Université de Montréal

Shape-keeping elements in tRNA

par

Ekaterina Zagriadskaia

Département de Biochimie

Faculté de Médecine

Thèse présentée à la Faculté des études supérieures

en vue de l'obtention du grade de Ph.D. en Biochimie

Decembre 2003

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée:

Shape-keeping elements in tRNA

présentée par
Ekaterina Zagriadskaia

A été évaluée par un jury composé des personnes suivantes:

Président-rapporteur	Dr. Luis Rokeach
Directeur de recherche	Dr. Serguei Chteinberg
Membre du jury	Dr. Joelle Pelletier
Examinateur externe	Dr. William McClain
Représentant du doyen de la FES	Dr. Elliot Drobetsky

SUMMARY

RNA molecules fulfill important functions in living cells. They serve as messengers in the decoding of genetic information from DNA into proteins, and also carry out some catalytic functions. Catalytic RNAs need to assume unique tertiary structures to become active. Tertiary structure of RNA is determined by its nucleotide sequence, however, the rules that govern RNA folding are not well understood. In the present study, the structural scaffold of transfer RNA (tRNA) was used to investigate how this particular RNA molecule forms and maintains its tertiary structure, and how the structure influences the function.

Transfer RNAs function in protein biosynthesis by matching amino acids to the corresponding codons on mRNA. All tRNAs have universal tertiary structure known as the 'L-shape', which is essential for their function. Conserved tertiary interactions are involved in the formation of the L-shape, many of which concentrate in the 'elbow' region, formed by interaction of the D- and T-loops of tRNA. I have investigated sequence and structural requirements for this region of tRNA. For this, I created a tRNA gene library, where several positions in the D- and T-loops were randomized, and screened this library *in vivo* for the presence of functional suppressor tRNAs. Screening revealed the importance of a particular tertiary interaction, reverse-Hoogsteen (RH) base pair between nucleotides 54 and 58 in the T-loop, for the tRNA structure, while other conserved tertiary interactions in the elbow region turned to be more tolerant of changes and could be replaced by alternative sets of interactions without destroying tRNA function. To find out the exact role of the RH base pair, I constructed the second tRNA gene library. Screening of this library showed, that the RH base pair is involved into formation of the L-shaped structure by providing correct juxtaposition of the two helical domains of tRNA. Only those molecules that were able to maintain the L-shape were functional.

My results contribute to the understanding of the role of different types of tertiary interactions in the formation of RNA structure and demonstrate that combinatorial approach can be successfully used *in vivo* to investigate the principles of RNA folding.

Key words: tRNA structure, T-loop, L-shape, combinatorial gene library, selection in vivo

RÉSUMÉ

Les ARNs jouent un rôle important dans les cellules vivantes. Ils servent comme messagers dans le décodage de l'information génétique et ils ont aussi des fonctions catalytiques. L'activité catalitique de l'ARN nécessite la formation d'une structure tertiaire unique. Cette structure tertiaire est déterminée par la séquence primaire de l'ARN, mais les règles qui gèrent ce repliment sont encore mal connues. Dans le travail présenté, nous avons utilisé l'ARN de transfert (ARNt) comme échafaudage pour étudier comment cet ARN particulier forme et maintient sa structure tertiaire, et comment cette dernière influence la fonction de l'ARNt.

Les ARNt fonctionnent dans la synthèse protéique. Leur rôle est d'apporter sur le ribosome l'acide aminé correspondant au codon de l'ARN messager. Tous les ARNts ont une structure tertiaire commune appelée 'la forme en L'. Cette structure est maintenue par de nombreuses interactions tertiaires, dont plusieurs se concentrent dans la région du 'coude' de l'ARNt, formé par l'interaction entre les boucles D et T. Nous avons étudié les exigences imposées sur la séquence et la structure de cette région de l'ARNt. Pour ce faire, nous avons construit une bibliothèque des gènes d'ARNt où plusieurs positions dans les boucles D et T ont été rendues aléatoires. Cette bibliothèque a été criblée *in vivo* pour sélectionner des ARNts suppresseurs. Ce criblage a révélé l'importance d'une interaction tertiaire en particulier, soit la paire de bases de type 'reverse-Hoogsteen' (RH) entre les nucléotides 54 et 58 dans la boucle T, pour la structure et la fonction de l'ARNt, tandis que d'autres interactions tertiaires dans la région 'coude' se sont avérées plus variables et pouvaient être remplacées par des interactions alternatives. Pour trouver le rôle exact de la paire de bases RH, nous avons construit une deuxième bibliothèque des gènes des ARNts. Le criblage de cette bibliothèque a montré que la paire de bases RH est impliquée dans la formation de la structure 'en L' par son rôle essentiel dans l'arrimage correct des deux domaines hélicoïdals de l'ARNt. Seulement les molécules capables d'adopter la structure 'en L' étaient fonctionnelles.

Ces résultats contribuent à la compréhension du rôle des différentes interactions tertiaires dans la formation de la structure de l'ARN et ils démontrent que l'approche utilisant des bibliothèques des gènes peut être très utile pour l'étude des principes du repliment de l'ARN.

Mots clés : structure d'ARNt, boucle T, forme L, bibliothèque des gènes, sélection *in vivo*

TABLE OF CONTENTS

LIST OF TABLES page

LIST OF FIGURES

LIST OF ABBREVIATIONS

Standard three-letter code is used for all amino acids

5S, 16S and 23S	5S, 16S and 23S subunits of bacterial ribosome
Å	angstrom (1^{-10} m)
DNA	deoxyribonucleic acid
RNA	ribonucleic acid
mRNA	messenger RNA
rRNA	ribosomal RNA
tRNA	transfer RNA
$tRNA^{Ser}_{GCU}$	tRNA serine with anticodon GCU
mit tRNA	mitochondrial tRNA
tmRNA	transfer-messenger RNA
RH	reverse-Hoogsteen base pair
ArcTGT	archaeosine tRNA – guanine transglycosylase
AspRS	aspartyl-tRNA synthetase
GlnRS	glutamyl-tRNA synthetase
PheRS	phenylalanyl-tRNA synthetase
SerRS	seryl-tRNA synthetase
RNase P	ribonuclease P
EF-Tu	elongation factor Tu
Y	pyrimidine base
R	purine base
Ψ	pseudouridine base
A	adenine base
C	cytosine base
G	guanine base
T	thymine base
U	uracil base

ACKNOWLEDGMENTS

This work would never be done without help and support from my advisor, Dr. Sergey Steinberg. His understanding was really remarkable. I also got a lot of technical help at the beginning of this project from graduate student Cesar Gomez. Thanks to Cesar, Sandra, Matthieu, Geneviève and Félix for the great time we had together at the lab. I'm also grateful to Prof. Léa Brakier-Gingras for discussions about my project and for advice on writing the thesis. Many department members were very helpful during the time of my studies, especially Mme Sylvie Beauchemin and Mme Denise Lessard. Thanks to Dominic Dulude and Martin Baril for their help with some experiments. I appreciate the support and encouragement from my family during the time I was writing this thesis. Finally, I'm grateful to Prof. Shana Kelley from Boston College, who allowed me to work in her lab while I was writing my thesis.

INTRODUCTION

Transfer RNA – a treasury of stereochemical information

Discovery of tRNA

Transfer RNA (tRNA) is probably one of the most studied classes of biological macromolecules. Almost 50 years of research have assembled an enormous amount of both functional and structural information about tRNAs. However, there are still questions waiting to be answered about them.

In 1955 Francis Crick formulated the "adaptor hypothesis" about the existence of adaptor molecules (possibly RNAs) in protein biosynthesis, which carry enzymatically attached amino acids and can specifically recognize codons on mRNA (Crick, 1966). In 1957-58 Paul Zamecnik, Mahlon Hoagland, Robert Holley and colleagues discovered the requirement for "soluble" RNA in order to incorporate amino acids into proteins (Hoagland *et al*, 1958; Holley, 1957). This RNA was later called transfer RNA. Transfer RNAs are adaptor molecules that translate genetic code from messenger RNA into amino acid sequence. Thus, each tRNA has two functional centers: the anticodon, which recognizes the codon on mRNA during translation, and the acceptor 3' end, where the cognate amino acid is attached to tRNA by the enzyme aminoacyl-tRNA synthetase. During translation, anticodon and acceptor end of tRNA interact with different subunits of the ribosome, and tRNA is involved in a complex series of movements and reactions on the ribosome, including translocation from A to P to E sites, decoding and peptidyl transfer. Besides ribosome and aminoacyl-tRNA synthetases, tRNA interacts with a lot of other factors during its life cycle: 5'- and 3'- processing enzymes, modifying enzymes, CCA-adding enzyme and elongation or initiation factors.

Some functions of tRNA not related to protein synthesis have been discovered. For example, $tRNA^{Lys}$ serves as a primer for reverse transcription in the HIV virus. Some plant RNA viruses have tRNA-like structures at the 3' ends of their genomes, which play the same role. Transfer RNAs are also involved in cell wall biosynthesis, as well as chlorophyll and heme biosynthesis, and in regulation of expression of some aminoacyl-tRNA synthetases (Soll and RajBhandary, 1995).

Cloverleaf secondary structure

In 1965 Holley and co-workers (Holley *et al*, 1965) managed to isolate and sequence the first tRNA, yeast $tRNA^{Ala}$. They proposed three alternative secondary structures for this molecule, one of them being the "cloverleaf". Later it became evident that cloverleaf structure is common to all prokaryotic, archaeal and eukaryotic cytosolic tRNAs. Cloverleaf (Fig. 1A, page 3) consists of four stems: acceptor, D (for dihydrouridine), anticodon and T (for ribothymidine). Three stems are capped with loops, called D-, T- and anticodon loops, while the acceptor stem is formed by the 5' and 3' ends of the molecule. There is an unpaired CCA sequence at the very 3' end of all tRNAs, which is necessary for aminoacylation. The length of the acceptor, T- and anticodon stems, as well as of T- and anticodon loops, is conserved in all tRNAs, while D- stem can vary from 3 to 4 base pairs, and the D-loop can have 7 to 11 nucleotides. Between anticodon and T-stems there is a region of variable length. In most tRNAs it is 4-5 bases long and is called variable loop, while in tRNAs leucine, serine and prokaryotic tRNA tyrosine it is much longer (10-24 bases) and forms a base paired stem with loop, called variable arm. There are more than 20 conserved or semi-conserved (restricted to purines or pyrimidines) residues in tRNAs, which are mostly located outside the stems (Dirheimer *et al*, 1995). It has been realized early that these nucleotides may be involved in the long-range interactions and formation of tertiary structure. However, modeling attempts were not able to predict the correct tertiary fold of tRNA, though some important details, like coaxial stacking of stems or tertiary base pair between nucleotides 15 and 48 were predicted correctly (Levitt, 1969).

L-shape of yeast $tRNA^{Phe}$

The first crystal structure of a tRNA at atomic resolution, that of yeast $tRNA^{Phe}$, was published in 1974 independently by the groups of Alexander Rich (Suddath *et al,* 1974) and Aaron Klug (Robertus *et al,* 1974). They were later refined to about 2.5 Å resolution, and recently both structures were re-determined using modern crystallographic methods with a resolution under 2 Å (Shi and Moore, 2000; Jovine *et al,* 2000). Though two structures result from different crystal forms, orthorhombic and monoclinic, the structures are almost identical. So, in 1974 it became clear, that tRNA had the shape of letter L (Fig. 1B, page 3), where one side of L was formed by coaxial stacking of the anticodon and D-stems and the other side - by stacking of acceptor and T-stems (Fig.2, page 4).

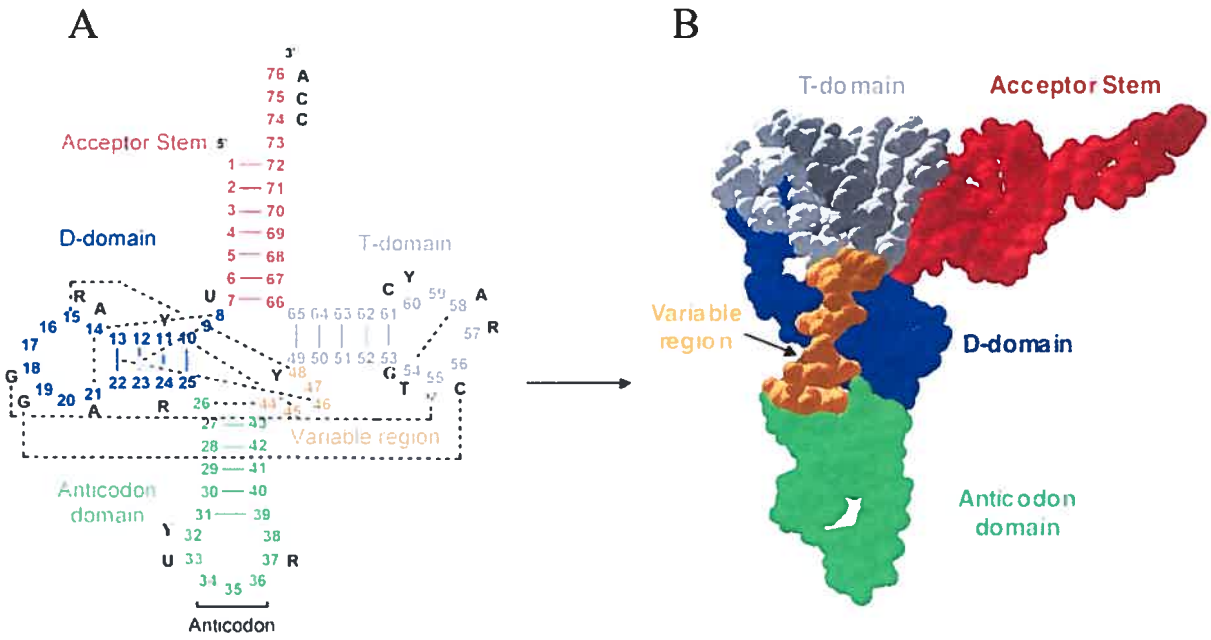


Figure 1. Cloverleaf and the L-shape of tRNA. **A**. Cloverleaf secondary structure of a canonical tRNA. Tertiary interactions are shown by dashed lines. Several conserved and modified nucleotides are shown: Ψ stands for pseudouridine, T for ribothymidine, R for purine, Y for pyrimidine. **B**. A space filling representation of the tertiary L-shaped structure of yeast $tRNA^{Phe}$. Domains are colored the same way as in **A**.

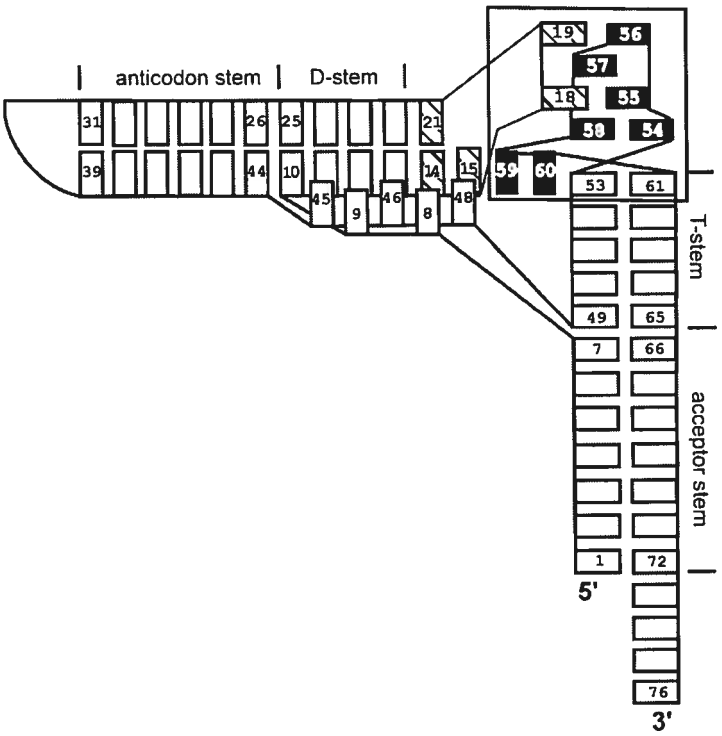


Figure 2. The scheme of the L-shape of $tRNA^{Phe}$. Nucleotides are represented by rectangles. Tertiary interactions in the core and the elbow region are shown as paired or overlapping rectangles. The anticodon loop is oriented to the left; CCA end is at the bottom. The elbow region is boxed, with T-loop nucleotides in black and D-loop nucleotides crosshatched. Nucleotides of the anticodon loop, nucleotide 47 and unstacked nucleotides of the D-loop are not shown.

However, D- and anticodon stems are not perfectly coaxial, as there is a kink of about 26° between them. Two sides of L are about the same length (60Å) and almost perpendicular to each other. Two functional centers of tRNA, the anticodon and acceptor terminus, are located at the two ends of the molecule. Acceptor end is exposed into solution, while all three nucleotides of the anticodon are stacked onto each other in the conformation suitable for recognition of the codon on mRNA.

The distance between the anticodon and the acceptor end, called 'primary axis', is about 80Å. The corner, or the 'elbow' of L is formed by the T-loop which interacts with the D-loop through formation of two base pairs, G18-Ψ55 and G19-C56 (elbow region is boxed on Fig.2). Other long-range interactions help to make the 'core' of the molecule, also called 'extended D-stem', because it consists of base triples formed by base pairs of the D-stem and nucleotides from the connector regions (nucleotides 8, 9 between the acceptor and D- stems and nucleotides 45-48 of the variable loop, see Fig 2). There are four base triples and a reverse-Watson-Crick base pair 15-48 in the core region. In addition, stacking interactions are very important for the formation of tertiary structure: 71 out of 76 nucleotides participate in stacking, even if they are not in the helices.

For a long time the $tRNA^{Phe}$ structure was the only one known for an RNA molecule, and was a source of valuable information on the structure of RNA helices, stacking, non-canonical base pairs, long-range interactions and sugar-phosphate backbone conformations. For this reason it was considered 'a treasury of stereochemical information' (Saenger, 1984).

Three other crystal structures of tRNA were solved later: yeast $tRNA^{Asp}$ (Moras *et al,* 1980), yeast initiator $tRNA^{Met}$ (Schevitz *et al,* 1979) and *E.coli* initiator $tRNA^{fMet}$ (Woo *et al,* 1980). All of them were later refined to atomic resolution. They all conserve the L-shape and have overall structures very similar to that of yeast $tRNA^{Phe}$. However, some smaller differences exist. For example, $tRNA^{Asp}$, which has been crystallized in two different forms, A and B, has a bigger angle between the two sides of L (about 110°, form A is shown in Fig. 3b, page 7). Also, D- and anticodon stems are more coaxial. The variable loop has four nucleotides instead of five, so residues 46 and 48 have slightly different positions, which induces different environment for base triple 8-14-21 and bases 9 and 45 (Westhof *et al,* 1985). Also, the anticodons of the two neighboring tRNA molecules in the crystal pair with each other (they are self-complementary), which stabilizes the anticodon arm, but destabilizes D- and T- loops, so that tertiary base pair 19-56 does not

exist in the B-form (B is the form obtained at lower temperature), while in the A-form it is present only in a fraction of molecules. All other interactions are the same as in $tRNA^{Phe}$.

Structures of initiator tRNAs also reveal some peculiarities that can be associated with their function in initiation, but not elongation, of protein synthesis (Basavappa and Sigler, 1991).

tRNA structure in complexes with proteins

The structure of $tRNA^{Gln}$ bound to its cognate aminoacyl-tRNA synthetase (GlnRS) was the first tRNA-protein complex to be solved in 1989 (Rould *et al*). Today, structures of 11 other tRNAs bound to their synthetases are known (for tRNAs Leu, Val, Ile, Arg, Phe, Asp, Cys, Tyr, Pro, Thr and Ser; in some of them tRNAs and synthetases are from different species, but form active complexes). Also, structures of tRNA complexes with elongation factor EF-Tu (Nissen *et al,* 1995) and several modification enzymes are known (Ishitani *et al,* 2003). Finally, ribosome crystal structures include mimics of parts of tRNAs or whole tRNAs (Yusupov *et al,* 2001; Ban *et al,* 2000; Wimberly *et al,* 2000). These structures give lots of information about tRNA function. One can also see the changes in tRNA structure induced by protein binding. It is impossible to discuss all these structures in detail, so I will mention only a few of them. As a rule, with only one exception known today, tRNA in these complexes preserves its L-shape, however, protein binding usually induces local structural changes in tRNA.

For example, yeast $tRNA^{Asp}$ bound to its synthetase, AspRS (Ruff *et al*, 1991), keeps the L-shape, but is nonetheless deformed compared to its free form (Fig.3, page 7). The anticodon arm is bent by as much as 20° towards the inside of the L, the anticodon loop is unfolded, with the anticodon bases unstacked in order to interact with the protein. The enzyme also interacts with the end of acceptor stem from the major groove side, which in RNA is much less accessible than the minor groove, by making it wider to accommodate protein side chains. The enzyme also specifically recognizes the G10-U25 wobble pair in the D- stem. However, most of the close contacts between protein and tRNA in this and other complexes are made with the backbone of tRNA, not the bases.

In $tRNA^{Gln}$-GlnRS complex (Rould *et al*, 1989) the anticodon stem is extended by two non-canonical base pairs, U32-Ψ38 and U33-A37, causing the anticodon to unstack, with its bases pointing into opposite directions and bound in different protein pockets.

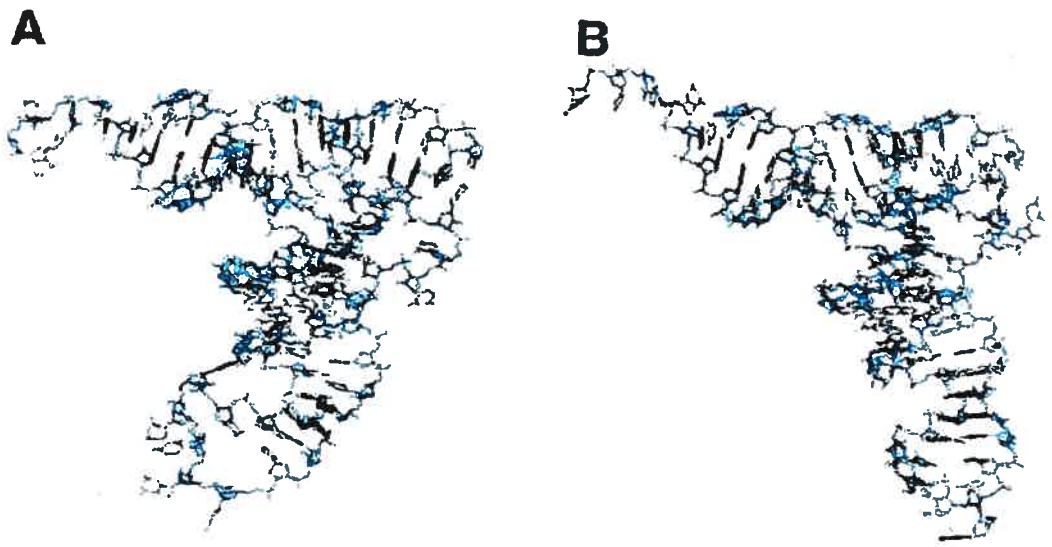


Fig. 3. Structure of yeast $tRNA^{Asp}$ in its free form (B) and complexed with the aminoacyl-tRNA synthetase (A, protein not shown). Anticodon loop is at the bottom, 3' end is oriented to the left. Differences between the two structures concern the anticodon arm, CCA end and D-T loop interaction.

Adapted from color plate 7 in Soll and RajBhandary (eds), *tRNA:structure, biosynthesis and function,* 1995.

The first base pair of the acceptor stem is disrupted by the protein. Unlike AspRS, GlnRS interacts with the acceptor stem from the more accessible minor groove side. It also recognizes base G10 in the D-stem. A peculiarity of the $tRNA^{Gln}$ is that it has a loosely packed core compared to yeast $tRNA^{Phe}$: nucleotide 46, instead of making a triple with base pair 13-22, is bulged out, and nucleotide 45 takes its place instead of interacting with base pair 10-25. This results in a 'hole' at the place usually occupied by nucleotide 45. Not surprisingly, the replacement of variable loop in this tRNA by a shorter one leads to the more compact core and increases the stability of the tRNA-synthetase complex (Bullock *et al*, 2000).

Another interesting example is the $tRNA^{Ser}$- SerRS complex (Biou *et al*, 1994). Since $tRNA^{Ser}$ has a long variable arm, several rearrangements are needed in the core of the tRNA at the place of the variable arm attachment. This attachment is mediated by two additional nucleotides in the D-loop of this tRNA, 20a and 20b. The variable arm protrudes from the body of the molecule at an angle of about 45 degrees to the plane of the L and serves as the major identity element for recognition by its cognate synthetase. The rest of the $tRNA^{Ser}$ molecule is normal, including the standard interactions between the T- and D-loops.

From this and other examples we can see that the L-shape of tRNA is universal, but flexible enough to allow tRNA to interact with many proteins, while keeping its general tertiary fold.

λ-form of tRNA

The only example known to date of a tRNA-protein complex, where the L-shape of tRNA is not maintained, is the recent structure of the modifying enzyme archaeosine tRNA – guanine transglycosylase (ArcTGT) bound to $tRNA^{Val}$ (Ishitani *et al*, 2003). This is also the first determined structure of a tRNA modification enzyme bound to a full-length tRNA. ArcTGT modifies nucleotide G15 into archaeosine in many archaeal tRNAs. In the L-shape, G15 forms tertiary base pair with nucleotide 48 and as a result is buried deep inside the tRNA core. The crystal structure shows, that in order to reach G15, the enzyme completely unfolds the D-arm of tRNA. However, this tRNA is not completely denatured, but forms an alternative structure, named λ-form (Fig. 4, page 9). While nucleotides 8 to 22 of the D-arm protrude, the rest of the D-stem (nucleotides 23 to 26) forms a new helical

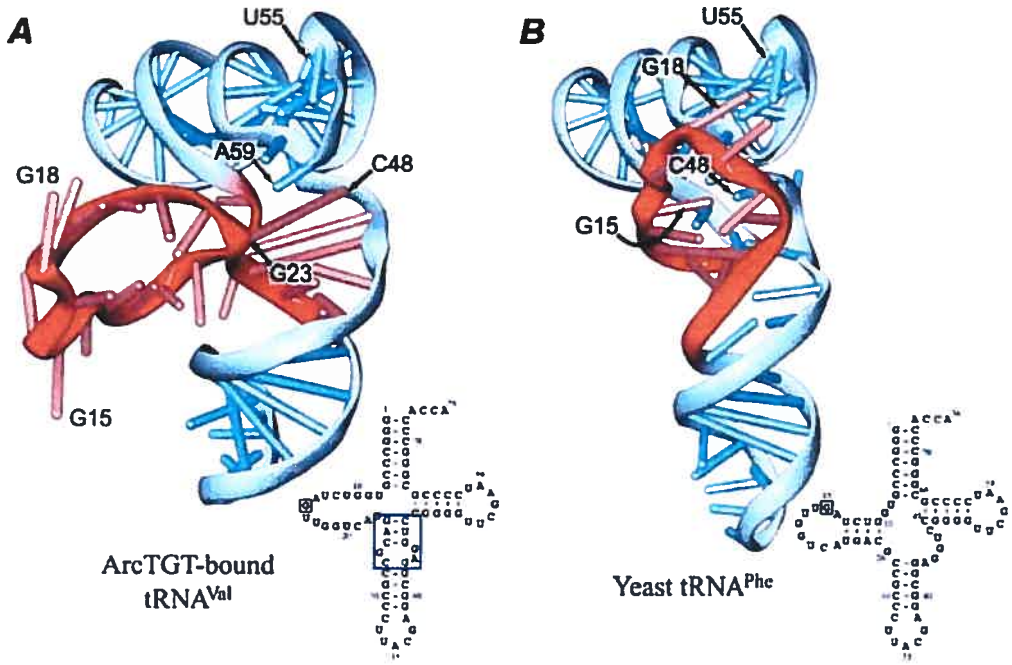


Figure 4. Comparison of λ-form (A) and L-form (B) of $tRNA^{Val}$. Sugar-phosphate backbone is represented as a ribbon, bases are shown as short sticks, Watson-Crick base pairs – as long sticks. The D-arm is colored in red. Corresponding secondary structures are shown for both tRNA forms. From Ishitani *et al*, 2003.

structure with the nucleotides of the variable loop, which authors call 'DV helix'. This helix is stacked on top of the anticodon stem. The acceptor/T domain remains folded and stacks to the last base pair 23-48 of the DV helix instead of the canonical base pair 15-48. Unfolded D-arm is extensively contacted by protein, which helps to position the G15 correctly for the catalysis.

It is possible, that λ-form may be common for tRNA binding to several modification enzymes that modify any of the core nucleotides 8, 9, 10, 13 or 22.

