMENTS

his work was supported by a grant from the Natural Science id Engineering Council of Canada. R.C. is Richard Ivey ellow of the Canadian Institute of Advanced Research.

REFERENCES

Acshel T, Gross HJ. 1993. Identity determinants of human tRNA^{Ser}: Sequence elements necessary for serylation and maturation of a tRNA with a long extra arm. *EMBO J 12*:3333–3338.

Amberg R, Takaharu M, Wu XQ, Gross HJ. 1996. Selenocysteine synthesis in mammalia: An identity of switch from tRNA^{Ser} to tRNA^{Sec}. *J Mol Biol 263*:8–19.

Böck A, Forchhammer K, Heider J, Baron C. 1991. Selenoprotein synthesis: An expansion of the genetic code. *Trends Biochem Sci 16*:463–467.

Diamond A, Dudock B, Hatfield D. 1981. Structure and properties of a bovine liver UGA suppressor serine tRNA with a triptophane anticodon. *Cell 25*:497–506.

loudovitch A, Steinberg SV. 1998. Modeling the tertiary interactions in the eukaryotic selenocysteine tRNA. *RNA 4*:in press.

Ohama T, Yang DCH, Hatfield DL. 1994. Selenocysteine tRNA and serine tRNA are aminoacylated by the same synthetase, but may manifest different identities with respect to the long extra arm. *Archives Biochem Biophys 315*:293–301.

Sprinzl M, Steegborn C, Hübel F, Steinberg S. 1996. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res 24*:68–72.

Stadtman TC. 1990. Selenium biochemistry. *Annu Rev Biochem 59*:111–127.

Steinberg S, Cedergren R. 1994. Structural compensation in atypical mitochondrial tRNAs. *Nature Struct Biol 1*:507–510.

Steinberg S, Leclerc F, Cedergren R. 1997. Structural rules and conformational compensations in the tRNA L-form. *J Mol Biol 266*:269–282.

Sturchler C, Westhof E, Carbon P, Krol A. 1993. Unique secondary and tertiary features of the eukaryotic selenocysteine tRNA^{Sec}. *Nucleic Acids Res 21*:1073–1079.

Sturchler-Pierrat C, Hubert N, Totsuka T, Mizutani T, Carbon P, Krol A. 1995. Selenocysteylation in eukaryotes necessitates the uniquely long aminoacyl acceptor stem of selenocysteine tRNA^{Sec}. *J Biol Chem 270*:18570–18574.

Wu XQ, Gross HJ. 1993. The long extra arms of human tRNA$^{(Ser)Sec}$ and tRNA^{Ser} function as major identity elements for serylation in an orientation-dependent, but not sequence specific manner. *Nucleic Acids Res 21*:5589–5594.

Wu XQ, Gross HJ. 1994. The length and the secondary structure of the D-stem of human selenocysteine tRNA are the major identity determinants for serine phosphorylation. *EMBO J 13*:241–248.

Zinoni F, Birkmann A, Leinfelder W, Böck A. 1987. Cotranslational insertion of selenocysteine into formate dehydrogenase from *Escherichia coli* directed by a UGA codon. *Proc Natl Acad Sci USA 84*:3156–3160.

Chapter III

Structural compensation in an archaeal selenocysteine transfer RNA

Anatoli Ioudovitch and Sergey V. Steinberg

Accepted for publication in J. Mol. Biol.

Structural compensation in an archael selenocysteine transfer RNA

Anatoli Ioudovitch & Sergey V. Steinberg*
Département de Biochimie, Université de Montréal,
Montréal, Québec, H3C 3J7 Canada

*Corresponding author
Tel. (514) 343-6111 ext. 5165
Fax (514) 343-2210
Email: Serguei.Chteinberg@Umontreal.CA

Running title: *Structural compensation in archael tRNA^{Sec}*

10 March 1999

SUMMARY

A new type of structural compensation between the lengths of two perpendicularly oriented RNA double helices was found in the archael selenocysteine tRNA from *Methanococcus jannascii* . This tRNA contains only four base pairs, one base pair less than in all other cytosolic tRNAs. Our analysis shows that such a T-stem in an otherwise normal tRNA cannot guarantee the formation of the normal interactions between the D and T-loops. The absence of these interactions would affect the juxtaposition of the two tRNA helical domains potentially damaging the tRNA function. In addition to the short T-stem, this tRNA possesses another unprecedented feature, a very long D-stem consisting of seven base pairs. Taken as such, a seven base pair D-stem will also disrupt the normal interaction between the D and T-loops. On the other hand, the presence of the universal nucleotides in both the D and T-loops suggests that in this tRNA these loops probably interact with each other in the same way as in other tRNAs. Here we demonstrate that the short T-stem and the long D-stem can naturally compensate each other thus providing the normal D/T interactions. Molecular modeling technique has helped suggest a detailed scheme of mutual compensation between these two unique structural aspects of the archael selenocysteine tRNA. In light of this analysis, other structural and functional characteristics of the selenocysteine tRNAs are discussed.

Keywords: *tRNA, tRNA structure, selenocysteine, RNA conformation, molecular modeling.*

INTRODUCTION

A very unusual $tRNA^{Sec}$ is found in prokaryotes and higher eukaryotes where it incorporates selenocysteine into the nascent peptide in response to the UGA codons otherwise assigned for the termination of translation (Zinoni *et al.*, 1987; Stadtman 1990; Böck *et al.*, 1991). This tRNA is delivered to the ribosome by a special elongation factor, which also recognizes particular elements of the mRNA secondary structure. The unusual functional pattern of the $tRNAs^{Sec}$ is determined by its unique structure. Both eubacterial and eukaryotic $tRNA^{Sec}$ contain an unprecedented six base pairs in the D-stem, which in the case of the eukaryotic $tRNA^{Sec}$ (euk-$tRNA^{Sec}$) have been shown to be an identity element for the selenylation and phosphorylation (Wu & Gross, 1994; Amberg *et al.*, 1996). The structure of Domain II (Fig. 1) is also abnormal in both $tRNAs^{Sec}$. In the eubacterial $tRNA^{Sec}$ (eub-$tRNA^{Sec}$, Fig. 2a), the acceptor stem contains an unusual eight base pairs, which together with the normal five base pair T-stem (8/5 structure) makes a total of thirteen base pairs in Domain II. In the euk-$tRNA^{Sec}$, the type of abnormality in Domain II depends on which of the two alternative secondary structures is taken (Diamond *et al.*, 1981; Böck *et al.*, 1991; Sturchler *et al.*, 1993). The first structure has the normal seven base pairs in the acceptor stem and five base pairs in the T-stem (7/5 structure, Fig. 2b), but contains an unpaired nucleotide in the middle of the T-stem. The second structure features the abnormal acceptor and T-stems with nine and four base pairs, respectively (9/4 structure, Fig. 2c). The fact that in the 9/4 structure the acceptor stem is longer than normal, as in the eub-$tRNA^{Sec}$ case, was considered as a factor favoring this structure over the 7/5 structure. A recently discovered nucleotide sequence of an archael $tRNA^{Sec}$ (arc-$tRNA^{Sec}$) from *Methanococcus jannaschii* (Bult *et al.*, 1996), also having the 9/4 structure (Fig. 2d), fitted to the hypothesis that a long acceptor stem is a key element determining the functionality of the $tRNA^{Sec}$ in all organisms (Böck *et al.*, 1991).

Such a 9/4 structure for the euk-$tRNA^{Sec}$ raises, however, some questions. The long acceptor stem in this structure comes together with a short T-stem. The existence of

only four base pairs in the T-stem creates problems for the normal juxtaposition of the two helical domains. The T-loop, rigidly connected to the rest of the molecule via the T-stem, is also involved in important tertiary interactions at the corner of the molecule, and a shorter T-stem will affect these interactions. Recently we investigated similar situations in structurally diverged mitochondrial tRNAs and suggested a set of compensatory rules that any changes in tRNA structure must satisfy if the normal juxtaposition of the helical domains is to be preserved (Steinberg *et al.*, 1997). According to these rules, a four base pair T-stem without corresponding compensations in other parts of the molecule does not provide the normal L-form. With this knowledge in mind, we recently analyzed a great body of experimental data on serylation, selenylation and phosphorylation of the euk-tRNA^{Sec} and its numerous mutants (Steinberg *et al.*, 1998). This analysis revealed a synergy between their activity in these three enzymatic processes and their ability to comply with our L-form compensatory rules within the 7/5 structure. In other words, the euk-tRNA^{Sec}, at least in serylation, selenylation and phosphorylation, behaved as though it had the usual 7/5 structure, thus avoiding the problems of a short T-stem associated with the 9/4 structure. A spatial model corresponding to the 7/5 secondary structure, characterized by the normal juxtaposition of the two helical domains and an elaborated system of the tertiary interactions resembling that observed in the eub-tRNA^{Ser}, was suggested (Ioudovitch & Steinberg, 1998).

The problem with the arc-tRNA^{Sec}, the other tRNA supposedly having the 9/4 structure, cannot, however, be resolved in the same way. The 9/4 structure is the only one possible for this tRNA, unlike for its eukaryotic counterpart. The existence of the 9/4 secondary structure in a tRNA without simultaneous changes in other parts of the molecule does not, however, fit to our compensatory rules for the tRNA L-form (Steinberg *et al.*, 1997), and therefore, puts in question the validity of these rules and their applicability to tRNAs other than mitochondrial.

In this paper we analyze the structure of the tRNA^{Sec} from *M. jannaschii* and note that in addition to the shortened T-stem, this tRNA possesses another unique feature which has not been discussed so far, an extraordinary long D-stem made of seven base pairs. We present evidence that the ability of a tRNA to form the normal D/T-loop

interactions affected by a deletion of a base pair in the T-stem, can be restored by additional extension of the D-stem for one more base pair. In this way, the two unique features of the arc-tRNA^{Sec}, the short T-stem and the long D-stem, would compensate each other, thus providing for the normal juxtaposition of the helical domains. In the light of this analysis some structural features of all tRNAs^{Sec} are discussed.

BACKGROUND

The L-form, describing the spatial arrangement of the two helical domains, is common to all known tRNA crystal structures (Ladner *et al.* 1975; Quigley *et al.*, 1975; Moras *et al.,* 1980). Within the L-form, Domain I sticks perpendicularly to the side of Domain II (Fig 1). This arrangement provides the proper juxtaposition of the two tRNA functional centers, the anticodon and the acceptor terminus, and is stabilized by two main interactions between the D and T-loops. In the first interaction, the two universal guanines G18 and G19 of the D-loop form base pairs with Ψ55 and C56 of the T-loop, respectively. A mutual intercalation of nucleotides of the two loops provides a continuous stack of purines A58-G18-R57-G19. G18 and G19 are connected to Domain I by two conformationally flexible regions 16-17-17a and 20-20a-20b. Because of this flexibility, the interaction of the two guanines with the T-loop does not fix the juxtaposition of Domains I and II. To maintain the interaction, the connectors need simply to be long enough, and in the standard tRNA structure one nucleotide in each of the two regions is sufficient for the normal connection (Sprinzl *et al.*, 1998).

In the second interaction, nucleotide 59 of the T-loop stacks to the tertiary base pair 15-48, which constitutes the last stacking layer of Domain I. The T-loop has a special conformation, which is determined by the universal nucleotide sequence G53-T54-Ψ55-C56-R57-A58-N59-N60-C61 (R and N stand for a purine and for any nucleotide, respectively; underlined nucleotides form a base pair in the T-stem) and is stabilized by intensive base-base stacking and H-bonding. Due to this conformation, which is virtually identical in all known tRNA crystal structures, nucleotide 59 has a fixed position with

respect to the whole Domain II. The position of the last stacking layer of Domain I is rigid, in turn, with respect to Domain I. The direct interaction between nucleotide 59 and pair 15-48 thus plays the crucial role in fixing the juxtaposition of Domains I and II. The conservation of this interaction is ensured by the standard lengths of Domain I and of the T-stem. Any deletion of a layer from Domain I or T-stem would affect the standard D/T-loop interaction and/or juxtaposition of the domains, consequently impairing the function of the molecule. It is not surprising, therefore, that Domain I in all non-selenocysteine cytosolic tRNAs consists strictly of twelve stacked layers, of which the first six and the next four are base pairs of the anticodon and D-stems, respectively, and the last two are built as tertiary interactions 8-14-21 and 15-48. The T-stem, in its turn, is also extremely conservative consisting exclusively of five base pairs (Sprinzl et al., 1998).

Interestingly, among mitochondrial tRNAs one can find species that challenge almost every structural aspect found to be invariable in cytosolic tRNAs. For example, one can find mitochondrial tRNAs with Domain I having less than 12 stacked layers, or with the T-stem composed of only four base pairs. These tRNAs were the object of our recent analysis (Steinberg & Cedergren, 1994; Steinberg *et al.*, 1997). We showed that a special role in compensating for these abnormalities could be played by unpaired nucleotides either in the antocodon stem or in the variable region. The intercalation of these nucleotides between base pairs in the anticodon stem or between Domain I and the T-loop could effectively extend Domain I and restore the normal tRNA geometry. Although the L-form compensatory rules have been derived from the analysis of mitochondrial tRNAs, they can be also applied to any other tRNA, in which the normal D/T-loop interactions are to be maintained.

ANALYSIS OF THE arc-tRNA^{Sec} STRUCTURE

The secondary structure of the *M. jannaschii* arc-tRNA^{Sec}, as deduced from its gene sequence, is characterized by a very unusual T-stem with only four base pairs instead of the normal five (Fig. 2d). The deletion of a base pair from the T-stem results in the displacement of the T-loop as a whole from its original position in the conventional

tRNA structure. This displacement can be represented as a shift of 2.8 Å along and a rotation of 33° around the axis of the T-stem double helix. As a result, nucleotide 59 becomes more distant from the anticodon loop and looses the key interaction with the last layer of Domain I (see Fig. 3). The restoration of this interaction will require unfavorable conformational changes in the T and D-domains not existing in any tRNA crystal structure. The presence of the two guanines in the D-loop and of sequence GUUCAAUUC in the T-loop fitting to the universal pattern indicates, on the other hand, that the conformation of the T-loop and the system of the D/T-loop interactions most likely remain intact. In this case, there should be additional aspects in the arc-tRNA^{Sec} structure able to compensate for the absence of a base pair in the T-stem. These aspects, as in the aforementioned structures of the abnormal mitochondrial tRNAs, should provide effective extension of Domain I. However, unlike in the mitochondrial tRNAs, the absence of unpaired nucleotides either in the anticodon stem or in the variable region does not allow the same mechanism of compensation. Surprisingly, the arc-tRNA^{Sec} contains another unique characteristics able to play the compensatory role.

The gap between Domain I and nucleotide 59 of the T-loop can be filled via formation of an additional base pair in the D-stem. As one can see in Fig 2d, nucleotides U16 and A20a constitute a Watson-Crick combination. If they form a base pair, it will increase the length of the D-stem up to seven base pairs making a total of thirteen base pairs in Domain I. Additional structural aspects favor the formation of this base pair. Firstly, it would leave the D-loop with four nucleotides, providing the same sequence pattern 5'-YGGU-3' (Y stands for a pyrimidine) as in all other tRNAs^{Sec}. Secondly, a pyrimidine-purine base pair U16-A20a stacks to the previous pair G15-C20b much better than an alternative, purine-pyrimidine pair would do, thus contributing to the stabilization of the whole D-stem. Finally, this pair would extend Domain I, ensuring its comfortable interaction with the energetically optimal conformation of the T-loop. Needless to say, the formation of this pair makes sense only in a view of the T-loop displacement due to the short T-stem. If the T-domain were normal, this pair would have had to occupy the space assigned for nucleotide 59, forcing the whole T-domain to move from its normal position with consequences potentially detrimental for the tRNA function. Figures 3 and 4 provide

a detailed view of how the formation of U16-A20a pair in the arc-tRNAs^{Sec} compensates for the absence of a base pair in the T-stem.

DISCUSSION

Arc-tRNA^{Sec} among other cytosolic tRNAs

Here we demonstrate that the two parts of the tRNA structure, Domain I and the T-stem, are found to be in a mutually compensatory relationship. The existence of this relationship is not obvious from analysis of the normal cytosolic tRNAs, which share the same universal structural pattern. It is revealed, however, in the *M. jannaschii* arc-tRNA^{Sec}, where both elements experience unprecedented for cytosolic tRNAs deviations from the conventional tRNA structure. The T-stem in the arc-tRNA^{Sec} is only four base pair long, one pair shorter than normal. Domain I, on the contrary, contains thirteen layers of stacked nucleotides, one layer more than in all other cytosolic tRNAs. Each of these two features, taken separately, poses a problem for the formation of the normal D/T loop interaction, making impossible for the tRNA to have the standard L-form. Nonetheless, appearing in the same molecule, they naturally compensate each other, providing the same juxtaposition of the two helical domains as in the other cytosolic tRNAs.

