Université de Montréal

Étude des interactions entre les boucles D et T chez les ARNs de transfert (ARNts)

par

Félix Doyon

Département de Biochimie (Bio-informatique)

Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures

en vue de l'obtention du grade de Maître ès Sciences (M.Sc.)

en Bio-informatique

Août, 2007

Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé :

Étude des interactions entre les boucles D et T chez les ARNs de transfert (ARNts)

présenté par :

Félix Doyon

a été évalué par un jury composé des personnes suivantes :

, président-rapporteur

Serguei Chteinberg, directeur de recherche

, membre du jury

Résumé

Afin d'élucider les contraintes générales imposées sur la structure des boucles D et T (région DT) chez les ARNs de transfert (ARNts), nous avons sélectionné, à partir d'une librairie combinatoire d'ARNts où plusieurs nucléotides étaient randomisés, une série d'ARNts suppresseur *in vivo*. L'analyse de ces séquences démontre que, parmi les nucléotides randomisés, les plus conservés sont les nucléotides 54 et 58 de la boucle T. La plupart du temps, ils forment une combinaison U54-A58 qui permet la formation d'une paire de base reverse-Hoogsteen. De façon surprenante, les autres séquences arborent quant à eux soit G54-A58, soit G54-G58. Nous avons pu montrer par modélisation moléculaire que ces deux alternatives permettent la formation de paires de base très similaires à la paire de base reverse-Hoogsteen U-A et qui pourraient la remplacer dans le contexte d'un ARNt fonctionnel. Ceci élève la paire de base reverse-Hoogsteen 54-58 au rang des éléments structuraux importants ayant un impact sur la fonction de la molécule. Nous suggérons que son rôle principal est de préserver la conformation du dinucléotide 59-60 et par conséquent de maintenir l'architecture générale en forme de « L » des ARNts. Dans un autre ordre d'idée, on peut remarquer que toutes les séquences tombent dans deux catégories, chacune caractérisée par un type particulier de séquence. L'analyse de ces deux groupes nous a permis de proposer deux arrangements pour la structure de la région DT.

Nous avons nommé le premier « piège à purine spécifique». Il est représenté par 12 séquences et correspond à une version généralisée des interactions entre les boucles D et T, où la purine 18 de la boucle D s'intercale entre la paire de base reverse-Hoogsteen U54-A58 et la purine 57 de la boucle T. L'identité de la purine 18 est contrôlée par un appariement spécifique avec le nucléotide 55 : elle est un G ou un A selon si ce dernier est un U ou un G, respectivement. Nous avons nommé le deuxième arrangement « piège à purine non-spécifique ». Il est représenté par 16 séquences c'est un arrangement qui n'a jamais été décrit auparavant dans la littérature. Il est caractérisé par l'intercalation de la purine 18 entre deux paires de base reverse-Hoogsteen U54-A58 et A55-C57. Par contre, il nécessite la présence d'un huitième nucléotide dans la boucle T et, à l'inverse du « piège à purine spécifique », la purine 18 ne forme pas d'interactions avec le nucléotide 55. Ainsi, nous avons décrit et explicité la logique derrière certains éléments de structure chez l'ARNt, et ce savoir, appliqué à de nouvelles molécules, nous permettra d'approfondir nos connaissances de la structure d l'ARN.

Mots-clés : motif ARN, modélisation moléculaire, Forme en « L », sélection *in vivo*

Abstract

To elucidate the general constraints imposed on the structure of the D- and T-loops in functional tRNAs, active suppressor tRNAs were selected *in vivo* from a combinatorial tRNA gene library in which several nucleotide positions of these loops were randomized. Analysis of the nucleotide sequences of the selected clones demonstrates that among the randomized nucleotides, the most conservative are nucleotides 54 and 58 in the T-loop. In most cases, they make the combination U54-A58, which allows the formation of the normal reverse Hoogsteen base pair. Surprisingly, other clones have either the combination G54-A58 or G54-G58. However, molecular modeling shows that these purine-purine base pairs can very closely mimic the reverse-Hoogsteen base pair U-A and thus can replace it in the T-loop of a functional tRNA. This places the reverse Hoogsteen base pair 54-58 as one of the most important structural aspects of tRNA functionality. We suggest that the major role of this base pair is to preserve the conformation of dinucleotide 59-60 and, through this, to maintain the general architecture of the tRNA L-form. With only one exception, all of the clones fall into two groups, each characterized by a distinct sequence pattern. Analysis of these two groups has allowed us to suggest two different types of nucleotide arrangement in the DT region. The first type, the so-called specific purine trap, is found in 12 individual tRNA clones and represents a generalized version of the standard D-T loop

interaction. It consists of purine 18 sandwiched between the reverse-Hoogsteen basepair U54-A58 and purine 57. The identity of purine 18 is restricted by the specific base-pairing with nucleotide 55. Depending on whether nucleotide 55 is U or G, purine 18 should be, respectively, G or A. The second structural type, the so-called non-specific purine trap, corresponds to the nucleotide sequence pattern found in 16 individual tRNA clones and is described here for the first time. It consists of purine 18 sandwiched between two reverse-Hoogsteen base-pairs U54-A58 and A55-C57 and, unlike the specific purine trap, requires the T-loop to contain an extra eighth nucleotide. Since purine 18 does not form a base-pair in the nonspecific purine trap, both purines, G18 and A18, fit to the structure equally well. The important role of both the specific and non-specific purine traps in the formation of the tRNA L-shape is discussed.

Keywords : RNA motif, molecular modeling, L-shape, selection *in vivo*

Table des matières

Liste des tableaux

Liste des figures

À l'homme qui représente la Science

pour moi, Serguei Chteinberg.

Remerciements

J'aimerais remercier mon directeur de recherche, Serguei Chteinberg, pour sa patience, sa compréhension et la confiance qu'il m'a accordée.

J'aimerais aussi remercier tous les membres du laboratoire, particulièrement mes coauteurs Ekaterina Zagryadskaya et Jianhong Chen.

J'aimerais aussi remercier ma copine Julie, sans l'aide et le support de laquelle je n'aurais probablement pas terminé ce mémoire.

Finalement, j'aimerais remercier le programme stratégique de bourse de formation des IRSC en bio-informatique, les bourses d'excellence bit, ansi que le programme de bourse du Fonds de la Recherche en Santé du Québec (FRSQ) pour leur support financier.

Chapitre 1 : Introduction

L'humanité se questionne depuis bien longtemps sur la question de l'origine de la vie. Plusieurs différentes théories ont vu le jour à travers les différentes époques, allant du Créationnisme à la théorie du « Big Bang ». Plus près de la réalité que l'on connaît aujourd'hui, beaucoup de découvertes ont été faites dans le domaine de la biologie depuis l'hypothèse de l'évolution émise par Darwin en 1859. Le 20^{e} siècle a depuis été particulièrement prolifique pour l'avancement de la Science. La découverte de la structure de l'ADN par Watson et Crick a généré un engouement sans précédent, non seulement chez la communauté scientifique mais aussi pour la société en général [WC1953]. Ce grand pas vers la compréhension de ce que nous sommes a ensuite ouvert la porte à de nombreuses découvertes dont certaines ont révolutionné notre vision de l'organisme vivant et son fonctionnement. Pour n'en nommer que quelques-unes, il y eut la détermination de la structure du premier ARN de transfert, il y a plus de 30 ans [QSMKSR1974] et, plus récemment, la détermination de la structure du ribosome [BNHMS2000, WBCMWCVHR2000, YYBLECN2001].

À ce jour, on sait que l'acide désoxyribonucléique (ADN) est la molécule encodant la presque totalité des processus vitaux. Il a été établi qu'il doit d'abord être transcrit en acide ribonucléique (ARN), ce qui s'effectue avec la collaboration essentielle de l'ARN polymérase. Par la suite, l'ARN est à son tour traduit en protéine par l'entremise du complexe ribosomal. Cette synthèse de peptides est un

processus élaboré qui implique une multitude de participants, allant des acides aminés jusqu'aux différentes enzymes. Bien que la science ait apporté de nombreuses réponses au cours des dernières décennies, beaucoup reste encore à éclaircir. D'ailleurs, nous découvrons et caractérisons sans cesse de nouveaux joueurs dans le processus de la synthèse protéique. Leurs fonctions recèlent de nombreux mystères que les scientifiques tentent d'élucider encore à ce jour.

Le séquençage en entier du génome humain, en 2004, a entre autre permis d'estimer le nombre total de gènes entre 20 000 et 25 000 [IHGSC2004]. Suite à cette révélation, il a été admit que plusieurs des gènes séquencés correspondent fort probablement à des protéines. Néanmoins, certains de ces gènes codent pour des ARNs et nous savons déjà que ceux-ci jouent un rôle prépondérant dans plusieurs processus biologiques. Le plus évident de ces rôles est de servir de messager (ARNm) entre l'ADN et le ribosome, qui est le centre de la synthèse des protéines. D'autre part, l'ARN joue un rôle au niveau de la composition du complexe ribosomal (ARNr). Il est également impliqué dans le transport des acides aminés destinés à être incorporés lors de la synthèse des protéines (ARNt). Par conséquent, l'ARN, sous ses différentes formes, assure plusieurs fonctions vitales à la cellule et de nos jours encore on lui découvre de nouveaux rôles et applications. Un des plus récents exemples est sans doute la découverte et

l'utilisation des siRNAs, qui révolutionne présentement le monde de la biologie moléculaire [HB1999, EHLYWT2001].

D'un point de vue structurel, les ARNs sont des molécules hautement organisées. La figure 1.1 est un exemple tridimensionnel de plusieurs molécules d'ARN et illustre bien leur complexité. Leur arrangement spatial est si crucial que leurs fonctions ne peuvent être assurées que dans des conditions bien précises, telles qu'un repliement adéquat de la molécule et ce, dans une fenêtre de temps bien précise. On comprend alors pourquoi il est primordial de bien comprendre les différents aspects qui régissent le repliement des ARNs. Ces connaissances nous permettront éventuellement d'être en mesure de mieux cerner leurs fonctions et d'en comprendre les mécanismes.

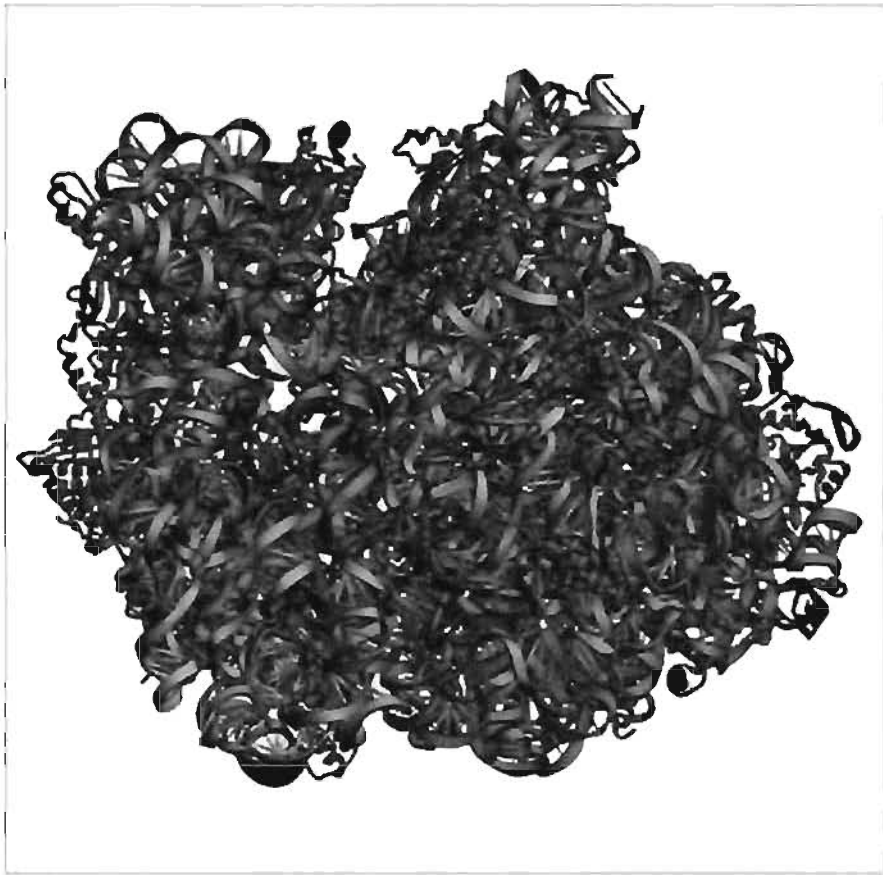

Figure 1.1 Structure tridimensionnelle du ribosome contenant des ARNts. L'ARN est représenté sous forme de bâtonnets (représentant les bases) et de rubans (représentant le squelette sucre-phosphate). La grosse sous-unité (23S) est en gris, la petite (16S) en cyan, l'ARN 5S en mauve, alors que les ARNts sont en jaune, orange et rouge. Les protéines ribosomales sont aussi présentes en magenta (grosse sous-unité) et en mauve foncé (petite sous-unité) [YYBLECN2001].