Crystal structure of the tRNA-like domain of tmRNA

One more recent crystal structure, that of tRNA-like domain of the transfer-messenger RNA (tmRNA) bound to the SmpB protein (Gutmann *et al*, 2003), is another example of variability, which can be achieved within a tRNA-like fold.

tmRNA is a hybrid bacterial RNA that combines functions of both transfer and messenger RNAs. It rescues ribosomes stalled at the ends of truncated mRNAs lacking stop codons. tmRNA is recognized by alanyl-tRNA synthetase, charged with alanine and delivered by elongation factor Tu to the A-site on the ribosome together with SmpB. Then, the incompletely synthesized polypeptide chain is transferred from the P-site tRNA to the alanyl-tmRNA, and the ribosome resumes translation, switching to the open reading frame of 10 amino acids encoded by the mRNA region of the tmRNA. Termination takes place normally at the stop codon on tmRNA, and the added peptide serves as a degradation tag for the truncated protein.

The tRNA-like domain of tmRNA has canonical acceptor and T-stems and T-loop (Fig. 5a, page 11). However, the D-arm is replaced by a loop, which does not have any base pairing. Helix 5 (H5) is an equivalent of the anticodon stem, but it is longer (with 8 base pairs) and ends with internal loop, after which the rest of the molecule follows. In the crystal structure, this loop was replaced by a tetraloop, thus cutting off the tRNA-like domain from the rest of the tmRNA.

The crystal structure shows, that the tRNA-like domain adopts an open L-shaped conformation, with the single-stranded D-loop bound by SmpB protein. However, this structure is significantly different from conventional L-shape (Fig. 5b). First, the angle between the two helical domains is much larger than in canonical tRNA. Second, the T-arm is rotated by 90° around its axis, resulting in significantly different orientation of this arm

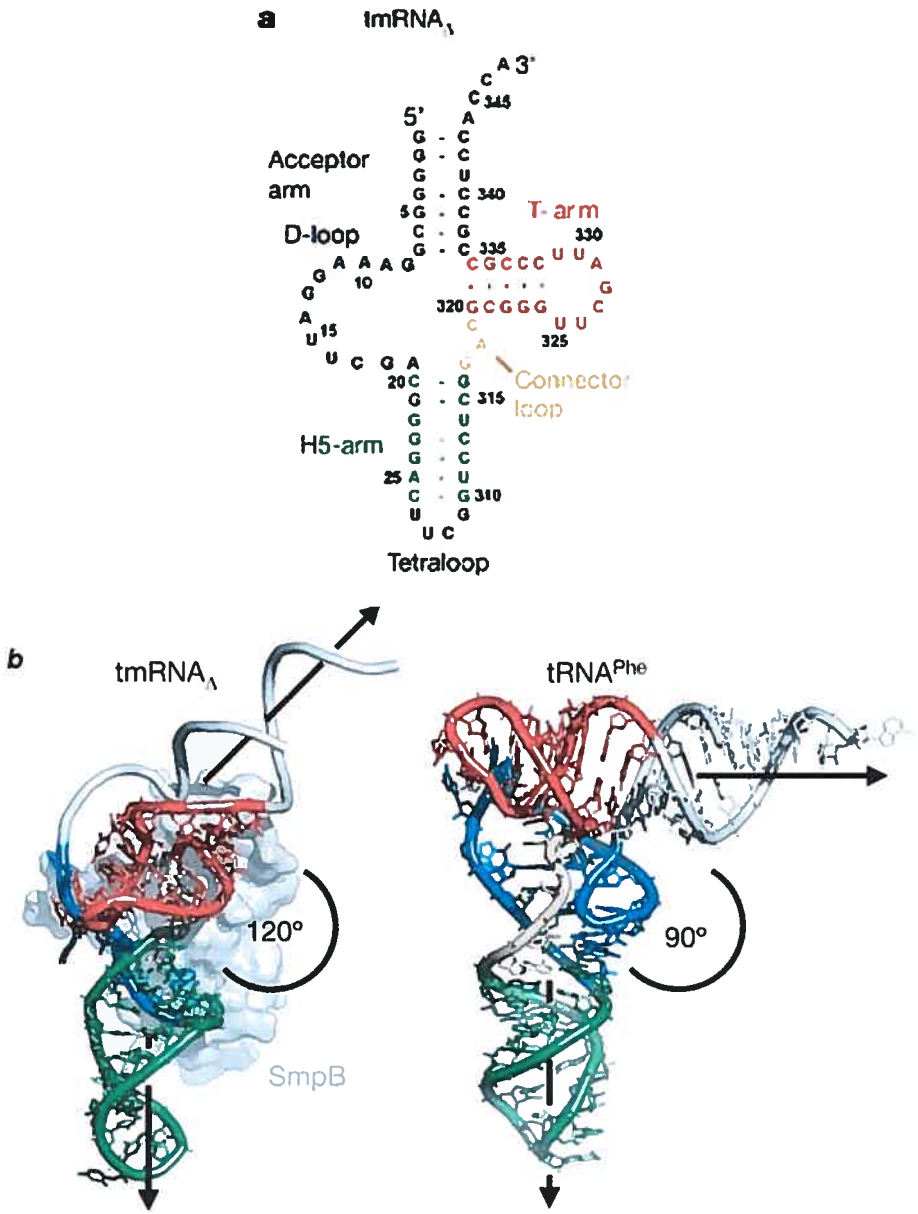


Fig. 5. Structure of the tRNA-like domain of tmRNA (from Gutmann *et al*, 2003).

a Cloverleaf-like secondary structure of the tRNA-like domain of the tmRNA.

b Comparison of the tertiary structure of the tRNA-like domain of the tmRNA with SmpB protein (left) and the yeast $tRNA^{Phe}$ (right). Color code of different domains is the same as in ***a***. In tmRNA, the poorly ordered acceptor stem and five nucleotides in the D-loop were modeled (shown in grey). tmRNA has a significantly larger angle between acceptor and anticodon stems (120°) than $tRNA^{Phe}$ (90°), while its T-arm is rotated approximately 90° around its helical axis, compared to $tRNA^{Phe}$.

relative to the rest of the molecule. In this conformation, T-loop cannot interact with the D-loop. The latter is stabilized by contacts with SmpB. Due to rotation of the T-arm, the H5 arm of tmRNA points in the direction, almost orthogonal to that of the anticodon stem in the tRNA, when the two structures are superimposed by their T-stems. This is certainly related to the functional differences between tmRNA and tRNA: tmRNA does not participate in codon-anticodon interaction, instead, it needs to accommodate the rest of its molecule and the SmpB protein, avoiding their collision with ribosomal subunits. This can be achieved due to the rotation of the T-arm. Thus, the structure of tmRNA is adapted to its unique function in translation. However, it is not clear from this crystal structure, given the absence of interactions between the T- and D-loops, why tmRNA has conserved the canonical tRNA sequence of the T-loop and the GG dinucleoitide in the D-loop, i.e., elements, which normally interact with each other. Maybe, tmRNA can adopt an alternative conformation (in the absence of SmpB), where the D- and T-loops interact with each other as in normal tRNA.

Mitochondrial tRNAs

Many animal mitochondrial tRNAs (mit tRNAs) present unusual primary and secondary structures (Helm *et al*, 2000). Almost any given mit tRNA, except those from plants, has some 'weird' features, like the absence of a conserved nucleotide, mismatches in the stems, or, in the extreme case, the lack of the entire D- or T- domain (in mit tRNAs from nematoda worms). Mit tRNAs are usually shorter than canonical tRNAs and have higher A-U base pair contents, which makes them less stable. Shortening usually affects D- and T-loops and the D-stem. How this influences the tertiary structure of such tRNAs is not completely understood. No crystal structure of a mit tRNA is known at present. However, modeling and structural probing results suggest that mit tRNAs manage to preserve the overall tRNA fold by means of different structural compensations and alternative interactions. For example, many mit $tRNAs^{Ser}_{GCU}$ lack the D-domain, which should substantially influence their tertiary structure. However, these tRNAs usually have longer anticodon stems (at least 9 base pairs), and structural modeling showed, that they can be folded into 'boomerang' structures (Fig. 6, page 13) with the bigger angle between acceptor and anticodon domains compared to the L- shape, but with a distance between the anticodon and the CCA end close to that in canonical tRNAs (Steinberg *et al*, 1994b).

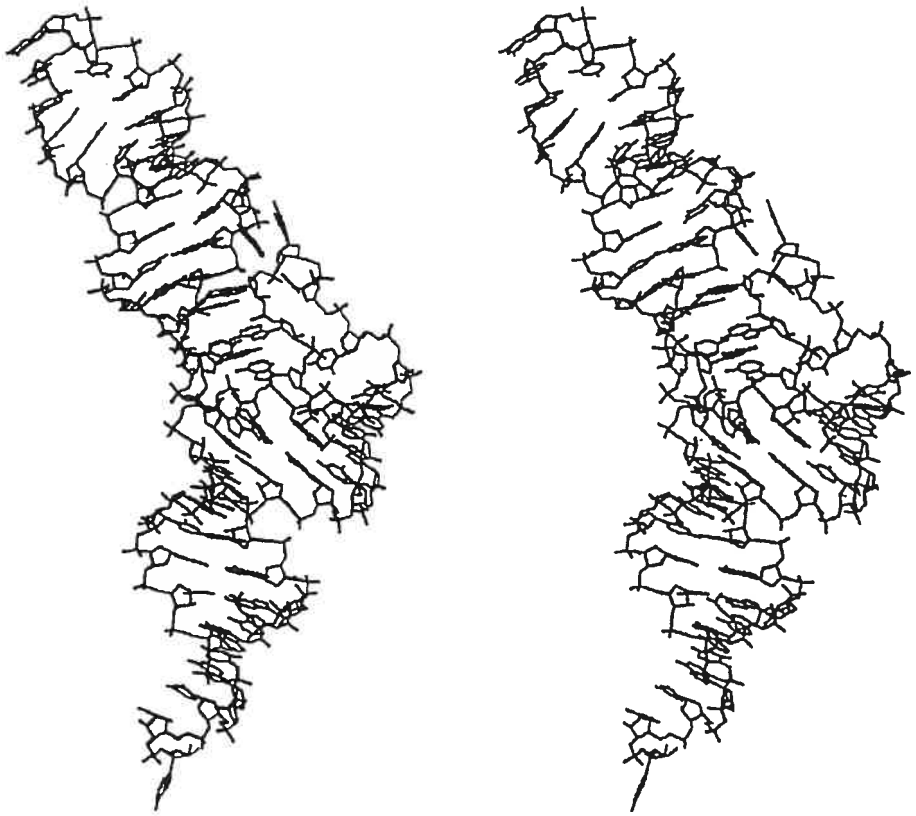


Figure 6. Stereo view of a 'boomerang' model of $tRNA^{Ser}_{GCU}$ from chimpanzee (Steinberg *et al*, 1994b). The angle between the two helical domains is much larger than in canonical tRNA, but the distance between the anticodon and acceptor end is similar to standard.

This finding suggests that alternative tRNA structures are possible, as long as some crucial parameters required for the tRNA function are conserved. Mit tRNAs may represent the molecules that fulfill minimal requirements for functioning in protein synthesis.

More generally, there is a correlation between the lengths of the D- and anticodon stems and the number of connector nucleotides in atypical mit tRNAs, which allows molecules with longer anticodon and shorter D-stems to fold into the same L-shaped tertiary structure, while the point of connection with the acceptor/T domain can be moved along the anticodon/D domain like in a zipper (Steinberg and Cedergren, 1994a). The general rule established from the analysis of atypical mit tRNAs states that the acceptor/D domain should contain 12 layers of stacked nucleotides to ensure its stacking with nucleotide 59 from the T-loop (Steinberg *et al*, 1997). Stacked layers can be made not only by Watson- Crick base pairs, but also by mismatch pairs and even by single unpaired nucleotides intercalated into a helical stem. There is also a restriction on the minimal length of the two connector regions in mit tRNAs: for N base pairs in the anticodon stem (N can vary from 5 to 10), the minimal length of connector 1 is 8-N and that of connector 2 is 9-N nucleotides.

Detailed structure of the DT (elbow) region

My thesis research was focused on the structure of the elbow region of tRNA. This region, which we also call the DT region, is where many tertiary interactions concentrate and where the two helical domains come together and interact through the D- and T-loops (see Fig. 2, page 4). It is known from the crystal structures of tRNA that two base pairs are formed between D- and T-loops: G18-Ψ55 (non-Watson-Crick) and G19-C56 (Watson-Crick), with the base of G18 from the D-loop being intercalated into the T-loop (Fig.7, page 15). There is an unusually large 'gap' between nucleotides 57 and 58, where G18 can fit. The dinucleotide 59-60 is bulged out from the T-loop and stacks to the tertiary base pair 15-48, the last layer of the D-domain. There is a reverse-Hoogsteen (RH) base pair between bases T54 and A58 in the T-loop, which is believed to stabilize the unique conformation of the T-loop, through its stacking to the base pair 53-61 of the T-stem and stabilization of hydrogen bond network from phosphate of nucleotide 60 to the amino group of C61 and ribose 58 (Romby *et al*, 1987). Conformation of the T-loop is also characterized by a U-turn (sharp turn in the sugar-phosphate backbone) between Ψ55 and

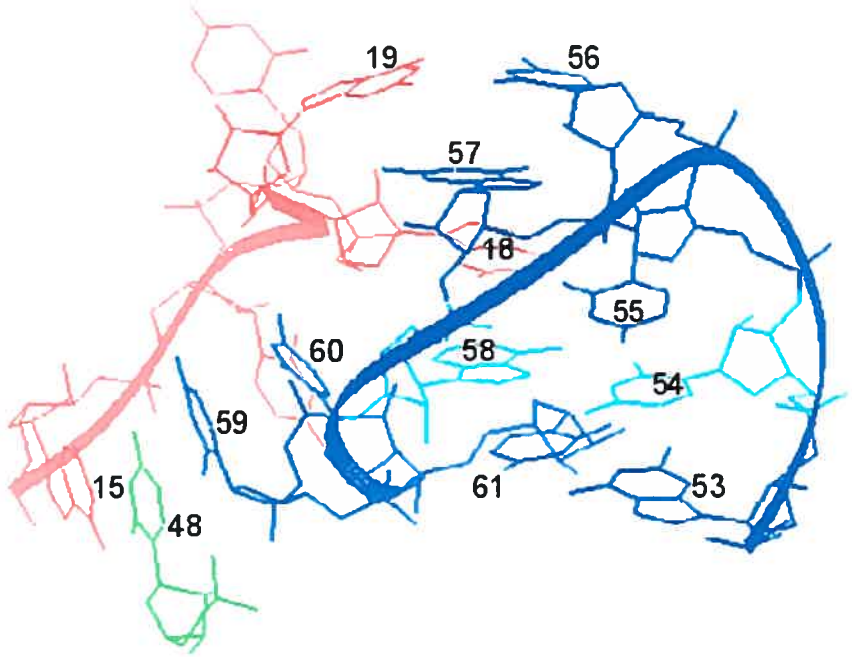


Fig.7. Structure of the DT region from yeast $tRNA^{Phe}$ (Shi and Moore, 2000, PDB entry 1EHZ). RNA backbone is traced as a ribbon. D-loop is colored red; the last base pair of the T-stem, G53-C61, and the T-loop are blue, with the reverse-Hoogsteen base pair T54-A58 in cyan; nucleotide 48 is green.

C56. All these interactions are believed to be common to all canonical tRNAs, since the nucleotide sequence of this region is very conserved. T-loop has the sequence $T_{54}\Psi_{55}C_{56}R_{57}A_{58}N_{59}Y_{60}$ (with base A58 often methylated) and is always closed by a G53-C61 base pair, while the D-loop has a conserved dinucleotide GG. The rest of the D-loop, as well as the position of the GG sequence, vary from one tRNA to another, but since the D-loop is flexible, it is believed that GG can be always arranged to interact with Ψ55 and C56.

Among canonical tRNAs, only eukaryotic initiator $tRNA^{Met}$ has some deviations from the standard in the structure of the T-loop. Thus, in yeast initiator $tRNA^{Met}$ T54 is replaced by adenine, nucleotide 55 is not modified to pseudouridine, and there is an adenine instead of usual dihydrouridine at position 20 in the D-loop. However, the crystal structure shows only minor differences from the canonical structure. G19-C56 and G18-U55 interactions are conserved, with G18 intercalating between bases 57 and 58, and the base pair A54-A58 is formed. An additional feature of yeast $tRNA^{Met}$ is that A20 interacts with base 57, unlike the canonical case. Such subtle differences in the structure of the T-loop may help distinguish initiator tRNA from elongator ones. Animal mitochondrial tRNAs do not always conserve canonical interactions between the T- and D-loops (for example, in mammalian mitochondria, only three tRNAs have them all; Helm *et al*, 2000). However, this does not mean that the T- and D- loops do not interact with each other in these tRNAs. It only means that they should have a different pattern of interaction.

It is believed that the T- loop of tRNA has its own intrinsic conformation, which is further stabilized by interactions with the D- loop (contrary to the D- loop, which does not have any particular conformation by itself). This is supported by studies in solution (Romby *et al*, 1987), as well as by recent crystal structures of tRNA complexes with pseudouridine synthase and ArcTGT.

Pseudouridine synthase is an enzyme that modifies the base of U55 in the T-loop into pseudouridine. This modification is highly conserved in all canonical tRNAs. Since Ψ55 is involved in a tertiary interaction with G18 from the D-loop, it is not readily accessible for the modification enzyme. Crystal structure of this enzyme complexed to the T-stem-loop RNA fragment reveals that upon binding to the pseudouridine synthase Ψ55 is flipped out from the T-loop into the protein binding pocket (Hoang and Ferre-D'Amare, 2001). In such conformation, formation of base pairs G18-Ψ55 and G19-C56 is not

possible. How this affects the L-shape is unknown, since only a fragment of tRNA was present in the complex. However, the structure of the T-loop is very similar to that in folded tRNA (even in the absence of modified bases): the reverse-Hoogsteen base pair T54-A58 is formed and nucleotides 59 and 60 are bulged. The differences exist in positions of bases 55, 56 and 57, which are extensively contacted by the enzyme, but this can result from 'unfolding', induced by protein binding. The authors conclude that the enzyme recognizes and binds a pre-formed T-loop structure.

In the ArcTGT complex described above (Ishitani *et al,* 2003), T-loop does not make any contact with the D-loop, since the D-arm is unfolded. Still, it almost perfectly preserves its folded conformation (Fig.4), with the only exception of nucleotide 57, which is bulged out. Remarkably, nucleotide 59 from the bulge, which is normally stacked to the tertiary base pair 15-48, still makes a stacking interaction, though now it is with base pair 23-48 of the newly formed DV helix.

However, the structure of the T-loop can be rearranged upon protein binding, as seen in the *Thermus thermophilus* $tRNA^{Phe}$-PheRS complex (Goldgur *et al*, 1997). In this complex, the synthetase makes contacts with the base pair 19-56 and with Ψ55. As a result, the reverse-Hoogsteen base pair T54-A58 is broken and instead A58 intercalates between T54 and Ψ55. G18 still makes a hydrogen bond with Ψ55, but is not intercalated between nucleotides 57 and 58, stacking only on G57. The rearranged structure is stabilized by interaction of nucleotide U59 with U16 from the D-loop.

Numerous examples of T-loop conformation have been recently identified in the ribosomal RNAs (one case in 5S, seven in 16S, and fifteen in 23S; Lee *et al*, 2003), as well as in the S domain of RNase P (two cases; Krasilnikov and Mondragon, 2003b). They all share the following structural features: a non-canonical base pair, three nucleotides in between, having similar conformation in all motifs, and bulged nucleotides or a sharp turn of the backbone at the 3' end of the motif. Interestingly, though a non-canonical base pair is in most cases a reverse-Hoogsteen U-A, as in tRNA, other base pairs are possible, like reverse-Hoogsteen G-G or C-A. Bulged nucleotides at the 3' end, as well as the nucleotides in between the base pair are often used to make tertiary contacts with other parts of the molecule. Thus, T-loop motif appears to be a stable, wide spread RNA motif suitable for providing long-range interactions in RNA.

The DT region plays a major structural role in the tRNA: it maintains the perpendicular arrangement of the two helical domains. This structural role of the DT region

is extensively discussed in the following chapters. However, this region is involved in the tRNA function through several other ways, being important for maturation of the 5' and 3' ends (Levinger *et al,* 1995 and 1998), recognition by some aminoacyl-tRNA synthetases (Puglisi *et al,* 1993; Du and Wang, 2003) and CCA-addition to the 3' end of the tRNA (Li *et al*, 1996).

Problems addressed in the thesis and an overview of the experimental method

In the literature review I have tried to illustrate two major principles: first, that it is important for tRNA to preserve its global tertiary fold, and second, that this goal can be achieved by different ways, giving a possibility for the existence of alternative structures. My thesis research has been focused on further investigation of these two aspects of tRNA structure.

Analysis of known RNA structures (not limited to tRNA) reveals that there are many common structural motifs widely used in RNA folding. Often, structurally similar motives share the same consensus sequence, like GNRA tetraloops, but there are also numerous examples of similar structures formed by very different sequences, as in the case of the T-loop motif mentioned above. Anyway, it is assumed that similar to protein structure, the tertiary structure of RNA is determined by its nucleotide sequence. However, predicting RNA tertiary fold from its primary sequence is presently an extremely challenging task, because the general rules of RNA folding are not well understood. We believe that unveiling these rules is an important scientific problem, solving which can help us better understand RNA structure and functions. We try to find such rules using tRNA as a model structural scaffold.

Transfer RNA is a good model system to study RNA structure for several reasons. First, its canonical L-shape structure is well known. Second, there is a lot of functional and structural information accumulated about tRNA itself and its complexes with other macromolecules. Third, several thousand tRNA sequences are known, both of canonical as well as non-canonical (mostly mitochondrial) tRNAs. Fourth, tRNA genes are relatively short and can be easily manipulated, including complete chemical synthesis. Finally, a large number of experimental systems has been developed to study different aspects of tRNA function, both *in vitro* and *in vivo*.

One powerful approach, which has proved useful, is the combinatorial approach. Initially, it has been developed *in vitro* and is known as SELEX (reviewed by Wilson and Szostak, 1999). It permitted to select RNA molecules with various binding properties (aptamers) from large pools of partially randomized sequences, called combinatorial libraries. Reverse transcription and PCR were used to amplify selected molecules, thus the pool could be enriched over several rounds of selection. Later this method was adapted for selection of RNA with catalytic activities (Lorsch and Szostak, 1996; Lee *et al*, 2000; Murakami *et al*, 2003). *In vitro* combinatorial approach has been successfully applied to tRNA, mostly for studying aminoacylation determinants (Peterson *et al*, 1993; Asahara *et al*, 1998; reviews by Vortler *et al,* 2001, and by Baskerville *et al*, 1998). However, combinatorial approach has not been used *in vivo* to study tRNA until recently. Our lab was the first to perform such experiments using the amber suppressor tRNA system, which proved to be an efficient tool for investigating the rules of tRNA structure formation (Bourdeau *et al*, 1998). Since then, several more studies using combinatorial tRNA libraries *in vivo* have been reported (Choi *et al,* 2003).

The combinatorial approach has several advantages over directed and random mutagenesis: first, much larger number of sequences can be screened simultaneously (*in vivo* limit is about 10^8 sequences due to cell transformation efficiency); second, no starting hypothesis about the role of particular nucleotides is required, instead, a whole region of a molecule can be targeted. In other words, the whole sequence and structural space, available for RNA with a given function, can be explored by this method.

The system that I have used allows selection of tRNA molecules that are functional as suppressor tRNAs in protein synthesis *in vivo*. Thus, selected tRNAs should be functional at each particular step of their metabolism, i.e. synthesis, processing, aminoacylation, translation. Compared to *in vitro* studies, *in vivo* selection criteria are usually more numerous and more strict, and they should allow to establish general rules imposed on tRNA sequence and structure in living cells.

In the present work, I applied this method to elucidate general rules that govern formation of an important region of tRNA molecule - the DT (elbow) region. Two different combinatorial libraries of tRNA genes containing randomized nucleotide positions in the D- and T- loops have been screened (they are referred to as K- and M-libraries). Analysis of selected suppressor tRNAs allowed us to identify what elements are primarily responsible for the folding of this region. They include the reverse-Hoogsteen base pair 54-

58 in the T-loop and the stacking of the T-loop bulge 59-60 to the D-domain of tRNA. Our approach also proved efficient in selecting sequences that have alternative types of interaction between the D- and T-loops. Our results contribute to the understanding of the role of different types of tertiary interactions in the formation of RNA structure and demonstrate that combinatorial approach can be successfully used *in vivo* to investigate the principles of RNA folding.

Chapter 1. Article:

Importance of the reverse Hoogsteen base pair 54-58 for tRNA function

Ekaterina I. Zagryadskaya, Felix R. Doyon and Sergey V. Steinberg

Département de Biochimie, Université de Montréal, Montréal, Québec H3C 3J7, Canada

Nucleic Acids Research, 2003, vol. 31, no. 14, pp. 3946-3953

Contribution of each author :

Ekaterina Zagriadskaia : developed detailed experimental design and did all experiments, participated in data analysis and preparation of manuscript and figures

Felix Doyon : did structural modeling, participated in preparation of figures

Sergey Steinberg : developed general experimental design, analyzed results, participated in structural modeling and preparation of manuscript and figures

ABSTRACT

To elucidate the general constraints imposed on the structure of the D and T-loops in functional tRNAs, active suppressor tRNAs were selected *in vivo* from a combinatorial tRNA gene library in which several nucleotide positions of these loops were randomized. Analysis of the nucleotide sequences of the selected clones demonstrates that among the randomized nucleotides, the most conservative are nucleotides 54 and 58 in the T-loop. In most cases, they make combination U54-A58, which allows the formation of the normal reverse-Hoogsteen base pair. Surprisingly, other clones have either combination G54-A58 or G54-G58. However, molecular modeling shows that these purine-purine base pairs can mimic very closely the reverse-Hoogsteen base pair U-A, and thus can replace it in the T-loop of a functional tRNA. This places the reverse-Hoogsteen base pair 54-58 as one of the most important structural aspects of the tRNA functionality. We suggest that the major role of this base pair is to preserve the conformation of dinucleotide 59-60, and through this, to maintain the general architecture of the tRNA L-form.

INTRODUCTION

One of the most conservative elements in the tRNA tertiary structure is the region at the outer corner of the tRNA L-form, where the T-loop interacts with the D-loop. This region, which we will henceforth call the DT region, is comprised of the whole T-loop, the first base pair of the T-stem 53-61, and nucleotides 18 and 19 of the D-loop, which interact respectively with nucleotides 55 and 56 of the T-loop (Fig.1). Out of 11 nucleotides of the DT region, only three, 57, 59 and 60, show a limited variability: 57 is always a purine, while 59 and 60 are pyrimidines in most cases (1). The other eight nucleotides of this region are invariable in cytosolic tRNAs. The DT region is involved in several important tRNA functions. First, it plays a major role in maintaining the perpendicular arrangement of the two helical domains called the L-form, which provides the proper juxtaposition of the two functional centers, the acceptor terminus and the anticodon. Also, this region is important for correct and efficient maturation of the termini of the molecule (2-4). Finally, it harbors recognition elements for interaction with different tRNA-binding enzymes, including some aminoacyl-tRNA synthetases (5-10).

The tertiary structure of the DT region is of special interest and has been the subject of a number of studies (11-14). The presence of such elements as the U-turn between Ψ55

and C56, the unusual non-Watson-Crick base pairs T54-A58 and G18-Ψ55, the mutual intercalation of fragments 57-58 and 18-19, the bulging of nucleotides 59-60, and the interaction of phosphate 60 with the amino group of C61 makes this region one of the most structurally diversified in the whole tRNA. This diversity raises questions concerning the role played by each of these elements in the structure of the DT region and of the whole molecule and the limits within which these elements could be modified without destroying tRNA structure and function. These questions become even more important in view of the recent finding that ribosomal RNA also contains motifs resembling the structure of the DT region (15). Thus, elucidation of the role and the sequence requirements for formation of the elements constituting this region in such a relatively small molecule as tRNA would contribute to understanding of structure-function relationships in other RNAs and RNA-protein complexes, including the ribosome. To address this problem, we undertook here an analysis of the general constraints imposed on the structures of the D and T-loops in a tRNA functioning *in vivo*. For this, we selected suppressor tRNAs from a specially designed combinatorial tRNA library in which a number of positions in the D and T-loops were randomized. Analysis of the nucleotide sequences of the successful tRNA clones sheds light on the role of particular elements of the DT region in the global tRNA structure.

MATERIALS AND METHODS

Strains

The *E.coli* strains TOP10 (Invitrogen) and XAC-1 (F' $lacI_{373}lacZ_{u118\,am}$ $proB^{+}/F^{-}$ $\Delta(lac\text{-}proB)_{x111}$ *nal*A *rif arg*E_{am} *ara*) were used respectively for cloning and selection of the suppressor tRNAs. XAC-1 strain contains amber mutations in genes *lac*Z and *arg*E (16).

