Université de Montréal

The study of RNA tertiary interactions in tRNA structure and function

Par
Tetsu M. Ishii

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 Bio-informatique
Mars, 2014

Université de Montréal

Faculté des études supérieures et postdoctorales

Cette thèse intitulée:

The study of RNA tertiary interactions in tRNA structure and function

présentée par :

Tetsu M. Ishii

a été évaluée par un jury composé des personnes suivantes:

Dr. Pascal Chatrand, président-rapporteur

Dr. Serguei Chteinberg (Sergey Steinberg), directeur de recherche

Dr. Marlene Oeffinger, membre du jury

Dr. Martin Schmeing, examinateur externe

Dr. François Robert, représentant du doyen

RÉSUMÉ

Le rôle des deux paires de bases universelles inverse Hoogsteen U : A (RHUAs) présentent chez les ARNt standards , une dans la boucle T et l'autre dans le noyau de la forme en L , a été étudiée. Pour chacun des RHUAs , un criblage génétique spécialisé in vivo chez les bactéries , le système suppresseur ambre (pour l'étude de la RHUA dans la boucle T) et le système d'ARNt de la sélénocystéine (tRNASec) (pour l'étude de la RHUA dans le noyau) , ont été utilisé pour générer des variants fonctionnels à partir de multiples librairies combinatoires . Ces variants ont ensuite été séquencé et soumis à une analyse systématique qui comprend la modélisation informatique et un type d'analyse phylogénétique. Les résultats du système suppresseur ambre ont montré un ensemble de variants fonctionnels qui ne nécessitent pas le motif RHUA dans la boucle T et qui ont remplacé la méthode standard de l'interaction entre les boucles D et T avec une double hélice interboucle , ILDH . D'autres études ont abouti à la détermination d'un modèle In silico de l'alternative à la norme standard de la boucle T, sous le nom de type III . Les résultats du système tRNASec ont révélé que pour cette ARNt exceptionnel, l'absence de RHUA (dans le noyau) assure une flexibilité accrue qui est spécifiquement nécessaire pour la fonction de tRNASec . Ainsi, les ARNt standards , à la différence de tRNASec , avec la présence universelle de RHUA dans le noyau , a été naturellement sélectionnée pour être rigide . Pris ensemble, la RHUA joue un rôle essentiel dans la stabilisation des interactions tertiaires.

Mots-clés : Evolution instantanée , la modélisation informatique , ARNt , sélénocystéine , tRNASec , T- boucle , RHUA , ARN pliage

SUMMARY

The role of two universally present reverse Hoogsteen U:A base pairs (RHUAs) in the T-loop and in the core of the L-shape of standard tRNA was studied. To study each of the RHUAs, bacterial *in vivo* genetic screens were used including the amber suppressor system (for the study of the RHUA in the T-loop) and the selenocysteine tRNA(tRNASec) system (for the study of the RHUA in the core). These screens generated functional variants from multiple combinatorial libraries. These variants were subsequently sequenced and subjected to a systematic analysis which included computer modeling and a type of phylogenetic analysis. The results from the amber suppressor system showed a set of functional variants which did not require the RHUA motif in the T-loop, and had replaced the standard way of interaction between the D and T loops with an interloop double helix, ILDH. Further study culminated in the determination of an *insilico* model of the alternative to the standard T-loop known as type III. The results from the tRNASec system revealed that for this exceptional tRNA, the absence of RHUA (in the core) ensures an enhanced flexibility that is specifically required for tRNASec function. Thus standard tRNAs, unlike tRNASec, with the universal presence of RHUA in the core have been naturally selected to be rigid. Taken together, RHUA plays an essential role in the stabilization of tertiary interactions.

Keywords: Instant Evolution, computer modeling, tRNA, selenocysteine, tRNASec, T-loop, RHUA, RNA folding

Table of contents page

List of Tables page

List of Figures page

List of Abbreviations

3D	three dimensional
A	Adenosine
Amp	ampicillin
Å	Angstrom
β-gal	beta-galactosidase
bp	base pairs
C	Cytidine
DNA	deoxyribonucleic acid
DT region	region where the D and T-loops meet in the tRNA
E.coli	*Escherichia coli*
EF-Tu	elongation factor Tu
FDH_N	formate dehydrogenase N
FDH_H	formate dehydrogenase H
ILDH	interloop double helix
LB	Luria-Bertani
G	Guanosine
GTP	guanine tri phosphate
MD	Molecular Dynamics
mRNA	messenger RNA
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	protein data bank
Ψ	pseudouridine
R	purine
RH	reverse Hoogsteen base pair
RH-RWC	four base pairs boxed in figure 4 from chapter III

RHUA	reverse Hoogsteen U:A base pair
RNA	ribonucleic acid
rRNA	ribosomal RNA
RWC	reverse Watson-Crick base pair
S	Svedberg
Sec	selenocysteine
SelB	selenocysteine tRNA elongation factor
SECIS	selenocysteine insertion sequence
T	5-methyluridine
tRNA	transfer RNA
$tRNA^{Ala}$	alanine tRNA
$tRNA^{Ala}_{CUA}$	alanine tRNA with CUA anticodon
tRNAPhe	phenylalanine tRNA
tRNASec/$tRNA^{Sec}$	selenocysteine tRNA
U	Uridine
WC	Watson-Crick
WT	wildtype
Y	pyrimidine

Acknowledgements: I would like to express my sincere gratitude to Dr. Serguei V. Chteinberg for showing me how his science is done. Despite having some experience in graduate-level research prior to entering Serguei's lab, the methods and the structural reasoning that was used here was completely foreign to me and it has taken a long time to even feel comfortable and be able to comprehend fully his previously published papers. I most value the time I was able to discuss structural concepts. He has a very vivid and unique vision as to how science should be done, and I appreciate that he has shared his strong views with me. Thank you for motivating me to find a structural reasoning out of a seemingly pattern-less pile of sequences, by telling me with conviction and then showing me later the pattern that was there, saying there is always a reason behind a set of sequences. I greatly appreciate and feel lucky to have been able to see the process of determining how possibly the ribosome has evolved. I would like to thank all colleagues that I was lucky enough to have been able to work with, which made my journey through the Serguei lab experience more rich (Kostya, Jianghong, Felix, Matthieu, Natalia, Luda, Benoit, Fatou, François, Geneviève, Alka, Katia). The times I spent supervising students were also very rewarding for me and thus thank those students who put up with me (Franck, Damien, Katerina-Beata, Stephane). I would especially like to thank my colleague and friend Yury. Thank you Stephane for translating the summary into french. Elaine, the secretary of the department, is the best. I appreciate Marie Pageau's support for all bioinformatic software related problems. Thank you to Ginette and Ernest for the maintenance of the wet lab. I would also like to thank my parents and brothers for all their support. Finally, I don't think I would have made it to the end if not for the support from my wife Zarah; thank you.

CHAPTER I: Introduction

Ribonucleic acids (RNA) are polymeric macromolecules made from the polymerization of monomer ribonucleotides. A ribonucleotide has three basic components: a ribose sugar, a phosphate group, and a nitrogenous base. There are 4 chemically distinct bases: G (guanosine), A (adenosine), C (cytidine), and U (uridine); the first two are known as purines while the latter two are known as pyrimidines. As such, because of the many possible variations in length and in sequence identity in the polynucleotide chain, a very wide array of different RNA can be formed.

Despite the multiple cellular functions assigned to RNA (Alberts et al., 2002), the predominant role of RNA is in the synthesis of proteins as almost all the RNA that resides in a cell (both in prokaryotes and eukaryotes) is comprised of ribosomal RNA (rRNA), transfer RNA (tRNA), and messenger RNA (mRNA; Milo et al., 2010); each type of RNA playing an integral part in the formation of proteins (Liljas and Ehrenberg, 2013; Milo et al., 2010). Outside the cell, RNA is found in some viruses as the carrier of genetic material instead of DNA (Alberts et al., 2002).

RNA are commonly categorized into either coding (mRNA) or non-coding RNA (all other RNA). Unlike the coding RNA, which generally only require a linear arrangement for the coding of protein on the ribosome, the non-coding RNA are known to require a specific folded conformation for their function as has been well documented for rRNA and tRNA (Liljas and Ehrenberg, 2013). Being less than 100 nucletides, the tRNA is a small non-coding RNA that deliver specific amino acids to the ribosome in response to a cognate codon on the mRNA (Liljas and Ehrenberg, 2013). Despite over five decades of study on a relatively small non-coding RNA such as tRNA, how its primary sequence determines its shape is still not well understood (Crick, 1958). Thus, understanding the rules that govern how the primary sequence determines the final 3D fold of RNA remains as one of the fundamental questions in biology.

Phylogenetic Analysis

RNA secondary structure refers to the helical parts of the RNA molecule formed through Watson-Crick (WC) base pairing. Such structures have been identified through an approach known as the phylogenetic comparative analysis based on the examination of aligned homologous RNA genes from many species. Through such an approach, a compensatory evolution can be observed through the identification of RNA secondary structural elements found by the detectable pattern of nucleotide substitutions in such an alignment. This detectable pattern are covarying sites that differ between two or more species but keep the potential for WC base-pairing in each of the species (i.e. a AU dinucleotide pair in species 1 is replaced by an GC dinucleotide pair in species 2). The phylogenetic comparative method has been an effective approach to identifying conserved RNA secondary structures of many non-coding RNAs and has been used, for instance, to generate a consensus cloverleaf structure for tRNAs (Holley et al., 1965).

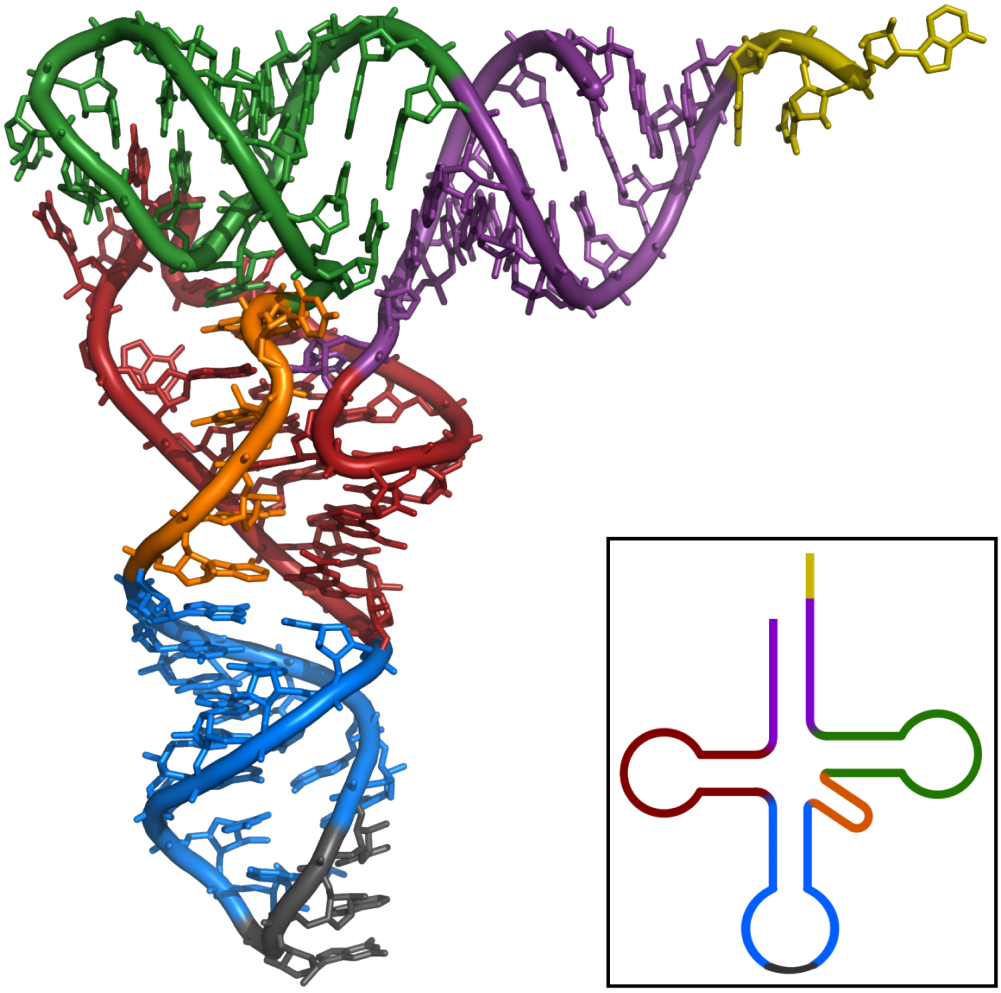


Figure 1: tRNA structure as cloverleaf secondary structure (boxed panel) and in the 3D L-shape (main figure): CCA tail (yellow), Acceptor stem (purple), Variable loop (orange), D arm (red), Anticodon arm (blue), Anticodon (black), T arm (green).

RNA motifs

A large portion of our understanding on RNA structure comes from the study of recurrent structural motifs. These RNA motifs, are found in the same and/or different molecules and have identical or similar 3D conformation. The fact that such motifs can form in mutiple structural instances strongly suggest that they can fold on their own. Therefore, elucidating the aspects important for RNA motif formation is key to the understanding of how the greater RNA molecule folds. RNA motifs can range in size with some having more than 13 nucleotides and being quite complex like the G-ribo RNA motif (Steinberg and Boutorine, 2007), while others can be smaller like the U-turn motif (Quigley and Rich, 1976). The U-turn motif has a three nucleotide UNR consensus sequence (where N is any nucleotide and R is a purine; figure 3, middle magenta). Another small RNA motif is the UA handle motif where an unusual U:A reverse Hoogsteen base pair (RHUA; figure 2) is always found stacked to a classic Watson-Crick base pair, and has a bulge of one or more nucleotides that can act as a handle for making different types of long-range interactions (Jaeger et al., 2009). As both the U-turn and the UA handle motifs are relatively small, it is not so surprising to find them as part of other bigger RNA motifs. Such an example is found in the RNA motif known as the T-loop motif; labelled because of its discovery in the T-loop of tRNAs (Krasilnikov and Mondragon, 2003). This motif is composed of five-nucleotides that includes a U-turn motif flanked by a noncanonical base pair (Krasilnikov and Mondragon, 2003). All 3 mentioned small RNA motifs are found highly conserved in an overlapping way in the T-loop and T-stem of cytosolic tRNAs (Jaeger et al., 2009; Krasilnikov and Mondragon, 2003; Quigley and Rich, 1976).

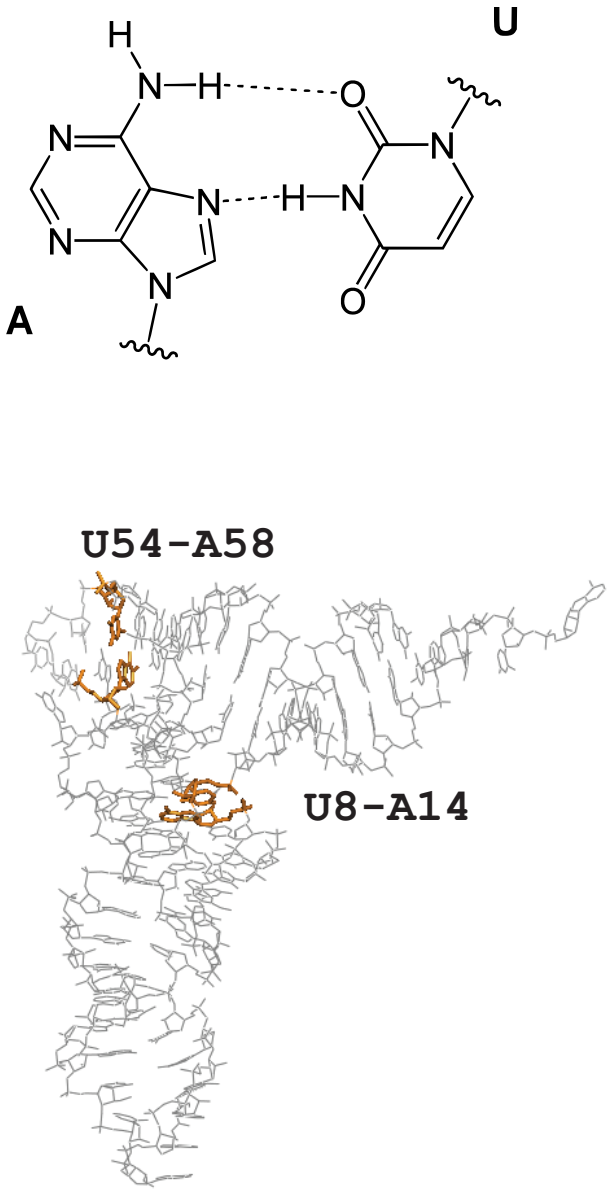


Figure 2: The top panel shows the reverse Hoogsteen base pair U A (RHUA). In the bottom panel, within the L-shape of the tRNA the two highly conserved RHUA in orange are shown; one in the T-loop formed by nucleotides U54 and A58, and the other RHUA located in the core of the tRNA formed by nucleotides U8 and A14.
(pdb file: 4TNA)
http://commons.wikimedia.org/wiki/File:Hoogsteen_base_pair_AU_reverse.svg

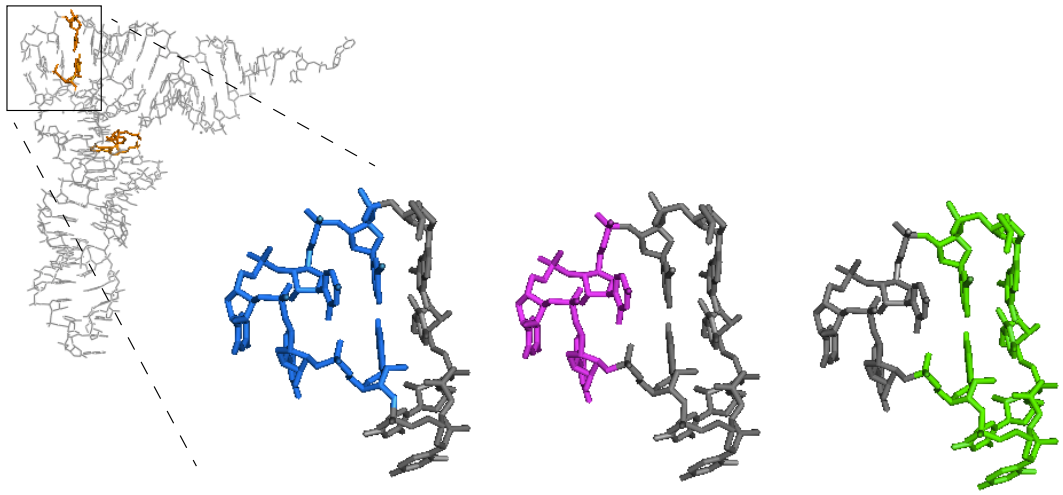


Figure 3: The L-shape tRNA is in the upper left. The T-loop and the first base pair of the T-stem is boxed. The boxed area is enlarged and shown three times to show: the T-loop RNA motif (blue, left), the U-turn motif (magenta, center), and the UA-handle motif (green, right).

In silico RNA studies: the use of Insight II

Another recent approach to the study of RNA structure is the modeling of RNA *in silico*. Insight II is a commonly utilized graphic molecular modeling program that is used in conjunction with the molecular mechanics and dynamics program such as Discover or CHARMm. Such a software allows the building and manipulating of 3D virtual biomolecules and to study their molecular properties. This provides an opportunity to test particular RNA models of interest *in silico* and in this way can help determine which aspects of the model may contribute or oppose in the formation of an optimal structure. An optimal stucture *in silico* can be defined as those structures with the lowest energy under energy minimization simulations, while under molecular dynamics simulations, a particular RNA model would be considered stable if it can maintain its structure and withstand the designed energy fluxuations placed over a given time (Biosym/MSI, 1995).

RNA folding through the study of tRNA

Despite being less than 100 nucleotides in length, the tRNA contains 3 known RNA motifs (T-loop, UA-handle, and U-turn) that reside in the well characterized L-shape structure demonstrated repeatedly by atomic resolution crystals of different tRNAs from different species (Söll et al., 1995). The distinct L-shape allows the tRNA to simultaneously interact with the codon on the mRNA and with the peptidyl transferase center in the large ribosomal subunit during protein synthesis (Liljas and Ehrenberg, 2013). These known aspects of tRNA coupled with the ease of manipulation by molecular cloning techniques due to its short gene length makes tRNA a good candidate for the study of RNA folding.

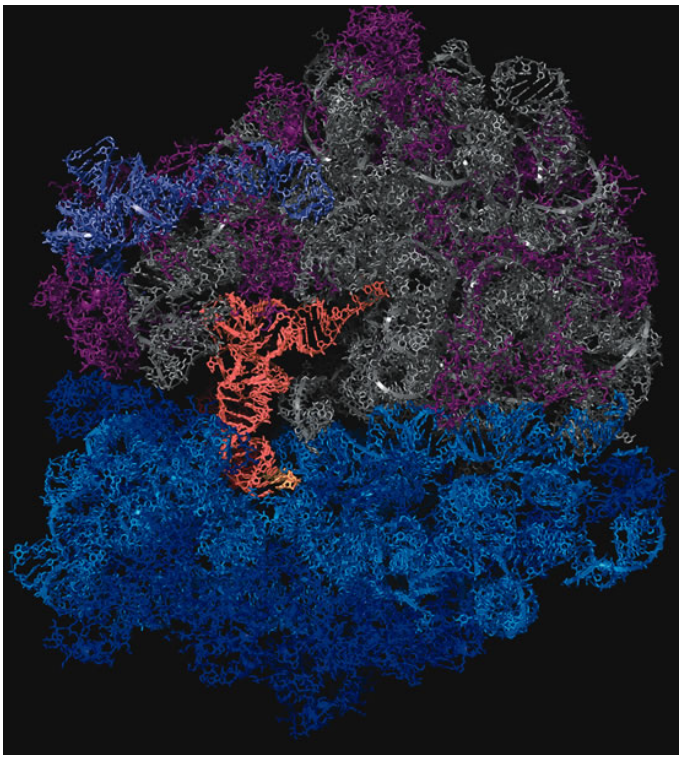


Figure 4: Cross-section of the bacterial ribosome with a tRNA in the P-site. Large rRNA (grey), ribosomal proteins associated with the 23s rRNA (magenta), 5s RNA (ocean blue), 16s rRNA (light blue), ribosomal proteins associated with the small rRNA (blue), mRNA (yellow), and P-site tRNA in the orientation shown above (orange).(Korostelev et al., 2006)

tRNA L-shape and the two unusual RHUAs

Within the L-shape structure of tRNA there are two highly conserved RHUAs; the first, as previously mentioned, is in the T-loop as part of the UA-handle submotif formed by nucleotides U54 and A58 (Juhling et al., 2009), while the other RHUA forms the core of the molecule by nucleotides U8 and A14 (figure 2, bottom). The presence of such a unique base pair in these two specific locations in the tRNA suggests that it plays an important role in the folding of the tRNA L-shape.

The problem of using extant molecules to study RNA folding

Even though a particular tRNA gene may be functional *in vivo*, it may not be functional enough for the viability of the organism that harbours it, and therefore, the molecule will be lost through natural selection. In other words, the RNA examples that we observe today in crystals and sequences identified by genome sequencing, are only but a small subset of all that is possible for a functional variant of a particular RNA molecule, and therefore represent a narrow range of examples that can be used for study. For the purpose of understanding the subtleties of how a particular sequence can affect the RNA stucture and function, it would be useful to study variants that are still functional yet are structually far from the wildtype. For instance, to understand the importance of each of the RHUAs in the tRNA L-shape, it would be useful to identify variants that deviate from this requirement but still be able to deliver on their primary function.

Never before seen functional tRNA variants by Instant Evolution

Instant Evolution refers to the experimental *in vivo* screening of functional variants of a particular gene of interest from a designed pool of combinatorial gene variants (Lee et al., 1997; Zagryadskaya et al., 2003). In my case, two genetic approaches were used to obtain multiple sets of functional tRNA variants. In the first case, to address the role of the RHUA present in the T-loop of tRNAs, an amber suppressor tRNA alanine system was used (Normanly et al., 1986; Zagryadskaya et al., 2003).

The anticodons of alanine suppressor tRNAs have mutations that complement stop codons which allow them to deliver amino acids in response to an UAG stop signal on the mRNA (Normanly et al., 1986). In our case, we have chosen

this particular type of suppressor tRNA. Not all tRNAs can be manipulated in this way, as many synthetases require the interaction with the anticodon in order for the tRNA to be charged with an amino acid (Schimmel, 1987). However for the alanine synthetase, the sole identity element resides in a non-standard GU base pair positioned at 3-70 in the acceptor-stem of the tRNA and therefore results in the proper charging with alanine despite changes to the anticodon(Hou and Schimmel, 1988; McClain and Foss, 1988). Moreover, alanine suppressor tRNA is an ideal tRNA for our studies since the synthetase does not interact with either the D or the T-loop and therefore, the outcome of the combinatorial libraries that target these regions in the tRNA should not be influenced by the way this tRNA is charged.

Thus, in this setup, a pool of alaninyl suppressor tRNAs combinatorially randomized in the D and T-loops were cloned into an expression plasmid and introduced into the XAC-1 bacterial strain which has been genetically modified to have a constitutively expressed beta-galactosidase gene with an in-frame UAG stop codon in the genome (Normanly et al., 1986). Therefore, from the pool of randomized tRNAs, those variants that were still able to function, i.e. deliver alanine in response to the UAG stop codon to the ribosome, were identified by the colonies that showed beta-galactosidase activity and were subsequently sequenced and analyzed in the ways described above.

Because the D and T-loop structure is highly conserved among all living things, it is difficult to understand what elements in either the D or the T-loop are important for structure and function and for what reason. Therefore, using the screening system above we hoped to gain some structural insight by obtaining variants widely different from the wildtype sequence yet were still able to function. The results of those studies are described in chapters II and III.

For the second section of my thesis, I studied an unusual tRNA that carries an equally unusual amino acid Selenocysteine (Sec). Sec is known as the 21st amino acid cotranslationally incorporated into a growing peptide. This amino acid is similar to serine (Ser) and cysteine, differing only by the presence of the selenium atom which replaces oxygen and sulphur, respectively (Yoshizawa and Bock, 2009). All three elements reside in the same column on the periodic table of which

selenium having the greatest ability to donate electrons. Like all other amino acids, Sec has its own tRNA called selenocysteine tRNA (tRNASec). However unlike standard amino acids, the mechanism of delivery to the ribosome is highly unusual. In the first step in bacteria, tRNASec is charged with Ser by the standard serine synthetase and is followed by the conversion to Sec by protein SelA(Yoshizawa and Bock, 2009). Once tRNASec is charged by Sec, it can now bind to a specialized elongation factor SelB which specifically recognizes the presence of the Sec on the tRNA(Forchhammer et al., 1989). This unique factor is an ortholog of EF-tu and therefore also depends on the presence of GTP. This ternary complex (Sec-tRNASec/SelB/GTP) then is delivered to the ribosome in response to a UGA stop on the mRNA in conjunction with a downstream(Zinoni et al., 1987; Zinoni et al., 1990) secondary structure called the Sec insertion sequence (SECIS) which has been suggested to interact with the last protein domain of SelB(Li et al., 2000; Yoshizawa and Bock, 2009). Proteins that are made with Sec are known as selenoproteins which are known to catalyze reduction/oxidation type reactions. The relatively lower electronegative character of selenium in Sec appears to play an integral part in the mechanism of catalysis of selenoproteins as Sec resides in the catalytic pocket and performs 100 fold better than when cysteine is present(Arner, 2010; Yoshizawa and Bock, 2009).

In line with the unusual way Sec is brought to the ribosome, the structure of the tRNASec is also very unusual. Unlike standard tRNAs, tRNASec has a long acceptor stem, a long D-stem, a long extra-arm but most intriguingly, tRNASec lacks the universally conserved tertiary interaction 8-14 made from a RHUA arrangement as well as tertiary interaction 15-48(Schön et al., 1989). We hypothesized that the absence of the universally present and centrally located tertiary interactions in the L-shape of the tRNA, would make a tRNA less rigid, and therefore would suggest tRNASec function may require a certain level of flexibility greater than that of standard tRNAs.

To test such hypothesis, we have generated an *in vivo* genetic screen that would allow to find functional yet different from the wildtype tRNASec in bacteria. To screen for functional variants of tRNASec, I cloned multiple combinatorial

libraries randomized in certain key positions in the tRNASec gene into an expression vector and introduced it into a bacterial strain devoid of the endogenous tRNASec (WL81460 strain; Zinoni et al., 1990). Those variants still able to deliver selenocysteine in response to the UGA stop and SECIS were screened based on the expression of the endogenous selenoprotein, formate dehydrogenase N (FDHn; (Barrett et al., 1979). The sets of *in vivo* functional variants were again analyzed in the ways described above and the results of those investigations are described in detail in chapters IV, V, and VI.

CHAPTER II

Active Suppressor tRNAs with a Double Helix between the D- and T-loops

Natalia Kotlova, Tetsu M. Ishii, Ekaterina I. Zagryadskaya[1], Sergey V. Steinberg*

Département de Biochimie, Université de Montréal,

Montréal, Québec, Canada H3C 3J7

*Corresponding author

Natalia Kotlova and Ekaterina I. Zagryadskay performed the wet experiments.
Tetsu M. Ishii analyzed and created the computer models and wrote the paper.
Sergey V. Steinberg designed experiments, performed analysis and wrote the paper.
This paper was published. J Mol Biol. 2007 Oct 19;373(2):462-75.

Abstract

One of the most conserved elements of the tRNA structure is the reverse-Hoogsteen base-pair T54–A58 in the T-loop, which plays a major role in the maintenance of the standard L-shape conformation. Here, we present the results of in vivo selection of 51 active suppressor tRNA clones, none of which contains base-pair T54–A58. In 49 clones, we found two regions in the D and T-loops that are complementary to each other. This finding suggests the existence of an inter-loop double helix consisting of three base-pairs, which could have the same role as base-pair T54–A58 in the fixation of the juxtaposition of the two helical domains within the L-shape. From this point of view, the appearance of the inter-loop double helix represents a compensatory effect for the absence of base-pair T54–A58. The results shed new light on the role of different elements of the tRNA structure in the formation of the standard L-shape conformation and on the possibility of synonymous replacements of one arrangement by another in functional RNA molecules.

Abbreviations used

WC, Watson–Crick; RH, reverse Hoogsteen

Keywords

L-shape; molecular modeling; selection *in vivo*; T-loop; tRNA

Introduction

Elucidation of the principles that govern the formation of RNA tertiary structure and enable this structure to perform a particular function is fundamental for molecular biology. Transfer RNA is a very useful tool for the study of these problems. On one hand, tRNA is a relatively small molecule that can be easily manipulated genetically. On the other hand, it is big enough to possess certain characteristics usually attributed to large RNAs with distinct tertiary structure. In particular, tRNA is characterized by a specific perpendicular arrangement of the two helical domains D/anticodon and acceptor/T, known as the L-shape. This arrangement guarantees that the two functional centres of the tRNA, the anticodon and the acceptor terminus, are juxtaposed in the way required for the normal tRNA function. Analysis of the tRNA structure suggests that the major events leading to the formation of the L-shape occur in the so-called DT region located at the elbow of the molecule, where the D and T-loops meet. This region contains several highly conserved elements, including the U-turn between Ψ55 (Ψ stands for pseudouridine) and C56, the unusual non-Watson–Crick (WC) base-pairs T54–A58 (T is 5-methyluridine) and G18–Ψ55, the mutual intercalation of fragments 57–58 and 18–19, and the bulging of nucleotides 59–60 (Figure 1). The presence of such elements makes the DT region one of the most structurally diverse in the whole tRNA. Given that arrangements similar to the DT region have been identified in tRNA and in RNase P and ribosomal RNA,[1., 2., 3., 4. and 5.] the question of how the communication between these elements results in the formation of the tRNA L-shape can be important for understanding structure–function relationships in these molecules as well.

To clarify the role of each element of the DT region in the L-shape formation, we recently undertook an *in vivo* selection of functional mutants of the amber suppressor $tRNA^{Ala}_{CUA}$ from two combinatorial tRNA gene libraries K and M (Figure 2(a)), in which 12 nucleotides in total in both loops D and T were randomized. [6. and 7.] The choice of $tRNA^{Ala}$ as the prototype molecule to study structure–function relationships in tRNA was determined by the fact that the aminoacylation of $tRNA^{Ala}$ depends almost exclusively on the presence of base-pair G3-U70 in the acceptor stem and is practically insensitive to the juxtaposition of the two helical domains or

even the existence of the D/anticodon domain. [8. and 9.] Therefore, any change in the tRNA performance caused by modifications at the DT region was not expected to be related to aminoacylation and thus would be attributed to the steps of the tRNA functional cycle common for most tRNAs. Analysis of the nucleotide sequences of the selected suppressor tRNA clones helped us to envisage the special role played by the reverse-Hoogsteen (RH) base-pair T54-A58 in the formation of the L-shape. This role was shown to be twofold. On one hand, the particular geometry of RH T54-A58 does not allow nucleotide 57 to stack to A58, which creates a niche between the two nucleotides, the so-called purine trap. The filling of this niche with purine 18 from the D-loop provides a stable inter-loop interaction (Figure 1).[10] On the other hand, the formation of RH T54-A58 allocates two unpaired nucleotides 59–60 for the bulge that fills the gap between the acceptor/T and D/anticodon helical domains (T-bulge). The properly arranged T-bulge keeps the required juxtaposition of the two domains. When in our experiments, the D/anticodon helical domain became shorter due to the elimination of its last stacking layer represented by the tertiary base-pair 15–48, RH in the T-loop also rearranged in a way that allocated an additional third nucleotide to the T-bulge, and the total number of stacked layers in the DT region remained the same.[6] On the basis of these findings, the properly arranged RH in the T-loop seemed to be indispensable for the normal juxtaposition of the helical domains.

There were, however, a few exceptional clones that could not form RH that would be able to allocate the proper number of nucleotides to the T-bulge and thus to guarantee the normal tRNA L-shape.[6] The fact that these tRNAs were functional meant that the fixation of the L-shape in these molecules was achieved by the use of a different strategy independent of the presence of RH in the T-loop. Here, we show that this new strategy pertains to the formation of an inter-loop double helix (ILDH) consisting of three base-pairs. Based on the selection of 51 suppressor tRNA clones from two different combinatorial libraries, we demonstrate that in the absence of RH, ILDH becomes a major factor determining the tRNA functionality. ILDH constitutes a new type of arrangement of the DT region able to keep the standard L-

shape conformation and to make the tRNA functional. The results presented here are important for understanding the general constraints imposed on the tertiary structure of functional tRNA; they can also be applicable to the regions in other RNA molecules whose structure is similar to the DT region in tRNA.

Results

Background: abnormal tRNA clones selected from the M-library

The design of the initial K-library is shown in Figure 2(a).[7] Six nucleotides were randomized in each of loop D and loop T. Compared to the standard tRNA structure, an additional eighth nucleotide was introduced into the T-loop to provide the latter with more conformational freedom and to increase the chance for selection of unusual structural forms. The presence of an additional nucleotide in the T-loop creates problems for its alignment with the standard T-loop. For this reason, for nucleotides of an eight-membered T-loop, we use numbering with asterisks from 1* to 8*. Analysis of the clones selected from this library (K-clones) showed that the most conserved nucleotides in the randomized regions were uridine in position 1* and adenosine in position 6* of the T-loop (group 1 in Table 1). The presence of these nucleotides allowed the formation of RH between them.7. and 10. Additional analysis showed that the special conformation of the adenosine within this base-pair did not allow it to stack to the nucleotide in position 5*. This provided a niche between them, a so-called purine trap, which could be filled with a purine from the D-loop (Figure 3(a)).[10] This arrangement is virtually identical with that observed in normal tRNAs, where guanosine 18 of the D-loop intercalates between purine 57 and adenosine 58 of the T-loop (Figure 1). Such intercalation allows the two contacting loops to keep the juxtaposition of the two helical domains close to the standard.

Another consequence of the formation of RH in the T-loop is that it allocates nucleotides 7* and 8* for the T-bulge. These two nucleotides are equivalent to nucleotides 59 and 60 in the normal tRNA; they fill the gap between the D-domain and the T-stem, which fixes the arrangement of the two helical domains. To check

the importance of this interaction for the tRNA structure and function, we designed another M-library6 (Figure 2(a)), which was almost identical with the K-library, with the only difference consisting of a wider gap between the D-domain and the T-stem. For this, we replaced the conservative purine G15 by U and deleted nucleotide C48, which effectively eliminated the last layer of the D-domain (Figure 3(a)).[6. and 7.] In most clones selected from the M-library, there was also a possibility to form RH in the T-loop between uridine and adenosine. However, unlike in the K-clones, where the adenosine occupied position 6*, it stayed in position 5* in the M-clones (group 2 in Table 1). The repositioning of the adenosine to position 5* allowed the formation of RH between nucleotides 1* and 5*, providing three nucleotides for the T-bulge instead of two (Figure 3(b)). The extension of the T-bulge by one nucleotide was seen as a compensation for the loss of layer 15–48 in the D-domain that allowed the maintenance of the standard L-shape.

Nine M-clones, however, could not form such RH (group 3 in Table 1). In some of these exceptional clones, the UA base-pair in the T-loop was not in the proper position, while in others such a base-pair could not be formed at all. The suppressor activity of these clones varied between 2.7% and 75%. Surprisingly, one of these clones (M12) outperformed all other tRNA suppressor mutants that we have selected so far from the two combinatorial libraries[6. and 7.] with relative activity that was close to the control suppressor $tRNA^{Ala}_{CUA}$.

The fact that these exceptional clones could not form the proper RH in the T-loop, but still had substantial suppressor activity led us to the hypothesis that these clones used an unknown alternative strategy to keep the normal juxtaposition of the two helical domains. Analysis of the nucleotide sequences of these clones provided a clue to what this alternative strategy could be. In the D and T-loops of all exceptional clones, we found regions that were complementary to each other, suggesting the existence of a double helix between them (ILDH). In all but two clones, it was possible to form three or more consecutive Watson–Crick (WC) base-pairs (group 3 in Table 1). The two exceptions pertained either to a GU base-pair in the middle of

the helix (clone M29), or to an interruption in the string of three WC base-pairs with a CC mismatch (clone M25). The exact position of the complementary regions varied within the randomized positions in the loops between the clones, which did not allow us to easily grasp the common pattern. However, for clones showing an average to high activity, a strong tendency was observed for these regions to occur in the central parts of both loops.

The H-library

To explore further the possibility of the ILDH formation and to understand the role it has in the tRNA structure and function, we designed another combinatorial library in which clones with ILDH would be selected as a rule rather than as an exception. This new library (H-library, Figure 2(b)) was based on the nucleotide sequence of clone M12, the most active M-clone, which had also a potential to form four WC inter-loop base-pairs. In the H-library, the eight nucleotides that comprised the two complementary regions in clone M12 were randomized. To prevent the formation of RH between the first and fifth nucleotides of the T-loop, the variation of the fifth nucleotide was limited to pyrimidines. Additionally, to give the region between the anticodon and T-stems more freedom in adaptation to new structures, nucleotide U48 was added to the design. If ILDH was indeed essential for the tRNA function, the new library was expected to reinforce the tendency toward its formation.

The tRNA gene harboring seven completely and one partially randomized position was purchased commercially, amplified by PCR and cloned into the pGFIB-1 plasmid, as described,[7. and 11.] producing a library with the sequence complexity of 3×10^4. The selection of active suppressor tRNA clones was based on a particular feature of the Escherichia coli XAC-1 strain having nonsense amber mutations in genes argE and lacZ.[12] The suppression of the nonsense mutation in the lacZ gene promotes the synthesis of the β–galactosidase, which provides blue colonies in the presence of X-Gal. Plasmids from the blue colonies were isolated and retransformed into XAC-1 to avoid the possibility that the suppressor phenotype was a result of a host mutation. The activities of all selected clones were measured by the traditional

colorimetric assay, which was based on the suppression of the amber mutation in the lacI/lacZ gene of the XAC-1 strain.[13]

Out of 3×10^4 screened variants, 42 positive clones with unique nucleotide sequences of the plasmid-borne tRNA gene were selected and characterized (Table 2). The activities of the selected clones in the suppression of the nonsense mutation in the lacI/lacZ gene varied between 2% and 72% and, on average, were higher than the activities of the clones selected from the two previous libraries. [6. and 7]. The presence of the second amber mutation in the argE gene of the XAC-1 cells12 allowed us to determine the efficiency of the selected suppressor tRNAs in enabling these cells to grow in the absence of arginine. Among the nine tRNA mutants tested, the three most active, H120, H122 and H128, as well as the suppressor $tRNA^{Ala}_{CUA}$ allowed the cells to reach the saturation level within 24 h (data not shown). For the less active tRNA mutants H106, H109, H121, H136 and H141, the growth to the level of saturation took between two and three days. Finally, H142, the weakest of the mutants examined, did not demonstrate any sign of growth after five days of incubation. These results correlate reasonably well with those obtained through the suppression of the nonsense mutation in the lacI/lacZ gene (Figure 4).

For the same group of nine clones, the suppressor activity was measured by a luminometric assay employing an additional plasmid pLux that carried the luxB gene from Vibrio harveyi with the amber stop codon in position 13.[14] As can be seen in Figure 4, the measured activities corroborate those obtained with the other two methods and attribute the effect of suppression to the particular features of the suppressor tRNA rather than to a measuring system.

For ten strong and weak clones, the *in vivo* level of the suppressor tRNA and of its aminoacylated and deacylated forms was determined by the acid PAGE followed by hybridization with a specific DNA probe complementary to the anticodon stem–loop.[11] Although all clones had a detectable steady level in the cytosol (Figure 5), this level varied over among the samples tested. While for more active clones, the

steady level was comparable with that of the suppressor $tRNA^{Ala}_{CUA}$, for less active clones, it was notably lower. The suppressor activities for the corresponding clones were normalized by the level of tRNA expression (Table 2). For the five strong clones, the normalized activities were similar to the overall activities, while for the five weak clones, the normalized activities were substantially higher than the overall ones. This result clearly shows that the low activity of the least active clones should be attributed primarily to their low presence in the cytosol and not to the inability per se to deliver on the tRNA primary function. Neither can it be attributed to insufficient aminoacylation, because, as can be seen in Figure 5, all selected clones existed in the cytosol predominantly in the aminoacylated form.

Analysis of the nucleotide sequences of the selected M and H clones

The nucleotide sequences of all 42 selected H clones and of those nine M clones presented in group 3 of Table 1 are aligned in Table 2. Although most H clones fit to the original library design, some sequences contain deviations from it, which deal exclusively with replacements, insertions and deletions of individual nucleotides. Most mutations occurred in the D-loop and in position 47. Mutations took place in the helical regions only in clones H150 and H20. In H150, a deletion of C51 and C64 resulted in a rearrangement of the T-stem and in the recruiting U48 for base-pairing with A65 (Figure 6(a)). In H20, mutation U12G in the D-stem provided for a secondary structure with an extended anticodon stem (Figure 6(b)). Similar secondary structures were observed in different mitochondrial tRNAs and analyzed in detail a few years ago.[15., 16., 17. and 18.]

All selected H-clones demonstrate the ability to form a double helix between two regions in the D and T-loops, although the length and the stability of this double helix can be different in different clones. In all clones, the ILDH nucleotides coming from the T-loop are found within the four randomized nucleotides 3*, 4*, 5* and 6*. The nucleotides of the D-loop forming the complementary region are correspondingly named 3#, 4#, 5# and 6#. Within the D-loop, the positions of

fragment 6#–3# can vary, although in most cases, it leaves at least two and three unpaired nucleotides on, respectively, the 3′ and 5′ side from itself.

The statistics on the occurrence of different dinucleotide combinations at each of the four positions of ILDH is shown in Table 3. Three of the four base-pairs, [3#,3*], [4#,4*] and [5#,5*] are predominantly WC. The most conserved of them are [5#,5*] and [4#,4*], in which deviations from the WC type occur only in five and six clones, respectively. Most exceptions deal with the presence of GU and UG combinations, which are known to be tolerated in WC double helices, and only three clones, M25, H158 and M4, contain mismatches CC, AA or CA. It is important, however, that all three exceptional clones have a low level of activity, which suggests that the presence of a mismatch in either [4#,4*] or [5#,5*] can seriously damage the tRNA function. Base-pair [3#,3*] is less conserved than [4#,4*] and [5#,5*]; in seven clones, it is neither WC nor GU/UG. Here also, mismatches occur in relatively less efficient clones, although their damaging effect on the tRNA function is milder than the effect of the mismatches in either [4#,4*] or [5#,5*]. All three base-pairs [3#,3*], [4#,4*] and [5#,5*] show a substantial level of nucleotide co-variation, which indicates the interdependence of the nucleotides within each base-pair and additionally supports the existence of the latter (Table 3). Also, in all three base-pairs, combinations GC/CG substantially outnumber all other combinations, thus providing for a stable ILDH.

Finally, while base-pair [4#,4*] harbours combinations RY and YR (R, purine; Y, pyrimidine) almost equally, base-pairs [3#,3*] and [5#,5*] demonstrate clear preferences toward, YR and RY combinations, respectively. Nucleotides 3# and 3* are positioned, respectively, on the 3′ and on the 5′ side from nucleotides 4# and 4*, which form the neighbouring base-pair [4#,4*]. For base-pair [3#,3*] in such a position, its predominant identity YR provides essentially better stacking with base-pair [4#,4*] than RY (compare patterns B and C in Figure 7).[19] In the same way, an RY base-pair in [5#,5*] makes its stacking with [4#,4*] essentially more effective than YR. Similarly, there are eight clones in which either base-pair [3#,3*] or [5#,5*]

is GU. In six of these clones, it is either GU for position [5#,5*] or UG for position [3#,3*]. In both cases, such a base-pair would stack to base-pair [4#,4*] much better than the alternative one in which G and U are replaced in their position (compare the patterns in Figure 7(a) and (d)). The GU base-pair has the inverted orientation in only two clones, M29 and M32, which substantially weakens its interaction with base-pair [4#,4*]. These two exceptional clones will be discussed later. In total, the observed preferences for [3#,3*] and [5#,5*] lead to a more stable arrangement of all three base-pairs within ILDH.