An interesting aspect of this compensation deals with the fact that the axes of the D and T-stems are perpendicular to each other. Until now the possibility that the reduction of one double helix could be compensated for by the extension of another helix has been considered as an indication of their coaxiality. This was used for elucidation of coaxial double helical regions in the ribosomal RNAs (Woese *et al.*, 1983) and later for understanding the structure of unusual mitochondrial tRNAs (Steinberg & Cedergren, 1994). The new type of structural compensation presented here shows that the existence of mutual compensation in the lengths of two helical regions is not necessarily associated with their coaxiality.

The fact that a cytosolic tRNA has this sort of structural compensation implies that the structural rules for the tRNA L-form derived from the analysis of mitochondrial

tRNAs are also valid for tRNAs of other origins. Since all non-selenocysteine cytosolic tRNAs have the standard length of the T-stem and of Domain I, they all obey the L-form compensatory rules by definition. The selenocysteine tRNAs, however, in all organisms where they have been found, display deviations from the standard cloverleaf secondary structure and do not necessarily seem to fit the same rules. Nevertheless, a detailed analysis shows that in fact, all these tRNAs do obey the L-form compensatory rules, which allows them to have the normal juxtaposition of the helical domains.

Each selenocysteine tRNA, however, is characterized by its own peculiarities. Thus, for the eub-tRNA^{Sec}, which contains the standard twelve layers in Domain I and five base pairs in the T-stem, to have the normal D/T interactions has never been a problem. For the euk-tRNA^{Sec}, the two secondary structures, 7/5 and 9/4, have been suggested, with only the former obeying the L-form rules. Recently we showed that at least in three enzymatic processes of serylation, phosphorylation and selenylation, the euk-tRNA^{Sec} behaves as having the 7/5 structure (Steinberg *et al.*, 1998). The last tRNA^{Sec} from archaebacteria is shown here to obey the same rules as well, although in a somewhat different way. Thus the ability the arc-tRNA^{Sec} to have the normal L-form unifies it with all cytosolic tRNAs including all other tRNAs^{Sec}. It also serves as an additional argument in favor of the structural compensation shown here, which was predicted based on analysis of mitochondrial tRNAs (Steinberg *et al.*, 1997) and now for the first time is found in a cytosolic tRNA.

The case of the euk-tRNA^{Sec}

We should admit, however, that whether the eukaryotic tRNA^{Sec} has the 7/5 or 9/4 secondary structure, is still under discussion. Since the same L-form compensatory rules that are used here for analysis of the arc-tRNA^{Sec}, were previously applied for elucidation of the secondary structure of its eukaryotic counterpart, the clarification of the euk-tRNA^{Sec} case is necessary to justify the general applicability of this approach. Our analysis strongly supports 7/5 structure for the euk-tRNA^{Sec} versus 9/4 (Steinberg *et al.*, 1998). Hubert *et al.* (1998), however, argued recently against the 7/5 secondary structure for this tRNA, referring to the archael tRNA^{Sec} as also having the 9/4 secondary structure.

From the analysis presented here it is clear that the existence of the seventh base pair in the D-stem of the arc-tRNA^{Sec} and its absence in the euk-tRNA^{Sec} makes these two cases essentially different. This pair enables the archael molecule to have the standard D/T-loop interactions and juxtaposition of the helical domains even within the 9/4 secondary structure, which is not possible for the eukaryotic molecule.

Hubert *et al.* (1998, #3), however, contested the results of our analysis of the euk-tRNA^{Sec}, saying that "the 9/4 structure does provide for a normal D/T loop interaction, in contrast to what was claimed by Steinberg et al. (1998)". Here they referred to their own 3D model of the euk-tRNA^{Sec} (Sturchler et al., 1993), in which, they said, the D/T interaction was correct, even though the T-stem had only four base pairs. The best way to resolve this controversy would be to compare the ways the interactions between an unmoved D-domain and a displaced T-loop were built in our case and in the model of Sturchler *et al.* (1993). From the stereo-drawing presented by Sturchler *et al.* (1993, Fig. 6a) one can judge that the positions of the junctions between the acceptor and T-stems and between the anticodon and D-stems in that model overlap well with those in the yeast tRNA^{Asp} (Westhof *et al.*, 1985). Following our logic for structures like the 9/4 euk-tRNA^{Sec} that has a short T-stem, this will inevitably lead either to disruption of the D/T interactions, or, if these interactions are to be preserved, to unjustifiable conformational changes in the T and D-domains. It is however unclear, which of these two options was chosen by Sturchler *et al.* (1993), because neither they, nor Hubert et al. (1998) discussed this issue.

The note of Hubert *et al.* (1998, #3) that the ability of the 9/4 structure to have the normal D/T interaction was well attested by protection of N3-C56 from DMS and by their own 3D model of the euk-tRNA^{Sec} (Sturchler *et al.*, 1993), can be accepted only partly. As it was just shown, this model is not detailed enough to support or dismiss any statement on the D/T interactions. The protection of N3-C56 *per se* suggests that the D/T loop interaction, indeed, may be normal, although it says nothing about the secondary structure of this tRNA. The latter may very well be of the 7/5 type, which is obviously consistent with the normal D/T interactions.

As mentioned, the euk-tRNA^{Sec} 7/5 secondary structure is characterized by the presence of an unpaired nucleotide C64 or C64a in the T-stem (Fig. 2d; C64a in our nomenclature corresponds to C66 in that used by Sturchler *et al.* (1993) and by Hubert *et al.* (1998)). If either of these two nucleotides is experimentally shown to be bulged, it would be seen as a strong indication of the 7/5 structure. Hubert et al. (1998), having found no Pb^{+2}-induced cleavage in this region of the euk-tRNA^{Sec}, considered this as an argument against the 7/5 structure. It is worth mentioning, however, that Ciesiolka *et al.* (1998), whom Hubert *et al.* (1998) acknowledged as the establishers of the Pb^{+2} cleavage approach, admitted that "experimental data collected thus far reveal that patterns of hydrolysis induced by Pb^{+2} in different RNA molecules do not always correspond precisely to their secondary structure models". In another place of the same paper the authors were even more specific, saying that "only the U bulge is weakly hydrolized at its 3'-side, the other bulges are not detected by Pb^{+2}". This shows that the treatment of the euk-tRNA^{Sec} with Pb^{+2} is probably not the most adequate procedure to detect whether or not C64 or C64a is bulged. DMS treatment, on the other hand, does not seem to have such obvious drawbacks. Hubert *et al.* showed that C64a was not sensitive to DMS and on this ground again dismissed the 7/5 structure (Hubert *et al.*, 1998, #2). However, the other cytidine, C64, displayed in the same experiment a remarkable sensitivity toward DMS even under native conditions (Hubert *et al.*, 1998, Fig. 2A), which can be seen as an indication of the 7/5 secondary structure with C64 bulged and C64a paired to G50. Taken together, the results of probing experiments can be interpreted in favor of the 7/5 model. We should say, however, that whatever the results of the probing experiments, they would have a limited value with respect to the functional tests, which, as we showed (Steinberg *et al.*, 1998), strongly support the 7/5 model.

Concluding remarks

The formation of the normal D/T interactions, however, does not end the structural problems of the prokaryotic tRNAs^{Sec}. The long acceptor stem found in these

$tRNAs^{Sec}$ will not allow them to fit properly to the ribosomal A and P-sites. These ribosomal sites, shared by all tRNAs, require the exact position of the two tRNA functional centers, the anticodon and the acceptor terminus, and would not tolerate the extension of the acceptor stem even for one (eub-$tRNA^{Sec}$) or two (arc-$tRNA^{Sec}$) base pairs. Therefore, at least during their association with the ribosome, all $tRNAs^{Sec}$ are expected to have the normal seven base pairs in the acceptor stem. One can envisage two alternative strategies for fitting the prokaryotic $tRNAs^{Sec}$ to this general constraint, by disruption of the excess base pairs either at the end of acceptor stem proximate to the acceptor terminus or at the other end proximate to the T-stem. The most important disadvantage of the first strategy is that the nucleotides of the disrupted pairs will still occupy their places close to the ribosomal peptidyl-transferase center, potentially affecting the transpeptidation reaction. This would not happen, however, if the disruption occured at the other end of the acceptor stem. In this case the excessive nucleotides of the 5'-strand of the acceptor stem could be involved in tertiary interactions at the core of the molecule in a way similar to that suggested recently for the euk-$tRNA^{Sec}$ (Ioudovitch & Steinberg, 1998).

The excessive pairs of the acceptor stem proximal to the T-stem could be open not only during the association of the $tRNA^{Sec}$ with the ribosomal A and P sites, but also during some other steps of the $tRNA^{Sec}$ functional cycle. An indirect indication in favor of this possibility comes from the experiment of Rudinger et al. (1996), who showed that the base pairs at the junction point between the acceptor and T-stems of the *Escherichia coli* $tRNA^{Sec}$ constitute a specific structural element not found in any other prokaryotic elongator tRNA that hinders binding of this tRNA to EF-Tu-GTP. On the other hand, the strong complementarity in the acceptor stem of these tRNAs suggests that these pairs may be formed at some steps of their functional cycle not shared by other tRNAs. In this conformation the standard mutual position of the acceptor terminus and the anticodon is no longer maintained. Whether it is true or not, and at which steps such conformational perturbation can happen is a matter for further analysis.

REFERENCES

Amberg, R., Muzitani, T., Wu, X. Q. & Gross, H.J. (1996). Selenocysteine synthesis in mammalia: An identity of switch from $tRNA^{Ser}$ to $tRNA^{Sec}$. *J. Mol. Biol.* **263**, 8-19.

Böck, A., Forchhammer, K., Heider, J. & Baron, C. (1991). Selenocysteine sythesis: An expansion of the genetic code. *Trends Biochem. Sci.* **16**, 463-467.

Bult, C. J. *et al.* (23 authors) (1996). Complete genome sequence of the methanogenic Archaeon, *Methanococcus jannaschii*. *Science* **273**, 1058-1073.

Ciesiolka, J., Michalowski, D., Wrzesinski, J., Krajewski, J. & Krzyzosiak, W. J. (1998). Patterns of cleavage induced by lead ions in defined RNA secondary structure motifs. *J. Mol. Biol.* **275**, 211-220.

Diamond, A., Dudock, B. & Hatfield, D. (1981). Structure and properties of a bovine live UGA supressor serine tRNS with a triptophane anticodon. *Cell* **25**, 497-506.

Ioudovitch, A. & Steinberg S.V. (1998). Modeling the tertiary interactions in the eukaryotic selenocysteine tRNA. *RNA* **4**, 365-373.

Hubert, N., Sturchler, C., Westhof, E., Carbon P. & Krol, A. (1998). The 9/4 secondary structure of eukaryotic selenocysteine tRNA: More pieces of evidence. *RNA* **4**, 1029-1033.

Ladner, J. E., Jack, A., Robertus, J. D., Brown, R. S., Rhodes, D., Clark, B. F. & Klug, A. (1975). Structure of yeast phenylalanine transfer RNA at 2.5 Å resolution. *Proc. Natl. Acad. Sci. USA* **72**, 4414-4418.

Moras, D., Comarond, M. B., Fisher, J., Weiss, R., Thiery, J. C., Ebel, J. P. & Geige, R. (1980). Crystal structure of yeast $tRNA^{Asp}$. *Nature (Lond)* **288**, 669-674.

Quigley, G. J., Seeman, N. C., Wang, A. H. J., Suddath, F. L. & Rich, A. (1975). Yeast phenylalanine tranfer RNA: Atomic coordinates and torsion angles. *Nucleic Acids Res.* **4**, 1649-1665.

Pearlman, D.I. *et al.* (1995). AMBER 4.1. San Francisco: University of California.

Rudinger, J., Hillenbrandt, R., Sprinzl, M. & Giege, R. (1996). Antideterminants present in minihelix(Sec) hinder its recognition by prokaryotic elongator factor Tu. *EMBO J.* **15**, 650-657.

Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A. & Steinberg, S. (1998). Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* **26**, 148-153.

Stadtman, T. C. (1990). Selenium biochemistry. *Annu. Rev. Biochem.* **59**, 111-127.

Steinebrg, S. & Cedergren, R. (1994) Structural compensations in atypical mitochondrial tRNAs. *Nature Struct. Biol.* **1**, 507-510.

Steinberg, S., Leclerc, F. & Cedergen, R. (1997). Structural rules and conformational compensations in the tRNA L-form. *J. Mol. Biol.* **266**, 269-282.

Steinberg, S. V., Ioudovitch, A. & Cedergren, R. (1998). The secondary structure of eukaryotic selenocysteine tRNA: 7/5 *versus* 9/4. *RNA* **4**, 241-245.

Sturchler, C., Westhof, E., Carbon, P. & Krol, A. (1993). Unique secondary structure and tertiary features of the eukaryotic selenocysteine $tRNA^{Sec}$. *Nucleic Acids Res.* **21**, 1073-1079.

Wu, X. Q. & Gross, H. J. (1994). The length and the secondary structure of the D-stem of human selenocysteine tRNA are the major identity determinants for serine phosporylation. *EMBO J.* **13**, 214-248.

Westhof, E., Dumas, P. & Moras, D. (1985). Crystallographic refinement of yeast aspartic acid transfer RNA. *J. Mol. Biol.* **184**, 119-145.

Woese, C. R., Gutell, R., Gupta, R. & Noller H. F. (1983). Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiol Rev.* **47**, 621-69

Zinoni, F., Birkmann, A., Leinfelder, W. & Böck, A. (1987). Cotranslational insertion of selenocysteine into formate dehydrogenase from *Escherichia coli* directed by a UGA codon. *Proc. Natl. Acad. Sci. USA* **84**, 3156-3160.

LEGENDS TO THE FIGURES

Figure 1

The standard tRNA L-form. Open rectangles represent base paired nucleotides; filled and crosshatched rectangles stand for nucleotides of the D and T-loops, respectively. Checkered rectangles represent the unpaired nucleotides between helical domains and at the amino acid terminus. The small figures 1 to 12 refer to the layers of stacked nucleotides starting from the base pair closest to the anticodon loop. Numbers 59 and 60 refer to the T-loop nucleotides in the standard tRNA nomenclature (Sprinzl *et al.*, 1998). Nucleotide 59 stacks to the last, twelfth layer of Domain I. Unstacked nucleotides in the D-loop are not shown.

Figure 2

Clover-leaf secondary structures of the eubacterial (a), eukaryotic (b, c) and archaebacterial (d) $tRNAs^{Sec}$. For the eukaryotic tRNA two possible secondary structures, 7/5 (b) and 9/4 (c), are shown. For the eubacterial and archaebacterial RNAs, the presented 8/5 and 9/4 structures are the only possible ones. An additional seventh base pair U16-A20 in the arc-$tRNA^{Sec}$ (d) is marked by the broken line. The nomenclature of nucleotides is taken from the Compilation of tRNA sequences and sequences of tRNA genes (Sprinzl *et al.*, 1998).

Figure 3

Mechanism of the compensation of the short T-stem in the arc-$tRNA^{Sec}$ by the extension of the D-stem. The structures of Domain I are shown in the same way as in Figure 1. In addition, open circles represent unstacked nucleotides. The anticodon stem consists of six layers numbered from 1 to 6. The D-stem covers layers 7-13, 7-12 and 7-10 in the archael (a), eubacterial or eukaryotic (b) $tRNA^{Sec}$ and the normal cytosolic tRNA (c), respectively. In the normal cytosolic tRNA (c) layers 11 and 12 are formed by

nucleotides of the D-loop and of the interdomain connector regions, while in the $tRNA^{Sec}$ (a and b) they are base pairs of the D-stem. In addition, the D-stem of the arc-$tRNA^{Sec}$ (a) contains a base pair in layer 13. The formation of this base pair compensates for the displacement of the T-loop due to the short T-stem (T = 4). As a result, T-loop stacks properly to Domain I in spite of the absence of a base pair in the T-stem.

Figure 4.

Stereo representation of the D/T loop interactions in the model of the *M. jannaschii* $tRNA^{Sec}$ (a), in the 7/5 model of the euk-$tRNA^{Sec}$ (b, Ioudovitch & Steinebrg, 1998) and in the yeast $tRNA^{Phe}$ (c, Ladner *et al.*, 1975). In each structure the D-loop is positioned in the middle, the proximate part of the D-stem is shown on the left, while the T-loop is shown on the right. The nucleotides constituting the last layer of Domain I (layer 13 in a and layer 12 in b and c) and nucleotide 59 of the T-loop (layer 14 in a and layer 13 in b and c) are shown in black. The displacement of nucleotide 59 in (a) from layer 13 to 14 is accompanied by the corresponding extension of Domain I from 12 to 13 layers. The model (a) has been subjected to partial energy minimization in the AMBER forcefield (Pearlman *et al.*, 1995) to resolve steric clashes.