Construction of the combinatorial library and selection of suppressor tRNAs

The template oligonucleotide coding for the combinatorial tRNA library (Fig. 2) was synthesized at BioCorp Inc (Montreal, Canada), amplified by PCR to produce the double-stranded DNA using primers 5'GCGAATTCGGGGCTATA3' and 5'GACTGCAGTGGTGGAGT3', and cloned into plasmid pGFIB-1 using *Eco*RI and *Pst*I restriction sites, as described previously (17). This plasmid provides a constitutive high-level expression of a cloned tRNA gene (18). All enzymes were from New England Biolabs. Of 20 μl of the ligation mixture, 5 μl were electroporated into the competent

TOP10 cells, providing 4.5×10^6 colonies, i.e. about a quarter of the sequence complexity of the library. The plasmid DNA was recovered using the Qiafilter Midiprep kit (Qiagen) and then transformed into the competent XAC-1 cells. The positive clones were selected as blue colonies when grown on the LB-agar containing ampicillin (100 μg/ml) and X-Gal (200 μl of the 20 mg/ml solution spread on top of each 150×15 mm plate). The plasmid DNA of these clones was extracted and retransformed into the XAC-1 cells to confirm the dependence of the phenotype on the presence of the plasmid. The ability of the selected tRNAs to suppress the nonsense mutation in gene *argE* was checked by plating the retransformed XAC-1 cells on the minimal A media without arginine.

Sequencing

Sequencing of the selected tRNA genes was performed on the LI-COR DNA sequencing system (Département de Biochimie, Université de Montréal) using primers 5'-GCTTCTTTGAGCGAACGATCAAAAATAAGT-3' and 5'-GGGTTTTCCCAGTCACGACGTTGTAAAACG-3' labeled at 5' with IRDye 800 (LI-COR Biosciences).

Measurements of the β-galactosidase activity

β-galactosidase activity of clones with suppressor tRNA genes was determined as described by Miller (19) using overnight cultures grown in the A medium containing 0.4% glucose and 1 mM $MgSO_4$ to an A_{600} of 0.8-0.9. The values were obtained by averaging the measurements from three independent cultures and calculated as a percentage of the activity of the control $tRNA^{Ala}_{su+}$.

Presence of the suppressor tRNAs in the cytosol and their aminoacylation level

To preserve the aminoacylated form of the tRNAs, the total cellular RNA was extracted under acidic conditions, as described previously (17). To obtain the deacylated tRNA, 4 μg of the total RNA was mixed with 1.5 μl of 0.5 M Tris (pH 9.0), incubated for 30 min at 37°C, and deposited on the acid polyacrylamide gel (6.5 % polyacrylamide, 8M urea, 0.1 M sodium acetate) together with the untreated fraction. The gel was run for 24 hours at 300 V at 4°C in 0.1 M sodium acetate, after which the part of the gel around the xylene cyanol dye was transferred by electroblotting to a Hybond-N nylon membrane

(Amersham). The membrane was hybridized with two radiolabeled DNA probes, one complementary to region 26-44 of the suppressor tRNAs consisting of the anticodon stem and loop, and the other to region 34-53 of the *E.coli* 5S rRNA. The 5S rRNA probe was used to monitor the amount of total RNA in each sample. The hybridization was performed overnight at 37°C in [7% SDS, 0.25 M Na_2HPO_4 (pH 7.4), 1mM EDTA (pH 8.0), 1% BSA] using the Robbins hybridization incubator.

Computer modeling

Preliminary modeling was done interactively, using InsightII/Discover package (Version 2000, Accelrys Inc., San Diego, CA). The X-ray structure of the yeast $tRNA^{Phe}$ (20) was used as a starting conformation. The presumed structures of RH-GA or RH-GG were appended to the T-stem replacing base pair U54-A58. The other randomized nucleotides were arranged in a way to resemble the structure of the DT region in the normal tRNAs and, at the same time, to provide a reasonable system of H-bonds and base-base stacking interactions. Each model was submitted to unrestrained energy minimization in vacuum using the AMBER forcefield (21) until an energy minimum was reached. Visualizations were done in a Silicon Graphics O2 computer.

RESULTS

The library design

The library was built from *E.coli* $tRNA^{Ala}_{UGC}$ as a scaffold (Fig. 2). The choice was determined by the fact that the most important $tRNA^{Ala}$ identity element for the cognate alanyl-tRNA synthetase, the G3-U70 base pair, was located in the acceptor stem, that is neither at the DT region, nor in the anticodon, the sites that were modified in this study (22, 23). This would minimize the role of interaction with a particular aminoacyl-tRNA synthetase as a factor in the tRNA selection. To enable the tRNAs to recognize the amber stop codon UAG, the anticodon TGC in the gene was replaced by CTA. All five nucleotides of region 54-58 of the T-loop, which were known to be involved in conserved interactions either within the loop or with nucleotides of the D-loop, were fully randomized. Correspondingly, four nucleotides 16-19 of the D-loop, which could be involved in interactions with the T-loop, were also fully randomized. To prevent nucleotide G20 from substituting either G18 or G19 in their interactions with the nucleotides of the T-

loop, it was replaced by T20. To stimulate the formation of alternative structural patterns in the DT region, we added one and two nucleotides to the randomized regions of the T and D-loop, respectively. Thus, in the design, the T-loop contained eight nucleotides, one more than in the standard tRNA structure, while the D-loop had ten nucleotides, which is not unprecedented for the cytosolic tRNAs (1). Each loop had six randomized positions, providing for the total sequence complexity of a library of 1.7×10^7 variants.

Cloning and selection of functional clones

The tRNA gene library was synthesized chemically, amplified by PCR, and cloned into the pGFIB-1 plasmid, as described previously (17). The selection of active suppressor tRNA clones was done in the XAC-1 strain of *E.coli*, which had nonsense amber mutations in genes *lacZ* and *argE*. A successful suppression of the first mutation in presence of 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-Gal) provides blue colonies, which was used for the primary identification of functional tRNA clones. Out of 3×10^4 clones screened, several dozens positive clones were selected, whose suppressor activity was confirmed by a subsequent retransformation and by a suppression of the second mutation in gene *argE* that converts the arginine-dependent cells into prototrophs. The β-galactosidase activity was evaluated quantitatively for each clone and compared to that of the control $tRNA^{Ala}_{su+}$. The latter tRNA was derived from the normal $tRNA^{Ala}$ by changing the anticodon from UGC to CUA, and cloned in the same plasmid as the other suppressor tRNAs.

The nucleotide sequences of the selected tRNA clones, as deduced from their genes, are presented in Table I. Only the sequences of those clones whose activity was at least 1% of the control are given. Comparison to the original design revealed six clones with a nucleotide deletion in the T-loop and three clones with a deletion in the D-loop, providing respectively for a seven-member T-loop or a nine-member D-loop. In two clones, K25 and K30, mutations affected the non-randomized part of the T-loop, deleting respectively U59 and C60. No other mutation outside the randomized regions was found. For eight clones arbitrarily chosen from Table I, the *in vivo* level of the suppressor tRNA and of its aminoacylated form was determined by acid polyacrylamide gel electrophoresis followed by hybridization with a specific probe complementary to the anticodon stem and loop. For all suppressor tRNAs tested, the level in cytosol was detectable although relatively low

compared to that of $tRNA^{Ala}_{su+}$ (Fig. 3). For each clone, most of the tRNA was found in the aminoacylated form.

Analysis of the nucleotide sequences

In the experiments described above, on average, only one in every 1,000 clones showed a detectable level of the nonsense suppression activity. This indicated that the nucleotide sequence space available for the DT region was rather small. A systematic analysis of the sequences of the selected clones could help to reveal the rules imposed on the structure of this region in functional tRNAs.

We started the analysis with the "quasi-normal" clones, those that had the normal seven nucleotides in the T-loop. Henceforth, we will call such molecules T7-tRNAs, in contrast to T8-tRNAs having eight nucleotides in this loop. Analysis showed that in T7-tRNAs, the fifth position of the T-loop was always occupied by A and was the only invariable position in both randomized regions (Table I). The second most conservative nucleotide was the first one of the T-loop, which in all sequences except K6 was U. The presence of U and A respectively in the first and the fifth position of the T-loop would allow the formation of the reverse-Hoogsteen base pair U54-A58 (RH-UA), as in the normal tRNAs. Although in the normal tRNAs U54 is always modified to T, it is not yet known whether it is the case in the selected tRNAs. On the other hand, because this modification does not interfere with the ability of the base to form H-bonds, its absence would not affect the formation of base pair U54-A58. Another conservative feature consistent with wild-type tRNA is the presence of a purine in position 57 of all but one T7-tRNAs. Other randomized nucleotides, including all six nucleotides in the D-loop, were essentially more diversified and did not seem to provide for a common structural pattern.

Among T8-tRNAs, half of the sequences (11 out of 22) also contained U in the first position of the T-loop (Table I). If this U plays the same role as it does in T7-tRNAs, there should be an A few nucleotides later that is able to form RH-UA with this U. Generally, this A could occupy either the fifth or the sixth position of the T-loop, depending on the position of the additional eighth nucleotide. Analysis showed that in those T8-tRNAs whose T-loop started with U, the only other conservative nucleotide was the A occupying the sixth position of the same loop. Thus, the formation of RH-UA of these two nucleotides would place the additional nucleotide in the region between them.

In all other T8-tRNAs, the first position of the T-loop was occupied by G (Table I). If this G plays a structural role analogous to that played by U, its possible partner would occupy either the fifth or the sixth position of the T-loop. None of the two positions were conserved in these sequences: the fifth nucleotide was allowed to be either C or A, while the sixth was either A or G. To explore the abilities of both the fifth and the sixth nucleotides to pair with the first G, we looked for possible arrangements of three different dinucleotide combinations, GC, GA and GG that would be close to the arrangement of U and A in RH-UA. For GC, we did not find any satisfactory arrangement. However, for both combinations GA and GG we found arrangements that are presented in Fig. 4. In these arrangements, the G that is equivalent to U in RH-UA donates two hydrogen atoms for formation of H-bonds with atom N7 of the other purine. This purine can be either A or G. In the latter case, an additional H-bond can be formed between N2-H and O6 of the first and the second G, respectively. The two arrangements GA and GG are superimposable in the sense that if one overlaps the positions of the glycosidic bonds of the first nucleotides, the glycosidic bonds of the second nucleotides in both arrangements would occupy about the same position. In the same sense, these two arrangements are fairly close to RH-UA. Accommodation of any of these arrangements based on the standard RH-UA would require a shift and rotation of one of the bases by only 1.5 Å and 20°, respectively. Therefore, a replacement of RH-UA in the T-loop by either GA or GG would require only relatively minor changes in the position of the neighboring nucleotides. To reflect the closeness of these GA and GG arrangements to RH-UA, we will call them RH-GA and RH-GG, respectively.

Further analysis revealed few additional nucleotide combinations like CA and AA seen in Fig. 4 that could also be arranged relatively closely to RH-UA while having a reasonable system of H-bonds. Still, all these additional combinations were more distant from RH-UA than RH-GA or RH-GG and therefore, their incorporation into the T-loop instead of RH-UA would cause greater changes in the conformation of the whole DT region. This latter aspect was expected to render these combinations less preferable in this place than RH-GA or RH-GG.

The fact that GG and GA can be accommodated close to RH-UA, while GC cannot, makes the sixth rather that the fifth nucleotide of the T-loop in T8-tRNAs the most probable partner to form a base pair with the G occupying the first position of this loop. As to the other randomized nucleotides in both loops of T8-tRNAs, they , like in T7-tRNAs,

were much more diversified and did not seem to provide for a common structural pattern. Finally we can consider K6, the only T7-tRNA clone, whose T-loop starts with G rather than with U. This clone has also A in the fifth position of the T-loop, which would allow these G and A to form RH-GA, analogous to RH-UA existing in all other T7-tRNAs.

From this analysis, a clear picture emerges: in all selected tRNAs the first and the last randomized positions of the T-loop are always able to form a RH base pair, i.e. either RH-UA, RH-GA or RH-GG. The last randomized position is either the fifth in the T7-tRNAs or the sixth in the T8-tRNAs. The region between the first and the last position varies in length and sequence and does not seem to have a common pattern.

Modeling of the tRNA structures

To confirm that the exchange of RH-UA for either RH-GA or RH-GG in the T-loop did not cause any steric problem, we modeled the structure of the DT region for several clones having either RH-GA or RH-GG. After unrestrained energy minimization, the bases constituting the RH base pair always retained their juxtapositions and the inter-base H-bonds, as one can see in the example of the model for clone K31 (Fig. 5). Comparison of the models with the structure of the yeast $tRNA^{Phe}$ (20) showed that the whole region that included the T-stem, the RH base pair (RH-GA or RH-GG in the models and RH-UA in $tRNA^{Phe}$), as well as nucleotides 59 and 60 was superimposable in all structures.

DISCUSSION

The results presented here show that a tRNA could be functional even if the structure of its DT region is substantially modified compared to the standard. Although for all selected clones the efficiency of the nonsense codon suppression was lower than for the $tRNA^{Ala}_{su+}$, it was strong enough to provide a level of the β-galactosidase synthesis sufficient to change the color of the colonies in the presence of X-Gal and to allow cell growth without external arginine. Additional examination of several clones showed that suppressor tRNAs had a detectable *in vivo* level and existed mainly in the aminoacylated form.

The nucleotide sequences of the selected suppressor tRNAs demonstrated a range of diversity never seen in the natural cytosolic tRNAs. In spite of this, the selected tRNAs constituted only a tiny fraction of the whole tRNA gene library, which implied the

existence of strong constraints imposed on the structure of functional tRNAs. To elucidate these constraints, we undertook a comparative analysis of the nucleotide sequences of the selected tRNAs. It may be a little surprising that among the selected clones there were no clones having the wild-type sequence pattern. On the other hand, the wild-type sequence G18-G19-…-U54-U55-C56-R57-A58 is expected to appear on average only once in 8 000 clones. Moreover, this probability can easily get beyond the technically detectable level if some additional requirements are imposed on the identity of the nucleotides flanking the conservative dinucleotide G18-G19 in the D-loop and on the additional eighth nucleotide in the T-loop. For most of the randomized nucleotides, our analysis did not find any obvious common pattern. The only exception consisted of the first and the last randomized positions in the T-loop, which were always able either to form RH-UA, analogous to base pair U54-A58 in the normal tRNAs, or to mimic it closely via formation of RH-GA or RH-GG. Modeling experiments showed that a replacement of RH-UA with either RH-GA or RH-GG did not cause any major rearrangement in the conformation of the DT region and provided for stable, sterically reasonable tRNA structures. Because such an RH base pair can be formed in all selected tRNAs, its existence is judged to be one of the most important requirements imposed on the structure of the DT region in a functional tRNA. In fact, this requirement has been the only one satisfied in all selected tRNAs, which allows us to conclude that the preservation of an RH base pair in the T-loop is more important for the tRNA function than that of other universal elements including inter-loop base pairs G18-Ψ55 and G19-C56.

Different explanations of the importance of the RH base pair for the tRNA function can be suggested. For example, this base pair could be involved in a specific, vitally important interaction with a protein or other factor and thus should be preserved as such. However, a specific interaction like this would probably not tolerate an exchange of RH-UA for either RH-GA or RH-GG, because the juxtaposition of the glycosidic bonds and therefore, the conformation of the backbone in the two latter base pairs, however close it is to that in RH-UA, is still notably different. Moreover, the three base pairs have different chemical groups exposed on the surface and thus can hardly be recognized by the same factor. In another, more probable explanation, an RH base pair is needed to stabilize a particular conformation of a neighboring region in the tRNA structure and thus to enable this region to serve its function. We do not expect this region to include the top of the T-loop closed by the RH base pair or the randomized part of the D-loop. Indeed, in the

selected tRNAs, these regions vary in length and in nucleotide sequence and do not seem to have a universal structure. There is, however, another region, dinucleotide 59-60 at the end of the T-loop, whose position may need a particular structure of the RH base pair. This dinucleotide bulges from the double helical stem between base pairs 54-58 and 53-61. Therefore, its conformation is to a great extent determined by the positions of the flanking nucleotides 58 and 61, which, in turn, depend on the structures of base pairs 54-58 and 53-61, respectively. Thus, the presence of base pair 54-58 with the correct juxtaposition of the bases is necessary for the proper positioning of dinucleotide 59-60. The reverse-Hoogsteen base pair U54-A58 perfectly suits this purpose. However, as our results show, combinations GA and GG can also be arranged in an appropriate way. In the modeled tRNAs containing these combinations, the position of nucleotide 59-60 remains virtually the same as in the standard tRNA structure (Fig. 5).

We have already suggested (24) that the bulged dinucleotide 59-60 plays a crucial role in maintenance of the general shape of tRNA. Indeed, in the normal tRNAs, nucleotide 59 stacks to the tertiary base pair 15-48, which constitutes the last stacked layer of the D/anticodon helical domain (Fig. 1, 5). This interaction determines the juxtaposition of the two domains, the D/anticodon and the acceptor/T, i.e. the general geometry of the tRNA L-form, and is invariable in all normal tRNAs (24, 25). Thus, the importance of maintaining the standard juxtaposition of the helical domains within the tRNA L-form would justify the necessity to preserve the conformation of dinucleotide 59-60 via formation of the reverse-Hoogsteen base pair 54-58. Because the bases of nucleotides 59 and 60 are mainly involved in stacking interactions between themselves and with nucleotides 15 and 48, and not in H-bonding, their identity is not that important. This reflects the partial variability of nucleotides 59 and 60 in the normal tRNAs (1), and also fits to the fact that replacements U59A in clones K25 and K30 and C60U in clone K30 did not impair the tRNA function (Table I).

The importance of this base pair for tRNA function, especially comparing to other randomized elements of the DT region, correlates well with the data of Nazarenko *et al* (26) on the efficiencies of mutants of the yeast Phe-tRNA^{Phe} in different partial reactions of the tRNA functional cycle. According to these data, the efficiencies of the mutants in the tertiary complex formation with factor Tu and GTP, in binding to the A and P site of the poly(U)-programmed ribosome and in peptide formation are generally more sensitive to nucleotide replacements in pair 54-58 than in pairs 18-55 and 19-56. There have also been

reports of using tRNA libraries with randomized positions in the DT region for selection of clones *in vitro* by affinity to either the phenylalanyl or glutamyl-tRNA synthetase and to the EF-Tu factor (27, 28). Interestingly, in none of these studies the necessity for an RH base pair in the T-loop has been detected. However, because different steps of the tRNA functional cycle are probably not equally dependent on the proper position of the two helical domains, concentration on only some steps of this cycle would not necessarily reveal all sequence requirements for a fully functional tRNA. Moreover, the mutations in the tRNA clones selected by affinity to a particular protein may be detrimental not only for other steps of the functional cycle not involved in this selection, but even for this very step, if, for example, they hinder the dissociation of the complex (29, 30). We used here an alternative approach of tRNA selection *in vivo* based on its suppressor activity. This guarantees that the selected clones are correctly transcribed, processed and folded, that they are able to interact productively with all necessary factors of the protein biosynthesis machinery, including the aminoacyl-tRNA synthetase, EF-Tu, as well as the UAG-charged ribosome. The sequence requirements revealed in this way have been balanced between all these steps.

Another essential aspect of our approach relates to the original design of the library. As one can see in Table I, among T7-tRNAs only clone K6 does not have RH-UA, while among T8-tRNAs there are more than half of such clones. It may reflect the existence of certain constraints on fitting either RH-GA or RH-GG into the seven-member T-loop, which would relax when an additional eighth nucleotide is added. Therefore, if in the T-loop of the original design the normal seven nucleotides were preserved, most probably, we would have seen an overwhelming majority of the selected clones having RH-UA only sporadically intermingling with those having RH-GA or RH-GG. This would have made it more difficult to understand that it is the conformation of the sugar-phosphate backbone rather than base pair U54-A58 *per se* which is crucial for the tRNA function. Although the introduction of an additional base into the T-loop has probably some negative effect on the suppressor activity, most importantly, it increases the chance of selecting alternative structures. The elucidation of the common structural pattern in the selected tRNAs designed in this way has allowed us to recognize the RH base pair as the most important aspect that must be preserved, even when the structure substantially deviates from the standard.

In contrast to the nucleotides composing the RH base pair, other randomized nucleotides do not seem to have a common structural pattern. This means that the existence of a particular universal structure of this region is not required for tRNA function. On the other hand, it does not mean that this region is not structured or does not play any functional role. Instead, there are indications that it has a particular structure and that this structure is important for the tRNA function, even if it is not the same in all selected tRNAs. First, as discussed above, the seven-member T-loop may limit the use of RH-GA and RH-GG, so that these latter base pairs are able to replace RH-UA only in a fraction of all successful T7-tRNAs. This ability will thus depend on the identity of other randomized nucleotides. Also, the activities among the selected clones differ by almost forty times, despite the presence of a RH base pair in all of them. This indicates the existence of other factors within the tRNA structure that affect the activity. Because these clones differ one from another only in the randomized regions, we have to conclude that other randomized nucleotides not involved in the RH base pair play a role in the tRNA function. Still, the overall inspection of the nucleotide sequences shown in Table I has provided no obvious consensus pattern able to explain the differences in the activity. In the normal tRNAs, the randomized nucleotides not involved in the RH base pair take part in the interactions between the D and T-loops. These interactions affect the overall stability of the tRNA tertiary structure, but may also be directly involved in a particular tRNA-related process. Similar interactions could exist in the selected suppressor tRNAs as well, although they would be different in different clones. An example of such interactions is seen in the model of K31 in Fig. 5. A systematic analysis of the possibility to form these interactions in the selected tRNA clones is now in progress and will be published elsewhere. One should admit, however, that this analysis may not be able to explain the existing differences in the tRNA activity. Indeed, there are many reasons that could affect negatively the tRNA activity, like formation of an alternative secondary structure, a higher susceptibility to a nuclease activity, a too low or too high affinity to the cognate aminoacyl-tRNA synthetase or to the elongation factor Tu, etc., and for each clone, the real reason can be different. A complete understanding of this phenomenon needs the analysis of the behavior of individual tRNA clones at each step of their functional cycle.

ACKNOWLEDGMENTS

We thank Drs. L. Brakier-Gingras and S.W. Michnick for critical reading of the manuscript. S.V.S. acknowledges grants from National Science and Engineering Research Council of Canada and from Human Frontiers Science Program and fellowships from Canadian Institutes of Health Research and from Fond de la Recherche en Santé du Québec.

REFERENCES

1. Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A. and Steinberg, S. (1998) Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.*, **26**, 148-153.
2. Tuohy, T.M., Li, Z., Atkins, J.F. and Deutscher, M.P. (1994) A functional mutant of tRNA(2Arg) with ten extra nucleotides in its TΨC arm. *J. Mol. Biol.*, **235**, 1369-1376.
3. Altman, S., Kirsebom, L. and Talbot, S. (1995) Recent studies of RNase P. In Söll,D. and RajBhandary,U. (eds.), *tRNA: Structure, Biosynthesis, and Function.* American Society for Microbiology, Washington, DC, pp.67-78.
4. Nashimoto, M., Tamura, M. and Kaspar, R.L. (1999) Minimum requirements for substrates of mammalian tRNA 3` processing endoribonuclease. *Biochemistry*, **38**, 12089-12096.
5. McClain, W.H. and Foss, K. (1988) Changing the acceptor identity of a transfer RNA by altering nucleotides in a "variable pocket". *Science*, **241**, 1804-1807.
6. McClain, W.H., Foss, K., Jenkins, R.A. and Schneider, J. (1991) Four sites in the acceptor helix and one site in the variable pocket of $tRNA^{Ala}$ determine the molecule's acceptor identity. *Proc. Natl Acad. Sci. USA*, **88**, 9272-9276.
7. Peterson, E.T. and Uhlenbeck, O.C. (1992) Determination of recognition nucleotides for *Escherichia coli* phenylalanyl-tRNA synthetase. *Biochemistry*, **31**, 10380-10389.
8. Biou, V., Yaremchuk, A., Tukalo, M. and Cusack, S. (1994) The 2.9 Å crystal structure of *T. thermophilus* seryl-tRNA synthetase complexed with $tRNA^{Ser}$. *Science,* **263**, 1404-1410.

9. Pleiss, J.A, Wolfson, A.D. and Uhlenbeck, O.C. (2000) Mapping contacts between Escherichia coli alanyl tRNA synthetase and 2` hydroxyls using a complete tRNA molecule. *Biochemistry*, **39**, 8250-8258.

10. Nomanbhoy, T., Morales, A.J., Abraham, A.T., Vortler, C.S., Giege, R. and Schimmel, P. (2001) Simultaneous binding of two proteins to opposite sides of a single transfer RNA. *Nat. Struct. Biol.*, **8**, 344-348.

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

12. de Bruijn, M.H.L. and Klug, A. (1983) A model for the tertiary structure of mammalian mitochondrial transfer RNAs lacking the entire "dihydrouridine" loop and stem. *EMBO J.* **2**, 1309-1321.

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

14. Ogata, H., Akiyama, Y., and Kanehisa, M. (1995) A genetic algorithm based molecular modeling technique for RNA stem-loop structures. *Nucl. Acids Res.* **23**, 419-426

15. Nagaswamy, U. and Fox, G.E. (2002) Frequent occurrence of the T-loop RNA folding motif in ribosomal RNAs. *RNA*, **8**, 1112-1119.

16. Normanly, J., Masson, J.M., Kleina, L.G., Abelson, J. and Miller, J.H. (1986) Construction of two Escherichia coli amber suppressor genes: tRNAPheCUA and tRNACysCUA. *Proc Natl Acad Sci USA*, **83**, 6548-6552.

17. Bourdeau, V., Steinberg, S.V., Ferbeyre, G., Emond, R., Cermakian, N. and Cedergren, R. (1998) Amber suppression in *Escherichia coli* by unusual mitochondria-like transfer RNAs. *Proc. Natl. Acad. Sci. USA*, **95**, 1375-1380.

18. Masson, J.M. and Miller, J.H. (1986) Expression of synthetic suppressor tRNA genes under the control of a synthetic promotor. *Gene*, **47**, 179-183.

19. Miller, J.H. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor Lab. Press, Plainview, NY, pp.352-355.

20. Shi, H. and Moore, P.B. (2000) The crystal structure of yeast phenylalanine tRNA at 1.93 A resolution: a classic structure revisited. RNA, **6**,1091-1105.

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

22. McClain, W.H. and Foss, K. (1988) Changing the identity of a tRNA by introducing a G-U wobble pair near the 3' acceptor end. *Science*, **240**, 793-796.

23. Hou, Y.M. and Schimmel, P. (1988) A simple structural feature is a major determinant of the identity of a transfer RNA. *Nature*, **333**, 140-145.

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

25. Ioudovitch, A. and Steinberg, S.V. (1999) Structural compensation in an archaeal selenocysteine transfer RNA. *J. Mol. Biol.*, **290**, 365-371.

26. Nazarenko, I.A., Harrington, K.M. and Uhlenbeck, O.C. (1994) Many of the conserved nucleotides of $tRNA^{Phe}$ are not essential for ternary complex formation and peptide elongation. *EMBO J.*, **13**, 2464-2471.

27. Peterson, E.T., Blank, J., Sprinzl, M. and Uhlenbeck, O.C. (1993) Selection for active *E.coli* $tRNA^{Phe}$ variants from a randomized library using two proteins. *EMBO J.*, **12**, 2959-2967.

28. Bullock, T.L., Sherlin, L.D. and Perona, J.J. (2000) Tertiary core rearrangements in a tight binding transfer RNA aptamer. *Nat. Struct. Biol.*, **7**, 497-504.

29. LaRiviere, F.J., Wolfson, A.D. and Uhlenbeck, O.C. (2001) Uniform binding of aminoacyl-tRNAs to elongation factor Tu by thermodynamic compensation. *Science*, **294**, 165-168.

30. Asahara, H. and Uhlenbeck, O.C. (2002) The tRNA specificity of *Thermus thermophilus* EF-Tu. *Proc. Natl Acad. Sci. USA*, **99**, 3499-3504.

FIGURE LEGENDS

Figure 1. The standard tRNA L-form.

Rectangles represent individual nucleotides. The DT region at the outer corner of the molecule is boxed. Crosshatched and filled rectangles represent nucleotides of the D and T-loop, respectively. Unpaired nucleotides as well as nucleotides at the beginnings and the ends of the helical regions are numbered in accordance with the standard tRNA nomenclature (Sprinzl *et al*, 1998). Nucleotides of the anticodon loop, non-stacked nucleotides of the D-loop, and nucleotide 47 are not shown. There are two base pairs G18-Ψ55 and G19-C56 formed between the D and T-loops. The reverse-Hoogsteen base pair T54-A58, whose structure is seen in Fig. 4, is formed within the T-loop. Dinucleotide 59-60 bulges from the double helical stem between base pairs G53-C61 and T54-A58. Nucleotide 59 stacks to the tertiary base pair 15-48 constituting the last layer of the D/anticodon helical domain. This interaction fixes the perpendicular arrangement of the two helical domains called L-form.

Figure 2. Construction of the tRNA gene library.

In the nucleotide sequence of the *E. coli* $tRNA^{Ala}_{UGC}$, each of the two enclosed regions, 16-19 in the D-loop and 54-58 in the T-loop, was replaced by six fully randomized positions, while nucleotide G20 and the anticodon TGC (boxed) were replaced by T20 and CTA, respectively. Nucleotides 54 and 58, which form the reverse-Hoogsteen base pair in the T-loop, are connected by a line. The EcoRI and PstI restriction sites that are seen flanking the 5' and 3' termini were used for cloning the library into the pGFIB-1 plasmid.