1.1 STRUCTURE DE L'ARN

À la base, l'ARN est un polymère formé d'un squelette sucre-phosphate et de quatre différentes bases azotées : l'adénine, la cytosine, la guanine et l'uracile (Figure 1.2). Il existe des molécules d'ARN de toutes les tailles : de l'ARN de

transfert (ARNt), qui est composé habituellement d'environ 76 nucléotides, jusqu'au ribosome, pour lequel la petite sous-unité à elle seule comporte plus de 1500 nucléotides chez les procaryotes et quelques centaines de plus chez les eucaryotes [BA2002].

L'ARN ressemble beaucoup à l'ADN, mis à part quelques différences déterminantes : dans un premier temps, le sucre du nucléotide est un ribose plutôt qu'un désoxyribose. Dans un deuxième temps, l'uracile remplace la thymine de l'ADN. En effet, la figure 1.2 montre que l'ADN possède un groupement hydroxyl en position 3' de son pentose, alors que l'ARN en possède aussi un en position 2'. L'ARN dispose, tout comme les protéines, d'une structure hautement hiérarchisée : la structure primaire correspond à la séquence nucléotidique composant la molécule; la structure secondaire correspond à l'appariement Watson-Crick entre les différents nucléotides; tandis que la structure tertiaire correspond à l'arrangement adopté par la molécule en trois dimensions. Dépendamment du type d'ARN et de ses fonctions, certains éléments de structure sont conservés et nécessaires au bon fonctionnement de la molécule. Nous voyons plus en détails ces différents aspects dans les sections 1.1.2, 1.1.3 et 1.1.4.

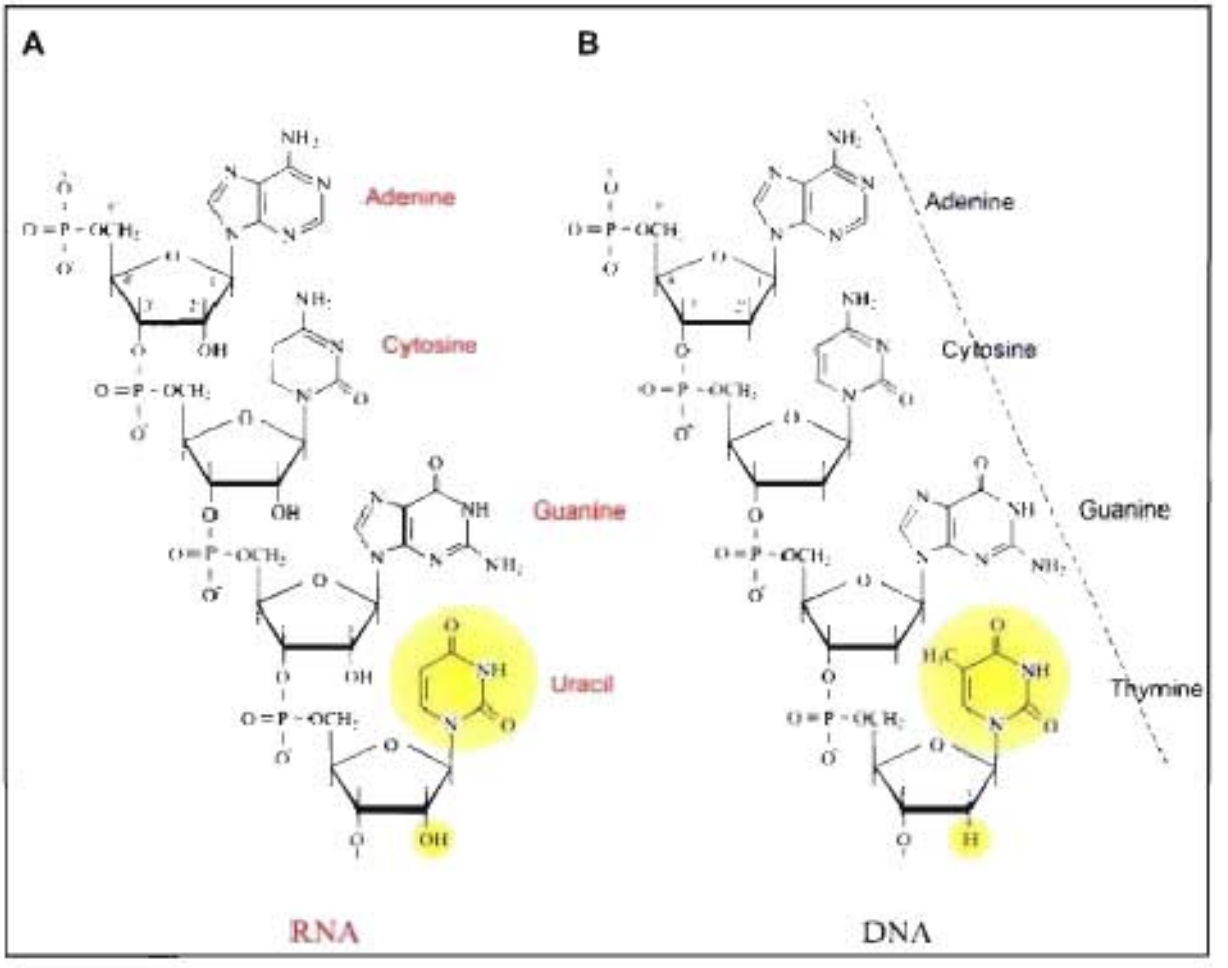

Figure 1.2 Structure polymérique des acides nucléiques. Pour les deux polymères, l'extrémité 5' est située en haut alors que l'extrémité 3' est en bas. **A)** Polymère d'ARN illustrant les 4 bases azotées de l'ARN, l'adénine, la cytosine, la guanine et l'uracile. **B)** polymère d'ARN illustrant les 4 bases azotées de l'ARN, l'adénine, la cytosine, la guanine et la thymine. Les différences entre l'ADN et l'ARN sont illustrées en jaune. On peut y voir que la thymine est remplacée par l'uracile dans l'ADN et que l'ARN possède un groupement hydroxyl en 2'. [KTF].

1.1.1 Le nucléotide

Le nucléotide est l'unité de base du polymère que sont les acides nucléiques. Il est formé de trois parties, soit d'une base azotée, d'un pentose, et d'un groupement phosphate (Figure 1.3). Les bases canoniques (ou conventionnelles) de l'ADN et de l'ARN sont des purines (adénine et guanine) ou des pyrimidines (cytosine, thymine, uracile). Le pentose est un ribose chez l'ARN tandis qu'il est un désoxyribose chez l'ADN. Le nucléotide possède trois arêtes hydrophiles via lesquelles il forme des ponts hydrogènes (pont-H) avec d'autres nucléotides et constitue ainsi des paires de bases. Ces trois arêtes sont la face Watson-Crick, la face Hoogsteen et la face Sucre [LW2001]. Dans une hélice double régulière, qui est un élément de structure fréquemment retrouvé, les bases se font face par leur côté Watson-Crick, mais une multitude de paires de bases dites non-Watson-Crick sont possibles.

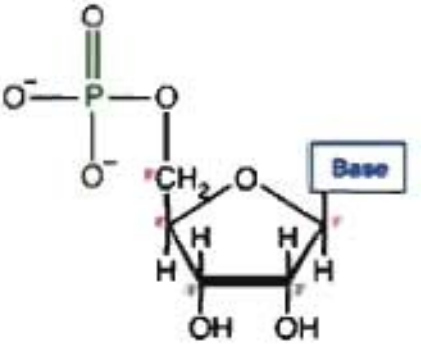

Figure 1.3 Structure d'un nucléotide de l'ARN. Le nucléotide est formé d'une base azotée (représentée par un rectangle bleu), d'un ribose (en noir; les carbones sont numérotés de 1' à 5') et d'un groupement phosphate (en vert) [ND].

Plusieurs nucléotides peuvent être modifiés au stade post-transcriptionnel. Par exemple, dans l'ARNt, la boucle T comporte une pseudouridine (ψ) en position 55. Ce nucléotide s'apparente à l'uracile, mais où le lien entre le sucre et la base se fait entre le C1 du sucre et le C5 de la base plutôt qu'avec le N1 [CG2000]. Un autre nucléotide modifié communément rencontré est l'inosine. Il s'agit en fait d'une guanine qui a subi une désamination. L'inosine est impliquée dans l'hypothèse Wobble, selon laquelle le nucléotide en 5' de l'anticodon jouit d'une plus grande liberté conformationnelle, lui permettant de former des paires de bases non-conventionnelles. Cette hypothèse expliquerait l'existence de 40 ARNt pour 60 codons [WC1966]. Cet exemple démontre bien qu'il est clair que ces nucléotides modifiés jouent un rôle important tant au niveau de la structure que de la fonction. Ainsi, dû à leur importance relative, on les retrouve souvent dans les centres fonctionnels des molécules d'ARN.

1.1.2 La structure primaire de l'ARN

La structure primaire consiste en la séquence nucléotidique de l'ARN, c'est-à-dire la suite des nucléotides la composant, de l'extrémité 5' jusqu'à l'extrémité 3'. L'ADN est d'abord transcrit en ARN, qui subit ensuite plusieurs étapes de maturation au cours desquelles il est épissé et où certains nucléotides peuvent être modifiés [GE2000] (tel que décrit à la section 1.1.1).

1.1.3 La structure secondaire de l'ARN

La structure secondaire consiste en l'appariement Watson-Crick des nucléotides (section 1.1.1). Ce positionnement forme différents domaines qui peuvent être composés d'hélices doubles, de tige-boucles (« hairpin loop »), de renflements (« bulges ») et de boucles internes (« internal loop »). Les tige-boucles sont constituées d'une hélice double bouclée par un petit nombre de nucléotides non-appariés. Les renflements correspondent à un ou plusieurs nucléotides non appariés dans un seul des brins d'une hélice double. Les boucles internes quant à elles sont formées de nucléotides non-appariés dans les deux brins d'une hélice double. Les renflements et les boucles internes confèrent une certaine flexibilité à la région. Les nucléotides les composants sont souvent impliqués dans les interactions au niveau de la structure tertiaire. La Figure 1.4 montre quelques-uns de ces motifs structurels.