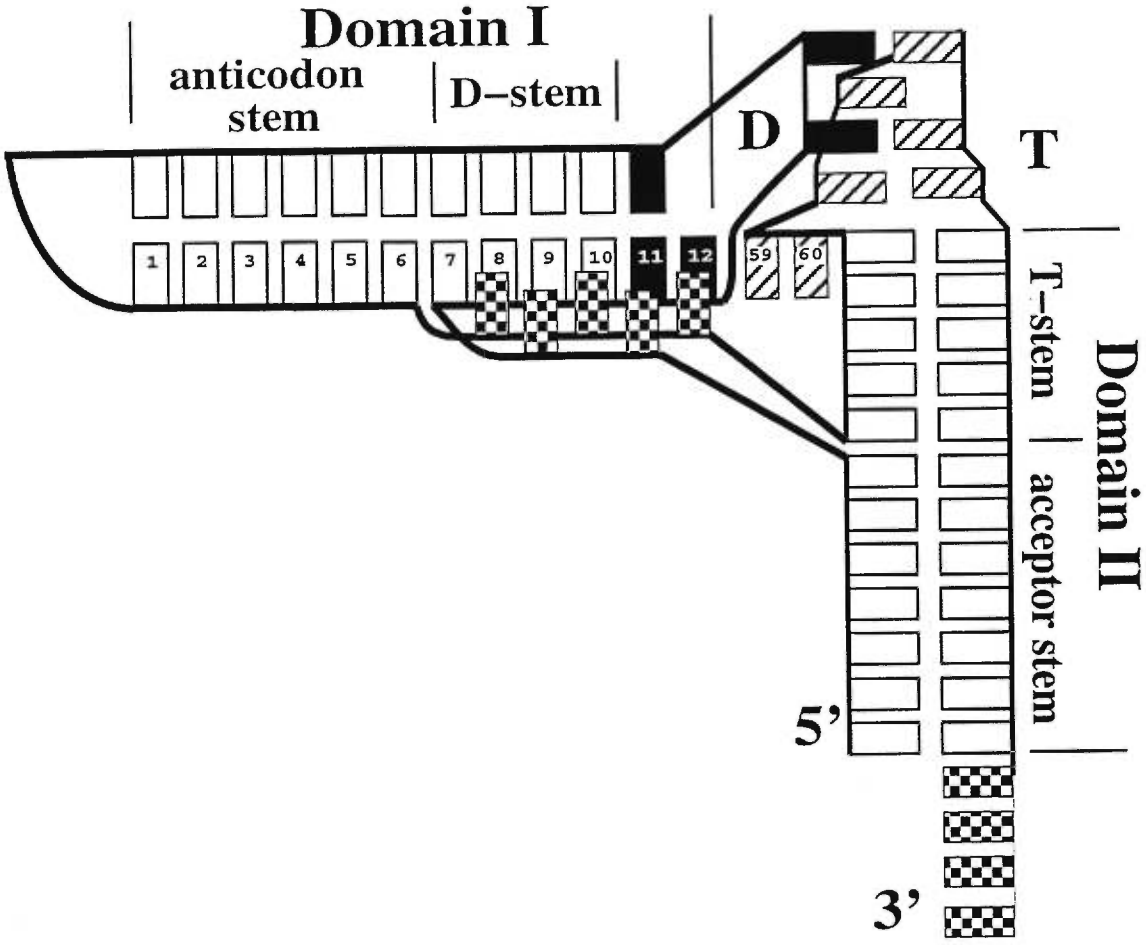
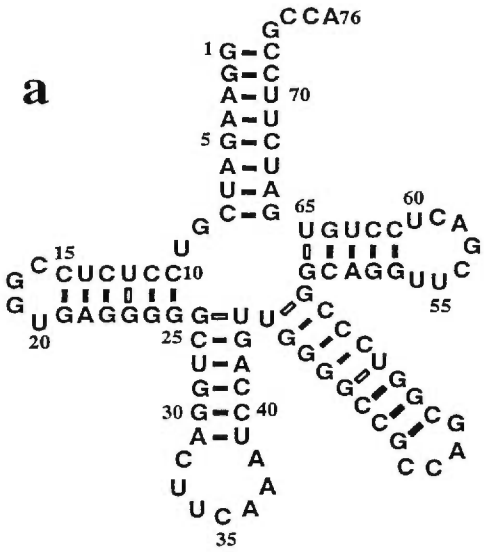
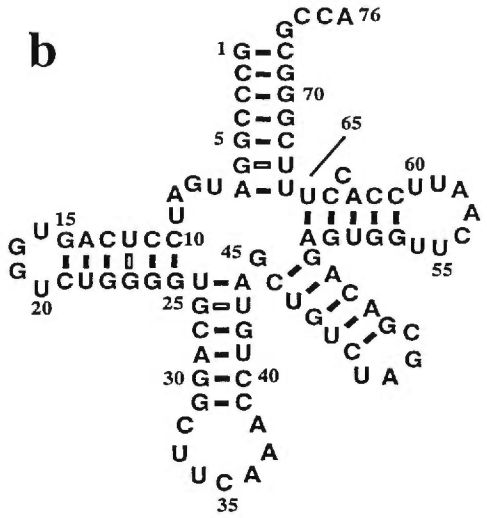

Domain I
anticodon stem
D–stem
D
T
1
2
3
4
5
6
7
8
9
10
11
12
59
60
T–stem
acceptor stem
Domain II
5'
3'

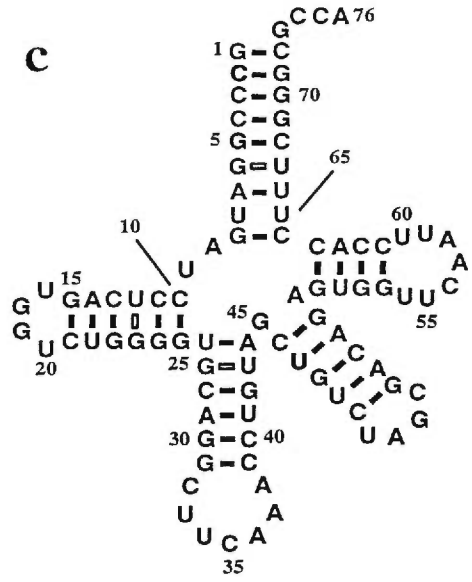
Figure 1.



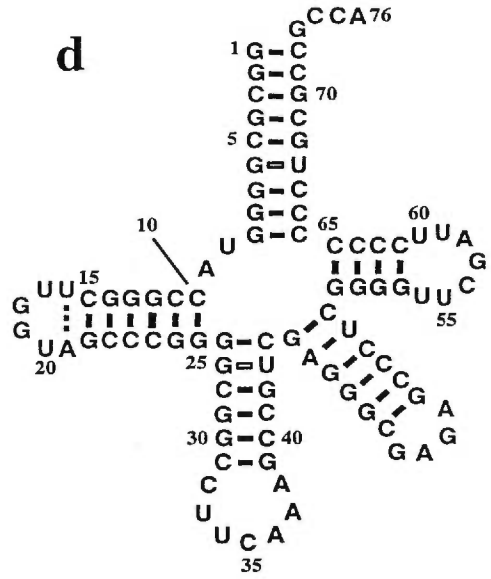
8/5 eub–tRNA^{Sec} *Escherichia coli*



7/5 euk–tRNA^{Sec} *Homo sapiens*



9/4 euk–tRNA^{Sec} *Homo sapiens*



9/4 arc–tRNA^{Sec} *Methanococcus janaschii*

Figure 2.

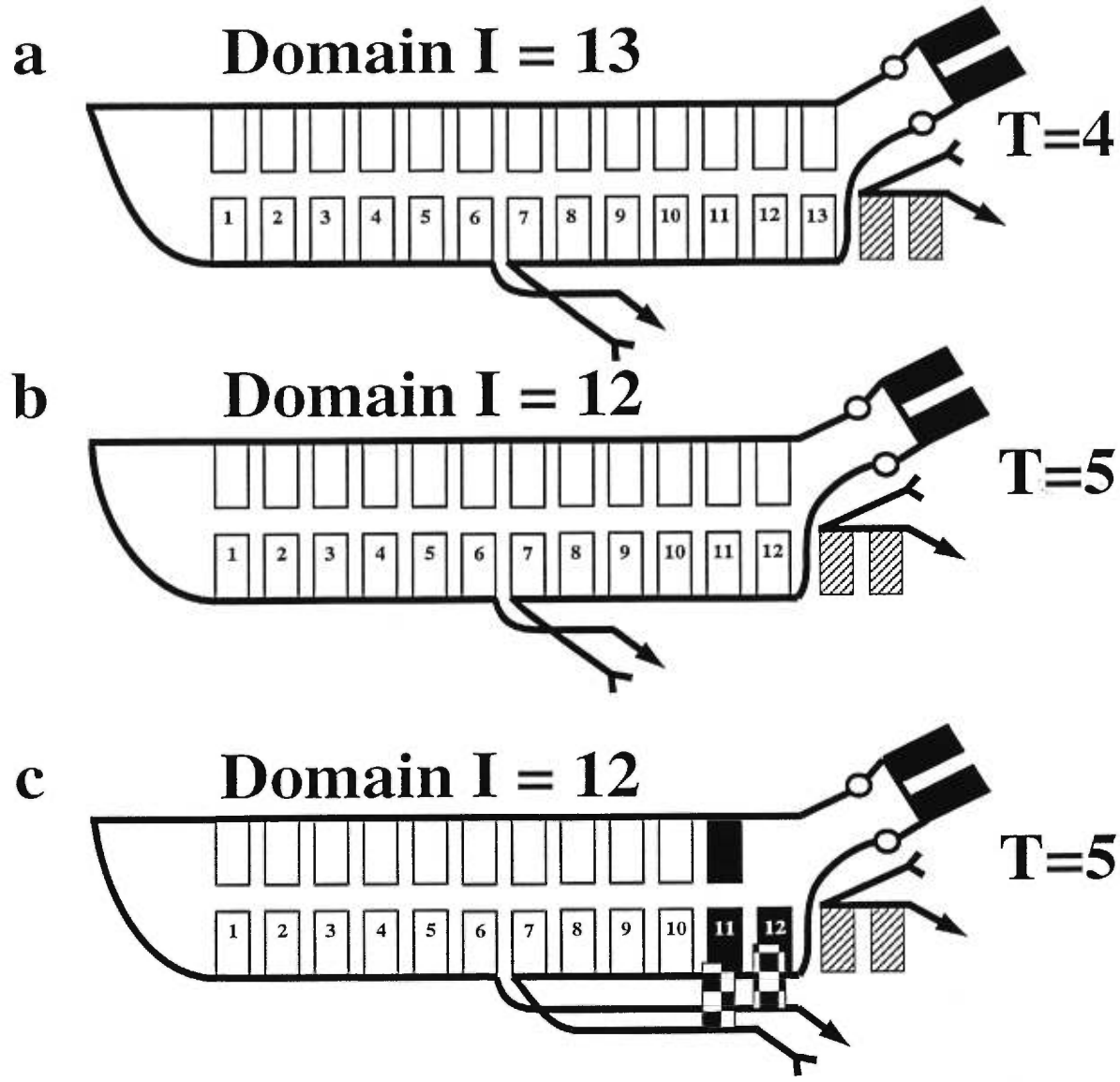

a
Domain I = 13
T=4
1 2 3 4 5 6 7 8 9 10 11 12 13
b
Domain I = 12
T=5
1 2 3 4 5 6 7 8 9 10 11 12
c
Domain I = 12
T=5
1 2 3 4 5 6 7 8 9 10 11 12

Figure 3.

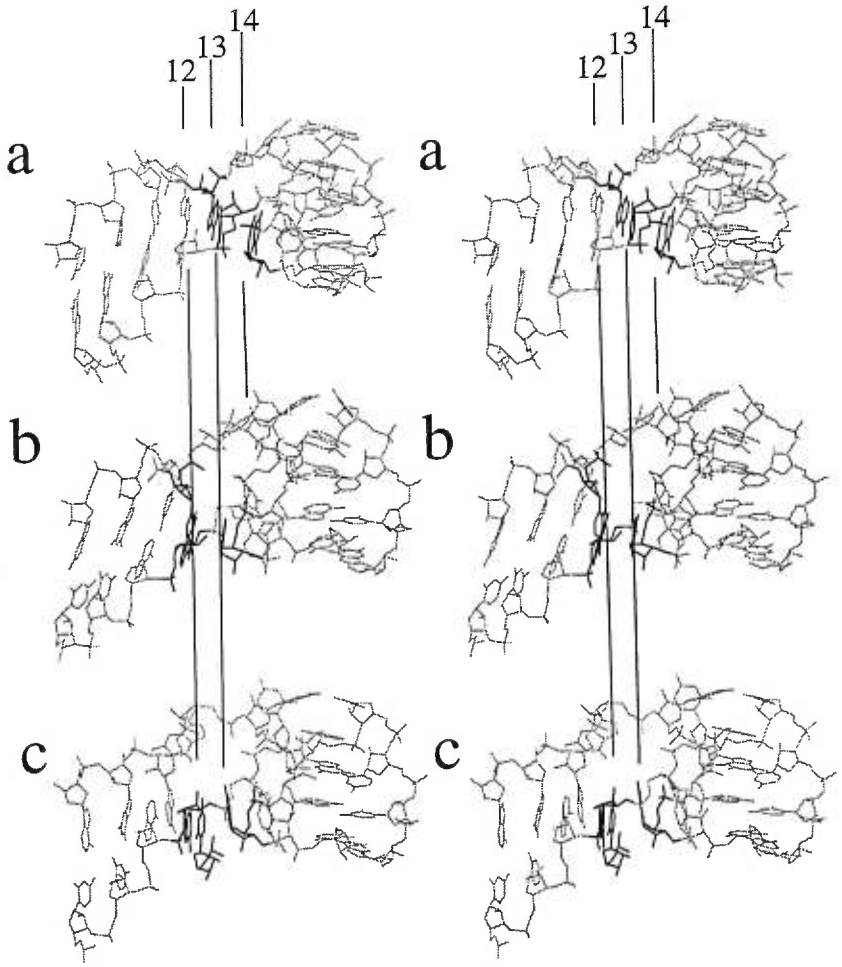


Figure 4.

Chapter IV

Modeling the tertiary interactions in the selenocysteine tRNA

Anatoli Ioudovitch and Sergey S. Steinberg

RNA 1998, **4**, 365-373

RNA (1998). 4:365–373. Cambridge University Press. Printed in the USA.

Modeling the tertiary interactions in the eukaryotic selenocysteine tRNA

ANATOLI IOUDOVITCH[1] and SERGEY V. STEINBERG[1,2]
[1]Département de Biochimie. Université de Montréal. Montréal. Québec. H3C 3J7 Canada

ABSTRACT

A novel three-dimensional model of tertiary interactions in the core region of the eukaryotic selenocysteine tRNA is proposed based on the analysis of available nucleotide sequences. The model features the 7/5 tRNA^{Sec} secondary structure characterized by seven and five base pairs in the acceptor and T-stems, respectively, and four nucleotides in the connector region between the acceptor and D-stems. The model suggests a unique system of tertiary interactions in the area between the major groove of the D-stem and the first base pair of the extra arm that provides a rigid orientation of the extra arm and contributes to the overall stability of the molecule. The model is consistent with available experimental data on serylation, selenylation, and phosphorylation of different tRNA^{Sec} mutants. The important similarity between the proposed model and the structure of the tRNA^{Ser} is shown. Based on this similarity, the ability of some tRNA^{Ser} mutants to be serylated, selenylated, and phosphorylated was evaluated and found to be in a good agreement with experimental data.

Keywords: computer graphics; computer modeling; RNA conformation; RNA structure; selenocysteine; tRNA structure

INTRODUCTION

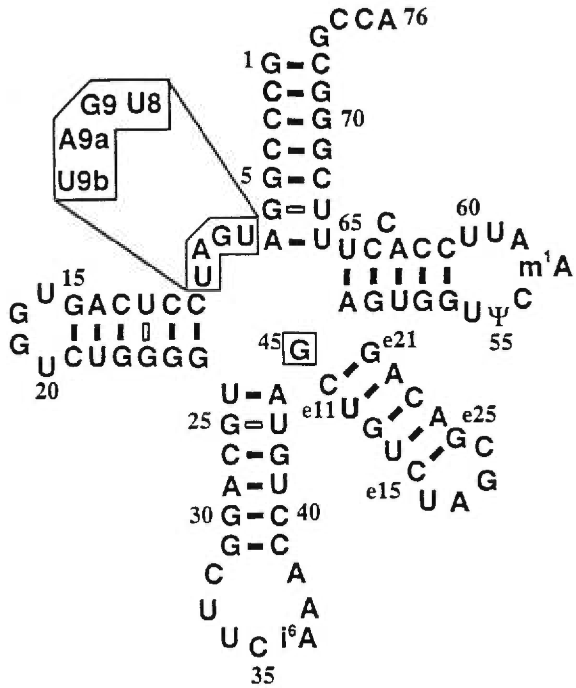
Selenocysteine tRNA, found in prokaryotes and higher eukaryotes, incorporates selenocysteine in response to the UGA stop codons (Zinoni et al., 1987; Stadtman, 1990; Böck et al., 1991). This unusual functional pattern is thought to be a result of its unusual structure: both prokaryotic and eukaryotic selenocysteine tRNAs (prok-tRNA^{Sec} and euk-tRNA^{Sec}) have an unprecedented six-base pair D-stem, which, in the case of the euk-tRNA^{Sec}, has been shown to serve as a major identity element for the selenocysteine synthase and kinase, converting the attached seryl residue into selenocysteine and phosphoserine, respectively (Wu & Gross, 1994; Amberg et al., 1996).