Unlike nucleotide combinations [3#,3*], [4#,4*] and [5#,5*], and contrary to what was expected on the basis of the nucleotide sequence of clone M12 used for the design of the H-library, nucleotides 6# and 6* do not seem to form a base-pair. The [6#,6*] combination is WC in only nine clones (always UA), and it is UG in 18 clones. The other base-pairs are AG (13 cases), AA (three cases), GA and UU (two cases each), CA and AC (one case each). Even in those clones where combination [6#,6*] is either UA or UG, this base-pair, most probably, does not exist. Indeed, base-pair UA is expected to be weak because it has only two hydrogen bonds, and because a pyrimidine at the 5′-end of a double helix stacks to the next base-pair very poorly. Such a poor stacking would severely compromise the stability of the WC base-pair in which it is involved (Figure 7(c)). For combination UG, the destabilizing effect is even stronger than for UA (Figure 7(d)). It becomes especially strong if the following nucleotide in position 5# is a purine, which is the case for almost all clones presented in Table 2. Statistical analysis shows that the number of different dinucleotide combinations is very close to what would be expected if nucleotides 6# and 6* were combined together in a random way (Table 3). In other words, the nucleotide identities of positions 6# and 6* vary independently from each other and, most probably, these nucleotides play different roles in the tRNA structure. Nucleotide 6# is most frequently U, and in a third of the cases it is A. Nucleotide 6* is most frequently G, and in a third of all cases it is A. Comparison of the nucleotide sequences of the negative and positive clones reveals a strong selective pressure

towards U and against G for position 6#, and towards G and against U for position 6* (data not shown).

Analysis of the standard tRNA tertiary structure shows that the juxtaposition of regions 18–19 and 56–58, which form the inter-loop interaction, is already close to that of two opposite strands in the A-RNA conformation, so that the replacement of this interaction by ILDH would require only minor rearrangements in other parts of the DT region. Molecular modeling of the tertiary structure for one of the selected clones (clone H1, Figure 3 and Figure 8) shows that the formation of ILDH is perfectly consistent with the standard position of both D and T-stems as well as of unpaired nucleotides U8, A14, A21, U54, U55, U59 and C60. Comparison of this model with the structure of the DT region in the yeast tRNAPhe shows that the conformation of the polynucleotide chain in both structures is very similar (Figure 8(b)).[20] The axis of ILDH is about parallel with the axis of the T-stem. As in the yeast tRNAPhe, the base of the second nucleotide of the T-loop (nucleotide 2* in our numbering and nucleotide 55 in the standard nomenclature) stacks to the phosphate of the fourth nucleotide (4* and 57) and forms a hydrogen bond with the phosphate of the fifth nucleotide (5* and 58). Strand 5#-3# is connected to adenosine residues 14 and 21 at the beginning and at the end of the D-loop by flexible connector regions. For each of these regions, one nucleotide will suffice the proper connection, which is the case for all sequences shown in Table 2.

For nucleotide 6*, one can envisage two different possibilities of accommodation within the DT region. The first possibility pertains to the formation of a base-pair between 6* and 1* analogous to 58–54 in the normal tRNAs. However, because position 6* in the selected tRNAs is occupied predominantly by guanosine, while RH 54–58 requires that nucleotide 58 be adenosine, base-pair [1*,6*] cannot form in the standard way. Another problem related to formation of the base-pair between 1* and 6* is that it would leave two nucleotides for the T-bulge. In view of the absence of the tertiary base-pair 15–48, the existence of only two nucleotides in the T-bulge can create problems for the normal juxtaposition of the helical domains. We thus

prefer the other possibility for accommodation of 6* in which this nucleotide is involved in the T-bulge (Figure 3(c)). Like in the M-clones containing RH (Figure 3(b)), the extension of the T-bulge to three nucleotides would compensate for the absence of the 15–48 tertiary base-pair, thus providing for the proper juxtaposition of the two helical domains. The fact that nucleotide 6* is a purine in almost all selected clones would provide an essential stabilizing effect for the position and the conformation of the T-bulge in the situation when the rigidity of the T-loop is weakened due to the absence of an internal base-pair analogous to RH 54–58.

Finally, we want to come back to the two exceptional clones M29 and M32 in which either [3#,3*] or [5#,5*] is a GU base-pair positioned in such a way that its stacking with [4#,4*] is severely compromised. These two clones share another feature that is not observed in any other clone shown in Table 1: the T-loop in both M29 and M32 contains the normal seven nucleotides instead of eight. A deletion of a nucleotide of the T-loop would additionally destabilize ILDH and the whole arrangement at the DT region. Therefore, it seems probable that these two exceptional clones do not contain ILDH. The normal length of the T-loop makes the standard pattern more probable for these tRNAs, although they do not completely fit to this pattern either. The elucidation of the structure of these clones goes beyond the scope of this study and will be the subject of further analysis.

Discussion

The results presented here demonstrate that tRNAs with a DT arrangement essentially different from the standard one can be active in vivo. The 51 tRNA clones presented in this study are unique in the sense that they do not have analogues among wild-type molecules. The activities of these clones are strong and, for some clones, they become close to that of the suppressor $tRNA^{Ala}_{CUA}$, which is characterized by the standard tertiary structure. These activities do not depend on the context in which the codon appears or on the measuring system, as was demonstrated by the effective suppression of the nonsense mutations in the genes of three different reporting proteins. The fact that the experiments are done *in vivo*

guarantees that all selected clones have successfully passed all steps of the tRNA functional cycle, including transcription, maturation, folding, aminoacylation, association with the elongation factor EF-Tu and passage through the ribosome. Even if the overall level of activity of a clone is low, most probably, it is due to the low steady level of the tRNA molecule in the cytosol rather than to its inability *per se* to deliver on the tRNA primary function.

Earlier, we demonstrated that in the standard tRNA structure, RH is primarily responsible for the maintenance of the proper juxtaposition of the two helical domains known as the L-shape and that such juxtaposition is required for the normal tRNA function.[6., 7. and 10.] Here, for the first time, we show that a tRNA can be functional in vivo even without this key element. The design of the library affected the formation of RH by replacing the adenosine in the fifth position of the T-loop with a pyrimidine. One may have expected that the presence of a pyrimidine in this position would provide for a similar DT arrangement with a less stable pyrimidine-pyrimidine base-pair between the same nucleotides. In reality, however, the replacement of a single nucleotide resulted in a complete rearrangement of the structure of the DT region. This result shows clearly that the integrity of RH is critical for the role it has in the tRNA structure. It also demonstrates the importance of this role for the tRNA function, so that if the formation of RH is compromised, the whole DT arrangement must be replaced by another one that does not rely on RH.

Without RH, all clones except two have a new feature, a complementarity between the two loops. In most clones, ILDH would consist of three WC base-pairs, and in the others it would have two WC base-pairs. The fact that ILDH can be formed in almost all selected clones strongly supports its existence. Indeed, because in the selection from combinatorial libraries, each clone appears independently of others, it represents an independent experiment. As such, this clone can serve as a control for the verification of patterns found in the analysis of other clones. Statistical analysis also shows that if the WC base-pairs composing ILDH were not essential for the tRNA function and appeared as a result of a random choice, the probability of

receiving so many of them within the selected clones would have been less than 10^{-30}. The fact that this number is so low proves beyond reasonable doubt the importance of these base-pairs for a tRNA clone to be selected as effective suppressor of nonsense mutations.

The existence of ILDH is supported also by the WC co-variation within each identified base-pair. The existence of this co-variation proves that the tRNA function depends not on the identity of particular nucleotides, but on the maintenance of the WC relationship between the two nucleotides within each base-pair. Because all three base-pairs composing ILDH are predominantly GC or CG, the helix should be stable. Moreover, the fact that in most clones base-pairs [3#,3*] and [5#,5*] are, respectively, YR and RY, strengthens their stacking with base-pair [4#,4*], thus stabilizing the whole ILDH even more. The predominant selection of combinations YR and RY for base-pairs [3#,3*] and [5#,5*] demonstrates that the stability of ILDH is one of the most important factors that determine the functionality of the selected clones. The level of conservation of the WC type in different base-pairs of ILDH is different. Base-pairs [4#,4*] and [5#,5*] are packed within the arrangement, which can explain their only slight tolerance of non-WC combinations. Base-pair [3#,3*], on the contrary, stays at the border of the structure and, correspondingly, shows a somewhat higher presence of non-WC combinations. Although originally, based on clone M12, we expected to find four base-pairs in ILDH, the selection clearly shows that the fourth base-pair, [6#,6*], does not exist. Nucleotides 6# and 6* vary independently of each other and seem to have different structural roles.

The modeling experiments show that introduction of ILDH into the DT region does not face any steric problem. Already in the standard tRNA structure, regions 18–19 and 56–58, which form the inter-loop contact, are arranged in an anti-parallel manner, so that the formation of a double helix between them would need only minor conformational rearrangements in the flanking regions. In the modelled tRNA structure containing ILDH, the conformation of the T-loop is very similar to the standard one (Figure 8(a)). The positions of the first three nucleotides of the T-loop

are very close to the positions of T54, Ψ55 and C56 in the normal tRNAs (Figure 8(b)). Nucleotide 2* of the T-loop still forms the same interactions within the T-loop, thus stabilizing its conformation, while nucleotide 3* also forms a WC base-pair with a nucleotide of the D-loop. Such similarity to normal tRNAs allows the selected clones to interact efficiently with different molecules, including RNase P, alanyl-tRNA synthetase and the ribosome, which explains their high activity.

The fact that ILDH appears almost every time when a selected clone does not have RH makes ILDH the most probable alternative to RH. It suggests also that the DT arrangement containing ILDH has the same role in the fixation of the standard juxtaposition of the two helical domains, as does the arrangement built with the help of RH. The ability of the ILDH to establish a strong tie between the two loops is obvious. Moreover, the modeling experiments show that the formation of such a double helix would keep the two helical domains reasonably close to their standard juxtaposition. This, however, is not sufficient: in a properly functional tRNA, the juxtaposition of the two domains should be fixed through the interaction between the last stacked layer of the D/anticodon domain and the T-loop bulge. In the standard tRNA structure, the T-loop bulge consisting of two nucleotides, 59 and 60, stacks to base-pair 15–48 of the D/anticodon domain. However, due to the absence of base-pair 15–48 from the designs of both the M and H-libraries, the functional clones require an extension of the T-bulge, which would secure the standard juxtaposition of the helical domains. In most clones of the M-library, this extension has been shown to occur due to the extension of the T-bulge by one nucleotide (Figure 3(b)).[6] Here, we suggest that the same mechanism is used for the maintenance of the L-shape in those M and H clones that contain ILDH. The modelled structure of the DT-region seen in Figure 3 and Figure 8 corresponds to an energy minimum and is thus characterized by a certain level of stability. Although the model fits to the experimental data, its verification will need additional analysis.

The necessity to allocate exactly three nucleotides to the T-loop bulge can explain why region 3*–5*, which forms base-pairs with the D-loop, occupies the same

position in all clones. In the D-loop, on the contrary, region 5#-3# is connected to A14 and A21 by two connectors of a variable length. The situation in both loops is virtually identical with that in the standard tRNA structure, in which each nucleotide of the T-loop is fixed in its position, while dinucleotide G18–G19 is connected to the beginning and to the end of the D-loop by the two connectors of variable length. In both the normal and selected suppressor tRNAs, these connectors are rich in uridine and have flexible conformations. The similarity of the general organisation of both loops in the normal and selected tRNAs shows that the major principles of the interaction between the D and T-loops as well as of the fixation of the tRNA L-shape remain the same, despite the differences at the level of particular interactions.

Because in the modelled structure of clone H1 (Figure 3(c)) each nucleotide of the T-loop plays a particular role, one can expect that a normal-length T-loop would have problems fitting to this structure. Indeed, the two exceptional clones M29 and M32 characterized by the presence of only seven nucleotides in the T-loop do not seem to form ILDH. Thus, the success of the M and H-libraries in providing for ILDH-containing tRNA species should be, to a great extent, attributed to their initial designs in which the T-loop was extended by one nucleotide. Whether the formation of ILDH is possible for a seven nucleotide T-loop will need further analysis.

Although the overall activity of each clone presented in Table 2 was sufficiently high to be detected with use of the β-galactosidase test, this activity varied by almost two orders of magnitude for different clones. Partly, this variation can be attributed to the different steady level of different tRNAs in cytosol. Still, even on the level of individual molecules, different tRNAs perform differently. One can name several potential reasons for low activities of some clones, like ineffective processing of the 5′ or 3′ termini, self-cleavage, misfolding, absence of some post-transcriptional covalent modifications, instability of the structure because of the imperfections in ILDH discussed above, inefficient aminoacylation etc. Currently, it is not possible to say which reason played a major role in each particular clone; answering this question could be a matter of further analysis. We want to stress, however, that

most probably all these reasons have a relatively minor impact on the tRNA function compared to the presence or absence of ILDH. Indeed, because we have not selected any H-clone that does not contain ILDH, without this element the tRNA has very little chance, if any, to be active.

The fact that the activity of some selected clones approached that of wild-type tRNA raises the question of why ILDH does not exist in normal tRNAs. There could be several reasons for this fact. First, thermodynamic characteristics of a tRNA molecule containing ILDH are unknown. Being effective at 37 °C, such tRNA could be less efficient at other temperatures, which would diminish the adaptability of the organism to different conditions. Also, the efficiency of the aminoacyl incorporation into the nascent peptide in response to the cognate codon–anticodon interaction is only one of important tRNA characteristics. An equally important aspect pertains to the unwillingness of tRNA to interact productively with non-cognate codons. Whether the tRNAs containing ILDH are different from the normal tRNAs in their stability, in the accuracy of the codon–anticodon interaction or in any other functionally important aspect will be a matter of further analysis.

As we mentioned earlier, arrangements closely resembling the DT region have been found in different RNA molecules.[1., 2., 3., 4. and 5.] The fact that in a functional tRNA, the standard DT arrangement formed with help of RH can be replaced by that mediated by ILDH makes the two arrangements structurally equivalent. It would be interesting to know how universal this equivalence is and whether the RH-based DT arrangements found in other molecules can also be replaced by ILDH. Answering these questions will be a matter of further analysis. In more general terms, the possibility of a replacement of one arrangement within a functional tRNA by another one suggests that synonymous arrangements can be quite common in RNA structures. Identification of such arrangements and elucidation of the conditions under which they become interchangeable is important for understanding the rules of RNA architecture and of structure–function relationships in RNA.

Experimental Procedures

Strains

The E. coli strain XAC-1 (ara Δ(lac-proB)XIII nalA rif argE(Am)/F' lacI373lacZU118(Am) proB+)12 containing an amber mutation in gene lacZ[28] was used for cloning and selection of suppressor tRNAs. For the luciferase experiments, clone MY649 [Δ(lac-pro) ara thi recA56 F-] was used.[14]

Construction of the combinatorial library and selection of suppressor tRNAs

The template oligonucleotide coding for the H-library (Figure 2(b)) GGGGCTATAGCTCATNNNNTAGAGCGCCTGCTTCTAACGCAGGAGGTTTGCGGTTNNYN TCCCGCATAGCTCCACCA was synthesized at Bio-Corp 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 EcoRI and PstI, as described.[6] This plasmid provides a constitutive high-level expression of a cloned tRNA gene. VentR DNA polymerase, restriction enzymes and the bacteriophage T4 DNA ligase were from New England Biolabs. The ligation mixture (ca10 ng of DNA) was electroporated into the electrocompetent XAC-1 cells. After 1 h incubation at 37 °C in a shaker, the transformation mixture was spread on LB-agar plates containing ampicillin (100 mg/ml) and 5-bromo-4-chloro-3-indolyl β-d-galactopyranoside (X-Gal), which provided about 50,000 colonies. XAC-1 carries an amber nonsense mutation in the β-galactosidase gene and forms white colonies on the X-Gal plates. The effective suppression of the nonsense mutation makes the colonies blue, which allows a rapid detection of suppressor tRNAs. The clones containing effective suppressor tRNAs were selected after overnight growth at 37 °C. The plasmid DNA of the selected clones was extracted and retransformed into XAC-1 cells to confirm the dependence of the phenotype on the presence of the plasmid encoding the tRNA gene.

Sequencing

DNA was sequenced with the help of primers
5′-GCTTCTTTGAGCGAACGATCAAAAATAAGT-3′ and

5′-GGGTTTTCCCAGTCACGACGTTGTAAAACG-3′
used, respectively, for the forward and reverse reading of the tRNA gene. For the sequencing, the BigDye Terminator Cycle Sequencing Kit v.3.1 (Applied Biosystems) was used. Electrophoresis of each product was performed on an ABI 3730 Genetic Analyser at IRIC, Université de Montréal.

Measurements of the overall suppressor activity by the β-galactosidase assay
The overall suppressor activity of selected tRNAs was evaluated by measuring the β-galactosidase activity in XAC-1 cells carrying suppressor tRNA genes, which resulted from suppression of the amber stop codon in the lacZ gene. The β-galactosidase activity was determined as described,13 with overnight cultures grown to A600 of 0.8–0.9 in minimal A medium containing 0.4% (w/v) glucose, 1 mM MgSO4, 100 μg/ml of ampicillin and 80 μg/ml of arginine. The activity values were obtained by averaging the measurements from three independent cultures and calculated as a percentage of the activity of the control suppressor $tRNA^{Ala}_{CUA}$ that had the standard tertiary structure and differed from the wild-type E. coli tRNAAla by the anticodon changed from UGC to CUA. The background β-gal activity in the XAC-1 cells transformed with the pGFIB plasmid without suppressor tRNA gene was below 0.01%.

Measurements of the suppressor activity by luciferase assay
Plasmids pGFIB-1 and pLUX carrying, respectively, the genes of a suppressor tRNA and of the luxB gene from Vibrio harveyi with the amber stop codon in position [13,14] were electroporated into MY649 cells. A 30 μl sample of the saturated overnight culture was diluted in 1 ml of the LB medium supplemented by ampicillin (50 μg/ml) and/or chloramphenicol (31 μg/ml) and were grown at 37 °C to A600 between 0.2 and 0.7. n-Decyl aldehyde emulsion was freshly prepared by the 1:1000 dilution in water followed by ultrasonication. 20 μl of the n-decyl aldehyde substrate was added to 200 μl of the medium, and the mixture was quickly vortex mixed. Light emission at room temperature was immediately determined in a Lumat LB 9507 Single Tube Luminometer (BERTHOLD TECHNOLOGIES GmbH and Co KG,

Bad Wildbad, Germany). The absorbance density of the culture was assessed immediately after the light measurement.

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 at pH 5.2, as described.[6] Then, half of each RNA sample was deacylated and run on acid PAGE in parallel with the untreated half of the sample. RNA samples were transferred from the gel to a Hybond-N nylon membrane (Amersham) by electroblotting, and the membrane was hybridized with two different radiolabelled DNA probes. One probe was complementary to region 26–44 of the suppressor tRNAs consisting of the anticodon stem–loop, while the other one was complementary to region 34–53 of the E. coli 5 S rRNA. The 5 S rRNA probe was used to monitor the amount of total RNA in each sample. The quantification of the bands on the gel was done with Quantity One (Bio Rad) software program. The normalized activities of the clones were calculated with use of formula:

NA=OA[Q-tRNA−control/Q-tRNA−mutant]/[Q-5S−control/Q-5S−mutant]

where NA and OA stand, respectively, for the normalized and overall activity of the mutant tRNA, while Q-values are the quantified amounts of the control suppressor $tRNA^{Ala}_{CUA}$, of the mutant suppressor tRNA, and of 5 S rRNA in the corresponding lanes.

Computer modeling

Preliminary modeling was done interactively, using the InsightII/Discover package (version 2000, Accelrys Inc., San Diego, CA). The X-ray structure of the yeast $tRNA^{Phe}$ was used as a starting conformation, to which the elements different from the standard tRNA structure were appended.[20] The model was submitted to energy minimization using the AMBER force field† until the energy minimum was reached. During the minimization, all secondary structure elements as well as nucleotides

U8-A14-A21 in the D-domain preserved their positions as in the yeast $tRNA^{Phe}$. Visualizations were done on a Silicon Graphics Fuel computer.

Acknowledgements

We thank Dr Michael Yarus for providing clone MY649 and the pLUX plasmids. This work was supported by a grant from the National Science and Engineering Research Council of Canada. S.V.S. acknowledges fellowships from the Canadian Institutes of Health Research and from Fond de la Recherche en Santé du Québec. T.M.I. acknowledges the Strategic Training fellowship in Bioinformatics from Canadian Institutes of Health Research.

References

1. A.S. Krasilnikov, X. Yang, T. Pan, A. Mondragon
Crystal structure of the specificity domain of ribonuclease P
Nature, 421 (2003), pp. 760–764

2. A.S. Krasilnikov, A. Mondragon
On the occurrence of the T-loop RNA folding motif in large RNA molecules
RNA, 9 (2003), pp. 640–643

3. L. Levinger, R. Bourne, S. Kolla, E. Cylin, K. Russell, X. Wang, A. Mohan
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 (1998), pp. 1015–1025

4. J.C. Lee, J.J. Cannone, R.R. Gutell
The lonepair triloop: a new motif in RNA structure
J. Mol. Biol., 325 (2003), pp. 65–83

5. U. Nagaswamy, G.E. Fox
Frequent occurrence of the T-loop RNA folding motif in ribosomal RNAs

Rna, 8 (2002), pp. 1112–1119

6. E.I. Zagryadskaya, N. Kotlova, S.V. Steinberg
Key elements in maintenance of the tRNA L-shape
J. Mol. Biol., 340 (2004), pp. 435–444

7. E.I. Zagryadskaya, F.R. Doyon, S.V. Steinberg
Importance of the reverse Hoogsteen base-pair 54–58 for tRNA function
Nucl. Acids Res., 31 (2003), pp. 3946–3953

8. W.H. McClain, K. Foss
Changing the identity of a tRNA by introducing a G-U wobble pair near the 3′ acceptor end
Science, 240 (1988), pp. 793–796

9. Y.M. Hou, P. Schimmel
A simple structural feature is a major determinant of the identity of a transfer RNA
Nature, 333 (1988), pp. 140–145

10. F.R. Doyon, E.I. Zagryadskaya, J. Chen, S.V. Steinberg
Specific and non-specific purine trap in the T-loop of normal and suppressor tRNAs
J. Mol. Biol., 343 (2004), pp. 55–69

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

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

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

14. D.W. Schultz, M. Yarus
A simple and sensitive in vivo luciferase assay for tRNA-mediated nonsense suppression
J. Bacteriol., 172 (1990), pp. 595–602

15. S. Steinberg, R. Cedergren
Structural compensation in atypical mitochondrial tRNAs
Nature Struct. Biol., 1 (1994), pp. 507–510

16. K. Wakita, Y. Watanabe, T. Yokogawa, Y. Kumazawa, S. Nakamura, T. Ueda et al.
Higher-order structure of bovine mitochondrial tRNA(Phe) lacking the 'conserved' GG and T psi CG sequences as inferred by enzymatic and chemical probing
Nucl. Acids Res., 22 (1994), pp. 347–353

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

18. D.V. Lavrov, W.M. Brown, J.L. Boore
A novel type of RNA editing occurs in the mitochondrial tRNAs of the centipede Lithobius forficatus
Proc. Natl Acad. Sci. USA, 97 (2000), pp. 13738–13742

19. D.H. Turner, N. Sugimoto
RNA structure prediction
Annu. Rev. Biophys. Biophys. Chem., 17 (1988), pp. 167–192

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

21. U. Varshney, C.P. Lee, U.L. RajBhandary
Direct analysis of aminoacylation levels of tRNAs in vivo. Application to studying recognition of Escherichia coli initiator tRNA mutants by glutaminyl-tRNA synthetase
J. Biol. Chem., 266 (1991), pp. 24712–24718

1 Present address: E.I. Zagryadskaya, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

† http://amber.ch.ic.ac.uk/amber/

Clone	D-loop	T-loop	Overall activities (%)
Group 0. The standard tRNA structure			
wt tRNA^{Ala}	AG CUGG GA	UU CGA UC	100
Group 1. Examples of K-clones[7]			
K25	AG GAACGC UA	UG AAA AC	17.7±6.2
K15	AG GCAUAU UA	UG AAA UC	11.0±1.4
K2	AG AAAGAC UA	UGACGA UC	7.9±1.7
K23	AG UAAGGU UA	UGCCAA UC	5.9±1.1
K42	AG UAGACAUA	UGACAA UC	3.8±0.9
Group 2. Examples of M-clones with the properly positioned RH[6]			
M2	AU GUGGUU UA	UU CGAUUC	18.0±4.0
M8	AU AUGGCA UA	UU CGAGUA	10.7±0.4
M11	AU AUUGG UA	UU CGACUA	6.1±0.7
M1	AU AGAGGA UA	UU CGAUUC	4.8±0.5
M22	AUCAUUGG UA	UU CGAUUC	4.2±0.8
M6	AU CAGGAU UA	UU CGAGUC	3.2±0.1
M10	AA UACGGCA UA	UU CAAUUC	2.5±0.2
Group 3. M-clones in which RH cannot be properly arranged[6]			
M12	AU UACCU UA	UUGGUA UC	75.0±0.5
M5	AU UCCUUA UA	CUAGGU UC	18.0±0.3
M30	AU UACCA UA	UUGGUG UC	15.0±0.3
M32	AU CUGGCU UA	UUAGA CU	7.9±0.4
M29	AU UUGUAA UA	UUCGA UU	6.4±0.3
M27	AU GGGCGA UA	CUGCCA UC	5.4±0.4
M23	AU UUUCC UA	UUGGAA UC	5.2±0.4
M25	AU UCCUAG UA	UGACGA UC	3.4±0.3
M4	AA UACGUAU UA	UUCGCG UC	2.7±0.2

Table 1

The activity of each clone is presented as the percentage of the activity of the control suppressor tRNAAla. Group 0: the nucleotides of the T-loop forming the reverse-Hoogsteen base-pair (RH) U54–A58 are yellow. Grey is purine 57, which together with A58 forms the purine trap. Cyan and red are base-pairs G18–U55 and G19–C56, respectively. Blue is G15, which together with C48 forms the last stacked layer of the D/anticodon stacked domain. The nucleotides forming the T-loop bulge are green. In groups 0, 1 and 2, the nucleotides involved in equivalent interactions are coloured in the same way. The same colours are also used for the equivalent nucleotides in Figure 3(a) and (b). Group 1: In all clones of this group, RH is arranged between the first and the sixth positions of the T-loop.[10] RH arranged in this way allocates two nucleotides for the T-loop bulge, which is required for the normal L-shape conformation. Group 2: In the clones of this group, base-pair 15–48 is destroyed, which necessitates the T-loop bulges to have three nucleotides instead of the normal two. Correspondingly, RH is formed between the first and the fifth positions of the T-loop. The inter-loop base-pair AG (cyan) is equivalent to base-pair G18–Ψ55 in the normal tRNAs.[10] Group 3: RH cannot be arranged properly in any of these

clones. On the other hand, all clones of this group are characterized by the existence of regions in the two loops that could form ILDH containing at least three WC or GU base-pairs. Yellow, the nucleotides in both loops that can form WC base-pairs. The nucleotides that can form a GU or UG base-pair are shown in cyan. Clones M27, M29 and M32 are as described.[6. and 7.] Other clones of group 3 are presented here for the first time.

Clone	D-loop 6#5#4#3#				T-loop 3*4*5*6*		nt47	Overall activity (%)	Normalized activity (%)
H122	AU	UGCC	UUA	UU	GGCG	UC		80±0.6	
M12	AU	UACC	UUA	UU	GGUA	UC		75±0.5	67±3
H120	AA	UAUC	AUUA	UU	GGUA	UC		72±0.4	93±4
H128	AAU	AGGU	UA	UU	ACCG	UC		64±0.5	58±2
H36	AU	AGCC	UUA	UU	GGCG	UC		50±0.5	
H1	AU	UGCC	UUA	UU	GGCG	UC		50±0.3	51±3
H121	AU	GGUC	UUA	UU	GACA	UC		40±0.5	
H111	AAU	UGCU	UUA	UU	GGCA	UC		36±0.4	
H141	AG	UGCC	UUA	UU	GGCG	UC		30±0.4	
H109	AU	AGGU	UUA	UU	GCCG	UC		23±0.3	
H106	AU	UGCU	UUA	UU	AGCG	UC		23±0.5	
H136	AU	AGCC	UUA	UU	GGCA	UC		20±0.6	
H129	AU	NGCC	UUA	UU	GGCA	UC		18±0.5	18±2
M5	AU	UCCU	UAUA	CU	AGGU	UC		18±0.3	
H116	AU	UGGC	UUA	UU	GCCG	UC		17±0.4	
H110	AAU	AGCA	UUA	UU	UGCA	UC		16±0.5	
M30	AU	UACC	AUA	UU	GGUG	UC		15±0.3	
H112	AU	UGGG	UUA	UU	CCCG	UC		15±0.4	
H102	AU	AGGC	UUA	UU	GCCG	UC		14±0.5	
H160	AAU	UGGC	UUA	UU	GCUG	UC		13±0.6	
H100	AU	UGGG	GGA	UU	AGCG	UC		13±0.4	
H154	AU	AGGU	UUA	UU	ACCG	UC		11±0.4	
H21	AU	AGGG	UUA	UU	GCCG	UC		11±0.3	
H20	UAG	CGGA	UGG	UU	GGCG	UC		10±0.5	
H4	AU	UGUC	UUA	UU	GACA	UC		9.0±0.4	
H107	AU	AGGG	UUA	UU	ACCG	UC		8.7±0.5	
M32	A	UCUGGC	UUA	U–	UAGA	CU		7.9±0.4	21±2
H131	AU	AGGC	UUA	UU	GCCG	UC		7.2±0.4	
H40	AU	CGCC	UUA	UU	GGCA	UC		7.0±0.3	
H146	AAU	AGUU	UUA	UU	AACG	UC		6.7±0.3	
M29	AU	UUGU	AAUA	U–	UCGA	UU		6.4±0.3	19±2
H157	AU	AGGG	UUA	UU	CCCG	UC		6.0±0.4	56±3
M27	AU	GGGC	GAUA	CU	GCCA	UC		5.4±0.4	
M23	AUU	UUCC	UUA	UU	GGAA	UC		5.2±0.4	
H6	AU	UACC	UUA	UU	GGUG	UC		5.0±0.3	
H148	AA	UAUU	CUUA	UU	AGUG	UC		5.0±0.2	
M25	AU	UCGU	AGUA	UG	ACGA	UC		3.4±0.3	
H150	AU	AGGG	UUA	UU	CCCA	U–		3.1±0.2	
H145	AU	UGUC	UUA	UU	GGCG	UC		3.0±0.3	
H35	AU	UGGC	UUA	UU	GCUG	UC		3.0±0.3	18±2
H44	AUU	AGCA	UUA	UU	UGCG	UC		3.0±0.2	
H152	AA	UACC	UUUA	UU	GGUU	UC		3.0±0.3	
H158	A	UGAU	GUA	UU	AACG	UC		2.9±0.2	
H142	AUG	AGGU	UA	UU	ACCG	UC		2.7±0.3	42±2
M4	AA	UACGU	AUUA	UU	CGCG	UC		2.7±0.2	
H46	AUA	UGGU	UA	UU	GCCG	UC		2.5±0.3	
H124	AAUU	GGUA	UUA	UU	AACG	UC		2.3±0.2	
H137	AU	AGGG	UUA	UU	ACCG	UC		2.0±0.2	
H105	AU	UGCC	UUA	UU	GGUA	UC		2.0±0.1	
H153	AU	UAUG	UUA	UU	CGUG	UC		2.0±0.2	30±3
H48	AU	UGUU	UUA	UU	AACG	UC		1.0±0.2	6±1

Table 2

Column nt47 indicates the clones in which nucleotide 47 is deleted. The overall and normalized suppressor activities were calculated as the percentage of the activity of the suppressor $tRNA^{Ala}_{CUA}$. Nucleotides forming WC, GU or other combinations in [3#,3*], [4#,4*] and [5#,5*] are, respectively, yellow, cyan and red. Purines in position 6* are green. N in position 6# of clone H129 stands for a non-identified nucleotide. The normalized suppressor activity of clones M29 and M32 is as described.[6]

	[3#, 3*]	[4#, 4*]	[5#, 5*]	[6#, 6*]
GC	5(0.9)	18(6.7)	35(26.8)	0(0)
CG	23(13.2)	21(10.8)	3(0.2)	1(1.3)
AU	3(0.3)	0(0)	7(1.6)	0(0.6)
UA	9(3.8)	6(1.4)	1(0)	9(9.3)
GU	1(0.7)	0(0)	3(7.5)	0(0.1)
UG	3(7.7)	4(4.9)	1(0.2)	18(18.8)
AC	0(0.4)	0(0.40)	1(5.0)	0(0)
CA	1(6.6)	0(3.0)	0(0.1)	1(0.6)
GG	1(4.9)	0(8.8)	0(3.0)	1(2.0)
GA	2(2.5)	0(2.5)	0(0.7)	2(1.0)
AG	1(2.2)	0(0.5)	0(0.6)	13(10.6)
AA	1(1.1)	1(0.1)	0(0.2)	3(5.1)
UU	1(1.1)	0(0)	0(0.4)	2(1.2)
UC	0(1.4)	0(3.7)	0(1.0)	0(0)
CU	0(1.9)	0(0)	0(0.6)	0(0.1)
CC	0(2.4)	1(8.2)	0(2.1)	0(0)
Unclear cases	0	0	0	1

Table 3

The statistic analysis was done for all of the clones included in Table 1. The values in parentheses represent the expected numbers of occurrences of each particular dinucleotide combination under the assumption that the nucleotides in the corresponding positions are combined at random. The calculations were made with the help of the formula E=K(L/N), where E is the expected value for combination αβ, K is the number of clones in which the first nucleotide has identity α, L is the number of clones in which the second nucleotide has identity β, and N is the total number of clones. The real numbers that are larger than the expected values by at least five or three points are coloured red and yellow respectively. The dinucleotide identities related to these numbers correspond, respectively, to a strong or mild positive selective pressure. The real numbers that are smaller than the expected values by at least five or three points are coloured blue and cyan, respectively. The dinucleotide identities related to these numbers correspond, respectively, to a strong or mild negative selective pressure. For base-pairs [3#,3*], [4#,4*] and [5#,5*], a strong preference is observed towards WC base-pairs and against all other base-pairs. More precisely, for base-pairs [3#,3*] and [5#,5*] a clear preference is observed toward YR and RY combinations, respectively, while base-pair [4#,4*] has a preference toward GC/CG. For combination [6#,6*], the expected numbers are close to the real ones, which has been taken as an indication that positions 6# and 6* vary independently of each other.

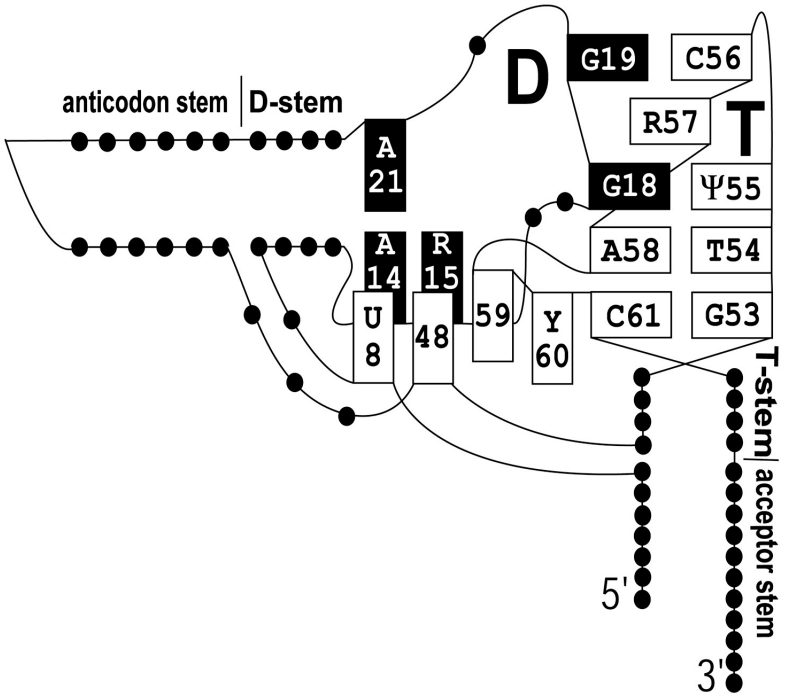


Figure 1. A conventional representation of the DT region in the standard tRNA structure. Each nucleotide of the DT region involved in stacking or base-pairing with another nucleotide is represented by a rectangle. The rectangles representing nucleotides of the D-loop are black, while those representing nucleotides of the T stem–loop as well as nucleotides 8 and 48 are white. All other nucleotides are shown as black dots. The identities of the conservative and semi-conservative nucleotides are indicated. R, purine; Y, pyrimidine; T, 5-methyluridine; and Ψ pseudouridine. A central role in the DT arrangement is played by the reverse-Hoogsteen base-pair (RH) T54–A58 (see the text).

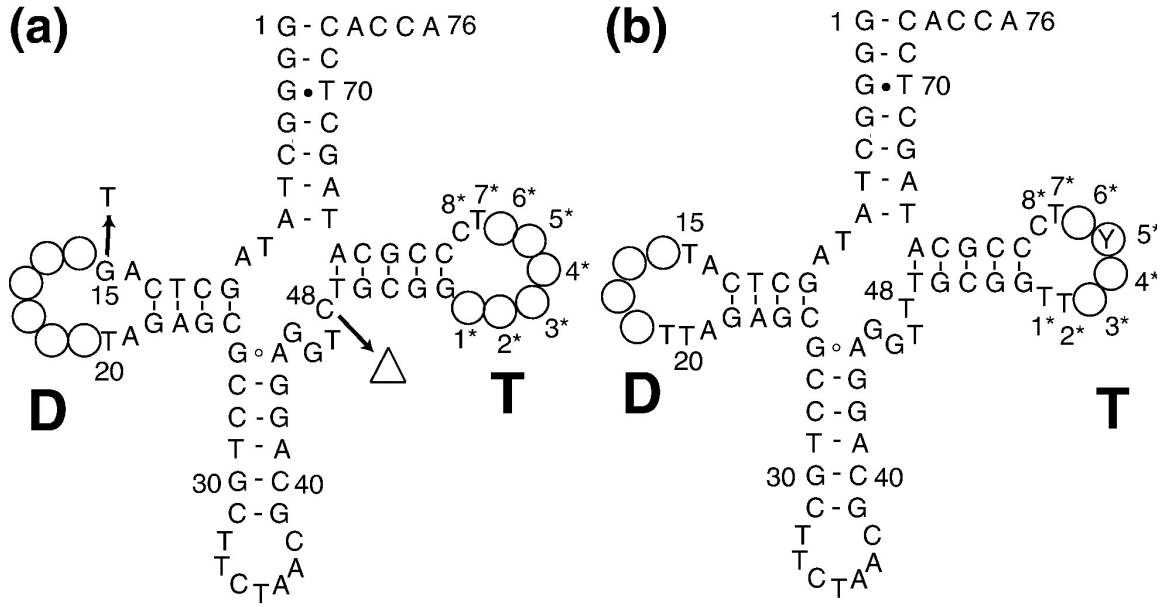


Figure 2. Designs of the three combinatorial tRNA gene libraries K, M and H. (a) The design of the K-library was based on the sequence of $tRNA^{Ala}_{CUA}$.7 The alanine anticodon was replaced by the amber anticodon CTA. The D-loop was extended by two nucleotides and the T-loop was extended by one nucleotide. The randomized positions are indicated by circles. The two additional mutations G15T and ΔC48 introduced into the M-library are indicated by arrows.6 The nucleotides of the T-loop have their own numbering with an asterisk (*). (b) The design of the H-library. The variation of nucleotide 5* of the T-loop was limited to pyrimidines (Y), and nucleotide U48 was added to the design (see the text).

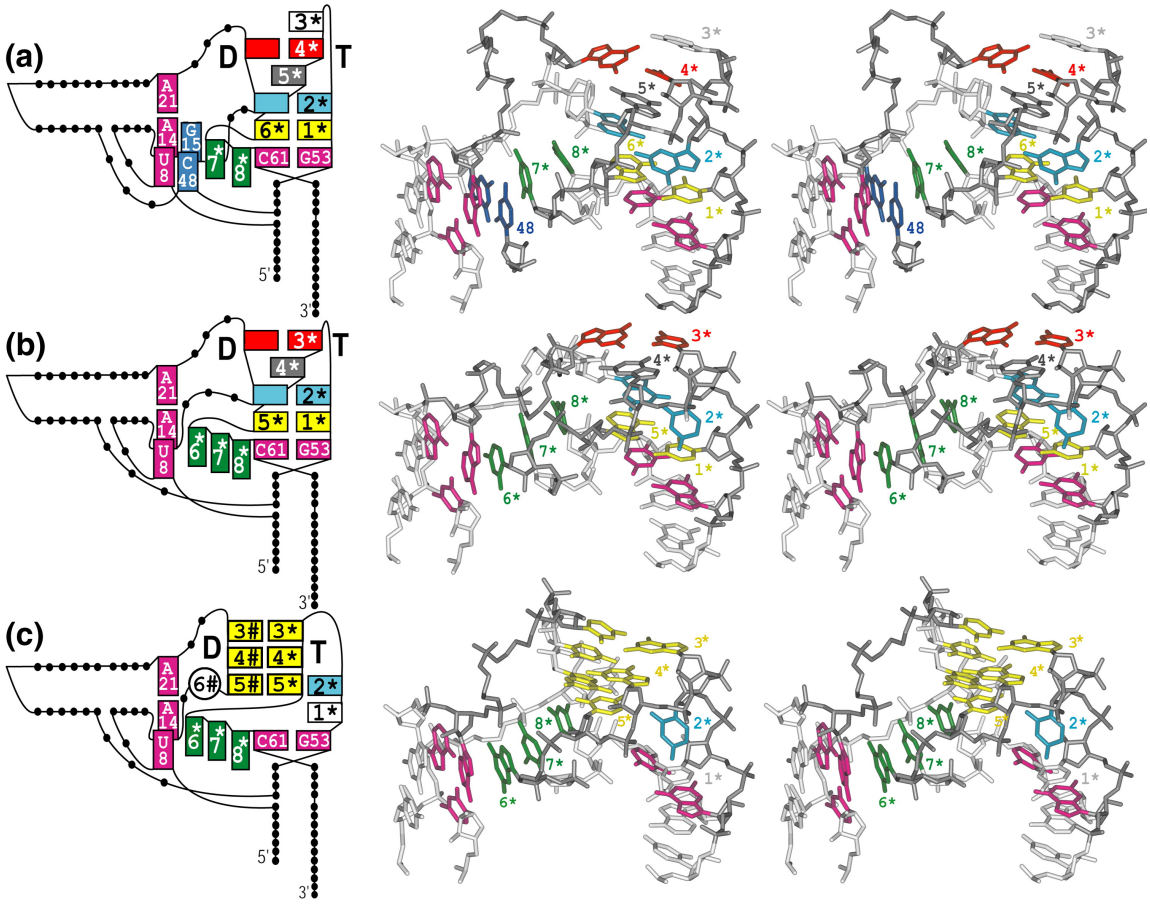


Figure 3. Comparison of the DT region structure in the tRNAs selected from the K, M and H-libraries. On the left: a representation of the DT region in the context of the whole tRNA L-shape. On the right: a stereo view of the DT region in the corresponding three-dimensional model. The nucleotides that play an important role in the maintenance of the standard L-shape conformation are represented by rectangles on the left and are shown explicitly on the right. The nucleotides outside the DT region carry the standard numbers. Nucleotides of the T-loop are numbered from 1* to 8*. For the same nucleotide, the same colour is used in both the right-hand and left-hand panels. (a) The model of clone K23 from group 1 of Table 1.[10] In this tRNA, RH (yellow) is formed between nucleotides 1* and 6* of the T-loop, thus allocating two nucleotides for the T-loop bulge. Base-pair 15–48 (dark blue) and the two nucleotides forming the T-loop bulge (green) are positioned exactly as in the crystal structure of the yeast $tRNA^{Phe}$.[20] The purine trap in the T-loop is arranged between nucleotides 5* (grey) and 6* (yellow). It harbours an adenosine (cyan) from the D-loop, which also forms base-pair AG with the guanosine in position 2*

(cyan). This AG base-pair is equivalent to base-pair G18-Ψ55 in the normal tRNAs.[10] At the top of the arrangement, two nucleotides from both loops form a base-pair (red) equivalent to base-pair G19–C56 in the standard tRNA structure. Except for nucleotide 3* at the top of the T-loop, clone K23 is considered as an analogue of the standard tRNA. (b) The model of clone M2 from group 2 of Table 1.[6] In this tRNA, tertiary base-pair 15–48 is absent, and RH is formed between nucleotides 1* and 5* of the T-loop (yellow). The purine trap is formed by nucleotides 4* (grey) and 5* (yellow). Nucleotide 6* of the T-loop joins the T-loop bulge and fits to the place normally occupied by base-pair 15–48. Except for nucleotide 6* at the top of the T-bulge, clone M2 is considered as an analogue of the standard tRNA. (c) The model of clone H1 from Table 2. Like M2 (b), this clone does not contain tertiary base-pair 15–48. However, unlike M2, it does not have a properly arranged RH, either. Instead, clone H1 contains an inter-loop double helix (ILDH) consisting of three WC base-pairs (yellow). Like RH in clone M2, ILDH allocates three unpaired nucleotides 6*–8* for the T-loop bulge, which occupy the same place as in clone M2. In clone H1, the general fold of the polynucleotide chain within the T-loop is very similar to that in RH-containing structures, including the standard tRNA (see also Figure 8(b)).

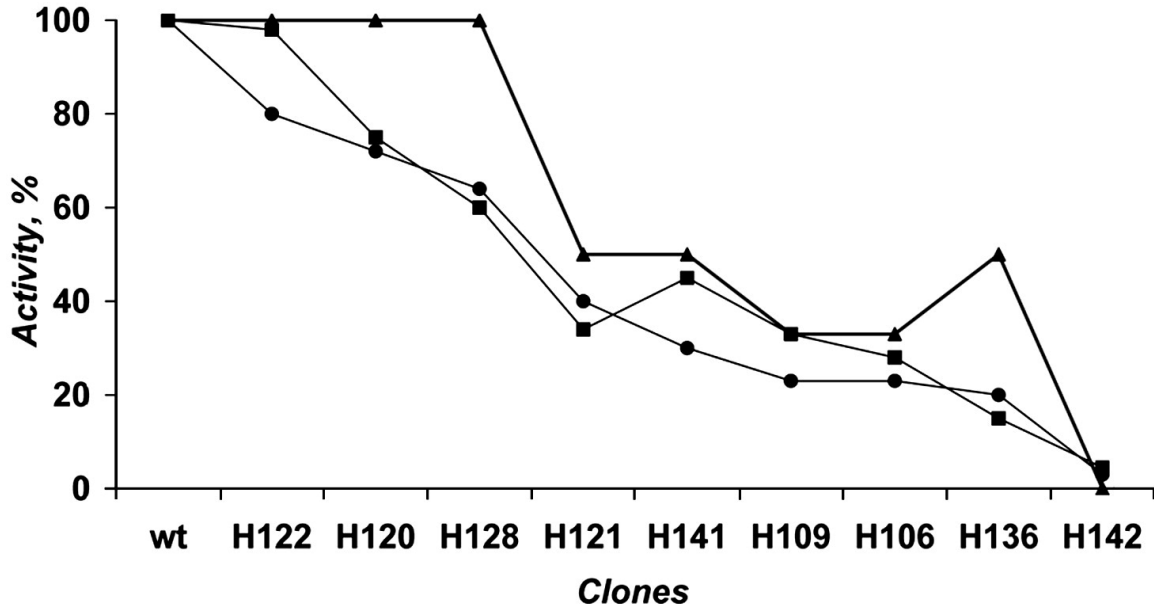


Figure 4. Comparison of different methods for measurement the activities of H-clones. The activities of all selected clones were measured by the β-gal assay (circles, see also Table 2).[13] In addition, for nine clones shown in the Figure, as well as for the suppressor tRNAAla, the activity was measured with use of the lux assay[14] (squares) and by the ability of the cells to grow in the minimal medium without Arg (triangles). In all cases, the activity of the wild-type suppressor $tRNA^{Ala}_{CUA}$ was taken as 100%. For the cell ability to grow in the minimal medium without Arg, the activity was considered to be reversely proportional to the time needed for the cells to reach the saturation level (data not shown). The Figure shows a general consensus between all three types of measurement, which allows us to conclude that the measured activities reflect the properties of the selected tRNAs rather than of the measuring systems or the contexts in which the amber stop codons appear.