Deux forces primordiales interviennent au niveau de la structure secondaire de l'ARN et ont un impact sur sa structure et ses fonctions. Celle ayant le plus grand impact est régit par des interactions hydrophobes et favorise l'empilement des orbitales π des cycles aromatiques des bases azotées l'un par-dessus l'autre. Ce phénomène est appelé empilement (« stacking ») et joue un énorme rôle dans la stabilité de la molécule d'ARN [FPHT1983]. Le deuxième élément primordial à

intervenir au niveau de la structure secondaire de l'ARN est le pont-H. Ce type d'interaction implique le partage d'un atome d'hydrogène par deux atomes électronégatifs. C'est d'ailleurs ce qui détermine la complémentarité entre l'adénine et l'uracile, et entre la guanine et la cytosine [J1997]. La présence de ces types d'interactions entre nucléotides a une grande influence au niveau de l'arrangement tridimensionnel de la molécule d'ARN et joue donc indirectement un rôle certain dans sa capacité à assurer ses fonctions de façon adéquate.

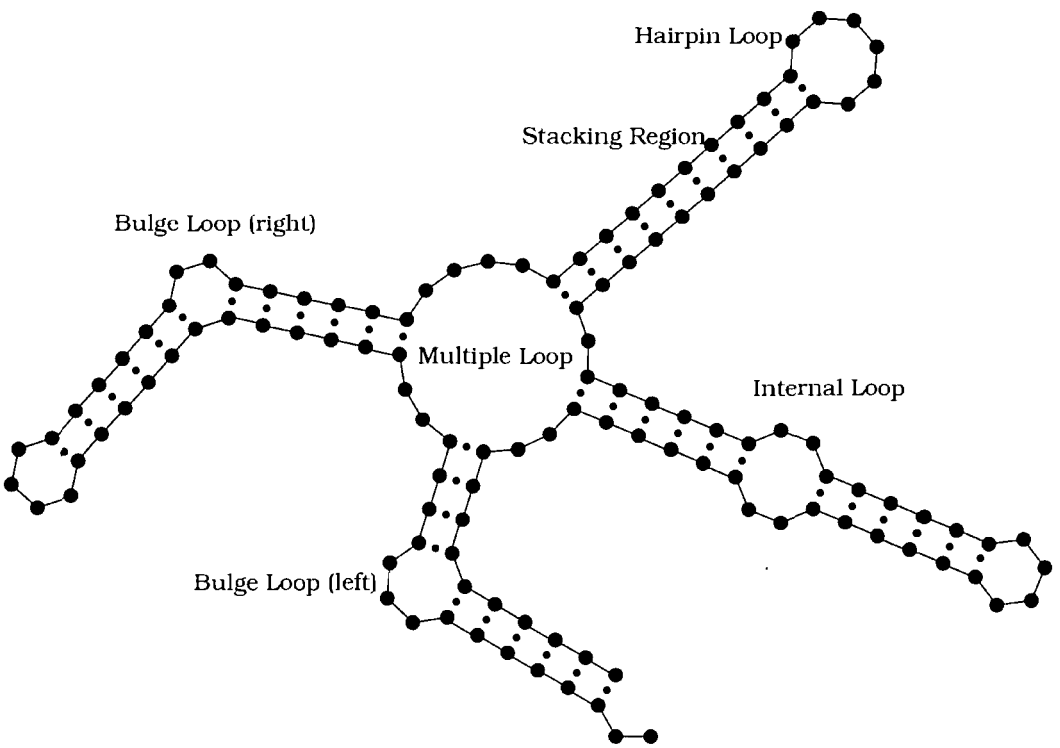

Figure 1.4 Différents éléments de structure secondaire retrouvés dans l'ARN. Plusieurs éléments de structure peuvent être retrouvés au sain de l'ARN. Notamment, les tige-boucle, les régions d'empilement, les renflements, les boucles internes et les boucles multiples [SG2005].

1.1.4 La structure tertiaire de l'ARN

La structure tertiaire de l'ARN diffère grandement de celle de l'ADN. Dans le cas de l'ADN, la structure tertiaire est majoritairement constituée de longues hélices doubles régulières. Dans le cas de l'ARN, il a plutôt tendance à replier ses éléments de structure secondaire l'un sur l'autre, de façon très compacte, un peu à la manière des protéines. Ces éléments de structure secondaire sont les hélices doubles, les tige-boucles, les régions connectrices simple brin, les paires de bases tertiaires et/ou non-Watson-Crick. Il n'est pas rare de voir des éléments distants en termes de structure secondaire se rapprocher dans la structure tertiaire et interagir ensemble.

La façon dont les ARNs se replient demeure un mystère, mais plusieurs motifs structuraux ont été déterminés. Ces différents éléments de structure pourraient former un ensemble de « blocs » structuraux bien défini à partir duquel les ARN adoptent la multitude de structures qu'on leur attribue. Les interactions de type tertiaire impliquent aussi parfois les atomes qui ne sont pas impliqués dans l'appariement entre les bases, c'est-à-dire les atomes du ribose et du groupement phosphate [HBH2005]. D'autres recherches sur la nature des nucléotides qui composent la molécule d'ARN ainsi que sur leurs interactions sont bien entendues nécessaires afin de pouvoir élucider les mécanismes de repliement de l'ARN. De

plus, l'étude de l'arrangement spatial en relation avec la composition des molécules d'ARN est essentielle afin de permettre d'identifier certaines causes du mauvais repliement des ARNs. Bien que les récents développements technologiques dans ce domaine de recherche aient permis de grandes percées, beaucoup reste encore à être élucidé quant aux ARN.

1.2 L'ARN DE TRANSFERT (ARNt)

À ce jour, les différentes études portant sur l'ARNt ont permis de déterminer qu'il est un participant important du processus de synthèse des protéines. En effet, les ARNts sont impliqués au niveau du ribosome et agissent en tant que molécules adaptatrices entre l'ARN messager et la séquence de la protéine nouvellement synthétisée. C'est donc dire qu'ils ont une fonction vitale au sein de notre organisme. Sa structure, bien qu'elle ait été séquencée, comporte de nombreux éléments déterminants dont le rôle et l'importance relative restent encore à déterminer.

1.2.1 La structure tertiaire en forme de «L» et les deux centres fonctionnels

Afin que l'ARNt puisse jouer son rôle au sein de la synthèse protéique, deux centres fonctionnels ont été identifiés: l'anticodon et l'extrémité acceptrice. Ces centres interagissent respectivement avec l'ARN messager et le ribosome, sur son

site de réaction peptidyltransférase. Pour que ces interactions soient possibles, la molécule d'ARNt doit adopter une structure particulière et précise afin que ses deux centres fonctionnels soir bien positionnés pour lui permettre d'assumer ses fonctions cellulaires vitales. C'est donc dire que les structures secondaire et tertiaire sont extrêmement importantes pour que la fonctionnalité des ARNtssoit conservée.

À ce sujet, la Figure 1.5 illustre une molécule d'ARNt sous sa forme secondaire et sous sa forme tertiaire. La structure secondaire est dite en « feuille de trèfle ». Cette appellation est due à la présence de trois tige-boucles (D, T et anticodon) et d'une quatrième tige, dite acceptrice. Cette dernière forme un domaine coaxial avec la tige-boucle T (domaine II) et elle a la capacité et le rôle de lier l'acide aminé. La tige-boucle de l'anticodon forme quant à elle un domaine coaxial avec la tige boucle D (domaine I).

L'orientation mutuelle de ces deux domaines est d'une importance primordiale afin que les ARNts puissent assumer correctement leur fonction. Il a été déterminé que les deux domaines coaxiaux doivent être orientés de façon perpendiculaire, ce qui confère la forme de « L » à la structure tertiaire de l'ARNt. Les domaines I et II se retrouvent ainsi aux extrémités du « L », soit au bout de la tige-boucle de l'anticodon et au bout de la tige acceptrice, comme on peut le voir à

la figure 1.5(B). Cette structure tridimensionnelle rigide universelle est très conservée chez les ARNt cytosoliques [SHBIS1998].

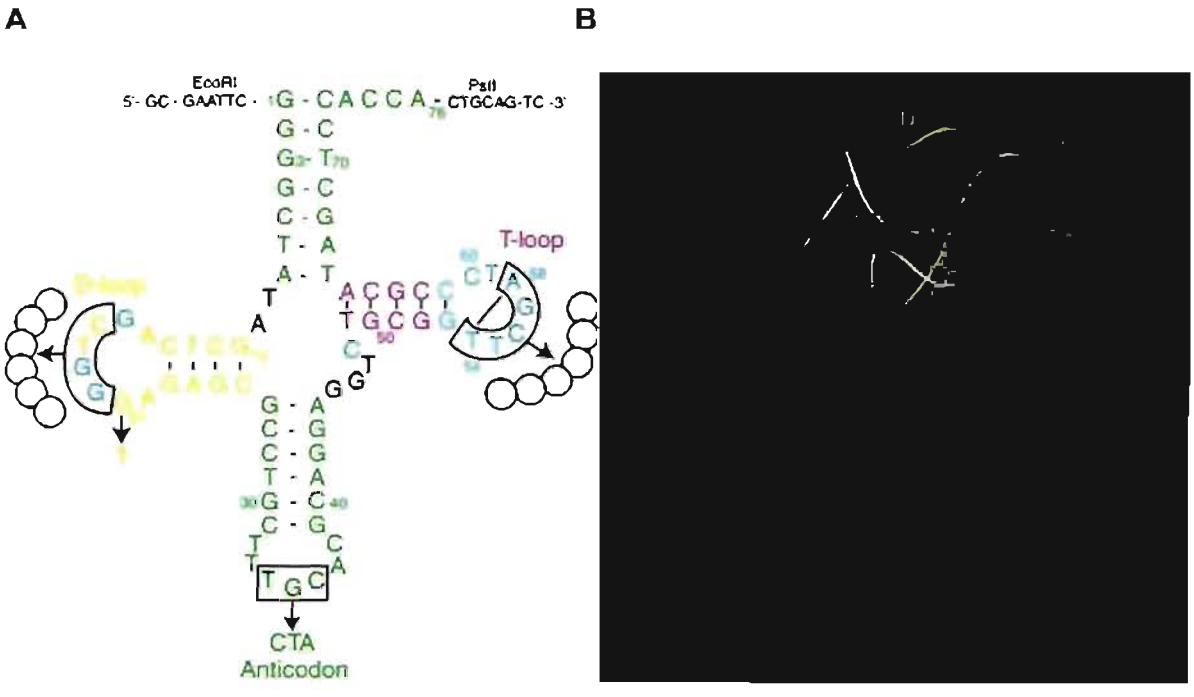

Figure 1.5 Structures secondaire (A) et tertiaire (B) de l'ARN de transfert. La tige acceptrice (ACC) ainsi que la tige-boucle de l'anticodon (A/C) sont en vert, la tige-boucle D en jaune et la tige-boucle T en magenta dans les deux structures. Les nucléotides en cyan sont ceux composant la région DT. Les nucléotides encadrés dans la structure secondaire sont les nucléotides qui ont été mutés dans la librairie-K [ZDS2003].

1.2.2 La région charnière et les nucléotides la composant

Les deux centres fonctionnels sont situés à une distance précise et cruciale de 70 Å l'un de l'autre [SSA1994]. D'un point de vue tridimensionnel, seule la forme en « L » permet d'obtenir cette distance. Cette particularité suggère donc

que la conformation adoptée par la région charnière au coude du « L » joue indubitablement un rôle primordial dans ce maintien et par conséquent dans la fonctionnalité de la molécule, qui constitue le sujet principal de ce travail.

La région charnière de la molécule est majoritairement composée des nucléotides des boucles D et T, et sera référée tout au long de ce mémoire en tant que région DT. Plusieurs des nucléotides de la région DT sont extrêmement conservés, ce qui suggère fortement leur implication dans la fixation des deux domaines hélicoïdaux. Par exemple, parmi les onze nucléotides de la région DT, huit sont invariables, du moins chez les ARNts cytosoliques. Les trois autres possèdent seulement un degré limité de variabilité : le nucléotide 57 doit être une purine, alors que les deux nucléotides formant l'appendice, 59 et 60, sont des pyrimidines dans la plupart des cas [SHBIS1998]. De plus, la région DT est composée d'interactions conservées entre ces nucléotides, malgré qu'il soit difficile d'évaluer l'importance relative de chacune de ces interactions ponctuelles.