Two alternate secondary structures of the euk-tRNA^{Sec} able to accommodate all phylogenetically related nucleotide sequences have been proposed having seven and five (7/5 structure, Fig. 1) or nine and four (9/4 structure) base pairs in the acceptor and T-stems, respectively (Diamond et al., 1981; Böck et al., 1991; Sturchler et al., 1993). As we showed recently (Steinberg et al., 1998), the available experimental data on serylation, selenylation, and phosphorylation of different euk-RNA^{Sec} mutants support the 7/5 rather than the 9/4 structure, because many euk-tRNA^{Sec} mutants unable to fold into the 9/4 structure are active in these distinctive enzymatic processes. On the other hand, the loss of the ability to fold into the 7/5 structure is associated with loss of these activities. In the 7/5 secondary structure, the connector region between the acceptor and D-stems (Connector 1) has four nucleotides, twice as many as in any other cytosolic tRNA. From the known tRNA crystal structures, it is not obvious how such a long four-nucleotide Connector 1 could be arranged. We suggest a novel three-dimensional model for the core region of the euk-tRNA^{Sec} where the connector nucleotides form a unique system of tertiary interactions in the area between the extra arm and the D-stem. These interactions provide a rigid orientation of the extra arm and contribute to the overall stability of the molecule. In view of this model, important characteristics of the euk-tRNA^{Sec}, euk-tRNA^{Ser}, and their mutants are discussed.

THE MODEL

With reference to the standard tRNA structure (Ladner et al., 1975; Quigley et al., 1975; Moras et al., 1980; Biou et al., 1994), the four nucleotides of Connector 1

Reprint requests to: Sergey V. Steinberg. Département de Biochimie. Université de Montréal. Montréal. Québec. H3C 3J7 Canada; e-mail: serguei.chteinberg@umontreal.ca.
[2]On the leave from Engelhardt Institute of Molecular Biology, Vavilova 32. Moscow. Russia.



IGURE 1. Nucleotide sequence of the human tRNA^{Sec} folded into ıe 7/5-type cloverleaf secondary structure. Tertiary nucleotides U8, i9, A9a, and U9b of Connector 1 and nt G45 are boxed. Numbering f the nucleotides is taken from Sprinzl et al. (1996). Nucleotides G9, '20, and C64 are followed by A9a and U9b, by C20b, and by C64a, ›spectively.

.nd nt G45 should be confined to the area between the ınticodon stem, the major groove of the D-stem, the ›xtra arm, and the T-loop, where they would be ex-ıected to form tertiary interactions (hence we will call hem tertiary nucleotides and the area where they are ɔcated, the tertiary area). Four tertiary nucleotides U8, ʒ9, A9a, and G45 are conserved in all known euk-RNAs^{Sec}, whereas nucleotide 9b can be U, C, or A Table 1).

For the modeling, we first generated all possible ar-angements of the tertiary nucleotides in which each of he nucleotides U8, G9, A9a, and G45 either stacks to ınother tertiary nucleotide and forms H-bonds with the)-stem, or stacks to the last base pair e11–e21 of the ›xtra arm. This included 8, 30, and 12 arrangements vith zero, one, and two tertiary nucleotides, respec-ively, stacked to pair e11–e21. Because in the known RNA structures, nt 26 and 44 never form a Watson–Crick pair, different schemes of H-bonding for pair U26–\44 were considered. All arrangements were subjected o a partial energy minimization in the AMBER force-ield (Pearlman et al., 1995) to select those that could ıave the standard geometry for the polynucleotide chain ınd H-bonds.

Two solutions were found from this analysis that have /ery similar positions of the tertiary nucleotides while ɟiffering in the structure of pair U26–A44. In one case, J26–A44 is a Watson–Crick pair, whereas in the other,

TABLE 1. Nucleotide sequences of the euk-tRNASec.[a]

DZ7560	*Caenorhabditis eleg.*	GCCCGGA	TGAA	CCATGG	CGGT	CTGTGG	TGCAGACTTCAAATCTGTA	G	GCGGT	TAGC	GCCGC	AGTGGTTCGACTCCACCT	TTCGGGTG---
		=====		===**=		=**===	=*==== ====*=		====*		*====	===== ==== =	=*====*
DZ7742	*Drosophila melano.*	GCCCCAC	TGAA	CTTCGG	TGGT	CCGGGG	TGCGGACTTCAAATCCGTA	G	TCGAT	TTGC	GTCGA	AGTGGTTCGATTCCACCT	GGGGGGCG---
		===== =		=**===		===**=	=*==== ====*=		====*		*====	===== ==== =	======
DZ7920	*Xenopus laevis*	GCCCGGA	TGAC	CCTCAG	TGGT	CTGGGG	TGCAGGCTTCAAACCTGTA	G	CTGTC	TAGC	GACAG	AGTGGTTCAATTCCACCT	TTCGGGCG---
		=====*=		==*===		===*==	=*==== ====*=		=====		=====	===== ==== =	=*=====
DZ8040	Chicken	GCCCGGA	TGAC	CCTCAG	TGGT	CTGGGG	TGCAGGCTTCAAACCTGTA	G	CTGTC	TAGC	GACAG	AGTGGTTCAATTCCACCT	TTCGGGCG---
		=====*=		==*===		===*==	=*==== ====*=		=====		=====	===== ==== =	=*=====
DZ9281	Bovine	GCCCGGA	TGAT	CCTCAG	TGGT	CTGGGG	TGCAGGCTTCAAACCTGTA	G	CTGTC	TAGC	GACAG	AGTGGTTCAATTCCACCT	TTCGGGCG---
		=====*=		==*===		===*==	=*==== ====*=		=====		=====	===== ==== =	=*=====
RZ8100	Mouse liver	GCCCGGA	UGAU	CCUCAG	UGGU	CUGGGG	UGCAGGCUNCA+ACCUGUA	G	CUGUU	UAGC	GACAG	AGUGGUPCA"UUCCACCU	UUCGGGCGCCA
		=====*=		==*===		===*==	=*==== ====*=		====*		*====	===== ==== =	=*=====
RZ9990	Human HeLa cells	GCCCGGA	UGAU	CCUCAG	UGGU	CUGGGG	UGCAGGCUNCA+ACCUGUA	G	CUGUC	UAGC	GACAG	AGUGGUPCA"UUCCACCU	UUCGGGCGCCA
		=====*=		==*===		===*==	=*==== ====*=		=====		=====	===== ==== =	=*=====

[a]ID index of each tRNA corresponds to that of the tRNA compilation (Sprinzl et al., 1996). Sequences determined by gene sequencing contain T, whereas those determined by sequencing of mature tRNAs contain U and modified nucleotides. Modified nucleotides are designated as in Sprinzl et al. (1996). Complementary regions in the secondary structure are shown at the top. Nucleotides of the secondary structure regions that form a Watson–Crick or G-U (G-T) pair are underlined by "=" and "*", respectively.

own in Figure 2, it is a Hoogsteen pair. The central ement of the arrangement, identical in both struc- es, consists of perpendicularly oriented purines G9 d A9a (Fig. 3A). G9 stacks to the first base pair of the tra arm (Fig. 3B), whereas A9a forms two H-bonds h the pair U13–A22 of the D-stem via its N6-H and atoms (Fig. 3C). Two other tertiary nucleotides, U8 d G45, stabilize the G9-A9a juxtaposition by stack- to A9a on both sides. U8 forms an H-bond with pair 4–U21 and interacts with C20b and the phosphate A49. G45, on the other hand, forms two H-bonds h G23. The last tertiary nucleotide U9b, being bulged t, is not involved in any specific interactions.

Although two different arrangements are sterically pos- le, one of them, harboring a Hoogsteen pair U26– 4, is clearly preferable, because, in this case, the cleotide surfaces are much better protected. Thus, this structure, C10 and U26 stack to A44 and G27, spectively. In the alternate structure with a Watson– ick pair U26–A44, nt C10, U26, and A44 are essen- lly exposed to the solvent. Another important feature the Hoogsteen pair-containing structure is that re- n 44–45 of the backbone protects nt G45 and inter- ts with amino groups of C10 and C11 (Fig. 2C). Atom P and the base of G45 interact in a similar way to interaction of O1P atom of A152 and the base of 150 in the structure of group I intron (Cate et al., 96) and to the interaction of O1P atom of AL3 with base of GL1 in the hammerhead ribozyme (Pley al., 1994). Atom O1P, in turn, forms a hydrogen bond th the amino groups of C10 and C11.

SCUSSION

ıe model of tertiary interactions in the euk-tRNA^{Sec} re region presented here is based on the compara- e analysis of the available euk-tRNA^{Sec} nucleotide quences. The most important features of the sec- ıdary structure are the D-stem having six base pairs ıd Connector 1 containing four nucleotides. Three rtiary nucleotides, U8, A9a, and G45, directly interact th the D-stem while stacking to each other. Another rtiary nucleotide, G9, is positioned perpendicularly to ese nucleotides. The identity of the central tertiary ıcleotide, A9a, is important not only for the H-bonding th the D-stem, but also crucial for the interaction with 9. A replacement of A9a by G would cause a collision its NH_2-group with G9, altering the whole structure. This arrangement of the tertiary nucleotides pro- ides a comfortable dock to moor the extra arm, en- ıling it to fix its orientation with respect to the rest of e molecule. This aspect differs from the model sug- ısted by Sturchler et al. (1993), where the extra arm ıes not interact with either the D-domain or Connec- r 1. The idea of a fixed extra arm gains support from e available experimental data showing that at least ıme aspects of the Class II tRNA function strongly depend on the orientation of the extra arm (Himeno et al., 1990; Wu & Gross, 1993; Asahara et al., 1994; Biou et al., 1994). The model was built for the human tRNA^{Sec}; however, it fits other vertebrate tRNAs^{Sec} and with only minor modifications accommodates all other eukaryotic tRNAs^{Sec}.

Comparison with euk-tRNA^{Sec} mutants

The suggested model of the tertiary interactions in the euk-tRNAs^{Sec} is consistent with the data on euk-tRNAs^{Sec} mutants. Thus, mutants X14, X17, X19, X29, X35,[3] in which base pair replacements in the D-stem did not have any dramatic effect on the arrangement of tertiary nucleotides, were effectively serylated, selenylated, and phosphorylated (Wu & Gross, 1994; Amberg et al., 1996).

In our model, unlike in that of Sturchler et al. (1993), nt G9 is squeezed between A9a and the first base pair of the extra arm and is not involved in specific H-bonding. Its special position benefits more from the fact that it is a purine than from its H-bonding potentials. Accordingly, mutation G9 → A affects neither serylation nor phosphorylation (Wu & Gross, 1994). Even cytidine in this position is possible, although it makes the serylation less efficient (Ohama et al., 1994).

U9b does not play a decisive role in the model. This correlates with the variability of this position in different euk-tRNAs^{Sec} (Table 1; Sprinzl et al., 1996). In addition, Ohama et al. (1994) demonstrated that mutation U9b → C did not affect the serylation. Our modeling experiments also indicate that a minor reorientation allows A9a to be connected directly to C10. Correspondingly, the deletion of U9b (mutant X12C, Amberg et al., 1996) does not seriously affect either selenylation of phosphorylation. Also, a double mutation [C11 → G; G23 → C] causing U9b to be involved in base pairing in the D-stem (Ohama et al., 1994; Steinberg et al., 1998) improved serylation. The redundancy of U9b does not fit the model of Sturchler et al. (1993), where U9b is essential for the connection between the acceptor and D-stems. A deletion of two nucleotides from Connector 1 would seriously weaken the tertiary interactions and reorient the extra arm. Correspondingly, in mutants X12D and X12G, the level of the selenylation and phosphorylation (in X12D) was notably decreased (Amberg et al., 1996).

The loop-like conformation of the G45 backbone enables A44 to be directly connected to the extra arm in case G45 is deleted. Indeed, mutant X2 with such a deletion displayed only a minor decrease in serylation and phosphorylation (Wu & Gross, 1994). The deletion of two nucleotides, A44 and G45, creates serious prob-

[3]All mutants retain the names given to them in the original articles (Acshel & Gross, 1993; Wu & Gross, 1993, 1994; Amberg et al., 1996) from which the data are derived.

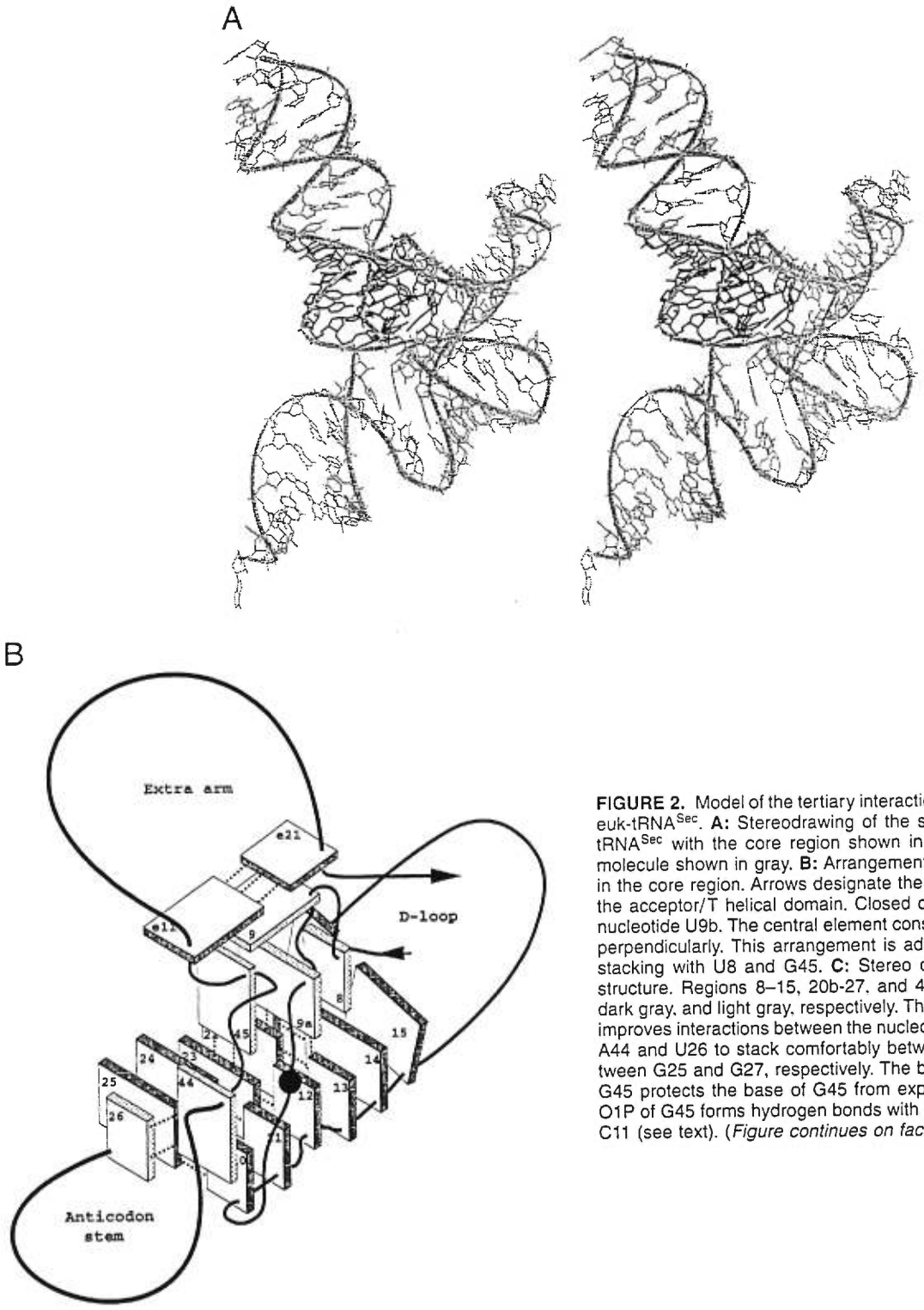


FIGURE 2. Model of the tertiary interactions in the core region of the euk-tRNA^{Sec}. **A:** Stereodrawing of the structure of the whole euk-tRNA^{Sec} with the core region shown in black and the rest of the molecule shown in gray. **B:** Arrangement of the tertiary nucleotides in the core region. Arrows designate the regions of connection with the acceptor/T helical domain. Closed circle stands for the bulged nucleotide U9b. The central element consists of nt G9-A9a arranged perpendicularly. This arrangement is additionally stabilized by A9a stacking with U8 and G45. **C:** Stereo drawing of the core region structure. Regions 8–15, 20b-27, and 43-e21 are shown in black, dark gray, and light gray, respectively. The Hoogsteen pair U26–A44 improves interactions between the nucleotides in the area. It enables A44 and U26 to stack comfortably between C10 and G27 and between G25 and G27, respectively. The backbone between A44 and G45 protects the base of G45 from exposure to the solvent. Atom O1P of G45 forms hydrogen bonds with the NH_2-groups of C10 and C11 (see text). (*Figure continues on facing page.*)

lems for the connection between the anticodon stem and the extra arm, which can explain a very poor serylation of X4 (Wu & Gross, 1994). Such problems, however, did not arise when these two nucleotides, instead of being deleted, were replaced by pyrimidines U44–C45 (mutant X5). This mutant was serylated only a little less efficiently that the wt euk-tRNA^{Sec} (Wu & Gross, 1994). These experimental data are in agreement with our analysis in that, although the interactions in which A44 and G45 are involved contribute to stabilization of the tRNA structure, they are not critical for the tRNA tertiary structure.