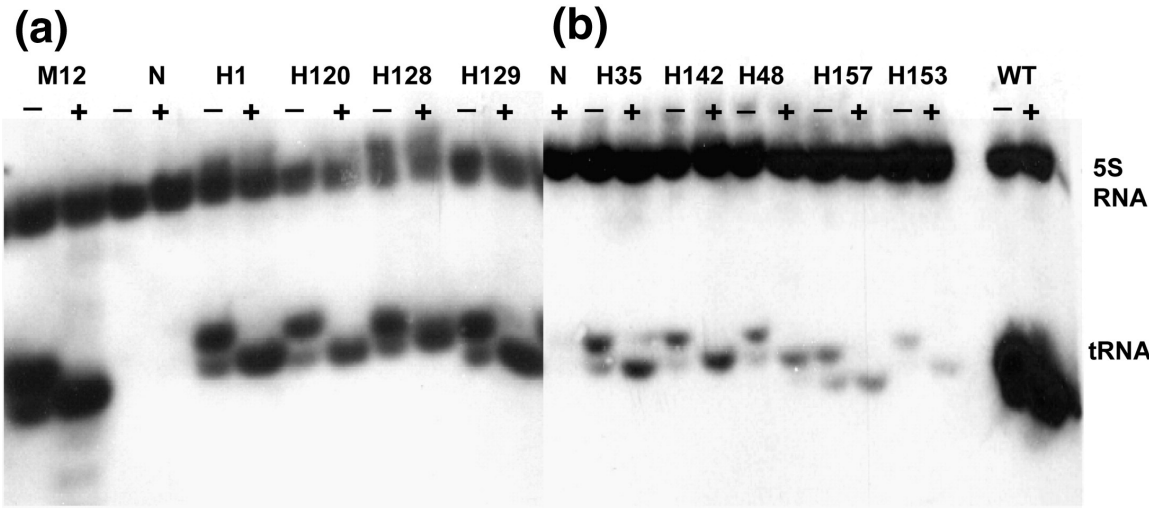


Figure 5. The steady and aminoacylation levels of clone M12, of some (a) strong and (b) weak H-clones, and of the suppressor $tRNA^{Ala}_{CUA}$. Total cellular RNA was fractionated by acid PAGE,21 transferred to a membrane, and the Northern blot hybridization was performed using a specific DNA probe complementary to the anticodon stem–loop of the cloned tRNAs.[6] 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. N refers to negative control, representing the pGFIB plasmid lacking a tRNA gene. Additional probe specific to 5 S rRNA was used to monitor the amount of total RNA in the samples (the upper band in both (a) and (b)).

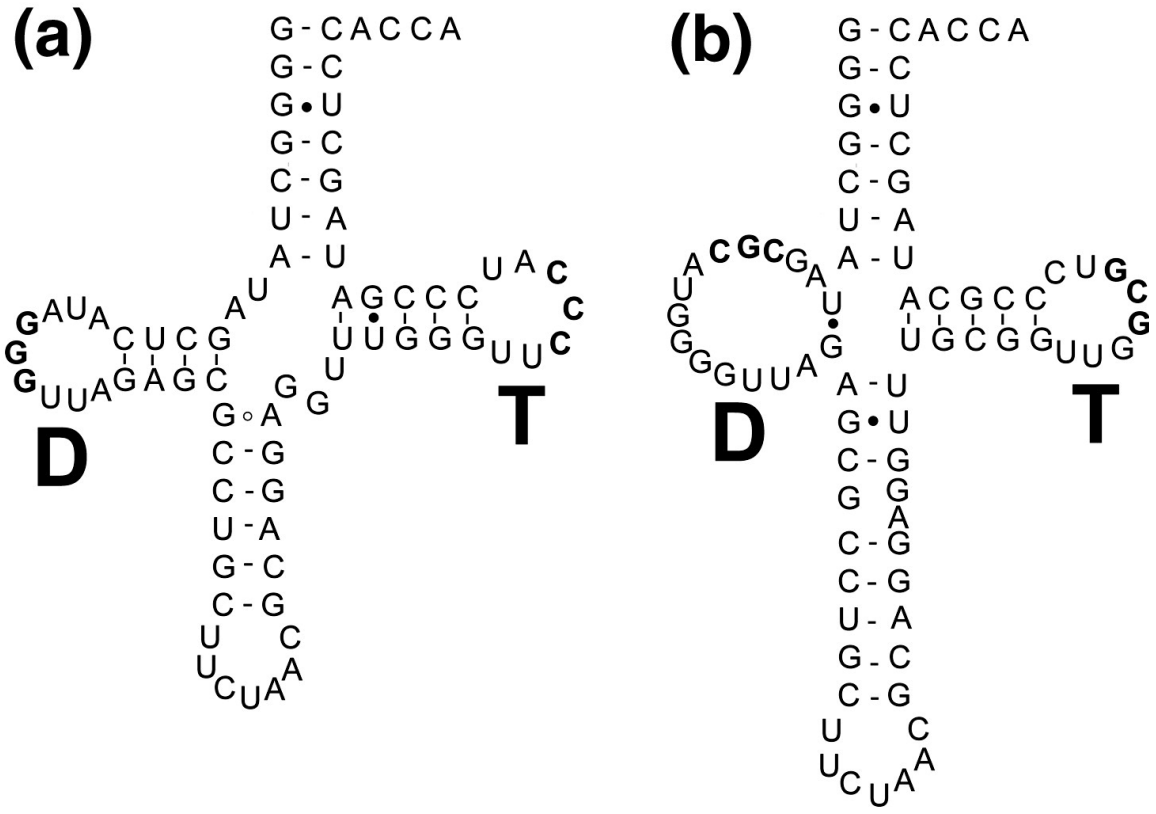


Figure 6. The cloverleaf secondary structures of clones (a) H150 and (b) H20, in which rearrangements occur within the D or T-stems. In H150, a deletion of C51 and C64 resulted in a rearrangement of the T-stem and in the recruitment of U48 for base-pairing with A65. In H20, mutation U12G in the D-stem provided for a secondary structure with an extended anticodon stem containing nine base-pairs. The bold letters indicate the nucleotides involved in ILDH.

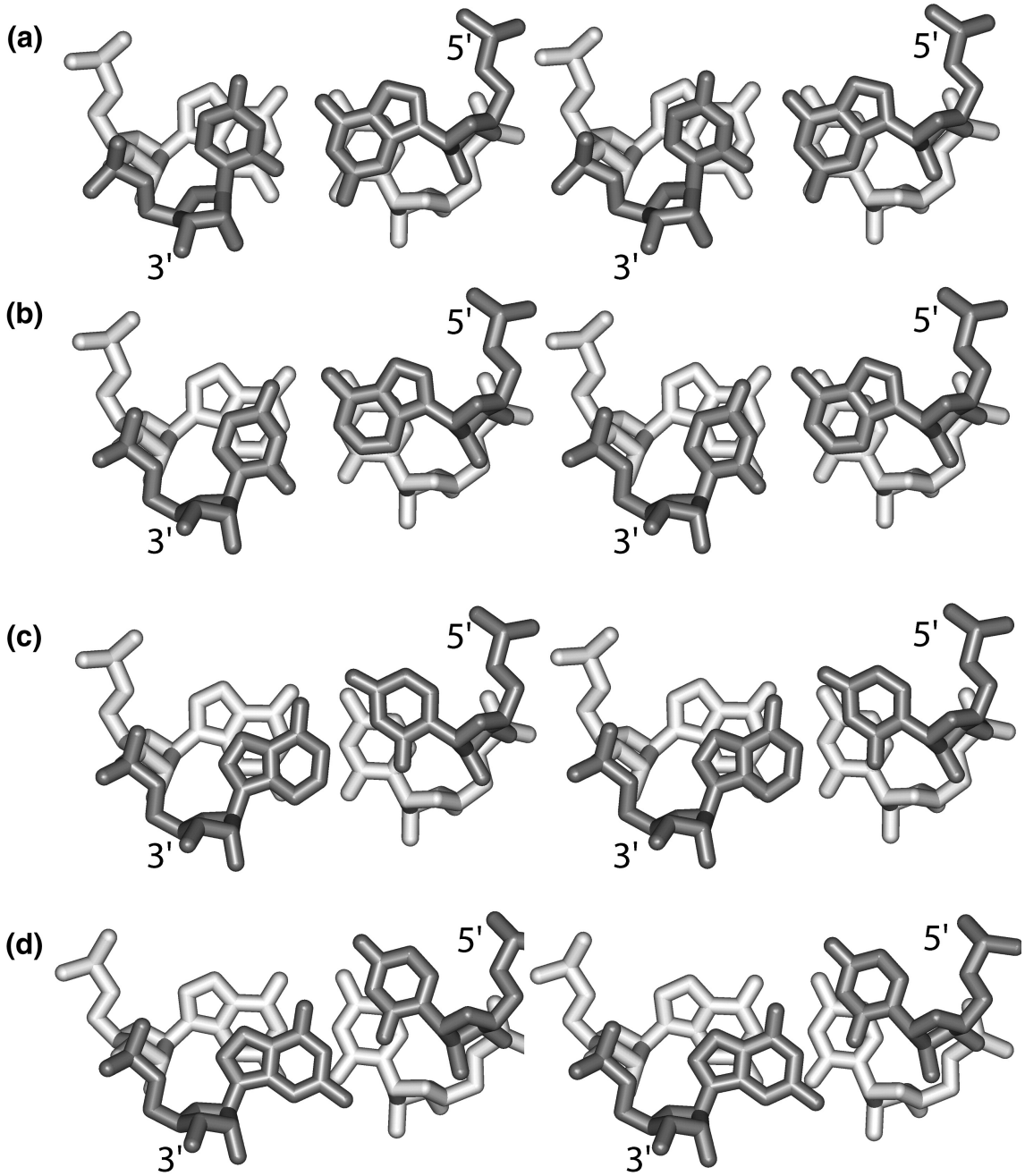


Figure 7. The stacking patterns of the last base-pair of a double helix (black) to the second last base-pair (white). (a), (b), (c) and (d) correspond to the last base-pair 5′-GU-3′, 5′-AU-3′, 5′-UA-3′ and 5′-UG-3′. From pattern (a) to (d), the stacking between the two base-pairs gradually weakens. Patterns (a) and (b) correspond to a very strong and strong stacking. Pattern (c) is characterized by a marginal stability. Pattern (d) is unstable.

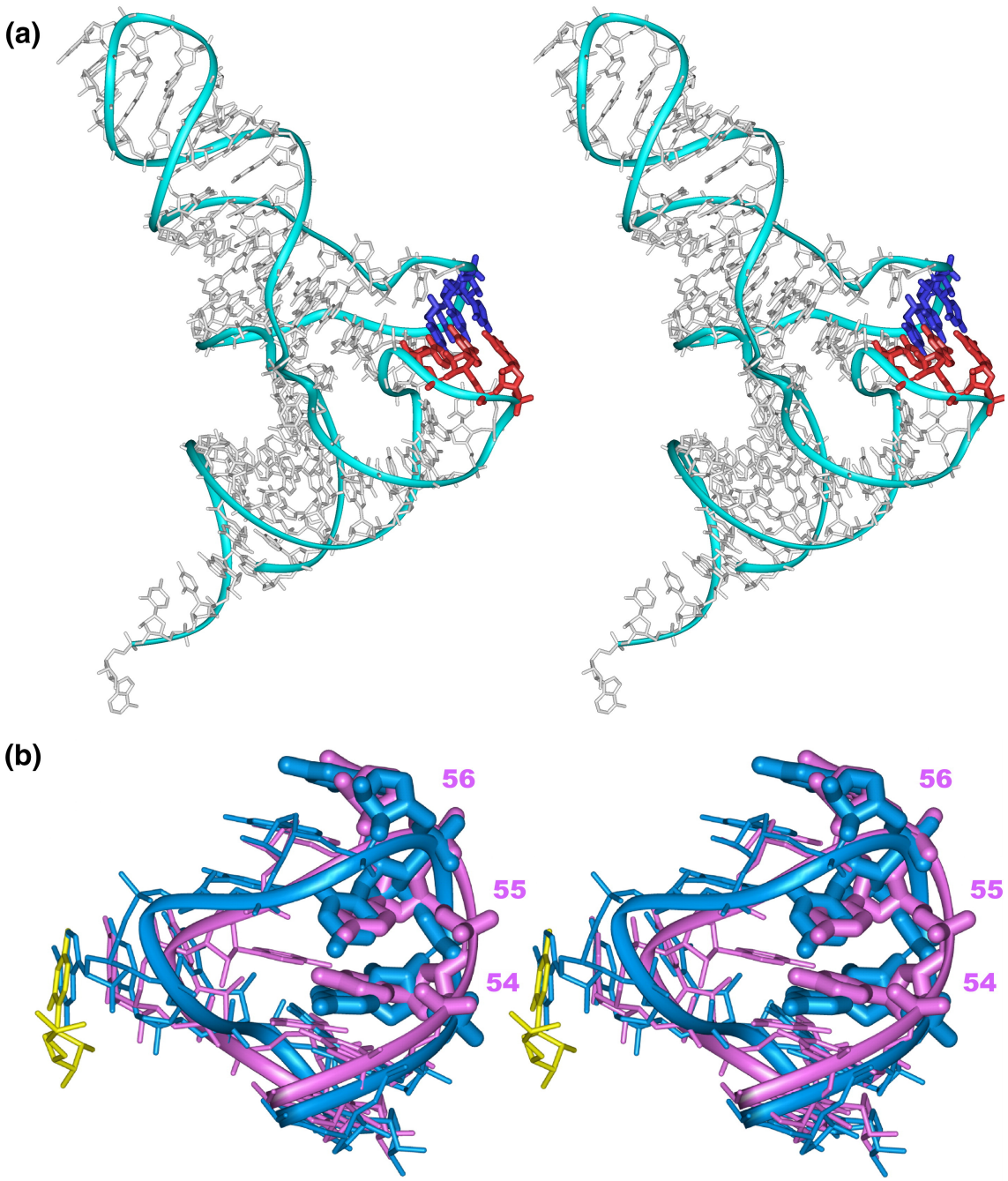


Figure 8.

The modelled tertiary structure for clone H1. (a) The general view of the L-shape. Nucleotides 3*–5* and 5#–3#, which form ILDH, are red and blue, respectively. (b) The superposition of the T-loop in the yeast tRNAPhe (magenta)20 and in the model of clone H1 (blue). Nucleotides 54–56 of the yeast tRNAPhe are numbered. The first three nucleotides in both structures (nucleotides 54–56 in the yeast tRNAPhe and nucleotides 1*–3* in H1) are thick. Their positions and the contacts in which they

are involved in both structures are very close. The major difference between the two structures occurs in the T-loop bulge, mostly due to the presence of the additional nucleotide 6* in the model of clone H1. This nucleotide takes the position normally occupied by C48 (yellow), which does not exist in clone H1.

CHAPTER III

A new strategy for fixation of the L-shape in functional tRNAs

Tetsu M. Ishii, Natalia Kotlova, Ekaterina I. Zagryadskaya* , & Sergey V. Steinberg†

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

*Present address: Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139 †To whom the correspondence should be addressed: Tel: 1 (514) 343-6320; FAX: 1 (514) 343-2210

Tetsu M. Ishii designed and performed the in silico experiments, modeled the structures, and wrote the paper.

Natalia Kotlova and Ekaterina I. Zagryadskaya performed the wet experiments.

Sergey V. Steinberg designed the combinatorial libraries, performed analysis and wrote the paper.

The paper is not yet published.

Keywords: RNA structure; T-loop; molecular modeling; combinatorial library

Running Title: D-T interaction and tRNA L-shape

ABSTRACT

To elucidate the general constraints imposed on the structure of the D- and T-loops in functional tRNAs, active suppressor tRNAs were obtained through *in vivo* screening of a combinatorial tRNA gene library in which several nucleotide positions in both loops were randomized. Analysis of nucleotide sequences of the 53 isolated clones demonstrated that in most of them, the T-loop contained such standard elements as the reverse-Hoogsteen base pair UA and the purine trap. However, 26 sequences did not have these features, but instead, were characterized by the presence of four unusual elements: the extended T-stem composed of six base pairs, the shortened T-bulge having only one nucleotide, an unusual reverse-Hoogsteen-reverse-Watson-Crick base pair within the T-loop, and a double helix between the D- and T-loops composed of two or three Watson-Crick base pairs. Molecular modeling and molecular dynamic simulations of the tertiary structure of these clones showed that the four unusual elements are involved in complex cause-effect relationships with each other, providing together for the proper and stable arrangement of the acceptor/T and D/anticodon helical domains known as the tRNA L-shape.

INTRODUCTION

tRNA plays a central role in the protein synthesis. It also represents a useful object to study the rules governing the formation of RNA tertiary structure. On one hand, tRNA is a relatively small molecule, which makes genetic manipulations relatively easy. On the other hand, it is large enough to possess a distinct tertiary structure, the so-called L-shape, which consists in the specific perpendicular arrangement of the acceptor/T and D/anticodon helical domains. The correctly arranged L-shape guarantees the proper juxtaposition of the two functional centers, the anticodon and the acceptor terminus, and is thus critical for the tRNA function (Söll and Rajbhandary 1995; Pan et al. 2006). Analysis of the tRNA tertiary structure shows that the major interactions that lead to the formation of the L-shape take place in the so-called DT region located at the corner of the structure, where the two loops D and T form a specific arrangement (Figure 1). Because similar arrangements are also found in other RNAs, like the RNase P and the ribosomal RNA (Levinger 1998; Nagaswamy and Fox 2002; Krasilnikov and Mondragon 2003; Krasilnikov et al. 2003; Lee et al. 2003;), the elucidation of the rules governing the formation of the DT region in tRNA can be important for understanding structure-function relationships in these molecules as well.

To understand the role of different elements of the DT region in the formation of the tRNA L-shape, we previously undertook a search of active suppressor tRNAs from several combinatorial gene libraries in which vast areas in both loops D and T were randomized (Zagryadskaya et al. 2003; Zagryadskaya et al. 2004; Doyon et al. 2004; Kotlova et al. 2007). Analysis of the obtained clones allowed us to formulate the general requirements imposed on the structure of the DT region in functional tRNAs. The central role in the structuring of this region was shown to be played by the reverse-Hoogsteen base pair (RH) UA formed by the first (U) and the third last (A) nucleotides of the T-loop (base pair U54-A58 *trans* W.C./Hoogsteen in the standard tRNA structure). This RH has two major functions. First, the particular geometry of RH-UA prohibits stacking between the adenosine of this base pair (A58) and its 5'-neighbor (purine-57) (Figure 1) (Doyon et al. 2004). The inability of these two nucleotides to stack to each other allows them to form a niche at the top of the T-loop, the so-called purine trap, that can harbor a purine

from the D-loop (G18 in the standard tRNA structure). The filling of the purine trap in the T-loop by a purine from the D-loop would bring the two loops close to each other in the fluctuating tRNA secondary structure (Doyon et al. 2004). The second role of RH consists in the allocation of two unpaired nucleotides (normally, nucleotides 59-60) for the T-bulge that fills the gap between the acceptor/T and D/anticodon helical domains (Figure 1), thus fixing their proper juxtaposition (Shi and Moore 2000).

The design of the first combinatorial tRNA gene library (K-library) that initiated our analysis of the DT region is shown in Figure 2. This library was based on the *E.coli* $tRNA^{Ala}$, in which the anticodon was modified to recognize the amber stop codon. In each of the two loops, six nucleotide positions were fully randomized, while an additional nucleotide was added to the normally seven-nucleotide T-loop. The initial expression of the K-library in the XAC-1 *E. coli* cells containing the amber mutation in the β-galactosidase gene yielded 28 active suppressor tRNAs (K-clones) (Zagryadskaya et al. 2003). Further exploration of the same combinatorial library increased the total number of K-clones to 53. Due to the presence of the additional nucleotide, the nucleotide sequences of the T-loop in these clones cannot be aligned with that in the standard tRNA structure. For identification of the T-loop nucleotides, we thus use numbers with asterisks, varying between 1* and 8*. Among the 53 obtained K-clones, 27 contained the RH U1*-A6* and the purine trap (Doyon et al. 2004). These clones fell into two distinct groups, with the nucleotide sequence of the T-loop fitting to formula [UGNN($^{G}/_{A}$)AUC] (Type I) or [UA($^{G}/_{A}$)($^{A}/_{C}$)CAUC] (Type II). Analysis of these two tRNA groups allowed us to model the tertiary structure for both of them (Figure 3). The two groups corresponded to the so-called specific purine trap (Type I) and non-specific purine trap (Type II). Both groups have been previously analyzed in detail (Doyon et al. 2004) and are not the subject of the current paper.

As mentioned earlier, among the 53 obtained K-clones only 27 fitted to either Type I or Type II pattern, while the other 26 K-clones were unable to do so. The tertiary structure of the latter tRNAs thus remained obscure. Here we present the results of the nucleotide sequence analysis and the molecular modeling of tertiary structures for those K-clones that fit to neither Type I nor Type II pattern. Our analysis shows that these clones use an alternative strategy for the fixation of the L-shape, which is based on the

presence of several unique elements. In particular, all analyzed clones contain such unprecedented features as a six base-pair T-stem, one nucleotide in the T-bulge instead of the normal two, and either a reverse-Watson-Crick pyrimidine-purine or reverse-Hoogsteen purine-purine base pair in the T-loop. In addition, these clones are characterized by a complementarity between the D- and T-loops that would allow the formation of an inter-loop double helix (ILDH). Inserted individually into a tRNA structure, any of these features would severely compromise the integrity of the L-shape. We demonstrate, however, that when all these abnormal elements coexist in the same molecule, they provide for the normal juxtaposition of the two helical domains. This aspect allows the tRNA to become functionally active in spite of its unusual structure. In more general terms, the results of our analysis demonstrate that in the RNA world, a particular function can be achieved by different structural means, which thus can be considered as functional synonyms.

RESULTS

K-clones containing G1*

At the bottom of Table 1 one can see the nucleotide sequences of those 26 K-clones that fit to neither Type I nor Type II pattern. The 13 clones marked by asterisks have been collected as a result of additional exploration of the K-library and are shown here for the first time. One can notice that with exception of K30, in all K-clones that fit to neither Type I nor Type II, nucleotide 1* is guanosine. All clones containing G1* form the Type III group. Because in all previously analyzed K-clones, nucleotides 1* and 6* were able to form RH-UA, we initially suggested that a similar base pair could be formed even if U1* was replaced by G1*(Zagryadskaya et al. 2003). In such cases, G1* would mimic uridine in its interaction with nucleotide 6*, thus forming a base-pair structurally equivalent to RH U1*-A6*. Further analysis, however, showed that the guanosine identity of nucleotide 1* was not the only unusual element of these clones. Assuming G1* formed a base pair with 6* similar to RH, the other nucleotides of the T-loop still would not be able to fit either to the Type I or to the Type II pattern. The inability of Type III clones to form either Type I or Type II purine traps suggested that they had an

essentially different T-loop structure and, correspondingly, a different strategy for the fixation of the L-shape.

Although, in principle, this alternative strategy could also include the formation of a 1*-6* base pair, analysis of the Type III clones provided a strong argument against it. In particular, we found that the nucleotide identity of 6* co-varied with the identity of nucleotide 2*. As one can see in Table 1, with the only exception of clone K10, which contains combination U2*-A6*, the identity of nucleotide 6* in all Type III clones relates to the identity of nucleotide 2* according to a simple pattern: C or A in 2* corresponded to A in 6*, while U or G in 2* corresponded to G in 6*. Such relationship between nucleotides 2* and 6* suggests that they form a base pair in which certain structural elements are preserved in spite of the variation of the identities of both nucleotides. An attempt to form all four 2*-6* combinations, CA, AA, UG and GG as RH base pairs in order to mimic the standard RH base pair UA (Figure 4, left column) failed because it did not allow the formation of base pair UG. Additional difficulty for this arrangement relates to the fact that the purine-purine combinations AA and GG built as RH base pairs would have a larger size than the pyrimidine-purine combination CA. Alternatively, all four dinucleotide combinations could be arranged as reverse-WC base pairs (RWC, *trans* W.C.), using the WC edges of both nucleotides with the trans-orientation of the glycosidic bonds (Figure 4, right column). Although all RWC base pairs allow the formation of two hydrogen bonds, in base pairs AA and GG formed in this way, the distance between the glycosidic bonds is notably longer than in base pairs CA and UG, which makes these arrangements also unlikely.

Comparison of the RH and RWC base pairs shown in Figure 4 revealed a new unexpected possibility for the 2*-6* base pairing. We found that in two RH base pairs, AA and GG, and in two RWC base pairs, CA and UG, (boxed in Figure 4) the juxtaposition of the glycosidic bonds is almost the same, which can also be seen when all four base pairs are superimposed (Figure 4, bottom). Additional analysis shows that such an arrangement is the only one that makes all four base pairs practically isosteric and that no other dinucleotide combination can be arranged in a way to be isosteric to these base pairs. Because two of these base pairs form as RH, while the other two form as RWC, the combination of all four base pairs arranged in this way will henceforth be called RH-

RWC. Thus, the formation of the 2*-6* base pair according to the RH-RWC scheme can explain both the presence of the four combinations CA, AA, UG and GG and the virtual absence of clones with any other dinucleotide combinations. The only exceptional clone K10 containing combination UA will be discussed later. Compared to the standard RH base pair, in the RH-RWC base pairs the glycosidic bonds are positioned in about the same orientation, although farther from each other by about 3Å. There is one more aspect related to the formation of the RH-RWC base pairs. Comparison of all four base pairs shows that in base pairs CA and UG nucleotide 6* has a different orientation compared to that in base pairs AA and GG. The transition between the two orientations consists in the rotation of the base around the glycosidic bond for 180º, so that if one orientation represents a more common *anti*-conformation, the other one will corresponds to a relatively rare *syn*-conformation. Although the *syn*-conformation is energetically less favourable than the *anti*-conformation, it is known to be much more acceptable for purines than for pyrimidines. Because in the Type III clones, position 6* is occupied exclusively by adenosines and guanosines, the *syn*-conformation would occur only in purines. This aspect further corroborates the suggestion of the RH-RWC arrangement for base pair 2*-6*.

Further analysis revealed another common feature of all Type III clones, the existence of a Watson-Crick (WC) correspondence between the D- and T-loops. As one can see in Table 1, the three-nucleotide region 3*-5* of the T-loop can always form two or three consecutive WC base pairs with a region in the D-loop. Nucleotide 4* would form a WC base pair in all clones, and for nucleotide 5*, it would happen in all clones except K4 and K31. However, if combination GA is considered acceptable, the two exceptional clones K4 and K31 will follow the same pattern with other clones. For position 3* the number of exceptions is higher, which, most probably, indicates that the formation of this base pair is not critically important for the tRNA function. Still, if combinations GU and GA are acceptable, nucleotide 3* would form a base pair in most clones. Interestingly, while in the nucleotide sequence of the T-loop, the position of the strand of ILDH is fixed within nucleotides 3*-5*, the opposite strand in the D-loop does not have a fixed position and can in different Type III clones be closer to or farther from the beginning of the loop.

The formation of base pair 2*-6* raises a question concerning the role of nucleotide 1*. Although this nucleotide could stay unpaired, the fact that its identity is restrained to guanosine suggests that it is also involved in a specific interaction for which its guanosine identity is important. Because the last nucleotide of the T-loop is unpaired and also because in almost all Type III clones it is cytidine, the most natural way to employ guanosine G1* would be to form a WC base pair G1*-C8*. The formation of such a base pair would extend the T-stem to six base pairs, thus leaving only one unpaired nucleotide 7* for the T-bulge.

Modeling the tertiary structure for Type III K-clones

The analysis of the nucleotide sequences of the K-clones presented in the previous part suggests the existence of four unusual elements, the particular RH-RWC base pair between nucleotides 2* and 6*, a double helix containing two or three base pairs between nucleotides 3*-5* of the T-loop and a region in the D-loop, the extended T-stem containing six base pairs, and finally, the reduced T-bulge containing only one nucleotide 7*. The scheme of the DT region that incorporates all these elements is shown in Figure 5 (left). Because these elements were suggested based on the properties of individual nucleotides without taking into account the structural context of the DT region, it was not clear, whether it was possible to integrate all of them into the same tRNA tertiary structure. Also, even if such integration is possible, it is unclear whether the resulting structure would provide for the proper juxtaposition of the two helical domains. Finally, it is important to know why the DT region of a functional tRNA would need to harbour simultaneously so many unusual elements.

For answering these questions, we modeled the tertiary structure for the Type III K-clones. The original construct was made interactively based on the available crystal structure of yeast $tRNA^{Phe}$ (Shi and Moore 2000) and on the nucleotide sequence of clone K32, which demonstrated the highest activity (Table 1). The additional base pair G1*-C8* was appended to the last base pair G53-C61 of the T-stem. The RH-RWC U2*-G6* base-pair was arranged on top of G1*-C8* similarly to the arrangement of RH T54-A58 with respect to base pair G53-C61 in canonical tRNAs. U7* was arranged in the available space between base pairs G1*-C8* and G15-C48 and was connected to G6*

and C8*. A double helix consisting of three WC base pairs was positioned on top of base pair U2*-G6*, and the connectors integrating both strands of this helix into the D- and T-loop were arranged accordingly to the nucleotide sequence of clone K32. The energy minimisation resulted in the model shown in Figure 5, right.

In this model, due to the natural twist of the double helix for about 32° between the fifth and additional sixth base pair of the T-stem, nucleotide 8* is located 3.2 Å closer to the D-domain compared to the position of the analogous nucleotide 61 in the standard tRNA structure. As a result, nucleotide 7* stays at virtually the same position as nucleotide 59 in the standard structure, and stacks to the tertiary base pair G15-C48.

The RH-RWC base pair 2*-6* stacks on base-pair C1*-G8*, to which it is directly connected on one side and indirectly, with mediation of the bulged nucleotide 7*, on the other side. The particular structure of the RH-RWC, which is similar to the standard RH base pair AU except that it is 3Å longer, allows this base pair to fit effortlessly to the given context. An essential element of the model pertains to the fact that here, nucleotide 5* comfortably stacks to nucleotide 6* (Compare Figures 3 and 5). As we showed previously, such interaction would not have been possible if base pair 2*-6* formed as RH-UA (Doyon et al. 2004). However, the fact that the structure of RH-RWC is different from RH-UA eliminates obstacles toward this stacking. Finally, in the model, nucleotides 3*-5* of the T-loop are involved in the WC base pairing with three nucleotides of the D-loop, forming together the ILDH. Among the three inter-loop base pairs, the one that involves nucleotide 3* is positioned at the top of the structure. Therefore, an alternative arrangement of the nucleotides within this base pair, or even the absence of base pairing would not seem to affect other interactions in the molecule. This can explain the fact that the inter-loop base pair including nucleotide 3* is less conserved than the other two base pairs of ILDH.

Additional modifications of the model

Additional modeling experiments were performed in order to check the adaptability of the proposed model to particular features of individual nucleotide sequences.

1. In five clones K14, K43, K44, K50 and K61, the top base pair of ILDH contains an AG combination. Our analysis shows that the AG base pair in this position can be formed in the head-to-head way (A-G cis W.C./W.C.) without affecting the positions of neighbouring nucleotides (Figure A in the Supplementary data).

2. The absence of nucleotide 3*, which happens in clones K6 and K37, does not interfere with the proper connection between nucleotides 2* and 4* and does not affect other elements of the tRNA structure (Figure B in the Supplementary data).

3. While the base pair formed by nucleotide 4* is always WC, the next base pair involving nucleotide 5* is GA in two clones K4 and K31. Our modeling experiments show that the formation of such a base pair as RH optimized its stacking interactions with its neighbouring flanking base-pairs (Figure C in the Supplementary data).

4. While in most clones, the 1*-8* base pair is GC, in clone K37 it has the GU identity. Our modeling experiments show that the replacement of C8* by U has a relatively minor local effect and does not compromise the integrity of the whole structure (Figure D in the Supplementary data).

5. As mentioned above, in clone K10 the 2*-6* base pair has a unique nucleotide combination UA. When this base pair is formed as RWC, its structure will be relatively close to, although still different from the structures of base pairs CA and UG. The modeling experiments show that although acceptable, such a base pair would stack to the previous base pair G1*-C8* less effectively as compared to base pairs CA or UG and thus should be considered as a compromising variant (Figure E in the Supplementary data).

The relation between RH and RH-RWC base pairs

Although the structure of the suggested base pair RH-RWC is similar to that of RH-UA existing in the normal tRNAs, our results clearly show that among Type III clones, RH-UA was persistently avoided in favour of RH-RWC. To understand the reasons of this phenomenon, we modeled the 3D structure of clone K10 with base pair U2*-A6* formed as RH. Our modeling, however, showed that such conformation cannot be structurally sound because of the interference between the base of A6* and the atom of hydrogen H1′ of C8*. In Figure 6A, corresponding to the RH-UA case, atom H1′ of

C8* can be seen as a sort of pimple protruding in the direction of nucleotide A6*. The specific position of this atom does not allow the normal stacking interaction between the bases of A6* and C8*. Interestingly, in a similar situation that takes place in the standard tRNA structure, when RH U54-A58 stacks on WC G53-C61, such interference is not observed (Figure 6B). The difference between the two cases is linked to the length of the T-bulge, which in the Type III clones and in the standard tRNA structure contains one and two nucleotides, respectively. The longer bulge observed in the standard tRNA structure would allow the base of A58 to be additionally displaced from the ribose of C61 for about 5Å and to avoid the unfavourable contact with its atom H1′. It is also interesting that while the formation of RH-UA in the Type III clone K10 interferes with the stacking between A6* and C8*, for the alternatively formed base pair RWC-UA such interference is essentially less pronounced (Figure 6C). Also, when RH U2*-A6* is replaced by either RWC C2*-A6* (Figure 6D) or by RH G2*-G6* (Figure 6E), the hydrogen atom H1′ of C8* would no longer interfere with the stacking of purine 6* and C8*. We thus can conclude that the presence of only one nucleotide in the T-bulge would favour the formation of the RH-RWC base pairs and disfavour the formation of RH-AU. This aspect provides an additional support for our model of the T-loop structure in the Type-III clones, in which the RH-RWC base pair 2*-6* plays the central role.

Another missed opportunity pertains to combination C2*-G6*, which has never happened among K-clones. Although C and G, like all other pyrimidine-purine combinations, can form a RWC base pair, the latter is not isosteric to the RH-RWC (Figure 4). The different structure of this base pair compared to the RWC base pairs U2*-G6* and C2*-A6* will make nucleotide G6* strongly interfering with the hydrogen atom H1′ of C8* (Figure 6F). This interference will be notably stronger than in the discussed above case of the RWC U2*-A6*. This aspect can explain the absence of combination C2*-G6* among K-clones.

Molecular Dynamics of the Type III structure

To evaluate the stability of the DT region in the Type III tRNAs, the models of the DT region containing different variants of the 2*-6* base pair were subjected to the 1-nanosecond MD simulations as described in the Materials and Methods (Figure 7 and

Figure F in the Supplementary data). The integrity of the DT region was monitored by following the maintenance of base pairs G53-C61, G1*-C8* and 2*-6*. The part of the whole simulation time when all three base-pairs were simultaneously maintained was taken as a measure of the stability of the given structure. At the 2*-6* position, we tested the RH base pairs UA, AA and GG, as well as the RWC base pairs UG, CA, UA and CG. As Figure 7 shows, the arrangements containing base pairs RH-GG, RWC-UG, RWC-CA and RH-AA maintained their integrity for 100%, 85%, 70% and 60% of the simulation time, respectively. The arrangement containing RWC-UA base pair kept its integrity for only 20% of the time, while the remaining two arrangements containing RH-UA and RWC-CG demonstrated a complete inability to have all three base pairs at the same time.

These results allow us to establish a clear correlation between the overall stability of the arrangement and the quality of the stacking between base pairs G1*-C8* and 2*-6*. Indeed, the four arrangements in which the stacking of these base pairs did not face any problem turned out to be most stable. When the interference between purine 6* and atom H1′ of C8* in the arrangement containing RWC-UA made the stacking of base pairs G1*-C8* and 2*-6* partly compromised, the whole arrangement became substantially less stable. Finally, the two most unstable arrangements were exactly those in which the strong interference between purine 6* and atom H1′ of C8* made the normal stacking of base pairs G1*-C8* and 2*-6* impossible. The MD simulations thus confirmed our initial suggestion that the interference between the base of purine 6* and atom H1′ of C8* is the major factor of instability of the analyzed arrangements.

DISCUSSION

In this study, we demonstrate that tRNA molecules with a structure of the DT region essentially different from the standard can function *in vivo*. Although in general, the activity of the collected clones is lower than that of the tRNA with the normal tertiary structure, it is high enough to be reliably detected with the β-gal assay. The nucleotide sequences of the collected suppressor tRNAs demonstrate a range of diversity never seen in the natural cytosolic tRNAs. In spite of this, the obtained tRNA clones constitute only a small fraction of the combinatorial possibilities provided by the tRNA gene library,

which infers the existence of strong constraints imposed on the structure of functional tRNAs.

The comparative analysis of the nucleotide sequences of all found tRNAs has been very helpful in elucidating the nature of these constraints. First, it has shown that almost all clones fall into three distinct groups, each having its own nucleotide sequence signature and a particular tertiary structure. While the structure of the DT region in the clones of the first two groups is based on the presence of such standard elements as the RH base pair and the two-nucleotide T-bulge, Type III clones contain none of these elements. The fact that the Type III clones are functional leads us to the important conclusion that the tRNA functionality does not directly depend on the presence of particular nucleotides or particular interactions within the DT region, like the RH-UA base pair and/or the two-nucleotide T-bulge, even if they exist in all cytosolic tRNAs. Instead, the functionality of tRNA seems to be determined by a more general aspect of its structure that goes beyond particular elements or interactions. Analysis of the three major groups of collected K-clones demonstrated that, despite the notable differences in the structure of the DT region, all of them share one important feature: in all groups, the obtained nucleotide sequences of the D- and T-loops guarantee the standard juxtaposition of the two helical domains, acceptor/T and D/anticodon. Thus, the proper juxtaposition of these domains seems to be a major factor determining the ability of the tRNA to be processed from larger precursor RNAs, to maintain intracellular stability and to deliver on its function.

While in the Type I and II clones, the juxtaposition of the two domains is fixed with use of standard elements, namely, the properly arranged RH, the purine trap and the two-nucleotide T-bulge, in the Type III clones, this fixation is achieved in a different way. The characteristic elements of this arrangement are the RH-RWC base pair, the ILDH, the T-bulge consisting of only one nucleotide and the T-stem containing six base pairs. The modeling has shown that, despite such an unusual structure of the DT region in the Type III clones, the general strategy of the L-shape fixation is always the same. In all tRNAs, this strategy consists in two steps, the formation of an interaction between the D- and T-loops due to their affinity to each other, and the final fixation of the L-shape through the filling of the gap between the D-domain and the T-stem with the properly

arranged T-bulge. While in normal tRNAs and in both Type I and II groups, the interaction of the D- and T-loops is achieved through the intercalation of a purine from the D-loop between nucleotides 57 (or 5*) and 58 (or 6*) of the T-loop, in the Type III clones, it proceeds through the formation of the ILDH. Also, in normal tRNAs as well as in both Type I and II groups, the gap between the D-domain and the T-stem is suitable for two nucleotides, and exactly this number of nucleotides come from the T-bulge. In the Type III clones, the additional sixth base pair in the T-stem makes the gap narrower and, correspondingly, only one nucleotide is provided by the T-bulge. Thus, despite the presence of such unusual elements in the DT region of the Type III clones, the general strategy used for the fixation of the juxtaposition of the helical domains in these clones follows essentially the same principles as in all other tRNAs.

The fact that all four elements specific to the Type III arrangement are found in all clones of this group and none of these elements existed in any other K-clone supports the idea that these elements are connected to each other through a chain of cause-effect relationships. We can suggest that all unusual features of the Type III clones originate from the appearance of guanosine in position 1* and the consecutive formation of the sixth base pair G1*-C8* in the T-stem. Then, the extended T-stem takes the space normally occupied by the last nucleotide of the T-loop, which leads to the reduction of the T-bulge by one nucleotide. Base pair 2*-6*, in its turn, should stack on base pair 1*-8* and be simultaneously connected to nucleotides 1* and 7*. We showed that the formation of the canonical RH-UA base pair at this place would cause interference between adenosine 6* and atom H1′ of C8*, while the alternative RH-RWC can be formed without problems. The replacement of RH-UA by RH-RWC creates, in its turn, another problem. The particular geometry of the RH base pair U2*-A6* disallowed its stacking to nucleotide 5*, which in the normal tRNAs and in the Type I and II clones resulted in the formation of either the specific or non-specific purine trap (Doyon et al. 2004). The RH-RWC base pair, despite the similarity to RH, does not share this feature, which opens the possibility for nucleotide 5* to comfortably stack to 6*. As a result, the formation of the purine trap becomes compromised. In the absence of the purine trap, the affinity between the two loops would also be compromised unless the purine trap is replaced by another type of inter-loop interaction. The formation of ILDH seems to be a

reasonable alternative way to restore the affinity between the loops. Thus, if the T-stem is extended by one base pair, the requirement for the proper juxtaposition of the two helical domains will with a high probability lead to the appearance of all other abnormal elements of the DT structure specific to the Type III clones. The performed MD simulations of the modeled structures demonstrate their stability compared to possible alternative arrangements.

The experimental data and the accompanying analysis presented here teach us some lessons concerning the general rules that govern the formation of RNA tertiary structure. First, the isostericity of base pairs is not necessarily associated with the same type of base pairing. Indeed, as it is shown in Figure 4, different RH base pairs, like different RWC base pairs are not isosteric to each other. At the same time, one can find isosteric base pairs across different types of base pairing. An example of this phenomenon is the RH-RWC base pairs, two of which form as RH, while the other two form as RWC. Despite this difference, in all RH-RWC base pairs the juxtaposition of the glycosidic bonds is practically the same, which makes all these base pairs isosteric. Second, additional minor aspects can strongly influence RNA structure. For example, the presence of a hydrogen atom at a particular place, like atom H1′ attached to atom C1′ of C8*, can strongly affect the formation of a base pair. So far, such aspects have never been taken into account or even considered essential. Our results show that for understanding the principles of RNA folding, such aspects should be taken into consideration.

Finally, a relatively minor change in the shape of a base pair can have major consequences for the whole structure. Thus, a shift in the position of the glycosidic bond of nucleotide 6* by only 3Å due to the replacement of the RH base pair U2*-A6* by RH-RWC allows this nucleotide to avoid the collision with C8*. As a result, the formation of the RH-RWC base pair becomes possible, even though the formation of the RH base pair U2*-A6* was not. The change in the structure of the 2*-6* base pair, in its turn, results in the replacement of the purine trap by the ILDH, thus making a complete rearrangement of the structure at the top of the T-loop. This phenomenon can be seen as an indication that the existence of interdependence between different elements goes well beyond their immediate contact. Given that these effects are found in tRNA, which is a relatively small

molecule, in larger RNAs with a more sophisticated tertiary structure the problem of interdependence between different elements is expected to be even more important and more complex. Revealing the facts of such interdependence and elucidating its nature is a necessary step toward understanding the mechanisms of RNA folding and function.

MATERIALS AND METHODS

Cloning, screening and measuring the suppressor tRNA activity

The design of the K-library is shown in Figure 2. It was based on the nucleotide sequence of the *E.coli* alanine tRNA and contained six fully randomized positions in each of the two loops, D and T. The choice of $tRNA^{Ala}$ as the prototype molecule to study structure–function relationships in tRNA was determined by the fact that the aminoacylation of $tRNA^{Ala}$ depends almost exclusively on the presence of base-pair G3-U70 in the acceptor stem and is practically insensitive to the juxtaposition of the two helical domains or even the existence of the D/anticodon domain (Prather et al. 1984; Hou and Schimmel 1988; McClain and Foss 1988). Therefore, any change in the tRNA performance caused by modifications at the DT region was not expected to be related to aminoacylation and thus would be attributed to steps of the tRNA functional cycle shared by other tRNAs. Compared to the *E.coli* alanine tRNA, the anticodon in this design was replaced by sequence CUA corresponding to the amber stop codon, while the T- and D-loops were extended by one and two nucleotides, respectively. The extensions of the loops were intended to provide for additional conformational flexibility, which would increase the probability of finding of new yet unknown structural forms.

The tRNA gene harbouring twelve randomized positions was PCR amplified and cloned into the pGFIB-1 plasmid, as described previously (Zagryadskaya et al. 2003), producing the library with the theoretical sequence complexity of 1.7×10^7. This library was then introduced into XAC-1 cells, having an amber nonsense mutation in the β-galactosidase gene. Functional clones were screened based on the appearance of blue colonies in the presence of X-gal, which amounted, on average, to one positive clone for every one thousand colonies screened. The details of the cloning and screening of the tRNA library as well as of the measuring the suppressor activity of the obtained tRNAs were described earlier (Zagryadskaya et al. 2003; Zagryadskaya et al. 2004).

The process of screening was stopped after a total of 53 non-redundant active clones had been obtained. Plasmids of active clones were isolated and retransformed into XAC-1 cells to verify that the suppression activity was due to the presence of the plasmid and not to a spontaneous host mutation. For each clone, the suppressor activity was measured and calculated as the percentage of the activity of the suppressor $tRNA^{Ala}$ having the normal structure of the DT region. Although the activities of the isolated clones were to some extent lower than that of the suppressor $tRNA^{Ala}$ (Table 1), they were essentially higher than the background level and thus could be reliably detected by the β-gal assay.

Computer modeling and Molecular Dynamics

Preliminary modeling was done interactively, using the InsightII/Discover package (Version 2000, Accelrys Inc., San Diego, CA). The X-ray structure of the yeast $tRNA^{Phe}$ (Shi and Moore 2000) was used as a starting conformation, to which the elements different from the standard tRNA structure were appended. Each model was submitted to the conjugate-gradient energy minimization in the AMBER forcefield with the distance-dependent dielectric constant set at 4.0 (Pearlman et al. 1995). The elements of the modeled structures identical to the corresponding elements in the yeast $tRNA^{Phe}$ were fixed during the minimization. The minimization continued until a clear energy minimum was reached.