Figure 3. Northern blot showing the presence in the cytosol and the level of aminoacylation of some suppressor tRNAs.

For each clone, the "-" and "+" lanes correspond to the samples not treated and treated with Tris. In the "-" lanes the aminoacylated and deacylated forms of the suppressor tRNA move as individual bands, while in the "+" lanes the total tRNA is deacylated and the suppressor tRNA moves as one band. In all "-" lanes, the bands corresponding to the aminoacylated form of the tRNA are much larger than those corresponding to the deacylated form and are comparable to the bands in the "+" lanes, representing the total amount of the suppressor tRNA. This indicates that in all clones, most of the tRNA is present in the aminoacylated

form. A smaller size of the bands of the suppressor tRNAs compared to $tRNA^{Ala}_{su+}$ indicates a notably lower presence of the selected tRNAs in the cytosol. 5S rRNA was visualized to monitor the amount of total RNA in each sample. Because the signal from 5S rRNA was much stronger than that from suppressor tRNAs, the upper and lower parts of the same membrane have been exposed, respectively, for 4 hours and overnight. The nucleotide sequence of clone K8, due to its low β-galactosidase activity, is not included in Table I and is available upon request.

Figure 4. Juxtaposition of the bases in RH-GA, RH-GG and other alternative base pair candidates for replacement of RH-UA.

A: Positions of the glycosidic bonds in the alternative base pairs comparing to that in RH-UA. In each base pair, the position of the glycosidic bond corresponding to the base on the right is superimposed on that of U in RH-UA. The glycosidic bond of the other nucleotide will thus occupy a particular place depending on the structure of the base pair. The numbers indicating particular positions of the glycosidic bonds correspond to the base pairs in part B.

Figure 5. The model of the structure of the DT region for clone K31 (red) superimposed on the corresponding region in the yeast $tRNA^{Phe}$ (green).

The figure also includes the T-stem and the tertiary base pair 15-48. For both tRNAs, the ribbon follows the sugar-phosphate backbone. Explicitly shown are base pairs 15-48, 54-58 and nucleotide 59 in $tRNA^{Phe}$, as well as all nucleotides of the DT region and pair 15-48 in K31. Comparison of the modeled structure with $tRNA^{Phe}$ demonstrates a good superposition of the T-stem, base pairs RH and 15-48 as well as nucleotide 59. The proper arrangements of the nucleotides in the RH base pair thus guarantees the proper position of nucleotide 59, whose stacking to base pair 15-48 would fix the juxtaposition of the two helical domains known as the L-form. Still, one can notice a difference in the conformation of the backbone in the two structures, which is the highest between nucleotides 58 and 59. Such a difference makes a universal interaction of this region with a particular protein factor unlikely.

Table I. The nucleotide sequences and the β-galactosidase activity of the selected tRNA clones

Clone	D-loop	T-loop	% β-gal activity
$tRNA^{Ala}_{su+}$	AGCUGGGA	UUCGAUC	100
T7-tRNAs			
	16 19	54 58	
K30	AGUGAGGAUA	UCCAAAU	10.8 ± 1.1
K25	AGGAACGCUA	UGAAAAC	17.7 ± 6.2
K3	AGAACGAAUA	UGAAAUC	4.1 ± 0.3
K15	AGGCAUAUUA	UGAAAUC	11.0 ± 1.4
K29	AGGAAAAAUA	UGGGAUC	5.1 ± 1.0
K6	AGAGGGAGUA	GCACAUC	25.0 ± 5.0
T8-tRNAs			
	16 19	54 58	
K26	AGAACGACUA	UAAACAUC	3.9 ± 0.8
K18	AGAACAAAUA	UAAACAUC	2.5 ± 0.6
K1	AGGAGAACUA	UAACCAUC	1.3 ± 0.1
K7	AGGACAAAUA	UAACCAUC	1.3 ± 0.2
K24	AGAAAAACUA	UAGCCAUC	6.0 ± 0.8
K5	AGCGAAGAUA	UAGCCAUC	1.7 ± 0.3
K20	AGGAGAUCUA	UAGCCAUC	3.2 ± 0.2
K27	AGUGAAAUUA	UAGCCAUC	9.9 ± 2.0
K19	AGA-CAACUA	UAUACAUC	2.0 ± 0.4
K2	AGAAAGACUA	UGACGAUC	7.9 ± 1.7
K23	AGUAAGGUUA	UGCCAAUC	5.9 ± 1.1
K9	AGAGCGAAUA	GACGCAUC	1.3 ± 0.3
K17	AGAGGCCAUA	GAGCCAUC	6.5 ± 1.8
K4	AGA-CGGGUA	GCACAAUC	1.1 ± 0.1
K28	AGGGCAAAUA	GCAGCAUC	2.9 ± 0.6
K21	AGUGAAAGUA	GCCACAUC	2.8 ± 0.5
K31	AGAGAGGGUA	GCCCAAUC	6.3 ± 0.8
K10	AGA-AGGAUA	GUACCAUC	5.9 ± 2.2
K14	AGGAGGGAUA	GGACCGUC	4.1 ± 0.5
K13	AGAGGAAAUA	GUACCGUC	1.3 ± 0.2
K16	AGGGGGAUUA	GUCAAGUC	4.8 ± 1.2
K32	AGUCGGUAUA	GUCGAGUC	38.5 ± 1.8

The sequences are deduced from the genes. Only the D and T-loops, where the sequences differ one from the other, are given. Positions of nucleotides 16-19 in the D-loop and 54-58 in the T-loop mark the beginning and the end of the each randomized region. Nucleotides forming the RH base pair in the T-loop are underlined. The activity of $tRNA^{Ala}_{su+}$ is taken for 100%.

Figure 1

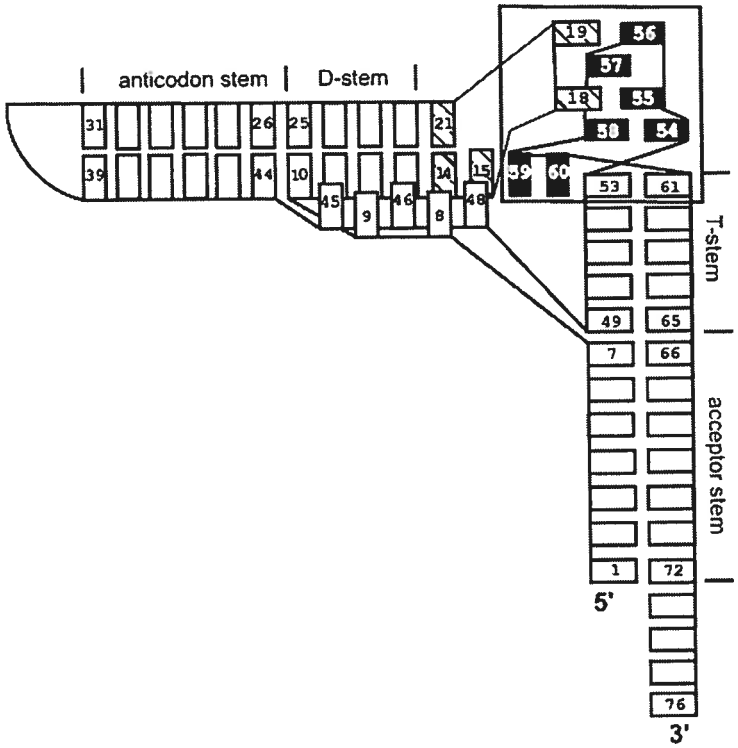


Figure 2

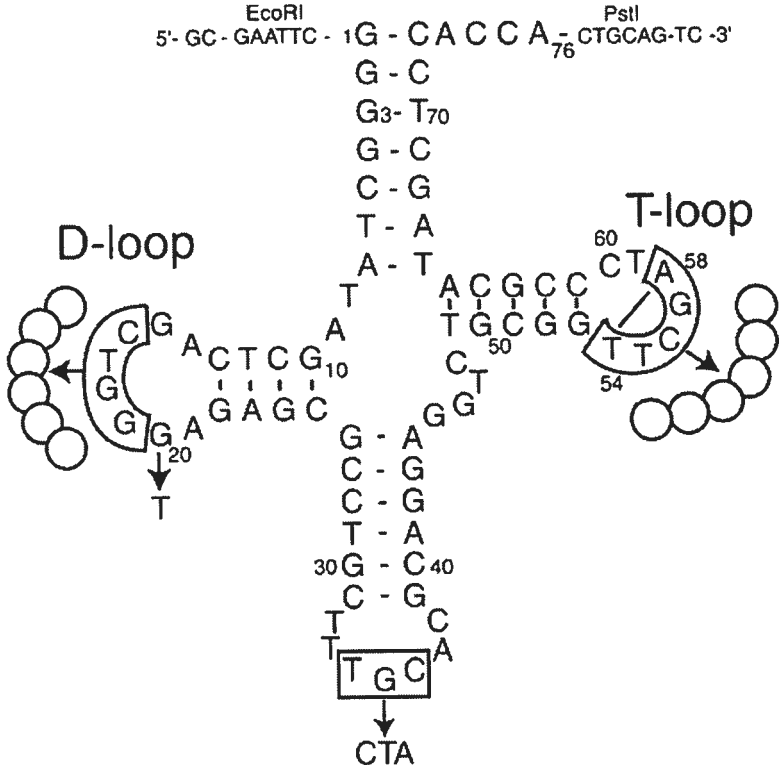


Figure 3

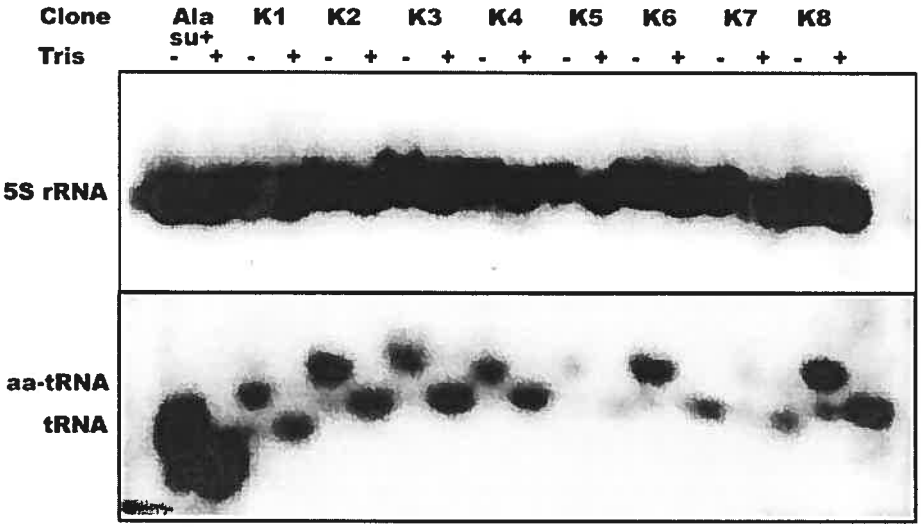


Figure 4

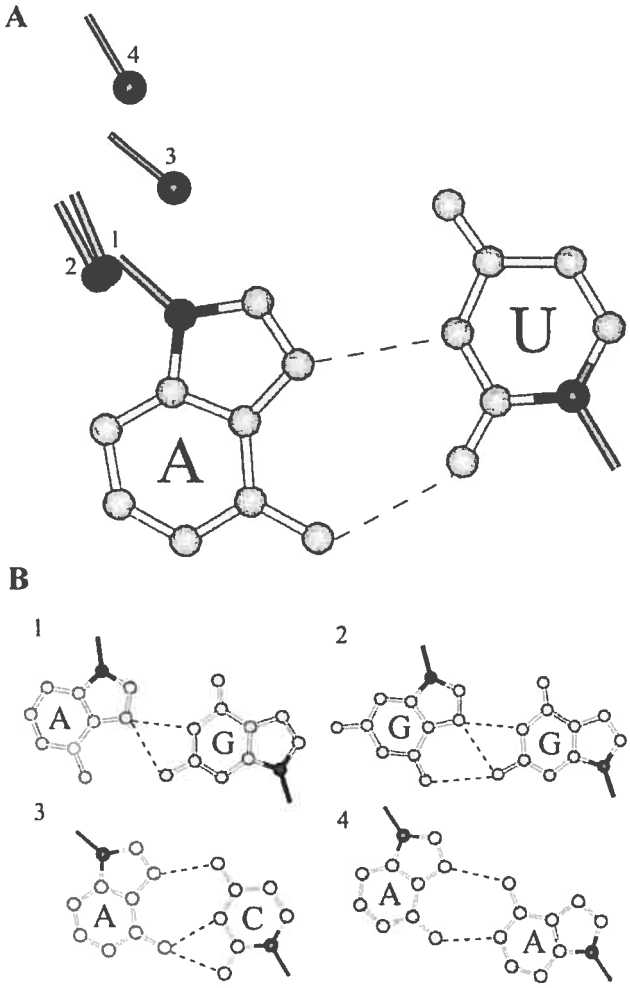
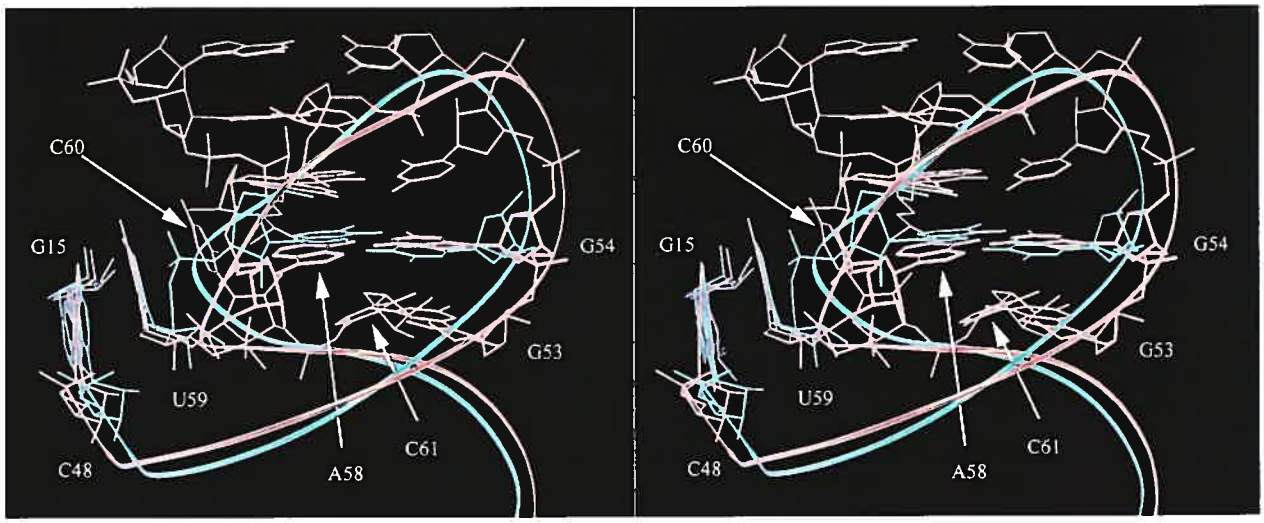


Figure 5



Chapter 2. Article:

Keeping the tRNA L-shape with the help of the reverse-Hoogsteen base pair 54-58

Ekaterina I. Zagryadskaya, Natalia Kotlova and Sergey V. Steinberg

Département de Biochimie, Université de Montréal, Montréal, Québec H3C 3J7, Canada

Submitted to *Proceedings of the National Academy of Sciences of USA*

Contribution of each author:

Ekaterina Zagriadskaia: developed detailed experimental design, did most of the experiments, participated in data analysis and preparation of manuscript and figures

Natalia Kotlova: did part of the experimental work

Sergey Steinberg: developed general experimental design, analyzed results, did structural modeling, participated in preparation of manuscript and figures

ABSTRACT

Based on *in vivo* selection of effective suppressor tRNAs from two different combinatorial gene libraries in which several nucleotides in the D- and T-loops were randomized, we show through analysis of selected tRNA sequences and computer modeling of their structures that the position of the reverse-Hoogsteen base pair in the T-loop co-varies with the length of the D-domain. When the D-domain has the normal length, the position of the reverse-Hoogsteen base pair in the T-loop is always such that it allocates two unpaired nucleotides for the bulge that fills the space between the D- and T-domains. However, when the D-domain becomes shorter, the position of the reverse-Hoogsteen base pair changes in such a way that more nucleotides are now allocated to the bulge in the T-loop, so that the total length of the D-domain and of the bulge remains the same. Such compensation guarantees that in all tRNAs, the D- and T-domains are always juxtaposed in the standard way. It also demonstrates the major role played by the reverse-Hoogsteen base pair in the formation of the canonical tRNA L-shape conformation.

INTRODUCTION

In spite of the recent progress in RNA crystallography, our understanding of how a particular RNA nucleotide sequence folds into a distinct tertiary structure and how the latter defines the function is rather limited. This is true not only for large ribonucleo-protein complexes like the ribosome, but even for the relatively small tRNA. The canonical tRNA conformation, the L-shape, which is characterized by a perpendicular arrangement of the two helical domains, has already been known for three decades. Still, despite the enormous amount of accumulated structural and functional data (1-2), one cannot specify precisely which elements of the tRNA structure determine the L-shape conformation.

It is generally accepted that the major interactions maintaining the L-shape occur at the corner of the molecule, where the D- and T-loops meet. This region, henceforth called the DT region, contains several unique elements, including the reverse-Hoogsteen base pair T54-A58 and the Ψ55-mediated U-turn in the T-loop, the inter-loop base pairs G18-Ψ55 and G19-C56, and the stack of four mutually intercalated purines A58-G18-R57-G19 (R stands for a purine) (3, Fig. 1). The bulge formed by two unpaired nucleotides 59-60 in the T-loop stacks on the last nucleotide layer of the D-domain formed by the tertiary base pair 15-48. The fact that all of these interactions show practically no variation in cytosolic tRNAs indicates their importance for the tRNA function.

The existing experimental data, though rather fragmentary, also support the importance of the DT region. Thus, among the mutants obtained by random mutagenesis of the *E.coli* $tRNA^{Ala}_{CUA}$ and having at least one of the above-mentioned interactions disrupted, none was functional (4). Also, an insertion of additional nucleotides into the T-loop or deletion of G19 from the D-loop affected the accuracy of the codon-anticodon recognition and resulted in a frameshift (5-6). There have also been some *in vitro* studies of the role of the DT region structure at different steps of the tRNA functional cycle. Depending on the particular step and on the enzyme involved, the role of the structure of this region can vary from important to not essential (7-12).

Both the extreme conservation of the DT region in the cytosolic tRNAs and the impossibility to obtain functional mutants containing mutations in this region by either directed or random mutagenesis hamper the analysis of the particular role played by each of its elements in the tRNA structure and function. To circumvent the problem, we use here an alternative approach of *in vivo* tRNA selection from combinatorial gene libraries, which has allowed us to select suppressor tRNAs with the structure of the DT region very different from the standard. Analysis of the nucleotide sequences of the selected clones and molecular modeling of their conformations has helped us to identify the most important elements in the DT region and to determine their role in the formation of the tRNA L-shape. In particular, our analysis revealed the key role played by the reverse-Hoogsteen base pair 54-58 in positioning the two unpaired nucleotides 59-60 in the space between the D/anticodon and acceptor/T helical domains. The proper arrangement of these nucleotides was found to be essential for the correct juxtaposition of the two helical domains.

MATERIALS AND METHODS

Strains

The *E.coli* strain XAC-1 (F' $lacI_{373}lacZ_{u118\ am}$ $proB^{+}$ /F^{-} $D(lac\text{-}proB)_{x111}$ *nal*A *rif* $arg E_{am}$ *ara*) containing amber mutation in gene *lac*Z (13) was used for cloning and selection of suppressor tRNAs.

Construction of the combinatorial library and selection of suppressor tRNAs

The template oligonucleotide coding for the combinatorial tRNA library (Fig. 2) was synthesized at BioCorp Inc (Montreal, Canada), amplified by PCR to produce the

double-stranded DNA using primers 5'GCGAATTCGGGGCTATA3' and 5'GACTGCAGTGGTGGAGT3', and cloned into plasmid pGFIB-1 using restriction sites *Eco*RI and *Pst*I, as described previously (14). This plasmid provides a constitutive high-level expression of a cloned tRNA gene. Vent DNA polymerase, restriction enzymes and T4 DNA ligase were from New England Biolabs. Ligation mixture was electroporated into electrocompetent XAC-1 cells. The transformation mixture was spread on twenty 150x15 mm LB-agar plates containing ampicillin (100 mg/ml) and 5-bromo-4-chloro-3-indolyl b-D-galactopyranoside (X-Gal), resulting in about 50 000 colonies. XAC-1 harbors an amber nonsense mutation in the β-galactosidase gene, which results in white colonies on the X-Gal plates. The effective suppression of this nonsense mutation makes the colonies blue, which allows a rapid detection of suppressor tRNAs. Positive clones containing suppressor tRNAs were selected as blue colonies after overnight growth at 37°C or after an additional incubation of 8 hours at 4°C. The plasmid DNA of these clones was extracted and retransformed into the XAC-1 cells to confirm the dependence of the phenotype on the presence of the plasmid encoding the tRNA gene.

Sequencing

Sequencing of both strands of the selected tRNA genes was performed on the LI-COR DNA sequencing system (Département de Biochimie, Université de Montréal). Primers 5'-GCTTCTTTGAGCGAACGATCAAAAATAAGT-3' and 5'-GGGTTTTCCCAGTCACGACGTTGTAAAACG-3' were used, respectively, for the forward and reverse reading of the tRNA gene. Both primers were labelled at the 5'-terminus with IRDye 800 (LI-COR Biosciences).

Measuremenst of the suppressor activity

Suppressor activity of selected tRNAs was evaluated by measuring the β-galactosidase activity in XAC-1 cells carrying suppressor tRNA genes, which results from suppression of the amber stop codon in the *lacZ* gene. β-galactosidase activity was determined as described by Miller (15), using overnight cultures grown in minimal A medium containing 0.4% glucose and 1 mM $MgSO_4$ to an OD_{600} of 0.8-0.9. The values were obtained by averaging the measurements from three independent cultures and calculated as a percentage of the activity of the control $tRNA^{Ala}_{CUA}$.

Detection of suppressor tRNAs by Northern blot and evaluation of their aminoacylation levels

To preserve the aminoacylated form of the tRNAs, the total cellular RNA was extracted under acidic conditions at pH 5.2, as described previously (14). Subsequently, part of each RNA sample was deacylated and run on acid PAGE side by side with untreated RNA sample (16). RNA from the gel was transferred by electroblotting to a Hybond-N nylon membrane (Amersham), and the membrane was hybridized with two radiolabeled DNA probes, one complementary to region 26-44 of the suppressor tRNAs consisting of the anticodon stem and loop, and the other to region 34-53 of the *E.coli* 5S rRNA (16). The 5S rRNA probe was used to monitor the amount of total RNA in each sample.

Computer modeling

Preliminary modeling was done interactively, using InsightII/Discover package (Version 2000, Accelrys Inc., San Diego, CA). The X-ray structure of the yeast $tRNA^{Phe}$ (3) was used as a starting conformation, to which the elements different from the standard tRNA structure were appended. Each model was submitted to energy minimization in vacuum using the AMBER forcefield (17) to avoid steric clashes and to guarantee the standard values of covalent bonds and angles. Visualizations were done on a Silicon Graphics O2 computer.

RESULTS

Background: the first K-library

To understand the structure-function relationships in the tRNA DT region, we recently undertook an *in vivo* selection of active suppressor tRNAs from a combinatorial gene library (16) in which six nucleotides in each loop of the *E. coli* suppressor $tRNA^{Ala}_{CUA}$ were randomized (K-library, Figure 2). $tRNA^{Ala}$ was chosen because its aminoacylation largely depends on the presence of base pair G3-U70 in the acceptor stem, while the DT region harbors only weak identity elements (18-21), which would make the library variants good substrates for the alanyl-tRNA synthetase. To provide the T-loop with more conformational freedom and to increase the probability of selection of alternative structural patterns, we extended the T-loop by one nucleotide compared to the standard seven. Henceforth, the six randomized nucleotides of such extended T-loop are numbered from 1* to 6*. In total, 28 positive clones were selected (16), of which some representatives are

shown in Table 1A. Analysis of the nucleotide sequences of the selected clones showed that in sixteen clones, positions 1* and 6* were occupied respectively by U and A (Fig. 2). In the other twelve, either G1*-G6* or G1*-A6* combination was found. The presence of the U1* and A6* allowed the formation of the reverse-Hoogsteen base pair between them analogous to base pair T54-A58 in the standard tRNAs. It was also shown that GG and GA were able to form arrangements almost isosteric to the reverse-Hoogsteen base pair U A (16). To stress the similarity between the three base pairs, we call all of them reverse-Hoogsteen pairs (RH). In some K-clones, a deletion of a nucleotide led to the normal length T-loop. Alignment of the sequences of the T-loop with and without this deletion showed that the additional eighth nucleotide fell into region 2*-5*, which corresponds to the top of the T-loop. This region not only had a different length in different clones, but also showed considerable sequence variability among the selected clones. Thus, the only common aspect that was preserved in the randomized regions of all clones with respect to the standard tRNA structure was the ability to form RH U54-A58 or its analogs GA or GG.

The revealed conservation of the RH suggests that it plays a crucial, although unknown role in the tRNA structure and function. We can argue, however, that the properly positioned RH is critical for the tRNA L-shape. Thus, we noted that the formation of base pair 1*-6* makes the two last nucleotides of the T-loop unpaired in the same way as the formation of base pair 54-58 does in the normal tRNAs (Fig. 1, 4A). In the standard tRNA structure, bulge 59-60 stacks between the T-stem and the tertiary base pair 15-48 constituting the last stacked layer of the D-domain. In this position, the bulge mediates the interaction between the D/anticodon and acceptor/T helical domains and thus plays the key role in their specific perpendicular juxtaposition known as the L-shape. According to our hypothesis, the major role of RH 54-58 (or 1*-6*) is to allocate a certain number of nucleotides for the bulge that will guarantee the maintenance of the standard tRNA L-shape. If this hypothesis is correct, all three elements, including RH, the bulge, and base pair 15-48, should be all important for the maintenance of the L-shape and for the tRNA function. Moreover, if the normal tRNA shape is to be preserved, a structural modification of one of these elements should be compensated for by the corresponding modifications in the others.

The second M-library

To test this hypothesis, we created another tRNA combinatorial library (M-library). Its design was based on the design of the first K-library, with the only changes consisted in deletion ΔC48 and replacement G15T, as indicated by arrows in Fig. 2. These changes were intended to eliminate the last stacking layer in the D/anticodon domain, 15-48, and thus to create problems for the normal interaction of this domain with bulge 59-60.

In total, 40 positive M-clones were selected, of which 21 showed the suppressor activity higher than 2% of the control suppressor $tRNA^{Ala}_{CUA}$. The nucleotide sequences of the randomized regions of these tRNAs are shown in Table 1B. For five clones arbitrarily chosen from Table 1B, the *in vivo* level of the suppressor tRNA and of its aminoacylated form was determined by the acid polyacrylamide gel electrophoresis followed by hybridization with a specific DNA probe complementary to the anticodon stem and loop. As one can see in Fig. 3, all tested suppressor tRNAs had a substantial level of presence in the cytosol, which however was somewhat lower than that of control $tRNA^{Ala}_{CUA}$. For each clone, most of the tRNA was found in the aminoacylated form.

Analysis of the nucleotide sequences of the selected M-clones

Based on common characteristics, all nucleotide sequences shown in Table 1B are divided into three major groups. The largest group (Group 1) consists of twelve sequences with an almost identical T-loop sequence. In these sequences, the first five nucleotides 1*-5* of the T-loop perfectly fit to the standard sequence pattern U54-U55-C56-R57-A58. In addition, in all twelve sequences, the randomized region of the D-loop contains dinucleotide GG, which corresponds to the conserved dinucleotide G18-G19. Based on this observation, we can suggest a similar pattern for all Group 1 tRNAs, in which nucleotides 1*-5* of the T-loop and the dinucleotide GG of the D-loop form the structure identical to that of the corresponding region 54-58 in the standard tRNAs. This includes the formation of RH U1*-A5* equivalent to U54-A58 and of the inter-loop base pairs equivalent to G18-U55 and G19-C56.