C

FIGURE 2. (*continued.*)

Comparison with other Class II tRNAs

The secondary structure of the core region in the euk-tRNA^{Sec} differs fundamentally from that in other Class II tRNAs. Six base pairs in the D-stem instead of normal three, four nucleotides in Connector 1 instead of two, an unpaired nucleotide just before and no unpaired nucleotides right after the extra arm make the euk-tRNA^{Sec} unique both in terms of secondary and tertiary structure. In spite of this, important similarities have been revealed between our model and other Class II tRNAs in the arrangement of their tertiary interactions.

The *Thermus thermophilus* tRNA^{Ser} is the only Class II tRNA for which an X-ray structure has been determined (Biou et al., 1994). The key elements of this structure include the three-base pair D-stem, the two-nucleotide Connector 1, triple G13-A22-G9, and nt G21 from the D-loop squeezed between Connector 1 and the extra arm (Fig. 4C). As our modeling experiments show (S.V. Steinberg & A. Ioudovitch, unpubl.), the euk-tRNA^{Ser}, although somewhat different, easily fits this pattern if its nt G46 replaces G21 (Fig. 4B). Surprisingly, our model of the euk-tRNA^{Sec}, in spite of much greater differences, fits the same structural pattern (Fig. 4A) if its nt U8, G9, A9a, and C20b correspond to nt A22, G20b, G9, and C48, respectively, in the *T. thermophilus* tRNA^{Ser}. Thus, the overall structure of the core region in the presented model of the euk-tRNA^{Sec} is very similar to that found in the *T. thermophilus* tRNA^{Ser} and suggested for the euk-tRNA^{Ser}, even though it is built of the elements taken from different parts of the tRNA nucleotide sequence.

Comparison of the euk-tRNA^{Sec} model with the known structure of *T. thermophilus* tRNA^{Ser} reveals complex interrelations between different elements of the tRNA structure. In particular, G15 pairs with either C20b or C48. Inability to form any of these pairs is expected to affect the tRNA function. This can explain that, as a rule, the tRNA^{Ser} mutants deprived of C48 were poorly serylated (Wu & Gross, 1994, 1993; Amberg et al., 1996). Also, the specific conformation of the four-nucleotide Connector 1 in the model of the tRNA^{Sec} strongly depends on the presence of the six-base pair D-stem. We argue below that the stability of the D-stem is also influenced by Connector 1, which is critical for some mutants when the base pairing is not perfect.

From euk-tRNA^{Ser} to euk-tRNA^{Sec}

The analogies revealed between the presented model for the euk-tRNA^{Sec} and the structure of the tRNA^{Ser} show that the role of each element in a tRNA structure is understandable only in the context of other elements. This helped explain the behavior of the mutant euk-tRNAs^{Ser} that harbored different combinations of four complex mutations AA, D, T, and E (Amberg et al., 1996). These mutations, being introduced together into the euk-tRNA^{Ser}, enabled it to be effectively serylated, selenylated, and phosphorylated. They included insertions of AU just before the D-domain and of CU between the T- and acceptor stems (mutation AA), a double mutation [U20b → C; A21 → U] facilitating the formation of the six-base pair D-stem (mutation D), a deletion of pair G53–C61 from the T-stem (mutation T), and a double mutation [U44 → (AGC); C48 → A] (mutation E). There are 16 possible combinations of these four mutations, ranging from $AA^-D^-T^-E^-$ (wt euk-RNA^{Ser}) to $AA^+D^+T^+E^+$ (mutant Y23). The majority of these combinations have been studied experimentally (Amberg et al., 1996).

Our analysis was based on the following factors, of which the first two were discussed in the previous section, whereas the others were discussed elsewhere.

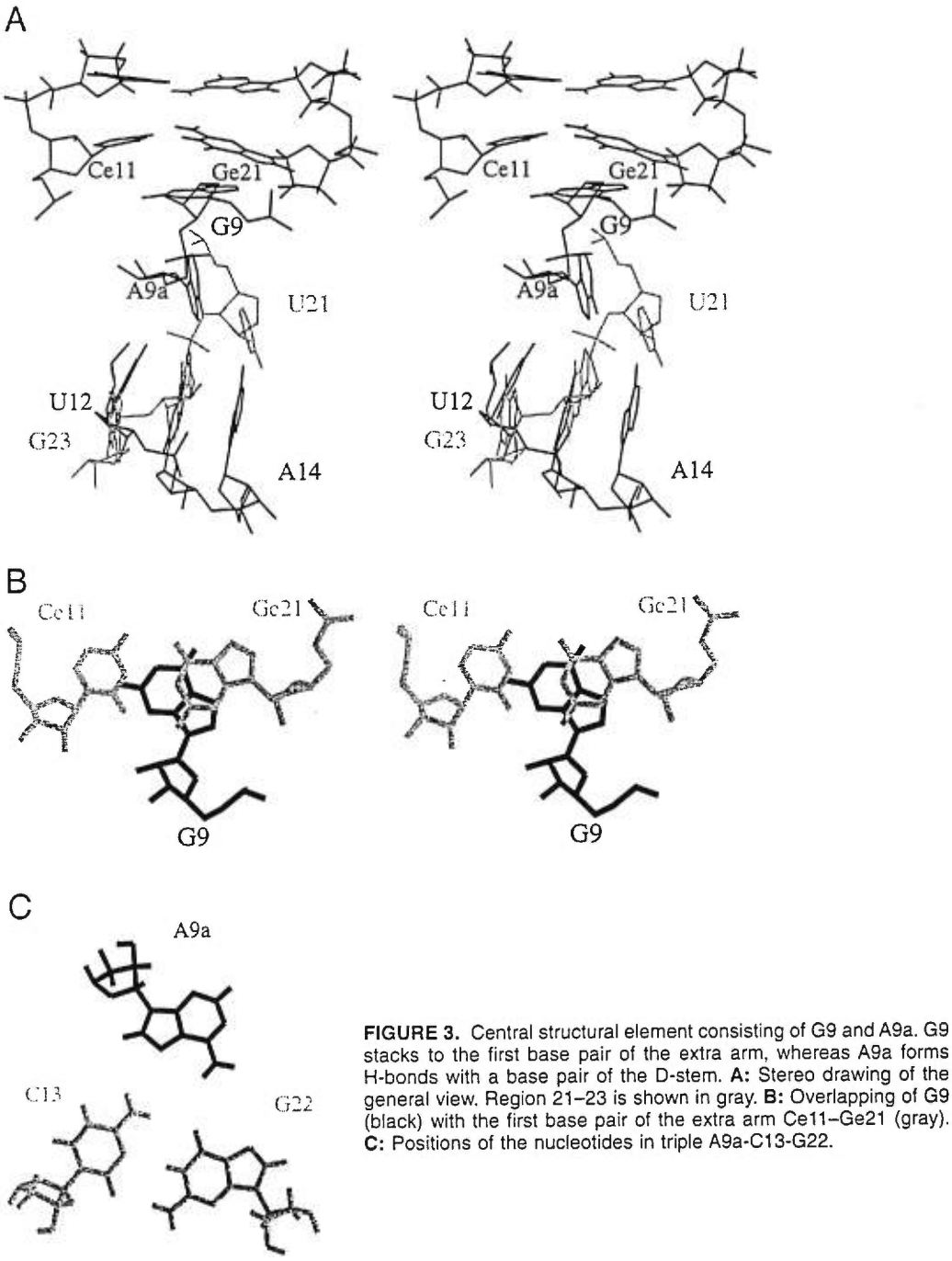


FIGURE 3. Central structural element consisting of G9 and A9a. G9 stacks to the first base pair of the extra arm, whereas A9a forms H-bonds with a base pair of the D-stem. **A:** Stereo drawing of the general view. Region 21–23 is shown in gray. **B:** Overlapping of G9 (black) with the first base pair of the extra arm Ce11–Ge21 (gray). **C:** Positions of the nucleotides in triple A9a-C13-G22.

1. Inability of G15 to form a pair with either C20b or :48 renders a mutant tRNA nonfunctional. Therefore, the ınctionality of the mutants with less than six base pairs ı the D-stem depends on the presence of C48. The ıtter exists only in the mutants with the E^- genotype.

2. Mutation D does not change pair G13–A22, which, ɔgether with the two-nucleotide Connector 1, is essenal for the $tRNA^{Ser}$-specific tertiary interactions. We ssume therefore, that mutation D provides a six-base pair D-stem only in the presence of a four-nucleotide Connector 1, which comes with mutation AA.

3. Inability to form six base pairs in the D-stem eliminates selenylation and phosphorylation and does not affect serylation (Wu & Gross, 1994; Amberg et al., 1996).

4. The two-nucleotide bulge between the acceptor and T-stems, which is a result of mutation AA, damages serylation without affecting either selenylation of

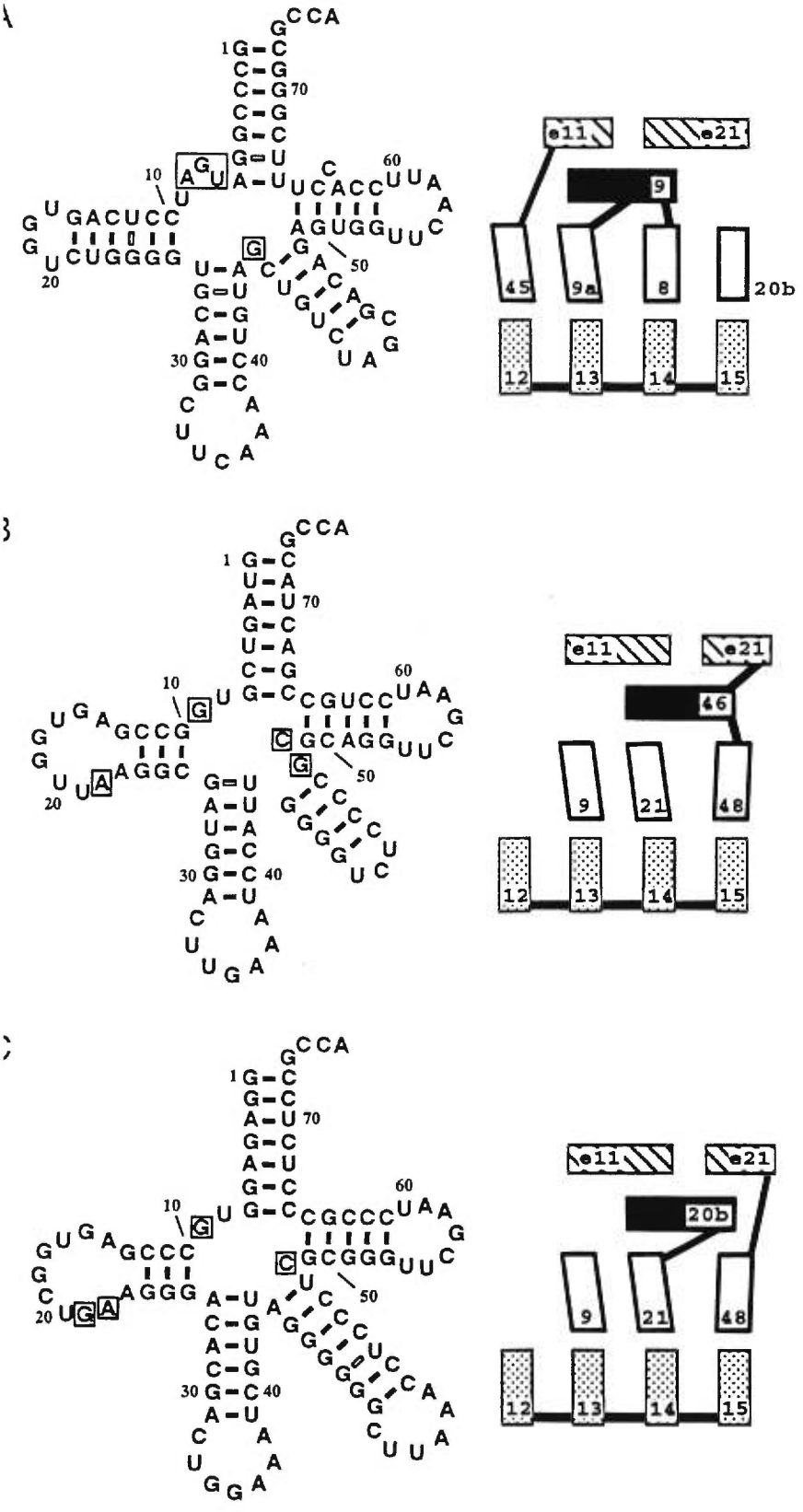


FIGURE 4. Comparison of the cloverleaf secondary structures and of tertiary nucleotide arrangements in (**A**) euk-tRNA^{Sec}, (**B**) euk-tRNA^{Ser}, and (**C**) prok-tRNA^{Ser}. In the cloverleafs, tertiary nucleotides involved in the tertiary areas are boxed. In the tertiary nucleotide arrangements, nucleotides that are neighbors in the polynucleotide chain are connected. Arrangements A and C correspond to the model of the euk-tRNA^{Sec} (A) and to the X-ray conformation of the *T. thermophilus* tRNA^{Ser} (C, Biou et al., 1994). Arrangement B was deduced from C based on the comparison of the tRNA nucleotide sequences and molecular modeling experiments (see text). In all structures, a guanine (G9 in A, G46 in B, and G20b in C, shown in black) stacks to the first base pair e11–e21 (cross-hatched) of the extra arm. Other tertiary nucleotides (white) form H-bonds with the base pairs of the D-domain (stippled) while stacking to each other. This stack consists of three (B, C) or four (A) layers and starts from the nucleotide interacting with G15 (C20b in A and C48 in B, C). C20b in tRNAs^{Sec} (A) is part of the D-stem, whereas the corresponding nt C48 in the tRNAs^{Ser} (B, C) belongs to the variable region.

TABLE 2. Correspondence between the predicted and experimentally determined activities of different euk-tRNA^{Ser} mutants.[a]

Genotype				Phenotype						
				Prediction			Experimental data			
AA	D	T	E	S	L	P	S	L	P	Mutant
–	–	–	–	+	–	–	+	–	–	wt tRNA^{Ser}
–	–	–	+	–	–	–				
–	–	+	–	–	–	–				
–	–	+	+	–	–	–				
–	+	–	–	+	–	–	+	–	–	Y8
–	+	–	+	–	–	–				
–	+	+	–	–	–	–	–	–	–	Y11
–	+	+	+	–	–	–	–	–	+	Y15
+	–	–	–	–	–	–	–	–	–	Y22
+	–	–	+	–	–	–				
+	–	+	–	–	–	–	–	–	–	Y21
+	–	+	+	–	–	–				
+	+	–	–	–	+	+	–	+	+	Y8I
+	+	–	+	–	+	+				
+	+	+	–	–	+	+	–	+	+	Y11H
+	+	+	+	+	+	+	+	+	+	Y23

[a] AA, D, T, and E designate the complex mutations described in the text (Amberg et al., 1996). S, L, and P stand for the ability of a mutant to be serylated, selenylated, and phosphorylated, respectively. Based on the five factors discussed in the text, the phenotypes of the 16 tRNAs (the wt euk-tRNA^{Ser} and 15 mutants) were predicted. In all nine cases where the phenotypes had been determined experimentally (Amberg et al., 1996), they corresponded to the predictions.

hosphorylation (see #12 in Steinberg et al., 1998). ormation of this bulge could be suppressed by two dditional mutations T and E, working in concert.

5. Mutation T per se impairs the normal interaction etween the D and T-loop, thus rendering a tRNA non- ınctional (see Steinberg et al., 1997; and #1, 10, 11 in ;teinberg et al., 1998). This effect can be suppressed y mutation AA.

Based on these considerations, it was possible to redict the ability of the clones to be serylated, sele- ylated, or phosphorylated. The results of the analysis resented in Table 2 show a very good correspon- ence with the existing experimental data, which sup- orts both the model of the euk-tRNA^{Sec}, and the uggested relationships between its different elements.