1.2.3 Les interactions intéressantes relevées à ce jour

Plusieurs interactions révélées par la détermination de la structure de l'ARNt semblent dignes d'intérêt. Tout d'abord, mentionnons la paire de base universelle dite « reverse-Hoogsteen » U54-A58. Cette paire est formée par l'interaction entre

le premier et le cinquième nucléotide de la boucle T. Puis, une autre intéressante interaction dans la région DT implique le nucléotide 18 de la boucle D qui s'empile entre les nucléotides 57 et 58 de la boucle T. Cet arrangement permet de former une paire de base non-Watson-Crick avec le nucléotide 55. On dénote aussi une paire de base tertiaire, mais de type Watson-Crick cette fois. Elle est formée par l'interaction des nucléotides 19 et 56 au-dessus du nucléotide 57.

La région DT est également composée d'autres éléments. Parmi ceux-ci, on y retrouve l'appendice, qui est formé de deux nucléotides non-appariés en 3' du nucléotide 58. Il est situé entre la paire de base « reverse-Hoogsteen » et la dernière paire de bases de la tige T. L'appendice vient s'empiler sur la dernière couche du domaine I. Cette couche est elle-même constituée d'une paire de base tertiaire formée par les nucléotides 15 et 48. Un autre élément présent digne de mention est un motif « U-turn » qui est régi par le nucléotide ψ55.

1.3 DIFFÉRENTES MÉTHODES D'ANALYSE DE LA STRUCTURE DE L'ARN

Même si le présent ouvrage se concentre sur l'analyse de la structure de l'ARNt, il est important de mentionner les différents procédés par lesquels ces structures ont été obtenues. Plusieurs méthodes ont été développées au fil des années, chacune possédant ses avantages et ses limites. Ensemble, ces

différents outils se sont complémentés de façon à nous permettre d'obtenir une banque de structures qui correspond au point de départ de notre analyse.

1.3.1 La cristallographie de rayon-X

La cristallographie est une méthode basée sur l'analyse du patron de diffraction d'un cristal pur provenant d'une molécule. Ce patron de diffraction est obtenu en faisant passer un faisceau de rayons X à travers le cristal. La toute première macromolécule à être cristallisée à la fin des années cinquante fût la protéine myoglobine. Plus tard, cette technique s'est raffinée et a permis, entre autres, de déterminer la structure tridimensionnelle de l'ARNt^{Phe}.

1.3.2 La résonnance magnétique nucléaire (RMN)

Cette autre technique permet la détermination de la structure de macromolécules biologiques. Dépendamment de leur environnement physico-chimique, différents protons résonnent à des fréquences différentes. Ce « déplacement chimique » peut être enregistré et analysé. Il permet de distinguer les atomes selon leur environnement physico-chimique à l'intérieur de la molécule, et d'ainsi obtenir différentes informations à propos de la molécule à l'étude. Cette méthode permet d'obtenir, avec une haute résolution, la structure de molécules mal structurées (pour lesquelles la structure secondaire n'est pas stable). Par

contre, en comparaison avec la méthode de cristallographie de rayon-X, la RMN ne peut être utilisée que pour des molécules relativement petites [L2001].

1.3.3 La modélisation moléculaire interactive

Cette méthode capitalise sur l'existence de structures tridimensionnelles et fait intervenir le jugement humain. L'objectif est d'analyser *in silico* certains éléments ponctuels de ces structures. Cette technique englobe aussi les différentes méthodes informatiques de prédiction de structures et de leurs comportements à différentes températures et dans différents environnements. La modélisation moléculaire interactive est un outil prédictif qui doit être utilisé de façon très critique. Par opposition avec les méthodes physico-chimiques de détermination de la structure, il ne constitue pas une preuve en soi mais plutôt un argument en faveur d'un modèle.

1.3.4 Les librairies combinatoires D'ARNt

Les librairies combinatoires sont aussi connues sous l'appellation « évolution dirigée ». Elles constituent une méthode puissante qui permet d'obtenir des séquences et des structures qui ne sont pas retrouvées dans la nature. L'avantage indéniable de cette méthode est qu'elle permet l'obtention d'une plus

grande variété de séquences qui n'ont pas été favorisées *a priori*, comme c'est le cas lorsque la technique de mutagénèse dirigée est utilisée.

1.4 HYPOTHÈSE ET OBJECTIFS

Malgré la connaissance de sites d'interaction sur la région DT de l'ARNt, plusieurs questions restent à ce jour sans réponse. Par exemple, pourquoi la paire de base « reverse-Hoogsteen » est-elle absolument nécessaire afin de préserver la fonctionnalité de la molécule? En quoi est- elle si spécifique et qu'est-ce qui la rend unique? Ces questions ne peuvent êtres résolues par une approche biochimique standard. C'est pourquoi nous avons choisi de coupler une analyse théorique poussée qui impliquera de la modélisation moléculaire *in silico* à la collecte de dorinées qui se fera de manière biochimique par l'entremise de librairies combinatoires d'ARNts.

L'objectif global de mon projet consiste à déterminer les règles qui gouvernent le repliement des ARNs structurés en trois dimensions. Plus spécifiquement, mon objectif est de déterminer les conditions nécessaires et suffisantes à la formation de motifs d'ARN. L'ARNt est une molécule-modèle de premier choix pour réaliser cette étude. En effet, beaucoup de données de

séquence et de structure sont disponibles. De plus, c'est une molécule assez complexe du point de vue structural tout en ayant une taille raisonnable.

CHAPITRE 2: Importance of the reverse Hoogsteen base pair 54-58 for tRNA function

Ekaterina I. Zagryadskaya, Félix R. Doyon, Sergey V. Steinberg

Département de Biochimie, Université de Montréal, Mtl, Qc H3C 3J7, Canada

DOI: 10.1093/nar/gkg448 Nucleic Acids Research, 2003, Vol. 31, No. 14 3946-3953

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

dinucleotide 59-60 and, through this, to maintain the general architecture of the tRNA L-form.

2.1 INTRODUCTION

One of the most conservative elements in the tRNA tertiary structure is the region at the outer corner of the tRNA L-form, where the T-loop interacts with the D-loop. This region, which we will henceforth call the DT region, is comprised of the whole T-loop, the first base pair of the T-stem 53-61 and nucleotides 18 and 19 of the D-loop, which interact, respectively, with nucleotides 55 and 56 of the T-loop (Fig. 1). Out of 11 nt of the DT region, only three, 57, 59 and 60, show a limited variability: 57 is always a purine, while 59 and 60 are pyrimidines in most cases [SHBIS1998]. The other 8 nt of this region are invariable in cytosolic tRNAs. The DT region is involved in several important tRNA functions. First, it plays a major role in maintaining the perpendicular arrangement of the two helical domains called the L-form, which provides the proper juxtaposition of the two functional centers, the acceptor terminus and the anticodon. Also, this region is important for correct and efficient maturation of the termini of the molecule [TLAD1994, AKT1995, NTK1999]. Finally, it harbors recognition elements for interaction with different tRNA-binding enzymes, including some aminoacyl-tRNA synthetases [MF1988a, MFJS1991, PU1992, BYTC1994, PWU2000, NMAVGS2001].

The tertiary structure of the DT region is of special interest and has been the subject of a number of studies [QR1976, dBK1983, RCWEEEG1987, OAK1995]. The presence of such elements as the U-turn between T54 and C55, the unusual non-Watson-Crick base pairs T54-A58 and G18-Y55, the mutual intercalation of fragments 57-58 and 18-19, the bulging of nucleotides 59-60 and the interaction of phosphate 60 with the amino group of C61 makes this region one of the most structurally diversified in the whole tRNA. This diversity raises questions concerning the role played by each of these elements in the structure of the DT region and of the whole molecule and the limits within which these elements could be modified without destroying tRNA structure and function. These questions become even more important in view of the recent finding that rRNA also contains motifs resembling the structure of the DT region [NF2002]. Thus, elucidation of the role and the sequence requirements for formation of the elements constituting this region in such a relatively small molecule as tRNA would contribute to understanding of structure-function relationships in other RNAs and RNA-protein complexes, including the ribosome. To address this problem, we here undertook an analysis of the general constraints imposed on the structures of the D- and T-loops in a tRNA functioning in vivo. For this, we selected suppressor tRNAs from a specially designed combinatorial tRNA library in which a number of positions in the D- and T-loops were randomized. Analysis of the nucleotide sequences of the

successful tRNA clones sheds light on the role of particular elements of the DT region in the global tRNA structure.

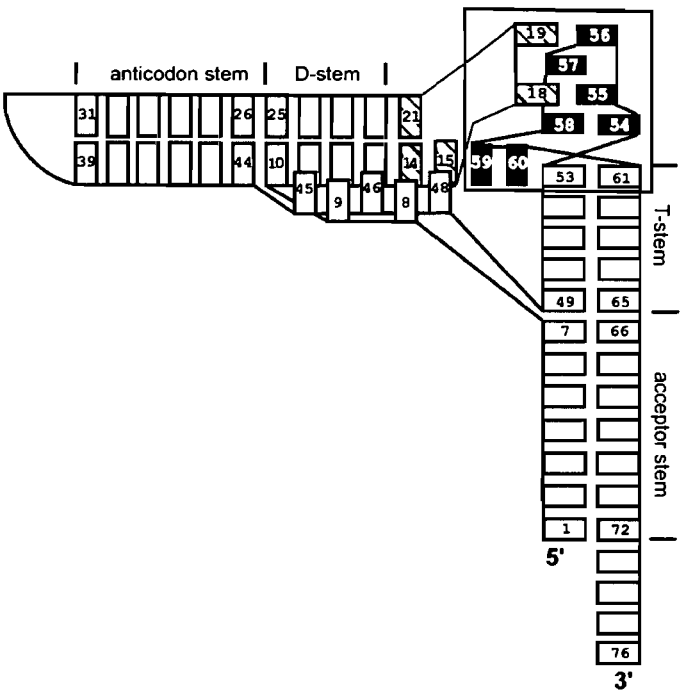

Figure 2.1 The standard tRNA L-form. Rectangles represent individual nucleotides. The DT region at the outer corner of the molecule is boxed. Cross-hatched and filled rectangles represent nucleotides of the D- and T-loop, respectively. Unpaired nucleotides as well as nucleotides at the beginning and the end of the helical regions are numbered in accordance with the standard tRNA nomenclature [SHBIS1998]. Nucleotides of the anticodon loop, non-stacked nucleotides of the D-loop and nucleotide 47 are not shown. There are two base pairs, G18-Y55 and G19-C56, formed between the D- and T-loops. The reverse Hoogsteen base pair T54-A58, whose structure is seen in Figure 2.4, is formed within the T-loop. Dinucleotide 59-60 bulges from the double helical stem between base pairs G53-C61 and T54-A58. Nucleotide 59 stacks on the tertiary base pair 15-48, constituting the last layer of the D/anticodon helical domain. This interaction fixes the perpendicular arrangement of the two helical domains called the L-form.