Molecular dynamics simulations (MD) were performed in vacuum at 300ºK. All other conditions were the same as in the energy minimizations. Each structure contained nucleotides G15 and C48 in the D-domain, as well as nucleotides G53, 1*-8*, and C61 in the T-loop. In addition, the D-loop included the three consecutive nucleotides that formed the ILDH with nucleotides 3*-5* of the T-loop. During the MD simulations, the positions of nucleotides G15, C48, and G53 were fixed. The integrity of the base pairs within the ILDH was kept with use of the distant constraints imposed on the lengths of inter-base hydrogen bonds. The positions of all other nucleotides were unrestrained. Each complex was submitted to a 1-nanosecond MD simulation. Visualizations were done on a Silicon Graphics Fuel computer.

ACKNOWLEDGEMENTS

We thank Dr. Ludmila Chirai for help in characterization of some K-clones.

FUNDINGS

This work was supported by a discovery grant from the Natural Science and Engineering Research Council of Canada to S.V.S..

REFERENCES

Doyon FR, Zagryadskaya EI, Chen J, Steinberg SV. 2004. Specific and non-specific purine trap in the T-loop of normal and suppressor tRNAs. *Journal of Molecular Biology* **343**: 55-69.

Hou YM, Schimmel P. 1988. A Simple Structural Feature Is a Major Determinant of the Identity of a Transfer-RNA. *Nature* **333**: 140-145.

Kotlova N, Ishii TM, Zagryadskaya EI, Steinberg SV. 2007. Active suppressor tRNAs with a double helix between the D- and T-loops. *Journal of Molecular Biology* **373**: 462-475.

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

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

Lee JC, Cannone JJ, Gutell RR. 2003. The lonepair triloop: A new motif in RNA structure. *Journal of Molecular Biology* **325**: 65-83.

Levinger L, Bourne R, Kolla S, Cylin E, Russell K, Wang XD, Mohan A. 1998. Matrices of paired substitutions show the effects of tRNA D/T loop sequence on Drosophila RNase P and 3 '-tRNase processing. *Journal of Biological Chemistry* **273**: 1015-1025.

Mcclain WH, Foss K. 1988. Changing the Identity of a Transfer-RNA by Introducing a G-U Wobble Pair near the 3' Acceptor End. *Science* **240**: 793-796.

Nagaswamy U, Fox GE. 2002. Frequent occurrence of the T-loop RNA folding motif in ribosomal RNAs. *RNA* **8**: 1112-1119.

Pan D, Kirillov S, Zhang CM, Hou YM, Cooperman BS. 2006. Rapid ribosomal translocation depends on the conserved 18-55 base pair in P-site transfer RNA. *Nature Structural and Molecular Biology* **13**: 354-359.

Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, Debolt S, Ferguson D, Seibel G, Kollman P. 1995. Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Computer Physics Communications* **91**: 1-41.

Prather NE, Murgola EJ, Mims BH. 1984. Nucleotide Substitution in the Amino-Acid Acceptor Stem of Lysine Transfer-RNA Causes Missense Suppression. *Journal of Molecular Biology* **172**: 177-184.

Shi HJ, Moore PB. 2000. The crystal structure of yeast phenylalanine tRNA at 1.93 angstrom resolution: A classic structure revisited. *RNA* **6**: 1091-1105.

Söll D, Rajbhandary UL. 1995. Editors of tRNA: Structure, Biosynthesis, and Function, ASM Press, Washington, DC.

Zagryadskaya EI, Doyon FR, Steinberg SV. 2003. Importance of the reverse Hoogsteen base pair 54-58 for tRNA function. *Nucleic Acids Research* **31**: 3946-3953.

Zagryadskaya EI, Kotlova N, Steinberg SV. 2004. Key elements in maintenance of the tRNA L-shape. *Journal of Molecular Biology* **340**: 435-444.

Table 1. Nucleotide sequences of the D-and T-loops and suppressor activities of the selected tRNA clones

Clone	D-LOOP	T-LOOP	Activity
$tRNA^{Ala}_{cua}$	AG CUGG GA	UU CGA UC	100.0%
Type I			
K25	AG GAACGC UA	UG AAA AC	17.7%
K15	AG GCAUAU UA	UG AAA UC	11.0%
K2	AG AAAGAC UA	UGACGA UC	7.9%
K23	AG UAAGGU UA	UGCCAA UC	5.9%
K29	AG GAAAAA UA	UG GGA UC	5.1%
K3	AG AACGAA UA	UG AAA UC	4.1%
K42	AG UAGACA UA	UGACAA UC	3.8%
K39	AG GAAGAA UA	UGACGA UC	3.4%
K48	AG ACAGAC UA	UG GAA AC	2.5%
K49	AG AAAAGA UA	UGCAAA UC	2.2%
K34	AG AG A	UGGCGA UC	1.4%
Type II			
K27	AG UGAAAU UA	UAGCCA UC	9.9%
K24	AG AAAAAC UA	UAGCCA UC	6.0%
K41	AG A CA	UAAACA UC	4.4%
K26	AG AACGAC UA	UAAACA UC	3.9%
K47	AG AGAGCA UA	UAGCCA UC	3.3%
K20	AG GAGAUC UA	UAGCCA UC	3.2%
K40	AG AAACAC UA	UAGCCA UC	2.8%
K36	AG AAAAAA UA	UAGCCA UC	2.7%
K18	AG AACAAA UA	UAAACA UC	2.5%
K33	AG GAAGAA UA	UAGUCA UC	2.0%
K19	AG ACAAC UA	UAUACA UC	2.0%
K5	AG CGAAGA UA	UAGCCA UC	1.7%
K38	AG AAAUAC UA	UAAACA UC	1.5%
K51	AG ACCAAA UA	UAACCA UC	1.4%
K7	AG GACAAA UA	UAACCA UC	1.3%
K1	AG GAGAAC UA	UAACCA UC	1.3%
Type III		1* 2* 3* 4* 5* 6* 7* 8*	
K32	AG UCGGUAUA	GU CGAG UC	38.5%
K6	AG AGGGAGUA	GC ACA UC	25.0%
K45*	AG GCCACAUA	GC GGCA UC	18.7%
K43*	AG CGGGAUUA	GC ACCA UC	17.0%
K64*	AG UAGGU UA	GC GCCA UC	15.6%
K37*	AG AUGAUUUA	GG AAG GU	13.0%
K44*	AG ACGGACUA	GC GCCA UC	12.3%
K61*	AG GGCAU UA	GA GGCA UC	11.1%
K66*	AG GGGGU UA	GC ACCA UC	8.1%
K52*	AG GACCUAUA	GC GGUA UC	7.0%
K17	AG AGGCCAUA	GA GCCA UC	6.5%
K31	AG AGAGGGUA	GC CCAA UC	6.3%
K10	AG AAGGA UA	GU ACCA UC	5.9%

K16	AG GGGGAUUA	GU CAAG UC	4.8%
K35*	AG AAAGGAUA	GG ACCG UC	4.4%
K62*	AG ATGGU UA	GG ACCG UC	4.2%
K14	AG GAGGGAUA	GG ACCG UC	4.1%
K63*	AG GGGU UA	GG GCCG UC	3.3%
K28	AG GGCAAAUA	GC AGCA UC	2.9%
K21	AG UGAAAGUA	GC CACA UC	2.8%
K65*	AG AUGGU UA	GT ACCG UC	2.1%
K50*	AG AAGGGAUA	GG ACCG UC	1.8%
K9	AG AGCGAAUA	GA CGCA UC	1.3%
K13	AG AGGAAAUA	GU ACCG UC	1.3%
K4	AG A CGGGUA	GC ACAA UC	1.1%
Non-classified clone			
K30	AG UGAGGAUA	U CCAA AU	10.8%

The nucleotide sequences of the K-clones. Only the sequences of the D- and T-loops are shown. Outside these loops, all sequences are identical and correspond to that shown in Fig 2. In the wt $tRNA^{Ala}_{CUA,}$ the interactions important for the structuring of the DT region are shown as lines connecting two interacting nucleotides. The nucleotides in the D-loop that interact with those from the T-loop are underlined. Type I follows closely the interactions shown by the wt $tRNA^{Ala}_{CUA}$, while for Type II and Type III clones, the interaction of nucleotides playing the equivalent structural role are shown by a different connection of lines. In more detail, Type I and II structures are discussed in (Doyon et al. 2004). In the Type III clones, the lines shows the interaction between G1* and C8*, the nucleotides forming RH-RWC, and those forming the base pairs within the ILDH. The suppressor activity of the $tRNA^{Ala}_{CUA}$ was taken for 100%. The nucleotide sequences presented here for the first time are marked by asterisks. A color version of this Table can be found in the supplemental material.

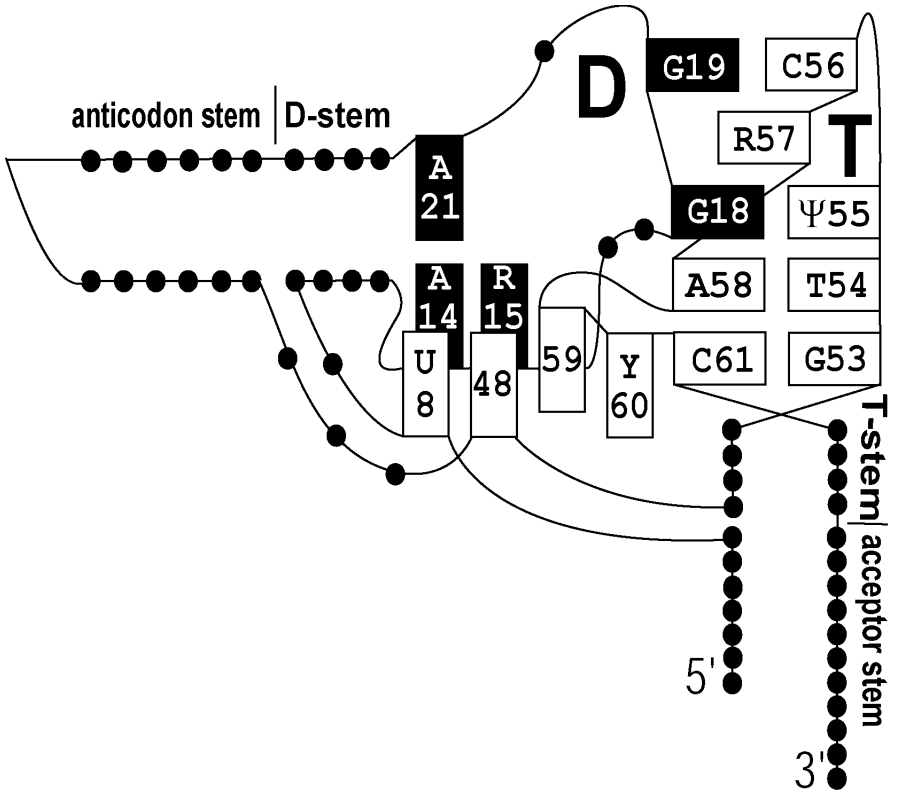


Figure 1

A conventional representation of the DT region in the standard tRNA structure. 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- loop as well as nucleotides 8, 48, 53 and 61 are white. All other nucleotides are shown as black dots. Nucleotides of the anticodon loop are not represented. The identities of the conservative and semi-conservative nucleotides are indicated as follows: R, purine; Y, pyrimidine; T, 5-methyluridine; and Ψ pseudouridine. T and Ψ are post-transcriptional modifications of U. The central role in the DT arrangement is played by the reverse-Hoogsteen base-pair (RH) T54–A58 (see the text).

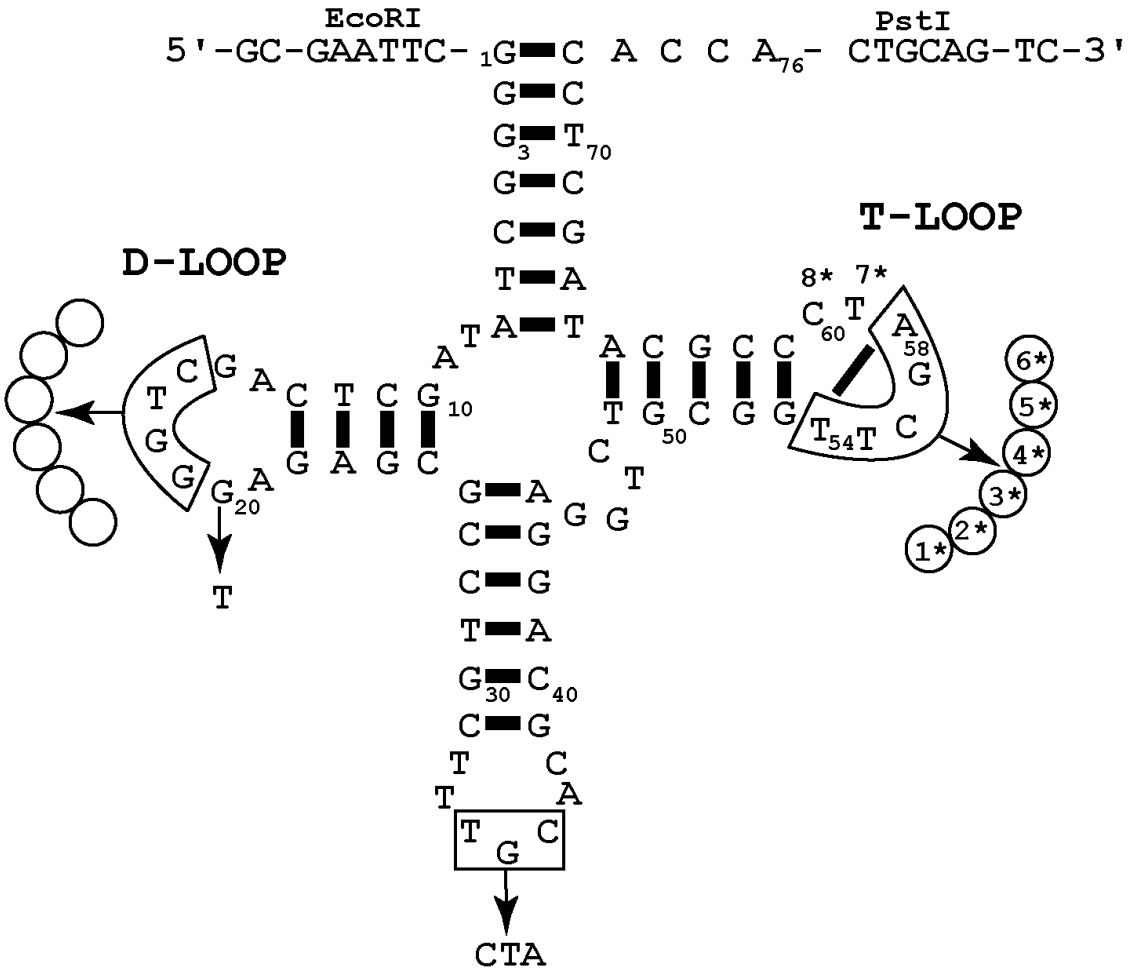


Figure 2

Design of the K-library. The design was based on the sequence of $tRNA^{Ala}_{CUA}$ (Zagryadskaya et al. 2003). The alanine anticodon TGC was replaced by the amber anticodon CTA. The D- and T-loops were extended by two and one nucleotide, respectively. The randomized positions are indicated by circles. The nucleotides of the T-loop have their own numbering with an asterisk. The EcoRI and PstI restriction sites that are seen flanking the 5'- and 3'-termini were used for the cloning of the library into the pGFIB-1 plasmid.

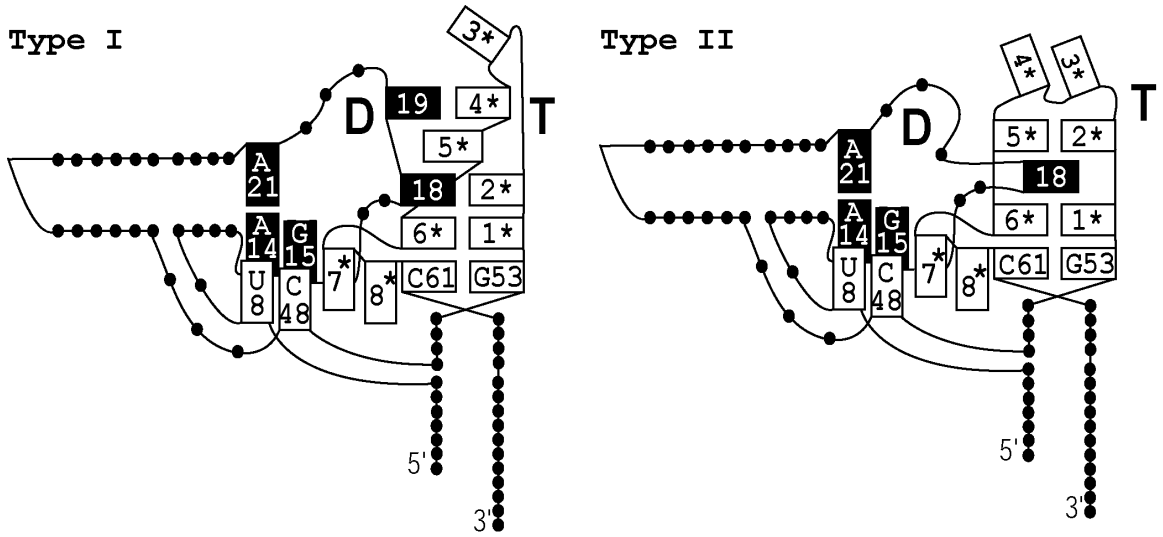


Figure 3

Comparison of the structure of the DT region in Type I and Type II tRNAs. The DT region is shown in the same way as in Figure 1. Nucleotides outside the T-loop carry the standard numbers, while nucleotides of the T-loop are numbered from 1* to 8*. In both structures, RH is formed between nucleotides 1* and 6* of the T-loop, thus allocating two nucleotides for the T-loop bulge. The two nucleotides 7* and 8* form the T-bulge, which stacks to base-pair 15–48 and fills the gap between the D-domain and T-stem. The specific purine trap (left) is arranged between nucleotides 5* and 6*. It harbors an adenosine from the D-loop, which also forms base-pair AG with the guanosine in position 2*. This AG base-pair is equivalent to base-pair G18-Ψ55 in the normal tRNAs (Doyon et al. 2004). At the top of the arrangement, two nucleotides from both loops can form a base-pair equivalent to base-pair G19–C56 in the standard tRNA structure. This base pair, however, does not exist in some Type I clones. In the non-specific purine trap (right), there are two reverse-Hoogsteen base-pairs, U1*-A6* and A2*-C5* that are positioned at a double distance from each other. Purine 18 of the D-loop intercalates between these two base-pairs, while nucleotides 3* and 4* close the loop (Doyon et al. 2004).

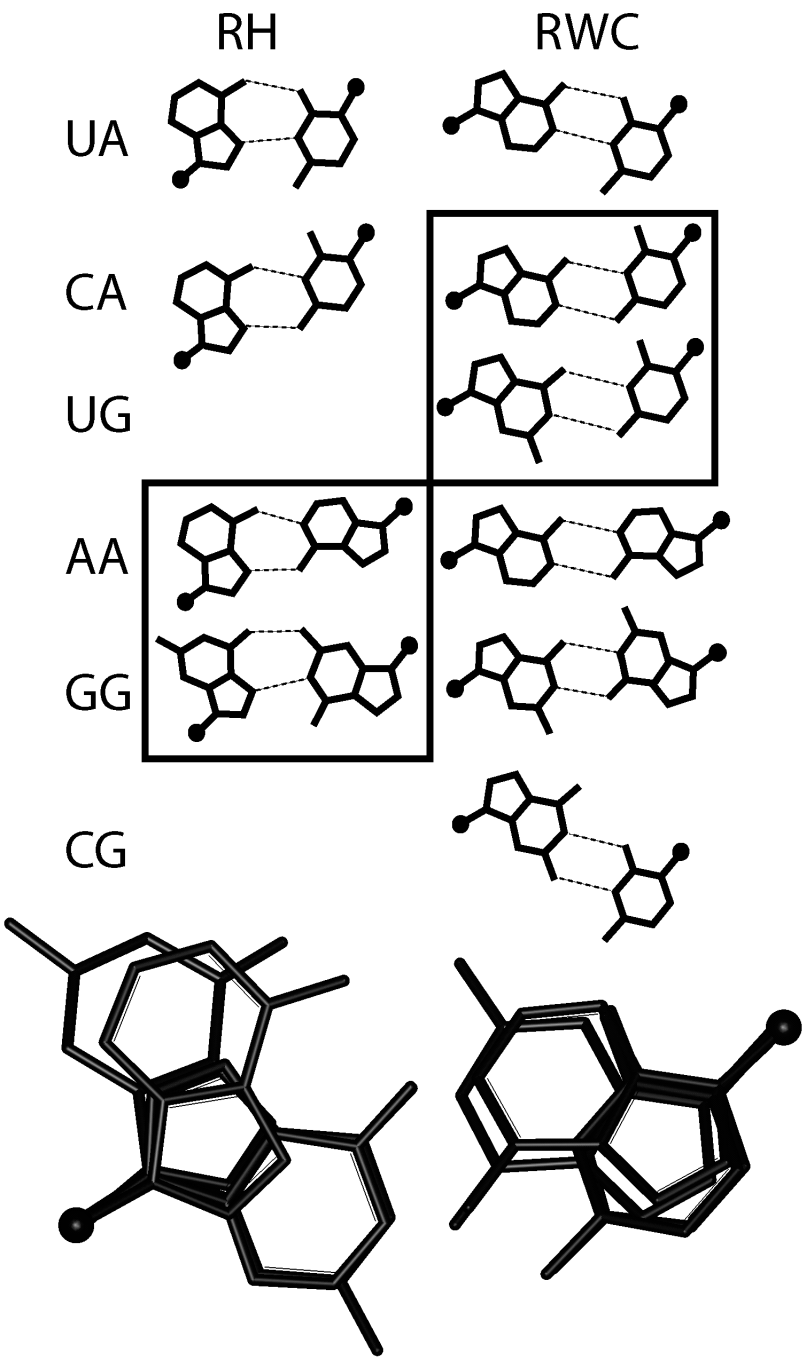


Figure 4

Comparison of different dinucleotide arrangements formed as reverse-Hoogsteen (RH, left) and reverse-Watson-Crick (RWC, right) base pairs. The dotted lines represent hydrogen bonds. The positions of the C1'-atoms are shown as black spheres. Boxed base pairs are those four RH and RWC base pairs that have about the same juxtaposition of the glycosidic bonds (shown at the bottom). In the paper, these four base pairs are called RH-RWC.

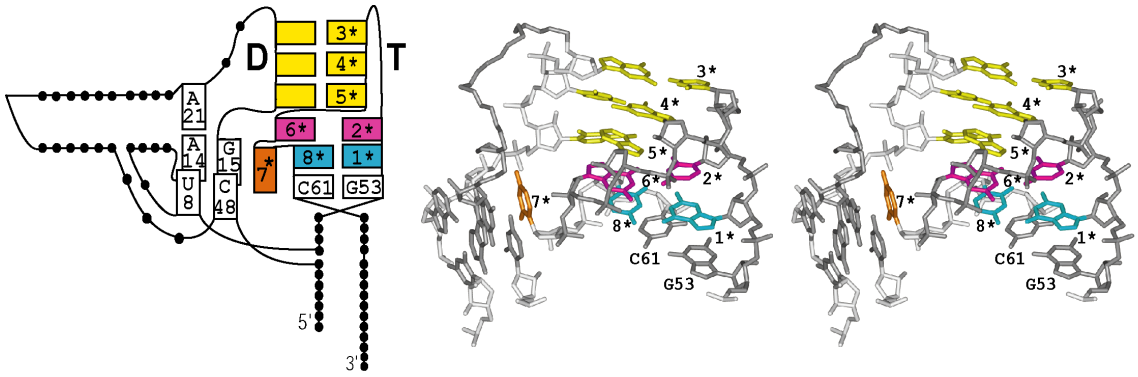


Figure 5

The structure of the DT region in the Type III tRNAs. On the left: a representation of the DT region in the context of the whole tRNA L-shape as in Figures 1 and 3. On the right: a stereo view of the DT region in the three-dimensional model of clone K32. For the same nucleotides, the same color is used in both right and left panels. Compared to Type I and Type II clones, the DT region in Type III clones has four unusual elements: the sixth base pair 1*-8* in the T-stem (cyan), the one-nucleotide T-bulge 7* (orange), the RH-RWC base pair 2*-6* (magenta), and the ILDH (yellow).

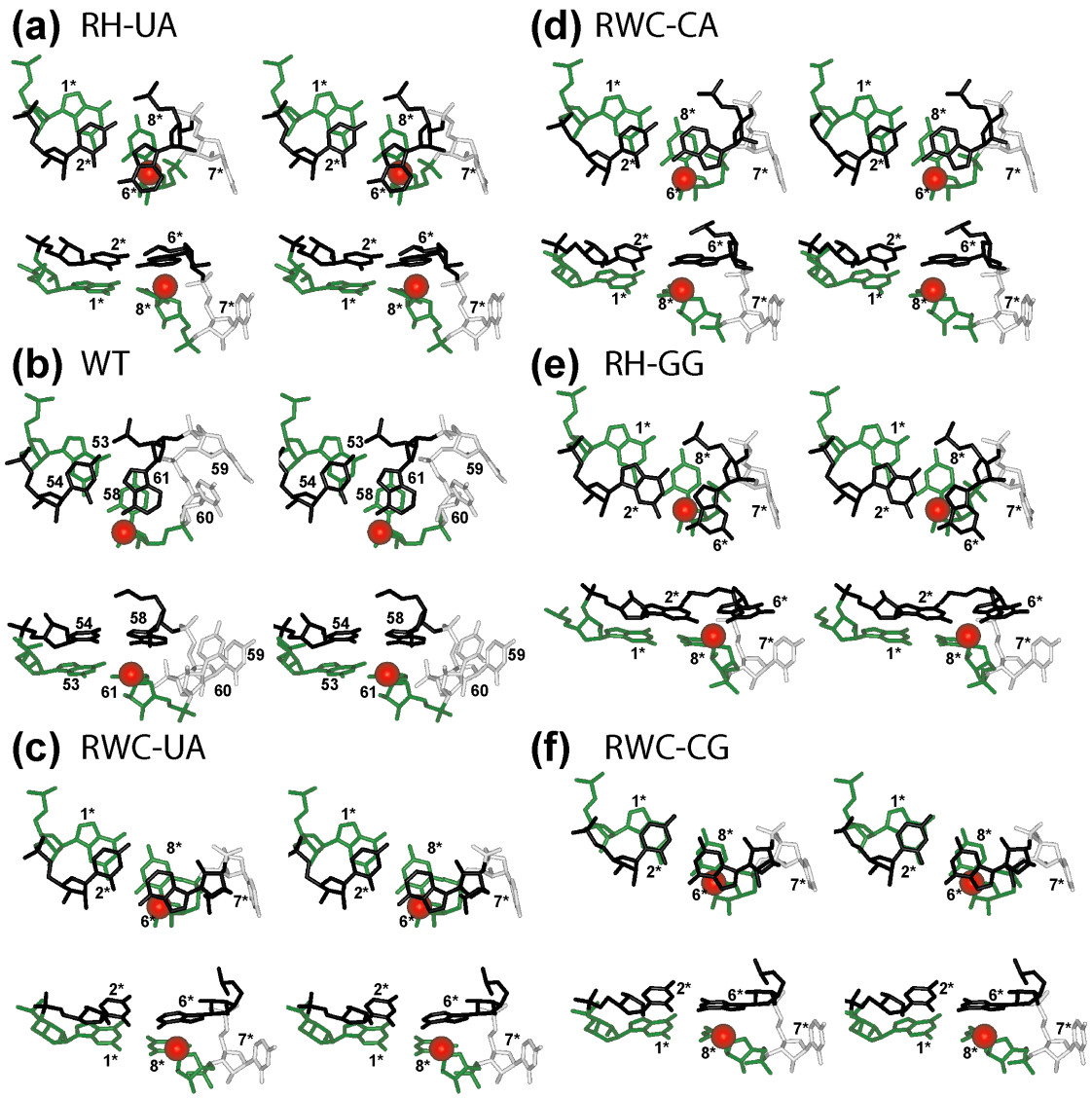
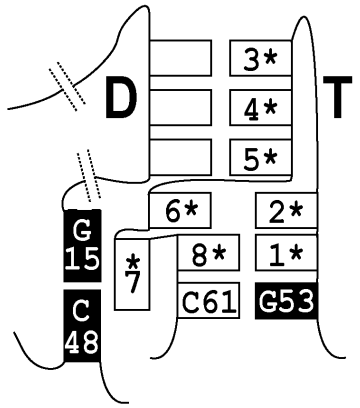


Figure 6

Role of the H1′ atom of nucleotide C8* in discrimination against base pairs RH U2*-A6*, RWC U2*-A6* and RWC C2*-G6*. In all panels, base pair 1*-8* is green, base pair 2*-6* is black, the T-bulge is gray, while atom H1′ of C8* is red. The radius of the ball representing atom H1′ of C8* corresponds to the van der Waals radius of hydrogen. In each case, the upper and lower panels show the same structure from two different perspectives. In the cases of RH-RWC base pairs (E,F) atom H1′ of C8* does not interfere with the stacking of nucleotides 6* and 8*. For base pair RWC-UA (C), the interference is notable, while for base pairs RWC-CG (F) and RH-UA (A) it is unbearable. The latter case (A) can be compared to the WT tRNA (B), in which the

presence of two nucleotides 59-60 in the T-bulge allows A58 to avoid the interference with atom H1′ of C61.



	2*-6*	1*-8*	G53-C61	ALL
RH-GG	100%	100%	100%	100%
RWC-UG	90%	85%	85%	85%
RWC-CA	100%	70%	97%	70%
RH-AA	100%	60%	100%	60%
RWC-UA	100%	96%	20%	20%
RWC-CG	15%	96%	20%	<1%
RH-UA	<1%	15%	100%	<1%

Figure 7

The summary of the results of the MD simulations for different Type III arrangements. On the left: the structure of the DT region shown as in Figure 5. Only the nucleotides represented by white rectangles were allowed to move during the simulations. The integrity of the ILDH was kept through introduction of distance constraints on the hydrogen bonds within the base pairs. The simulations were performed for the arrangements containing base pairs RH-UA, RH-AA, RH-GG, RWC-UG, RWC-UA, and RWC-CG in position 2*-6*. Numbers show the fraction of the total simulation time (1 nanosecond) when the integrities of each of the three monitored base pairs individually and all of them simultaneously were maintained. The detailed results of the simulations are presented in the supplemental material.

Supplemental material

Table 1. Nucleotide sequences of the D-and T-loops and suppressor activities of the selected tRNA clones

Clone	D-LOOP	T-LOOP	Activity
$tRNA^{Ala}_{cua}$	AG CUGG GA	UU CGA UC	100.0%
Type I			
K25	AG GAACGC UA	UG AAA AC	17.7%
K15	AG GCAUAU UA	UG AAA UC	11.0%
K2	AG AAAGAC UA	UGACGA UC	7.9%
K23	AG UAAGGU UA	UGCCAA UC	5.9%
K29	AG GAAAAA UA	UG GGA UC	5.1%
K3	AG AACGAA UA	UG AAA UC	4.1%
K42	AG UAGACA UA	UGACAA UC	3.8%
K39	AG GAAGAA UA	UGACGA UC	3.4%
K48	AG ACAGAC UA	UG GAA AC	2.5%
K49	AG AAAAGA UA	UGCAAA UC	2.2%
K34	AG AG A	UGGCGA UC	1.4%
Type II			
K27	AG UGAAAU UA	UAGCCA UC	9.9%
K24	AG AAAAAC UA	UAGCCA UC	6.0%
K41	AG A CA	UAAACA UC	4.4%
K26	AG AACGAC UA	UAAACA UC	3.9%
K47	AG AGAGCA UA	UAGCCA UC	3.3%
K20	AG GAGAUC UA	UAGCCA UC	3.2%
K40	AG AAACAC UA	UAGCCA UC	2.8%
K36	AG AAAAAA UA	UAGCCA UC	2.7%
K18	AG AACAAA UA	UAAACA UC	2.5%
K33	AG GAAGAA UA	UAGUCA UC	2.0%
K19	AG ACAAC UA	UAUACA UC	2.0%
K5	AG CGAAGA UA	UAGCCA UC	1.7%
K38	AG AAAUAC UA	UAAACA UC	1.5%
K51	AG ACCAAA UA	UAACCA UC	1.4%

K7	AG GACAAA UA	UAACCA UC	1.3%
K1	AG GAGAAC UA	UAACCA UC	1.3%

Type III

K32	AG UCGGUAUA	GU CGAG UC	38.5%
K6	AG AGGGAGUA	GC ACA UC	25.0%
K45*	AG GCCACAUA	GC GGCA UC	18.7%
K43*	AG CGGGAUUA	GC ACCA UC	17.0%
K64*	AG UAGGU UA	GC GCCA UC	15.6%
K37*	AG AUGAUUUA	GG AAG GU	13.0%
K44*	AG ACGGACUA	GC GCCA UC	12.3%
K61*	AG GGCAU UA	GA GGCA UC	11.1%
K66*	AG GGGGU UA	GC ACCA UC	8.1%
K52*	AG GACCUAUA	GC GGUA UC	7.0%
K17	AG AGGCCAUA	GA GCCA UC	6.5%
K31	AG AGAGGGUA	GC CCAA UC	6.3%
K10	AG AAGGA UA	GU ACCA UC	5.9%
K16	AG GGGGAUUA	GU CAAG UC	4.8%
K35*	AG AAAGGAUA	GG ACCG UC	4.4%
K62*	AG ATGGU UA	GG ACCG UC	4.2%
K14	AG GAGGGAUA	GG ACCG UC	4.1%
K63*	AG GGGU UA	GG GCCG UC	3.3%
K28	AG GGCAAAUA	GC AGCA UC	2.9%
K21	AG UGAAAGUA	GC CACA UC	2.8%
K65*	AG AUGGU UA	GT ACCG UC	2.1%
K50*	AG AAGGGAUA	GG ACCG UC	1.8%
K9	AG AGCGAAUA	GA CGCA UC	1.3%
K13	AG AGGAAAUA	GU ACCG UC	1.3%
K4	AG A CGGGUA	GC ACAA UC	1.1%

Non-classified clone

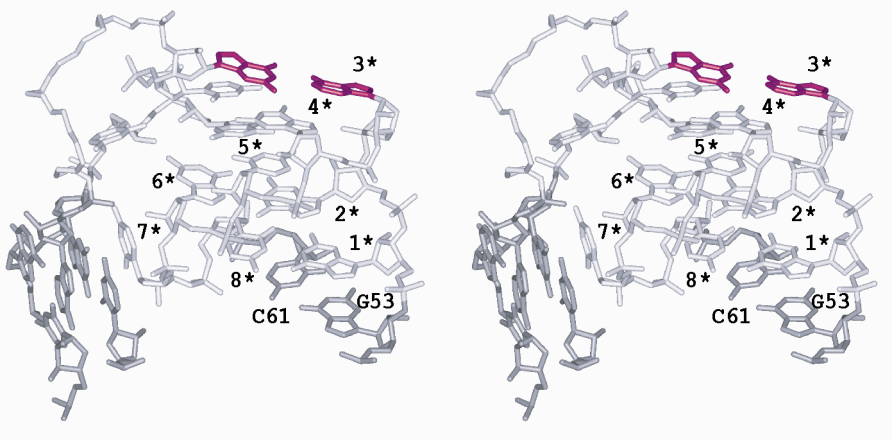
K30	AG UGAGGAUA	U CCAA AU	10.8%

Legend to the Color version of Table 1

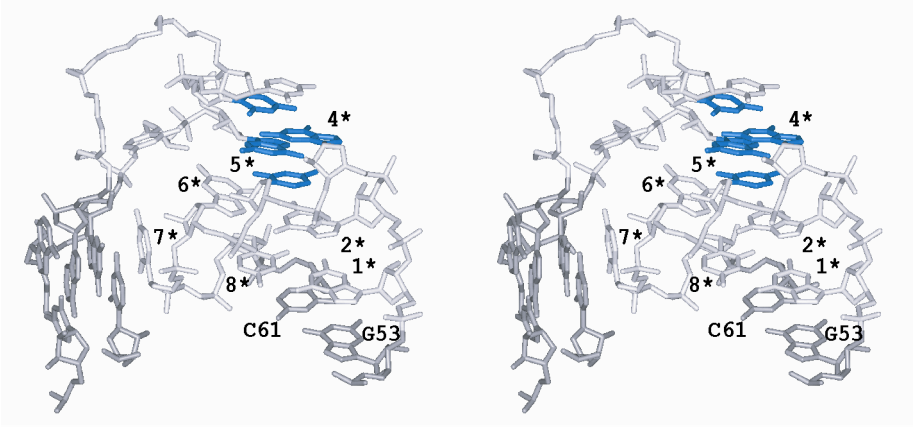
The nucleotide sequences of the K-clones. Only the sequences of the D- and T-loops are shown. Outside these loops, all sequences are identical and correspond to that shown in Fig 2. In the wt $tRNA^{Ala}_{CUA}$ and in the Type I and II clones, the nucleotides playing the

equivalent structural role are shown with the same color. In more detail, Type I and II structures are discussed in (Doyon et al. 2004). In the Type III clones, G1* and C8* are cyan; the nucleotides forming RH-RWC are magenta; those forming WC, AG and GU base pairs within the ILDH are yellow, green and green, respectively. The nucleotides of the T-bulge are orange in all sequences. The suppressor activity of the $tRNA^{Ala}_{CUA}$ was taken for 100%. The nucleotide sequences presented here for the first time are marked by asterisks.

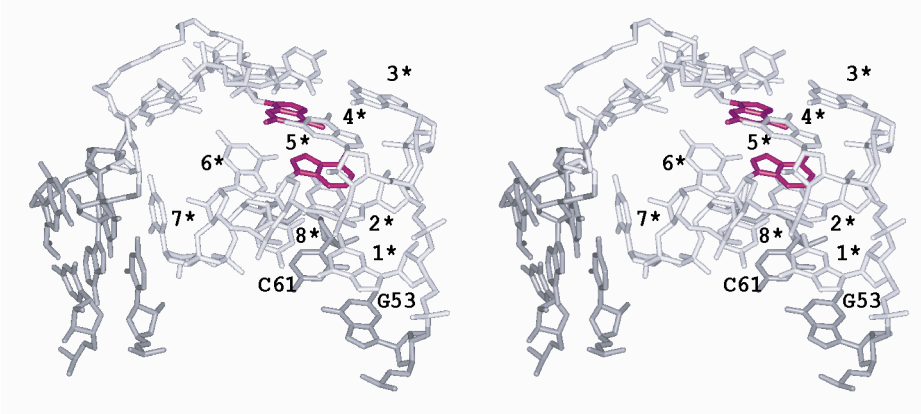
The structure of the DT region modeled for particular Type III clones.



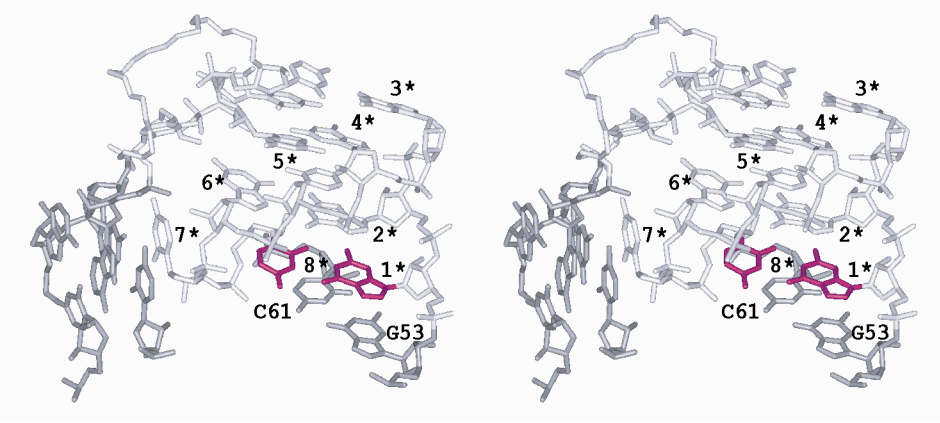
(A) In five clones K14, K43, K44, K50, and K61 the base pair in which nucleotide 3* is involved is AG. Our analysis shows that the AG base pair in this position can be formed in the head-to-head (cis W.C/W.C, magenta) way without affecting the positions of neighbouring nucleotides.



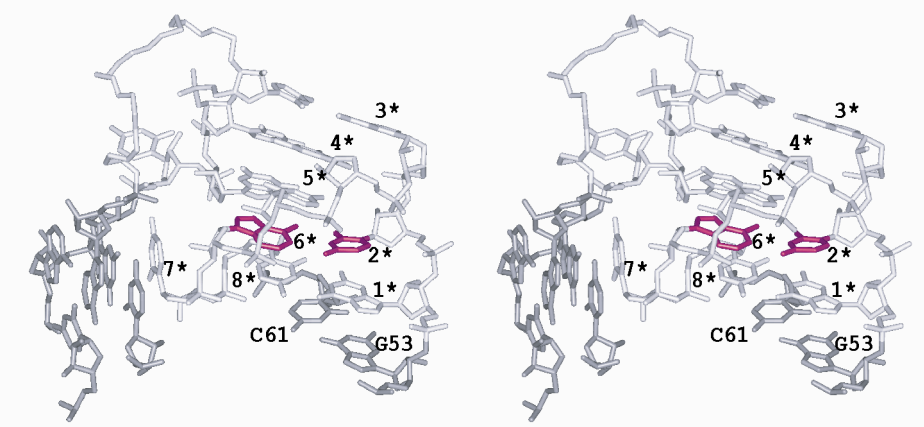
(B) The absence of nucleotide 3*, which happens in clones K6 and K37, does not interfere with a proper connection between nucleotides 2* and 4*. The two base pairs forming ILDH are blue.



(C) While the base pair formed by nucleotide 4* is always WC, the next base pair involving nucleotide 5* is GA in two clones K4 and K31 (magenta). Our modeling experiments show that the formation of such a base pair as RH (cis W.C/Hoogsteen) optimized its stacking interactions with its neighbouring stacking base-pairs.



(D) While in most clones, the 1*-8* base pair is GC, in clone K37 it has the GU identity (magenta). Our modeling experiments show that the replacement of C8* by U has only minor local effect and does not compromise the integrity of the whole structure.



(E) As mentioned above, in clone K10 the 2*-6* base pair has a unique nucleotide combination UA. When this base pair is formed as RWC, its structure will be relatively close to, although still different from that of base pairs CA and UG (magenta). The modeling experiments show that although acceptable, such base pair would stack to the previous base pair G1*-C8* not as good as base pairs CA or UG and thus should be considered as a compromising variant.

The MD simulations of the DT region for particular Type III models

(F) The Type III models containing RH-AA, RH-GG, RH-UA, RWC-UA, RWC-UG, RWC-CG, and RWC-CA in position 2*-6* were subjected to MD simulations for 1 ns. During the simulation, those nucleotides that were allowed to move freely and those the positions of which were frozen are shown in Fig 7 of the main text (Left). For each simulation, the integrity of the hydrogen bonds within three base pairs 2*-6*, 1*-8*, and C53-G61 was followed. Correspondingly, for each simulation, we have three graphs showing how the distance between the corresponding electro-negative atoms was changing with time. In all graphs, the measured lengths range from 2 Å (the bottom border of each panel) to 8 Å (the top border of each panel) with the increment of 1 Å corresponding to the distance between neighboring horizontal lines. In all graphs, the time spans from 0 to 1 nanosecond.

For RH-GG, all three base pairs were maintained for 100% of the time.

For RWC-UG, base pair 2*-6* was maintained for 90%. Base pairs 1*-8* and G53-C61 were both maintained for 85%. All three base-pairs were maintained for approximately 85%.

For RWC-CA, base pairs 2*-6*, 1*-8* and G53-C61 were maintained for 100%, 70% and 97%, respectively. All three base-pairs were maintained for 70% of the time.

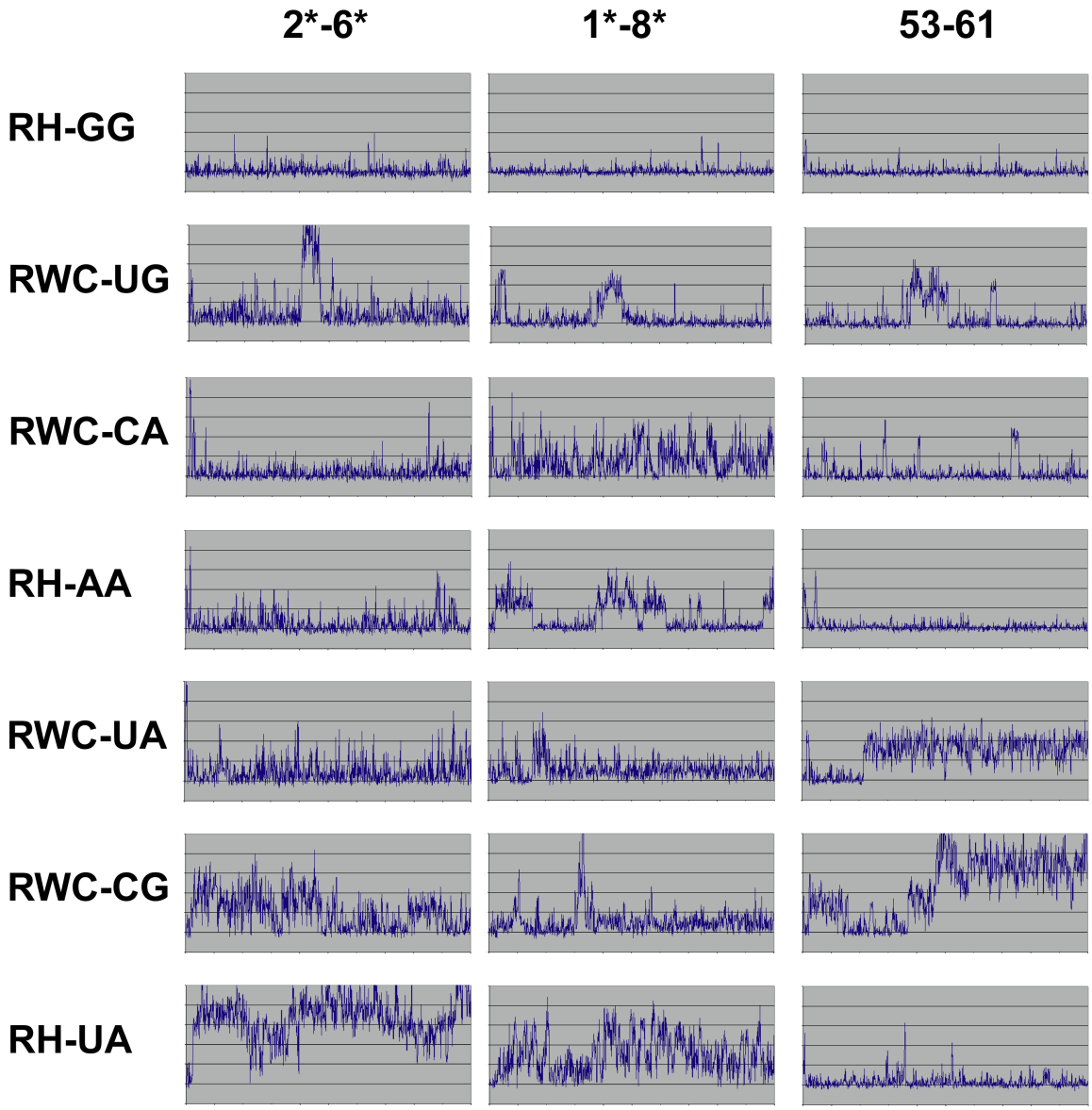
For RH-AA, base pairs 2*-6* and G53-C61 were maintained throughout the simulation, while base pair 1*-8* was maintained for 60% of the time. All three base pairs are maintained for 60% of the time.