ıCKNOWLEDGMENTS

his work was supported by a grant from the Natural Science nd Engineering Council of Canada to Robert Cedergren. Ve are grateful to Robert Cedergren for numerous discus- ions, reading the manuscript, and moral support.

Received December 3, 1997; returned for revision December 31, 1997; revised manuscript received January 8, 1998

REFERENCES

ımberg R, Takaharu M, Wu XQ, Gross HJ. 1996. Selenocysteine synthesis in mammalia: An identity of switch from tRNA^{Ser} to tRNA^{Sec}. *J Mol Biol 263*:8–19.

Asahara H, Himeno H, Tamura K, Nameki N, Hasegawa T, Shimizu M. 1994. *Escherichia coli* seryl-tRNA synthetase recognizes tRNA^{Ser} by its characteristic tertiary structure. *J Mol Biol 236*:738–748.

Biou V, Yaremchuk A, Tukalo M, Cusak S. 1994. The 2.9 Å crystal structure of *T. thermophilus* seryl-tRNA synthetase complexed with tRNA^{Ser}. *Science 263*:1404–1410.

Böck A, Forchhammer K, Heider J, Baron C. 1991. Selenoprotein synthesis: An expansion of the genetic code. *Trends Biochem Sci 16*:463–467.

Cate JH et al. 1996. Crystal structure of a group I ribozyme domain: Principles of RNA packing. *Science 273*:1678–1685.

Diamond A, Dudock B, Hatfield D. 1981. Structure and properties of a bovine liver UGA suppressor serine tRNA with a tryptophane anticodon. *Cell 25*:497–506.

Himeno H, Hasegawa T, Ueda T, Watanabe K, Shimizu M. 1990. Conversion of aminoacylation specificity from tRNA^{Tyr} to tRNA^{Ser} in vitro. *Nucleic Acids Res 18*:6815–6819.

Ladner JE, Jack A, Robertus JD, Brown RS, Rhodes D, Clark BFC, Klug A. 1975. Structure of yeast phenylalanine transfer RNA at 2.5 Å resolution. *Proc Natl Acad Sci USA 72*:4414–4418.

Moras D, Comarond MB, Fisher J, Weiss R, Thiery JC, Ebel JP, Giege R. 1980. Crystal structure of yeast tRNA^{Asp}. *Nature (Lond) 288*:669–674.

Ohama T, Yang DCH, Hatfield DL. 1994. Selenocysteine tRNA and serine tRNA are aminoacylated by the same synthetase, but may manifest different identities with respect to the long extra arm. *Arch Biochem Biophys 315*:293–301.

Pearlman DA et al. 1995. *AMBER 4.1*. San Francisco: University of California.

Pley HW, Flaherty KM, McKay DB. 1994. Three-dimensional structure of a hammerhead ribozyme. *Nature 372*:68–74.

Quigley GJ, Seeman NC, Wang AHJ, Suddath FL, Rich A. 1975. Yeast phenylalanine transfer RNA: Atomic coordinates and torsion angles. *Nucleic Acids Res 4*:1649–1665.

Sprinzl M, Steegborn C, Hübel F, Steinberg S. 1996. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res 24*:68–72.

Stadtman TC. 1990. Selenium biochemistry. *Annu Rev Biochem 59*:111–127.

Steinberg S, Leclerc F, Cedergren R. 1997. Structural rules and con-

formational compensations in the tRNA L-form. *J Mol Biol 266*:269–282.

Steinberg SV, Ioudovitch A, Cedergren R. 1998. The secondary structure of eukaryotic selenocysteine tRNA: 7/5 versus 9/4. *RNA 4*:241–245.

Sturchler C, Westhof E, Carbon P, Krol A. 1993. Unique secondary and tertiary features of the eukaryotic selenocysteine $tRNA^{Sec}$. *Nucleic Acids Res 21*:1073–1079.

Wu XQ, Gross HJ. 1993. The long extra arms of human $tRNA^{(Ser)Sec}$ and $tRNA^{Ser}$ function as major identity elements for serylation in an orientation-dependent, but not sequence specific manner. *Nucleic Acids Res 21*:5589–5594.

Wu XQ, Gross HJ. 1994. The length and the secondary structure of the D-stem of human selenocyteine tRNA are the major identity determinants for serine phosphorylation. *EMBO J 13*:241–248.

Zinoni F, Birkmann A, Leinfelder W, Böck A. 1987. Cotranslational insertion of selenocysteine into formate dehydrogenase from *Escherichia coli* directed by a UGA codon. *Proc Natl Acad Sci USA 84*:3156–3160.

Chapter V

A role for the bulged nucleotide 47 in the facilitation of tertiary interactions in the tRNA structure

Sergey Steinberg and Anatoli Ioudovitch[3]

RNA 1996, **2**, 84-87

[3] Contributions by each author:
Sergey Steinberg – sequence analysis and general supervision
Anatoli Ioudovitch – modeling experiments

/A (1996), 2:84-87. Cambridge University Press. Printed in the USA.

ı role for the bulged nucleotide 47 in the facilitation of ertiary interactions in the tRNA structure

ERGEY STEINBERG[1] AND ANATOLI IOUDOVITCH
partement de Biochimie, Université de Montréal, Montréal, Quebéc H3C 3J7, Canada

BSTRACT

ased on computer modeling and with the use of energy minimisation procedure, we show that the bulged ucleotide 47 in the yeast tRNA^{Phe} structure plays an important steric role, allowing the formation of canonil tertiary interactions 15-48 and 22-46 within the D-domain. The absence of nucleotide 47 can be compented by the presence of a wobble pair U13-G22, whose unusual stereochemistry permits as well the formation the canonical tertiary interactions. The tRNA database shows that the vast majority of the cytosolic tRNAs ave either a nucleotide at position 47 or a wobble pair U13-G22. On the contrary, many mitochondrial tRNAs, aving a Watson-Crick pair 13-22, do not have a nucleotide in position 47, which suggests that their tertiary teractions within the D-domain must differ from those in cytosolic tRNAs.

eywords: tRNA; tRNA structure; nucleic acids conformation; models, molecular

ITRODUCTION

ne uncovering of unsuspected structural motifs in bioolymers can lead to the revelation of new sequence orrelations that, in turn, leads to a deeper understandg of the correspondence between primary and tertiary ructure. Here we present an example of structural orrelation between nt 47 from the variable loop region tRNA and the nature of pair 13-22 from the D-stem. ucleotide 47 has not as yet been considered essential r tRNA structure, however, in this article, we argue at it is called upon to play a crucial role in the the foration of intermolecular tertiary interactions.

The supposedly banal role of nt 47 in tRNA could be educed from the following observations. (1) Nucleode 47 is not present in all tRNAs (Steinberg et al., 993; for examples, see Fig. 1). (2) In all known X-ray ructures of tRNAs containing nt 47, including the yeast RNA^{Phe} (Ladner et al., 1975; Quigley et al., 1975; Woo al., 1980; Rould et al., 1989), this nucleotide is bulged ut and does not participate in interactions with other arts of the molecule. (3). In the yeast tRNA^{Asp}, abence of nt 47 is responsible for only minor conformaonal differences from the yeast tRNA^{Phe} (Moras et al., 1980, 1986), whereas the general scheme of nucleotide-nucleotide contacts in the both molecules is essentially the same.

RESULTS AND DISCUSSION

If nt 47 does not influence the tRNA structure and function, it is not clear why it is present in the majority of cytosolic tRNAs (Steinberg et al., 1993). To elucidate this question, we decided to determine whether indeed the loss of this nucleotide leads to any serious consequences for the RNA structure. We initiated a modeling study of the yeast tRNA^{Phe} in which nt 47 had been deleted. Atomic coordinates of the yeast tRNA^{Phe} X-ray conformation were taken, from which nt 47 was removed. We then tried to connect nt 46 and 48; however, the distance between nt 46 and 48 in the tRNA^{Phe} was larger than the distance that could be spanned by a phosphodiester bond. Keeping in mind that in the tRNA^{Asp} these nucleotides are normally connected, we started displacing nt 46 and 48 as well as some of their neighbors from the D-domain in order to achieve a satisfactory connection.

Surprisingly, we found that there was no way to make this connection, because, when nt 46 and 48 were arranged as in the tRNA^{Asp}, nt 46 seriously collided with nt 22 (Fig. 2). Any attempt to use energy minimization in order to avoid this collision and simultaneously preserve the connection between nt 46 and 48

Reprint requests to: Sergey Steinberg, Département de Biochimie, niversité de Montréal, Montréal, Quebéc H3C 3J7, Canada; e-mail iteinbe@medcn.umontreal.ca.
[1] On leave from Engelhardt Institute of Molecular Biology, 32 ul avilova, 117984 Moscow, Russia.

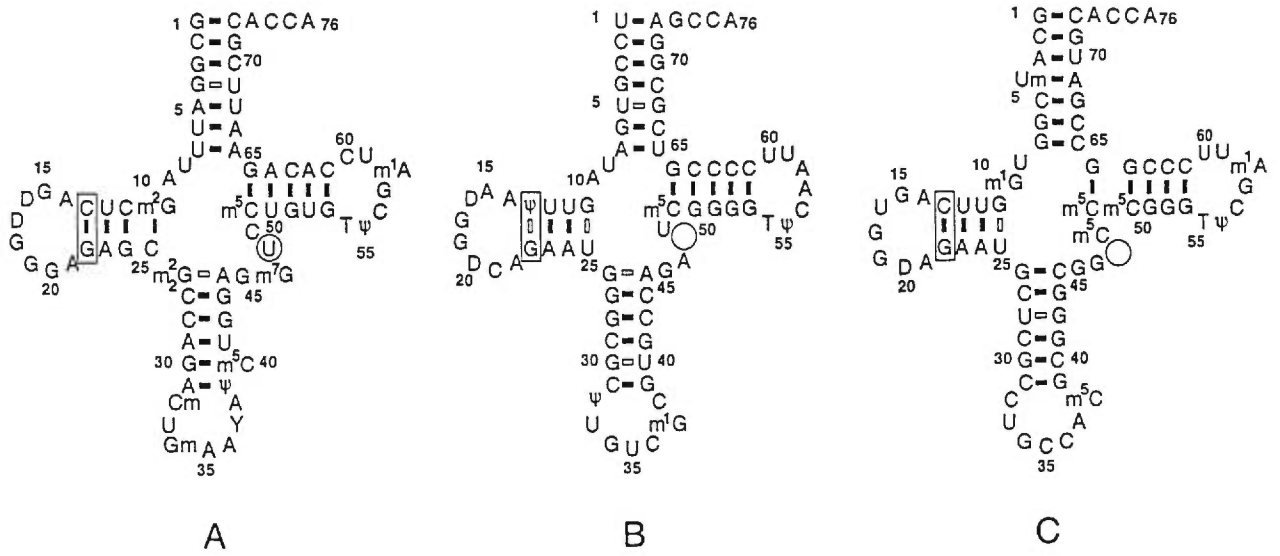


FIGURE 1. Nucleotide sequence of (**A**) yeast tRNA^{Phe} (ID number in the tRNA Compilation RF6280), (**B**) yeast tRNA^{Asp} (RD6280), and (**C**) *B. mori* tRNA^{Gly} (RG7680) folded into the cloverleaf secondary structure. Pair 13-22, whose identity is C-G in the tRNA^{Phe} and tRNA^{Gly} and Ψ in the tRNA^{Asp}, is boxed. Position 47, which is occupied by D in the tRNA^{Phe} and is empty in the tRNA^{Asp} and tRNA^{Gly}, is circled. The tRNA^{Gly} does not have nt 47 in spite of the fact that it contains a Watson-Crick pair 13-22. This does not allow the formation of the standard tertiary interaction 22-46 (see text). The same situation occurs in tRNAs^{Gly} RG1310, RG1380, RG1381, RG1660, RG1670, DG1820, DG7740, RG9991.

resulted in disruption of at least one of the important secondary and tertiary interactions 13-22, 22-46, or 15-48, depending on the strategy of the energy minimization. In other words, the deletion of nt 47 of the tRNA^{Phe} led to major structural perturbations not observed in the tRNA^{Asp}.

The failure to form a 46–48 connection suggests the possibility that the D-domain in the tRNA^{Asp} is not simply a version of the D-domain found in the tRNA^{Phe}. We reasoned that there must be some aspect of the tRNA^{Asp}, absent in the tRNA^{Phe}, that allows the tRNA^{Asp} to form the proper connection between nt 46 and 48 while maintaining all the secondary and tertiary interactions within the D-domain. The comparison of the nucleotide sequences of both tRNAs showed that the absence of nt 47 in the tRNA^{Asp} is not the only difference in this region. In particular, we noticed that, although position 13 in the tRNA^{Phe} is occupied by a C, the tRNA^{Asp} contains Ψ in the same position. The resulting Ψ13-G22 base pair in tRNA^{Asp}, formed in the same way as a U-G wobble pair, has the effect of shifting the purine about 2 Å toward the minor groove in comparison to its position in the Watson–Crick pair C13-G22 of the tRNA^{Phe}. Figure 2 shows that it is precisely this shift that helps to avoid the collision of nt 22 and 46 and thus allows the formation of the tRNA^{Phe}-like tertiary interaction pattern.

If this tertiary interaction is essential for tRNA function, we would expect that tRNAs, which have a similar pattern of secondary and tertiary interactions in this region based on their nucleotide sequences, should contain either a 13U-22G pair or a nucleotide at position 47. Indeed, our screening of the tRNA Compilation (Steinberg et al., 1993) showed that this rule is satisfied in the vast majority of the cytosolic tRNAs having all other potential to form the tRNA^{Phe}-like tertiary interaction pattern (Table 1). Although the absence of nt 47 generally correlates with pair U13-G22 in these tRNAs, only 9 of 444 cytosolic tRNAs having all other potentials to form the tRNA^{Phe}-like tertiary interaction pattern fail to obey this rule. Although the

TABLE 1. Occurrence of Watson–Crick (WC) and U-G pairs 13-22 and nt 47 in cytosolic and mitochondrial tRNAs.[a]

		Nucleotide 47	
	Pair 13-22	+	−
Cytosolic TRNAs	WC	361	9
	U-G	17	57
Mitochondrial	WC	103	180
tRNAs	U-G	25	57

[a] tRNAs sequences were selected from the tRNA Compilation (Steinberg et al., 1993) based on their ability to accept the tRNA^{Phe}-like pattern of tertiary interactions. Sequences containing features responsible for formation of alternate tRNA^{Gln} (Rould et al., 1989) or tRNA^{Ser}-like (Biou et al., 1994) tertiary interaction patterns, or disfavoring the tRNA^{Phe}-like pattern, i.e., a long extra arm, either non-Watson–Crick combination 15-48 or G9-G23; neither Watson–Crick nor U-G pair 13-22; no purine in either position 22 or 46 were removed from the analysis. If in the tRNA Compilation a sequence existed both as that of the gene and that of the mature tRNA, only one of them was taken for the statistics.