The formation of base pair U1*-A5* will assign the additional eighth nucleotide of the T-loop to the bulge between this base pair and the T-stem, which will now contain three nucleotides. This drastically contrasts to what was observed in the K-library, where RH consistently appeared between positions 1* and 6* of the T-loop, while the eighth nucleotide was located in the top part of the T-loop. Because the only difference between

the designs of the K- and M-libraries consisted in the absence of base pair 15-48 in the latter, we can argue that the relocation of the additional nucleotide of the T-loop from the top part to the bulge represents a structural rearrangement which compensates for the absence of this base pair. We can suggest that in the tertiary structure of the Group 1 tRNAs, the first five (1*-5*) and the last two (59-60) nucleotides of the T-loop stay at the positions equivalent to those of the seven nucleotides 54-60 in the standard tRNA structure, while nucleotide 6* takes the place normally occupied by nucleotide 48 (Fig. 4B). The location of nucleotide 6* at this place will compensate for the absence of base pair 15-48 and will guarantee that the number of stacked layers in the region between the two helical domains corresponds exactly to that needed for the standard domain juxtaposition. Although the distance between nucleotides 5* and 6* in this case is larger than that between nucleotides 58 and 59 in the standard tRNA, modeling experiments show that the conformational potential of the polynucleotide chain is sufficient to make the proper connection without additional movement of any nucleotide in the region. The resulting structure of the T-loop is stable and sterically reasonable (Fig. 4B). Additional stabilization of this arrangement can come from base pairing between nucleotides 15 and 6*, which are now located in the same stacked layer. In the most favorable case of a Watson-Crick combination 15-6*, the reverse-Watson-Crick base pair will be formed, although some other nucleotide combinations providing for a reasonable H-bonding, including UG, UU, UC or AA can also be accommodated at this place.

Though we propose that all tRNAs from Group 1 have similar structural pattern, we can see that their suppressor activities differ substantially (9-fold between the most and least active). We attribute this fact to the differences between their D-loop sequences, which can result in small structural changes between individual tRNAs and influence their function. Based on data in Table 1 and Fig. 3 we can suggest that the D-loop sequence can influence tRNA metabolism prior to aminoacylation (different expression levels), as well as post aminoacylation (higher activity of less abundant tRNA: compare clones M32 and M36).

In two clones M23 and M31 comprising Group 1a, one can also form RH between nucleotides 1* and 5*, respectively UA and GG. Again, such a base pair will allocate three nucleotides for the bulge, thus compensating for the absence of the last stacked layer in the D-domain. However, in these two clones, unlike in the Group 1 clones, the top of the T-loop between nucleotides 2* and 5* does not correspond to the standard sequence pattern.

This is not surprising because, for example, in the K-library, no clone had the standard sequence pattern and still, all selected clones showed a detectable level of activity (16). The variability of the top part of the T-loop once again indicates that the particular structure of this region is not critical for the RNA function. In clone M23, RH can be formed in two different ways, as U1*-A5* or U1*-A6*. Only the first type of base pairing provides for the conformation with the proper juxtaposition of the two helical domains. We think, however, that as long as at least one such conformation is possible, the tRNA species will be able to function.

In none of the seven sequences forming Groups 2 and 3, RH between positions 1* and 5* is possible. In the four sequences of Group 2, such a base pair can be formed between positions 1* and 6*. The 1*-6* base pair leaves only two nucleotides for the bulge region, as in the standard tRNA structure and in the K-library, which in the absence of base pair 15-48 would create problems for the correct interaction of the helical domains. Interestingly, all four sequences of Group 2, due to a spontaneous mutation in the D-loop, contain adenine in position 15. As a purine, this adenine can form a stacked layer on its own or make a reverse-Watson-Crick base pair with nucleotide U47, which in the absence of C48 would take its place. In either case, the stacked layer corresponding to base pair 15-48 in the normal tRNAs will be restored, which would require only two nucleotides for the bulge to make the proper juxtaposition of the two helical domains. Thus, in Group 2, the inability of the bulge to have more than two nucleotides is compensated by the restoration of the deleted stacked layer in the D-domain.

The three clones constituting Group 3 can form none of base pairs UA, GG or GA in the T-loop that would allocate the necessary number of nucleotides for the bulge region to guarantee the normal juxtaposition of the two helical domains. This aspect places the Group 3 clones apart from almost fifty K and M-clones. Below, we discuss alternative configurations by which the nucleotides in the DT region can be arranged that would conserve the L-shape of these tRNAs.

DISCUSSION

The results presented here demonstrate that a tRNA can be functional even if the structure of its DT region is substantially modified compared to the standard. Indeed, none of the clones selected either from K- or M-library had the standard structure of the DT

region. Although for all clones the efficiency of the nonsense codon suppression was lower than for the $tRNA^{Ala}_{CUA}$, it was strong enough to provide a detectable level of β-galactosidase activity. Also, given that the cellular levels of the selected clones were somewhat lower than that of the control $tRNA^{Ala}_{CUA}$, their activities normalized by the concentrations are in reality higher than those reported in Table 1. Decrease in cellular levels of selected tRNAs most probably derives from less efficient 5' and 3' processing, since it has been shown that mutations in the D- and T-loops decrease tRNA processing (10). The fact that these clones were selected *in vivo*, guaranteed that all of them were correctly transcribed, processed and folded, and that they were able to interact productively with all necessary factors of the protein biosynthesis machinery. In particular, all tested suppressor tRNAs existed in the cell mostly in aminoacylated form (Fig. 3), which we attribute to the fact that in $tRNA^{Ala}$, the DT region plays a relatively minor role in the aminoacylation (18-21).

Selection of so many different unusual clones can be attributed to the power of the approach of using combinatorial tRNA gene libraries, which has allowed us to explore the whole sequence space of the DT region, including areas distant from the standard pattern. This also shows that the sequence space available to functional tRNAs is not limited to that which can be deduced from the analysis of the naturally occurring cytosolic molecules. One can conclude, therefore, that the uniformity of the cytosolic tRNAs is not driven by the fact that other possibilities have no chance to be functional, but rather by a higher efficiency of the standard tRNAs and/or by the tRNA uniformity itself, because the standardization of the structure of all tRNA species would optimize their interaction with such common factors as RNase P, EF-Tu or the ribosome.

The comparison of the nucleotide sequences of the K- and M-clones makes the central point of this work. While in the K-clones, RH consistently occurred between positions 1* and 6* of the T loop (Fig. 4a), in most M-clones (Groups 1 and 1a), it was found between positions 1* and 5* (Fig. 4b). Since the only aspect by which the design of the M-library differs from that of the K-library is the absence of the tertiary interaction 15-48, we should consider the switch of the base pairing from 1*-6* to 1*-5* as a direct response to the loss of base pair 15-48. This switch releases nucleotide 6* from pairing with nucleotide 1* and enables it to join the bulge where it forms a stacked layer at the place normally occupied by base pair 15-48. An alternative way of coping with the absence of the 15-48 base pair would be the restoration of the 15-48 layer and the return to the

standard type of interaction between the D-domain and the two-nucleotide T-loop bulge, as exemplified by the Group 2 clones. In these clones, the inability to allocate more than two nucleotides for the bulge coincides with the spontaneous appearance of A in position 15 of the D-loop. This again leaves the total number of the stacked layers in the DT region unchanged and, therefore, does not affect the juxtaposition of the helical domains.

These results clearly demonstrate the major role of RH in the maintenance of the tRNA L-shape. Upon formation, this base pair allocates a particular number of nucleotides for the bulge region, which would guarantee the standard juxtaposition of the helical domains. In the normal tRNAs, neither the length of the bulge region nor of the D-domain varies, thus the real role of RH is hidden. To uncover this role, one has to use an experimental scheme in which alternative conformations of the T-loop would be possible. The necessary level of the flexibility of the T-loop conformation was achieved by introduction of an additional eighth nucleotide. Depending on the particular structural context, i.e. on the presence or absence of base pair 15-48, this extra nucleotide is involved either in the bulge or in the top region of the T-loop. The exact location of this extra nucleotide in each clone thus depends on the position of RH. Although this role of RH could only be revealed when both the position of RH and the length of the bulge became variable, it is the same in the normal tRNAs as well, even though neither of the two aspects varies there.

Previously, based on the analysis of unusual mitochondrial tRNAs in which the anticodon stem, instead of containing the standard six base pairs, varies in length between five and ten base pairs, we suggested a rule for the normal tRNA shape (22). According to this rule, the total number of the stacked layers in the anticodon/D helical domain between the anticodon loop and the T-loop should be twelve. Taking into account the data presented here that the number of nucleotides in the T-loop bulge can also vary, we can generalize the rule, stating that the total number of stacked layers between the anticodon loop and the T-stem, including the anticodon stem, the D-stem, the stacked nucleotides of the D-loop as well as the nucleotides of the T-loop bulge, should be fourteen.

Our interpretation of the experimental results is based on the suggestion that the bulge region in the T-loop can incorporate an additional third nucleotide, which would stack to nucleotide 59 and take the place normally occupied by the tertiary base pair 15-48. For this to occur, the region between nucleotides 5* and 6* should accept an extended conformation comparing to that between nucleotides 58 and 59 in the standard tRNA

structure. Whether this is possible, was not evident at the outset. According to our molecular modeling experiments, this extension does not create steric problems. However, making one more step in the same direction, i.e. adding another nucleotide to the bulge region and positioning it at the next layer, normally occupied by triple 8-14-21, would require a stretch of the dinucleotide conformation beyond its natural capacity. To make the situation sterically reasonable, the introduction of a connector nucleotide may be necessary.

Our results reveal a strong tendency to have the number of the stacked layers in the DT region of functional tRNAs invariable with respect to that in the standard tRNA structure. This is manifested by allocation of an additional nucleotide for the T-loop bulge when base pair 15-48 is eliminated. If this additional nucleotide in the bulge did not appear, nucleotide 59, most probably, would have stacked on the last layer of the D-domain formed by triple 8-14-21. This would displace the acceptor/T helical domain as a whole with respect to the D/anticodon domain, potentially damaging the tRNA function. Although it is generally accepted that the L-shape is absolutely required for tRNA function, the precision with which it should be maintained was unknown. Our results show that the deficiency of even one stacked layer, that is only about 3 Å in the length of the D-domain, can make a difference between functional and non-functional tRNA.

The three clones from the M-library that constitute Group 3 do not follow the same rule as all other clones. In these clones it is not possible to form RH that would allocate the required number of nucleotides for the bulge region. Because it happens in only three out of 51 clones selected from both libraries, while all other clones seem to have the standard juxtaposition of the two helical domains, we can suggest that in the Group 3 clones, the helical domains are also properly arranged, although the mechanism of the domain fixation may be different. Several alternatives for the domain fixation could be proposed. For example, we noticed that in all Group 3 clones, a double helix consisting of at least four base pairs can be formed between the D and T-loops. Shared by all three Group 3 clones, this feature is not at all common for the other selected clones. If the formation of such a double helix is able to fix the positions of the two helical domains, these clones could function even without the properly positioned RH and the bulge of the proper length. Due to insufficient statistics, it is premature to deduce precisely how these clones can form the correct L-shape. The analysis of sequences from a larger set of clones will allow for a rigorous test of the alternative hypotheses.

In the standard tRNA structure, the interactions of the D- and T-domains are not

limited to the contact between base pair 15-48 and nucleotide 59. In addition to this contact, two base pairs 18-55 and 19-56 are formed between the D-loop and the top part of the T-loop. Because in the selected clones, the nucleotide sequence of the top part of the T-loop is variable, the presence of these two base pairs does not seem to be crucial for tRNA function. The probable role of these base pairs is to stabilize additionally RH T54-A58 and thus to contribute to the stability of the standard conformation. One can suggest that in those clones where the top of the T-loop is modified compared to the standard, these base pairs are replaced by other interactions playing the same stabilizing role. The nature of these interactions in different selected clones is a matter of further analysis. It would be of particular interest to investigate the exact nature of these interactions by using direct experimental methods, such as structural probing in solution, NMR or X-ray crystallography. However, such studies may not be possible if selected tRNA variants have reduced stability or highly dynamic structures.

Although the necessity for the tRNA to have the normal shape has been demonstrated, it is not yet clear which steps of its functional cycle will be damaged the most if the proper shape is not maintained. It is known that the elongation factor Tu does not interact with the D/anticodon domain or the T-loop (23). As for the different aminoacyl-tRNA synthetases, the role of the DT region varies from important to non-essential. In the first case, this region contains major identity elements, like A20 and A59 in *E. coli* $tRNA^{Arg}$ (24), while in the second, almost any nucleotide sequence for either the D- or the T-loop is acceptable (9). Between these two extremities there are also intermediate cases, when nucleotide replacements in this region do not abolish the aminoacylation, but only reduce its efficiency (7-8). Despite these differences, all cytosolic tRNAs have the same universal structure of the DT region. It is also known, that modifications of the DT region can influence the processing of the 3' and 5' termini, as well as the CCA-addition. For example, it was shown, that disruption of base pairs between the D- and T-loops had a negative effect on the 3' and 5' processing of $tRNA^{His}$ from *Drosophila* (10), while positions 57 and 58 of the T-loop were involved in the recognition by the CCA-adding enzyme (12). However, there has been no evidence that these enzymes are sensitive to the juxtaposition of the two helical domains in the tRNA rather than to the identity of particular nucleotides or the integrity of particular base pairs.

Although any tRNA-related process relies on the presence of some elements in the tRNA structure, we think that the part of the tRNA functional cycle, which is shared by all

tRNAs and is the most demanding on the tRNA shape, is the interaction with the ribosome. During the selection of tRNA based on the correct codon-anticodon interaction, and later when tRNA is moving from one ribosomal site to another via several distinct intermediate states, it performs a complex, thoroughly choreographed set of movements, where different elements of its structure, the anticodon, the acceptor terminus, the D and acceptor stems, at each particular moment are bound to particular places on the ribosome surface and must exchange these interactions in a very distinct way. It seems plausible that for these movements, even a deficiency of one stacked layer in the D/anticodon helical domain could be critical.

tRNA is a small RNA molecule, and the way it keeps its shape is relatively simple. Other, larger RNAs are expected to use their own, more elaborate strategies to maintain their active conformations. On the other hand, the fact that the elements resembling the DT region have recently been identified in RNase P and in the ribosome (25-26), suggests that these molecules may use, among others, the same conformation-fixing strategies as we have seen in tRNA. The systematic elucidation of the conformation-fixing strategies used by different RNAs will constitute a very important step toward understanding of how RNA tertiary structure forms and how it defines the function of the molecule.

ACKNOWLEDGMENTS

We thank Drs. S.W. Michnick and W.H. McClain for critical reading of the manuscript and F.R. Doyon for assistance in preparation of the figures. S.V.S. acknowledges a grant from National Science and Engineering Research Council of Canada and fellowships from Canadian Institutes of Health Research and from Fond de la Recherche en Santé du Québec.

REFERENCES

1. *tRNA: Structure, Biosynthesis, and Function* (1995), eds. Söll, D. & Rajbhandary, U.L. (ASM Press, Washington, DC).

2. Rich, A. & RajBhandary, U.L. (1976) *Annu. Rev. Biochem.*, **45**, 805-860.

3. Shi, H. & Moore, P.B. (2000) *RNA* **6**, 1091-1105.

4. Hou, Y.M. & Schimmel, P. (1992) *Biochemistry* **31**, 4157-4160.

5. O'Mahony, D.J., Mims, B.H., Thompson, S., Murgola, E.J. & Atkins, J.F. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7979-7983.

6. Tuohy, T.M.F., Thompson, S., Gesteland, R.F. & Atkins, J.F. (1992) *J. Mol. Biol.* **228**, 1042-1054.

7. Puglisi, J.D., Putz, J., Florentz, C. & Giege, R. (1993) *Nucleic Acids Res.* **21**, 41-49.

8. Du, X. & Wang, E.D. (2003) *Nucleic Acids Res.* **31**, 2865-2872.

9. Bullock, T.L., Sherlin, L.D. & Perona, J.J. (2000) *Nature Struct. Biol.* **7**, 497-504.

10. Levinger, L., Bourne, R., Kolla, S., Cylin, E., Russel, K., Wang, X. & Mohan, A. (1998) *J. Biol. Chem.* **273**, 1015-1025.

11. Nazarenko, I.A., Harrington, K.M. & Uhlenbeck, O.C. (1994) *EMBO J.* **13**, 2464-2471.

12. Li, Z., Gillis, K.A., Hegg, L.A., Zhang, J. & Thurlow, D.L. (1996) *Biochem. J.* **314**, 49-53.

13. Normanly, J., Masson, J.M., Kleina, L.G., Abelson, J. & Miller, J.H. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 6548-6552.

14. Bourdeau, V., Steinberg, S.V., Ferbeyre, G., Emond, R., Cermakian, N. & Cedergren, R. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 1375-1380.

15. Miller, J.H. (1972) *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory Press, Plainview, NY), pp. 352-355.

16. Zagryadskaya, E.I., Doyon, F.R., & Steinberg, S.V. (2003) *Nucleic Acids Res.* **31**, 3946-3953.

17. Pearlman, D.I., Case, D.A., Caldwell, J.W., Ross, W.S., Cheatham, T.E.,III, Ferguson, D.M., Seibel, G.L., Singh, U.C., Weiner, P.K. & Kollman, P.A. (1995) *AMBER 4.1* (University of California, San Fransisco, CA).

18. McClain, W.H. & Foss, K. (1988) *Science*, **240**, 793-796.

19. Hou, Y.M. & Schimmel, P. (1988) *Nature*, **333**, 140-145.

20. McClain, W.H., Foss, K., Jenkins, R.A. & Schneider, J. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 9272-9276.

21. Tamura, K., Asahara, H., Himeno, H., Hasegawa, T. & Shimizu, M. (1991) *J. Mol. Recognit.*, **4**, 129-132.

22. Steinberg, S., Leclerc, F. & Cedergren, R. (1997). *J. Mol. Biol.* **266**, 269-282.

23. Nissen, P., Kjeldgaard, M., Thirup, S., Polekhina, G., Reshetnikova, L., Clark, B.F. & Nyborg, J. (1995) *Science* **270**, 1464-1472.

24. McClain, W.H. & Foss, K. (1988) *Science* **241**, 1804-1807.

25. Krasilnikov, A.S. & Mondragon, A. (2003) *RNA* **6**, 640-643.

26. Lee, J.C., Cannone, J.J. & Gutell, R.R. (2003) *J. Mol. Biol.* **325**, 65-83.

LEGENDS TO THE FIGURES

Figure 1. A conventional representation of the DT region in the standard tRNA structure (3). Each nucleotide of the DT region involved in stacking or base pairing with another nucleotide is represented by a rectangle. The rectangles representing nucleotides of the D-loop are black, while those representing nucleotides of the T stem and loop as well as nucleotides 8 and 15 are white. All other nucleotides are shown as black dots. The identities of the conservative and semi-conservative nucleotides are indicated. R and Y stand for purine and pyrimidine, respectively.

Figure 2. Design of the two combinatorial tRNA gene libraries.

The design of the K-library was based on the sequence of $tRNA^{Ala}_{CUA}$. Six randomized nucleotide positions (indicated by circles) were introduced in each of the D and T-loops. In the T-loop, these positions are numbered from 1* to 6*. The two additional mutations G15T and ΔC48 introduced in the M-library are indicated by arrows.

Figure 3. The aminoacylation levels of selected tRNA clones.

Total cellular RNA was fractionated in acid polyacrylamide gel, transferred to a membrane, and Northern blot hybridization was performed using a specific DNA probe complementary to the anticodon stem-loop of the cloned tRNAs (16). For each clone, the sample in the + lane was deacylated by incubation with Tris, while the sample in the – lane was not. The aminoacylated form of each tRNA runs in the gel slower (higher band) than the deacylated form. wt stands for the amber suppressor $tRNA^{Ala}_{CUA}$, in which the DT region structure is the same as in the natural $tRNA^{Ala}$. N represents a negative control, standing for the pGFIB plasmid lacking a tRNA gene. Additional probe specific to 5S rRNA was used to monitor the amount of total RNA in the samples.

Figure 4. Comparison of the DT region structure in the tRNAs selected from the K- and M-libraries.

On the left: the schematic representation of the DT region in the context of the whole tRNA L-shape. On the right: a stereoview of the DT region in the corresponding three-dimensional model. Those nucleotides whose role in the juxtaposition of the two helical domains is discussed in the paper are represented by rectangles on the left and shown explicitly on the right. For the same nucleotide, the same color is used in both the right and left figures.

(*A*) The model of tRNA K2. In this tRNA, RH is formed between 1* and 6*, while the positions of base pair 15-48 and of the two nucleotides in the T-loop bulge are the same as in the standard tRNA structure (3).

(*B*) The model of tRNA M40. In this tRNA, tertiary base pair 15-48 is deleted, RH is formed between 1* and 5*, while nucleotide 6* joins the T-loop bulge and fits to the place normally occupied by base pair 15-48. One can see that in M40, the top region of the T-loop is shorter due to the fact that the additional eighth nucleotide switches from the top region to the bulge.

LEGEND TO TABLE 1.

The tRNA sequences outside the D- and T-loops were identical in all clones and are not shown. Randomized regions are enclosed in solid boxes. In the D-loop and T-loops, they correspond, respectively, to regions 16-19 and 54-58 of the standard tRNA structure. Nucleotides of the T-loop allocated for the bulge are enclosed in dashed boxes. The following nucleotides are underlined: (1) the nucleotides of the T-loop presumed to form the RH; (2) dinucleotide GG in the T-loop of the Group 1 clones; (3) A15 in the D-loop of the Group 2 clones; (4) the complementary regions in the two loops of the Group 3 clones. Suppressor activity of tRNAs was evaluated by measuring β-galactosidase activity in XAC-1 cells, which results from the suppression of the stop codon in the β-galactosidase gene. The suppressor activity of the amber suppressor $tRNA^{Ala}_{CUA}$ was taken for 100%. Background β-galactosidase activity in the XAC-1 cells transformed with the pGFIB plasmid without suppressor tRNA gene was below 0.01%.

Table 1. Sequences of the D-and T-loops and suppressor activities of tRNAs selected from the K- and M-libraries.

Clone		D-loop		T-loop		Activity
$tRNA^{Ala}_{CUA}$	AG	CUGG	GA	UU CGA	UC	100%
A . Examples of clones selected from the K-library						
K25	AG	GAACGC	UA	UG AAA	AC	17.7±6.2%
K29	AG	GAAAAA	UA	UG GGA	UC	5.1±1.0%
K30	AG	UGAGGA	UA	UC CAA	AU	10.8±1.1%
K2	AG	AAAGAC	UA	UGACGA	UC	7.9±1.7%
K23	AG	UAAGGU	UA	UGCCAA	UC	5.9±1.1%
K27	AG	UGAAAU	UA	UAGCCA	UC	9.9±2.0%
K26	AG	AACGAC	UA	UAAACA	UC	3.9±0.8%
K6	AG	AGGGAG	UA	GC ACA	UC	25.0±5.0%
K10	AG	A AGGA	UA	GUACCA	UC	5.9±2.2%
K17	AG	AGGCCA	UA	GAGCCA	UC	6.5±1.8%
K14	AG	GAGGGA	UA	GGACCG	UC	4.1±0.5%
K31	AG	AGAGGG	UA	GCCCAA	UC	6.3±0.8%
K32	AG	UCGGUA	UA	GUCGAG	UC	38.5±1.8%
B. Clones selected from the M-library						
Group 1						
M2	AU	GUGGUU	UA	UUCGA U	UC	18.0±4.0%
M21	AU	GUUGGU	UA	UUCGA A	UC	12.0±2.0%
M8	AU	AUGGCA	UA	UUCGA G	UC	10.7±0.4%
M11	AU	AUUGG	UA	UUCGA C	UC	6.1±0.7%
M3	AU	UAUGGA	UA	UUCGA A	UC	5.8±1.2%
M1	AU	AGAGGA	UA	UUCGA U	UC	4.8±0.5%
M22	AU	CAUUGG	UA	UUCGA U	UC	4.2±0.8%
M40	AA	UAGGUCC	UA	UUCAA A	UC	4.0±0.5%
M6	AU	CAGGAU	UA	UUCGA G	UC	3.2±0.1%
M33	AU	UGCUGG	UA	UUCAA G	UC	3.1±0.5%
M10	AA	UACGGCA	UA	UUCAA U	UC	2.5±0.2%
M39	AU	UCGGUC	UA	UUCGA A	UC	2.1±0.1%
Group 1a						
M23	AU	UUUCCU	UA	UUGGA A	UC	5.2±0.4%
M31	AU	UCGGGA	UA	GUCCG U	UC	2.2±0.1%
Group 2						
M36	AA	UUUCACC	UA	UAGACA	UC	5.9±1.0%
M9	AA	UUUCAAC	UA	UAUACA	UC	2.7±0.4%
M28	AA	ACGGU	UA	UUCCCA	UC	5.6±0.1%
M19	AA	UGUGUUA	UA	UUAACA	UC	4.7±2.0%
Group 3						
M32	AU	CUGGCU	UA	UU AGA	CU	7.9±0.4%
M29	AU	UUGUAA	UA	UU CGA	UU	6.4±0.3%
M27	AU	GGGCGA	UA	CUGCCA	UC	5.4±0.4%

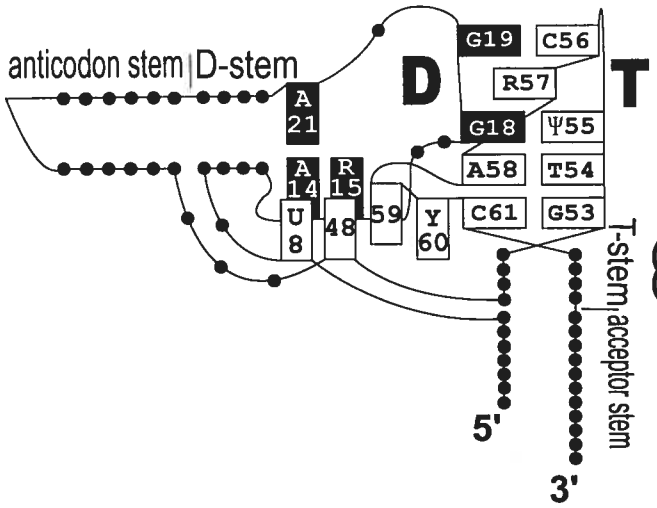


Figure 1

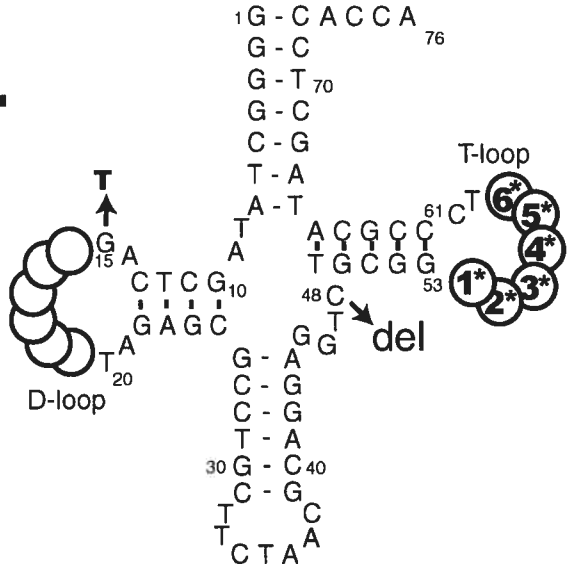


Figure 2

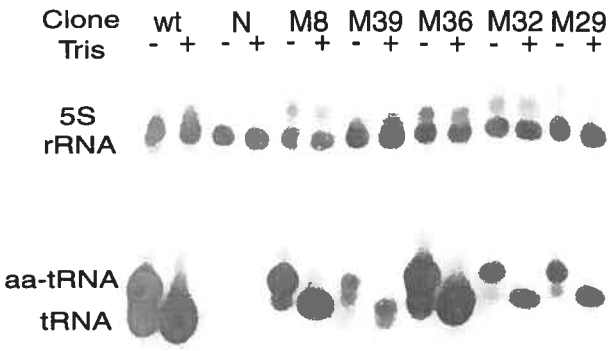


Figure 3

61

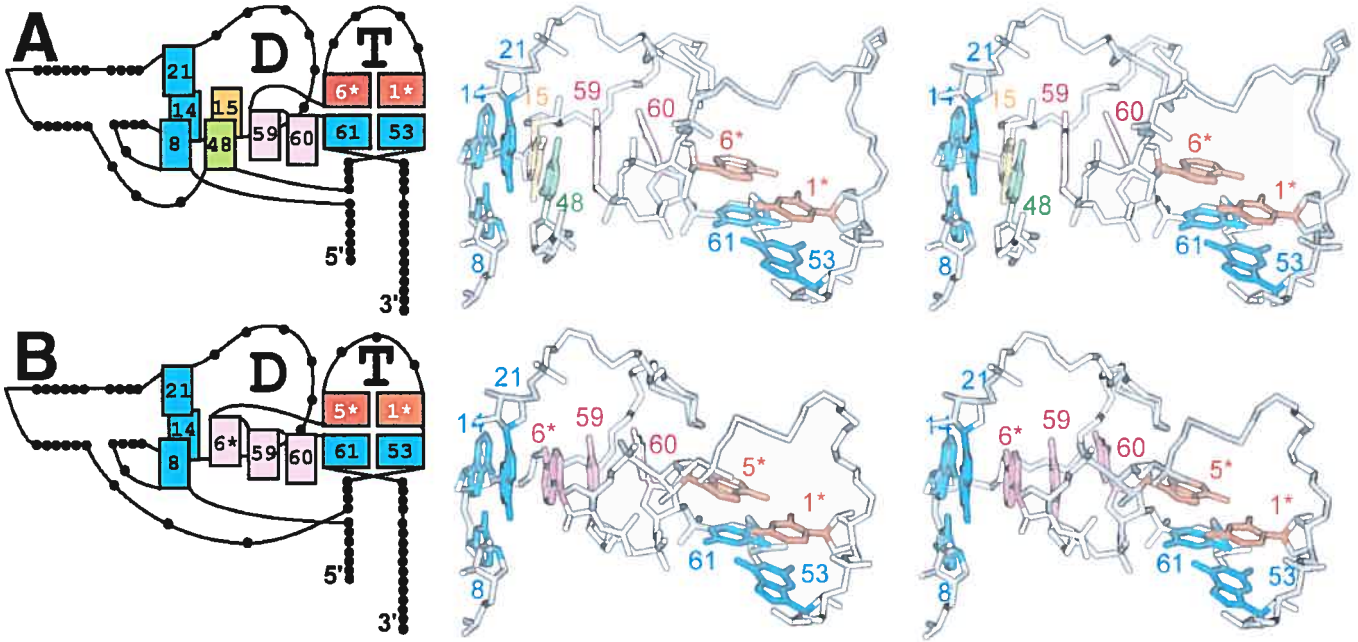

A
D
T
21
14
15
8
48
59
60
6*
1*
61
53
5'
3'
B
D
T
21
14
8
6*
59
60
5*
1*
61
53
5'
3'
21
1
15
59
60
6*
1*
48
8
61
53
21
1
6*
59
60
5*
1*
8
61
53

Figure 4

62

Chapter 3. Article:

Specific and non-specific purine trap in the T-loop of normal and suppressor tRNAs

Felix R. Doyon, Ekaterina I. Zagryadskaya and Sergey V. Steinberg

Département de Biochimie, Université de Montréal, Montréal, Québec H3C 3J7, Canada

Manuscript in preparation

Contribution of each author:

Felix Doyon: did structural modeling, participated in data analysis and preparation of manuscript and figures

Ekaterina Zagriadskaia: did the experimental part of work, participated in preparation of the manuscript

Sergey Steinberg: developed general experimental design, participated in structural modeling, data analysis and preparation of the manuscript

SUMMARY

To elucidate the general constraints imposed on the structure of the D and T-loops in functional tRNAs, active suppressor tRNAs were selected *in vivo* from a combinatorial tRNA gene library in which several nucleotide positions in these loops were randomized. Analysis of the nucleotide sequences of fifty selected clones demonstrates that 29 among them contain combination U54-A58 allowing the formation of the standard reverse-Hoogsteen base pair 54-58 in the T-loop. Among these 29 clones, two different types of the T-loop structure have been identified by sequence comparison and molecular modeling. The first type, so-called specific purine trap, is found in 12 individual tRNA clones and represents a generalized version of the standard D-T loop interaction. In this type of structure, the reverse-Hoogsteen base pair U54-A58 and the purine 57 are separated by the distance, which is twice as big as the usual distance between the two stacked bases in RNA, and the whole arrangment is stabilized by intercalation of purine 18 from the D-loop. The identity of purine 18 is restricted by specific base pairing with nucleotide 55. Depending on whether nucleotide 55 is U or G, purine 18 should be, respectively, G or A. The second structural type, so-called non-specific purine trap, is found in 16 individual tRNA clones and is described here for the first time. It consists of the two reverse-Hoogsteen base pairs UA and AC positioned at a similar double distance from each other. The intercalation of purine 18 between the two base pairs stabilizes the whole arrangement. The non-specific purine trap does not contain a nucleotide equivalent to nucleotide 55, so that both purines, G and A, at position 18 fit to the structure equally well. The role of both the specific and non-specific purine traps in the formation of the tRNA L-shape is discussed.