For RWC-UA, base pairs 2*-6*, 1*-8* and G53-C61 were maintained for 100%, 96%, and 20% of the time, respectively. All three base-pairs were maintained for 20% of the time.

For RWC-CG, base pairs 2*-6*, 1*-8* and G53-C61 were maintained for about 15%, 98%, and 20% of the time, respectively. However, in no moment all three base pairs were maintained simultaneously.

For RH-UA, base pair 2*-6* broke within the first 1% of the simulation, base pair 1*-8* was maintained for 15%, and 53-61 for 100%. All three base-pairs were thus maintained for less than 1% of the time.

Based on the length of time all three base-pairs were maintained, the following ranking of models can be made starting with the most stable: RH-GG (100%), RWC-UG (85%), RWC-CA (70%), RH-AA (60%), RWC-UA (20%), RWC-CG (<1%), and RH-UA (<1%).



CHAPTER IV

The long D-stem of the Selenocysteine tRNA provides Resilience at the expense of Maximal function

Tetsu M. Ishii, Natalia Kotlova, Franck Tapsoba and Sergey V. Steinberg[1]

Department of Biochemistry, Université de Montréal, H3C 3J7, Succ. Centre-ville, Montréal, Québec, Canada.

Tetsu M. Ishii designed the experiments, performed experiments, and wrote the paper.
Natalia Kotlova and Franck Tapsoba assisted in the collection of some of the wet data.
Sergey V. Steinberg designed the experiments, analyzed the data, and wrote the paper.
This chapter is published. J Biol Chem. 2013 May 10;288(19):13337-44.

*Running title: *Selenocysteine tRNA: a long D-stem*

[1]To whom correspondence should be addressed

Keywords: tertiary structure, *in vivo* screening, combinatorial library, instant evolution

---

Background: The selenocysteine tRNA ($tRNA^{Sec}$) has a uniquely long D-stem containing six base pairs.
Results: The extended D-stem is not essential for function but is required for stability.
Conclusion: Enhanced secondary structure in selenocysteine tRNA compensates the absence of canonical tertiary interactions.
Significance: The flexibility due to the absence of tertiary interactions are required for $tRNA^{Sec}$ function, while the enhanced secondary structure compensate the decreased stability.

SUMMARY

The D-stem of the selenocysteine tRNA ($tRNA^{Sec}$) contains two additional base pairs, which replace tertiary interactions 8-14 and 15-48 universally present in all other cytosolic tRNAs. To study the role of these additional base pairs in the $tRNA^{Sec}$ function we used the Instant Evolution approach. *In vivo* screening of six combinatorial gene libraries provided 158 functional variants of the *E. coli* $tRNA^{Sec}$. Analysis of these variants showed that the additional base pairs in the D-stem were not required for the $tRNA^{Sec}$ function. Moreover, at lower temperatures, these base pairs notably harmed the $tRNA^{Sec}$ activity. However, at elevated temperatures these base pairs became

essential, as they made the tRNA structure more stable. The alternative way to stabilize the structure through formation of the standard tertiary interactions was not an option for $tRNA^{Sec}$ variants, which suggests that the absence of these interactions and the resulting flexibility of the tertiary structure are essential for $tRNA^{Sec}$ function.

Selenocysteine (Sec) is the 21st amino acid co-translationally inserted into the nascent polypeptide chain. The insertion takes place at the UGA codon when the latter is accompanied by a particular downstream mRNA secondary structure termed SECIS (Yoshizawa and Bock, 2009). Selenocysteine is delivered to the ribosome by a special selenocysteine-tRNA ($tRNA^{Sec}$) in the ternary complex with GTP and a special elongation factor SelB (analog of EF-Tu) (Yoshizawa and Bock, 2009). Among all cytosolic tRNAs, the $tRNA^{Sec}$ is distinguished by its unusual secondary structure, of which the most documented feature is its long acceptor stem (Leinfelder et al., 1988; Sturchler et al., 1993). While in all other cytosolic tRNAs the acceptor stem has universally seven base pairs, in the $tRNA^{Sec}$ it has either eight (in bacteria) or nine (in archaea and eukaryotes) base pairs. Although the presence of the additional base pairs in the acceptor stem of the $tRNA^{Sec}$ has been clearly established (Burkard and Söll, 1988; Itoh et al., 2009; Leinfelder et al., 1988; Palioura et al., 2009), it remains unclear how a tRNA with such an unusual feature is able to share the ribosomal sites with all other tRNAs having the invariable seven base pair acceptor stem.

Another abnormal feature of the $tRNA^{Sec}$ is the long D-stem, which has at least six base pairs instead of the normal four. Although such an abnormal length of the D-stem is unprecedented among tRNAs, it is not expected to create major problems for the $tRNA^{Sec}$ function, as the two additional base pairs in the D-stem replace the universal tertiary base pairs 8-14 (trans-Watson-Crick/Hoogsteen base pair U8-A14) and 15-48 (trans Watson-Crick/Watson-Crick base pair) without changing the overall shape of the molecule. Still, it remains unclear how such an extension of the D-stem can be beneficial for the function of the $tRNA^{Sec}$. In the archaeal and eukaryotic $tRNAs^{Sec}$, the long D-stem may be important for the phosphorylation of the seryl moiety by O-phosphoseryl-tRNA kinase, an intermediate step required for the successful delivery of amino acid selenocysteine (Chiba et al., 2010; Wu and Gross, 1994). However, in bacteria this reaction does not exist (Carlson et al., 2004; Turanov et al.) and still, the D-stem of all bacterial $tRNA^{Sec}$ contains six base pairs. Thus, the particular role of the extended stem and how critical it is for the bacterial $tRNA^{Sec}$ function remain unknown.

Here, based on the analysis of 158 *in vivo* screened functional variants of the *E. coli* $tRNA^{Sec}$, we show that the presence of six base pairs in the D-stem is not a prerequisite for the $tRNA^{Sec}$ function. Some $tRNA^{Sec}$ variants with only five or four base pairs in the D-stem had

robust activity *in vivo*. Surprisingly, when the activity was measured at a lower temperature, some variants with a shortened D-stem significantly outperformed the wild-type $tRNA^{Sec}$ (WT). Therefore, the long D-stem is not only dispensable, but can even be inhibitory for the $tRNA^{Sec}$ function.

EXPERIMENTAL PROCEDURES

Bacterial Strains- The *E.coli* strain WL81460 (Δ(*argF-lac*)U169 *rpsL150*, *rpsL*$^{+}$ *rpsE13*, Δ(*srl-recA*)306::Tn*10*, Δ(*selC*)400::*Kan*)(Zinoni et al., 1990), having a deletion in the gene coding for the $tRNA^{Sec}$ (SelC), and WL81300(Δ(*argF-lac*)U169 *rpsL150*, *rpsL*$^{+}$ *rpsE13*, Δ(*srl-recA*)306::Tn*10*, Δ(*selB*)300::*Kan*)(Tormay et al., 1996) , having a deletion in the gene coding for the elongation factor SelB, were used in this study.

Combinatorial Library designs and cloning for Instant Evolution- Oligonucleotides used for all six libraries (Figure 1c) and for the primers used for the library amplification were ordered from Bio-Corp Inc. (Montreal, Canada). The nucleotide sequences of all combinatorial libraries are listed in the supplemental information. Each library was PCR amplified using the two primers 5'CGGAATTCGGAAGATC3' and 5'TTCCAATGCATTGGCTGCAGTGGCGGAA GATCACAGGAGTCGAACCTGC3' and was cloned into plasmid pGFIB-1 using restriction sites EcoRI and PstI as described before (Zagryadskaya et al., 2003).

In vivo screening of active $tRNA^{Sec}$ variants- Active colonies were screened after plating the ligation reaction onto MacConkey Nitrate Agar plates (Barrett et al., 1979) and incubation for 48 hours at 37°C under anaerobic conditions. MacConkey nitrate agar contained (per liter): 40 g of MacConkey agar base (Difco MacConkey Agar Base), 10 g of potassium nitrate, and 0.5 g of sodium formate. Bright white colonies were picked and sequenced.

Preparing samples for the Formate Dehydrogenase H assay- The formate dehydrogenase H (FDH_H) assay was performed as described elsewhere (Takahata et al., 2008), with some modifications. Plasmids of selected clones were transformed into the WL81460 strain, plated onto LB agar plates containing ampicillin and grown overnight. Colonies were picked and grown overnight in 2 mL of LB with ampicillin under aerobic conditions. Subsequently, 1.5 mL eppendorf tubes were filled up with Stadtman buffer containing ampicillin, inoculated with 40 microliters of the overnight culture, and incubated for 20 hours at 30˚C, 33˚C, 37˚C or 42˚C. The density of samples was measured at O.D. 600nm. Samples were sedimented by centrifugation, frozen and stored at -80˚C.

Formate Dehydrogenase H assay- Samples were defrosted at room temperature. The pellets were washed with 300 microliters of 0.5X TBE containing 5mM $MgSO_4$, re-suspended in 800 microliters of the same buffer and transferred to a glass cuvette containing 100 microliters of sodium formate [200mM] and 100 microliters of

benzyl viologen [20mM]. Both sodium formate and benzyl viologen solutions were prepared with pH 7.0 0.1M potassium phosphate buffer. The cuvette was covered with a rubber stopper, deoxygenated for 5 minutes with argon, and was left at room temperature for another 5 minutes. The FDH_H activity was assayed by measuring the rate of the increase in the absorbance at 600nm during 5 minutes. The activities were represented as the change in O.D. at 600nm per minute normalized by the density of the cell culture. Each assay was performed in triplicate and then averaged.

Northern Blot of $tRNA^{Sec}$ variants- The northern blot was performed as described before (Kotlova et al., 2007) with some modifications. For RNA extraction, the bacteria were grown under anaerobic conditions in Stadman buffer at 30˚C and 37˚C until the OD of absorbance at 600 nm reached 0.4. One probe was complementary to the extra-arm (5'-CGGGACCGCTGGCGGCCCCA-3') of the *E.coli* $tRNA^{Sec}$, while the other one was complementary to region 34–53 of the *E. coli* 5S rRNA(5'- TTCTGAGTTCGGCATGGGGT-3'). The 5S rRNA probe was used to monitor the amount of total RNA in each sample. Densitometry analysis of the bands were performed using the Quantity One® software (BioRad).

RESULTS

General approach- For analysis of the structure-function relationships in the *E.coli* $tRNA^{Sec}$ we used the approach known as Instant Evolution (Lee et al., 1997; Zagryadskaya et al., 2003). In our case, it consisted in the *in vivo* screening of active *E. coli* $tRNA^{Sec}$ variants originated from several combinatorial gene libraries expressed in the bacterial strain devoid of the $tRNA^{Sec}$ (Zinoni et al., 1990). Due to the nature of this approach, the appearance of each single functioning variant can be considered as an isolated genetic event independent of other screened variants and of the will of the researcher. Such independency would allow us to further analyze the screened clones using statistical methods. In fact, as one can see below, the bioinformatic analysis of the obtained sequences comprised a significant part of the results.

In total, we have explored six combinatorial gene libraries (F1-F6), the designs of which are shown in Figure 1. The designs of all libraries except F1 contained additional nucleotide positions inexistent in WT (explained below). The colonies containing active variants of the $tRNA^{Sec}$ were identified by the bright white color attributed to the presence of the active selenoprotein formate dehydrogenase N (FDH_N). This protein uses formate as a source of electrons for reduction of nitrate. The absence of FDH_N, on the contrary, would result in accumulation of formate and decrease of pH. The level of pH can be detected by the pH indicator present in the growth medium. Depending on whether the pH is high or low, the indicator will make bacterial colonies white and

red, respectively (Barrett et al., 1979). In this study, all non-white colonies were ignored.

For some selected variants, the level of activity was measured using the assay for another selenoprotein formate dehydrogenase H (FDH_H), which is able to reduce benzyl viologen. By following the rate of the benzyl viologen reduction, we quantitated the level of FDH_H in the cell (Takahata et al., 2008). (see Materials and methods)

In both proteins FDH_N and FDH_H, whose activities were used for assessment of the performance of $tRNA^{Sec}$ variants, the selenocysteine is a key amino acid of the catalytic center. We thus assume that all $tRNA^{Sec}$ variants screened and measured in this research have been able to successfully pass through all intermediate steps required for the proper delivery of selenocysteine. Therefore, the activities of both selenoproteins directly reflect the functionality of the $tRNA^{Sec}$ variant.

Library F1- To assess whether the 5th and 6th base pairs of the D-stem are required for the $tRNA^{Sec}$ function, we designed a combinatorial library in which the four nucleotides composing these base pairs as well as the unpaired nucleotides between the acceptor and D-stems (nucleotides 8-9, Connector 1) and between the extra arm and the T-stem (nucleotide 48, Connector 2) were fully randomized (Library F1, Figure 1). The expression of this library provided 63 variants different from WT (Table S1). The number of the screened colonies exceeded the number of unique nucleotide sequences generated by this library (4^7=16384).

For some $tRNA^{Sec}$ variants from library F1, their steady state levels inside the cell were verified by northern blot (Figure 2). All tested variants showed about the same presence in the cell and about the same level of aminoacylation as WT (Figure 2). The function of all tested variants required SelB, as none of them provided white colonies when expressed in the bacterial strain WL81300 lacking SelB (data not shown).

Analysis of the selected F1-variants showed that the 5th and 6th base pairs were Watson-Crick (WC) in 28 (44%) and 14 (22%) variants, respectively. Only in three cases (5%), the 5th and 6th base pairs were simultaneously WC. The WC identities of the 5th and 6th base pairs varied between the clones; in different F1-clones, the 5th base pair had all WC identities except G14-C21, while the 6th base pair had all four possible identities.

If the GU dinucleotide combination was included in the definition of a base pair, the 5th and 6th base pairs would then be formed in 34 (54%) and 20 (32%) clones, respectively. Even with this expanded definition, only 11 F1-variants (17%) contained both base pairs, while 20 variants (32%) had none of them. Even though most variants missed one or two base pairs in the D-stem, there was no indication for a compensatory formation of the standard tertiary interactions 8-14 or 15-48. The identities of the nucleotides composing the connector regions varied randomly without any

obvious relationship to the identities of the corresponding base pairs.

In the selected F1-clones, the 5th base pair was formed notably more often than the 6th base pair. Moreover, a co-variation analysis showed that the number of clones in which the 5th base pair is WC (28 clones) is substantially higher than the number expected if both nucleotide positions varied independently (18.4 clones, all details are provided in Supplemental Extended Experimental Procedure). The same analysis performed for the 6th base pair showed that the number of clones in which this base pair is WC (14 clones) is only marginally higher than one would expect if both nucleotide positions varied independently (12.9 clones). Therefore, the formation of the 5th base pair seemed to be more important for tRNA^{Sec} function than the formation of the 6th base pair.

To determine whether the existence of one base pair could influence the formation of the other one, we divided all F1-clones into two groups; those variants with (group 1) and without (group 2) a WC dinucleotide combination at the 5th base pair. Analysis showed that the chance of having the 6th base pair WC was notably higher in group 2 than in group 1 (33% versus 17%). In other words, the probability for formation of the 6th base pair became twice as high in the absence of the 5th base pair compared to the situation when it existed.

In a similar way, when the 6th base pair did not form, the probability for the 5th base pair to be formed was higher than in the clones having the 6th base pair (51% versus 40%). Comparable results were obtained when the definition for a base pair was expanded to include the GU dinucleotide combination. These findings suggest the existence of a compensatory cross-talk between the 5th and 6th base pairs; if one of them does not exist, its absence has a tendency to be compensated by the presence of the other base pair. The existence of such a cross-talk is more evident when the 5th base pair is unpaired compared to the alternative situation when the 6th base pair is unpaired.

Libraries F2 and F3- We next modified the design of the F1 library by extending both connector regions by one (nucleotides 9a and 48a, Library F2, Figure 1B and C) and two (nucleotides 9a-9b and 48a-48b, library F3) randomized nucleotide positions. Library F2 resulted in the screening of 33 functional variants shown in Table S2. We noticed that during the screening of library F2, the frequency of appearance of positive clones was substantially lower than during the screening of library F1. Compared to F1-clones, the proportion of F2-clones having base pairs in both the 5th and 6th base pairs of the D-stem became notably higher. In particular, among all 33 F2-clones, the 5th and the 6th base pairs were WC in 30 (91%) and 23 (70%) clones, respectively. If the GU combination is also considered as a base pair, the corresponding numbers would increase to 31 (94%) and 26 (79%) clones, respectively (Figure 3). Again, as observed among F1-clones, the 5th base pair in F2-clones occurred more frequently than the 6th one, while the nucleotide sequences of the connector regions varied randomly without

any notable relation to the nucleotides corresponding to the 5th and 6th base pairs.

The increased tendency for the 5th and 6th base pairs to be formed when the connector regions were extended was even more pronounced among the seven selected F3-variants, as all of them had a WC or GU base pair in the 5th layer, while all variants except one had a WC or GU base pair in the 6th layer (Figure 3, Table S3). Again, the frequency of appearance of positive F3 clones was substantially lower than of positive F2 clones.

Libraries F4, F5 and F6- As discussed above, analysis of the nucleotide sequences of the obtained F1-clones suggested the existence of a cross-talk between the 5th and the 6th base pairs. When either the 5th or the 6th base pair did not form, the other base pair had a greater tendency to form as WC or GU. To further verify the existence of such relationship, we designed three more combinatorial libraries F4-F6. In each of these libraries we took the design of library F2 as a prototype into which we forcefully introduced a purine-purine or pyrimidine-pyrimidine mismatch at the place of either the 5th or the 6th base pair of the D-stem (Figure 1C). Analysis of the screened variants showed that, compared to F2-clones, the clones screened from libraries F4-F6 were characterized by a very limited set of dinucleotide combinations in the other base pair, whose nucleotide identities were not limited during the experiment. Thus, the introduction of a pyrimidine-pyrimidine combination at the 5th base pair (library F5, Table S4) made the 6th base pair predominantly C15-G20. The exceptions happened in 7 out of the 27 collected variants and consisted of base pairs U15-A20 (one clone), G15-C20 (three clones) and mismatch G15-G20 (three clones). Also, the introduction of a purine-purine combination at the same place (library F4, Table S5) made the 5th and 6th base pairs predominantly A14-G21 and C15-G20, with exceptions happening in three out of the 11 obtained clones. Finally, the introduction of a pyrimidine-pyrimidine combination at the 6th base pair (library F6, Table S6) made the 5th base pair almost exclusively C14-G21 or A14-G21, and exceptions happened only in two out of 17 obtained variants. These data additionally support the existence of a cross-talk between the 5th and 6th base pairs of the D-stem, so that a particular type of mismatch in one base pair strongly limits the set of acceptable dinucleotide combinations in the other one.

The extended D-stem as a factor for the $tRNA^{Sec}$ stability- The results obtained so far clearly demonstrate that even though the 5th and 6th WC base pairs are uniquely present in the $tRNA^{Sec}$, neither of them plays a specific role in the $tRNA^{Sec}$ function. In different variants of the $tRNA^{Sec}$, each of the two base pairs assumes all possible WC identities AU, UA, GC and CG. Moreover, both base pairs can have extended identities GU and UG or contain mismatches. The absence of distinct requirements imposed on the structure of these base pairs makes improbable their involvement in specific interactions with factors assisting the $tRNA^{Sec}$ functioning.

Although neither the 5th nor the 6th base pair in the D-stem was required for the $tRNA^{Sec}$ function, both of them played a positive role: the presence of the 5th base pair strongly increased the chance for a $tRNA^{Sec}$ variant to be selected in our procedure, and in the absence of the 5th base pair, a similar effect was observed for the 6th base pair. Also, additional perturbing elements introduced into the nucleotide sequence of the $tRNA^{Sec}$ strongly favored the selection of variants containing these base pairs. Such perturbing elements included one (library F2) or two (library F3) extra nucleotides in the connector regions, as well as a mismatch at the place of either the 5th (libraries F4 and F5) or the 6th base pair (library F6) of the D-stem. Introduction of each of these elements sharply increased the chance for selection of clones with a D-stem containing more than four base pairs. The results of the screening of libraries F4-F6 also confirm the initial observation made based on analysis of the F1-library that the two base pairs are involved in a cross-talk between themselves. The absence of either the 5th or the 6th base pair notably increased the chance for the other one to be formed.

Although the designs of libraries F2-F6 were different, they all shared the same feature consisting in creation of obstacles for formation of the normal secondary structure. The obstacles consisted in the extension of the connector regions and/or in the introduction of mismatches into the D-stem. The introduction of mismatches into the D-stem would destabilize the standard secondary structure, while the extension of the connector regions by randomized nucleotides would increase the probability for alternative secondary structures. This general conclusion was also checked by folding simulations using program *mfold* (Zuker, 2003). These simulations demonstrated that indeed, both the extension of the connector regions and the introduction of mismatches into the D-stem notably decreased the probability for formation of the proper cloverleaf secondary structure (not shown). Therefore, the observed elevated level of base pairings at the end of the D-stem compared to that observed among F1-clones (Figure 3) should be seen as a compensatory effect able to restore the stability of the variants. This, in turn, leads us to the general conclusion that the major role of the additional 5th and 6th base pairs in the D-stem of the $tRNA^{Sec}$ consists in a non-specific stabilization of the D-stem and through this, of the whole $tRNA^{Sec}$ secondary structure.

Thermal stability of $tRNA^{Sec}$ variants- To further verify the stabilizing role played by the additional base pairs in the D-stem, we measured FDH_H activities for some of the variants screened from library F1 which represented the spectrum of base pairing strength found within the D-stem. In total, we studied twelve molecules, including WT. Among them, the 5th base pair had identities UA (WT + three variants), as well as CG, UG, CA and AG (two variants for each combination). The identities of the 6th base pair were CG (WT + 1 variant), UA (1), AU (2), AA (5), UU (1) and GA (1). For all variants analyzed, the activities were measured at 30˚C and at 37˚C (Table 1), while for

WT and for variant F1-24, (in F1-24, the 5th and the 6th base pairs were CA and UA, respectively), the activities were also measured at 33˚C and at 42˚C (Figure 4).

As anticipated, at 37˚C WT was the most active among all measured variants. Interestingly, when measured at 30˚C, WT performed better than at 37˚C. This fact, however, was expected and essentially repeated the result of a previous study (Takahata et al., 2008). All other variants followed the same pattern, having a higher activity at 30˚C than at 37˚C. Unexpectedly, a few variants performed at 30˚C significantly better than WT. In particular, clone F1-24 at 30˚C demonstrated higher than a two-fold (243%) activity than WT. Three more variants F1-4, F1-51 and F1-11 had activities that ranged from 158% to 184% of WT. When the temperature was raised from 30˚C to 37˚C, all these clones experienced a substantial drop in activity to become no better than other clones. The same tendency continued to 42˚C with a further substantial drop in activity, as shown for clone F1-24 (Figure 4). For the other clones that did not have such robust activity at 30°C, the drop in the activity associated with the increase in temperature, was substantially more modest than for the first-mentioned clones F1-24, F1-4, F1-51 and F1-11.

The difference in the behavior of the measured clones did not relate to their presence in the cytosol. At 30˚C clone F1-24 performed better than WT, even though its level in the cytosol was no higher than of the latter (compare lanes 5 and 6, to lanes 7 and 8 in Figure 2). Conversely, at 37˚C WT performed better than F1-24 despite having an equal to or even slightly lower presence in the cell (compare lanes 11 and 12, to 13 and 14 of Figure 2). This observation allows us to make a general suggestion that the differences in the activities of the measured clones are somehow linked to particular features of their structure.

To reveal such linkage, we assigned for each variant two parameters. The first parameter (parameter A, Table 1) represented the activity of the variant at 30˚C normalized by the corresponding activity of WT. Given that all clones demonstrated the highest activity at 30˚C, parameter A would thus represent the maximal activity measured for each clone. The second parameter (parameter B, Table 1) was calculated as the ratio between the activities at 30˚C and 37˚C and thus represented the loss in activity caused by the increase in temperature.

On the plot of B versus A (Figure 5), all measured clones form three distinct groups. Group 1 is composed of six clones, including WT, that had a low to average activity at 30˚C, and a small decline in activity when temperature was elevated (low values of both A and B). Group 2 consists of four above-mentioned clones F1-24, F1-4, F1-51 and F1-11 characterized by a high maximal activity and a significant decrease in activity caused by the increase in the temperature (high values of both A and B). Finally, group 3 included two clones F1-28 and F1-54, with a low value of A and a high value of B. In other words, clones F1-28 and F1-54 had at 30˚C activities similar to that of WT, which however dropped as sharply as

in clones of group 2 when the temperature was raised from 30˚C to 37˚C.

Analysis of the nucleotide sequences of the measured clones revealed a clear link between the position of a clone on the A/B plot and the identity of the 5th base pair in the D-stem. In all clones of group 1, this base pair was WC, being either UA (3 clones + WT) or CG (2 clones). In group 2, this base pair does not exist, being replaced by either AG (2) or CA (2) combination. Finally, in both variants of group 3, the 5th base pair is UG. Although each variant has five more variable nucleotides, which, in principle, could affect their level of activity, we do not think that the role of these nucleotides is essential. Between the clones, the identities of these five nucleotides vary widely and without any visible relation to the function of the clones. This makes the identity of the 5th base pair of the D-stem the major if not the only factor that determines the differences in the functional behavior of the measured clones.

How exactly do the identities in the 5th base pair of the D-stem determine the functionality of the tRNA^{Sec} variants? The presence of a WC base pair at this position (group 1) is linked to a relatively low level of activity loss when the temperature increases from 30˚C to 37˚C. In other words, if the 5th base pair is WC, it provides the tRNA with a higher resistance to heat (a low value of parameter B in group 1). Conversely, the absence of a WC base pair is associated with a substantial decline in activity when the temperature was increased from 30˚C to 37˚C, or with a lower resistance to heat (high value of B in groups 2 and 3). Therefore, a WC base pair in the 5th position of the D-stem appears to be the major factor for the tRNA stability when the temperature rose from 30˚C to 37˚C.

Surprisingly, at 30˚C the presence of either a WC or UG base pair in the 5th position of the D-stem, in spite of its stabilizing effect, makes a tRNA^{Sec} variant substantially less active (low values of A in groups 1 and 3, Figure 5) compared to the variants that do not have such a base pair (group 2, Figure 5). This means that at 30˚C, the additional stabilization of the tRNA structure caused by the presence of a WC or UG base pair is harmful for the tRNA^{Sec} function.

Thus, at lower temperatures, a higher activity of the tRNA^{Sec} is associated with the additional conformational flexibility in the D-stem, which can be achieved through the absence of the 5th base pair. At higher temperatures, the stability of the tRNA^{Sec} becomes critical, which thus requires that this base pair be formed. The UG base pair behaves in a peculiar way. At 30˚C, it behaves like a WC base pair, making the tRNA less effective (low level of parameter A). However, when the temperature is elevated, the UG base pair behaves as if it is inexistent (high level of parameter B). Such a dual behavior is consistent with the known mediocre stability of the UG base pair.

DISCUSSION

In this study, we analyzed the role of the extended D-stem in the function of the *E. coli* tRNA^{Sec}. For this purpose, we used the Instant

Evolution approach consisting in *in vivo* screening of a large number of functional variants originated from several combinatorial gene libraries. By focusing on positive variants, this approach is able to overcome the general problem, usually associated with *in vivo* studies, of attributing a poor functioning of a given variant to the particular aspect of the molecule's synthesis, maturation or functioning. As we screened only highly efficient clones, all of them should have had no major problems at any of these steps. Also, due to the inherent combinatorial nature of the approach, it allowed further statistical analysis of the nucleotide sequences of the screened clones.

The results presented in this paper clearly show that the 5th and 6th base pairs of the D-stem, which are uniquely present in the tRNA^{Sec}, do not play any specific functional role and are not required for formation of any essential interaction of this molecule with other elements of the Sec-incorporating machinery in *E. coli*. A tRNA can be functional even if neither of the two base pairs is formed. Moreover, the formation of the 5th base pair as WC or UG can even hamper the functionality of the tRNA^{Sec}, as it was demonstrated by the lower activity at 30˚C of the variants containing these base pairs compared to variants having a shorter D-stem. Only at higher temperatures the presence of this base pair becomes essential for the tRNA^{Sec} function.

The presented data allow us to conclude that the major role of the additional base pairs in the D-stem of the tRNA^{Sec} consists in providing for sufficient stability of the tRNA structure. This conclusion is based on three sets of experiments. First, analysis of the screened variants of the tRNA^{Sec} revealed the existence of a cross-talk between the 5th and 6th base pairs in the D-stem, so that the absence of one of them has a tendency to be compensated by the presence of the other one. Second, each time when the design of a combinatorial library contained additional obstacles for formation of the proper secondary structure, the percentage of the screened variants having these base pairs notably increased. Finally, the measurement of the activity of several tRNA^{Sec} variants demonstrated that the 5th base pair in the D-stem is required only at an elevated temperature, while at a lower temperature its formation can be harmful for the tRNA^{Sec} function.

A reasonable question would concern the reasons of why the stability of the tRNA^{Sec} requires the presence of the base pairs that do not exist in any other tRNA. We think that the requirement for these base pairs relates to the fact that while all other tRNAs contain universal tertiary interactions 8-14 and 15-48, in the tRNA^{Sec} these interactions do not exist. The formation of the additional base pairs in the D-stem of the tRNA^{Sec} should thus be considered as a compensatory measure that would stabilize the tRNA structure in the absence of the standard tertiary interactions.

If the major aspect limiting functionality of tRNA^{Sec} variants pertains to their stability, can the latter be restored through formation of the normal tertiary interactions existing in other

tRNAs? Inspection of the selected clones shows that out of the 63 F1-clones, only in three, F1-14, F1-36 and F1-1 the identities of nucleotides 8, 14, 15 and 48 allow the formation of the two canonical tertiary base pairs 8-14 and 15-48. However, in two clones F1-14 and F1-1 the formation of base pair U8-A14 has to compete with the 5th base pair of the D-stem A14-U21. Whether in these exceptional cases the tertiary interactions 8-14 and 15-48 indeed form is unknown. It is clear, however, that out of the two options of structural stabilization, either through extension of the D-stem or through formation of tertiary interactions 8-14 and 15-48, F1-clones demonstrate a strong preference toward the first one. Such preference is even more pronounced among clones selected from libraries F2 and F3. In the latter clones, the stabilization of the tRNA^{Sec} structure always proceeds through the extension of the D-stem, while the identities of the nucleotides of the connector regions do not provide even a theoretical opportunity for formation of the tertiary base pairs.

Based on the fact that between the two options for stabilization of the tRNA structure either through restoration of the long D-stem or through formation of tertiary interactions 8-14 and 15-48 the selected variants systematically choose the first one, we suggest that the tertiary interactions are harmful for the tRNA^{Sec} function. The absence of these interactions makes the tRNA^{Sec} conformation more flexible, compared to that of other cytosolic tRNAs. The existence of such flexibility has already been noticed when the available X-ray conformations of the tRNA^{Sec} were compared with each other (Chiba et al., 2010; Itoh et al., 2009).

The flexibility provided by the absence of standard tertiary interactions seems to be essential for the tRNA^{Sec} function. Moreover, the efficiency of WT can even be improved, if one introduced additional flexibility to its structure by disrupting the last base pairs of the D-stem. Such strategy, however, works only at low temperatures, when the modified tRNA is still able to maintain its secondary structure. At higher temperatures, when the stability of the whole tRNA^{Sec} structure diminishes, the additional base pairs in the D-stem become essential. This result shows that as long as the integrity of the tRNA^{Sec} secondary structure is not compromised, a higher conformational flexibility would correspond to a higher selenocysteine-incorporating activity. For the *E. coli* tRNA^{Sec}, normally functioning at 37˚C, the formation of the 5th and 6th base pairs in the D-stem seems to be the best compromise between flexibility and integrity.

The presented data strongly suggest that the observed conformational flexibility of the tRNA^{Sec} is required for some unique aspects of the tRNA^{Sec} function that this molecule does not share with other tRNAs. In particular, it may be linked to another unusual feature of the tRNA^{Sec}, the long acceptor stem. In the normal tRNAs, the standard tertiary interactions 8-14 and 15-48 make the acceptor/T helical domain rigidly attached to the D/anticodon domain. If the tRNA^{Sec} contained the standard tertiary interactions and thus were as

rigid as other tRNAs, the unusually long acceptor stem would have juxtaposed the CCA-terminus and the anticodon loop differently compared to other tRNAs. This, in turn, would create problems for the proper accommodation of the $tRNA^{Sec}$ at the standard ribosomal sites. However, due to the absence of the standard tertiary interactions in the D stem-loop, the juxtaposition of the two functional centers of the tRNA becomes rather flexible. Such flexibility thus would be helpful for the simultaneous accommodation of the anticodon loop and of the CCA-terminus at the proper places on the ribosome surface.

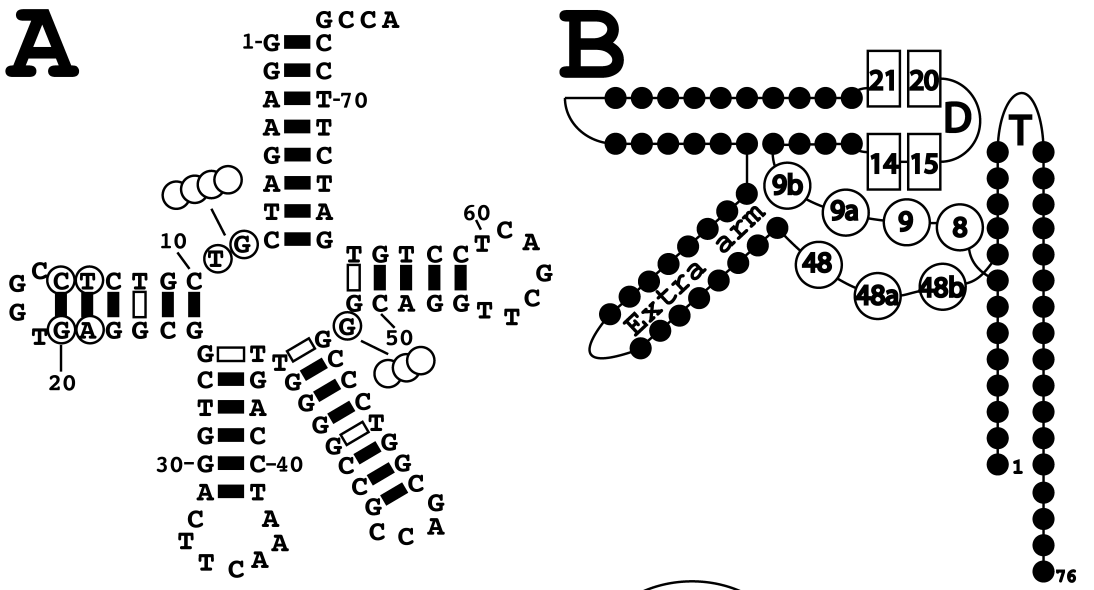
At the same time, the flexible characteristic of the $tRNA^{Sec}$ may also be related to its unusual functional pattern consisting in the theft of the UGA stop codon in response to the SECIS (Heider et al., 1992), the downstream secondary structure element essential for selenocysteine incorporation. How exactly the absence of the tertiary interactions in the $tRNA^{Sec}$ relates to its function will require further analysis.

REFERENCES

1. Yoshizawa, S., and Böck, A. (2009) The many levels of control on bacterial selenoprotein synthesis. *Biochim Biophys Acta* **1790**, 1404-1414
2. Leinfelder, W., Zehelein, E., Mandrand-Berthelot, M. A., and Böck, A. (1988) Gene for a novel tRNA species that accepts L-serine and cotranslationally inserts selenocysteine. *Nature* **331**, 723-725
3. Sturchler, C., Westhof, E., Carbon, P., and Krol, A. (1993) Unique secondary and tertiary structural features of the eucaryotic selenocysteine tRNA(Sec). *Nucleic Acids Res* **21**, 1073-1079
4. Palioura, S., Sherrer, R. L., Steitz, T. A., Söll, D., and Simonovic, M. (2009) The human SepSecS-tRNASec complex reveals the mechanism of selenocysteine formation. *Science* **325**, 321-325
5. Itoh, Y., Chiba, S., Sekine, S., and Yokoyama, S. (2009) Crystal structure of human selenocysteine tRNA. *Nucleic Acids Res* **37**, 6259-6268
6. Burkard, U., and Söll, D. (1988) The unusually long amino acid acceptor stem of Escherichia coli selenocysteine tRNA results from abnormal cleavage by RNase P. *Nucleic Acids Res* **16**, 11617-11624
7. Wu, X. Q., and Gross, H. J. (1994) The length and the secondary structure of the D-stem of human selenocysteine tRNA are the major identity determinants for serine phosphorylation. *EMBO J* **13**, 241-248
8. Chiba, S., Itoh, Y., Sekine, S., and Yokoyama, S. (2010) Structural basis for the major role of O-phosphoseryl-tRNA kinase in the UGA-specific encoding of selenocysteine. *Mol Cell* **39**, 410-420
9. Turanov, A. A., Xu, X. M., Carlson, B. A., Yoo, M. H., Gladyshev, V. N., and Hatfield, D. L. Biosynthesis of selenocysteine, the 21st amino acid in the genetic code, and a novel pathway for cysteine biosynthesis. *Adv Nutr* **2**, 122-128
10. Carlson, B. A., Xu, X. M., Kryukov, G. V., Rao, M., Berry, M. J., Gladyshev, V. N., and Hatfield, D. L. (2004) Identification and characterization of phosphoseryl-tRNA[Ser]Sec kinase. *Proc Natl Acad Sci U S A* **101**, 12848-12853

11. Zinoni, F., Heider, J., and Böck, A. (1990) Features of the formate dehydrogenase mRNA necessary for decoding of the UGA codon as selenocysteine. *Proc Natl Acad Sci U S A* **87**, 4660-4664
12. Tormay, P., Sawers, A., and Böck, A. (1996) Role of stoichiometry between mRNA, translation factor SelB and selenocysteyl-tRNA in selenoprotein synthesis. *Mol Microbiol* **21**, 1253-1259
13. Zagryadskaya, E. I., Doyon, F. R., and Steinberg, S. V. (2003) Importance of the reverse Hoogsteen base pair 54-58 for tRNA function. *Nucleic Acids Res* **31**, 3946-3953
14. Barrett, E. L., Jackson, C. E., Fukumoto, H. T., and Chang, G. W. (1979) Formate dehydrogenase mutants of Salmonella typhimurium: a new medium for their isolation and new mutant classes. *Mol Gen Genet* **177**, 95-101
15. Takahata, M., Tamura, T., Abe, K., Mihara, H., Kurokawa, S., Yamamoto, Y., Nakano, R., Esaki, N., and Inagaki, K. (2008) Selenite assimilation into formate dehydrogenase H depends on thioredoxin reductase in Escherichia coli. *J Biochem* **143**, 467-473
16. Kotlova, N., Ishii, T. M., Zagryadskaya, E. I., and Steinberg, S. V. (2007) Active suppressor tRNAs with a double helix between the D- and T-loops. *J Mol Biol* **373**, 462-475
17. Lee, K., Varma, S., SantaLucia, J., Jr., and Cunningham, P. R. (1997) In vivo determination of RNA structure-function relationships: analysis of the 790 loop in ribosomal RNA. *J Mol Biol* **269**, 732-743
18. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406-3415
19. Heider, J., Baron, C., and Böck, A. (1992) Coding from a distance: dissection of the mRNA determinants required for the incorporation of selenocysteine into protein. *EMBO J* **11**, 3759-3766
20. Jühling, F., Mörl, M., Hartmann, R. K., Sprinzl, M., Stadler, P. F., and Pütz, J. (2009) tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Res* **37**, D159-162

Acknowledgements- The authors thank Drs. Gary Sawers and August Böck for the gift of bacterial strains WL81460 and WL81300, Drs. Christian Baron and Patrick Hallenbeck for discussions and valuable advice, Damien Biot-Pelletier for technical assistance in the experiments.



C

	CONNECTOR I								CONNECTOR II		
	8	9	9a	9b	14	15	20	21	48	48a	48b
Library F1	N	N	-	-	N	N	N	N	N	-	-
Library F2	N	N	N	-	N	N	N	N	N	N	-
Library F3	N	N	N	N	N	N	N	N	N	N	N
Library F4	N	N	N	-	R	N	N	R	N	N	-
Library F5	N	N	N	-	Y	N	N	Y	N	N	-
Library F6	N	N	N	-	N	Y	Y	N	N	N	-

FIGURE 1. The secondary structure of the *E. coli* $tRNA^{Sec}$ and the designs of the combinatorial gene libraries expressed in this paper. The secondary structure is shown as the cloverleaf (A) and as the L-form (B). In the cloverleaf, WC and GU base pairs are indicated by black and open rectangles, respectively. In the cloverleaf diagram, the randomized nucleotide positions 8, 9, 14, 15, 20, 21, and 48 are circled, while the inserted nucleotides 9a, 9b, 48a, and 48b are indicated by overlapping circles. In the L-form, black circles stand for nucleotides of double helical regions. The positions of the D- and T-loops and of the extra arm are indicated. Nucleotides are numbered as in the tRNA database (Juhling et al., 2009). All nucleotides that have been variable in our combinatorial libraries are shown by open circles (in the connector regions) or open rectangles (at the end of the D-stem). Nucleotides 9a, 9b, 48a and 48b are present only in some of the libraries (see panel C). C: the designs of each of the six combinatorial gene libraries discussed in the paper. For all libraries, N, R and Y stand for a fully randomized nucleotide (N), for variation between A and G (R) and between C and U (Y). Dashes stand for nucleotide positions not present in the library.

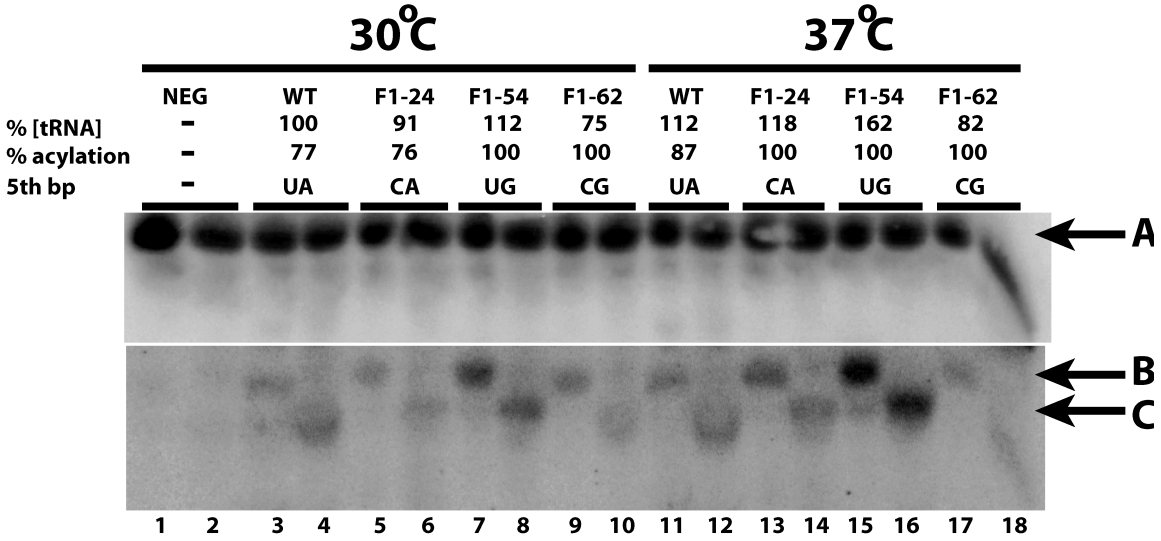


FIGURE 2. The steady state and aminoacylation levels of the WT $tRNA^{Sec}$ and of variants F1-24, F1-54, and F1-62 at 30˚C (lanes 3-10) and at 37˚C (lanes 11-18). The negative control containing the void vector is shown in lanes 1 and 2. Northern blot hybridization was performed using a specific DNA probe complementary to the extra-arm of the $tRNA^{Sec}$. For each variant checked, the presence in the cell at 30˚C and at 37˚C is comparable. The samples in the even lanes were deacylated by incubation with Tris pH 9.0, while the samples in the odd-lanes were not. The fact that the aminoacylated form migrated slower than the deacylated form allowed evaluation of the level of aminoacylation. Each variant existed in the cell predominantly in the aminoacylated form. An additional probe specific to 5S rRNA was used to monitor the amount of total RNA in the samples (indicated by arrow A). For each sample, the identity of the 5^{th} base pair of the D-stem (base pair 14-21) is provided. The upper part of the figure showing the 5S rRNA was taken from a short exposure of the northern blot, while the lower part showing $tRNA^{Sec}$ variants was taken from a longer exposure (the original blots with the short and long exposure are shown in the supplemental figure S1). The relative amounts of tRNA are shown for each variant as a percentage of the WT $tRNA^{Sec}$ at 30˚C (level C, lane 4) which was taken to be 100%. The aminoacylation levels for each variant was calculated by dividing the aminoacylated band density (level B, odd lane) by the total tRNA density (level C, even lane). All density calculations were normalized to the amount of 5S (level A).