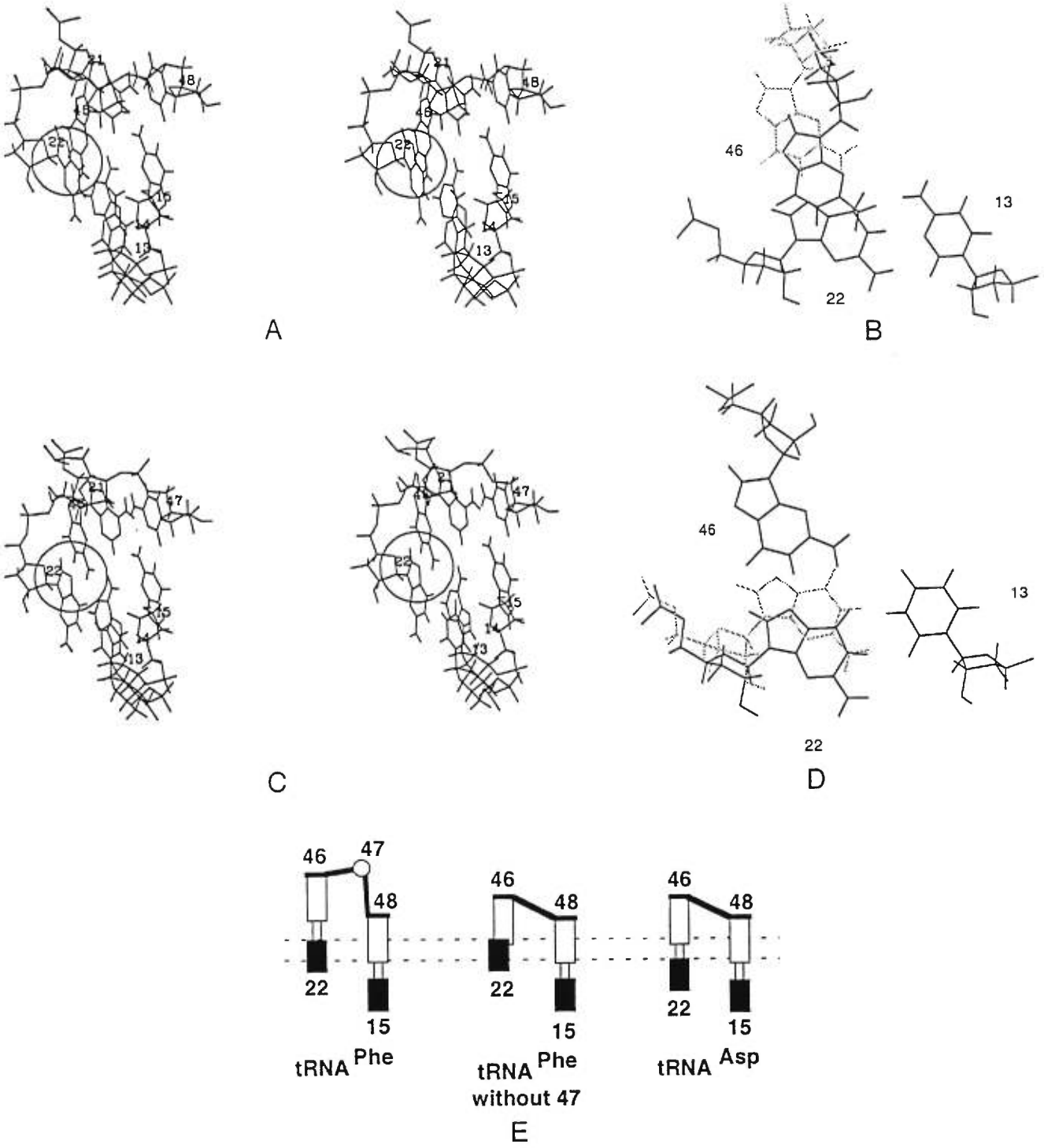


FIGURE 2. Standard tertiary interaction 22–46 cannot be formed in a tRNA with a Watson–Crick pair 13-22 and without nt 47. **A,C:** Stereoviews of nt 13–15, 21–22, and 46–48 of a tRNA with pair C13-G22 (A) or U13-G22 (C). The region of contact, 22–46, is circled. **B,D:** Mutual positions of nt 13, 22, and 46 in a tRNA with pair C13-G22 (B) or U13-G22 (D). The position of nt 46 is shown as dotted in a tRNA with nt 47 (B) and nt 22 in a tRNA with C13-G22 (D). In a tRNA with pair C13-G22, only the presence of nt 47 makes the interaction 22–46 possible, whereas without nt 47, nt 22 and 46 collide with each other (A,B). In a tRNA with pair U13-G22, nt G22 shifts in the direction of the minor groove, thus avoiding the collision (C,D) even if nt 47 is absent. Here the case of G22-G46 combination is presented. The cases of combinations G22-A46 and A22-A46, which also occur in tRNAs, provide essentially the same result (not shown). **E:** Schematic representation of the positions of nt 13, 22, 46, and 48 in the tRNA with nt 47 and a Watson–Crick pair 13-22 as in the $tRNA^{Phe}$ (left); without nt 47 and with a Watson–Crick 13-22 pair (center); and without nt 47 and with pair 13U-22G as in the $tRNA^{Asp}$ (right). Positions of nt 15 and 48 are shown the same in all three cases. Dashed lines between nt 22-46 and 15-48 represent the internucleotide H-bonds. Horizontal lines represent the upper level of nt 22 in the tRNA with a Watson–Crick (upper) and U-G (lower) pair 13-22. Deletion of nt 47 makes the connection 46–48 shorter, which, in turn, forces nt 46 to shift and collide with nt 22 (center). This collision can be avoided by shifting of nt 22 due to pair U13-G22, as in the $tRNA^{Asp}$ (right).

ceptions represent only a small portion of all tRNAs, ey are worthy of more detailed consideration. Deite the fact that the exceptions are tRNAs from such olutionarily distant organisms as eubacteria, lower d higher eukaryotes; surprisingly, all have a glycine ecificity (see the legend to Fig. 1C). Further analysis owed that two of these exceptional $tRNA^{Gly}$ from *aphylococcus epidermidis* (ID numbers in the tRNA ompilation RG1380 and RG1381) are known not to be volved in the ribosome-dependant protein synthesis, t rather in the synthesis of peptidoglycans (Stewart al., 1971; Roberts et al., 1973). Moreover, there is direct evidence that at least some of the other exceptnal tRNAs do not function in the protein biosynthe- either. In particular, the $tRNA^{Gly}$ from *Bombix mori* D number RG7680, anticodon GCC, see Fig. 1C) was t able to bind to the ribosome charged with any of ycine codons when tested under physiological conions, and further displayed an abnormal wobble patn at higher concentrations of Mg^{2+} (Kawakami et al., 80). We hypothesize, therefore, that these exceptnal $tRNAs^{Gly}$ represent a family of tRNA-like molules not involved in protein synthesis and that they rform an auxiliary function. Moreover, we suggest at their inability to interact with the ribosome reflects eir inability to form the standard tertiary interaction ttern within the D-domain. We also state that some her $tRNAs^{Gly}$ involved in the delivery of glycine to e cell wall have other unusual aspects in the D-stem, ch as a non-Watson–Crick combination 15-48 (Gaian et al., 1991).

Contrary to cytosolic tRNAs, almost half of the 365 itochondrial tRNAs contain the unusual motif of a -22 Watson–Crick pair without nt 47 (Table 1). Beuse there is no doubt about involvement of these NAs in protein synthesis, we conclude that tRNA nctionality in mitochondria may not depend on the rmation of the standard tertiary interactions within e D-domain. This suggestion is hardly surprising, owever, because it is well-known that some mitoondrial tRNAs have even more bizarre structures d are still able to perform their function (Steinberg Cedergren, 1994; Dirheimer et al., 1995).

The example presented here shows that, even if all e nucleotides that are involved in formation of terty interactions in an RNA molecule are present in the quence, steric factors may render some interactions possible. Auxiliary elements, such as bulged nucleides not interacting with the rest of the molecule, can ay an important structural role, allowing the formaon of other interactions. We suggest also a direct link etween the inability of some cytosolic tRNAs to form the standard tertiary interaction pattern and their inability to interact with the ribosome in the normal way. A particular question arises about "exceptional" cytosolic and mitochondrial tRNAs that have no nt 47 even though their pair 13-22 has a Watson–Crick type. The tertiary structures these tRNAs and their inability to form the important tertiary interactions is a matter for further theoretical and experimental analysis that is now being performed in our laboratory.

ACKNOWLEDGMENTS

We are grateful to Dr. R. Cedergren for reading the manuscript and for fruitful discussions. This work was supported by grant from NSERC of Canada to R. Cedergren.

Received November 3, 1995; returned for revision November 28, 1995; revised manuscript received December 1, 1995

REFERENCES

Biou V, Yaremchuk A, Tukalo M, Cusack S. 1994. The 2.9 Å crystal structure of *T. thermophilus* seryl-tRNA synthetase complexed with $tRNA^{Ser}$. *Science 263*:1404–1410.

Dirheimer G, Keith G, Dumas P, Westhof. 1995. Primary, secondary, and tertiary structures of tRNAs. In: Söll D, RajBhandary, U. eds. *tRNA: Structure, biosynthesis, and function*. Washington DC: ASM Press. pp 93–126.

Gamian A, Krzyzaniak A, Barciszewska MZ, Gawronska I, Barciszewski J. 1991. Specific incorporation of glycine into bacterial lipopolysaccharide. Novel function of specific transfer ribonucleic acids. *Nucleic Acids Res 19*:6021–6025.

Kawakami M, Tsonis PA, Nishio K, Takemura S. 1980. Abnormal codon recognition of glycyl-tRNA from the posterior silk glands of *Bombix mori*. *J Biochem (Tokyo) 88*:1151–1157.

Ladner JE, Jack A, Robertus JD, Brown RS, Rhodes D, Clark BFC, Klug A. 1975. Structure of yeast phenylalanine transfer RNA at 2.5 Å resolution. *Proc Natl Acad Sci USA 72*:4414–4418.

Moras D, Comarmond MB, Fisher J, Weiss R, Thiery JC, Ebel JP, Giege R. 1980. Crystal structure of yeast $tRNA^{Asp}$. *Nature (Lond) 288*:669–674.

Moras D, Dumas P, Westhof E. 1986. GU base pairs and variable loop in yeast $tRNA^{Asp}$. In: van Knippenberg PH, Hilbers CW, eds. *Structure and dynamics of RNA*. NATO ASI Series, Series A: Life Science Vol. 110. New York/London: Plenum Press. pp 113–124.

Quigley GJ, Seeman NC, Wang AHJ, Suddath FL, Rich A. 1975. Yeast phenylalanine tranfer RNA: Atomic coordinates and torsion angles. *Nucleic Acids Res 4*:1649–1665.

Roberts RJ. 1973. Staphylococcal transfer ribonucleic acids. *J Biol Chem 249*:4787–4796.

Rould MA, Perona JJ, Söll D, Steitz TA. 1989. Structure of *E. coli* glutaminyl-tRNA sythetase complexed with $tRNA^{Gln}$ and ATP at 2.8 Å resolution. *Science 246*:1135–1142.

Steinberg S, Cedergren R. 1994. Structural compensation in atypical mitochondrial tRNAs. *Nature Struct Biol 1*:507–511.

Steinberg S, Misch A, Sprinzl M. 1993. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res 21*:3011–3015.

Stewart TS, Roberts RJ, Strominger JL. 1971. Novel species of tRNA. *Nature (Lond) 230*:36–38.

Woo NH, Roe BA, Rich A. 1980. Three-dimensional structure of *E. coli* initiator $tRNA^{fMet}$. *Nature (Lond) 286*:346–351.

Discussion

tRNA database problems and perspective.

The "Compilation of the tRNA sequences and sequences of tRNA genes" is a result of the collaborative effort to create a comprehensive database of tRNA sequences aligned by their structural properties.

A comparative analysis of the sequences within this database will help better understanding of sequence-structure correlations. In addition, this analysis is expected to reveal inconsistencies and mistakes in the database. Although there are different sources of mistakes in the compilation, the most difficult ones to catch are errors in sequencing. The long extra arm of the $tRNA^{Tyr}$ from *Trypanosoma brucei* is an example of such an error, which was found only recently. Usually cytoplasmic eukaryotic $tRNAs^{Tyr}$ do not have a long extra arm; the only exception seen in the Compilation is this tRNA from *T. brucei*. However, if one takes into account the ten nucleotide intron in the anticodon loop, this tRNA "loses" additional nucleotides in the extra-arm and becomes a normal Class I tRNA. Such mistakes, if not corrected, pose a serious problem for the database analysis. Thus, verification and correction of mistakes in the data and ensuring the completeness of the database becomes a very important part of the analysis of the tertiary interactions and of structural motifs in general.

Progress in studies of the role of tertiary interactions will probably affect the presentation of the sequences in the future databases. Indexing and classification of the sequences according to certain established structural features and principles, such as a covariation between the absence of nucleotide 47 and the presence of non-canonical pair U13-G22, will provide additional information useful in experimental design and eventually leading to a better understanding of the tRNA structure.

Linker effect in RNA structure

As it has been shown here, even nucleotides, which are not directly involved in any contact as nucleotide 47, can influence dramatically the system of important tertiary

interactions. The identity of nucleotide 47 is relatively unimportant, except that it should not be able to participate in "parasitic" interactions such as intercalation between nucleotides of Connector I and II or base pairing within the D-stem. This idea gets a support from the fact that in many tRNAs the identity of nucleotide 47 is either a dihydrouridine or a uridine modified at C1. Dihydrouridine is unable to participate in any stacking interactions, while the modification C1 deprives the uridine of any base pairing abilities.

Our results show that the tertiary interactions 22-46 and 15-48 are essential for the tRNA structure, however, it is not absolutely clear yet how exactly the presence of these interactions affect the tRNA functional cycle. The primary function of tRNA, i.e. delivering of the amino acid to the ribosome and its incorporation into the nascent peptide in response to a given codon, does not seem to be affected by the absence of these interactions (Cermakian *et al.*, 1997). This, however, does not exclude the possibility that the increased flexibility of the molecule due to the absence of these interactions can result in a higher level of miscoding or can make it more susceptible to cellular ribonucleases. The cells containing tRNAs with disrupted tertiary interactions will lose, in the long run, to those that maintain them, as one can judge from the analysis of the tRNA compilation (Chapter V) and of the available experimental data (Cermakian *et al.*, 1997).

For the general RNA architecture, the case of nucleotide 47 is an interesting example of the "linker" effect. The nature of the linker is not very important, while its length, and, in extreme cases, its presence or absence affects the global molecular structure. A somewhat similar case is associated with the "double zipper" covariation mentioned in the Introduction. In mitochondria, the shortening of Domain I is usually compensated by intercalation of nucleotides of Connector 2 between the last stacking layer of the D-domain and nucleotide 59 of the T-stem (Steinberg *et al.*, 1997). In this situation Connector 1 is required to be long enough to guarantee a proper connection between the D and T-stems.

Structure-function relationship and tRNA architecture

From the inability of the $tRNA^{Sec}$ to fit into the standard secondary structure one can conclude that in general terms, the functionality of adapter molecules in the ribosome-dependant protein biosynthesis is not directly associated with particular elements of the secondary structure. Instead, it deals with the conservation of the canonical L-shaped architecture. The presented arguments in favor of the 7/5 secondary structure of the eukaryotic $tRNA^{Sec}$ and the discovered mutual compensation between the shortened T-stem and the enlarged D-stem in the archaeal $tRNA^{Sec}$ show the role played by the D/T tertiary interactions n the maintenance of the normal L-shape. However, the archaeal $tRNA^{Sec}$ has unprecedented nine base pairs in the acceptor stem. Such a long acceptor stem may not be tolerated in the transpeptidation for which the exact positions of the acceptor termini of both tRNAs are crucial. Following this logic, we can argue that at this step of translation the number of base pairs in the acceptor stem should be the same as in the normal cytosolic tRNAs. It was hypothesized in Chapter IV that the base pairs in the acceptor stem adjacent to the T-stem are probably the first to be sacrificed for the sake of the proper size. However, an experimental study is needed to prove this hypothesis.

The situation with the secondary structure of the $tRNAs^{Sec}$ highlights an interesting problem in the structural analysis. Because the functional pattern of different $tRNAs^{Sec}$ is very similar, it would be reasonable to expect a strong structural similarity among them. Thus, the ability of the eukaryotic $tRNA^{Sec}$ to have almost the same 9/4 secondary structure as the archaeal $tRNA^{Sec}$ has, could be considered as an argument in favor of the 9/4 structure for both molecules. This argument could be strengthened even more by the fact that Domain II in the 9/4 structure consists of thirteen base pairs, the same number of base pairs as in the 8/5 secondary structure of the prokaryotic $tRNA^{Sec}$. Thus, based only on the analysis of the secondary structures of the $tRNAs^{Sec}$ one could favor the 9/4 secondary structure for the eukaryotic $tRNA^{Sec}$. This chain of logic looks persuasive until we take into account the interactions in which the T-loop is normally involved. A four base pair T-stem *per se* does not provide for the normal D/T interactions

and needed such a special compensation as an extended D-stem. This is the case for the archaeal $tRNA^{Sec}$, and it does not happen in the eukaryotic $tRNA^{Sec}$. Thus, it is very difficult to draw the line between the conservative sequence patterns and "so-called" exceptions from them because they both can satisfy the same constraints if the structure is considered at a more detailed level.

Motifs, which include tertiary interactions

Tertiary interactions are very important part of the RNA architecture. They help arrange properly double helical regions and provide an essential rigidity for the whole structure. They may constitute binding and recognition sites for different ligands and form active sites for some biochemical processes.

Although the secondary structure of the core region in the eukaryotic $tRNA^{Sec}$ differs significantly from that in the other Class II tRNAs, important similarities have been observed on the level of tertiary interactions. A group of tertiary nucleotides form a "shed", in which two nucleotides interact with the D-domain by forming "walls", while another nucleotide stacks to the extra arm making a "roof". Such a structure allows a rigid docking of the extra arm to Domain I.

Despite the uniqueness of every RNA molecule, it appears that the tertiary structures of different RNAs are built from a limited number of elements or structural motifs that can be combined together in many unexpected ways. The modeled arrangement of tertiary nucleotides organized as the "shed" structure (Chapter IV) may be one of these elements. It can serve as a docking structure for two perpendicularly oriented helices in other RNAs structures. Clearly, the elucidation of the structural requirements for a particular motif in the transfer RNA will be very useful for understanding the properties of this motif in all other molecules where it can be found.

References

Aboul-Ela F., Karn J. & Varani G. (1995) The structure of the human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by TAT protein. *J. Mol. Biol.*, **253**, 313-332.

Aboul-Ela F., Karn J. & Varani G. (1996) Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. *Nucleic Acid Res.* **24**, 3974-3981.