INTRODUCTION

Analysis of the available crystal and solution structures of different RNA molecules and RNA-protein complexes clearly shows that although unpaired non-helical regions usually account for less than a half of the nucleotide sequences, their role in the formation of RNA structure and in its function is crucially important. Often, specific interactions of such regions with other parts of the molecule ensure the particular juxtaposition of double helices within the unique functionally active tertiary structure (1). On other occasions, irregular n on-helical regions can recognize particular elements in another molecule, thus

mediating intermolecular contacts (2-3). Therefore, the understanding of how nucleotide sequence forms RNA tertiary structure and how the latter defines the function is impossible without clarification of the particular role played by the non-helical irregular elements and of the rules that govern their formation. Our knowledge on the structure of such elements is quite fragmentary. Several structural motifs have been identified in the tertiary structures of different RNAs (4-13). Also, a few irregular motifs, mostly representing distorted helices, have been studied by X-ray crystallography and NMR-spectrometry (14-16). Much less is known about the sequence and context requirements for the formation of these motifs and the particular role they play in the tertiary structure of functionally active RNAs.

The importance of irregular elements in RNA was first recognized almost thirty years ago when the tertiary structure of the yeast $tRNA^{Phe}$ was elucidated (17-18). In that structure, the two loops, D and T, were shown to interact specifically at the outer corner of the molecule in the so-called DT region (Fig. 1). This interaction brings together two parts of the nucleotide sequence separated by more than thirty nucleotides and positioned in different domains at the opposite ends of the secondary cloverleaf structure. The resulting conformation is characterized by the perpendicular juxtaposition of the two helical domains, D/anticodon and acceptor/T, known as the tRNA L-shape. The interaction between the two loops is very conservative and has been the subject of a number of studies (19-23). The presence of such elements as the U-turn between Ψ55 (Ψ stands for pseudouridine, a post-transcriptional modification of U) and C56, the unusual non-Watson-Crick base pairs T54-A58 and G18-Ψ55, the mutual intercalation of fragments 57-58 and 18-19, the two-nucleotide bulge 59-60 sandwiched between the T-stem and the D-domain (Fig. 1) makes the DT region one of the most structurally diverse in the whole tRNA and raises questions as to the particular role played by each element in the structure of this region and of the whole tRNA. To address these questions, we recently undertook an *in vivo* selection of active suppressor tRNAs from a combinatorial gene library (24, K-library) in which six nucleotides in each of the two loops were randomized, while the T-loop contained an additional eighth nucleotide compared to the standard seven (Fig. 2). Henceforth, the six randomized nucleotides of such an extended T-loop are numbered from 1* to 6*. Analysis of the selected tRNA clones showed that the only element that appeared in the randomized regions of most of the sequences was base pair U1*-A6* equivalent to the reverse-Hoogsteen base pair T54-A58 in the canonical tRNA structure. The formation of this base pair allocates two unpaired nucleotides for the bulge region, exactly as in the

normal tRNAs. Additional study, which involved selection of suppressor tRNAs from another combinatorial library, showed that the bulge could be extended up to three nucleotides when the last stacked layer of the D-domain, represented by tertiary base pair 15-48, was eliminated (25). In all cases, the total number of stacked layers in the DT region was always the same, despite the changes in both the D-domain and the T-loop bulge. This would guarantee the standard juxtaposition of the two helical domains of the tRNA, which is believed to be a prerequisite of the tRNA functionality. The fact that the number of nucleotides in the T-loop bulge is determined by the position of the reverse-Hoogsteen base pair 54-58 demonstrates the key role played by this base pair in keeping the tRNA L-shape.

As to the top region of the T-loop, enclosed by the reverse-Hoogsteen base pair, it did not have the canonical sequence in any of the K-clones. Moreover, the top region demonstrated considerable sequence variability in the selected clones and did not reveal any common pattern. This allowed us to conclude that the particular conformation of this region was not crucially important for the function of suppressor tRNAs. It was not possible at the time to say whether this region contained any important elements, mainly because of the insufficient number of sequences. Since then, we have performed an additional selection of suppressor tRNAs from the same K-library, whose total number has now reached 51. Of these clones, 29 contained nucleotides U1* and A6* in the T-loop, which suggested the formation of the reverse-Hoogsteen base pair between them, while most of the other clones contained G1* and either A6* or G6*. Because the presence of the reverse-Hoogsteen base pair UA provides an important structural constraint that simplifies the problem of elucidating the structure of the T-loop and of the whole DT region, we decided, at the first step, to limit our analysis to the sequences containing the U1*-A6* combination. The presented here theoretical analysis of the sequences of these tRNAs and molecular modeling of their tertiary structures has allowed us to identify two major structural types of the conformation of the top part of the T-loop. We call these types the specific and non-specific purine traps. The two types have very different sequence requirements and are easily recognizable in the nucleotide sequences. Both types allow the T-loop to interact with a part of the D-loop, although each type does it in a different way. The canonical structure of the T-loop constitutes a particular case of the specific purine trap, while the non-specific purine trap is described here for the first time. Comparison of the two identified types has helped us to formulate the most general requirements imposed on the structure of the DT region in functional tRNAs.

METHODS

Cloning, selection and measuring the suppressor tRNA activity

The details of the cloning and screening of the tRNA library as well as of the measurement of the suppressor activity of the selected tRNAs were described earlier (24).

Computer modeling

Preliminary modeling was done interactively, using InsightII/Discover package (Version 2000, Accelrys Inc., San Diego, CA). The X-ray structure of the yeast $tRNA^{Phe}$ (26) was used as a starting conformation, to which the elements different from the standard tRNA structure were appended. Each model was submitted to energy minimization in vacuum using the AMBER forcefield (27). The elements of the modeled structures identical to the corresponding elements of the yeast $tRNA^{Phe}$ were fixed during the minimization. Only those structures for which the energy minimum was reached, were taken for further consideration. Visualizations were done on a Silicon Graphics O2 computer.

RESULTS

The selected clones

The design of the suppressor tRNA combinatorial library was based on the sequence of the *E. coli* $tRNA^{Ala}_{UGC}$ (Fig. 2). Compared to the standard tRNA structure, the T-loop contained an additional eighth nucleotide. Six positions in the D-loop and six positions in the T-loop were randomized. The randomized positions in the T-loop are numbered from 1* to 6*. In total, 51 clones of active suppressor tRNAs have been selected. 29 of these clones shown in Table 1 contained combination U1*-A6*. This combination allows the formation of the reverse-Hoogsteen base pair analogous to T54-A58 in the normal tRNAs.

Type I: a quasi normal pattern

For the reason discussed below, none of the selected clones had the canonical tRNA sequence pattern. On the other hand, the initial inspection of the sequences revealed that at least some clones contained elements identical or similar to those existing in the normal tRNAs. We started our analysis with the five "quasi-normal" clones that had the standard

seven nucleotides in the T-loop due to a spontaneous deletion. Henceforth, we will call such molecules T7-tRNAs, in contrast to T8-tRNAs, which have eight nucleotides in the T-loop, as presumed by the library design (Fig. 2). For the T7 clones we will use the standard nucleotide numbering, while in the T8 clones, the randomized nucleotides of the T-loop are numbered from 1* to 6*. In all T7 clones, nucleotide 57 is a purine, as in the normal tRNAs, while the identities of nucleotides 55 and 56 always differ from the standard U(Ψ)55 and C56 known to be involved in specific interactions respectively with nucleotides G18 and G19 of the D-loop. Interestingly, in all five sequences, position 55 is occupied by G. Although none of the cytosolic tRNA has G55, the existence of this nucleotide in the T-loop of some mitochondrial tRNAs (28), viral tRNA-like structures (29) and in the T-loop-like structures of the RNase P (30) and of the ribosomal RNA (31-32) is well-documented. In almost all these documented cases, replacement U55G coexists with replacement G18A, which infers the formation of base pair A18-G55 instead of the standard G18-U(Ψ)55. Analysis shows that A and G can form a two-hydrogen-bond base pair similar to the standard G18-U(Ψ)55 and replace the latter in the tRNA structure with only minimal distortions of the backbone conformation (Fig. 4). The G18-U55 ⇔ A18-G55 co-variation was also observed by Peterson *et al.* in the *in vitro* selection of $tRNA^{Phe}$-derived variants able to be aminoacylated by phenylalanine-tRNA synthetase and to bind to the elongation factor Tu (33). As one can see in Table 1, each of the five T7 clones has several adenines in the D-loop able to form the base pair with G55 equivalent to G18-U(Ψ)55. Henceforth, the particular adenine forming this base pair is referred to as A18. The purine present in position 57 of all T7 clones can stack on top of this base pair exactly as in the standard tRNA structure.

The top stacked layer of the DT region in the canonical tRNA structure is occupied by the Watson-Crick base pair G19-C56 (Fig. 1). In the T7 clones, the D-loop partner of nucleotide 56 in this base pair could be any nucleotide following an adenine. Only in two sequences, K15 and K48, there is a possibility to make a Watson-Crick base pair 19-56, respectively UA and CG, while in the three other sequences this base pair could be AA or CA (in clones K25 and K3) or AG (K29). There are, however, doubts whether the Watson-Crick base pair 19-56 exists even in clones K15 and K48, where, in principle, it can be formed. First, as one can judge from Table 1, the possibility to form this base pair in clones K15 and K48 does not directly correspond to a higher suppressor activity among T7 clones. Also, additional analysis of the structure of the DT region in the available tRNA X-ray

conformations shows that for the stability of base pair 19-56, nucleotide 56 should be a pyrimidine, as in the standard tRNA structure, because only in this case, the hydrogen bonds between the two bases are partly protected by purine 57. In the alternative case, when nucleotide 57 is a purine, these hydrogen bonds would be essentially weakened due to their complete exposure to the solvent. As one can see in Table 1, in all five T7 clones, position 56 is occupied by a purine, which indicates the weakness of any hydrogen bond in which this purine is involved. Based on this, we think that as a rule, base pair 19-56 does not exist in the T7 clones, and even if it forms in some clones, this would not improve their activity. In the absence of this base pair, the stacking position of purine 56 on top of purine 57 seems to be the most stable conformation (Fig. 3A).

Extension of the quasi normal pattern for T8- tRNAs

Analysis of the T8-tRNA sequences shown in Table 1 revealed six of them, which could be aligned well with the T7 sequences if one assigns the extra eighth nucleotide in the T8 T-loop to position 3* between nucleotides 55(2*) and 56(4*). Indeed, all these T8 clones contain G2* equivalent to G55 in the T7 clones and purine 5* equivalent to purine 57. In addition, they have at least one adenine in the D-loop able to form a base pair with G2* equivalent to A18-G55. This similarity allows us to combine all these T7 and T8 sequences into Type I. As to nucleotide 4*, we can argue that it is equivalent to nucleotide 56 in the canonical tRNA structure. Indeed, while in the T7 tRNAs this nucleotide was exclusively a purine, among the T8 tRNAs, it is cytidine or adenine, respectively, in four and two cases. Coincidently, in all four sequences containing C4*, one can find dinucleotide AG in the D-loop, which would play the role of dinucleotide 18-19 in the canonical tRNA structure and would allow the formation of base pair G19-C4* equivalent to base pair G19-C56. As one can judge from Table 1, the suppressor activity of all four clones having C4* is higher than that of the two clones having A4*. This demonstrates that in the T8-tRNAs of Type I, unlike in the T7-tRNAs, the existence of base pair G19-C4* correlates with higher suppressor activity, and therefore, position 4* in T8 clones corresponds to position 56 in the normal tRNAs. The extra nucleotide 3* could stack on top of nucleotide 4* (Fig. 3B). In two sequences, K23 and K49, position 3* is occupied by C, which could form a base pair with the G following nucleotide 19. However, we have doubts that this base pair actually exists: first, its formation does not correlate with the suppressor activity of the T8-tRNAs and second, the modeling of these structures showed that

formation of such base pair would place nucleotide 3* too far from G2*, which would force the latter to move away from its optimal position of stacking on the reverse-Hoogsteen base pair U1*-A6*.

Finally, in another clone, K34, the T-loop fits well to other Type I T8 sequences, but the D-loop contains a deletion of five nucleotides. Due to this deletion, the D-loop becomes so short, that the only sterically reasonable interaction that can be formed between the two loops would consist of the base pair between the third nucleotide of the D-loop, coincidently A, and G2*, arranged in the same way as base pair A18-G55(2*) in all other Type I clones. We should admit, however, that this interaction is not optimal, because when the model of this tRNA was submitted to the unrestrained energy minimization, the two helical domains had a tendency to get closer to each other, thus optimizing the inter-loop interactions at the DT region, but changing the standard juxtaposition of the tRNA helical domains (data not shown). This may be the reason for the relatively low suppressor activity of clone K34.

The specific purine trap in the T-loop of the Type I tRNAs

According to the analysis of the Type I sequences presented in the previous part, different Type I clones can form either one or two base pairs between the D- and T-loops corresponding to base pairs 18-55 and 19-56 of the standard tRNA structure. However, only the first base pair has been found in all clones, while the second one exists only in less than half of the Type I clones. This indicates a relative importance of base pair 18-55 for the tRNA structure and function. Intuitively, such difference between the two base pairs seems obvious, because base pair 18-55 makes a part of the core of the T-loop, and therefore, its disruption will affect the whole structure of the DT region. Base pair 19-56, on the contrary, is positioned on the periphery of the T-loop, so that its disruption would have only a mild, local effect. Only for the T8 clones the presence of base pair 19-56(4*) may be essential, because here, the stacking of nucleotide 3*, predominantly adenine, on top of this base pair makes it a part of the internal structure of the DT region. The higher relative importance of base pair 18-55 compared to base pair 19-56 fits also to the fact that in the crystal structure of the yeast $tRNA^{Asp}$ (34), the disruption of base pair 19-56 does not affect the structure of the rest of the DT region.

The formation of base pair 18-55(2*) can be seen from another perspective. One can say that in the T-loop, there is a special trap made of base pair 54(1*)-58(6*) and purine

57(5*) arranged at a double distance from each other. This trap attracts purine 18 from the D-loop to intercalate between them and to form hydrogen bonds with nucleotide 55(2*) at the bottom of the trap. The presence of the latter nucleotide makes the trap specific: in the standard tRNA structure, nucleotide 55 is pseudouridine, and the trap 'catches' guanine. In the Type I clones, nucleotide 55(2*) is guanine, and correspondingly, the trap 'catches' adenine. Because this specific purine trap exists in all Type I clones, it seems to be indispensable for the function of these tRNAs.

Purine trap and reverse-Hoogsteen base pair UA

If the locking of purine 18 into the T-loop purine trap is so critical for the tRNA structure and function, it would be important to know how the trap is formed in the first place. In particular, we would like to know whether the extended distance between purines 57(5*) and 58(6*), which allows purine 18 to get in between, is arranged independently of or only upon the intercalation of the latter. The existing data support the idea that the gap between nucleotides 57 and 58 forms before the intercalation of nucleotide 18 takes place. Thus, in the recently published crystal structure of the modifying enzyme archaeosine tRNA–guanine transglycosylase bound to $tRNA^{Val}$ (35), the D-domain of the tRNA was found to be strongly deformed. In particular, no nucleotides of the D-loop were involved in interaction with the T-loop. The only element of the T-loop that was in contact with the rearranged D-domain was nucleotide 59 of the bulge region. This nucleotide stacked on base pair 23-48, thus forming the contact analogous to that between nucleotide 59 and base pair 15-48 in the standard tRNA structure. Despite this, the conformation of the T-loop was characterized by the double distance between nucleotide 57 and 58, as in the standard tRNA structure, even without any intercalating nucleotide at the place normally occupied by G18.

Our analysis showed that the double distance between nucleotides 57 and 58 originates from the particular conformation of the sugar-phosphate backbone of nucleotides 58 and 59 when the first of them forms the reverse-Hoogsteen base pair with U54, while the second one is involved in the T-loop bulge. In our modeling system, we took the structure of the T stem-and-loop from the yeast $tRNA^{Phe}$ and deleted nucleotides 55 and 56, leaving the other nucleotides untouched. Then, we tried to change the position of purine 57 to make it stacked on base pair 54-58. We found, however, that as long as nucleotides 54 and 58 were involved in the standard reverse-Hoogsteen base pair, while nucleotides 59 and

60 formed the T-loop bulge, the stacking of nucleotide 57 on base pair 54-58 was not sterically possible. Such stacking could only be achieved at the expense of the integrity of the region 58-60, for example, when nucleotides 59-60 were also deleted form the structure. In other words, the standard conformation of region 58-60 does not allow nucleotide 57 to stack on base pair 54-58. Although the position of nucleotide 57 at the double distance from this base pair, characteristic of the canonical tRNA structure, is not the only conformation possible, it has an advantage of allowing a nucleotide from the D-loop to intercalate in the created gap, thus optimizing the stacking interactions. This aspect of the T-loop structure is important for Type II structures, which are discussed in the next part.

Type II as a non-specific purine trap

Among the clones from Table 1 that do not fit to Type I, all but one clone (K30) have A in position 2*. We combine these A2*-containing clones into Type II. If the structure of the T-loop in these clones were formed in the same way as in the Type I clones, there should be a nucleotide in the D-loop at the place of purine 18 able to form a base pair with A2* similar to the standard G18-Ψ55 and to A18-G55, specific to the Type I clones. However, our analysis failed to find a reasonable scheme of base pairing for A2* that would satisfy this requirement. This was taken as an indication that in the Type II clones, the T-loop does not form in the way it forms in the standard and in the Type I tRNAs. Also, in none of the Type II clones, position 5* is occupied by a purine. Instead, among Type II sequences, this position is exclusively occupied by C. Neither A2* nor C5* has ever been seen among the standard or Type I tRNAs, and the fact that they coincide in the Type II clones suggests the existence of an alternative structural pattern. The following considerations were crucial for suggestion of the T-loop structure in the Type II clones. First, as it was just mentioned, if A2* occupies the same position as G2* in the Type I tRNAs, it is unable to find a base pairing partner from the D-loop that would fit reasonably into the structure of the T-loop. Second, for the reasons discussed in the previous section, C5* cannot stack directly on base pair U1*-A6* and has to be positioned at a double distance from it. A reasonable solution that takes both these aspects into account would consist of the formation of a base pair between A2* and C5* within the layer positioned at the double distance from the reverse-Hoogsteen base pair U1*-A6*. This arrangement would allow a nucleotide from the D-loop to intercalate between the two base pairs. The only possible stable structure that can be modeled according to this description is shown in

Fig. 3C. In this structure, base pair A2*-C5* is formed according to the reverse-Hoogsteen scheme, similar to base pair U1*-A6* (Fig. 5). A minimum of two unpaired nucleotides, 3* and 4*, are needed to close the T-loop, which fits well to the fact that among Type II clones, unlike Type I clones, no T7-tRNAs have been found. None of nucleotides 3* and 4* is conserved among the Type II clones. We noticed, however, that nucleotide 3* is almost exclusively a purine, while nucleotide 4*, with only one exception, is either C or A. This observation fits well to the suggested structure of a Type II tRNA shown in Fig. 3C. A purine in position 3* allows an extensive stacking interaction with nucleotide 4* as well as a contact with A2*. Depending on whether nucleotide 3* is G or A, the hydrogen bond can be formed between its extra-cyclic atom O6 or N6 and either N1 or N6 of A2*. The presence of either C or A in position 4* allows the formation of the hydrogen bond between amino group of 4* and atom O2 of C5*. Another hydrogen bond is possible between the amino group of C5* and phosphate 4*. All these interactions make the very top region of the T-loop stable on its own, without any additional interactions with a nucleotide of the D-loop. The absence of such interactions with the D-loop is also supported by the fact that we have not found any significant nucleotide co-variation between nucleotides of the two loops.

The central part of the T-loop, on the contrary, is absolutely conserved among the Type II clones. The sandwiching of a nucleotide from the D-loop between the two reverse-Hoogsteen base pairs U1*-A6* and A2*-C5* provides a strong stabilizing effect for the whole arrangement. The energy minimization of this arrangement in the context of the whole tRNA structure optimized some atom-atom contacts and hydrogen bonds, but did not move substantially any nucleotide. This indicates that the proposed arrangement of the nucleotides is stable and corresponds to an energy minimum. Additional analysis shows that the stability of the arrangement is very sensitive to the structure of both reverse-Hoogsteen base pairs. For example, when we replaced base pair A2*-C5* in this arrangement by a relatively close base pair A2*-U5*, the same energy minimization test broke the structure. In other words, the particular geometry of the reverse-Hoogsteen base pair A2*-C5* is essential for the stability of the arrangement, which can explain the absolute conservation of these nucleotides among the Type II clones.

The nucleotide of the D-loop that intercalates between the two reverse-Hoogsteen base pairs plays an important stabilizing role for the structure. The stabilizing effect comes primarily from the stacking interaction of this nucleotide with each of the two base pairs.

To ensure that this interaction is sufficiently strong to keep the integrity of the arrangement, the intercalating nucleotide must be a purine. Correspondingly, the energy minimization of the structure in which this nucleotide is a pyrimidine destroyed this arrangement. On the other hand, it did not seem to matter whether the intervening nucleotide was A or G, because both purines fitted equally well to the gap between the two reverse-Hoogsteen base pairs. An additional stabilization of the structure would come from interaction of the amino group of this purine with the nearest ribose 2`OH group between the two reverse-Hoogsteen base pairs. As one can see in Table 1, in each Type II clone, there are several purines in the D-loop, and there is no reason to prefer any one of them to another as a candidate for the intercalation.

Finally, among the Type II clones, like in Type I, there is a clone, K41, characterized by a deletion of five nucleotides in the D-loop. Again, this deletion makes the D-loop so short that it is able to provide only one nucleotide, a purine, for the intercalation between the two reverse-Hoogsteen base pairs. This only interaction may be sufficient to keep the integrity of the tRNA and to make it functional. It is worth mentioning, that clone K41, as well as K34, resembles some mitochondrial tRNAs, which also have deletions in the D-loop. Due to this deviation from canonical sequences, such tRNAs are not able to form standard interactions between the D- and T-loops, and their detailed structures are not known. They may have alternative interactions between the D- and T-loops, like the one proposed here for clones K34 and K41, or may just lack these interactions.

We see that the structural block consisting of a D-loop purine, sandwiched between the two T-loop reverse-Hoogsteen base pairs, is shared by all Type II clones. By analogy with the Type I clones, we can look at the gap between the two reverse-Hoogsteen base pairs positioned at the double distance from each other as a trap for a purine from the D-loop. However, in the Type II clones, unlike in the Type I clones, this trap is nonspecific, mostly due to the absence of a nucleotide analogous to nucleotide 55(2*) in Type I tRNAs, which would form hydrogen bonds with the intercalating nucleotide and restrict its identity. This allows us to name the structural block characteristic of the Type II clones as the nonspecific purine trap, which, due to its existence in all Type II clones, seems to be indispensable for their function. For the intervening purine, also by analogy with the Type I clones, we assign number 18.

DISCUSSION

This work represents a new step in our study of the general structural constraints imposed on functional tRNAs. Previously, we have shown that the standard juxtaposition of the two tRNA helical domains requires the presence of exactly two unpaired nucleotides 59-60 in the T-loop bulge, which is achieved by the formation of the reverse-Hoogsteen base pair between positions 54 and 58 (25). However, the fact that the top of the T-loop did not seem to have a common motif in all selected clones meant that a particular conformation of this region was not critical for the function of the suppressor tRNAs. On the other hand, due to the insufficient number of sequences, it was not possible to say whether this region played any important role in the tRNA structure and function and whether any sequence constraints were imposed on it. After the additional selection of more than twenty new suppressor tRNA clones and analysis of their nucleotide sequences, we can conclude that the top of the T-loop is involved in specific interactions with the D-loop. Although more than one conformation has been found for this region, it does not mean that the constraints imposed on the nucleotide sequences of the two loops in functional tRNAs are loose or do not exist. On the contrary, the analysis shows that these constraints are so strict that for those tRNAs having the reverse-Hoogsteen base pair U-A equivalent to base pair 54-58 in the canonical tRNA structure, only two distinct families of sequences are allowed. Each family corresponds to a particular type of the T-loop conformation. The first type, so-called specific purine trap, is a generalized version of the standard T-loop conformation, while the second, the nonspecific purine trap, is described here for the first time.

The specific purine trap is based on two key elements, base pair 18-55(2*) and purine 57(5*). In the tertiary structure, it consists of a gap in the T-loop between the reverse-Hoogsteen base pair 54(1*)-58(6*) and nucleotide 57(5*), which is filled by purine 18 from the D-loop. The requirement for a purine in position 57(5*) comes from the fact that the stacking interactions play the most important role in the formation of this arrangement. The Type I purine trap is characterized by the specificity with respect to the identity of purine 18, which is restricted via formation of base pair 18-55(2*). Only two combinations are allowed for this base pair, GU and AG. These base pairs have been shown to have similar juxtaposition of the glycosidic bonds, which allows the replacement of one

pair by the other without major changes in the backbone conformation. The cytosolic tRNAs contain only GU base pairs, while the AG base pair is found on many other occasions: among mitochondrial tRNAs, in viral tRNA-like structures, as well as in the T-loop-like motifs existing in different tRNAs. Interestingly, in the selected Type I tRNAs we found only the AG base pair, and no GU. We attribute this to the fact that in the randomized positions of the commercially purchased K-library, the four standard nucleotides were not present in equal proportions. In fact, when we determined the nucleotide sequences of forty randomly chosen negative clones (data not shown), which characterize better the distribution of the identities of each nucleotide due to the fact that negative clones constitute about 99.9% of the library, we found that at each randomized position, adenine appeared more often than the other three nucleotides. This disproportion would increase the chance to be selected for those combinations that contain more adenines, like base pair AG versus GU in our case. Compared to the K-library, our next library (26) was randomized more evenly, and correspondingly, the selected tRNAs predominantly contained the U55-G18 base pair. The other inter-loop base pair 19-56(4*), which exists in the canonical tRNA structure, is not present in all selected clones. This base pair is not directly involved in the formation of the purine trap and is positioned on the periphery of the structure. This allows the disruption of base pair 19-56 without jeopardizing the integrity of the whole DT region. The stability of this base pair becomes essential for the tRNA function only when the additional eighth nucleotide 3* in the T-loop stacks on top of this base pair.