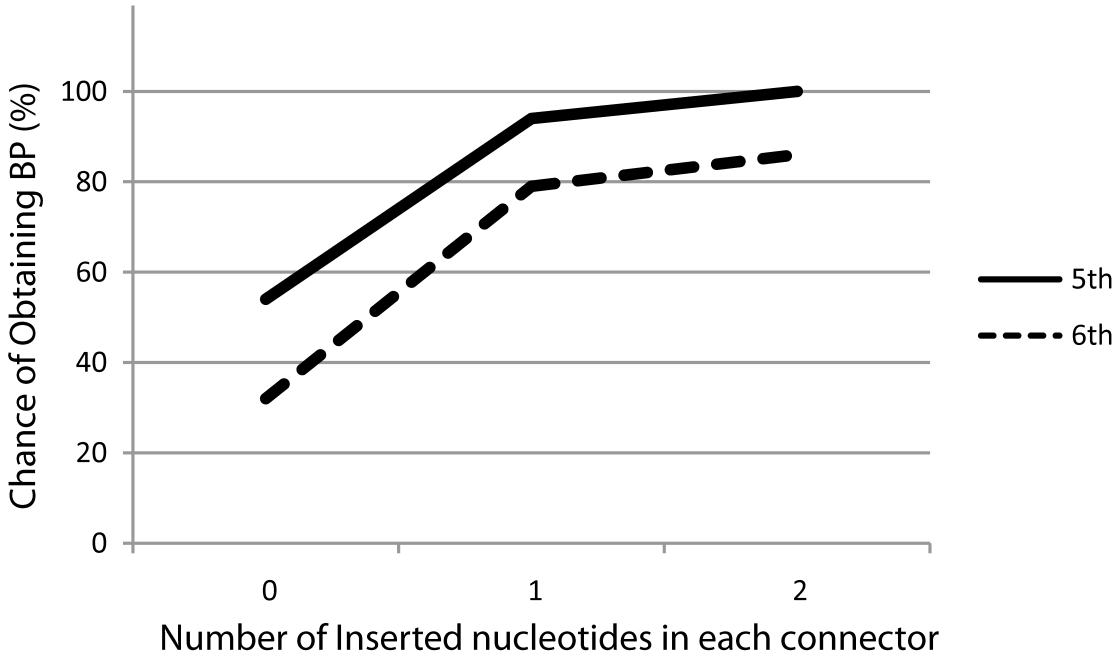


FIGURE 3. The proportion of variants having the 5th (solid line) and 6th (dotted line) base pair WC or GU depending on the number of additional nucleotides incorporated into each of the two connector regions (0, 1 and 2 nucleotides for libraries F1, F2 and F3, respectively). For both base pairs, the probability to be formed notably increased when the connector regions were extended. In all screenings, the 5th base pair formed more often than the 6th base pair. The plot demonstrates that the relative instability of the secondary structure caused by the extension of the connector regions is generally compensated by a more stable D-stem.

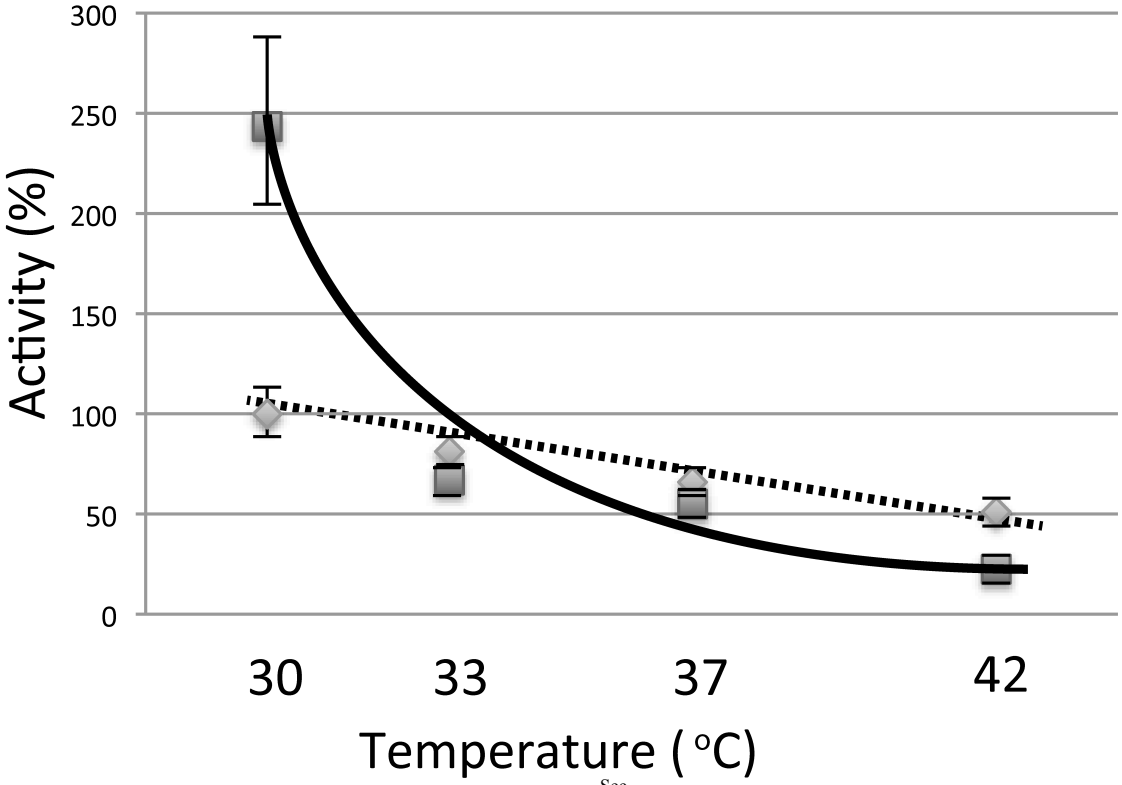


FIGURE 4. The FDH_H activities of the WT $tRNA^{Sec}$ (diamond, dotted line) and of variant F1-24 (square, solid line) at different temperatures. The activity of the WT $tRNA^{Sec}$ at 30˚C was taken as 100%. Compared to WT, the activity of variant F1-24 at 30˚C was notably higher, but at higher temperatures declined significantly faster. Each plot represents the activity averaged over three independent measurements. The standard deviations are indicated.

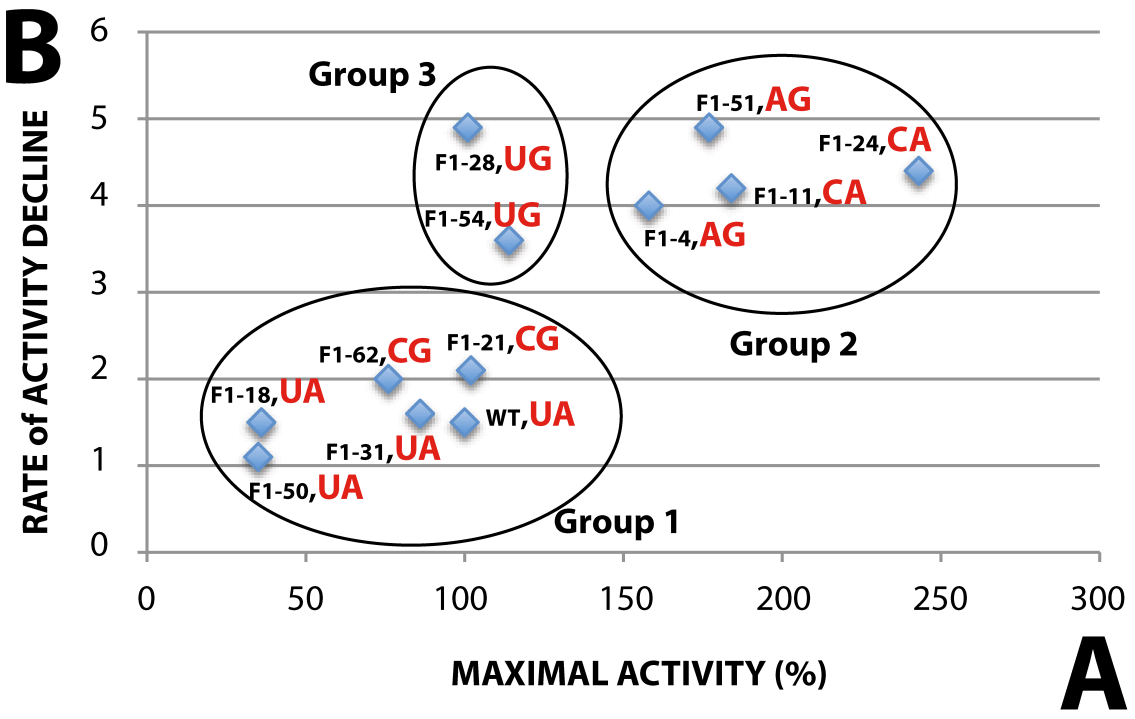


FIGURE 5. The positions of twelve selected F1 variants on the A/B plot. Parameter A represents the activity of a variant at 30˚C normalized by the activity of the WT $tRNA^{Sec}$. Parameter B characterizes the rate of the FDH_H activity decline of a variant when the temperature increases from 30˚C to 37˚C (see the text). On the plot, the positions of the variants become grouped depending on the identity of the 5^{th} base pair. Each point is marked by the clone number (black) followed by the identity of the fifth base pair in the D-stem (red).

Clone	30°C PARAMETER A	37°C	PARAMETER B	5th (nts 14-21)	6th (nts 15-20)	1st (nts 8-9)	2nd (nt48)
F1-50	35%	31%	1.1	UA	UU	CU	A
F1-18	36%	23%	1.5	UA	AA	CA	A
F1-62	76%	38%	2.0	CG	CG	UC	G
F1-31	86%	53%	1.6	UA	UA	AU	C
WT	100%	66%	1.5	UA	CG	GU	G
F1-21	102%	49%	2.1	CG	AA	CU	G
F1-28	101%	21%	4.9	UG	AA	UA	A
F1-54	114%	31%	3.6	UG	AU	UA	G
F1-4	158%	40%	4.0	AG	GA	GC	U
F1-51	177%	36%	4.9	AG	AA	AG	U
F1-11	184%	44%	4.2	CA	AA	CA	A
F1-24	243%	55%	4.4	CA	AU	GA	U

TABLE 1. Relative FDH_H activities of selected variants performed at 30°C and 37°C.
The FDH_H activities of selected variants performed at 30°C (parameter A) and 37°C are shown with the WT activity at 30°C being 100%. Paramater B is the ratio between the activity at 30°C divided by the activity of 37°C. The corresponding nucleotide identities for the 5th base pair, 6th base pair, 1st Connector, and the 2nd Connector is shown. In the 5th base pair column, variants are boxed based on the dinucleotide combinations that form either WC, GU, or non-WC base pairs.

Supplemental Data:

Table S1: List of Library F1 variants

	ACCEPTOR-STEM	C1	D-STEM		D-STEM	ANTICODON-STEM		ANTICODON-STEM	EXTRA-ARM		EXTRA-ARM	C2	T-STEM		T-STEM	ACCEPTOR-STEM	.
WT	GGAAGATC	GT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
LibF1	GGAAGATC	NN	CGTCNN	CGGT	NNGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	N	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-1	GGAAGATC	TA	CGTC**A**A	CGGT	A**T**GGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	T	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-2	GGAAGATC	AG	CGTCT**G**	CGGT	**C**GGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	A	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-3	GGAAGATC	CT	CGTC**C**A	CGGT	G**G**GGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGGCCG	CCAG	CGGTCCCG	G	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-4	GGAAGATC	GC	CGTCAG	CGGT	AGGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	T	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-5	GGAAGATC	AG	CGTCAG	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	A	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-6	GGAAGATC	TA	CGTCTA	CGGT	AGGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-7	GGAAGATC	CC	CGTCAA	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	T	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-8	GGAAGATC	TA	CGTC**C**G	CGGT	G**G**GGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	A	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-9	GGAAGATC	TG	CGTCAA	CGGT	AGGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-10	GGAAGATC	TT	CGTCAT	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-11	GGAAGATC	CA	CGTCCA	CGGT	AAGGCG	ACTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	A	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-12	GGAAGATC	CT	CGTCA**A**	CGGT	**T**AGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-13	GGAAGATC	AA	CGTC**A**G	CGGT	G**T**GGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	T	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-14	GGAAGATC	TC	CGTC**A**G	CGGT	T**T**GGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-15	GGAAGATC	CT	CGTCAA	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	T	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-16	GGAAGATC	TG	CGTCGA	CGGT	AGGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-17	GGAAGATC	GC	CGTC**A**G	CGGT	G**T**GGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-18	GGAAGATC	CA	CGTC**T**A	CGGT	A**A**GGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	A	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-19	GGAAGATC	CT	CGTC**T**G	CGGT	T**A**GGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	A	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-20	GGAAGATC	TA	CGTC**T**A	CGGT	G**A**GGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	A	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-21	GGAAGATC	CT	CGTC**C**A	CGGT	A**G**GGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-22	GGAAGATC	TT	CGTCAA	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	A	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-23	GGAAGATC	AT	CGTCA**A**	CGGT	**T**AGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	T	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-24	GGAAGATC	GA	CGTCC**A**	CGGT	**T**AGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	T	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-25	GGAAGATC	AG	CGTCAA	CGGT	AAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	T	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-26	GGAAGATC	AC	CGTC**C**A	CGGT	G**G**GGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	C	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-27	GGAAGATC	GT	CGTCC**A**	CGGT	**T**TGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-28	GGAAGATC	TA	CGTCTA	CGGT	AGGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	A	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-29	GGAAGATC	TT	CGTC**C**A	CGGT	G**G**GGCG	GCTGGA	CTTCAAA	TCCAGT	TGGG CCG	CCAG	CGGTCCCG	A	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-30	GGAAGATC	CG	CGTC**T**G	CGGT	T**A**GGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-31	GGAAGATC	AT	CGTC**TT**	CGGT	**AA**GGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	C	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-32	GGAAGATC	TA	CGTC**T**G	CGGT	A**A**GGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	T	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-33	GGAAGATC	TA	CGTC**T**A	CGGT	C**A**GGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	C	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-34	GGAAGATC	GA	CGTCAA	CGGT	AAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	T	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-35	GGAAGATC	TG	CGTC**C**A	CGGT	A**G**GGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F1-36	GGAAGATC	TA	CGTCAA	CGGT	GGGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	T	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA

```
F1-37   GGAAGATC AA  CGTCTA  CGGT TTGGCG  GCTGGA CTTCAAA TCCAGT    TGGGGCCG  CCAG CGGTCCCG G  GCAGG TTCGACT CCTGT GATCTTCC GCCA
F1-38   GGAAGATC CT  CGTCAA  CGGT AGGGCG  GCTGGA CTTCAAA TCCAGT    TGGGGCCG  CCAG CGGTCCCG A  GCAGG TTCGACT CCTGT GATCTTCC GCCA
F1-39   GGAAGATC CT  CGTCAA  CGGT AGGGCG  GCTGGA CTTCAAA TCCAGT    TGGGGCCG  CCAG CGGTCCCG T  GCAGG TTCGACT CCTGT GATCTTCC GCCA
F1-40   GGAAGATC TT  CGTCAG  CGGT TAGGCG  GCTGGA CTTCAAA TCCAGT    TGGGGCCG  CCAG CGGTCCCG G  GCAGG TTCGACT CCTGT GATCTTCC GCCA
F1-41   GGAAGATC TC  CGTCCA  CGGT AGGGCG  GCTGGA CTTCAAA TCCAGT    TGGGGCCG  CCAG CGGTCCCG C  GCAGG TTCGACT CCTGT GATCTTCC GCCA
F1-42   GGAAGATC GC  CGTCCG  CGGT GAGGCG  GCTGGA CTTCAAA TCCAGT    TGGGGCCG  CCAG CGGTCCCG A  GCAGG TTCGACT CCTGT GATCTTCC GCCA
F1-43   GGAAGATC TT  CGTCTG  CGGT TAGGCG  GCTGGA CTTCAAA TCCAGT    TGGGGCTG  CCAG CGGTCCCG T  GCAGG TTCGACT CCTGT GATCTTCC GCCA
F1-44   GGAAGATC TA  CGTCTA  CGGT AAGGCG  GCTGGA CTTCAAA TCCAGT    TGGGGCCG  CCAG CGGTCCCG G  GCAGG TTCGACT CCTGT GATCTTCC GCCA
F1-45   GGAAGATC TT  CGTCAA  CGGT GGGGCG  GCTGGA CTTCAAA TCCAGT    TGGGGCCG  CCAG CGGTCCCG A  GCAGG TTCGACT CCTGT GATCTTCC GCCA
F1-46   GGAAGATC CA  CGTCAA  CGGT AAGGCG  GCTGGA CTTCAAA TCCAGT    TGGGGCCG  CCAG CGGTCCCG A  GCAGG TTCGACT CCTGT GATCTTCC GCCA
F1-47   GGAAGATC CT  CGTCAA  CGGT TAGGCG  GCTGGA CTTCAAA TCCAGT    TGGGGCCG  CCAG CGGTCCCG T  GCAGG TTCGACT CCTGT GATCTTCC GCCA
F1-48   GGAAGATC TC  CGTCTA  CGGT CAGGCG  GCTGGA CTTCAAA TCCAGT    TGGGGCCG  CCAG CGGTCCCG T  GCAGG TTCGACT CCTGT GATCTTCC GCCA
F1-49   GGAAGATC TA  CGTCTC  CGGT AAGGCG  GCTGGA CTTCAAA TCCAGT    TGGGGCCG  CCAG CGGTCCCG G  GCAGG TTCGACT CCTGT GATCTTCC GCCA
F1-50   GGAAGATC CT  CGTCTT  CGGT TAGGCG  GCTGGA CTTCAAA TCCAGT    TGGGGCCG  CCAG CGGTCCCG A  GCAGG TTCGACT CCTGT GATCTTCC GCCA
F1-51   GGAAGATC AG  CGTCAA  CGGT AGGGCG  GCTGGA CTTCAAA TCCAGT    TGGGGCCG  CCAG CGGTCCCG T  GCAGG TTCGACT CCTGT GATCTTCC GCCA
F1-52   GGAAGATC TA  CGTCAA  CGGT AGGGCG  GCTGGA CTTCAAA TCCAGT    TGGGGCCG  CCAG CGGTCCCG G  GCAGG TTCGACT CCTGT GATCTTCC GCCA
F1-53   GGAAGATC TT  CGTCCA  CGGT GGGGCG  GCTGGA CTTCAAA TCCAGT    TGGGGCCG  CCAG CGGTCCCG C  GCAGG TTCGACT CCTGT GATCTTCC GCCA
F1-54   GGAAGATC TA  CGTCTA  CGGT TGGGCG  GCTGGA CTTCAAA TCCAGT    TGGGGCCG  CCAG CGGTCCCG G  GCAGG TTCGACT CCTGT GATCTTCC GCCA
F1-55   GGAAGATC AT  CGTCTA  CGGT AAGGCG  GCTGGA CTTCAAA TCCAGT    TGGGGCCG  CCAG CGGTCCCG T  GCAGG TTCGACT CCTGT GATCTTCC GCCA
F1-56   GGAAGATC GG  CGTCAT  CGGT ATGGCG  GCTGGA CTTCAAA TCCAGT    TGGGGCCG  CCAG CGGTCCCG C  GCAGG TTCGACT CCTGT GATCTTCC GCCA
F1-57   GGAAGATC AG  CGTCAC  CGGT ATGGCG  GCTGGA CTTCAAA TCCAGT    TGGGGCCG  CCAG CGGTCCCG A  GCAGG TTCGACT CCTGT GATCTTCC GCCA
F1-58   GGAAGATC GA  CGTCCT  CGGT ACGGCG  GCTGGA CTTCAAA TCCAGT    TGGGGCCG  CCAG CGGTCCCG T  GCAGG TTCGACT CCTGT GATCTTCC GCCA
F1-59   GGAAGATC TG  CGTCTG  CGGT CGGGCG  GCTGGA CTTCAAA TCCAGT    TGGGGCCG  CCAG CGGTCCCG C  GCAGG TTCGACT CCTGT GATCTTCC GCCA
F1-60   GGAAGATC GG  CGTCAA  CGGT AAGGCG  GCTGGA CTTCAAA TCCAGT    TGGGGCCG  CCAG CGGTCCCG T  GCAGG TTCGACT CCTGT GATCTTCC GCCA
F1-61   GGAAGATC TG  CGTCTA  CGGT TGGGCG  GCTGGA CTTCAAA TCCAGT    TGGGGCCG  CCAG CGGTCCCG A  GCAGG TTCGACT CCTGT GATCTTCC GCCA
F1-62   GGAAGATC TC  CGTCCC  CGGT GGGGCG  GCTGGA CTTCAAA TCCAGT    TGGGGCCG  CCAG CGGTCCCG G  GCAGG TTCGACT CCTGT GATCTTCC GCCA
F1-63   GGAAGATC AG  CGTCAG  CGGT AAGGCG  GCTGGA CTTCAAA TCCAGT    TGGGGCCG  CCAG CGGTCCCG T  GCAGG TTCGACT CCTGT GATCTTCC GCCA
```

Table S2: List of Library F2 variants

	ACCEPTOR-STEM		D-STEM		ANTICODON-STEM		EXTRA-ARM		T-STEM	
		C1						C2		
LibF2	GGAAGATC	NNN	CGTCNN	CGGT	NNGGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	NN	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-1	GGAAGATC	TGG	CGTC**TA**	CGGT	**TA**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	CG	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-2	GGAAGATC	TAA	CGTCAA	CGGT	GAGGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	AG	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-3	GGAAGATC	TAT	CGTC**C**A	CGGT	G**G**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	AA	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-4	GGAAGATC	CAA	CGTC**AA**	CGGT	**TT**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	TA	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-5	GGAAGATC	GGA	CGTC**C**A	CGGT	A**G**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	GA	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-6	GGAAGATC	CTG	CGTC**TT**	CGGT	**AA**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	TT	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-7	GGAAGATC	AAA	CGTC**A**G	CGGT	T**T**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	TG	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-8	GGAAGATC	TTT	CGTC**A**G	CGGT	A**T**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	AT	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-9	GGAAGATC	GAT	CGTC**AA**	CGGT	**TT**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	TA	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-10	GGAAGATC	AAT	CGTC**CC**	CGGT	**GG**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	TG	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-11	GGAAGATC	TGG	CGTC**GG**	CGGT	**CC**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	CG	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-12	GGAAGATC	TT[	CGTC**AA**	CGGT	**TT**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	AG	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-13	GGAAGATC	GAT	CGTC**AT**	CGGT	**AT**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	GC	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-14	GGAAGATC	AAC	CGTC**C**A	CGGT	A**G**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	CG	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-15	GGAAGATC	TA[	CGTC**A**T	CGGT	G**T**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	TG	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-16	GGAAGATC	TGA	CGTC**AT**	CGGT	**AT**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	TT	[CAGG TTCGACT CCTGT GATCTTCC GCCA
F2-17	GGAAGATC	TAC	CGTC**TT**	CGGT	**AA**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGACCG CCAG	CGGTCCCG	TG	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-18	GGAAGATC	ATT	CGTC**TA**	CGGT	**TA**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	GG	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-19	GGAAGATC	TCA	CGTC**C**A	CGGT	G**G**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	GT	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-20	GGAAGATC	TTA	CGTC**TC**	CGGT	**GA**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	AG	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-21	GGAAGATC	ACC	CGTC**G**C	CGGT	**GT**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	TT	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-22	GGAAGATC	TCA	CGTC**CG**	CGGT	**CG**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	AA	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-23	GGAAGATC	GTC	CGTC**G**A	CGGT	A**C**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	CA	GCAGG TTCGATT CCTGT GATCTTCC GCCA
F2-24	GGAAGATC	GCT	CGTCA**T**	CGGT	**A**GGGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	AA	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-25	GGAAGATC	CGA	CGTC**AT**	CGGT	**AT**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	GT	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-26	GGAAGATC	TGT	CGTC**GA**	CGGT	**TC**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	GG	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-27	GGAAGATC	GTA	CGTCT**T**	CGGT	G**A**GACG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	AG	GCAGG TTCAACT CCTGT GATCTTCC GCCA
F2-28	GGAAGATC	TCG	CGTC**GC**	CGGT	**GC**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	AT	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-29	GGAAGATC	GTA	CGTC**GA**	CGGT	**TC**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	TA	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-30	GGAAGATC	ATA	CGTC**CA**	CGGT	**TG**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	TG	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-31	GGAAGATC	GTG	CGTC**CA**	CGGT	**TG**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	AT	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-32	GGAAGATC	TTC	CGTC**GG**	CGGT	**CC**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	AG	GCAGG TTCGACT CCTGT GATCTTCC GCCA
F2-33	GGAAGATC	TTT	CGTC**AA**	CGGT	**TT**GGCG	GCTGGA CTTCAAA TCCAGT	TGGGGCCG CCAG	CGGTCCCG	CT	GCAGG TTCGACT CCTGT GATCTTCC GCCA

Table S3: List of Library F3 variants

	ACCEPTOR-STEM	C1	D-STEM		D-STEM	ANTICODON-STEM		ANTICODON-STEM	EXTRA-ARM		EXTRA-ARM	C2	T-STEM		T-STEM	ACCEPTOR-STEM	
LibF3	GGAAGATC	NNNN	CGTCNN	CGGT	NNGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	NNN	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F3-1	GGAAGATC	AGAG	CGTC**AT**	CGGT	**AT**GGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	GGG	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F3-2	GGAAGATC	GTCC	CGTC**AC**	CGGT	**GT**GGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	ATA	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F3-3	GGAAGATC	ATAC	CGTCT**A**	CGGT	**T**GGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	CGT	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F3-4	GGAAGATC	GGTT	CGTC**AG**	CGGT	**CT**GGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	GTG	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F3-5	GGAAGATC	ATAC	CGTC**A**A	CGGT	G**T**GGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	TGT	GCAGG	TTCGACT	CCAGT	GATCTTCC	GCCA
F3-6	GGAAGATC	TAGA	CGTCT**C**	CGGT	**G**GGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	TTT	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F3-7	GGAAGATC	TGAC	CGTCT**C**	CGGT	**G**GGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	TAG	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA

Table S4: List of Library F5 variants

	ACCEPTOR-STEM	C1	D-STEM		D-STEM	ANTICODON-STEM		ANTICODON-STEM	EXTRA-ARM		EXTRA-ARM	C2	T-STEM		T-STEM	ACCEPTOR-STEM	.
LibF5	GGAAGATC	NNN	CGTCYN	CGGT	NYGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	NN	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F5-1	GGAAGATC	ATA	CGTCC**C**	CGGT	**G**TGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	AG	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F5-2	GGAAGATC	TTT	CGTCC**G**	CGGT	**C**TGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCC_	AT	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F5-3	GGAAGATC	TTA	CGTCT**C**	CGGT	**G**CGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	TG	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F5-4	GGAAGATC	GTC	CGTCC**T**	CGGT	**A**CGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	AA	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F5-5	GGAAGATC	AAC	CGTCCG	CGGT	GTGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	GG	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F5-6	GGAAGATC	TAC	CGTCT**C**	CGGT	**G**TGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	TG	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F5-7	GGAAGATC	AAA	CGTCCG	CGGT	GTGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	_G	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F5-8	GGAAGATC	AAG	CGTCT**C**	CGGT	**G**CGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	AG	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F5-9	GGAAGATC	GTA	CGTCT**C**	CGGT	**G**CGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	TG	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F5-10	GGAAGATC	CTT	CGTCT**C**	CGGT	**G**TGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	CG	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F5-11	GGAAGATC	GTT	CGTCT**C**	CGGT	**G**CGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	GT	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F5-12	GGAAGATC	GCC	CGTCT**C**	CGGT	**G**TGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	TG	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F5-13	GGAAGATC	AAC	CGTCT**C**	CGGT	**G**TGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	GG	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F5-14	GGAAGATC	GAC	CGTCC**C**	CGGT	**G**CGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	TG	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F5-15	GGAAGATC	GTG	CGTCT**C**	CGGT	**G**TGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	AG	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F5-16	GGAAGATC	GTC	CGTCT**C**	CGGT	**G**CGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	TG	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F5-17	GGAAGATC	CCC	CGTCT**C**	CGGT	**G**CGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	TG	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F5-18	GGAAGATC	GGA	CGTCC**C**	CGGT	**G**CGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	AG	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F5-19	GGAAGATC	GTA	CGTCT**C**	CGGT	**G**TGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	T_	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F5-20	GGAAGATC	GAC	CGTCC**C**	CGGT	**G**TGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	TG	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F5-21	GGAAGATC	ATA	CGTCC**C**	CGGT	**G**TGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	AT	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F5-22	GGAAGATC	TGA	CGTCT**C**	CGGT	**G**CGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	AG	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F5-23	GGAAGATC	GTC	CGTCT**G**	CGGT	**C**TGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	GT	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F5-24	GGAAGATC	TCC	CGTCT**C**	CGGT	**G**CGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	AA	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F5-25	GGAAGATC	GAC	CGTCCG	CGGT	GTGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	TG	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F5-26	GGAAGATC	GAC	CGTCC**C**	CGGT	**G**CGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	TA	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
F5-27	GGAAGATC	GTC	CGTCT**G**	CGGT	**C**TGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	GG	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA

Table S5: List of Library F4 variants

	ACCEPTOR-STEM		D-STEM		ANTICODON-STEM		EXTRA-ARM		T-STEM	
		C1						C2		
LibF4	GGAAGATC	NNN	CGTCRN	CGGT NRGGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG CGGTCCCG	NN	GCAGG TTCGACT	CCTGT GATCTTCC GCCA
F4-1	GGAAGATC	CTC	CGTCA**C**	CGGT **G**GGGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG CGGTCCCG	TG	GCAGG TTCGACT	CCTGT GATCTTCC GCCA
F4-2	GGAAGATC	GAG	CGTCA**C**	CGGT **G**GGGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG CGGTCCCG	AT	GCAGG TTCGACT	CCTGT GATCTTCC GCCA
F4-3	GGAAGATC	AGA	CGTCA**C**	CGGT **G**GGGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG CGGTCCCG	AG	GCAGG TTCGACT	CCTGT GATCTTCC GCCA
F4-4	GGAAGATC	GAA	CGTCAA	CGGT GGGGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG CGGTCCCG	AG	GCAGG TTCGACT	CCTGT GATCTTCC GCCA
F4-5	GGAAGATC	GGC	CGTCA**C**	CGGT **G**GGGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG CGGTCCCG	AG	GCAGG TTCGACT	CCTGT GATCTTCC GCCA
F4-6	GGAAGATC	GTG	CGTCA**C**	CGGT **G**GGGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG CGGTCCCG	GT	GCAGG TTCGACT	CCTGT GATCTTCC GCCA
F4-7	GGAAGATC	CAA	CGTCA**G**	CGGT **C**AGGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG CGGTCCCG	GT	GCAGG TTCGACT	CCTGT GATCTTCC GCCA
F4-8	GGAAGATC	TAA	CGTCA**C**	CGGT **G**GGGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG CGGTCCCG	GG	GCAGG TTCGACT	CCTGT GATCTTCC GCCA
F4-9	GGAAGATC	GAT	CGTCG**T**	CGGT **A**GGGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG CGGTCCCG	AA	GCAGG TTCGACT	CCTGT GATCTTCC GCCA
F4-10	GGAAGATC	TCG	CGTCA**C**	CGGT **G**GGGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG CGGTCCCG	GA	GCAGG TTCGACT	CCTGT GATCTTCC GCCA
F4-11	GGAAGATC	GAT	CGTCA**C**	CGGT **G**GGGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG CGGTCCCG	TT	GCAGG TTCGACT	CCTGT GATCTTCC GCCA

Table S6: List of Library F6 variants

	ACCEPTOR-STEM		D-STEM		ANTICODON-STEM		EXTRA-ARM			T-STEM		
		C1							C2			
LibF6	GGAAGATC	NNN	CGTCNY	CGGT YNGGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG	CGGTCCCG	NN	GCAGG TTCGACT	CCTGT	GATCTTCC GCCA
F6-1	GGAAGATC	TAT	CGTC**C**C	CGGT T**G**GGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG	CGGTCCCG	AT	GCAGG TTCGACT	CCTGT	GATCTTCC GCCA
F6-2	GGAAGATC	AAT	CGTC**A**T	CGGT T**G**GGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG	CGGTCCCG	AA	GCAGG TTCGACT	CCTGT	GATCTTCC GCCA
F6-3	GGAAGATC	AAT	CGTC**C**C	CGGT T**G**GGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG	CGGTCCCG	CG	GCAGG TTCGACT	CCTGT	GATCTTCC GCCA
F6-4	GGAAGATC	AAA	CGTC**T**C	CGGT T**G**GGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG	CGGTCCCG	CG	GCAGG TTCGACT	CCTGT	GATCTTCC GCCA
F6-5	GGAAGATC	AAG	CGTC**C**T	CGGT T**G**GGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG	CGGTCCCG	GT	GCAGG TTCGACT	CCTGT	GATCTTCC GCCA
F6-6	GGAAGATC	GGG	CGTC**C**T	CGGT T**G**GGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG	CGGTCCC-	TG	GCAGG TTCGACT	CCTGT	GATCTTCC GCCA
F6-7	GGAAGATC	TTT	CGTCAT	CGGT TGGGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG	CGGTCCCG	A-	-CAGG TTCGACT	CCTGT	GATCTTCC GCCA
F6-8	GGAAGATC	AAT	CGTC**C**T	CGGT T**G**GGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG	CGGTCCCG	GC	GCAGG TTCGACT	CCTGT	GATCTTCC GCCA
F6-9	GGAAGATC	TAT	CGTC**C**T	CGGT C**G**GGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG	CGGTCCCG	GA	GCAGG TTCGACT	CCTGT	GATCTTCC GCCA
F6-10	GGAAGATC	AAG	CGTC**C**T	CGGT T**G**GGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG	CGGTCCCG	GT	GCAGG TTCGACT	CCTGT	GATCTTCC GCCA
F6-11	GGAAGATC	TAG	CGTC**C**T	CGGT C**G**GGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG	CGGTCCCG	CA	GCAGG TTCGACT	CCTGT	GATCTTCC GCCA
F6-12	GGAAGATC	TAA	CGTCAT	CGGT TGGGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG	CGGTCCC-	G-	GCAGG TTCGACT	CCTGT	GATCTTCC GCCA
F6-13	GGAAGATC	AAA	CGTC**C**T	CGGT C**G**GGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG	CGGTCCCG	TT	GCAGG TTCGACT	CCTGT	GATCTTCC GCCA
F6-14	GGAAGATC	TAG	CGTCAT	CGGT TGGGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG	CGGTCCCG	AG	GCAGG TTCGACT	CCTGT	GATCTTCC GCCA
F6-15	GGAAGATC	AAA	CGTC**A**T	CGGT T**T**GGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG	CGGTCCCG	AA	GCAGG TTCGACT	CCTGT	GATCTTCC GCCA
F6-16	GGAAGATC	AAT	CGTC**C**T	CGGT T**G**GGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG	CGGTCCCG	CG	GCAGG TTCGACT	CCTGT	GATCTTCC GCCA
F6-17	GGAAGATC	TAT	CGTCAT	CGGT TGGGCG	GCTGGA CTTCAAA TCCAGT		TGGGGCCG CCAG	CGGTCCCG	A-	GCAGG TTCGACT	CCTGT	GATCTTCC GCCA

Table S7: Covariation analysis for the 5th base pair from library F1 variants

UU(2.5)=1; UC(0.2)=0; **UA(9.5)=13**; UG(7.6)=6
CU(2.5)=1; CC(0.2)=1; CA(6.7)=3; **CG(5.3)=9**
AU(3.6)=6; AC(0.4)=0; AA(13.3)=14; AG(10.7)=8
GU(0.1)=0; **GC(0.0)=0**; GA(0.5)=0; GG(0.4)=1

Table S8: Covariation analysis for the 6th base pair from library F1 variants

UU(1.1)=1; UC(0.3)=0; **UA(2.2)=3**; UG(1.3)=1
CU(0.7)=0; CC(0.2)=0; CA(1.3)=2; **CG(0.8)=1**
AU(8.9)=8; AC(2.5)=2; AA(17.8)=20; AG(10.8)=10
GU(3.3)=5; **GC(1.0)=2**; GA(6.7)=3; GG(4.0)=5

Supplemental Extended Experimental Procedure

The ordered oligos for all libraries are as follows:

Library F1, 5'CGGAATTCGGAAGATCNNCGTCNNCGGTNNGGCGGCTGGACTTCAAATCCAGTTGGGGCCGCCAGCGGTCCCGNGCAGGTTCGACTCCTGTGATC3'
Library F2, 5'CGGAATTCGGAAGATCNNNCGTCNNCGGTNNGGCGGCTGGACTTCAAATCCAGTTGGGGCCGCCAGCGGTCCCGNNGCAGGTTCGACTCCTGTGATC3'
Library F3, 5'CGGAATTCGGAAGATCNNNNCGTCNNCGGTNNGGCGGCTGGACTTCAAATCCAGTTGGGGCCGCCAGCGGTCCCGNNNGCAGGTTCGACTCCTGTGATC3'
Library F4, 5'CGGAATTCGGAAGATCNNNCGTCRNCGGTNRGGCGGCTGGACTTCAAATCCAGTTGGGGCCGCCAGCGGTCCCGNNGCAGGTTCGACTCCTGTGATC3'
Library F5, 5'CGGAATTCGGAAGATCNNNCGTCYNCGGTNYGGCGGCTGGACTTCAAATCCAGTTGGGGCCGCCAGCGGTCCCGNNGCAGGTTCGACTCCTGTGATC3'
Library F6, 5'CGGAATTCGGAAGATCNNNCGTCNYCGGTYNGGCGGCTGGACTTCAAATCCAGTTGGGGCCGCCAGCGGTCCCGNNGCAGGTTCGACTCCTGTGATC3'

Calculation of the expected number of dinucleotide combinations if random.

The expected number of clones for a particular dinucleotide combination was calculated as follows: if the total number of screened clones is N_{all}, while the numbers of clones with A in position 14 and with U in position 21 are, respectively, N_{A14} and N_{U21}, then the expected number of clones containing dinucleotide combination A14-U21 under the presumption that the identities of the two positions varied independently would be $N_{A14,U21} = (N_{A14} * N_{U21})/N_{all}$. Then, the expected number of WC combinations 14-21 (N_{WC}) would be: $N_{WC} = N_{A14,U21} + N_{U14,A21} + N_{G14,C21} + N_{C14,G21}$.

Based on the above, we have tabulated the expected number of dinucleotide combinations for all possible dinucletide combinations for the 5th (Table 7) and the 6th (Table 8) layer of the D-stem if they were to occur randomly (the number in brackets). The number followed by '=' is the actual number of occurences.

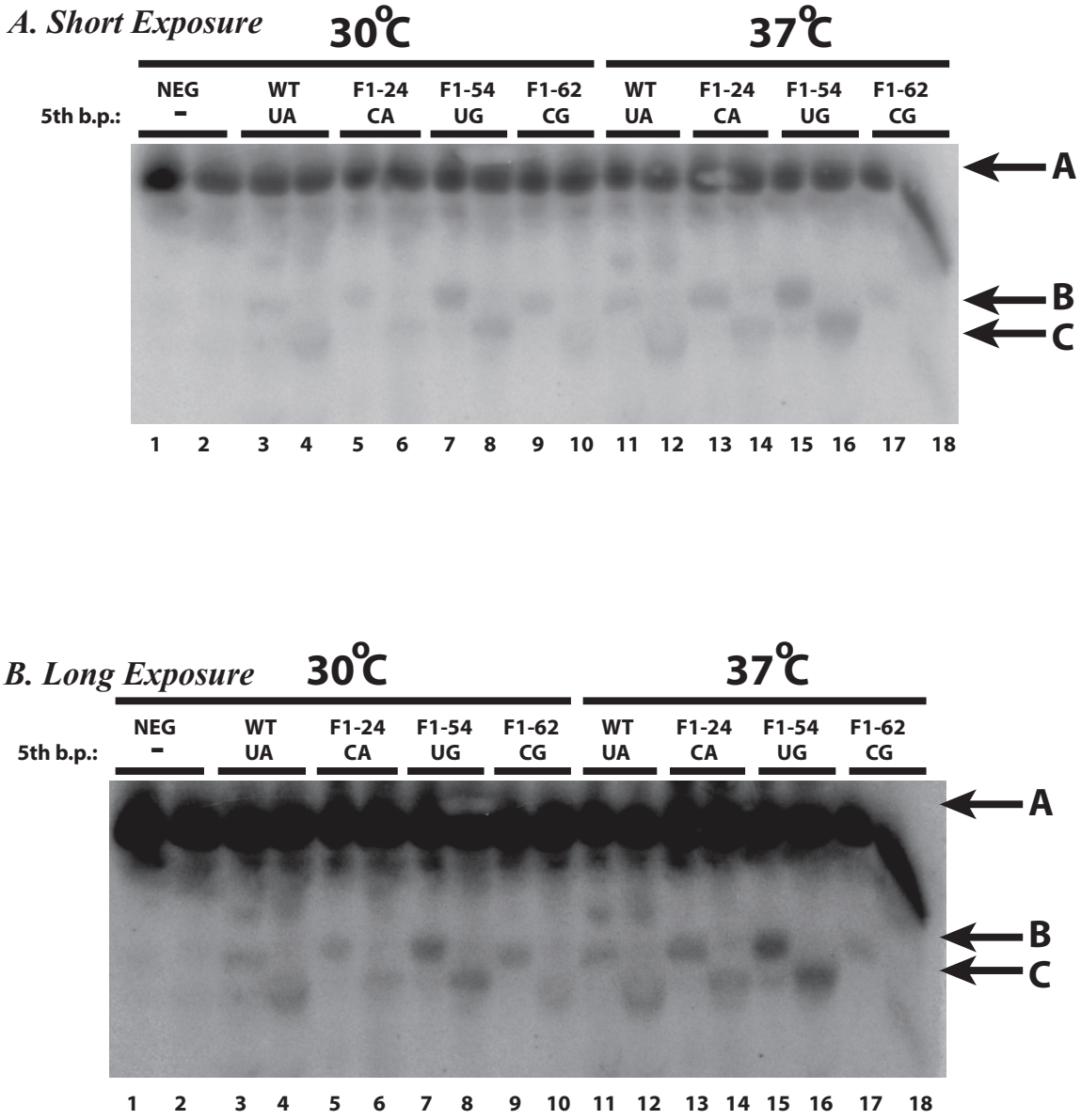
The expected WC combination outcome was 18.4 while the actual case was 28, thus indicating a strong bias for the formation of WC in the 5th layer of the D-stem.

The expected WC combination outcome for the 6th layer was 12.9 while the actual case was 14. At the 6th layer, there was not a strong bias for the formation of WC base pairs.

Figure Legend to Tables S1-S6

List of all screened clones. The name of the clone is followed by the DNA sequence of the $tRNA^{Sec}$. The designed library is found at the top with the name starting with 'Lib'. The representation of the following symbols are described as follows:C1, connector 1; C2, connector 2; G, Guanine; A, Adenine;T, Thymine;C, Cytosine; _ , Empty;N, Any; R, Purine; Y, Pyrimidine; Red, non-randomized regions; Black, Randomized regions; Black and Bold, WC base pair; Black and Underlined, GU base pair; Green highlight, spontaneous mutations.

Figure S1


A. Short Exposure
30°C
37°C
5th b.p.:
NEG
–
WT
UA
F1-24
CA
F1-54
UG
F1-62
CG
WT
UA
F1-24
CA
F1-54
UG
F1-62
CG
A
B
C
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
B. Long Exposure
30°C
37°C
5th b.p.:
NEG
–
WT
UA
F1-24
CA
F1-54
UG
F1-62
CG
WT
UA
F1-24
CA
F1-54
UG
F1-62
CG
A
B
C
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

CHAPTER V

The long acceptor/T domain and not the long acceptor stem is important for the function of the bacterial selenocysteine tRNA

Tetsu M. Ishii, Stéphane Nemours & Sergey V. Steinberg*

Tetsu M. Ishii designed and performed experiments, analyzed the results and wrote the paper.

Stéphane Nemours performed experiments.

Sergey V. Steinberg designed the experiments, analyzed the results and wrote the paper.

Université de Montréal, Département de Biochimie, C.P. 6128, succursale Centre-Ville
Montréal, PQ, Canada H3C 3J7.

* To whom correspondence should be addressed. Tel: 1-514-343-6320; Fax: 1- 514-343-2210

ABSTRACT

The selenocysteine tRNA ($tRNAs^{Sec}$) is renowned for its long acceptor stem, having either eight or nine base pairs (bp) instead of the standard seven. Also, the acceptor/T domain of the $tRNA^{Sec}$ always contains 13 bp instead of the standard 12. We show here that for a reliably detectable level of the bacterial $tRNA^{Sec}$ function *in vivo*, neither the long acceptor stem, nor the long acceptor/T domain is required. However, the variants with 13 bp in the acceptor/T domain regardless of the length of the acceptor stem, which can have 7, 8 or 9 bp, displayed higher activity. The indifference of the $tRNA^{Sec}$ function to the length of the acceptor stem infers the existence of substantial conformational flexibility in its tertiary structure, potentially caused by the absence of the standard tertiary interactions 8-14 and 15-48.

KEYWORDS

Selenocysteine, tRNA, combinatorial library, *in vivo* screen, instant evolution

INTRODUCTION

Selenocysteine is the 21st amino acid that is co-translationally incorporated into polypeptide chains. The incorporation takes place at a UGA codon in response to SECIS, a particular secondary structure element located in the mRNA downstream to this codon(Zinoni et al., 1987; Zinoni et al., 1990). Selenocysteine is delivered to the ribosome by the specific selenocysteine tRNA ($tRNA^{Sec}$) in the ternary complex with translation factor SelB (analog of EF-Tu) and GTP(Bock et al., 1991; Forchhammer et al., 1989). The $tRNA^{Sec}$ is known for its unusual cloverleaf secondary structure, the most prominent aspect of which is the long acceptor stem (Figure 1)(Bock et al., 1991; Leinfelder et al., 1988; Schön et al., 1989). While in all other cytosolic tRNAs the acceptor stem contains strictly seven base pairs (bp), in the $tRNA^{Sec}$ it has either eight(Itoh et al.) (in bacteria) or nine (in archaea and in eukaryotes) bp(Chiba et al.; Itoh et al., 2009). In addition, the T-stem of the archaeal and eukaryotic $tRNAs^{Sec}$ contains only four bp instead of the normal five. As a result, the acceptor/T domain in all $tRNAs^{Sec}$ universally contains 13 bp, i.e. one base pair more than in other cytosolic tRNAs. The structure of the acceptor/T domain can be denoted as M/N, where M and N stand for the number of bp in the acceptor and T-stems, respectively. Correspondingly, the structure of this domain in $tRNA^{Sec}$ is denoted as either 8/5 (in bacteria) or 9/4 (in archaea and in eukaryotes).