Achsel T. & Gross H. J. (1993) Identity determinants of human tRNA(Ser): sequence elements necessary for serylation and maturation of a tRNA with a long extra arm. *EMBO J.* **12**, 3333-3338.

Allain F. H. T. & Varani G. (1995) Structure of the P1 helix from group I self-splicing introns. *J. Mol. Biol.* **250**, 333-353.

Altona C. (1996) Classification of nucleic acid junctions. *J. Mol. Biol.* **263**, 568-581.

Arnez J. G. & Steitz T. A. (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. *Biochemistry* **33**, 7560-7567.

Arnott S., Hukins D. W. L., Dover S. D., Fuller W. & Hodgson A. R. (1973) Structures of synthetic polynucleotides in the A-RNA and A'-RNA conformations: X-ray diffraction analyses of the molecular conformations of polyadenylic acid-polyuridylic acid and polyinosinic acid-polycytidylic acid. *J. Mol. Biol.* **81**, 107-122.

Baeyens K. J., de Bondt H. L., Pardi A. & Holbrook S. R. (1996) A curved RNA helix incorporating an interanl loop with G-A and A-A non-Watson-Crick base pairing. *Proc. Natl. Acad. Sci. USA* **93**, 12851-12855.

Basavappa R. & Sigler P. B. (1991) The 3 A crystal structure of yeast initiator tRNA: functional implications in initiator/elongator discrimination. *EMBO J.* **10**, 3105-3110.

Battacharrya A., Murchie A. I. H. & Lilley D. M. J. (1990) RNA bulges and the helical periodicity of double-stranded RNA. *Nature* **343**, 484-487.

Battiste J. L., Mao H., Rao N. S., Tan R., Muhandiram D. R., Kay L. E., Frankel A. D. & Williamson J. R. (1996) Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science* **273**, 1547-1551.

Beier H., Lee M. C., Sekiya T., Kuchino Y. & Nishimura S. (1992) Two nucleotides next to the anticodon of cytoplasmic rat tRNA(Asp) are likely generated by RNA editing. *Nucleic Acid Res.* **20**, 2679-2683.

Biou V., Yaremchuk A. Tukalo M. & Cusak S. (1994) The 2.9Å crystal structure of *T. thermophylus* seryl-tRNA synthetase complexed with $tRNA^{Ser}$. *Science* **263**, 1404-1410.

Breitschopf K., Achsel T., Busch K. & Gross H. J. (1995) Identity elements of human tRNA(Leu): structural requirements for converting human tRNA(Ser) into a leucine acceptor in vitro. *Nucleic Acids Res.* **23**, 3633-3637.

Broitman S. L., Im D. D. & Fresco J.R (1987) Formation of the triple-stranded polynucleotide helix, poly(A.A.U). *Proc. Natl. Acad. Sci. USA* **84**, 5120-5124.

Brown J. W. (1998) The Ribonuclease P database. *Nucleic Acid Res.* **26**, 351-352.

Butcher S. E., Dieckmann T. & Feigon J. (1997a) Solution structure of the conserved 16 S-like ribosomal RNA UGAA tetraloop. *J. Mol. Biol.* **268**, 348-358.

Butcher S. E., Dieckmann T. & Feigon J. (1997b) Solution structure of a GAAA tetraloop receptor RNA. *EMBO J.* **16**, 7490-7499.

Cate J. H., Gooding A. R., Podell E., Zhou K., Golden B. L., Kundrot C. E., Cech T. R. & Doudna J. A. (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **273**, 1678-1685.

Cermakian N., McClain W. H. & Cedergren R. (1998) tRNA nucleotide 47: an evolutionary enigma. *RNA* **4**, 928-936.

Chang K.Y. & Tinoco I. Jr. (1997) The structure of an RNA "kissing" hairpin complex of the HIV TAR hairpin loop and its complement. *J. Mol. Biol.* **269**, 52-66.

Chastain M. & Tinoco I. Jr. (1991) Structural elements in RNA. *Prog. Nucleic Acid Res. Mol. Biol.* **41**, 131-177.

Cheong C & Moore B. P. (1992) Solution structure of an unusually stable RNA tetraplex containing G- and U-quartet structures. *Biochemistry* **31**, 8406-8414.

Correll C. C., Freeborn B., Moore P. B. & Steitz T. A. (1997) Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain. *Cell* **91**, 705-712.

Crain P. F. & McCloskey (1997) The RNA modification database. *Nucleic Acids Res.* **25**, 126-127.

Dallas A. & Moore P. B. (1997) The loop E-loop D region of *Escherichia coli* 5S RNA: the solution structure reveals an unusual loop that may be important for binding ribosomal proteins. *Structure* **5**, 1639-1653.

Davis P.W., Thurmes W. & Tinoco I. Jr. (1993) Structure of a small RNA hairpin. *Nucleic Acid Res.* **21**, 537-545.

Deutscher M. P. (1995) tRNA processing nucleases. In: *tRNA structure, biosynthesis and function*. Ed. Söll D. & RajBhandary U. L., ASM Press, Washington, DC.

Dieckmann T., Suzuki E., Nakamura G. K. & Feigon J. (1996) Solution structure of an ATP-binding aptamer reveals a novel fold. *RNA* **2**, 628-640.

Dietrich A., Weil J. H. & Marechal-Drouard L. (1992) Nuclear-encoded transfer RNAs in plant mitochondria. *Annu. Rev. Cell Biol.* **8**, 115-131.

Dock-Bregeon A. C. , Westhof E., Giege R. & Moras D. (1989) Solution structure of a tRNA with a large variable region: yeast $tRNA^{Ser}$. *J. Mol. Biol.* **206**, 707-722.

Du Z., Giedroc D. P. & Hoffman D. W. (1996) Structure of the autoregulatory pseudoknot within the gene 32 messenger RNA of bacteriophages T2 and T6: a model for a possible family of structurally related RNA pseudoknots. *Biochemistry* **35**, 4187-98

Dunstan H. M., Young L. S. & Sprague K. U. (1994) TFIIIR is an isoleucine tRNA. *Mol. Cell Biol.* **14**, 3588-3595.

Fan P., Suri A. K., Fiala R., Live D. & Patel D. J. (1996) Molecular recognition in the FMN-RNA aptamer complex. *J. Mol. Biol.* **258**, 480-500.

Ferré-D'Amaré A.R., Zhou K. & Doudna J.A. (1998) Crystal structure of a hepatitis delta virus ribozyme. *Nature* **395**, 567-574.

Greenbaum N.L., Radhakrishnan I., Patel D.J. & Hirsh D. (1996) Solution structure of the donor site of a trans-splicing RNA. *Structure* **4**, 725-733.

Hakoshima T., Fukui T., Ikehara M. & Tomita K. (1981) Molecular structure of a double helix with a non-Watson-Crick type base pairing formed by 2-substituted poly(A) and poly(U). *Proc. Natl. Acad. Sci. USA* **78**, 7309-7313.

Hall K., Cruz P., Tinoco I. Jr., Jovin T.M. & van de Sande J.H. (1984) 'Z-RNA'-A left-handed RNA double helix. *Nature* **311**, 584-586.

Harada F., Peters G. G. & Dahlberg J. E. (1979) The primer tRNA for Moloney murine leukemia virus DNA synthesis. Nucleotide sequence and aminoacylation of $tRNA^{Pro}$. *J. Biol. Chem.* **254**, 10979-10985.

Helm M., Brule H., Degoul F., Cepanec C., Leroux J.P., Giege R. & Florentz C. (1998) The presence of modified nucleotides is required for cloverleaf folding of a human mitochondrial tRNA. *Nucleic Acids Res.* **26**, 1636-1643.

Henkin T. M. (1994) tRNA-directed transcription antitermination. *Mol. Microbiol.* **13**, 381-387.

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

Heus H. A., Wijmenga S. S., Hoppe H. & Hilbers C. W. (1997) The detailed structure of tandem G.A mismatched base-pair motifs in RNA duplexes is context dependent. *J. Mol. Biol.* **271**, 147-158.

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

Jaeger J., Restle T. & Steitz T. A. (1998) The structure of HIV-1 reverse transcriptase complexed with an RNA pseudoknot inhibitor. *EMBO J.* **17**, 4535-4542.

Jiang F., Kumar R. A., Jones R. A. & Patel J. D. (1996) Structural basis of RNA folding and recognition in an AMP-RNA aptamer complex. *Nature* **382**, 183-186.

Jucker F. M. & Pardi A. (1995) Solution structure of the CUUG hairpin loop: a novel RNA tetraloop motif. *Biochemistry* **34**, 14416-14427.

Jucker F. M., Heus H. A., Yip P. F., Moors E. H. & Pardi A. (1996) A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J. Mol. Biol.* **264**, 968-980.

Kajava A. & Rüterjans H. (1993) Molecular modeling of the 3-D structure of RNA tetraloops with different nucleotide sequences. *Nucleic Acid Res.* **21**, 4556-4562.

Kang H., Hines J. V. & Tinoco I. Jr. (1996) Conformation of a non-frameshifting RNA pseudoknot from mouse mammary tumor virus. *J. Mol. Biol.* **259**, 135-47

Kanyo J. E., Duhamel J. & Lu P. (1996) Secondary structure of the r(CUUCGG) tetraloop. *Nucleic Acid Res.* 24, 4015-4022.

Kim S. H. (1978) Three-dimensional structure of transfer RNA and its functional implications. *Adv. Enzymol.* **46**, 279-315.

Kim S. H., Suddath F. L., Quigley G. J., McPherson A., Sussman J. L. Wang A. H. J., Seeman N. C. & Rich A. (1974) Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. *Science* **185**, 435-440.

Kolk M. H., van der Graaf M., Fransen C. T., Wijmenga S. S., Pleij C. W., Heus H. A. & Hilbers C. W. (1998) Structure of the 3'-hairpin of the TYMV pseudoknot: preformation in RNA folding. *EMBO J.* **17**, 7498-7504.

Kolk M. H., Heus H. A.& Hilbers C. W. (1997) The structure of the isolated, central hairpin of the HDV antigenomic ribozyme: novel structural features and similarity of the loop in the ribozyme and in free solution. *EMBO J.* **16**, 3685.

Lee A.J & Crothers D. M. (1998) The solution structure of an RNA loop-loop complex: the ColE1 inverted loop sequence. *Structure* **6**, 993-1005.

Leibowitz M. J. & Soffer R. L. (1969) A soluble enzyme from *Escherichia coli* which catalyzes the transfer of leucine and phenylalanine from tRNA to acceptor proteins. *Biochem. Biophys. Res. Commun.* **36**, 47-53.

Levitt M. (1969) Detailed molecular model for transfer ribonucleic acid. *Nature* **224**, 759-763.

Littauer U. Z. & Inouye H. (1973) Regulation of tRNA. *Annu. Rev. Biochem.* **42**, 439-470.

Luebke K. J., Landry S. M. & Tinoco I. Jr. (1997) Solution structure of a five-nucleotide RNA bulge loop from group I intron. *Biochemistry* **36**, 10246-10255.

Maidak B. I., Olsen G. J., Larsen N., Overbeck R., McCaughey M. J. & Woese C. R. (1997) The RDP (Ribosomal Database Project) *Nucleic Acids Res.* **25**, 109-110.

Michel F., Umesono K. & Ozeki H. (1989) Comparative and functional anatomy of group II catalytic introns--a review. *Gene* **82**, 5-30.

Michel F. & Westhof E. (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* **216**, 585-610.

Moras D., Comarmond M. B., Fisher J., Weiss R., Thiery J. C., Ebel J. P. & Giege R. (1980) Crystal structure of yeast $tRNA^{Asp}$. *Nature (Lond)* **288**, 669-674.

Pley H. W., Flaherty K. M. & McKay D. B. (1994) Three-dimensional structure of a hammerhead ribozyme. *Nature* **372**, 68-74.

Puglisi E. V. & Puglisi J. D. (1998) HIV-1 A-rich RNA loop mimics the tRNA anticodon structure. *Nat. Struct. Biol.* **5**, 1033-1036.

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

Puglisi J. D., Wyatt J. R. & Tinoco I. Jr. (1990) Solution conformation of an RNA hairpin loop. *Biochemistry* **29**, 4215-26.

Quigley G. J. & Rich A. (1976) Structural domains of transfer RNA molecules. *Science* **194**, 796-806.

Rao S. T. & Rossmann M. G. (1973) Comparison of super-secondary structure in proteins. *J. Mol. Biol.* **76**, 241-56

Rife J. P. & Moore P. B. (1998) The structure of a methylated tetraloop in 16S ribosomal RNA. *Structure* **6**, 747-756.

Robertus J. D., Ladner J. E., Finch J. R., Rhodes D., Brown R. S., Clark B.F.C. & Klug A. (1974) Structure of yeast phenylalanine tRNA at 3Å resolution. *Nature* **250**, 546-551.

Romby P., Carbon P., Westhof E., Ehresmann C., Ebel J. P., Ehresmann B. & Giege R. (1987) Importance of conserved residues for the conformation of the T-loop in tRNAs. *J. Biomol. Struct. Dyn.* **5**, 669-687.

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

Saenger W. (1984) *Principles of Nucleic Acid Structures*. Springer-Verlag, New-York.

SantaLucia J., & Turner D.H. (1993) Structure of $(rGGCGAGCC)_2$ in solution from NMR and restraint molecular dynamics. *Biochemistry* **32**, 12612-12623.

Schindelin H., Zhang M., Bald R., Furste J. P., Erdmann V.A. & Heinemann U. (1995) Crystal structure of an RNA dodecamer containing the *Escherichia coli* Shine-Dalgarno sequence. *J. Mol. Biol.* **249**, 595-603

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

Shen L. X., Cai Z. & Tinoco I. Jr. (1995) RNA structure at high resolution. *FASEB J.* **9**, 1023-33

Shen L. X. & Tinoco I. Jr. (1995) The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary tumor virus. *J. Mol. Biol.* **247**, 963-978

Shön A., Krupp G., Berry-Lowe S., Kannangara C. G. & Söll D. (1986) The RNA required in the first step of chlorophyll biosynthesis is a chloroplast glutamate RNA. *Nature* **322**, 281-284.

Sich C., Ohlenschlager O., Ramachandran R., Gorlach M. & Brown L. R. (1997) Structure of an RNA hairpin loop with 5'-CGUUUCG-3' loop motif by heteronuclear NMR spectroscopy and distance geometry. *Biochemistry* **36**, 13989-14002.

Stallings S. C. & Moore P. B. (1997) The structure of an essential splicing element: stem loop IIA from yeast U2 snRNA. *Structure* **5**, 1173.

Steinberg S. & Cedergren R. (1994) Structural compensation in atypical mitochondrial tRNAs. *Nat. Struct. Biol.* **1**, 507-510.

Steinberg S. & Cedergren R. (1995) A correlation between N2-dimethylguanosine presence and alternate tRNA conformers. *RNA* **1**, 886-891.

Steinberg S., Leclerc F. & Cedergren R. (1997) Structural rules and conformational compensations in the tRNA L-form. *J. Mol. Biol.* **266,** 269-282.

Studnicka G. M., Rahn G. M., Cummings I. W. & Salser W. A. (1978) Computer method for predicting the secondary structure of single-stranded RNA. *Nucleic Acid Res.* **5**, 3365-

Szewczak A. A. Moore P.B., Chang Y. L. & Wool I. G. (1993) The conformation of the sarcin/racin loop from 28S RNA. *Proc. Natl. Acad. Sci. USA* **90**, 9581-9585.

Tang R.S. & Draper D.E. (1990) Bulge loops used to measure the helical twist of RNA in solution. *Biochemistry* **29**, 5232-5237.

Varani G., Wimberly B. & Tinoco I. Jr. (1991) Structure of an unusually stable RNA hairpin. *Biochemistry* **30**, 3280-3289.

Wimberly B., Varani G. & Tinoco I. Jr. (1993) The conformation of loop E of eukaryotic 5S ribosomal RNA internal loop. *Biochemistry* **32**, 1078-1087.

Woese C. R., Gutell R., Gupta R. & Noller H. F. (1983) Detailed analysis of the higher-order structure of 16S-like ribosomal nucleic acids. *Microbiol Rev.* **42**, 621-669.

Woese C. R., Winker S. & Gutell R. R. (1990) Architecture of ribosomal RNA: constraints on the sequence of "tetra-loops". *Proc. Natl. Acad. Sci. USA* **87**, 8467-8471.

Wolstenholme D. R., Macfarlane J. L., Okimoto R., Clary D. O. & Wahleithner J. A. (1987) Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. *Proc. Natl. Soc. Acad. Sci. USA* **84**, 1324-1328.

Acknowledgment

I would like to express my gratitude to Dr. Sergey Steinberg for his mentorship and guidance through all these eventful years.

My special thanks are to Alice Rae, Marie Pageau and Veronique Bourdeau for their help in improving my English and French and correcting this manuscript. Last, but not least, I would like to thank all members of our lab, current and those who left, for their support, friendship and endless discussions.