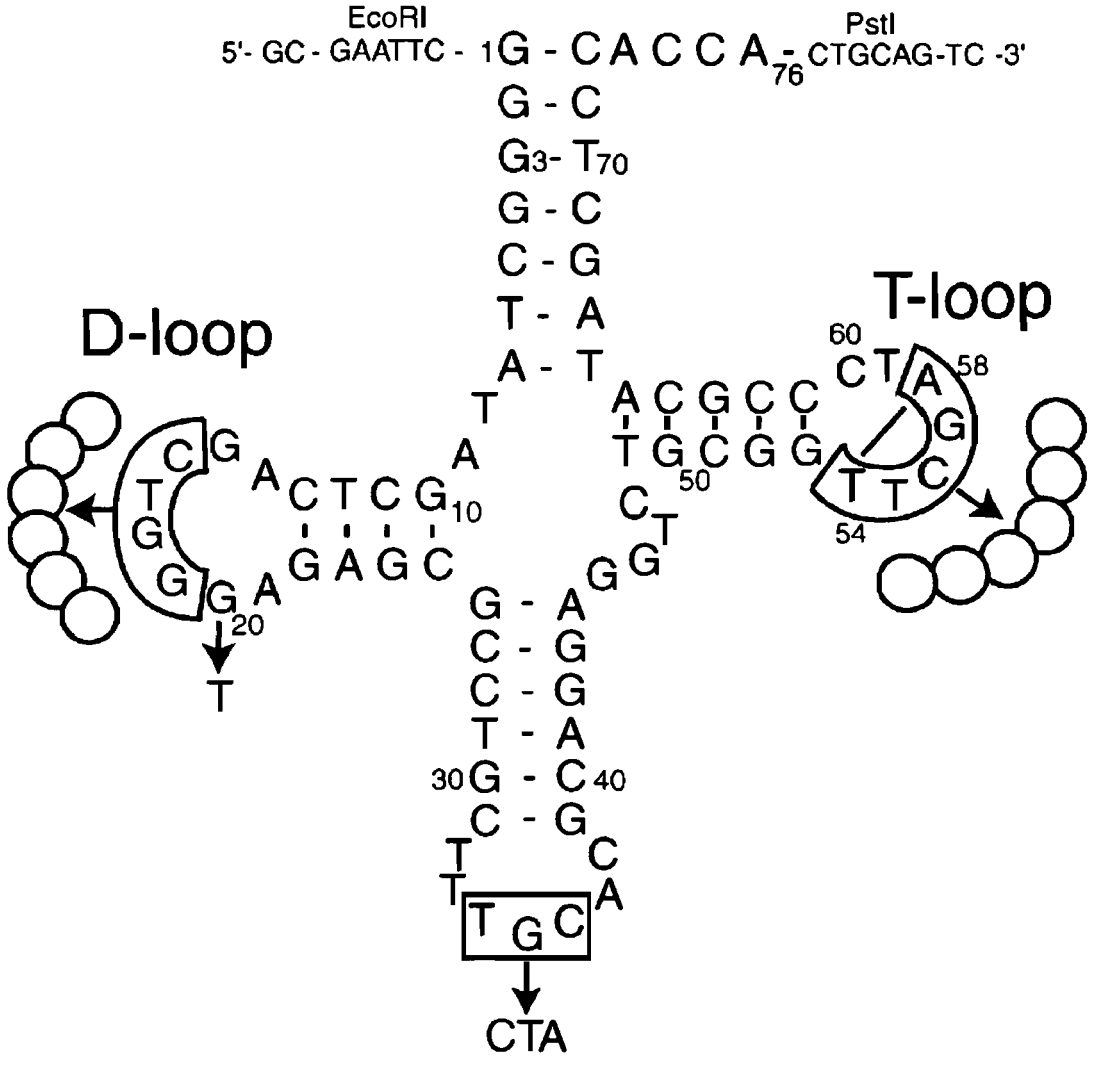

Figure 2.2 Construction of the tRNA gene library. In the nucleotide sequence of *E. coli* $tRNA^{Ala}_{UGC}$, each of the two enclosed regions, 16-19 in the D-loop and 54-58 in the T-loop, was replaced by six fully randomized positions, while nucleotide G20 and the anticodon TGC (boxed) were replaced by T20 and CTA, respectively. Nucleotides 54 and 58, which form the reverse Hoogsteen base pair in the T-loop, are connected by a line. The *EcoRI* and *PstI* restriction sites that are seen flanking the 5' and 3' termini were used for cloning the library into the pGFIB-1 plasmid.

2.2 MATERIALS & METHODS

2.2.1 Strains

The *Escherichia coli* strains TOP10 (Invitrogen) and XAC-1 (*F' lacI*$_{373}$*lacZ*$_{u118am}$ *proB*$^{+}$/*F*$^{-}$ *Δ(lac-proB)*$_{x111}$ *nalA rif argE*$_{am}$*ara*) were used, respectively, for cloning and selection of the suppressor tRNAs. The XAC-1 strain contains amber mutations in the genes *lacZ* and *argE* [NMKAM1986].

2.2.2 Construction of the combinatorial library and selection of suppressor tRNAs

The template oligonucleotide coding for the combinatorial tRNA library (Fig. 2) was synthesized at BioCorp Inc. (Montreal, Canada) amplified by PCR to produce the double-stranded DNA using primers 5'- CGAATTCGGGGCTATA- 3' and 5'-GACTGCAGTGGTGGAGT-3', and cloned into plasmid pGFIB-1 using *EcoRI* and *PstI* restriction sites, as described previously [BSFECC1998]. This plasmid provides a constitutive high level expression of a cloned tRNA gene [MM1986]. All enzymes were from New England Biolabs. Of 20 ml of the ligation mixture, 5 ml was electroporated into competent TOP10 cells, providing 4.5 X 10^6 colonies, i.e. about a quarter of the sequence complexity of the library. The plasmid

DNA was recovered using the Qiafilter Midiprep kit (Qiagen) and then transformed into competent XAC-1 cells. The positive clones were selected as blue colonies when grown on LB-agar containing ampicillin (100 mg/ml) and X-Gal (200 ml of the 20 mg/ml solution spread on top of each 150 X 15 mm plate). The plasmid DNA of these clones was extracted and retransformed into the XAC-1 cells to confirm the dependence of the phenotype on the presence of the plasmid. The ability of the selected tRNAs to suppress the nonsense mutation in gene *argE* was checked by plating the retransformed XAC-1 cells on minimal A medium without arginine.

2.2.3 Sequencing

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

2.2.4 Measurement of the β–Galactosidase activity

β-Galactosidase activity of clones with suppressor tRNA genes was determined as described by Miller [M1972] using overnight cultures grown in A

medium containing 0.4% glucose and 1 mM $MgSO_4$ to an A_{600} of 0.8-0.9. The values were obtained by averaging the measurements from three independent cultures and calculated as a percentage of the activity of the control $tRNA^{Ala}_{su+}$.

2.2.5 Presence of the Suppressor tRNAs in the cytosol and the aminoacylation level

To preserve the aminoacylated form of the tRNAs, the total cellular RNA was extracted under acidic conditions, as described previously [BSFECC1998]. To obtain the deacylated tRNA, 4 mg of the total RNA was mixed with 1.5 ml of 0.5 M Tris (pH 9.0), incubated for 30 min at 37°C and deposited on an acid polyacrylamide gel (6.5% polyacrylamide, 8 M urea, 0.1 M sodium acetate) together with the untreated fraction. The gel was run for 24 h at 300 V at 4°C in 0.1 M sodium acetate, after which the part of the gel around the xylene cyanol dye was transferred by electroblotting to a Hybond-N nylon membrane (Amersham). The membrane was hybridized with two radiolabeled DNA probes, one complementary to region 26-44 of the suppressor tRNAs, consisting of the anticodon stem and loop, and the other to region 34-53 of the *E. coli* 5S rRNA. The 5S rRNA probe was used to monitor the amount of total RNA in each sample. The hybridization was performed overnight at 37°C in 7% SDS, 0.25 M Na_2HPO_4 (pH 7.4), 1 mM EDTA (pH 8.0), 1% BSA using a Robbins hybridization incubator.

2.2.6 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 tRNAPhe [SM2000] was used as a starting conformation. The presumed structures of RH-GA or RH-GG were appended to the T-stem replacing base pair U54-A58. The other randomized nucleotides were arranged in a way to resemble the structure of the DT region in the normal tRNAs and, at the same time, to provide a reasonable system of hydrogen bonds and base-base stacking interactions. Each model was submitted to unrestrained energy minimization using the AMBER force field [PCCRCFSSWK1995] until an energy minimum was reached. Visualizations were done in a Silicon Graphics O2 computer.

2.3 RESULTS

2.3.1 The library Design

The library was built from *E.coli* $tRNA^{Ala}_{UGC}$ as a scaffold (Fig. 2.2). The choice was determined by the fact that the most important $tRNA^{Ala}$ identity element

for the cognate alanyl-tRNA synthetase, the G3-U70 base pair, was located in the acceptor stem, i.e. neither in the DT region nor in the anticodon, the sites that were modified in this study [MF1988, HS1988]. This would minimize the role of interaction with a particular aminoacyl-tRNA synthetase as a factor in the tRNA selection. To enable the tRNAs to recognize the amber stop codon UAG, the anticodon TGC in the gene was replaced by CTA. All five nucleotides of region 54-58 of the T-loop, which were known to be involved in conserved interactions either within the loop or with nucleotides of the D-loop, were fully randomized. Correspondingly, four nucleotides 16-19 of the D-loop, which could be involved in interactions with the T-loop, were also fully randomized. To prevent nucleotide G20 from substituting for either G18 or G19 in their interactions with the nucleotides of the T-loop, it was replaced by T20. To stimulate the formation of alternative structural patterns in the DT region, we added one and two nucleotides to the randomized regions of the T- and D-loops, respectively. Thus, in the design, the T-loop contained 8 nt, one more than in the standard tRNA structure, while the D-loop had 10 nt, which is not unprecedented for the cytosolic tRNAs [SHBIS1998]. Each loop had six randomized positions, providing for the total sequence complexity of a library of 1.7×10^7 variants.

2.3.2 Cloning and selection of functional clones

The tRNA gene library was synthesized chemically, amplified by PCR and cloned into the pGFIB-1 plasmid, as described previously [BSFECC1998]. The selection of active suppressor tRNA clones was done in the XAC-1 strain of *E. coli*, which had nonsense amber mutations in genes *lacZ* and *argE*. A successful suppression of the first mutation in the presence of 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-Gal) provides blue colonies, which was used for the primary identification of functional tRNA clones. Out of 3 X 10^4 clones screened, several dozen positive clones were selected, whose suppressor activity was confirmed by a subsequent retransformation and by suppression of the second mutation in gene *argE*, which converts the arginine-dependent cells into prototrophs. The β-galactosidase activity was evaluated quantitatively for each clone and compared to that of the control $tRNA^{Ala}_{su+}$. The latter tRNA was derived from the normal tRNAAla by changing the anticodon from UGC to CUA and cloned in the same plasmid as the other suppressor tRNAs. The nucleotide sequences of the selected tRNA clones, as deduced from their genes, are presented in Table 2.1. Only the sequences of those clones whose activity was at least 1% of the control are given. Comparison to the original design revealed six clones with a nucleotide deletion in the T-loop and three clones with a deletion in the D-loop, providing, respectively, for a seven member T-loop or a nine member D-loop. In two clones, K25 and K30, mutations affected the nonrandomized part of the T-loop, deleting, respectively, U59 and C60. No other mutation outside the randomized regions was found. For

eight clones arbitrarily chosen from Table 2.1, the in vivo level of the suppressor tRNA and of its aminoacylated form was determined by acid polyacrylamide gel electrophoresis followed by hybridization with a specific probe complementary to the anticodon stem and loop. For all suppressor tRNAs tested, the level in cytosol was detectable, although relatively low compared to that of $tRNA^{Ala}_{su+}$ (Fig. 3). For each clone, most of the tRNA was found in the aminoacylated form.

2.3.3 Analysis of the nucleotide sequences

In the experiments described above, on average, only one in every 1000 clones showed a detectable level of nonsense suppression activity. This indicated that the nucleotide sequence space available for the DT region was rather small. A systematic analysis of the sequences of the selected clones could help to reveal the rules imposed on the structure of this region in functional tRNAs.