Based on the universal presence of the abnormally long acceptor stem and acceptor/T domain in all $tRNAs^{Sec}$, it was suggested that these structural aspects or at least one of them is essential for the $tRNA^{Sec}$ function (Bock et al., 1991). It remains however unclear which of the two aspects is more important and how critical it is for the $tRNA^{Sec}$ function. To address

these questions, we used here the Instant Evolution approach (Ishii et al., 2013; Lee et al., 1997) consisting in the screening of functional *in vivo* variants of the $tRNA^{Sec}$ from specially designed combinatorial gene libraries in which certain areas of the nucleotide sequence were randomized. This approach allowed us to find functional $tRNA^{Sec}$ variants having non-standard structures of the acceptor/T domain. Analysis of 55 such variants revealed the limits within which the structure of the acceptor/T domain can be modified without ruining the $tRNA^{Sec}$ function. In particular, we demonstrate here that for a reliably detectable level of selenocysteine incorporation, neither the unusual length of the acceptor stem nor of the acceptor/T domain of the $tRNA^{Sec}$ variant was required. However, all variants with a robust activity approaching or even exceeding that of the wild-type $tRNA^{Sec}$ (WT), contained invariably 13 base pairs in the acceptor/T domain, irrespective of the length of the acceptor stem.

The experimental approach for *in vivo* screening of $tRNA^{Sec}$ variants

Functional variants of the $tRNA^{Sec}$ were obtained through *in vivo* screening of specially designed combinatorial gene libraries in which certain nucleotide positions of the $tRNA^{Sec}$ were modified or randomized. The libraries were cloned into the pGFIB-1 expression vector and transformed into the *E. coli* strain WL81460, in which the gene coding for the $tRNA^{Sec}$ (gene *SelC*) had been inactivated (Zinoni et al., 1990). Bacterial colonies expressing functional $tRNA^{Sec}$ variants were collected based on the activity of the endogenously expressed selenoprotein formate dehydrogenase N (FDH_N), in which the amino acid selenocysteine is the key component of the catalytic center (Ishii et al., 2013; Kramer and Ames, 1988). When cells were grown under anaerobic conditions, the color of the colonies varied between dark red and bright-white, indicating the range of FDH_N activities between low and relatively high, respectively. For our purpose, we screened exclusively bright-white

colonies, thus assuring a certain level of the FDH_N activity. On average, such variants appeared once in a thousand colonies screened. To verify whether the screened $tRNA^{Sec}$ variants were delivered to the ribosome by factor SelB and not by EF-Tu, ten arbitrary chosen effective variants screened from different libraries were expressed in the $SelB^-$ strain (Tormay et al., 1996). None of the chosen variants demonstrated any selenocysteine incorporating activity in the $SelB^-$ cells.

To verify the effectiveness of the screened $tRNA^{Sec}$ variants, some of the clones shown at the previous step to be SelB-dependent were subjected to the benzyl-viologen assay (Ishii et al., 2013; Takahata et al., 2008). This assay measured the activity of another intracellular selenoprotein, the formate dehydrogenase H (FDH_H), in which the amino acid selenocysteine is also the key component of the catalytic center. Therefore, the activity of FDH_H directly reflected the functionality of the $tRNA^{Sec}$ variants. The colony containing the void plasmid had no detectable FDH_H activity, while the colony containing the plasmid-based WT $tRNA^{Sec}$ demonstrated a robust activity, which was set as 100%.

The expression of the human $tRNA^{Sec}$ showed weak FDH_N activity based on the pink colored colonies, and was consistent with the FDH_H activity measurements based on the benzyl violgen assay which showed 11% of WT. Both the FDH_N and FDH_H activities determined for the human $tRNA^{Sec}$ corresponded well to the previously made measurements (Baron et al., 1994) (see also Supplementary Note).

RESULTS

The importance of the 8th base pair of the acceptor stem for the $tRNA^{Sec}$ function

To assess the importance of the long acceptor stem for the $tRNA^{Sec}$ function, we expressed a combinatorial gene library (library G1) in which the identities of the nucleotides forming

the last two base pairs of this stem as well as the two nucleotides preceding its 5'-strand and the five nucleotides following its 3'-strand were randomized (Figure 2). The screening of this library provided nine active G1-variants shown in Figure 2. In all of them, the dinucleotide combination corresponding to the eighth base pair of the acceptor stem was invariably GC, as in WT. The invariableness of this G(-1)-C72a base pair (Juhling et al., 2009) fits well to the previous work of Burkard et al. (Burkard and Söll, 1988), which demonstrated the importance of this base pair for the proper RNase P-mediated cleavage of the $tRNA^{Sec}$ transcript between positions -2 and -1 (Figure 1).

Other randomized nucleotides were less conserved, indicating that they are not critically important for the $tRNA^{Sec}$ function. In particular, the second last base pair of the acceptor stem (bp 1-72) was predominantly GC, as in WT, but could also be GT (1 variant) or AT (1 variant). The nucleotide following the last base pair of the acceptor stem (nt 73), known as the discriminator nucleotide, was mostly G (in six variants), as in WT, but could also be A (in three variants). The identities of nucleotides 74-76, corresponding to the CCA-3' sequence, varied substantially, thus demonstrating that they do not need to be encoded in the tRNA gene and can be added later by the CCA-adding enzyme (Betat et al.).

A functional $tRNA^{Sec}$ variant having the 7/5 structure

The results obtained from the screening of library G1 demonstrated that it is not possible to shorten the acceptor stem of the $tRNA^{Sec}$ by eliminating the base pair proximate to the acceptor terminus. This, in turn, prompted us to explore another way for shortening the acceptor stem that consisted in the elimination of a base pair at the opposite end of the acceptor stem adjacent to the T-stem. To check whether the acceptor stem can be shortened in this way, we designed two consecutive libraries G2 and G3. In library G2, the base pair in the acceptor stem adjacent to the T-stem was eliminated, while several nucleotide positions

in the T-stem and in the connector region between the acceptor and D-stems and in the D-loop were randomized (Figure 2, see also Supplementary Tables S3 and S4). The screening of library G2 provided three functional variants shown in Supplementary Table S3. The screened variant G2-3 was then used as a template for the subsequent library G3, which provided two functional variants, G3-1 and G3-2 (Figure 2 and Supplementary Table S4), containing five base pairs in the T-stem and only seven base pairs in the acceptor stem. Even though each of the two variants G3-1 and G3-2 had the 7/5 structure, which had never been observed before in a functional $tRNA^{Sec}$, both variants formed bright-white colonies, thus demonstrating the ability to incorporate selenocysteine at the specific UGA codon of the FDH_N mRNA.

To confirm the presence of only seven base pairs in the acceptor stem of the screened G3-variants, we used the nucleotide sequence of variant G3-1 to design another combinatorial library (library G4, Figure 2). In this library, the nucleotide position preceding nucleotide -1 and several nucleotide positions following nucleotide 72 were randomized. If the function of the screened G4-variants depended on the presence of the eighth base pair in the acceptor stem, the identities of the nucleotides forming this base pair were expected to co-vary in the way to demonstrate such dependence. The results showed however, that out of the nine collected G4-variants (Figure 2 and Supplementary Table S5), only in two cases (variants G4-5 and G4-9) nucleotides corresponding to the eighth base pair constituted a Watson-Crick combination, while in the other variants they were CA (2 variants), AG (2 variants) and GG (3 variants). Like in G1-variants, the discriminator nucleotide in the G4-variants was either G or A. Also, the nucleotide positions that corresponded to the CCA-3' sequence demonstrated the same wide variability, which again indicated the involvement of the CCA-adding enzyme in the maturation of the G4-variants. These results confirmed that the function of the G4-variants and, therefore, of variant G3-1

did not depend on the presence of the eighth base pair in the acceptor stem. Finally, the expression of variant G3-1 in the SelB$^-$ strain did not demonstrate any selenocysteine incorporating activity, which eliminated the possibility that this variant was delivered to the ribosome using an alternative SelB-independent way. This finding led us to the conclusion that although variant G3-1 had the 7/5 structure, its functional pattern closely followed that of the WT tRNA^{Sec}.

Functional variants of the tRNA^{Sec} having structures 8/4, 9/3, 9/4 and 7/6

At the next step, we used the same approach to generate tRNA^{Sec} variants having the 8/4 (library G5), 9/4 (libraries G6, G8, and G9) and 7/6 (library G7) structures. In each case, we started with the nucleotide sequence of WT, and then the lengths of the acceptor and T-stems were adjusted through elimination or insertion of base pairs in either the acceptor or T-stem. To increase the adaptability of the variants of the tRNA^{Sec} to these modifications, the nucleotides of several base pairs in both stems were randomized (see Supplementary Table S1). For the same reason, we introduced two additional randomized nucleotides at the beginning and at the end of the D-loop and, only in library G6, we randomized the terminal base pair of the extra arm. In total, we collected 55 new functional variants of the tRNA^{Sec}, which are presented in Supplementary Tables S2-S11. During the screening, all selected variants formed bright-white colonies, which proved their ability to incorporate selenocysteine at the corresponding codon of the FDH_N mRNA. Ten variants, five of which are shown in Figure 3, while the others are G3-2, G5-1, G6-1, G7-1, and G8-9, were expressed in the SelB$^-$ strain. None of the variants tested were functional in the SelB$^-$ strain (data not shown), which proved that all of them were delivered to the ribosome using the same path as the WT tRNA^{Sec}.

Activity, steady-state level and the level of aminoacylation of the screened variants having 12 and 13 bp in the acceptor/T domain

For the five tRNA^{Sec} variants shown in Figures 3 and 4, we measured the activity of the selenocysteine incorporation using the benzyl-viologen assay (Ishii et al., 2013; Takahata et al., 2008). The results of the measurements show that in spite of the bright-white color of the colonies formed by all these variants, the activities of the variants varied substantially. The highest activity, observed for clone G6-7 (130% of WT) was 26 times higher than of clone G3-1 (5%). Based on their activities, the measured variants can be divided in two distinct groups of those having the activity of 60% of WT or higher and of those whose activity did not exceed 10% of WT. The first group included WT (activity 100%), as well as variants G6-7 (130%) and G7-4 (60%), in all of which the acceptor/T stem contained 13 bp. The second group contained variants G5-4 (10%), G8-10 (9%), and G3-1 (5%), all of which had only 12 bp in the acceptor/T domain. Thus, the comparison of the two groups strongly suggests that the presence of 13 bp in this domain is a major requirement for the high efficiency of a tRNA^{Sec} variant.

The Northern blot of the same tested variants (Figure 4) showed that for each of them except G3-1, a detectable amount was present in the cytosol. The relative position of all these variants on the gel was predictable based on their molecular masses calculated from the known nucleotide sequences. Thus, variants G6-7 and G7-4 migrated slower than the WT tRNA^{Sec}, which reflected the presence of two additional nucleotides in the D-loop (Supplementary Tables S1, S7 and S8). Also, G7-4 migrated slower than G6-7 due to the fact that in three nucleotide positions, two in the D-loop and one between the extra arm and the T-stem (position 48), it contained purines, while G6-7 contained pyrimidines (Supplementary Tables S7 and S8). In the same way, one can explain the positions of

variants G5-4 and G8-10. The existing correlation between the nucleotide sequence of the clones and their positions on the gel proves that the measured variants were properly processed and kept their integrity. Also, all these variants demonstrated a substantial level of aminoacylation (shown at the top of Figure 4).

The Northern blot (Figure 4) also showed that the variants having 13 bp in the acceptor/T domain were present in the cell in higher amounts than the 12-bp-variants. For variant G3-1, its position on the gel was indistinct which is consistent with the fact that this variant had the lowest activity among all measured variants.

DISCUSSION

In this paper we demonstrate that for the effective functioning of a $tRNA^{Sec}$ variant, its acceptor/T domain should contain 13 bp, while the particular length of the acceptor stem is not so important. Although in all naturally selected $tRNAs^{Sec}$ the acceptor stem is always longer than in other tRNAs, we show here that a functional $tRNA^{Sec}$ variant can have as little as seven bp in the acceptor stem. However, for such a variant to be highly effective, the length of its T-stem must be extended to the unprecedented six bp, thus making 13 bp in the whole acceptor/T domain. The necessity to have a total of 13 bp in this domain unifies all naturally occurring $tRNAs^{Sec}$ despite the variation in their acceptor stem lengths observed across different evolutionary domains. The existence of a 13 bp acceptor/T domain in a $tRNA^{Sec}$ variant leads to its higher steady-state level in the cell. We can thus suggest that the proper interaction of a $tRNA^{Sec}$ variant with other factors of the selenocysteine incorporation machinery is favored by the long acceptor/T domain, which provides additional stabilization for the tRNA structure.

The requirement for the 13-base pair acceptor/T domain is not however absolutely rigid. A reliably detectable level of the selenocysteine-incorporating activity can be achieved

even if this domain contains only twelve bp. We found, however, that the shortening of the acceptor/T domain from 13 to 12 bp was always associated with a significant drop in the $tRNA^{Sec}$ activity, as measured by the benzyl viologen assay and a notably lower steady-state level, as shown by the Northern blot. Despite the low activity and the low presence, all screened variants demonstrated the ability to specifically incorporate selenocysteine *in vivo* at the assigned codons of two selenoproteins FDH_N and FDH_H. Ten arbitrarily chosen variants were shown to be SelB-dependent, which indicated that they incorporate selenocysteine into the polypeptide chains using the same functional path as the WT $tRNA^{Sec}$. Based on this, we can expect that most other screened variants were SelB-dependent as well. If a SelB-independent $tRNA^{Sec}$ variant were found, it would have meant that it was delivered to the ribosome as a regular elongator tRNA, using EF-Tu. Whether such variants can be found using the approach exercised here remains unknown. However, even if such variants existed, they would not have changed the conclusions of this paper, which were made based on the analysis of only those variants whose SelB-dependence had been demonstrated.

Compared to the length of the whole acceptor/T domain, the particular lengths of the acceptor and T-stems are of lesser importance. This aspect opens the possibility for redistribution of bp between the two stems in functional $tRNA^{Sec}$ variants. Our results clearly show that in the acceptor/T domain of the $tRNA^{Sec}$ the junction between the acceptor and T-stem does not need to be always located at the same place and can migrate towards the T-loop or towards the acceptor terminus, providing for the 9/4 and 7/6 structures, respectively.

The possibility for rearrangement of bp between the acceptor and T-stems makes the $tRNA^{Sec}$ unique among cytosolic tRNAs and reflects particular features of its structural

organization. In all normal cytosolic tRNAs, the position of the junction between the acceptor and T-stems is fixed with respect to the anticodon/D domain with help of the universal tertiary interactions 8-14 and 15-48. Due to this fixation, shortening or extending of the acceptor stem in a normal tRNA by only one base pair will result in the shift and in the rotation of the acceptor terminus with respect to the anticodon/D domain by 2.8Å and 33˚, respectively, potentially damaging the tRNA function. The $tRNA^{Sec}$, however, due to the absence of the above-mentioned tertiary interactions, does not possess equivalent structural tools able to fix the position of the acceptor-T junction with respect to the anticodon/D domain. In the absence of such tools, the two helical domains become arranged with respect to each other rather flexibly. Such flexibility makes the particular lengths of each of the acceptor and T-stems less important for the $tRNA^{Sec}$ function, as long as their total length remains unchanged. It also fits to the fact that in different evolutionary domains, the acceptor/T domain of the $tRNA^{Sec}$ has either the 8/5 or 9/4 structure.

A substantial level of conformational flexibility caused by the absence of the standard tertiary interactions 8-14 and 15-48 seems to be essential for the $tRNA^{Sec}$ function. In the $tRNA^{Sec}$, these tertiary interactions are replaced by two additional base pairs in the D-stem. We recently showed that the additional bp in the D-stem of the $tRNA^{Sec}$ do not play any specific role and are dispensable (Ishii et al., 2013). Moreover, at a lower temperature (30˚C) the existence of these bp was even harmful for the $tRNA^{Sec}$ function. Only at a higher temperature the additional bp in the D-stem became essential as they can stabilize the $tRNA^{Sec}$ structure without use of the above-mentioned tertiary interactions. The absence of the standard tertiary interactions in the D-domain makes the $tRNA^{Sec}$ intrinsically flexible and in this sense, unique among cytosolic tRNAs.

The two unique features of the $tRNA^{Sec}$, the internal flexibility of its tertiary structure and the long acceptor/T domain, are essential for its function. In fact, these features can be linked to each other. Indeed, a rigid tRNA molecule with an abnormally long acceptor/T domain would face difficulties while moving through the ribosome from one tRNA-binding site to another. However, the observed internal flexibility of the tRNA structure should essentially alleviate such a problem and allow a smooth passage. Most probably, both unique features of the $tRNA^{Sec}$ are linked to its unprecedented function, which consists in the recognition of the UGA codon not solely based on the cognateness of the codon-anticodon interaction, but in context of the secondary structure element SECIS located in the mRNA a few nucleotides downstream. How the two unique features of the $tRNA^{Sec}$ structure cooperate in providing for such an unusual function will require further analysis.

MATERIALS AND METHODS

Bacterial Strains- The *E.coli* strain WL81460 (Δ(*argF-lac*)U169 *rpsL150*, *rpsL*$^{+}$ *rpsE13*, Δ(*srl-recA*)306::Tn*10*, Δ(*selC*)400::*Kan*)(Zinoni et al., 1990), having a deletion in the gene coding for the $tRNA^{Sec}$ (SelC), and WL81300(Δ(*argF-lac*)U169 *rpsL150*, *rpsL*$^{+}$ *rpsE13*, Δ(*srl-recA*)306::Tn*10*, Δ(*selB*)300::*Kan*) (Tormay et al., 1996), having a deletion in the gene coding for the elongation factor SelB, were used in this study.

Combinatorial Library designs and cloning - Oligonucleotides for all library DNA templates (Supplementary Method and Supplementary Table S1) and for the primers required for the library amplification were ordered from Bio-Corp Inc. (Montreal, Canada). The designs of all combinatorial libraries (Supplementary Table S1) as well as all sequences of the ordered oligos for PCR are shown in the Supplementary methods. Each library was PCR-amplified using the corresponding pair of primers (Supplementary Methods) and was

cloned into plasmid pGFIB-1 using restriction sites EcoRI and PstI, as described before (Zagryadskaya et al., 2003).

In vivo screening of active tRNA^{Sec} variants- Colonies containing active variants of the tRNA^{Sec} were identified by the bright white color attributed to the presence of the active selenoprotein formate dehydrogenase N (FDH_N). This protein uses formate as a source of electrons for reduction of nitrate. The absence of FDH_N, on the contrary, would result in accumulation of formate and decrease of pH. The level of pH can be detected by the pH indicator present in the growth medium. Depending on whether the pH is high or low, the indicator will make bacterial colonies white and red, respectively (Barrett et al., 1979). In this study, all non-white colonies were ignored.

Active colonies were screened after plating the ligation reaction onto MacConkey Nitrate Agar plates and incubation for 48 hours at 37˚C under anaerobic conditions (Barrett et al., 1979). MacConkey nitrate agar contained (per liter): 40 g of MacConkey agar base (Difco MacConkey Agar Base), 10 g of potassium nitrate, and 0.5 g of sodium formate. Bright white colonies were picked and sequenced.

Preparing samples for the Formate Dehydrogenase H assay- For some selected variants, the level of activity was measured using the assay for another selenoprotein formate dehydrogenase H (FDH_H), which is able to reduce benzyl viologen. By following the rate of the benzyl viologen reduction, we quantitated the level of FDH_H in the cell. The FDH_H assay was performed as described elsewhere (Takahata et al., 2008), with some modifications. Plasmids of selected clones were transformed into the WL81460 strain, plated onto LB agar plates containing ampicillin and grown overnight. Colonies were picked and grown overnight in 2 mL of LB with ampicillin under aerobic conditions. Subsequently, 1.5 mL eppendorf tubes were filled up with Stadtman buffer containing

ampicillin, inoculated with 40 microliters of the overnight culture, and incubated for 20 hours at 30˚C. The density of samples was measured at O.D. 600nm. Samples were sedimented by centrifugation, frozen and stored at -80˚C.

Formate Dehydrogenase H assay- Samples were defrosted at room temperature. The pellets were washed with 300 microliters of 0.5X TBE containing 5mM $MgSO_4$, re-suspended in 800 microliters of the same buffer and transferred to a glass cuvette containing 100 microliters of sodium formate [200mM] and 100 microliters of benzyl viologen [20mM]. Both sodium formate and benzyl viologen solutions were prepared with pH 7.0 0.1M potassium phosphate buffer. The cuvette was covered with a rubber stopper, deoxygenated for 5 minutes with argon, and was left at room temperature for 5 more minutes. The FDH_H activity was assayed by measuring the rate of the increase in the absorbance at 600nm during 5 minutes. The activities were represented as the change in O.D. at 600nm per minute normalized by the density of the cell culture. Each assay was performed in triplicate and then averaged.

Northern Blot of $tRNA^{Sec}$ variants- The Northern blot was performed as described before (Kotlova et al., 2007) with some modifications. For RNA extraction, the bacteria were grown under anaerobic conditions in Stadtman buffer at 30˚C until the O.D. of absorbance at 600 nm reached 0.4. One probe was complementary to the extra-arm (5'-CGGGACCGCTGGCGGCCCCA-3') of the *E.coli* $tRNA^{Sec}$, while the other one was complementary to region 34–53 of the *E. coli* 5S rRNA(5'- TTCTGAGTTCGGCATGGGGT-3'). The 5S rRNA probe was used to monitor the amount of total RNA in each sample. Densitometry analysis of the bands was performed using the Quantity One® software (BioRad).

SUPPLEMENTARY DATA

Supplementary Tables S1-S11, Supplementary Methods, and Supplementary Note.

ACKNOWLEDGEMENTS

We thank Drs. Gary Sawers and August Böck for the gift of bacterial strains WL81460 and WL81300, Drs. Christian Baron and Patrick Hallenbeck for discussions and valuable advice.

FUNDING

This work was supported by the Canadian Institutes of Health Research.

REFERENCES

1. Milo R, Jorgensen P, Moran U, Weber G, Springer M (2010) BioNumbers--the database of key numbers in molecular and cell biology. Nucleic Acids Res 38: D750-753.
2. Yoshizawa S, Bock A (2009) The many levels of control on bacterial selenoprotein synthesis. Biochim Biophys Acta 1790: 1404-1414.
3. Leinfelder W, Zehelein E, Mandrand-Berthelot MA, Bock A (1988) Gene for a novel tRNA species that accepts L-serine and cotranslationally inserts selenocysteine. Nature 331: 723-725.
4. Sturchler C, Westhof E, Carbon P, Krol A (1993) Unique secondary and tertiary structural features of the eucaryotic selenocysteine tRNA(Sec). Nucleic Acids Res 21: 1073-1079.
5. Palioura S, Sherrer RL, Steitz TA, Söll D, Simonovic M (2009) The human SepSecS-tRNASec complex reveals the mechanism of selenocysteine formation. Science 325: 321-325.
6. Itoh Y, Chiba S, Sekine S, Yokoyama S (2009) Crystal structure of human selenocysteine tRNA. Nucleic Acids Res 37: 6259-6268.
7. Burkard U, Söll D (1988) The unusually long amino acid acceptor stem of Escherichia coli selenocysteine tRNA results from abnormal cleavage by RNase P. Nucleic Acids Res 16: 11617-11624.
8. Wu XQ, Gross HJ (1994) The length and the secondary structure of the D-stem of human selenocysteine tRNA are the major identity determinants for serine phosphorylation. EMBO J 13: 241-248.
9. Chiba S, Itoh Y, Sekine S, Yokoyama S (2010) Structural basis for the major role of O-phosphoseryl-tRNA kinase in the UGA-specific encoding of selenocysteine. Mol Cell 39: 410-420.
10. Turanov AA, Xu XM, Carlson BA, Yoo MH, Gladyshev VN, et al. Biosynthesis of selenocysteine, the 21st amino acid in the genetic code, and a novel pathway for cysteine biosynthesis. Adv Nutr 2: 122-128.

11. Carlson BA, Xu XM, Kryukov GV, Rao M, Berry MJ, et al. (2004) Identification and characterization of phosphoseryl-tRNA[Ser]Sec kinase. Proc Natl Acad Sci U S A 101: 12848-12853.
12. Zinoni F, Heider J, Bock A (1990) Features of the formate dehydrogenase mRNA necessary for decoding of the UGA codon as selenocysteine. Proc Natl Acad Sci U S A 87: 4660-4664.
13. Tormay P, Sawers A, Bock A (1996) Role of stoichiometry between mRNA, translation factor SelB and selenocysteyl-tRNA in selenoprotein synthesis. Mol Microbiol 21: 1253-1259.
14. 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.
15. Barrett EL, Jackson CE, Fukumoto HT, Chang GW (1979) Formate dehydrogenase mutants of Salmonella typhimurium: a new medium for their isolation and new mutant classes. Mol Gen Genet 177: 95-101.
16. Takahata M, Tamura T, Abe K, Mihara H, Kurokawa S, et al. (2008) Selenite assimilation into formate dehydrogenase H depends on thioredoxin reductase in Escherichia coli. J Biochem 143: 467-473.
17. Kotlova N, Ishii TM, Zagryadskaya EI, Steinberg SV (2007) Active suppressor tRNAs with a double helix between the D- and T-loops. J Mol Biol 373: 462-475.
18. Lee K, Varma S, SantaLucia J, Jr., Cunningham PR (1997) In vivo determination of RNA structure-function relationships: analysis of the 790 loop in ribosomal RNA. J Mol Biol 269: 732-743.
19. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31: 3406-3415.
20. Heider J, Baron C, Bock A (1992) Coding from a distance: dissection of the mRNA determinants required for the incorporation of selenocysteine into protein. EMBO J 11: 3759-3766.
21. Juhling F, Morl M, Hartmann RK, Sprinzl M, Stadler PF, et al. (2009) tRNAdb 2009: compilation of tRNA sequences and tRNA genes. Nucleic Acids Res 37: D159-162.
22. Zinoni F, Birkmann A, Leinfelder W, Bock A (1987) Cotranslational insertion of selenocysteine into formate dehydrogenase from Escherichia coli directed by a UGA codon. Proc Natl Acad Sci U S A 84: 3156-3160.
23. Forchhammer K, Leinfelder W, Bock A (1989) Identification of a novel translation factor necessary for the incorporation of selenocysteine into protein. Nature 342: 453-456.
24. Bock A, Forchhammer K, Heider J, Leinfelder W, Sawers G, et al. (1991) Selenocysteine: the 21st amino acid. Mol Microbiol 5: 515-520.
25. Schön A, Bock A, Ott G, Sprinzl M, Söll D (1989) The selenocysteine-inserting opal suppressor serine tRNA from E. coli is highly unusual in structure and modification. Nucleic Acids Res 17: 7159-7165.
26. Itoh Y, Sekine SI, Suetsugu S, Yokoyama S Tertiary structure of bacterial selenocysteine tRNA. Nucleic Acids Res.

27. Chiba S, Itoh Y, Sekine S, Yokoyama S Structural basis for the major role of O-phosphoseryl-tRNA kinase in the UGA-specific encoding of selenocysteine. Mol Cell 39: 410-420.
28. Ishii TM, Kotlova N, Tapsoba F, Steinberg SV The long D-stem of the Selenocysteine tRNA provides Resilience at the expense of Maximal function. J Biol Chem.
29. Kramer GF, Ames BN (1988) Isolation and characterization of a selenium metabolism mutant of Salmonella typhimurium. J Bacteriol 170: 736-743.
30. Baron C, Sturchler C, Wu XQ, Gross HJ, Krol A, et al. (1994) Eukaryotic selenocysteine inserting tRNA species support selenoprotein synthesis in Escherichia coli. Nucleic Acids Res 22: 2228-2233.
31. Betat H, Rammelt C, Morl M tRNA nucleotidyltransferases: ancient catalysts with an unusual mechanism of polymerization. Cell Mol Life Sci 67: 1447-1463.
32. Baron C, Westhof E, Bock A, Giege R (1993) Solution structure of selenocysteine-inserting tRNA(Sec) from Escherichia coli. Comparison with canonical tRNA(Ser). J Mol Biol 231: 274-292.
33. Biou V, Yaremchuk A, Tukalo M, Cusack S (1994) The 2.9 A crystal structure of T. thermophilus seryl-tRNA synthetase complexed with tRNA(Ser). Science 263: 1404-1410.
34. Normanly J, Ollick T, Abelson J (1992) Eight base changes are sufficient to convert a leucine-inserting tRNA into a serine-inserting tRNA. Proc Natl Acad Sci U S A 89: 5680-5684.
35. Baron C, Bock A (1991) The length of the aminoacyl-acceptor stem of the selenocysteine-specific tRNA(Sec) of Escherichia coli is the determinant for binding to elongation factors SELB or Tu. J Biol Chem 266: 20375-20379.
36. Itoh Y, Brocker MJ, Sekine S, Hammond G, Suetsugu S, et al. Decameric SelA*tRNA(Sec) ring structure reveals mechanism of bacterial selenocysteine formation. Science 340: 75-78.
37. Leibundgut M, Frick C, Thanbichler M, Bock A, Ban N (2005) Selenocysteine tRNA-specific elongation factor SelB is a structural chimaera of elongation and initiation factors. EMBO J 24: 11-22.

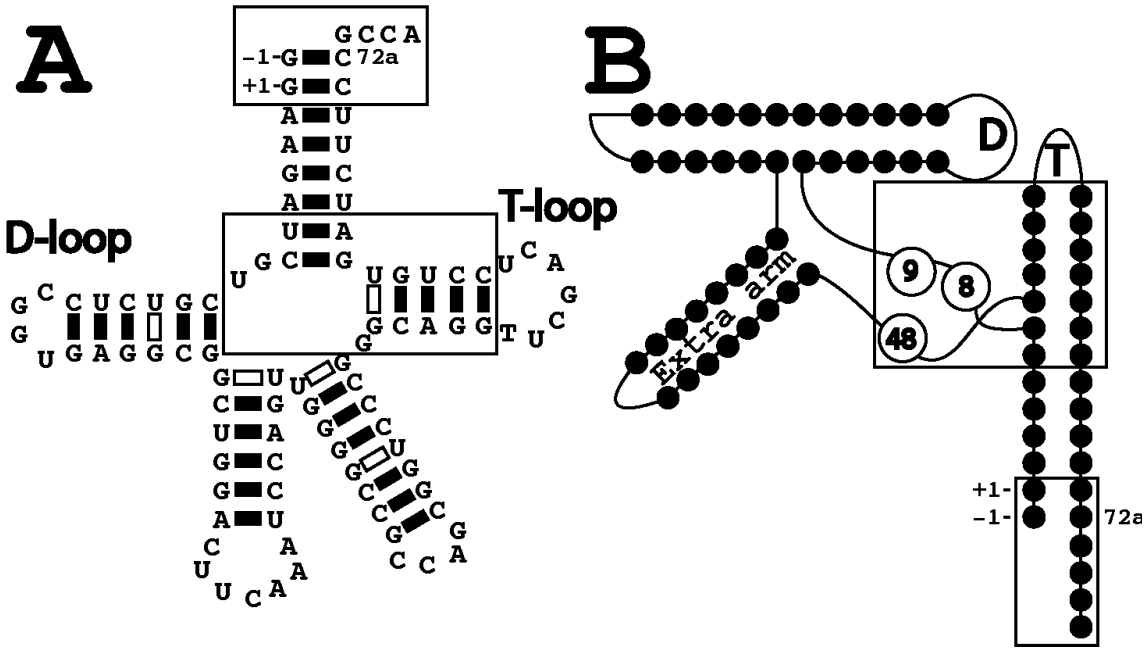


FIGURE 1. The secondary structure of the *E. coli* tRNA^{Sec}. *E. coli* tRNA^{Sec} presented as the cloverleaf (Baron et al., 1993) (A) and as the L-form (B). Nucleotides -1, +1, and 72a are indicated. The regions in the acceptor/T domain that have been modified in this work are boxed. (B): Nucleotides 8, 9 and 48, which in regular tRNAs are involved in tertiary interactions within the D-domain, are indicated. Nucleotides of the loops are not shown.

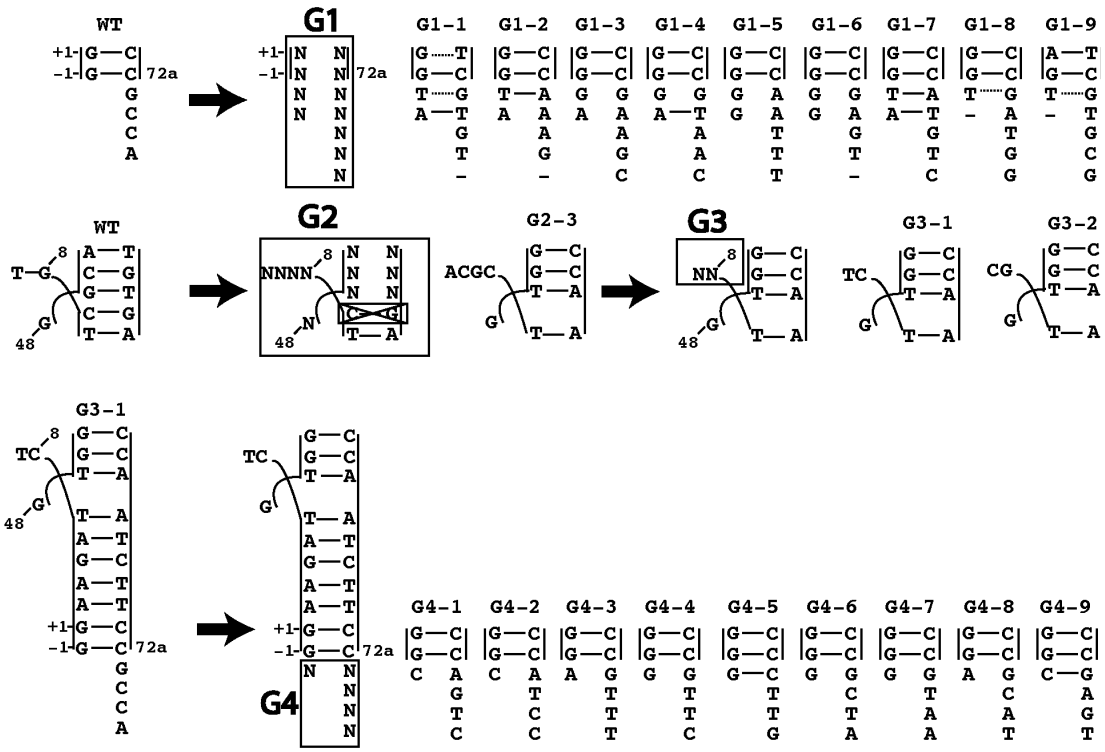


FIGURE 2. The results of the screening from combinatorial libraries G1, G2, G3 and G4. Arrows lead from the targeted region of the $tRNA^{Sec}$ variant for randomization (see Figure 1) to the library design. In libraries G1 and G4, we modified the region at the end of the acceptor stem, while in libraries G2 and G3, we modified the T-stem and the nucleotides that connect the acceptor/T domain to the rest of the molecule. N stands for a fully randomized nucleotide position. To the right of each library design, some of the screened variants are shown. The cross-checked box stands for the base pair eliminated for the design of library G2. The designs of all libraries and sequences of all variants, in the context of the whole nucleotide sequence, are shown in the Supplementary Data.

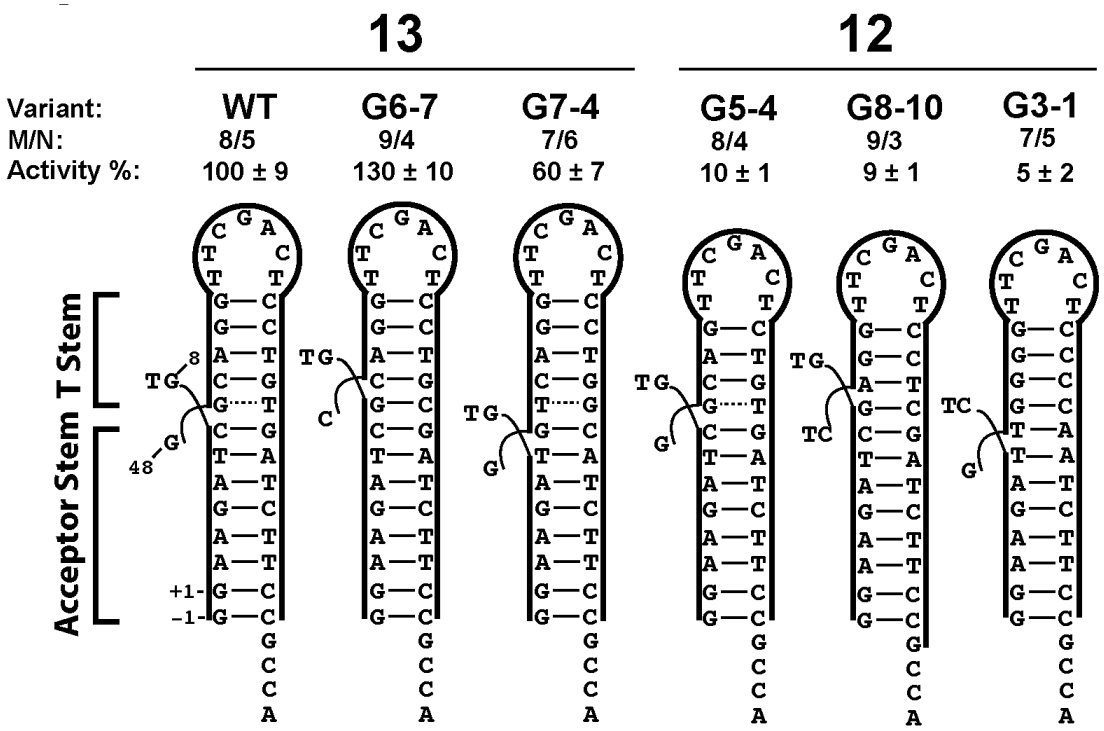


FIGURE 3. The structure of the acceptor/T domain in various tRNA^{Sec} variants.

The variants containing 13 and 12 bp in the acceptor/T domain are shown on the left and on the right, respectively. For each variant, the name, the type of the structure in the M/N notation form (M and N stand for the number of bp in the acceptor and T stems, respectively), as well as the FDH_H activity normalized by that of the WT tRNA^{Sec} are provided. While variants having structures 7/5, 8/4, 9/4 and 7/6 emerged due to the specifically arranged library designs, the only variant having the 9/3 structure (variant G8-10) appeared as a result of a spontaneous deletion of the last nucleotide of the T-stem proximate to the acceptor stem.

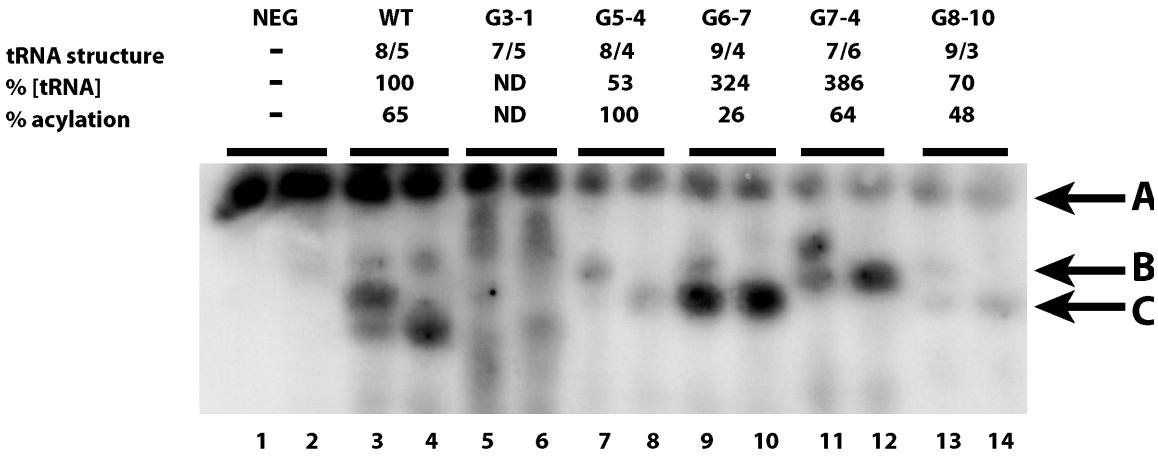


FIGURE 4. The steady state and aminoacylation levels of the screened tRNA^{Sec} variants.

The Northern blot was performed as described before (Kotlova et al., 2007) with some modifications. The negative control containing the void vector is shown in lanes 1 and 2. For each variant, the sample in the even lane was deacylated by incubation with Tris pH 9.0, while the sample in the odd lane was not. Northern blot hybridization was performed using a specific DNA probe complementary to the extra-arm of the tRNA^{Sec} (see Materials and methods). The fact that the aminoacylated form migrated slower than the deacylated form allowed us to evaluate the level of aminoacylation. An additional probe specific to 5S rRNA was used to monitor the amount of total RNA in the samples (indicated by arrow A, see Materials and methods). For each variant, the relative amounts of tRNA are shown as a percentage of that of the WT tRNA^{Sec} (level C, lane 4). For each variant, the aminoacylation level was calculated by dividing the aminoacylated band density (level B, odd lane) by the total tRNA density (level C, even lane). All density calculations were normalized by the amount of 5S (level A). For variant G3-1, the steady state level and the level of aminoacylation could not be determined reliably

(ND). The relative positions of all variants reflect their molecular masses, which were determined based on the known nucleotide sequences (a more detailed explanation is provided in the text).