The structural type of the DT region that is alternative to the specific purine trap is the non-specific purine trap. It represents a new structural motif never seen before in tRNAs or tRNA-like domains. The key sequence requirements that allow the formation of this motif consist of A2*, C5* and of at least two nucleotides between them. Nucleotides A2* and C5* form the specific reverse-Hoogsteen base pair positioned at the double distance from base pair U1*-A6*, thus allowing purine 18 from the D-loop to intercalate between them. While in the specific purine trap, the necessity of a large surface for sufficiently strong stacking interaction with nucleotide 18 required the presence of a purine in position 57(5*), in the nonspecific purine trap, the same surface is built of two paired nucleotides, 5* and 2*, which allows nucleotide 5* to be a pyrimidine. The additional eighth nucleotide in the T-loop plays a critical role in the integrity of the motif: without it, the T-loop cannot be closed. The predominant identities of nucleotides 3* (a purine) and 4*

(A or C) make their arrangement on top of base pair A2*-C5* (Fig. 3C) stable even without additional interactions with the D-loop. The absence of such additional interactions is also supported by the fact that we have not found any valuable nucleotide co-variation that would indicate the existence of a specific interaction between the two loops. While the nonspecific purine trap is not allowed for T7-tRNA, it becomes at least as probable as Type I for T8-tRNAs, when the additional nucleotide joins the top of the T-loop.

Although the non-specific purine trap is described here for the first time, among the known nucleic acid motifs one can find those that resemble some important aspects of its structure. As one can see in Fig. 6b, in both fragments U1*-A2* and C5*-A6* forming the nonspecific purine trap, the pyrimidines are shifted in the direction of the major groove, while the following adenines are shifted in the direction of the minor groove. This resembles the arrangement of four purines, GA versus GA or AA versus AA, within an anti-parallel duplex (16, 36, Fig. 6a). Here as well, the first purine in each chain is shifted in the direction of the major groove, while the following adenine is shifted to the minor groove. In this motif, as in the non-specific purine trap, specific base-base hydrogen bonds within each base pair are able to stabilize the structure. Another motif having a similarity to the non-specific purine trap is I-DNA, built of four poly(C) chains (37, Fig. 7a). In this structure, two consecutive nucleotides from the two opposite chains are positioned at the double distance from each other and form specific base pairs. Between the two base pairs, nucleotides from the other two chains intercalate. Like I-DNA, the nonspecific purine trap is very sensitive to the geometry of the base pairs, so that a replacement of base pair A2*-C5* by a similar although different base pair AU is very damaging for the structure. The similarity of the non-specific purine trap to these two well-known nucleic acid motifs serves as an additional argument in support of this structure.

In spite of obvious differences, the two structural types of the DT region share important common characteristics. Both have the same lower part of the T-loop, including bulge 59-60, its interaction with the D-domain, and the reverse-Hoogsteen base pair 54(1*)-58(6*). Also, as we showed, the formation of the reverse-Hoogsteen base pair 54(1*)-58(6*) forces nucleotide 57(5*) to be positioned at the double distance from A58(6*), which opens the possibility for nucleotide 18 from the D-loop to intercalate between them. This intercalation, specific in the Type I clones and nonspecific in the Type II clones, is shared by all clones of both groups and thus seems to be indispensable for the tRNA function.

The role of this intercalation for the whole tRNA is of a special interest. Modeling the tertiary structures of the selected tRNAs and their comparison with the standard tRNA structure allows us to suggest that this intercalation is crucially important for the proper juxtaposition of the two helical domains. From our previous experiments (25-26) we know that the key element in the fixation of the L-shape is the T-loop dinucleotide bulge 59-60. This bulge fills the space between the D-domain and the T-stem and thus mediates the interaction between them (Fig. 1). When in some of our mutant tRNAs the D-domain became shorter, correspondingly, the T-loop bulge extended in the way that the total number of the stacked layers in the DT region remained unchanged. This compensation provides the condition for the standard arrangement of the two helical domains. However, because the interaction between the T-loop bulge and the D-domain is based solely on the stacking of base 59 and base pair 15-48, it is able to keep the standard juxtaposition of the helical domains only if the domains are already arranged relatively closely to the standard L-shape. The interaction of nucleotide 18 with the purine trap can help to solve this part of the problem by giving the two loops a chance to associate with each other in the orientation close to the standard.

It is important, however, that although the interaction of nucleotide 18 with the purine trap can bring the two loops together and orient them similar to the standard, it is unable to fix the juxtaposition of the two helical domains. This inability is mostly due to the fact that neither in the canonical tRNA structure, nor in the selected clones is the position of the intervening nucleotide 18 fixed with respect to the D-domain. In the canonical tRNA structure, the G18-G19 dinucleotide is connected to the rest of the D-domain by two regions, 16-17a and 20-20b, which are conformationally flexible and are variable in length. In different selected clones, the intervening purine 18 occupies different position within the randomized region in the D-loop. Moreover, in many clones, there are several purines in the D-loop that can play the role of nucleotide 18, and there is no reason to prefer any one to another. Therefore, the fixation of the tRNA L-shape can be achieved only via cooperation between the interaction of purine 18 with the purine trap and the stacking of the T-loop bulge on the D-domain. In this cooperation, the first element arranges the two domains relatively closely to the standard L-shape, while the second one provides the final fixation.

The interaction of nucleotide 18 with the purine trap can affect the tRNA L-shape in one more way. The proper number of nucleotides in the T-loop bulge, which is critical for

the standard juxtaposition of the helical domains, is guaranteed via the formation of the reverse-Hoogsteen base pair 54-58. On the other hand, as has been shown here, the formation of this base pair forces nucleotide 57 to occupy the position at the double distance from base pair 54-58, creating a gap between the bases. The filling of this gap by nucleotide 18 provides an additional stability for the whole arrangement and, what is most important, for base pair 54-58. This, in turn, stabilizes the structure of bulge 59-60 and indirectly, the whole tRNA L-shape.

Although both structural types have demonstrated the ability to provide for a functional tRNA, only Type I is found in the naturally selected tRNA species. The obvious advantage of the Type I structure deals with the fact that it restricts both the identity of the intercalating purine and its orientation in the complex with the T-loop, which would help to avoid misfoldings and kinetic traps during the formation of the tRNA tertiary structure.

We have analyzed here only two types of the structure of the DT region in the suppressor tRNAs selected from the K-library. There are serious reasons to think that these two types do not exhaust all the possibilities of the structural organization of the DT region. Indeed, at least one clone in Table 1, K30, has combination U54-A58, but does not fit to any of the two structural types. Also, as mentioned in Introduction, 21 K-clone out of 51 selected from the K-library did not have combination U54-A58, but were still functional. Analysis of the nucleotide sequences of these clones could reveal new types of structure of the DT region. Although these types are yet unknown, we can anticipate that they will also, like Types I and II, provide mechanisms for the correct mutual fixation of the two helical domains. The elucidation of these mechanisms will constitute an important step in understanding of the particular role played by irregular unpaired regions in RNA structure and function.

ACKNOWLEDGMENTS

This work was supported by a grant from National Science and Engineering Research Council of Canada. S.V.S. also acknowledges fellowships from Canadian Institutes of Health Research and from Fond de la Recherche en Santé du Québec.

REFERENCES

1. Pley HW, Flaherty KM, McKay DB. (1994) Three-dimensional structure of a hammerhead ribozyme. Nature, 372, 68-74.

2. Hermann T, Patel DJ. (2000) RNA bulges as architectural recognition motifs. Structure Fold. Des., 8, R47-54.

3. Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Szewczak AA, Kundrot CE, Cech TR, Doudna JA. (1996) RNA tertiary structure mediation by adenosine platforms. Science, 273, 1696-1699.

4. Hermann T, Patel DJ. (1999) Stitching together RNA tertiary architecture. J. Mol. Biol., 294, 829-849.

5. Conn GL, Draper DE. (1998) RNA structure. Curr. Opin. Struct. Biol., 8, 278-285.

6. Batey RT, Rambo RP, Doudna JA. (1999) Tertiary motifs in RNA structure and folding. Angew. Chem. Int. Ed., 38, 2326-2343.

7. Moore PB. (1999) Structural motifs in RNA. Annu. Rev. Biochem., 68, 287-300.

8. Westhof E, Fritsch V. (2000) RNA folding: beyond Watson-Crick pairs. Structure Fold. Des., 8, R55-65.

9. Westhof E. (2002) Group I introns and RNA folding. Biochem. Soc. Trans., 30, 1149-1152.

10. Doherty EA, Batey RT, Masquida B, Doudna JA. (2001) A universal mode of helix packing in RNA. Nat.Struct. Biol., 8, 339-343.

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

12. Nissen P, Ippolito JA, Ban N, Moore PB, Steitz TA. (2001) RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. Proc. Natl. Acad. Sci. USA, 98, 4899-4903.

13. Klein DJ, Schmeing TM, Moore PB, Steitz TA. (2001) The kink-turn: a new RNA secondary structure motif. EMBO J., 20, 4214-4221.

14. Heus HA, Hilbers CW. (2003) Structure of non-canonical tandem base paires in RNA helices: review. Nucleosides, Nucleotides, Nucleic Acids, 22, 559-571.

15. SantaLucia J Jr, Kierzek R, Turner DH. (1991) Stabilities of consecutive A.C, C.C, G.G, U.C and U.U mismatches in RNA internal loops: evidence for stable hydrogen-bonded U.U and C.C+ pairs. Biochemistry, 30, 8242-8251.

16. SantaLucia J Jr, Turner DH. (1993) Structure of (rGGCGAGGCC)2 in solution from NMR and restrained molecular dynamics. Biochemistry, 32, 12612-12623.

17. Suddath FL, Quigley GJ, McPherson A, Sneden D, Kim JJ, Kim SH, Rich A. (1974) Three-dimensional structure of yeast phenylalanine transfer RNA at 3.0 angstroms resolution. Nature, 248, 20-24.

18. Robertus JD, Ladner JE, Finch JT, Rhodes D, Brown RS, Clark BF, Klug A. (1974) Structure of yeast phenylalanine tRNA at 3 A resolution. Nature, 250, 546-551.

19. Nazarenko, I.A., Harrington, K.M. and Uhlenbeck, O.C. (1994) Many of the conserved nucleotides of tRNAPhe are not essential for ternary complex formation and peptide elongation. EMBO J., 13, 2464-2471.

20. Puglisi JD, Putz J, Florentz C and Giege R. (1993) Influence of tRNA tertiary structure and stability on aminoacylation by yeast aspartyl-tRNA synthetase. Nucleic Acids Res., 21, 41-49.

21. Levinger L, Bourne R, Kolla S, Cylin E, Russel K, Wang X and Mohan A. (1998) Matrices of paired substitutions show the effects of tRNA D/T loop sequence on Drosophila RNase P and 3'-tRNase processing. J. Biol. Chem., 273, 1015-1025.

22. Du X. and Wang ED. (2003) Tertiary structure base pairs between D- and T?C-loops of Escherichia coli tRNALeu play important roles in both aminoacylation and editing. Nucleic Acids Res. 31, 2865-2872.

23. Bullock TL, Sherlin LD and Perona JJ. (2000) Tertiary core rearrangements in a tight binding transfer RNA aptamer. Nature Struct. Biol., 7, 497-504.

24. Zagryadskaya EI, Doyon FR, Steinberg SV. (2003) Importance of the reverse Hoogsteen base pair 54-58 for tRNA function. Nucleic Acids Res. 31, 3946-3953.

25. Zagryadskaia EI, Kotlova N, Steinberg SV. Keeping the tRNA L-shape with the help of the reverse-Hoogsteen base pair 54-48. Submitted.

26. Shi H and Moore PB. (2000) crystal structure of yeast phenylalanine tRNA at 1.93 A resolution: a classic structure revisited. RNA, **6**,1091-1105.

27. Pearlman DI. *et al.* (1995) AMBER 4.1, University of California, San Francisco, CA.

28. Dirheimer G, Keith G, Dumas P and Westhof E. Primary, secondary, and tertiary structures of tRNAs. (1995) In: tRNA: Structure, Biosynthesis, and Function, (Söll D and Rajbhandary UL, eds.), ASM Press, Washington, DC, pp. 93-126.

29. Florentz C and Giege R. tRNA-like structures in plant viral RNAs. (1995) In: tRNA: Structure, Biosynthesis, and Function, (Söll D and Rajbhandary UL, eds.), ASM Press, Washington, DC, pp. 141-163.

30. Krasilnikov AS and Mondragon A. (2003b) On the occurrence of the T-loop RNA folding motif in large RNA molecules. RNA, 9, 640-643.

31. Lee JC, Cannone JJ and Gutell RR. (2003) The lonepair triloop: a new motif in RNA structure. J. Mol. Biol. 325, 65-83.

32. Nagaswamy, U. and Fox, G.E. (2002) Frequent occurrence of the T-loop RNA folding motif in ribosomal RNAs. RNA, 8, 1112-1119.

33. Peterson, E.T., Blank, J., Sprinzl, M. and Uhlenbeck, O.C. (1993) Selection for active E. coli $tRNA^{Phe}$ variants from a randomized library using two proteins. EMBO J., 12, 2959-2967.

34. Moras D, Comarmond MB, Fischer J, Weiss R, Thierry JC, Ebel JP and Giege R. (1980) Crystal structure of yeast $tRNA^{Asp}$. Nature, 288, 669-674.

35. Ishitani R, Nureki O, Nameki N, Okada N, Nishimura S and Yokoyama S. (2003) Alternative tertiary structure of tRNA for recognition by a posttranscriptional modification enzyme. Cell, 113, 383-394.

36. Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA. (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science*, **273**, 1678- 1685.

37. Gehring K, Leroy JL, Gueron M. (1993) A tetrameric DNA structure with protonated cytosine.cytosine base pairs. *Nature*, **363**, 561-565.

LEGENDS TO THE FIGURES

Figure 1. The structure of the DT region in the context of the tRNA L-form.

Rectangles represent individual nucleotides of the DT region involved in stacking or base pairing with another nucleotide. All other nucleotides are shown as black dots. The identities of the conservative and semi-conservative nucleotides are indicated; R stands for purine, Y – for pyrimidine, Ψ – for pseudouridine. There are two base pairs G18-Ψ55 and G19-C56 formed between the D- and T-loops. The reverse-Hoogsteen base pair T54-A58, whose structure is seen in Fig. 5, is formed within the T-loop. Dinucleotide 59-60 bulges

from the double helical stem between base pairs G53-C61 and T54-A58. Nucleotide 59 stacks to the tertiary base pair 15-48 constituting the last layer of the D/anticodon helical domain. This interaction fixes the perpendicular arrangement of the two helical domains called L-form.

Figure 2. Design of the combinatorial tRNA gene library.

The design of the K-library was based on the sequence of $tRNA^{Ala}_{CUA}$. Six randomized nucleotide positions (indicated by circles) were introduced in each of the D- and T-loops. In the T-loop, these positions are numbered from 1* to 6*. The *EcoRI* and *PstI* restriction sites flanking the 5' and 3' termini were used for cloning the library.

Figure 3. Comparison of the DT region structure in the Type I and Type II tRNAs.

On the left: the schematic representation of the DT region in the context of the whole tRNA L-shape. On the right: a stereoview of the DT region in the corresponding three-dimensional model. Those nucleotides that are discussed in the paper are represented by rectangles on the left and shown explicitly on the right. For the same nucleotide, the same color is used on the scheme on the right and the model on the left.

a. T7-tRNA of Type I.

b. T8-tRNA of Type I.

c. tRNA of Type II.

Figure 4. Comparison of base pairs G18-U55 and A18-G55.

Structure of the base pair G18-U55 present in canonical tRNAs is shown on the left. In Type I tRNAs it is replaced by base pair A18-G55, for which one of the possible structures is shown on the right. Another possible structure may be formed by flipping the A base and forming a hydrogen bond between its N7 atom and the amino group of G. AG base pairs are not completely isosteric to the G18-U55 base pair, but they have similar positions of glycosidic bonds, and due to the flexibility of the D-loop structure, nucleotide 18 should be able to intercalate into the specific purine trap in the T-loop and form a base pair with G55.

Figure 5. Comparison of the two reverse-Hoogsteen base pairs existing in Type II tRNAs.

Reverse-Hoogsteen base pair U1*-A6*, equivalent to standard base pair T54-A58, is shown

to the left, while A2*-C5* base pair is shown on the right. Note that both base pairs have similar geometry, but in the A-C base pair C is shifted to the minor groove compared to U in the U-A pair. In the structure of the non-specific purine trap, the base pair A2*-C5* is positioned above the base pair A1*-U6*, with the purine 18 intercalated between them (see Figures 6b,c and 7).

Figure 6. Comparison of the non-specific purine trap (Type II) with the AA-AA motif.

a. Stereo view of the tandem AA-AA base pairs from the crystal structure of group I intron RNA (36, pdb entry 1GID, nucleotides A113-A114 and A206-A207). One RNA strand is shown in black, the opposite strand in grey. Both AA base pairs are formed in the same way. They are called 'sheared' base pairs, because one nucleotide in the pair is shifted towards the major groove, while the other one is shifted to the minor groove. Note extensive cross-strand stacking of two adenines from the opposite strands. Similar structure exists for tandem GA-GA base pairs.

b. Stereo view of dinucleotides U1*-A2* (in grey) and C5*-A6* (in black) forming the non-specific purine trap in Type II tRNAs. The nucleotide from the D-loop intercalating between the two base pairs is not shown. Base pair U1*-A6* is farther away, while base pair A2*-C5* is closer to the reader. Nucleotides forming base pairs are shifted in the same directions as in the AA-AA motif: pyrimidines are shifted to the major grove and purines to the minor grove. Stacking between nucleotides from opposite strands is replaced by stacking with intercalated purine 18.

c. Same structure as in **b,** but with intercalating purine 18 shown in grey.

Figure 7. Comparison of the non-specific purine trap (Type II) with I-DNA.

a. Stereo view of the fragment of the I-DNA (37, pdb entry 225D). Two paired DNA strands are shown in black; the other two paired strands, which intercalate between the base pairs of the first two strands, are shown in grey.

b. Stereo view of the non-specific purine trap in Type II tRNA. Nucleotides 1*-2* and 5*-6* of the T-loop are shown in black, the intercalating purine 18 is grey. This single intercalating purine plays the same structural role as the intercalating base pair in the I-DNA.

LEGEND TO TABLE 1

(a) These sequences have been reported previously (24).

Only those sequences from the K-library that had the U1*-A6* combination are shown. The tRNA sequences outside the D- and T-loops were identical in all clones and are not shown. Randomized regions are highlighted in bold. In the D-loop and T-loops, they correspond, respectively, to regions 16-19 and 54-58 of the standard tRNA structure. For the T8-tRNAs, alternative numbering of randomized positions in the T-loop is indicated (1* to 6*). The nucleotides of the T-loop presumed to form the reverse-Hoogsteen base pair U1*-A6* are underlined. Suppressor activity of tRNAs was evaluated by measuring β-galactosidase activity in XAC-1 strain of *E.coli*, which results from the suppression of the stop codon in the β-galactosidase gene. Reported activity is the mean value of three independent measurements. The suppressor activity of the amber suppressor $tRNA^{Ala}_{CUA}$ was taken for 100%. Background β-galactosidase activity in the XAC-1 cells was below 0.01%.

Table 1. Sequences of selected tRNAs and their suppressor activities.

Clone	D-loop	T-loop	% β-gal activity
$tRNA^{Ala}_{CUA}$	AG CUGG GA	UU CGAUC	100
Type I			
T7-tRNAs	14 / 15 21	54 56 58 60 / 55 57 59	
K25[a]	AG**GAACGC**UA	**UG AAA**C	17.7
K15[a]	AG**GCAUAU**UA	**UG AAA**UC	11.0
K29[a]	AG**GAAAAA**UA	**UG GGA**UC	5.1
K3[a]	AG**AACGAA**UA	**UG AAA**UC	4.1
K48	AG**ACAGAC**UA	**UG GAAA**C	2.5
T8-tRNAs	14 / 15 21	1* 3* 5* 7* / 2* 4* 6* 8*	
K2[a]	AG**AAAGAC**UA	**UGACGA**UC	7.9
K23[a]	AG**UAAGGU**UA	**UGCCAA**UC	5.9
K42	AG**UAGACA**UA	**UGACAA**UC	3.8
K39	AG**GAAGAA**UA	**UGACGA**UC	3.4
K49	AG**AAAAGA**UA	**UGCAAA**UC	2.2
K12[a]	AG**CUACAU**UA	**UGAAAA**UC	0.4
K34	AG**AG** A	**UGGCGA**UC	1.4
Type II			
	14 / 15 21	1* 3* 5* 7* / 2* 4* 6* 8*	
K27[a]	AG**UGAAAU**UA	**UAGCCA**UC	9.9
K24[a]	AG**AAAAAC**UA	**UAGCCA**UC	6.0
K26[a]	AG**AACGAC**UA	**UAAACA**UC	3.9
K47	AG**AGAGCA**UA	**UAGCCA**UC	3.3
K20[a]	AG**GAGAUC**UA	**UAGCCA**UC	3.2
K40	AG**AAACAC**UA	**UAGCCA**UC	2.8
K36	AG**AAAAAA**UA	**UAGCCA**UC	2.7
K18[a]	AG**AACAAA**UA	**UAAACA**UC	2.5
K33	AG**GAAGAA**UA	**UAGUCA**UC	2.0
K19[a]	AG**ACAAC** UA	**UAUACA**UC	2.0
K5[a]	AG**CGAAGA**UA	**UAGCCA**UC	1.7
K38	AG**AAAUAC**UA	**UAAACA**UC	1.5
K51	AG**ACCAAA**UA	**UAACCA**UC	1.4
K7[a]	AG**GACAAA**UA	**UAACCA**UC	1.3
K1[a]	AG**GAGAAC**UA	**UAACCA**UC	1.3
K41	AG**AC** A	**UAAACA**UC	4.4
Non-identified type			
	14 / 15 21	54 56 58 60 / 55 57 59	
K30[a]	AG**UGAGGA**UA	**UC CAAA**U	10.8

Figure 1

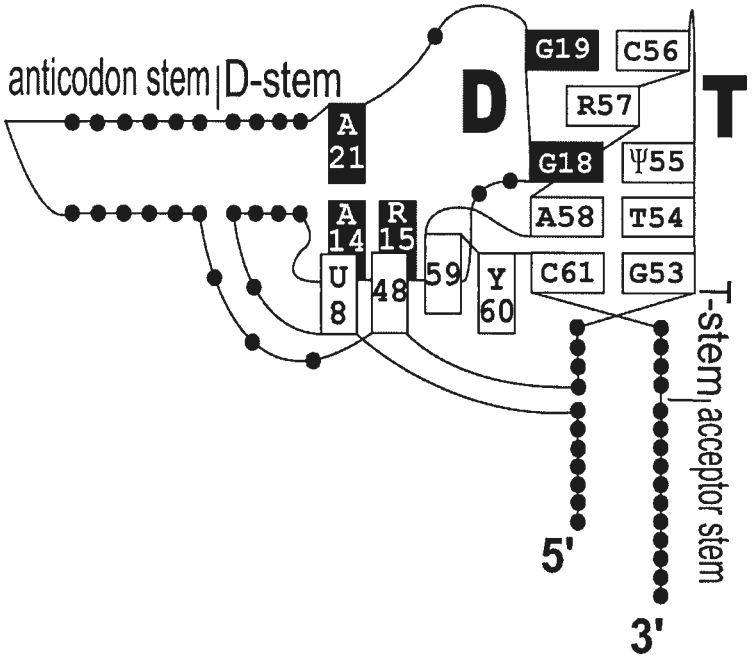


Figure 2

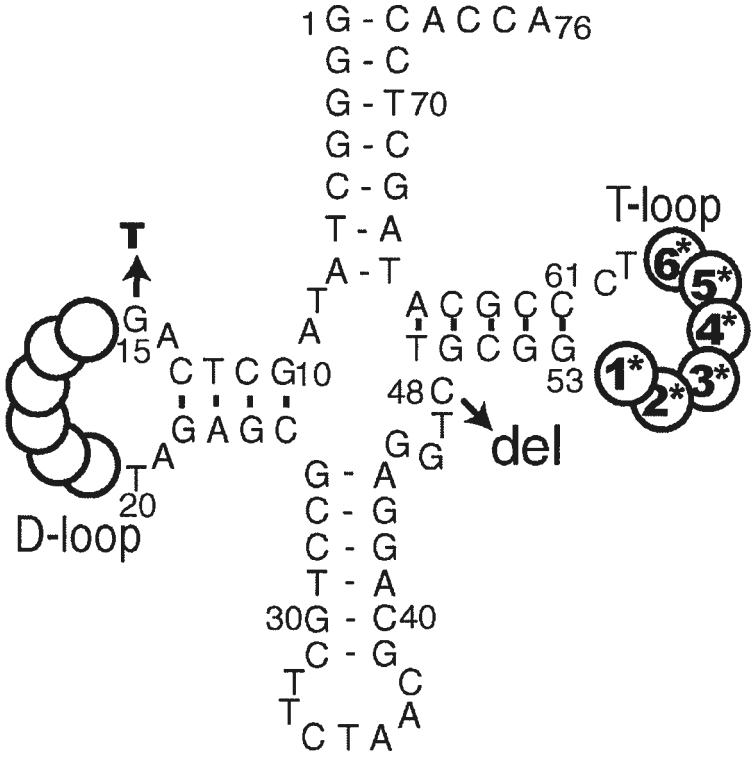


Figure 3

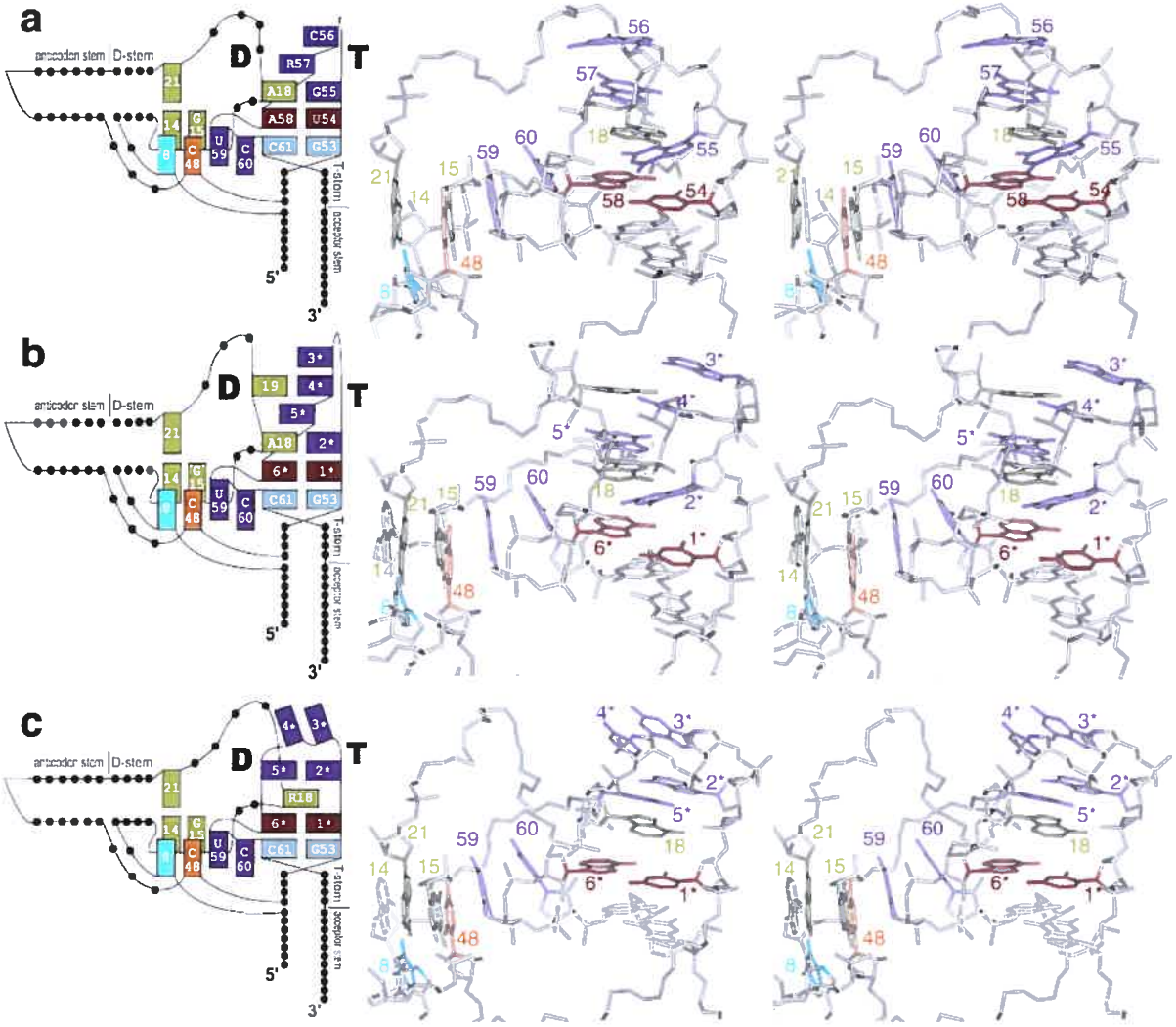

a
b
c
anticodon stem
D-stem
D
T
T-stem
acceptor stem
5'
3'
C56
R57
A18
G55
A58
U54
C61
G53
19
4*
5*
2*
6*
1*
R18
3*