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

Table 2.1 The nucleotide sequences and the β-galactosidase activity of the selected tRNA clones. The sequences are deduced from the genes. Only the D- and T-loops, where the sequences differ one from the other, are given. Positions of nucleotides 16-19 in the D-loop and 54-58 in the T-loop mark the beginning and the end of the each randomized region. Nucleotides forming the RH base pair in the T-loop are underlined. The activity of $tRNA^{Ala}_{su+}$ is taken as 100%.

We started the analysis with the 'quasi-normal' clones, those that had the normal 7 nt in the T-loop. Henceforth, we will call such molecules T7-tRNAs, in contrast to T8-tRNAs which have 8 nt in this loop. Analysis showed that in T7-tRNAs, the fifth position of the T-loop was always occupied by A and was the only invariable position in both randomized regions (Table 2.1). The second most conservative nucleotide was the first one of the T-loop, which in all sequences except K6 was U. The presence of U and A, respectively, in the first and the fifth position of the T-loop would allow the formation of the reverse Hoogsteen base pair U54-A58 (RH-UA), as in the normal tRNAs. Although in the normal tRNAs U54 is always modified to T, it is not yet known whether it is the case in the selected tRNAs. On the other hand, because this modification does not interfere with the ability of the base to form hydrogen bonds, its absence would not affect the formation of base pair U54-A58. Another conservative feature consistent with wild-type tRNA is the presence of a purine in position 57 of all but one T7-tRNA. Other randomized nucleotides, including all 6 nt in the D-loop, were essentially more diversified and did not seem to provide for a common structural pattern.

Among T8-tRNAs, half of the sequences (11 out of 22) also contained U in the first position of the T-loop (Table 2.1). If this U plays the same role as it does in T7-tRNAs, there should be an A a few nucleotides later that is able to form RH-UA with this U. Generally, this A could occupy either the fifth or the sixth position of the

T-loop, depending on the position of the additional eighth nucleotide. Analysis showed that in those T8-tRNAs whose T-loop started with U the only other conservative nucleotide was the A occupying the sixth position of the same loop. Thus, the formation of RH-UA by these two nucleotides would place the additional nucleotide in the region between them.

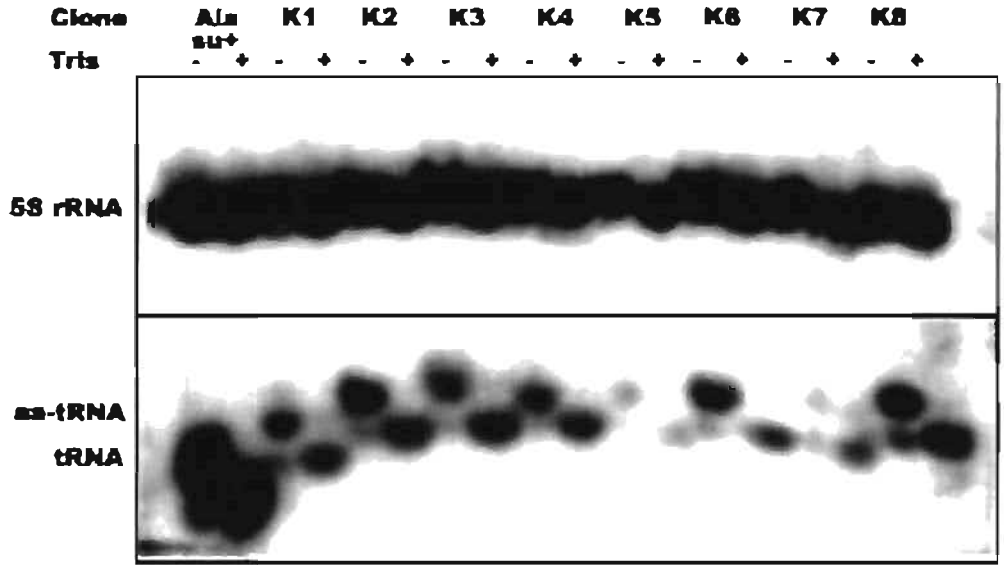

Figure 2.3 Northern blot showing the presence in the cytosol and the level of aminoacylation of some suppressor tRNAs. For each clone the - and + lanes correspond to the samples not treated and treated with Tris. In the - lanes the aminoacylated and deacylated forms of the suppressor tRNA move as individual bands, while in the + lanes the total tRNA is deacylated and the suppressor tRNA moves as one band. In all - lanes the bands corresponding to the aminoacylated form of the tRNA are much larger than those corresponding to the deacylated form and are comparable to the bands in the + lanes, representing the total amount of the suppressor tRNA. This indicates that in all clones most of the tRNA is present in the aminoacylated form. A smaller size of the bands of the suppressor tRNAs compared to $tRNA^{Ala}_{su+}$ indicates a notably lower presence of the selected tRNAs in the cytosol. 5S rRNA was visualized to monitor the amount of total RNA in each sample. Because the signal from 5S rRNA was much stronger than that from suppressor tRNAs, the upper and lower parts of the same membrane have been exposed, respectively, for 4 h and overnight. The nucleotide sequence of clone K8, due to its low β-galactosidase activity, is not included in Table 2.1, but is available upon request.

In all other T8-tRNAs, the first position of the T-loop was occupied by G (Table 2.1). If this G plays a structural role analogous to that played by U, its possible partner would occupy either the fifth or the sixth position of the T-loop. Neither of the two positions was conserved in these sequences: the fifth nucleotide was allowed to be either C or A, while the sixth was either A or G. To explore the abilities of both the fifth and the sixth nucleotides to pair with the first G, we looked for possible arrangements of three different dinucleotide combinations, GC, GA and GG, that would be close to the arrangement of U and A in RH-UA. For GC, we did not find any satisfactory arrangement. However, for both combinations GA and GG we found arrangements that are presented in Figure 2.4. In these arrangements, the G that is equivalent to U in RH-UA donates two hydrogen atoms for formation of hydrogen bonds with atom N7 of the other purine. This purine can be either A or G. In the latter case, an additional hydrogen bond can be formed between N2-H and O6 of the first and second G residues, respectively. The two arrangements GA and GG are superimposable in the sense that if one overlaps the positions of the glycosidic bonds of the first nucleotides, the glycosidic bonds of the second nucleotides in both arrangements would occupy about the same position. In the same sense, these two arrangements are fairly close to RH-UA. Accommodation of any of these arrangements based on the standard RH-UA would require a shift and rotation of one of the bases by only 1.5 Å and 20°,

respectively. Therefore, a replacement of RH-UA in the T-loop by either GA or GG would require only relatively minor changes in the positions of the neighboring nucleotides. To reflect the closeness of these GA and GG arrangements to RH-UA, we will call them RH-GA and RH-GG, respectively.

Further analysis revealed a few additional nucleotide combinations like CA and AA seen in Figure 2.4 that could also be arranged relatively closely to RH-UA while having a reasonable system of hydrogen bonds. Still, all these additional combinations were more distant from RH-UA than RH-GA or RH-GG and, therefore, their incorporation into the T-loop instead of RH-UA would cause greater changes in the conformation of the whole DT region. This latter aspect was expected to render these combinations less preferable in this place than RH-GA or RH-GG.

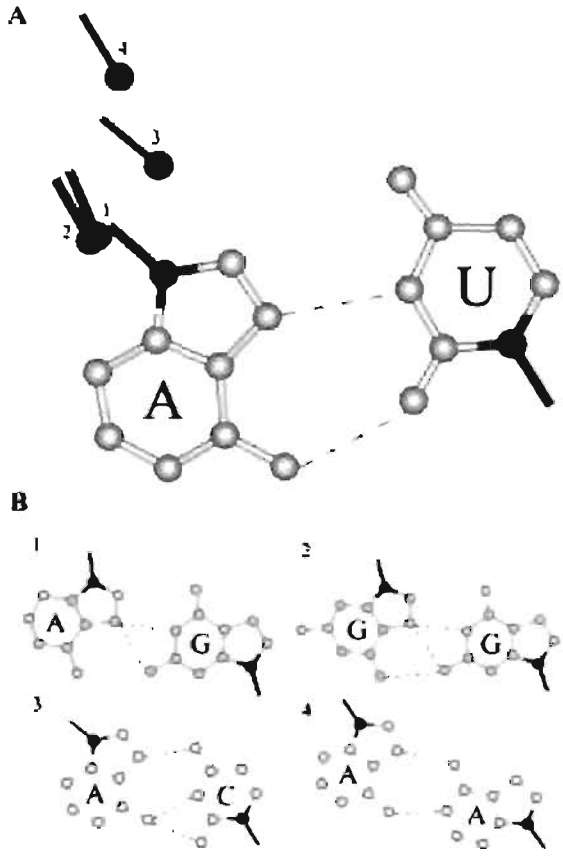

Figure 2.4 Juxtaposition of the bases in RH-GA, RH-GG and other alternative base pair candidates for replacement of RH-UA. (A) Positions of the glycosidic bonds in the alternative base pairs compared to that in RH-UA. In each base pair the position of the glycosidic bond corresponding to the base on the right is superimposed on that of U in RH-UA. The glycosidic bond of the other nucleotide will thus occupy a particular place depending on the structure of the base pair. The numbers indicating particular positions of the glycosidic bonds correspond to the base pairs in **(B)**.

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

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

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

2.3.4 Modeling of the tRNA structures

To confirm that the exchange of RH-UA for either RH-GA or RH-GG in the T-loop did not cause any steric problem, we modeled the structure of the DT region for several clones having either RH-GA or RH-GG. After unrestrained energy minimization, the bases constituting the RH base pair always retained their juxtapositions and the inter-base hydrogen bonds, as one can see in the example of the model for clone K31 (Fig. 2.5). Comparison of the models with the structure

of the yeast $tRNA^{Phe}$ [SM2000] showed that the whole region that included the T-stem and the RH base pair (RH-GA or RH-GG in the models and RH-UA in tRNAPhe), as well as nucleotides 59 and 60, was superimposable in all structures.

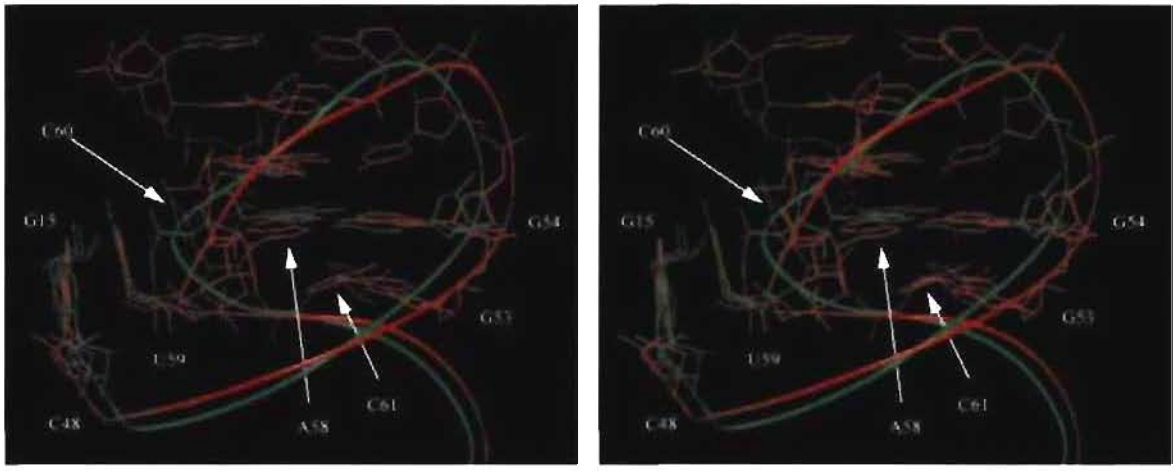

Figure 2.5 The model of the structure of the DT region for clone K31 (red) superimposed on the corresponding region in yeast tRNAPhe (green). The figure also includes the T-stem and the tertiary base pair 15-48. For both tRNAs, the ribbon follows the sugar-phosphate backbone. Explicitly shown are base pairs 15-48 and 54-58 and nucleotide 59 in $tRNA^{Phe}$, as well as all nucleotides of the DT region and pair 15-48 in K31. Comparison of the modeled structure with $tRNA^{Phe}$ demonstrates a good superposition of the T-stem and base pairs RH and 15-48, as well as nucleotide 59. The proper arrangements of the nucleotides in the RH base pair thus guarantees the proper position of nucleotide 59, whose stacking to base pair 15-48 would fix the juxtaposition of the two helical domains known as the L-form. Still, one can notice a difference in the conformation of the backbone in the two structures, which is highest between nucleotides 58 and 59. Such a difference makes a universal interaction of this region with a particular protein factor unlikely.