Supplementary Data:

Supplementary Methods: List of Oligonucleotide used for PCR

Human tRNASec: 5’GCCCGGATGATCCTCAGTGGTCTGGGGTGCAGG CTTCAAA
CCTGTAGCTGTCTAGCGACAGAGTGGTTCAATTCCACCTTTCGGGCGCCA3’
Primers: 5’CGGAATTCGCCCGGATGATCCTCA3’ and 5’TTCCAATGCATTGGCTGCAGTGGCGCCCGAAAGGTG3’

WT:5’GGAAGATCGTCGTCTCCGGTGAGGCGGCTGGACTTCAAATCCAGTTGGGGCCGCCAGCGGTCCCGGGCAGGTTCGACTCCTGTG
ATCTTCCGCCA3’
Primers: 5’CGGAATTCGGAAGATC3’ and 5’TTCCAATGCATTGGCTGCAGTGGCGGAAGATCACAGGAGTCGAACCTGC3’

Library G1:
5’GTCGTCTCCGGTGAGGCGGCTGGACTTCAAATCCAGTTGGGGCCGCCAGCGGTCCCGGGCAGGTTCGACTCCTGTGATCTTNNNNNN
NCTGCAGCCAATG3’
Primers: 5’CGGAATTCAANNNNAAGATCGTCGTCTCCGGTGA3’and
5’CATTGGCTGCAGNNNNNNNAAGATCACAGGAGTCGAACCTGCCC3’

Library G2:
5’GGAAGATNNNNCGTCTCNNGGNNGAGGCGGCTGGACTTCAAATCCAGTTGGGGCCGCCAGCGGTCCCGNNNNGGTTCGACTCCNNNA
TCTTCCGCCA3’
Primers: 5’CGGAATTCGGGAAGATNNNNCGTCTC3’and
5’TTCCAATGCATTGGCTGCAGTGGCGGAAGATNNNGGAGTCGAACCNNNNNNCGGGACCG3’

Library G3:
5’GGAAGATNNCGTCTCNGGNGAGGCGGCTGGACTTCAAATCCAGTTGGGGCCGCCAGCGGTCCCGGTGGGGTTCGACTCCCCAATCTT
CCGCCA3’
Primers: 5’CGGAATTCAAAGGAAGATNNCGTCTC3’and
5’TTCCAATGCATTGGCTGCAGTGGCGGAAGATTGGGGAGTCGAACCCCACCGGGACCG3’

Library G4:
5’GGAAGATCTCGTCTCCGGTGAGGCGGCTGGACTTCAAATCCAGTTGGGGCCGCCAGCGGTCCCGGTGGGGTTCGACTCCCCAATCTT
CC3’
Primers: 5’CGGAATTCAANGGAAGATCTCGTCTC3’and
5’TTCCAATGCATTGGCTGCAGNNNNGGAAGATTGGGGAGTCGAACCCCACCGGGACCG3’

Library G5:
5’GGAAGATCGTCGTCTCNNGGNNGAGGCGGCTGGACTTCAAATCCAGTTGGGGCCGCCAGCGGTCCCGGGCAGTTCNNNTCTGTGATC
TTCCGCCA3’
Primers: 5’CGGAATTCAAAGGAAGATCGTCGTCTC3’and
5’TTCCAATGCATTGGCTGCAGTGGCGGAAGATCACAGAGTCGAACTGCCCGGGACCG3’

Library G6:
5’GGAAGATCRGTCGTCTCNCGGTNGAGGCGGCTGGACTTCAAATCCAGTRGGGGCCGCCAGCGGTCCCYYCAGGTTCGACTCCTGYGA
TCTTCCGCCA3’
Primers: 5’CGGAATTCAAAGGAAGATC3’and 5’TTCCAATGCATTGGCTGCAGTGGCGGAAGATC3’

Library G7:
5’GGAAGATGTCGTCTCNNGGNNGAGGCGGCTGGACTTCAAATCCAGTTGGGGCCGCCAGCGGTCCCGGNNNAGGTTCGACTCCTNNNA
TCTTCCGCCA3’
Primers: 5’CGGAATTCAAAGGAAGATGTCGTCTC3’ and
5’TTCCAATGCATTGGCTGCAGTGGCGGAAGATNNNAGGAGTCGAACCTNNNCCGGGACCG3’

Library G8:
5’GGAAGATCRGTCGTCTCNCGGTNGAGGCGGCTGGACTTCAAATCCAGTTGGGGCCGCCAGCGGTCCCGYCAGGTTCGACTCCTGYGA
TCTTCCGCCA3’
Primers: Same ones used for Library G8

Library G9:
5’GGAAGANCTCGTCTCCGGTGAGGCGGCTGGACTTCAAATCCAGTTGGGGCCGCCAGCGGTCCCGNGGGGTTCGACTCCCCNNTCTTC
CGCCA3’
Primers: 5’CGGAATTCAAAGGAAGANCTCGTCTC3’and
5’TTCCAATGCATTGGCTGCAGTGGCGGAAGANNGGGGAGTCGAACCCCNCCGGGACCG3’

Library G10:
5’GGAAGANNNGTCGTCTCGCGGTCGAGGCGGCTGGACTTCAAATCCAGTTGGGGCCGCCAGCGGTCCCGGNNAGTTCGACTCTNNNNN
TCTTCCGCCA3’
Primers: 5’CGGAATTCAAAGGAAGANNNGTCGTCTC3’and
5’TTCCAATGCATTGGCTGCAGTGGCGGAAGANNNNNAGAGTCGAACTNNCCGGGACCG3’

Table S1: Sequence of WT, Human tRNASec, and designs of libraries

ACCEPTOR-STEM

			C1	D-STEM			ANTICODON-STEM			EXTRA-ARM			C2	T-STEM				
WT		GGAAGATC	GT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAGG	TTCGACT	CCTGT	GATCTTCC	GCCA
HUMAN		GCCCGGATG	AT	CCTCAG	TGGT	CTGGGG	TGCAGG	CTTCAAA	CCTGTA	GCTGTC	TAGC	GACAG	A	GTGG	TTCAATT	CCAC	CTTTCGGGC	GCCA
LIB.G1	NN	NNAAGATC	GT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAGG	TTCGACT	CCTGT	GATCTTNN	NNNNN
LIB.G2		GGAAGAT	NNNN	CGTCTC	NNGGNN	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	N	NNNGG	TTCGACT	CCNNN	ATCTTCC	GCCA
LIB.G3		GGAAGAT	NN	CGTCTC	NGGN	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	TGGGG	TTCGACT	CCCCA	ATCTTCC	GCCA
LIB.G4	N	GGAAGAT	CT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	TGGGG	TTCGACT	CCCCA	ATCTTCC	NNNN
LIB.G5		GGAAGATC	GT	CGTCTC	NNGGNN	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAG	TTCGACT	CTGT	GATCTTCC	GCCA
LIB.G6		GGAAGATCR	GT	CGTCTC	NCGGTN	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	RGGGGCCG	CCAG	CGGTCCCY	Y	CAGG	TTCGACT	CCTG	YGATCTTCC	GCCA
LIB.G7		GGAAGAT	GT	CGTCTC	NNGGNN	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	NNNAGG	TTCGACT	CCTNNN	ATCTTCC	GCCA
LIB.G8		GGAAGATCR	GT	CGTCTC	NCGGTN	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	Y	CAGG	TTCGACT	CCTG	YGATCTTCC	GCCA
LIB.G9		GGAAGAN	CT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG		NGGGG	TTCGACT	CCCCN	NTCTTCC	GCCA
LIB.G10		GGAAGANNN	GT	CGTCTC	GCGGTC	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	NNAG	TTCGACT	CTNN	NNNTCTTCC	GCCA

Table S2: Library G1

LIB.G1	NN	NNAAGATC	GT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAGG	TTCGACT	CCTGT	GATCTTNN	NNNNN
G1-1	AT	GGAAGATC	GT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAGG	TTCGACT	CCTGT	GATCTTTC	GTGT
G1-2	AT	GGAAGATC	GT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAGG	TTCGACT	CCTGT	GATCTTCC	AAAG
G1-3	AG	GGAAGATC	GT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAGG	TTCGACT	CCTGT	GATCTTCC	GAAGC
G1-4	GG	GGAAGATC	GT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAGG	TTCGACT	CCTGT	GATCTTCC	GTAAC
G1-5	GG	GGAAGATC	GT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAGG	TTCGACT	CCTGT	GATCTTCC	AATTT
G1-6	GG	GGAAGATC	GT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGTCG	CCAG	CGGTCCCG	G	GCAGG	TTCGACT	CCTGT	GATCTTCC	GAGT
G1-7	AT	GGAAGATC	GT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAGG	TTCGACT	CCTGT	GATCTTCC	ATGTC
G1-8	T	GGAAGATC	GT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAGG	TTCGACT	CCTGT	GATCTTCC	GATGG
G1-9	T	GAAAGATC	GT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAGG	TTCGACT	CCTGT	GATCTTTC	GTGCG

Table S3: Library G2

LIB.G2	GGAAGAT	NNNN	CGTCTC	NNGGNN	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	N	NNNGG	TTCGACT	CCNNN	ATCTTCC	GCCA
G2-1	GGAAGAT	TGCT	CGTCTC	GCGGCT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	AGAGG	TTCGACT	CCTCC	ATCTTCC	GCCA
G2-2	GGAAGAT	GGAA	CGTCTC	GCGGTA	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	T	GGTGG	TTCGACT	CCCCA	ATCTTCC	GCCA
G2-3	GGAAGAT	CGCA	CGTCTC	GGGGTG	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	TGGGG	TTCGACT	CCCCA	ATCTTCC	GCCA

Table S4: Library G3

LIB.G3	GGAAGAT	NN	CGTCTC	NGGN	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	TGGGG	TTCGACT	CCCCA	ATCTTCC	GCCA
G3-1	GGAAGAT	CT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	TGGGG	TTCGACT	CCCCA	ATCTTCC	GCCA
G3-2	GGAAGAT	GC	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGGCCGCCAG		CGGTCCCG	G	TGGGG	TTCGACT	CCCCA	ATCTTCC	GCCA

Table S5: Library G4

LIB.G4	N	GGAAGAT	CT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	TGGGG	TTCGACT	CCCCA	ATCTTCC	NNNN
G4-1	C	GGAAGAT	CT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	TGGGG	TTCGACT	CCCCA	ATCTTCC	AGTC
G4-2	C	GGAAGAT	CT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	TGGGG	TTCGACT	CCCCA	ATCTTCC	ATCC
G4-3	A	GGAAGAT	CT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	TGGGG	TTCGACT	CCCCA	ATCTTCC	GTTT
G4-4	G	GGAAGAT	CT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	TGGGG	TTCGACT	CCCCA	ATCTTCC	GTTC
G4-5	G	GGAAGAT	CT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	TGGGG	TTCGACT	CCCCA	ATCTTCC	CTTG
G4-6	G	GGAAGAT	CT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	TGGGG	TTCGACT	CCCCA	ATCTTCC	GCTA
G4-7	G	GGAAGAT	CT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	TGGGG	TTCGACT	CCCCA	ATCTTCC	GTAA
G4-8	A	GGAAGAT	CT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	TGGGG	TTCGACT	CCCCA	ATCTTCC	GCAT
G4-9	C	GGAAGAT	CT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	TGGGG	TTCGACT	CCCCA	ATCTTCC	GAGT

Table S6: Library G5

LIB.G5	GGAAGATC	GT	CGTCTC	NNGGNN	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAG	TTCGACT	CTGT	GATCTTCC	GCCA
G5-1	GGAAGATC	GT	CGTCTC	GCGGTC	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAG	TTCGACT	CTGT	GATCTTCC	GCCA
G5-2	GGAAGATC	GT	CGTCTC	ACGGTT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAG	TTCGACT	CTGT	GATCTTCC	GCCA
G5-3	GGAAGATC	GT	CGTCTC	CGGGTG	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAG	TTCGACT	CTGT	GATCTTCC	GCCA
G5-4	GGAAGATC	GT	CGTCTC	GCGGTG	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GCAG	TTCGACT	CTGT	GATCTTCC	GCCA

Table S7: Library G6

LIB.G6	GGAAGATCR	GT	CGTCTC	NCGGTN	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	RGGGGCCG	CCAG	CGGTCCCY	Y	CAGG	TTCGACT	CCTG	YGATCTTCC	GCCA
G6-1	GGAAGATCA	GT	CGTCTC	ACGGTT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	GGGGGCCG	_CAG	CGGTCCCT	T	CAGG	TTCGACT	CCTG	TGATCTTCC	GCCA
G6-2	GGAAGATCA	GT	CGTCTC	TCGGTA	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	GGGGGCCG	CCAG	CGGTCCCT	T	CAGG	TTCGACT	CCTG	TGATCTTCC	GCCA
G6-3	GGAAGATCA	GT	CGTCTC	TCGGTA	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	AGGGGCCG	CCAG	CGGTCCCT	T	CAGG	TTCGACT	CCTG	TGATCTTCC	GCCA
G6-4	GGAAGATCG	GT	CGTCTC	TCGGTG	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	AGGGGCCG	CCAG	CGGTCCCT	T	CAGG	TTCGACT	CCTG	TGATCTTCC	GCCA
G6-5	GGAAGATCG	GT	CGTCTC	GCGGTA	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	AGGGGCCG	CCAG	CGGTCCCT	T	CAGG	TTCGACT	CCTG	TGATCTTCC	GCCA
G6-6	GGAAGATCG	GT	CGTCTC	TCGGTA	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	AGGGGCCG	CCAG	CGGTCCCC	C	CAGG	TTCGACT	CCTG	TGATCTTCC	GCCA
G6-7	GGAAGATCG	GT	CGTCTC	GCGGTC	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	GGGGGCCG	CCAG	CGGTCCCT	C	CAGG	TTCGACT	CCTG	CGATCTTCC	GCCA
G6-8	GGAAGATCG	GT	CGTCTC	CCGGTG	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	GGGGGCCG	CCAG	CGGTCCCC	T	CAGG	TTCGACT	CCTG	TGATCTTCC	GCCA

Table S8: Library G7

LIB.G7	GGAAGAT	GT	CGTCTC	NNGGNN	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	NNNAGG	TTCGACT	CCTNNN	ATCTTCC	GCCA
G7-1	GGAAGAT	GT	CGTCTC	CTGGTG	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GACAGG	TTCGACT	CCTGTA	ATCTTCC	GCCA
G7-2	GGAAGAT	GT	CGTCTC	TCGGTC	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GAAAGG	TTCGACT	CCTTCC	ATCTTCC	GCCA
G7-3	GGAAGAT	GT	CGTCTC	ATGGTC	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	ACCAGG	TTCGACT	CCTGGC	ATCTTCC	GCCA
G7-4	GGAAGAT	GT	CGTCTC	AGGGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GTCAGG	TTCGACT	CCTGGC	ATCTTCC	GCCA
G7-5	GGAAGAT	GT	CGTCTC	GCGGCA	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	TGAAGG	TTCGACT	CCTTCC	ATCTTCC	GCCA

Table S9: Library G8

LIB.G8	GGAAGATCR	GT	CGTCTC	NCGGTN	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	Y	CAGG	TTCGACT	CCTG	YGATCTTCC	GCCA
G8-1	GGAAGATCG	GTT	CGTCTC	GCGGTA	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGGCCG	CCAG	CGGTCCCG	T	CAGG	TTCGACT	CCTG	TGATCTTCC	GCCA
G8-2	GGAAGATCG	GT	CGGTCTC	TCGGTG	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	T	CAGG	TTCGACT	CCTG	TGATCTTCC	GCCA
G8-3	GGAAGATC	A	GTCTCA	CGGT	TGAGGC	GGCTGGA	CTTCAAA	TCCAGTT	GGGGCCG	CCAG	CGGTCCC	GC	CAGG	TTCGACT	CCTG	CGATCTTCC	GCCA
G8-4	GGAAGATCG	GT	CGTCTC	CCGGTG	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCC_G	C	CAGG	TTCGACT	CCTG	CGATCTTCC	GCCA
G8-5	GGAAGATCG	GT	CGTCTC	GCGGTC	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	C	CAGG	TTCGACT	CCTG	CGATCTTCC	GCCA
G8-6	GGAAGATCG	GT	CGTCTC	GCGGTC	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	_	CAGG	TTCGACT	CCTG	CGATCTTCC	GCCA
G8-7	GGAAGATCG	GT	CGGTCTC	TCGGTA	GAGGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	C	CAGG	TTCGACT	CCTG	CGATCTTCC	GCCA
G8-8	GGAAGATCAG	GT	CGTCTC	GCGGTA	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	T	CAGG	TTCGACT	CCTG	CGATCTTCC	GCCA
G8-9	GGAAGATCA	GT	CGTCTTC	TCGGTT	GAGGCGG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	T	CAGG	TTCGACT	CCTG	CGATCTTCC	GCCA
G8-10	GGAAGATCG	GT	CGTCTC	TCGGTA	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	TC	AGG	TTCGACT	CCT_	CGATCTTCC	GCCA

Table S10: Library G9

LIB.G9	GGAAGAN	CT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG		NGGGG	TTCGACT	CCCCN	NTCTTCC	GCCA
G9-1	GGAAGAT	CT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	AGGGG	TTCGACT	CCCCA	ATCTTCC	GCCA
G9-2	GGAAGAA	CT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG		TGGGG	TTCGACT	CCCCT	CTCTTCC	GCCA
G9-3	GGAAGAG	CT	CGTCTC	CGGT	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	AGGGG	TTCGACT	CCCCT	CTCTTCC	GCCA

Table S11: Library G10

LIB.G10	GGAAGANNN	GT	CGTCTC	GCGGTC	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	NNAG	TTCGACT	CTNN	NNNTCTTCC	GCCA
G10-1	GGAAGAAGG	GT	CGTCTC	GCGGTC	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	GTAG	TTCGACT	CTCC	ACTTCTTCC	GCCA
G10-2	GGAAGATGA	GT	CGTCTC	GCGGTC	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	ACAG	TTCGACT	CTGA	TCGTCTTCC	GCCA
G10-3	GGAAGATGG	GT	CGTCTC	GCGGTC	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	CAAG	TTCGACT	CTTG	ACATCTTCC	GCCA
G10-4	GGAAGATAC	GT	CGTCTC	GCGGTC	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	TGAG	TTCGACT	CTCA	ATATCTTCC	GCCA
G10-5	GGAAGAGGG	GT	CGTCTC	GCGGTC	GAGGCG	GCTGGA	CTTCAAA	TCCAGT	TGGGGCCG	CCAG	CGGTCCCG	G	TGAG	TTCGACT	CTCA	CCCTCTTCC	GCCA

Figure Legend to Tables S1-S11

List of all screened clones. The name of the clone is followed by the DNA sequence of the $tRNA^{Sec}$. The designed library is found at the top with the name starting with 'LIB'. The representation of the following symbols are described as follows:C1, connector 1; C2, connector 2; G, Guanine; A, Adenine;T, Thymine;C, Cytosine; _ , Empty;N, Any; R, Purine; Y, Pyrimidine; Black, non-randomized regions; Red, Randomized regions.

Supplementary Note: Comment on the Methods assessing selenocysteine incorporation ability of variants of $tRNA^{Sec}$

The assessment of active $tRNA^{Sec}$ variants proceeded in two steps: the primary screening and activity measurement. Each step was based on the detection of the activity of the particular selenoprotein, FDH_N and FDH_H, respectively. One can see, however, that the results obtained using these two tests have not always been fully consistent with each other. In particular, according to the FDH_N-test, variants G5-4, G8-10, and G3-1, having 12 bp in the acceptor/T domain, formed bright-white colonies and thus were more active than the human $tRNA^{Sec}$, which formed a pink colony. However, the quantification of the $tRNA^{Sec}$ efficiency with use of the FDH_H-test demonstrated that the human $tRNA^{Sec}$ was somewhat more active than these variants (11% versus 10%, 9%, and 5%, respectively). Such selenoprotein gene dependent Sec incorporation by mutants of $tRNA^{Sec}$ has been known (8,11) and justifies the necessity of the simultaneous use of both of them for a more complete characterization of the screened variants. The existence of this discrepancy may reflect differences in the particular contexts surrounding the selenocysteine UGA codons in the mRNA transcripts of the two selenoprotein genes. How exactly these differences can affect the efficiency of the selenocysteine incorporation should be a matter of further analysis.

CHAPTER VI

Sel A measures the length of the acceptor/T domain of $tRNA^{Sec}$

This chapter was originally part of the paper (chapter V) submitted for consideration in JBC, but was left out in the most recent submission to Plos one. Nevertheless, I feel the findings here are significant and have therefore been included as part of this thesis as a separate chapter and should flow relatively smoothly from the previous chapter.

In the last chapter we analyzed the constraints imposed on the structure of the acceptor/T domain in functional variants of the $tRNA^{Sec}$. Using the combinatorial library search, we generated $tRNA^{Sec}$ variants with non-standard lengths of the acceptor and T-stems. In these variants, the length of the acceptor stem varied between 7 and 9 bp, while the length of the acceptor/T domain was either 12 or 13 bp. Although in some variants the lengths of the acceptor stem and/or of the acceptor/T domain were different from those observed in the WT $tRNA^{Sec}$, all variants demonstrated the ability to specifically incorporate selenocysteine *in vivo* at the assigned codons of two selenoproteins FDH_N and FDH_H. Ten arbitrarily chosen variants were shown to be SelB-dependent, which indicated that they incorporate selenocysteine into the polypeptide chains using the same functional path as the WT $tRNA^{Sec}$. Given that the SelB-dependence was observed in all ten variants tested, we can expect that most of the other screened variants were SelB-dependent as well. If a SelB-independent $tRNA^{Sec}$ variant were found, it would have meant that it was delivered to the ribosome as a regular elongator tRNA, using EF-Tu. Whether such variants can be found using the approach exercised here remains unknown. However, even if such variants existed, they would not have changed the conclusions from the previous chapter, which

were made based on the analysis of only those variants whose SelB-dependence had been demonstrated.

The inability of the measured variants to function as regular elongator tRNAs raises some questions. Previously, it was shown that the GU base pair existing in the T-stem of the WT $tRNA^{Sec}$ (Figure 1, Chapter V) could serve as an anti-determinant with respect to EF-Tu(Rudinger et al., 1996). The fact that such GU base pair existed only in some of the measured clones and still, all measured clones were SelB-dependent indicates that the structural constraints that prevent $tRNA^{Sec}$ to function as a normal elongator tRNA are more complex than was originally thought. The elucidation of these constraints can be a matter of further analysis.

In spite of the fact that all screened variants were able to incorporate selenocysteine at the specific codons of the two reporting selenoproteins, the activities of the variants were markedly different. The difference started with the presence of these variants in the cytosol. The Northern blot presented in Figure 4 (Chapter V) demonstrated that variants containing 13 bp in the acceptor/T domain always had a notably higher presence in the cell than variants containing only 12 bp, regardless of the particular length of the acceptor stem. This observation infers that the efficiency of the synthesis and/or maintenance of Sec-$tRNA^{Sec}$ variants depends on the length of the acceptor/T domain. It also suggests the existence of a supervising system filtering out those $tRNA^{Sec}$ variants in which this domain is not long enough. Most probably, this supervising system involves the proteins that participate in the synthesis and/or maintenance of the Sec-$tRNA^{Sec}$. The three most probable candidates for this role are SerRS, SelA and SelB. While the first two proteins are directly involved in the synthesis of the Sec-$tRNA^{Sec}$, the third one specifically stabilizes this molecule through binding to its acceptor stem and terminus, as well as to the selenocysteinyl residue. The

knowledge of the crystal structures of all three proteins allows us to suggest the mechanism of such selection.

The crystal structure of the SerRS complex with the $tRNA^{Ser}$ (Biou et al., 1994) shows a tight and specific interaction of SerRS with the tip of the tRNA T-loop. Unfortunately, in this structure the end of the acceptor stem of the $tRNA^{Ser}$ has not been properly resolved, which does not allow us to see how tightly and specifically it interacts with SerRS. However, additional biochemical studies showed that the discriminator nucleotide 73, which is adjacent to the acceptor stem, represents an important identity element of the $tRNA^{Ser}$ with respect to SerRS (Figure 1a)(Normanly et al., 1992). Such role of nucleotide 73 presumes the existence of a tight interaction between the nucleotide and the protein. The simultaneous interaction of SerRS with the tip of the T-loop and with nucleotide 73 should be sensitive to the juxtaposition of the two elements in the tRNA structure and, therefore, to the length of the whole acceptor/T domain. Given that in the WT $tRNA^{Sec}$, the acceptor/T domain is longer than in the $tRNA^{Ser}$, it is not surprising that the efficiency of the serylation of the $tRNA^{Sec}$ is about 100 times lower than of the $tRNA^{Ser}$ (Baron and Bock, 1991). Therefore, if the serylation of the $tRNA^{Sec}$ were responsible for the difference in the cytosolic level of the measured variants of the $tRNA^{Sec}$, we would have expected that variants having 12 bp in the acceptor/T domain be present in higher quantities than those having 13 bp. Given that our results show the opposite, it is unlikely that SerRS plays any role in the observed difference.

The recently published crystal structure of the SelA-$tRNA^{Sec}$ complex (Itoh et al. 2013) shows that the region of the SelA contacting the acceptor stem of the $tRNA^{Sec}$ is flexible and that its interaction with this stem is neither strong nor specific. Such flexibility undermines the ability of SelA to distinguish the $tRNA^{Sec}$ based on the length of the acceptor

stem. On the other hand, according to the recently published paper of Aldag et al., (Aldag et al. 2013), a notable selenylation level can be achieved only if the $tRNA^{Sec}$ variant contains an additional eighth base pair in the acceptor stem. Together, these two observations lead us to the suggestion that it is not the length of the acceptor stem per se, but the length of the whole acceptor/T domain that serves for the recognition of the $tRNA^{Sec}$ by SelA. Indeed, as it is shown in Figure 1b, both the tip of the T-loop and the discriminator nucleotide immediately preceding the 3'-terminal CCA-sequence form specific interactions with SelA. Therefore, the interaction of the $tRNA^{Sec}$ with SelA should be sensitive to the length of the whole acceptor/T domain. Thus, even though SelA seems to be unable to recognize the $tRNA^{Sec}$ through the measurement of the length of the acceptor stem, it probably can do it by checking the length of the whole acceptor/T domain. When the two extremities of this domain are bound to SelA, the interaction made by four nucleotides of the acceptor stem with SelA (shown in Figure 1b) will stabilize the protein-tRNA interaction in a non-specific manner.

Although SerRS and SelA are unrelated enzymes, both of them interact with the tip of the T-loop and with the discriminator nucleotide of the tRNA. Due to the latter aspect, both enzymes should be able to sense the length of the acceptor/T domain. SerRS preferably recognizes a 12bp domain, while SelA would prefer a 13bp domain. Interestingly, in none of the cases the recognition is absolute; SerRS is also able to aminoacylate the WT $tRNA^{Sec}$, having 13 bp in the acceptor/T domain, while SelA, according to results presented here, can selenylate $tRNA^{Sec}$ variants in which this domain contains only 12 bp. For both SerRS and SelA, the alternative way of functioning seems to be notably less efficient than the primary one.

Although currently, there is no available crystal structure for a tRNA*SelB complex, a convincing model of such complex was built based on superimposition of the crystal conformation of the archaeal SelB, a homolog of the bacterial SelB, on the conformation of EF-Tu within the crystal structure EF-Tu*tRNA^{Cys} (Leibundgut et al., 2005; Nissen et al., 1999). According to this model, SelB forms interactions with the three terminal base pairs of the acceptor stem and with the part of the T-stem proximate to the acceptor stem (Figure 1c). Given that outside this area SelB does not form any contacts with the acceptor/T domain, it should be unable to distinguish between tRNA^{Sec} variants having 12 and 13 base pairs in the whole acceptor/T domain. In particular, SelB would be expected to bind equally well to the WT and to variant G5-4 (Figure 3, chapter V) as the two molecules have the same nucleotide sequence of the acceptor/T domain, except for the CG bp proximate to the T-loop, which is missed in variant G5-4. In spite of this, the WT tRNA^{Sec} has a notably higher presence in the cell than variant G5-4 (Figure 4, chapter V). Therefore, factor SelB, like the SerRS, is unlikely to be primarily responsible for the observed differences in the presence of the tRNA^{Sec} variants in the cytosol.

The analysis of the structures of the three protein-tRNA complexes leads us to the conclusion that out of the three proteins involved in the synthesis and delivery of Sec-tRNA^{Sec}, SelA is the only one that is able to verify the length of the acceptor/T domain. We therefore suggest that it is factor SelA that is primarily responsible for a lower cellular level of Sec-tRNA^{Sec} having only 12 bp in the acceptor/T domain. The role of SelB, however, should not be underestimated. Even though SelB is unable to distinguish tRNA^{Sec} variants based on the length of the acceptor/T domain, it still can participate in the tRNA selection. Given that SelB binds only those variants that carry selenocysteine (Forchhammer et al., 1989), the inability of tRNA^{Sec} variants having 12 bp in the acceptor/T domain to be effectively selenylated will prevent their association with SelB, thus leading to degradation.

Another step of the $tRNA^{Sec}$ functional cycle in which the length of the acceptor stem and/or of the whole acceptor/T domain can make a difference is when $tRNA^{Sec}$ functions at the ribosome. The normalization of the activities of the measured clones by the presence in the cytosol demonstrated a clear preference for the 8/5 and 9/4 variants, while the 7/6 variant, despite the long acceptor/T domain, has about the same activity as the variants having only 12 bp in this domain. This observation indicates that the particular length of the acceptor stem can also be important for the optimal functioning of the $tRNA^{Sec}$ and corroborates the fact that in all domains of Life, the WT $tRNAs^{Sec}$ have always either the 8/5 or 9/4 structure. Despite this, it is undeniable that neither the requirements for the length of the acceptor stem nor the requirements for the length the acceptor/T domain are absolutely rigid: functional $tRNA^{Sec}$ variants having structures 7/6, 7/5, 8/4 and 9/3 can demonstrate notable Sec-incorporating activity. It is quite interesting that the presence of an extra bp in the anticodon can be required in different way at different steps of the functional cycle.

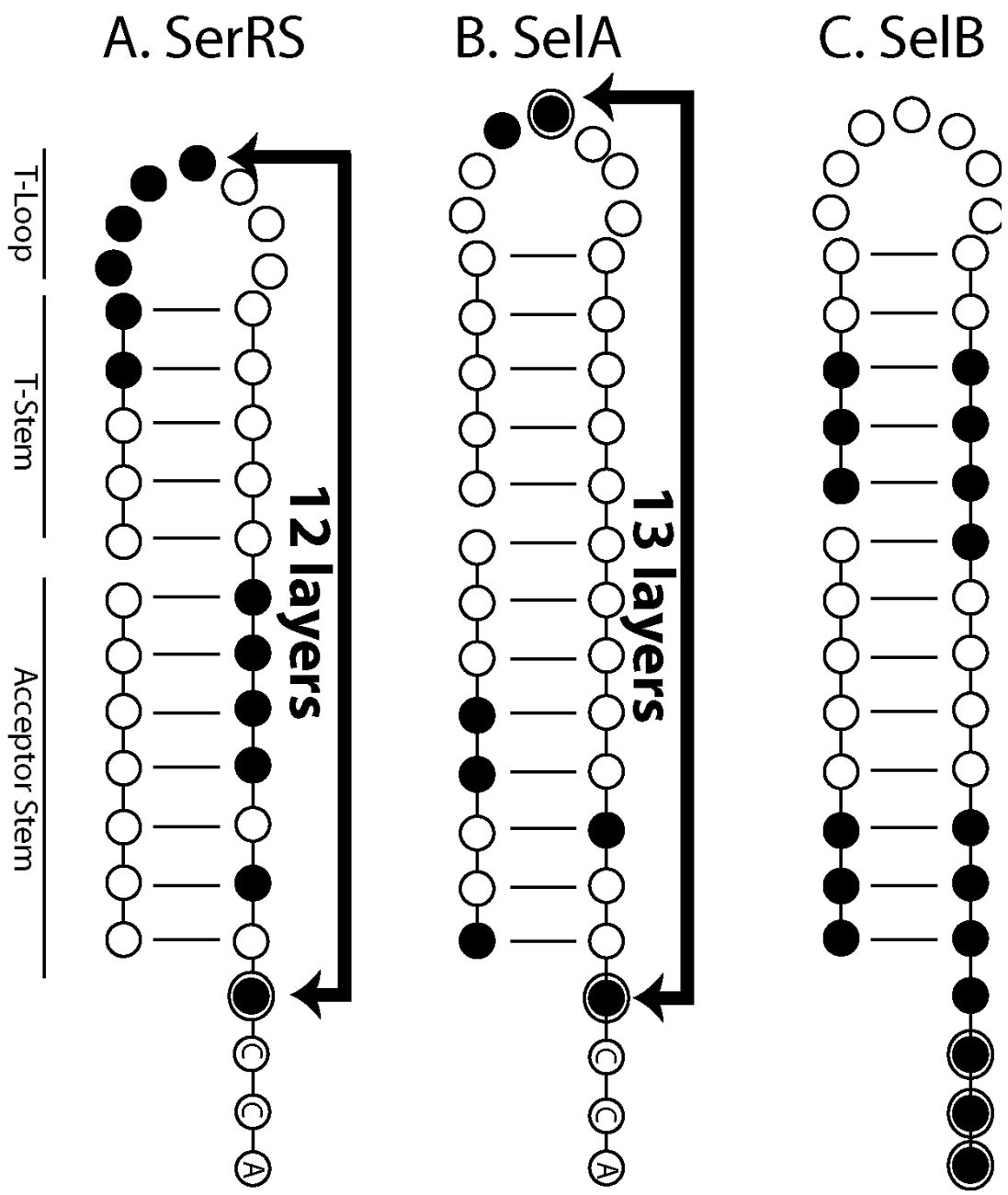


FIGURE 1. Nucleotides of the acceptor/T domain that are involved in interaction with proteins SerRS (A), SelA (B), and SelB (C).

Open, filled and double circles stand for nucleotides that do not contact the protein (open circles), form contacts only through the sugar-phosphate backbone (filled circles) or with participation of bases (double circles). Horizontal lines between the circles indicate base pairing. A: the scheme of the tRNA-SerRS interactions was deduced from the crystal structure of the SerRS-$tRNA^{Ser}$ complex (Biou et al., 1994), but also included the information

concerning the $tRNA^{Ser}$ identity elements (Normanly et al., 1992). The requirement to simultaneously maintain the interactions with the T-loop and with the discriminator nucleotide-73 would make SerRS sensitive to a 12-bp length of the acceptor/T domain, which explains the low serylation efficiency of the $tRNA^{Sec}$ (Baron and Bock, 1991). B: The scheme of the $tRNA^{Sec}$-SelA interactions was deduced from the available crystal structure of the corresponding complex (Itoh et al. 2013). Factor SelA specifically interacts with nucleotide 57 at the top of the T-loop and with the discriminator nucleotide adjacent to the acceptor stem. Like for the SerRS, the necessity to maintain both interactions should make the activity of SelA dependent to the total length of the acceptor/T domain. However, factor SelA, unlike SerRS, is expected to prefer the 13-bp acceptor/T domain present in the $tRNA^{Sec}$. The additional non-specific interactions between SelA and the acceptor stem serve for general stabilization of the tRNA-protein interaction (see the text) C: the scheme of the $tRNA^{Sec}$–SelB interaction proposed based on the crystal structure of the archaeal SelB (Leibundgut et al., 2005) (the details are provided in the text). Factor SelB interacts neither with the T-loop, nor with the proximate part of the T-stem. Therefore, its interaction with the $tRNA^{Sec}$ should be insensitive to the length of the acceptor/T domain.

References

Aldag, C., Brocker, M.J., Hohn, M.J., Prat, L., Hammond, G., Plummer, A., and Söll, D (2013). Rewiring translation for elongation factor Tu-dependent selenocysteine incorporation. Angew Chem Int Ed Engl *52*, 1441-1445.

Baron, C., and Bock, A. (1991). The length of the aminoacyl-acceptor stem of the selenocysteine-specific tRNA(Sec) of Escherichia coli is the determinant for binding to elongation factors SELB or Tu. J Biol Chem *266*, 20375-20379.

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

Forchhammer, K., Leinfelder, W., and Bock, A. (1989). Identification of a novel translation factor necessary for the incorporation of selenocysteine into protein. Nature *342*, 453-456.

Itoh, Y., Brocker, M.J., Sekine, S., Hammond, G., Suetsugu, S., Söll, D., and Yokoyama, S (2013). Decameric SelA*tRNA(Sec) ring structure reveals mechanism of bacterial selenocysteine formation. Science *340*, 75-78.

Leibundgut, M., Frick, C., Thanbichler, M., Bock, A., and Ban, N. (2005). Selenocysteine tRNA-specific elongation factor SelB is a structural chimaera of elongation and initiation factors. EMBO J *24*, 11-22.

Nissen, P., Thirup, S., Kjeldgaard, M., and Nyborg, J. (1999). The crystal structure of Cys-tRNACys-EF-Tu-GDPNP reveals general and specific features in the ternary complex and in tRNA. Structure *7*, 143-156.

Normanly, J., Ollick, T., and Abelson, J. (1992). Eight base changes are sufficient to convert a leucine-inserting tRNA into a serine-inserting tRNA. Proc Natl Acad Sci U S A *89*, 5680-5684.

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

CHAPTER VII: Discussion and Conclusion

The studies in chapter II demonstrated that tRNAs with T-loops different from the standard, with no RHUA in the T-loop, can still function *in vivo*. Moreover, the attraction between the D- and T-loops, normally provided by the purine trap, was shown to be replaceable by ILDH. Further analysis, described in chapter III, led us to the model for the alternative to the T-loop, type III. This alternative structure has 4 interdependent elements that differ from the wildtype and include: the extension of the T-stem by one GC base pair, one nucleotide in the T-bulge instead of the normal two, an RH-RWC base pair in the T-loop instead of RHUA, and an ILDH instead of the purine trap. The model shows that each elemental component seemingly fits together in a cascading interdependent network of structurally compensated building blocks. In addition, our analysis has led to the identification of a single H-atom (H1′ atom of nucleotide C8* from chapter III) as being one of the main constraining reasons for the covarying pattern in the dinucleotide sequence identities of RH-RWC base pair.

The identification of an *in vivo* functional T-loop motif alternative, type III, leads to some intriguing questions that require further study such as: Does type III exist naturally in other RNA molecules?; and Can type III replace other T-loop motifs found elsewhere?

More importantly, the study from chapter III reveals the existence of a tight steric interdependence between elements that make up type III. To see such a network of compensation identified above suggests that other RNA motifs also follow their own intricate network of constraints. Identifying such types of compensatory networks would add important information that could help solve the RNA folding problem. Currently, it seems there are really no methods available to elucidate such RNA compensatory behavior other than the one that we have used here.

In chapter IV, the use of the Instant Evolution approach to study the extended D-stem of tRNASec has revealed that the extension of the D-stem in

tRNASec introduces two important effects to this tRNA molecule: first it prevents the formation of tertiary interactions (8-14, 15-48) making this molecule more flexible, and second, the lost tertiary stability from the absence of the formerly mentioned tertiary interaction can be compensated by the increase in secondary structure stability through the addition of 2 WC base pairs in the D-stem. In addition, the fact that some variants without WC base pairs in the 5^{th} and 6^{th} layer of the D-stem functioned even better than wildtype at low but not high temperatures indicate that *Nature* has selected for tRNASec molecules that are resilient over a wider range of stressful environments (ie. higher temperature) rather than select those variants that perform better in a narrow range of environements (ie. only at low temperature).

Chapter V showed the length of acceptor/T domain can vary between 12 and 13 and still be functional, and moreover, the way in which this domain is composed can vary (8/5, 9/4, 7/5, 7/6, 8/4, and 9/3). Such experimental evidence suggests, based on the fact that even a single base pair diffence in a helical domain would cause a large change in the position of the acceptor end (a rotation of 33^{o} and an extension or contraction along the axis by 2.8Angstrom), that the tRNASec can remain functional despite such large differences in the structure of the tRNA L-shape. The ability for diverse tRNASec variants (different acceptor/T combinations) to remain functional implied that tRNASec is flexible as was predicted in chapter IV. Interestingly, variants with 13 layers were more active and had greater presence in the cytosol than those that had 12. We speculated that the higher presence of the former in the cytosol was through enhanced tRNA stability due to better interaction with one or more of the factors in the tRNASec functional cycle.

The theoretical study from chapter VI identified SelA as the likely factor that stabilizes variants with 13 layers in the acceptor/T domain over those with 12. Based on the process of elimination, we proposed SelA has the ability to directly select those variants with 13 layers as SelA interacts with the acceptor terminus and the T-loop, thus effectively measuring the length of the acceptor/T domain. Once the

tRNA variant is selenylated, SelB can play a secondary role in the stabilization of the tRNASec variant through binding which relies on the presence of Sec on the tRNA.

Although the studies on tRNASec has demonstrated the unprecedented flexibility available in this tRNA, for what reason such flexibiliy is needed remains unknown. As described in chapter V, we however speculate that such internal flexibility may be linked to the unusually long acceptor/T domain, as the latter aspect would likely cause steric problems during its time on the ribosome, and the former aspect would be expected to relieve it. The universal presence of tertiary interactions RHUA (8-14), as well as 15-48, in all standard tRNAs strongly suggests that a rigid tRNA molecular structure has been selected for standard tRNAs, which is in contrast to tRNAsec.

The studies undertaken here have shown the importance of providing an unbiased freedom to a given experimental system by the use of combinatorial libraries; thereby allowing the system to return viable alternatives to which we can learn from without prejudice. Most of the structural insights that were obtained based on screened variants were unexpected, yet proved extremely fruitful. In the future, where in a time when crystallography becomes routine, the joint use of such combinatorial gene library screens like the one used here and crystalizing the screened mutants for verification of structural models, can be expected to make this type of work more popular.

One of the most interesting parts of this work was to realize that there are steric clashes that are caused by the atoms inherent in the primary sequence of RNA which play a decisive role in the final outcome of a 3D structure, as we have seen with the 'pimple effect' from chapter III. Unfortunetly, the techniques that are available today that determine structure will not be able to document such phenomenon as only the final stable structure are manifested; a structure that has avoided such steric collisions. This implies that there is an unobservable parallel structural reality that dictates the way in which a particular RNA molecule folds in a 3D form. Perhaps in the future, with advances in such techniques as cryo-EM, we

may be able to catch such unfavorable folding moments that show clashes between atoms and help us understand why particular structures can not be formed. A systematic identification of steric collisions due to constraints in the primary sequence in this “unobservable parallel structural reality” would then be important if we are to truly understand the rules for RNA architecture.

References:

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). Molecular Biology of the Cell, 4th edition, 4th edn (New York, Garland Science).

Aldag, C., Brocker, M.J., Hohn, M.J., Prat, L., Hammond, G., Plummer, A., and Söll, D. Rewiring translation for elongation factor Tu-dependent selenocysteine incorporation. Angew Chem Int Ed Engl *52*, 1441-1445.

Arner, E.S. (2010). Selenoproteins-What unique properties can arise with selenocysteine in place of cysteine? Exp Cell Res *316*, 1296-1303.

Baron, C., and Böck, A. (1991). The length of the aminoacyl-acceptor stem of the selenocysteine-specific tRNA(Sec) of Escherichia coli is the determinant for binding to elongation factors SELB or Tu. J Biol Chem *266*, 20375-20379.

Baron, C., Sturchler, C., Wu, X.Q., Gross, H.J., Krol, A., and Böck, A. (1994). Eukaryotic selenocysteine inserting tRNA species support selenoprotein synthesis in Escherichia coli. Nucleic Acids Res *22*, 2228-2233.

Baron, C., Westhof, E., Böck, A., and Giege, R. (1993). Solution structure of selenocysteine-inserting tRNA(Sec) from Escherichia coli. Comparison with canonical tRNA(Ser). J Mol Biol *231*, 274-292.

Barrett, E.L., Jackson, C.E., Fukumoto, H.T., and Chang, G.W. (1979). Formate dehydrogenase mutants of Salmonella typhimurium: a new medium for their isolation and new mutant classes. Mol Gen Genet *177*, 95-101.

Betat, H., Rammelt, C., and Morl, M. tRNA nucleotidyltransferases: ancient catalysts with an unusual mechanism of polymerization. Cell Mol Life Sci *67*, 1447-1463. Biosym/MSI (1995). Insight II User Guide (San Diego).

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

Böck, A., Forchhammer, K., Heider, J., Leinfelder, W., Sawers, G., Veprek, B., and Zinoni, F. (1991). Selenocysteine: the 21st amino acid. Mol Microbiol *5*, 515-520.

Burkard, U., and Söll, D. (1988). The unusually long amino acid acceptor stem of Escherichia coli selenocysteine tRNA results from abnormal cleavage by RNase P. Nucleic Acids Res *16*, 11617-11624.

Carlson, B.A., Xu, X.M., Kryukov, G.V., Rao, M., Berry, M.J., Gladyshev, V.N., and Hatfield, D.L. (2004). Identification and characterization of phosphoseryl-tRNA[Ser]Sec kinase. Proc Natl Acad Sci U S A *101*, 12848-12853.

Chiba, S., Itoh, Y., Sekine, S., and Yokoyama, S. Structural basis for the major role of O-phosphoseryl-tRNA kinase in the UGA-specific encoding of selenocysteine. Mol Cell *39*, 410-420.

Chiba, S., Itoh, Y., Sekine, S., and Yokoyama, S. (2010). Structural basis for the major role of O-phosphoseryl-tRNA kinase in the UGA-specific encoding of selenocysteine. Mol Cell *39*, 410-420.

Crick, F.H. (1958). On protein synthesis. Symp Soc Exp Biol *12*, 138-163.

Forchhammer, K., Leinfelder, W., and Böck, A. (1989). Identification of a novel translation factor necessary for the incorporation of selenocysteine into protein. Nature *342*, 453-456.

Heider, J., Baron, C., and Böck, A. (1992). Coding from a distance: dissection of the mRNA determinants required for the incorporation of selenocysteine into protein. EMBO J *11*, 3759-3766.

Holley, R.W., Apgar, J., Everett, G.A., Madison, J.T., Marquisee, M., Merrill, S.H., Penswick, J.R., and Zamir, A. (1965). Structure of a Ribonucleic Acid. Science *147*, 1462-1465.

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.

Ishii, T.M., Kotlova, N., Tapsoba, F., and Steinberg, S.V. (2013) The long D-stem of the Selenocysteine tRNA provides Resilience at the expense of Maximal function. J Biol Chem.

Itoh, Y., Brocker, M.J., Sekine, S., Hammond, G., Suetsugu, S., Söll, D., and Yokoyama, S.(2013) Decameric SelA*tRNA(Sec) ring structure reveals mechanism of bacterial selenocysteine formation. Science *340*, 75-78.

Itoh, Y., Chiba, S., Sekine, S., and Yokoyama, S. (2009). Crystal structure of human selenocysteine tRNA. Nucleic Acids Res *37*, 6259-6268.

Itoh, Y., Sekine, S.I., Suetsugu, S., and Yokoyama, S. (2013). Tertiary structure of bacterial selenocysteine tRNA. Nucleic Acids Res.

Jaeger, L., Verzemnieks, E.J., and Geary, C. (2009). The UA_handle: a versatile submotif in stable RNA architectures. Nucleic Acids Res *37*, 215-230.

Juhling, F., Morl, M., Hartmann, R.K., Sprinzl, M., Stadler, P.F., and Putz, J. (2009). tRNAdb 2009: compilation of tRNA sequences and tRNA genes. Nucleic Acids Res *37*, D159-162.

Korostelev, A., Trakhanov, S., Laurberg, M., and Noller, H.F. (2006). Crystal structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements. Cell *126*, 1065-1077.

Kotlova, N., Ishii, T.M., Zagryadskaya, E.I., and Steinberg, S.V. (2007). Active suppressor tRNAs with a double helix between the D- and T-loops. J Mol Biol *373*, 462-475.

Kramer, G.F., and Ames, B.N. (1988). Isolation and characterization of a selenium metabolism mutant of Salmonella typhimurium. J Bacteriol *170*, 736-743.

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

Lee, K., Varma, S., SantaLucia, J., Jr., and Cunningham, P.R. (1997). In vivo determination of RNA structure-function relationships: analysis of the 790 loop in ribosomal RNA. J Mol Biol *269*, 732-743.

Leibundgut, M., Frick, C., Thanbichler, M., Böck, A., and Ban, N. (2005). Selenocysteine tRNA-specific elongation factor SelB is a structural chimaera of elongation and initiation factors. EMBO J *24*, 11-22.

Leinfelder, W., Zehelein, E., Mandrand-Berthelot, M.A., and Böck, A. (1988). Gene for a novel tRNA species that accepts L-serine and cotranslationally inserts selenocysteine. Nature *331*, 723-725.

Li, C., Reches, M., and Engelberg-Kulka, H. (2000). The bulged nucleotide in the Escherichia coli minimal selenocysteine insertion sequence participates in interaction with SelB: a genetic approach. J Bacteriol *182*, 6302-6307.

Liljas, A., and Ehrenberg, M. (2013). Structural Aspects of Protein Synthesis 2nd Edition, 2nd edn (London, World Scientific).

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.

Milo, R., Jorgensen, P., Moran, U., Weber, G., and Springer, M. (2010). BioNumbers--the database of key numbers in molecular and cell biology. Nucleic Acids Res *38*, D750-753.

Nissen, P., Thirup, S., Kjeldgaard, M., and Nyborg, J. (1999). The crystal structure of Cys-tRNACys-EF-Tu-GDPNP reveals general and specific features in the ternary complex and in tRNA. Structure *7*, 143-156.

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 U S A *83*, 6548-6552.

Normanly, J., Ollick, T., and Abelson, J. (1992). Eight base changes are sufficient to convert a leucine-inserting tRNA into a serine-inserting tRNA. Proc Natl Acad Sci U S A *89*, 5680-5684.

Palioura, S., Sherrer, R.L., Steitz, T.A., Söll, D., and Simonovic, M. (2009). The human SepSecS-tRNASec complex reveals the mechanism of selenocysteine formation. Science *325*, 321-325.

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

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

Schimmel, P. (1987). Aminoacyl tRNA synthetases: general scheme of structure-function relationships in the polypeptides and recognition of transfer RNAs. Annu Rev Biochem *56*, 125-158.

Schön, A., Böck, A., Ott, G., Sprinzl, M., and Söll, D. (1989). The selenocysteine-inserting opal suppressor serine tRNA from E. coli is highly unusual in structure and modification. Nucleic Acids Res *17*, 7159-7165.

Söll, D., RajBhandary, U.L., and RajBhandary, T.L. (1995). tRNA Structure, Biosynthesis, and Function, First edn (ASM Press).

Steinberg, S.V., and Boutorine, Y.I. (2007). G-ribo: a new structural motif in ribosomal RNA. RNA *13*, 549-554.

Sturchler, C., Westhof, E., Carbon, P., and Krol, A. (1993). Unique secondary and tertiary structural features of the eucaryotic selenocysteine tRNA(Sec). Nucleic Acids Res *21*, 1073-1079.

Takahata, M., Tamura, T., Abe, K., Mihara, H., Kurokawa, S., Yamamoto, Y., Nakano, R., Esaki, N., and Inagaki, K. (2008). Selenite assimilation into formate dehydrogenase H depends on thioredoxin reductase in Escherichia coli. J Biochem *143*, 467-473.

Tormay, P., Sawers, A., and Böck, A. (1996). Role of stoichiometry between mRNA, translation factor SelB and selenocysteyl-tRNA in selenoprotein synthesis. Mol Microbiol *21*, 1253-1259.

Turanov, A.A., Xu, X.M., Carlson, B.A., Yoo, M.H., Gladyshev, V.N., and Hatfield, D.L. Biosynthesis of selenocysteine, the 21st amino acid in the genetic code, and a novel pathway for cysteine biosynthesis. Adv Nutr *2*, 122-128.

Wu, X.Q., and Gross, H.J. (1994). The length and the secondary structure of the D-stem of human selenocysteine tRNA are the major identity determinants for serine phosphorylation. EMBO J *13*, 241-248.

Yoshizawa, S., and Böck, A. (2009). The many levels of control on bacterial selenoprotein synthesis. Biochim Biophys Acta *1790*, 1404-1414.

Zagryadskaya, E.I., Doyon, F.R., and Steinberg, S.V. (2003). Importance of the reverse Hoogsteen base pair 54-58 for tRNA function. Nucleic Acids Res *31*, 3946-3953.

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

Zinoni, F., Heider, J., and Böck, A. (1990). Features of the formate dehydrogenase mRNA necessary for decoding of the UGA codon as selenocysteine. Proc Natl Acad Sci U S A *87*, 4660-4664.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res *31*, 3406-3415.