Figure 4

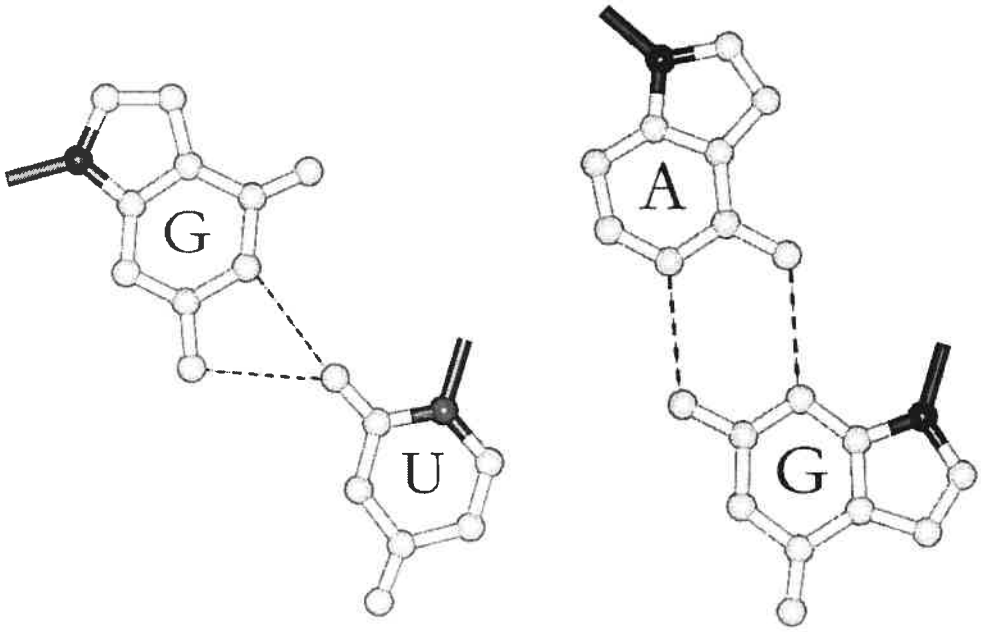


Figure 5

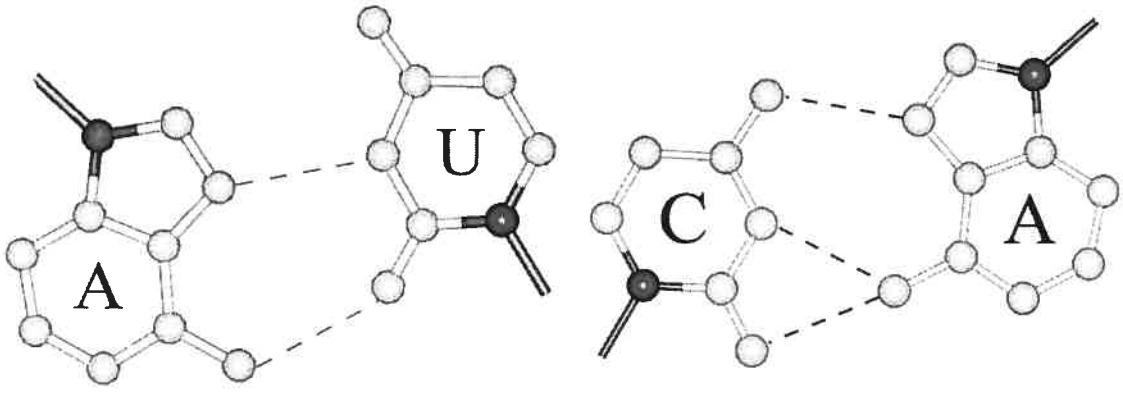


Figure 6

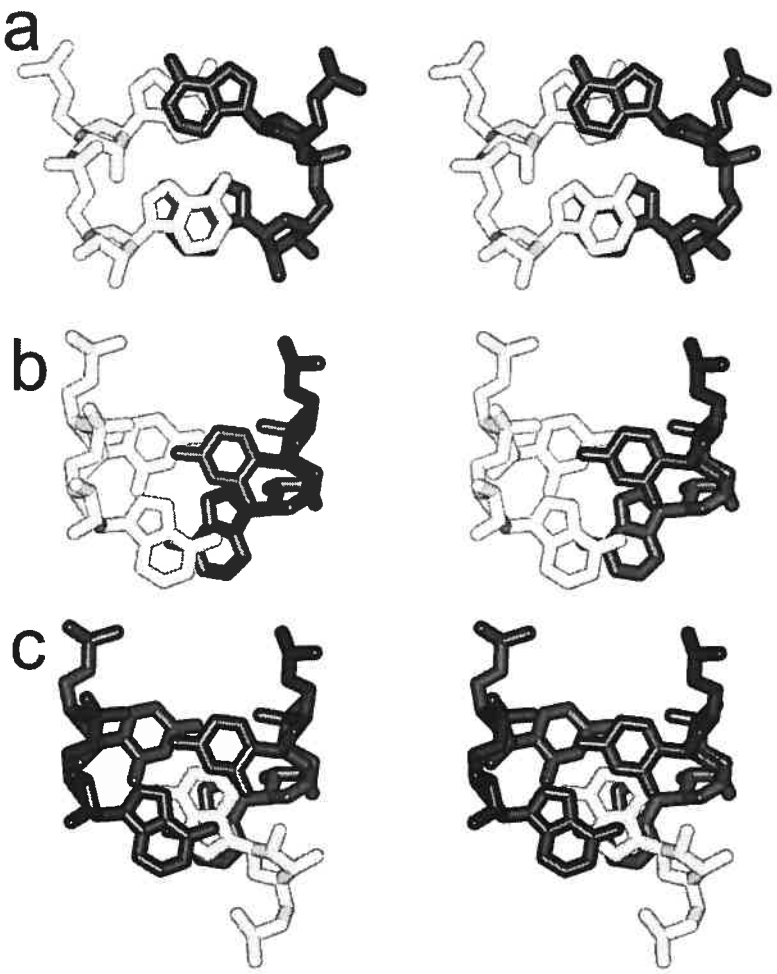
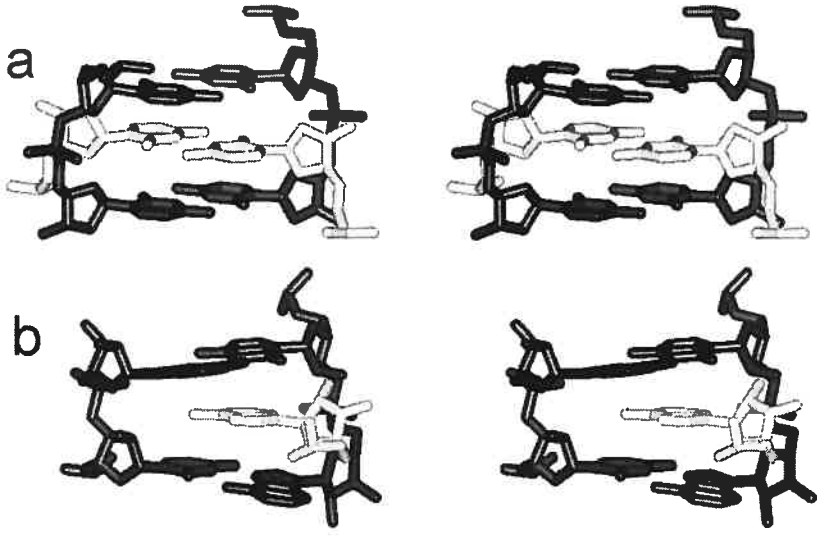


Figure 7



DISCUSSION

Conservation of the L-shape of tRNA

Our results obtained from screening of the M-library (Chapter 2) clearly demonstrate that tRNA has to preserve its L-shape in order to be functional. This conclusion arises from the fact that the overwhelming majority of selected tRNAs presented structural compensations to the artificial disruption of the L-shape by deletion of the tertiary base pair 15-48. L-shape has to be maintained with a very high degree of precision, since the deletion of even one stacked layer, i.e. shortening of about 3Å, is detrimental for tRNA function.

The idea about the importance of the L-shape for the tRNA function is as old as the L-shape itself: when the first tRNA crystal structure was determined, it was realized that L-shape fold was essential for tRNA to work in protein synthesis (Holbrook *et al*, 1978). Determination of several other tRNA structures had further strengthened this opinion. Since that time a lot of experimental evidence has been accumulated, which shows that disruption of the L-shape is detrimental for many steps in tRNA metabolism. Our study once again confirms this long known fact, but in addition, it also shows several alternative ways that tRNA can use to maintain the L-shape. Discovery of these ways is a completely new knowledge, since they have not been seen in naturally occurring tRNAs.

We do not know exactly how the tRNA L-shape evolved, and why it was preferred to other types of structures, but the advantages of all tRNAs having the same tertiary structure is obvious: it provides basis for uniform interaction of all tRNAs with the factors involved in tRNA metabolism and protein biosynthesis. At the same time, the specificity of recognition by different aminoacyl-tRNA synthetases can be achieved through differences in nucleotide sequences and in subtle structural features. It has been proposed, based on studies of aminoacylation of RNA mini- and microhelices, that the acceptor/T domain of tRNAs has evolved first, and that the anticodon/D domain evolved later on (Martinis and Schimmel, 1995). If this is true, there would be a problem at some point to find the way of connecting the anticodon/D domain to the existing acceptor/T domain. The way this connection is done in existing tRNAs is quite elaborated: the exact structure of the tRNA core is very complex, and it became known only after applying the methods of X-ray crystallography, since all earlier models were wrong and could not even predict the actual

shape of tRNA. Nowadays, many more structures of longer functional RNAs are known, like ribosomal RNA (Yusupov *et al*, 2001) or RNase P (Krasilnikov *et al*, 2003a), and they reveal even higher complexity of RNA tertiary folds. However, in these RNAs we can find some elements resembling parts of tRNA, like the T-loop motif (Lee *et al*, 2003; Krasilnikov and Mondragon, 2003b). It means that tRNAs use the same folding strategies common to all RNAs. This makes us believe that conclusions of our work can be extended from tRNA to RNA folding in general.

Importance of the reverse-Hoogsteen base pair for tRNA structure and function

Though it has been acknowledged since long that the L-shape was essential for tRNA function, it was believed that the most important role in keeping the L-shape was played by the two base pairs between the D- and T-loops, Ψ55-G18 and C56-G19. Several *in vitro* mutagenesis experiments were done to study the implication of these base pairs into tRNA function. It was discovered that disruption of these base pairs has destabilizing effects on tRNA structure, as revealed by chemical probing (Puglisi *et al*, 1993; Levinger *et al*, 1995 and 1998) and reduction in tRNA cleavage by lead (Behlen *et al,* 1990), as well as on the rates of aminoacylation (Puglisi *et al*, 1993; Du and Wang, 2003) and 5' or 3' processing (Levinger *et al*, 1995 and 1998).

With much attention drawn to the two inter-loop base pairs, the role of the reverse-Hoogsteen base pair (RH) 54-58 in the T-loop in keeping the tRNA L-shape has not been recognized. This base pair was known to be important for stabilizing the native conformation of the T-loop (Romby *et al,* 1987), but the role of the stacking interaction between the T-loop bulge and the D-domain was not realized. This may be due to the fact, that since this interaction is the same in all canonical tRNAs, its importance could not be discovered through sequence comparison. Only analyzing non-canonical mitochondrial tRNAs made it possible to reveal the importance of stacking of nucleotide 59 to the D-domain (Steinberg *et al,* 1997). But even then, the role of the RH was not anticipated. However, the results of selection of suppressor tRNAs from our first combinatorial tRNA gene library (K-library, Chapter 1), made it possible to reveal the conservation and importance of the RH. Trying to put this knowledge together with the idea of importance of stacking between nucleotide 59 and the tertiary base pair 15-48 in the D-domain led us to the hypothesis about the structural role of the RH in tRNA folding. According to this hypothesis, the RH contributes to the maintenance of the L-shape by keeping the correct

length and position of the two-nucleotide bulge of the T-loop, and thus provides the stacking of nucleotide 59 to the base pair 15-48. To test this hypothesis, the second tRNA gene library (M-library, Chapter 2) was designed, where the base pair 15-48 was deleted. Screening of this library confirmed that our hypothesis about the role of RH was right, as we observed exactly the type of structural compensation that was predicted by the hypothesis: the pattern of RH formation switched in such a way as to add one more nucleotide to the T-loop bulge to take the place of the deleted base pair 15-48 and to restore the stacking between the two helical domains. These results prove the importance of the stacking interaction involving nucleotide 59 and base pair 15-48 in keeping the L-shape of tRNA and highlight the role of the RH in providing for this stacking interaction. It is possible that other RNA molecules use similar strategy, that is, use unpaired nucleotides for stacking with helical stems to provide correct juxtaposition of different domains and attain final tertiary structure.

Alternative structures of the DT region of tRNA

tRNAs that we have selected from the two combinatorial gene libraries show remarkable variety in the sequences of the D- and T-loops. In the first K-library, for instance, none of selected tRNAs had the TΨC sequence in the T-loop, which is otherwise absolutely conserved in all eubacterial, plant, and eukaryotic cytoplasmic tRNAs, as well as in many mitochondrial tRNAs. Instead, structural modeling showed at least two distinct structural types among the tRNAs from the K-library (Chapter 3), one of which resembled closely the standard structure of the DT region (type I), while the other one was very different from it. In addition to these two types described in Chapter 3, the third one probably also exists in the tRNAs which have a purine-purine combination 54-58 in the T-loop. Structural modeling of these tRNAs is now in progress, and preliminary results show, that their DT region structures differ from both Type I and II. Thus, the standard structure of the DT region is not an absolute requirement for the tRNA functioning, at least in the regime of suppression of nonsense codons. However, since all selected tRNAs showed lower suppressor activities than the control tRNA, which had the canonical structure, we should admit, that having standard sequence and structure gives tRNA certain advantages over non-standard structures, because it evolved together with all other parts of the protein synthesis machinery and probably represents the best fit to them.

In contrast to the K-library, most clones selected from the M-library had the standard pattern of D-T loop interaction, though several sequences belonging to type II were also observed. This probably results from general weakening of the tRNA structure by deletion of the base pair 15-48. In such a situation, the standard structure of the DT region may provide additional stabilizing effect on the tRNA tertiary structure and is thus preferred to alternative structures.

Another interesting fact about alternative structures of the DT region is that their number seems to be quite limited. Based on this, one can conclude, that the possible conformational space for productive D-T loop interaction is very limited. However, it is not restricted solely to the canonical pattern.

Present work is the first study showing that tRNAs with alternative structures of the DT region can be functional in protein synthesis in bacteria. Before, only mitochondrial (mit) tRNAs were thought to be able to do this. Many mit tRNAs cannot form canonical structure because they lack nucleotides necessary for that. Little is known about the structure of the DT region in such mit tRNAs, but several modeling studies had suggested alternative structural patterns (de Bruijn and Klug, 1983; Steinberg *et al*, 1994b). We think that our experimental results can help understand the structures of the DT region in many mit tRNAs. It would be exiting to look whether the structures of the DT region in tRNAs selected from our combinatorial libraries overlap with possible mit tRNA structures.

Advantages of combinatorial method

The type of results that were obtained during this study and the conclusions that were made would not be possible without using the combinatorial approach. Only this approach allowed the screening of complete sequence space of the DT region and identification of general structural requirements for this region and of alternative structural patterns. Since alternative structures are very different from the canonical one and require multiple nucleotide substitutions, and even an increase in the length of the T-loop, they can hardly be detected by other methods, like random mutagenesis. Success in isolation of compensatory mutants from the M-library would also not be possible without the introduction of additional eighth nucleotide into the T-loop. Lengthening of the D-loop by two nucleotides has also contributed to the success of the experiment by broadening the possibility for structural adaptation.

We should mention, however, that one should be very careful while making a design of a combinatorial RNA library and constrain randomization to relatively short regions, to be able to interpret the results. The structural scaffold of RNA should also be maintained in order to provide basis for structural modeling and comparison. Randomization of longer regions can lead to problems with alignment of selected sequences and with structural interpretation of results.

The combinatorial approach has been successfully applied to chemical synthesis, drug design, protein selection and *in vitro* selection of nucleic acids. Our study shows that it can also be very useful for *in vivo* studies of functional RNAs, when combined with structural modeling. Development of new *in vivo* screening systems suitable for studies of functional RNAs will certainly increase our knowledge about RNA structure and function and help to understand general rules that govern RNA folding.

CONCLUSIONS

1. Screening of the two combinatorial tRNA gene libraries with randomized sequences in the D- and T- loops for the presence of functional suppressor tRNAs showed, that only tRNAs which maintained the L-shaped tertiary structure were functional. This confirms previous observations that conservation of the L-shape is a major requirement for tRNA functioning.
2. Reverse-Hoogsteen base pair between nucleotides 54 and 58 in the T-loop of tRNA plays essential role in tRNA architecture and maintenance of the L-shape by providing the bulged conformation of nucleotides 59 and 60 of the T-loop and their stacking to the D-domain of tRNA, which is necessary for the correct juxtaposition of the two helical domains of tRNA.
3. Canonical pattern of interaction of the nucleotides at the top of the T-loop with their partner nucleotides in the D-loop is not required for suppressor activity of tRNA. Canonical interactions can be replaced by alternative interactions as long as the L-shape of tRNA is maintained.
4. Use of combinatorial RNA libraries is a promising method to study the rules that govern RNA structure formation.

REFERENCES (Introduction and Discussion)

Asahara H, Nameki N and Hasegawa T. (1998) *In vitro* selection of RNAs aminoacylated by *Escherichia coli* leucyl-tRNA synthetase. *J. Mol. Biol.*, **283**, 605-618.

Ban N, Nissen P, Hansen J, Moore PB and Steitz TA. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 resolution. *Science*, **289**, 905-920.

Basavappa R and Sigler PB. (1991) The 3 crystal structure of yeast initiator tRNA: functional implications in initiatior/elongator discrimination. *EMBO J.*, **10**, 3105-3111.

Baskerville S, Frank D and Ellington AD. Directed evolutionary descriptions of natural RNA structures. In: *RNA structure and function* (Simons RW and Grunberg-Manago M, eds.) Cold Spring Harbor Laboratory Press, pp. 203-251, 1998.

Behlen LS, Sampson JR, DiRenzo AB and Uhlenbeck OC. (1990) Lead-catalyzed cleavage of yeast $tRNA^{Phe}$ mutants. *Biochemistry*, **29**, 2515-2523.

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

Bourdeau V, Steinberg SV, Ferbeyre G, Emond R, Cermakian N and Cedergren R. (1998) Amber suppression in *Escherichia coli* by unusual mitochondria-like transfer RNAs. *Proc. Natl. Acad. Sci. USA* **95**, 1375-1380.

de Bruijn MH and Klug A. (1983) A model for tertiary structure of mammalian mitochondrial transfer RNAs lacking the entire 'dihydrouridine' loop and stem. *EMBO J.*, **2**, 1309-1321.

Bullock TL, Sherlin LD and Perona JJ. (2000) Tertiary core rearrangements in a tight binding transfer RNA aptamer. *Nature Struct. Biol.*, **7**, 497-504.

Choi H, Gabriel K, Schneider J, Otten S and McClain WH. (2003) Recognition of acceptor-stem structure of $tRNA^{Asp}$ by *Escherichia coli* aspartyl-tRNA synthetase. *RNA*, **9**, 386-393.

Crick FHC. (1966) The genetic code - yesterday, today and tomorrow. *Sprig Harbor Symp. Quant. Biol.*, **31**, 3-9.

Dirheimer G, Keith G, Dumas P and Westhof E. Primary, secondary, and tertiary structures of tRNAs. In: *tRNA: Structure, Biosynthesis, and Function,* (Söll D and Rajbhandary UL, eds.), ASM Press, Washington, DC, pp. 93-126, 1995

Du X. and Wang ED. (2003) Tertiary structure base pairs between D- and TΨC-loops of *Escherichia coli* $tRNA^{Leu}$ play important roles in both aminoacylation and editing. *Nucleic Acids Res.* **31**, 2865-2872.

Goldgur Y, Mosyak L, Reshetnikova L, Ankilova V, Lavrik O, Khodyreva S, Safro M. (1997) The crystal structure of phenylalanyl-tRNA synthetase from *Thermus thermophilus* complexed with cognate $tRNA^{Phe}$. *Structure*, **5**, 59-68.

Gutmann S, Haebel PW, Metzinger L, Sutter M, Felden B and Ban N. (2003) Crystal structure of the transfer-RNA domain of transfer-messenger RNA in complex with SmpB. *Nature*, **424**, 699-703.

Helm M, Brule H, Friede D, Giege R, Putz J and Florentz C. (2000) Search for characteristic structural features of mammalian mitochondrial tRNAs. *RNA*, **6**, 1356-1279.

Hoagland MB, Stephenson ML, Scott JF, Hecht LI, Zamecnik PC. (1958) A soluble ribonucleic acid intermediate in protein synthesis *J. Biol. Chem.*, **231**, 241-257.

Hoang C and Ferre-D'Amare AR. (2001) Cocrystal structure of a tRNAΨ55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. *Cell*, **107**, 929-939.

Holbrook SR, Sussman JL, Warrant RW and Kim SH. (1978) Crystal structure of yeast phenylalanine transfer RNA. II. Structural features and functional implications. *J. Mol. Biol.* **123**, 631-660.

Holley RW. (1957) An alanine-dependente, ribonuclease-inhibited conversion of AMP to ATP, and its possible relationship to protein synthesis. *J. Am. Chem. Soc.*, **79**, 658-662.

Holley RW, Apgar J, Everett GA, Madison JT, Marquisee M, Merrill SH, Penswick JR and Zamir A. (1965) Structure of a ribonucleic acid. *Science*, **147**, 1462-1465.

Ishitani R, Nureki O, Nameki N, Okada N, Nishimura S and Yokoyama S. (2003) Alternative tertiary structure of tRNA for recognition by a posttranscriptional modification

enzyme. *Cell*, **113**, 383-394.

Jovine L, Djordjevic S, Rhodes D. (2000) The crystal structure of yeast phenylalanine tRNA at 2.0 A resolution: cleavage by Mg(2+) in 15-year old crystals. *J. Mol. Biol.*, **301**, 401-414.

Krasilnikov AS, Yang X, Pan T, Mondragon A. (2003a) Crystal structure of the specificity domain of ribonuclease P. *Nature*, 421, 760-764.

Krasilnikov AS and Mondragon A. (2003b) On the occurrence of the T-loop RNA folding motif in large RNA molecules. *RNA*, **9**, 640-643.

Lee JC, Cannone JJ and Gutell RR. (2003) The lonepair triloop: a new motif in RNA structure. *J. Mol. Biol.* **325**, 65-83.

Lee N, Bessho Y, Wei K, Szostak J and Suga H. (2000) Ribozyme-catalysed tRNA aminoacylation. *Nature Struct. Biol.*, **7**, 28-33.

Levinger L, Vasisht V, Greene V, Bourne R, Birk A and Kolla S. (1995) Sequence and structure requirements for Drosophila tRNA 5'- and 3'-end processing. *J. Biol. Chem.*, **270**, 18903-18909.

Levinger L, Bourne R, Kolla S, Cylin E, Russel K, Wang X and Mohan A. (1998) Matrices of paired substitutions show the effects of tRNA D/T loop sequence on Drosophila RNase P and 3'-tRNase processing. *J. Biol. Chem.*, **273**, 1015-1025.

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

Li Z, Gillis KA, Hegg LA, Zhang J and Thurlow DL. (1996) Effects of nucleotide substitutions within the T-loop of precursor-tRNAs on interaction with ATP/CTP:tRNA nucleotidyltransferases from Escherichia coli and yeast. *Biochem. J.*, **314**, 49-53.

Lorsch JR and Szostak JW. (1996) Chance and necessity in the selection of nucleic acid catalysts. *Accounts of Chemical Research,* **29**, 103-110.

Martinis SA and Schimmel P. Small RNA oligonucleotide substrates for specific

aminoacylation. In: *tRNA: Structure, Biosynthesis, and Function,* (Söll D and RajBhandary UL, eds.), ASM Press, Washington, DC, pp. 349-370, 1995.

Moras D, Comarmond MB, Fischer J, Weiss R, Thierry JC, Ebel JP and Giege R. (1980) Crystal structure of yeast $tRNA^{Asp}$. *Nature*, **288**, 669-674.

Murakami H, Saito H and Suga H. (2003) A versatile tRNA aminoacylation catalyst based on RNA. *Chemistry and Biology*, **10**, 655-662.

Nissen P, Kjeldgaard M, Thirup S, Polekhina G, Reshetnikova L, Clark BF, Nyborg J. (1995) Crystal structure of the ternary complex of Phe-tRNAPhe, EF-Tu, and a GTP analog. *Science*, **270**, 1464-1472.

Peterson ET, Blank J, Sprinzl M and Uhlenbeck OC. (1993) Selection for active *E.coli* $tRNA^{Phe}$ variants from a randomized library using two proteins. *EMBO J.*, **12**, 2959-2967.

Puglisi JD, Putz J, Florentz C and Giege R. (1993) Influence of tRNA tertiary structure and stability on aminoacylation by yeast aspartyl-tRNA synthetase. *Nucleic Acids Res.*, **21**, 41-49.

Robertus JD, Ladner JE, Finch JT, Rhodes D, Brown RS, Clark BF, Klug A. (1974) Structure of yeast phenylalanine tRNA at 3 A resolution. *Nature*, **250**, 546-551.

Romby P, Carbon P, Westhof E, Ehresmann C, Ebel JP, Ehresmann B and Giege R. (1987) Importance of conserved residues for the conformation of the T-loop in tRNAs. *J. of Biomolecular Structure and Dynamics*, **5**, 669-687.

Rould MA, Perona JJ, Soll D, Steitz TA. (1989) Structure of E. coli glutaminyl-tRNA synthetase complexed with tRNA(Gln) and ATP at 2.8 A resolution. *Science*, **246**, 1135-1142.

Ruff M, Krishnaswamy S, Boeglin M, Poterszman A, Mitschler A, Podjarny A, Rees B, Thierry JC, Moras D. (1991) Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA(Asp). *Science*, **252**, 1682-1689.

Saenger W. *Principles of nucleic acid structure.* Springer-Verlag, New York, 1984.

Schevitz RW, Podjarny AD, Krishnamachari N, Hughes JJ, Sigler PB and Sussman JL. (1979) Crystal structure of a eukaryotic initiator tRNA. *Nature*, **278**, 188-190.

Shi H and Moore PB. (2000) The crystal structure of yeast phenylalanine tRNA at 1.93 A resolution: a classic structure revisited. *RNA,* **6**, 1091-1105.

Soll D and RajBhandary UL. Transfer RNA in its fourth decade. In: *tRNA: Structure, Biosynthesis, and Function,* (Söll D and Rajbhandary UL, eds.), ASM Press, Washington, DC, pp. 1-4, 1995.

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

Steinberg S, Gautheret D and Cedergren R. (1994b) Fitting the structurally diverse animal mitochondrial $tRNAs^{Ser}$ to common three-dimensional constraints. *J. Mol. Biol.*, **236**, 982-989.

Steinberg S, Leclerc F and Cedergren R. (1997) Structural rules and conformational compensations in the tRNA L-form. *J. Mol. Biol.*, **226**, 269-282.

Suddath FL, Quigley GJ, McPherson A, Sneden D, Kim JJ, Kim SH, Rich A. (1974) Three-dimensional structure of yeast phenylalanine transfer RNA at 3.0 angstroms resolution. *Nature*, 248, 20-24.

Vortler S, Putz J and Giege R. (2001) Manipulation of tRNA properties by structure-based and combinatorial *in vitro* approaches. *Prog. Nucleic Acid Res. Mol. Biol.*, **70**, 291-334.

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

Wilson D S and Szostak JW. (1999) In vitro selection of functional nucleic acids. *Annu. Rev. Biochem.*, **68**, 611-647.

Wimberly BT, Brodersen DE, Clemons WM Jr, Morgan-Warren RJ, Carter AP, Vonrhein C, Hartsch T and Ramakrishnan V. (2000) Structure of the 30S ribosomal subunit. *Nature*, **407**, 327-339.

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

Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, Cate JH and Noller HF. (2001) Crystal structure of the ribosome at 5.5 resolution. *Science*, **292,** 883-896.