2.4 DISCUSSION

The results presented here show that a tRNA could be functional even if the structure of its DT region is substantially modified compared to the standard.

Although for all selected clones the efficiency of the nonsense codon suppression was lower than for $tRNA^{Ala}_{su+}$, it was strong enough to provide a level of β-galactosidase synthesis sufficient to change the color of the colonies in the presence of X-Gal and to allow cell growth without external arginine. Additional examination of several clones showed that suppressor tRNAs had a detectable in vivo level and existed mainly in the aminoacylated form.

The nucleotide sequences of the selected suppressor tRNAs demonstrated a range of diversity never seen in the natural cytosolic tRNAs. In spite of this, the selected tRNAs constituted only a tiny fraction of the whole tRNA gene library, which implied the existence of strong constraints imposed on the structure of functional tRNAs. To elucidate these constraints, we undertook a comparative analysis of the nucleotide sequences of the selected tRNAs. It may be a little surprising that among the selected clones there were no clones having the wild-type sequence pattern. On the other hand, the wild-type sequence G18-G19-...-U54-U55-C56-R57-A58 is expected to appear on average only once in 8000 clones. Moreover, this probability can easily get beyond the technically detectable level if some additional requirements are imposed on the identity of the nucleotides flanking the conservative dinucleotide G18-G19 in the D-loop and on the additional eighth nucleotide in the T-loop. For most of the randomized nucleotides, our analysis did not find any obvious common pattern. The only exception consisted of

the first and the last randomized positions in the T-loop, which were always able either to form RH-UA, analogous to base pair U54-A58 in the normal tRNAs, or to mimic it closely via formation of RH-GA or RH-GG. Modeling experiments showed that a replacement of RH-UA with either RH-GA or RH-GG did not cause any major rearrangement in the conformation of the DT region and provided for stable, sterically reasonable tRNA structures. Because such an RH base pair can be formed in all selected tRNAs, its existence is judged to be one of the most important requirements imposed on the structure of the DT region in a functional tRNA. In fact, this requirement has been the only one satisfied in all selected tRNAs, which allows us to conclude that the preservation of a RH base pair in the T-loop is more important for tRNA function than that of other universal elements, including inter-loop base pairs G18-Y55 and G19-C56.

Different explanations of the importance of the RH base pair for tRNA function can be suggested. For example, this base pair could be involved in a specific, vitally important interaction with a protein or other factor and thus should be preserved as such. However, a specific interaction like this would probably not tolerate an exchange of RH-UA for either RH-GA or RH-GG, because the juxtaposition of the glycosidic bonds and, therefore, the conformation of the backbone in the two latter base pairs, however close it is to that in RH-UA, is still notably different. Moreover, the three base pairs have different chemical groups

exposed on the surface and thus are unlikely to be recognized by the same factor. In another, more probable explanation, an RH base pair is needed to stabilize a particular conformation of a neighboring region in the tRNA structure and thus to enable this region to serve its function. We do not expect this region to include the top of the T-loop closed by the RH base pair or the randomized part of the D-loop. Indeed, in the selected tRNAs, these regions vary in length and in nucleotide sequence and do not seem to have a universal structure. There is, however, another region, dinucleotide 59-60 at the end of the T-loop, whose position may need a particular structure of the RH base pair. This dinucleotide bulges from the double helical stem between base pairs 54-58 and 53-61. Therefore, its conformation is to a great extent determined by the positions of the flanking nucleotides 58 and 61, which, in turn, depend on the structures of base pairs 54-58 and 53-61, respectively. Thus, the presence of base pair 54-58 with the correct juxtaposition of the bases is necessary for the proper positioning of dinucleotide 59-60. The reverse Hoogsteen base pair U54-A58 perfectly suits this purpose. However, as our results show, combinations GA and GG can also be arranged in an appropriate way. In the modeled tRNAs containing these combinations, the position of dinucleotide 59-60 remains virtually the same as in the standard tRNA structure (Fig. 2.5).

We have already suggested [SLC1997] that the bulged dinucleotide 59-60 plays a crucial role in maintenance of the general shape of tRNA. Indeed, in the normal tRNAs, nucleotide 59 stacks on the tertiary base pair 15-48, which constitutes the last stacked layer of the D/anticodon helical domain (Figs 2.1 and 2.5). This interaction determines the juxtaposition of the two domains, the D/anticodon and the acceptor/T, i.e. the general geometry of the tRNA L-form, and is invariable in all normal tRNAs [SLC1997, IS1999]. Thus, the importance of maintaining the standard juxtaposition of the helical domains within the tRNA L-form would justify the necessity to preserve the conformation of dinucleotide 59-60 via formation of the reverse-Hoogsteen base pair 54-58. Because the bases of nucleotides 59 and 60 are mainly involved in stacking interactions between themselves and with nucleotides 15 and 48, and not in hydrogen bonding, their identity is not that important. This reflects the partial variability of nucleotides 59 and 60 in the normal tRNAs [SHBIS1998], and also fits the fact that replacements U59A in clones K25 and K30 and C60U in clone K30 did not impair the tRNA function (Table 2.1).

The importance of this base pair for tRNA function, especially compared to other randomized elements of the DT region, correlates well with the data of Azarenko *et al.* [NHU1994] on the efficiencies of mutants of yeast Phe-tRNA^{Phe} in different partial reactions of the tRNA functional cycle. According to these data, the

efficiencies of the mutants in tertiary complex formation with factor Tu and GTP, in binding to the A and P site of the poly(U)-programmed ribosome and in peptide formation, are generally more sensitive to nucleotide replacements in pair 54-58 than in pairs 18-55 and 19-56. There have also been reports of using tRNA libraries with randomized positions in the DT region for selection of clones in vitro by affinity to either the phenylalanyl- or glutamyl-tRNA synthetase and to the EF-Tu factor [PBSU1993, BSP2000]. Interestingly, in none of these studies was the necessity for a RH base pair in the T-loop detected. However, because different steps of the tRNA functional cycle are probably not equally dependent on the proper position of the two helical domains, concentration on only some steps of this cycle would not necessarily reveal all sequence requirements for a fully functional tRNA. Moreover, the mutations in the tRNA clones selected by affinity for a particular protein may be detrimental not only for other steps of the functional cycle not involved in this selection, but even for this very step if, for example, they hinder the dissociation of the complex [LWU2001, AU2002]. We here used an alternative approach of tRNA selection in vivo based on its suppressor activity. This guarantees that the selected clones are correctly transcribed, processed and folded, that they are able to interact productively with all necessary factors of the protein biosynthesis machinery, including the aminoacyltRNA synthetase and EF-Tu, as well as the UAG-charged ribosome. The sequence requirements revealed in this way have been balanced between all these steps. Another essential aspect of

our approach relates to the original design of the library. As one can see in Table 2.1, among T7-tRNAs only clone K6 does not have RH-UA, while among T8-tRNAs there are more than half of such clones. It may reflect the existence of certain constraints on fitting either RH-GA or RH-GG into the seven member T-loop, which would relax when an additional eighth nucleotide is added. Therefore, if in the T-loop of the original design the normal 7 nt were preserved, most probably we would have seen an overwhelming majority of the selected clones having RH-UA, only sporadically intermingling with those having RH-GA or RH-GG. This would have made it more difficult to understand that it is the conformation of the sugar-phosphate backbone rather than base pair U54-A58 per se which is crucial for tRNA function. Although the introduction of an additional base into the T-loop probably has some negative effect on the suppressor activity, most importantly it increases the chance of selecting alternative structures. The elucidation of the common structural pattern in the selected tRNAs designed in this way has allowed us to recognize the RH base pair as the most important aspect that must be preserved, even when the structure substantially deviates from the standard.

In contrast to the nucleotides composing the RH base pair, other randomized nucleotides do not seem to have a common structural pattern. This means that the existence of a particular universal structure of this region is not required for tRNA function. On the other hand, it does not mean that this region is

not structured or does not play any functional role. Instead, there are indications that it has a particular structure and that this structure is important for tRNA function, even if it is not the same in all selected tRNAs. First, as discussed above, the seven member T-loop may limit the use of RH-GA and RHGG, so that these latter base pairs are able to replace RH-UA in only a fraction of all successful T7-tRNAs. This ability will thus depend on the identity of other randomized nucleotides. Also, the activities among the selected clones differ by almost 40 times, despite the presence of a RH base pair in all of them. This indicates the existence of other factors within the tRNA structure that affect the activity. Because these clones differ one from another only in the randomized regions, we have to conclude that other randomized nucleotides not involved in the RH base pair play a role in tRNA function. Still, an overall inspection of the nucleotide sequences shown in Table 2.1 has provided no obvious consensus pattern able to explain the differences in the activity. In normal tRNAs, the randomized nucleotides not involved in the RH base pair take part in the interactions between the D- and T-loops. These interactions affect the overall stability of the tRNA tertiary structure, but may also be directly involved in a particular tRNA-related process. Similar interactions could exist in the selected suppressor tRNAs as well, although they would be different in different clones. An example of such interactions is seen in the model of K31 in Figure 2.5. A systematic analysis of the possibility of forming these interactions in the selected tRNA clones is now in progress and will be

published elsewhere. One should admit, however, that this analysis may not be able to explain the existing differences in tRNA activity. Indeed, there are many reasons that could negatively affect tRNA activity, such as formation of an alternative secondary structure, a higher susceptibility to a nuclease activity, a too low or too high affinity for the cognate aminoacyl-tRNA synthetase or for the elongation factor Tu, etc., and for each clone, the real reason can be different. A complete understanding of this phenomenon needs an analysis of the behavior of individual tRNA clones at each step of their functional cycle.

2.5 ACKNOWLEDGEMENTS

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

CHAPITRE 3 : Apport scientifique de l'article

L'apport principal de cet article est l'identification de la paire de base reverse-Hoogsteen U54-A58 comme élément central de la région DT, et la détermination de son rôle essentiel dans la fixation des deux domaines hélicoïdaux.

Lorsque la détermination de la structure de l'ARNt a été établie, il y a plusieurs années, une série d'interactions a pu être observé pour la première fois. Toutefois, aucune allégation quant à l'importance relative ou au rôle de maintient de la structure en « L » n'avait été avancée ou suggérée. Suite à nos récentes études, nous sommes parvenus à associer un rôle certain pour cette paire de base au niveau de la structure tertiaire de l'ARNt ainsi qu'au niveau de sa fonction. En effet, nos travaux ont permis de démontrer qu'en allouant le bon nombre de nucléotides à l'appendice, cette paire de base fait ainsi en sorte que le nombre de couches du domaine I reste constant. Par le fait même, elle contribue au maintient de la structure tertiaire en forme de « L » qui est essentielle à la fonctionnalité de la molécule.

Ce lien entre la paire de base U54-A58 et la fonctionnalité de l'ARNt s'appuie sur le fait que, la conformation en « L » est nécessaire afin d'obtenir une

distance de 70 Å [SSA1994] entre les deux domaines fonctionnels de la molécule et pour leur permettre une orientation réciproque adéquate.

Ce qui fut d'abord intéressant de constater fut l'étonnante diversité parmi les clones sélectionnés de la librairie combinatoire. Jamais chez les ARNts cytosoliques une telle diversité n'a pu être observée. Malgré cette variété, la moitié des clones arboraient une paire de base U54-A58 analogue à la séquence et à la structure canonique. Chez les autres clones, cette paire de base est remplacée par deux alternatives : G54-A58 ou G54-G58. Par la modélisation moléculaire interactive, nous avons pu démontrer que les nucléotides composant ces deux paires de base pouvaient être orientés d'une façon très similaire à la façon dont sont orientés les nucléotides U54 et A58 dans la paire de base reverse-Hoogsteen caractéristique de la structure du type sauvage.

De cette façon, la structure de l'ARNt reste sensiblement la même et la molécule peut remplir sa fonction, certes pas aussi bien que le fait le type sauvage, mais assez pour permettre la survie de la cellule, ce qui est déjà un énorme accomplissement.

Ainsi, cet article apporte une nouvelle vision des règles régissant la structure des ARNs, suggérant également que cette nouvelle conception pourrait

être appliquée à d'autres molécules. Ce faisant, il élargit notre connaissance et pourra être utile à la compréhension des mécanismes régissant la formation et le maintient de la structure tertiaire d'autres ARNs plus imposants, tel que le ribosome, et par conséquent l'effet de ces mécanismes sur leurs fonctions.

Il est intéressant de mentionner que le choix des outils pour ces recherches ont permis une approche unique et avantageuse de l'hypothèse du départ. Effectivement, l'utilisation de librairies combinatoires a permis la sélection d'un plus grand nombre de clones et ainsi l'observation d'une multitude de possibilités. La stratégie informatique s'est également révélée un bon choix dans cette étude puisqu'elle a permis la validation de notre modèle de départ.

Les prochaines étapes intéressantes qui pourraient faire suite à ce travail pourraient impliquer l'étude des autres éléments de structures de l'ARNt. En effet, la détermination de la structure de l'ARNt a permis l'observation de d'autres interactions. En déterminant l'importance relative de ces autres paires de base nous pourrions faire un autre pas vers l'avant dans la compréhension des mécanismes régissant la structure et la fonction de l'ARNt. Tel que mentionné précédemment, il serait également intéressant d'étendre ces projets d'études de structure à d'autres molécules d'ARN.

Il est maintenant reconnu que le rôle de l'ARN s'avère déterminant dans plusieurs mécanismes vitaux de la cellule. Une meilleure compréhension de leur comportements, et ce dans différents environnement, nous amèneraient à approfondir nos connaissances sur ce sujet et, qui sait, à développer de nouveaux outils et applications.

CHAPITRE 4 : Specific and Non-Specific Purine Trap in the T-loop of normal and suppressor tRNAs (published article)

Félix R. Doyon, Ekaterina I. Zagryadskaya, Jianhong Chen, Sergey V. Steinberg

Département de Biochimie, Université de Montréal, Mtl, Qc H3C 3J7, Canada

doi:10.1016/j.jmb.2004.08.025 *J. Mol. Biol.* (2004) 343, 55–69

To elucidate the general constraints imposed on the structure of the D and T-loops in functional tRNAs, active suppressor tRNAs were selected in vivo from a combinatorial tRNA gene library in which several nucleotide positions in these loops were randomized. Analysis of the nucleotide sequences of the selected clones demonstrates that most of them contain combination U54-A58 allowing the formation of the standard reverse-Hoogsteen base-pair 54–58 in the T-loop. With only one exception, all these clones fall into two groups, each characterized by a distinct sequence pattern. Analysis of these two groups has allowed us to suggest two different types of nucleotide arrangement in the DT region. The first type, the so-called specific purine trap, is found in 12 individual tRNA clones and represents a generalized version of the standard D-T loop interaction. It consists of purine 18 sandwiched between the reverse-Hoogsteen basepair U54-A58 and purine 57. The identity of purine 18 is restricted by the specific base-pairing with nucleotide 55. Depending on whether nucleotide 55 is U or G, purine 18 should be, respectively,

G or A. The second structural type, the so-called non-specific purine trap, corresponds to the nucleotide sequence pattern found in 16 individual tRNA clones and is described here for the first time. It consists of purine 18 sandwiched between two reverse-Hoogsteen base-pairs U54-A58 and A55-C57 and, unlike the specific purine trap, requires the T-loop to contain an extra eighth nucleotide. Since purine 18 does not form a base-pair in the nonspecific purine trap, both purines, G18 and A18, fit to the structure equally well. The important role of both the specific and non-specific purine traps in the formation of the tRNA L-shape is discussed.

Keywords: RNA motif, molecular modeling, L-shape, selection *in vivo*

4.1 INTRODUCTION

Analysis of the available crystal and solution structures of different RNA molecules and RNA– protein complexes clearly shows that although unpaired non-helical regions usually account for less than a half of the nucleotide sequences, their role in the formation of RNA structure and in its function is crucially important. Often, specific interactions of such regions with other parts of the molecule ensure the particular juxtaposition of double helices within the unique functionally active tertiary structure [CGPZGSKCD1996, DBMD2001, QR1976, KSMS2001, NIBMS2001, W2002]. On other occasions, irregular non-helical regions can

recognize particular elements in another molecule, thus mediating intermolecular contacts [PFM1994, HP2000]. Therefore, the understanding of how nucleotide sequence forms RNA tertiary structure and how the latter defines the function is impossible without clarification of the particular role played by the non-helical irregular elements and of the rules that govern their formation. Our knowledge on the structure of such elements is quite fragmentary. Several structural motifs have been identified in the tertiary structures of different RNAs [CGPZGSKCD1996, DBMD2001, QR1976, KSMS2001, NIBMS2001, W2002, PFM1994, HP2000, HM1999, BRD1999, M1999, CD1998, WF2000]. Also, a few irregular motifs, mostly representing distorted helices, have been studied by X-ray crystallography and NMR-spectrometry [HH2003, SKT1991, ST1993]. Much less is known about the sequence and context requirements for the formation of these motifs and the particular role they play in the tertiary structure of functionally active RNAs.

Probably, for the first time, the importance of irregular elements in RNA was recognized 30 years ago when the tertiary structure of yeast $tRNA^{Phe}$ was elucidated [SQMSKKSR1974, RLFRBCK1974]. In that structure, the two loops, D and T, were shown to interact specifically at the outer corner of the molecule in the so-called DT region (Figure 4.1). This interaction brings together two parts of the nucleotide sequence, which are separated by more than 30 nucleotides and are positioned in different domains at the opposite ends of the secondary cloverleaf

structure. The resulting conformation is characterized by the perpendicular juxtaposition of the two helical domains, D/anticodon and acceptor/T, known as the tRNA L-shape. The interaction between the two loops is highly conserved and has been the subject of a number of studies [NHU1994, PPFG1993, LBKCRWM1998, DW2003, BSP2000]. The presence of such elements as the U-turn between ψ55 (ψ stands for pseudouridine, a post-transcriptional modification of U) and C56, the unusual non-Watson–Crick base-pairs T54-A58 and G18-Ψ55, the mutual intercalation of fragments 57-58 and 18-19, the two-nucleotide bulge 59-60 sandwiched between the T-stem and the D-domain (Figure 4.1) makes the DT region one of the most structurally diverse in the whole tRNA and raises questions as to the particular role played by each element in the structure of this region and of the whole tRNA.

To address these questions, we recently undertook an in vivo selection of active suppressor tRNAs from a combinatorial gene library [ZDS2003] (K-library) in which six nucleotides in each of the two loops were randomized, while the T-loop contained an additional eighth nucleotide compared to the standard seven (Figure 4.2). Henceforth, the six randomized nucleotides of such an extended T-loop are numbered from 1* to 6*.

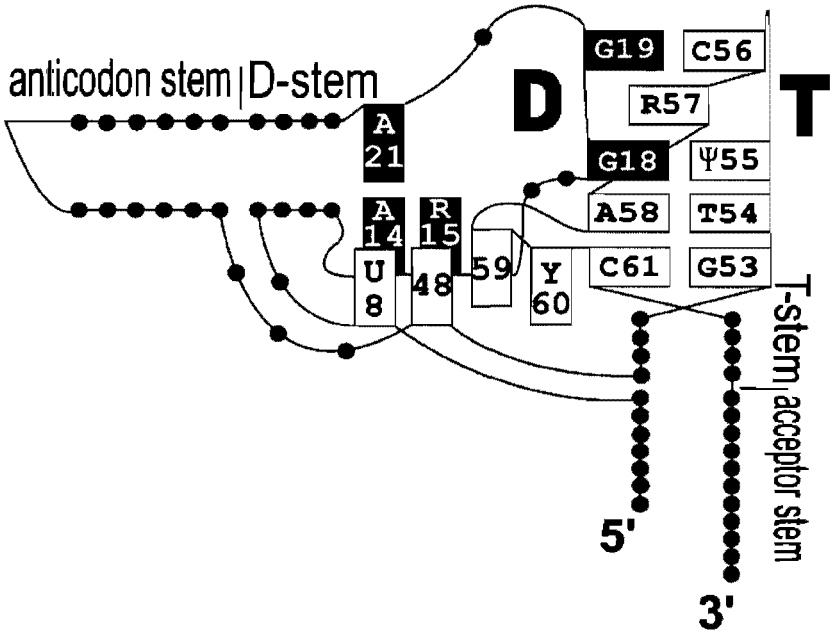

Figure 4.1 The structure of the DT region in the context of the tRNA L-shape. Rectangles represent individual nucleotides of the DT region involved in stacking or base-pairing with another nucleotide. Rectangles representing nucleotides of the D-loop are black, while those representing nucleotides of the T-stem and loop as well as nucleotides 8 and 48 are white. All other nucleotides are shown as black dots. The identities of conservative and semi-conservative nucleotides are indicated; R, Y and J stand for purine, pyrimidine and pseudouridine, respectively. Nucleotides T54 and A58 in the T-loop form the reverse-Hoogsteen base-pair, while nucleotides 59 and 60 form the T-loop bulge. This bulge is squeezed between the T-stem and the last stacking layer of the D/anticodon domain represented by base-pair 15–48 and plays a key role in the fixation of the perpendicular arrangement of the two helical domains in the tRNA L-shape. Purine 57 intercalates between the two inter-loop base-pairs G18-Ψ55 and G19-C56.

Analysis of the selected tRNA clones showed that the only element that appeared in the randomized regions of most sequences was base-pair U1*-A6* equivalent to base-pair T54-A58 in the canonical tRNA structure. This base-pair is formed accordingly to the scheme [UA *trans* W.C./Hoogsteen] [LW2001] and is called here reverse-Hoogsteen base-pair. The formation of this base-pair allocates

two unpaired nucleotides for the bulge region, exactly as in the normal tRNAs. Additional study, which involved selection of suppressor tRNAs from another combinatorial library, showed that the bulge could be extended up to three nucleotides when the last stacked layer of the D-domain, represented by tertiary base-pair 15–48, was eliminated [ZKS2004]. In all cases, the total number of stacked layers in the DT region was always the same, despite the changes in both the D-domain and the T-loop bulge. This would guarantee the standard juxtaposition of the two helical domains of the tRNA, which is a prerequisite of the tRNA functionality. The fact that the number of the nucleotides in the T-loop bulge is determined by the position of the reverse-Hoogsteen base-pair 54–58 demonstrates the key role played by this base-pair in keeping the tRNA L-shape.

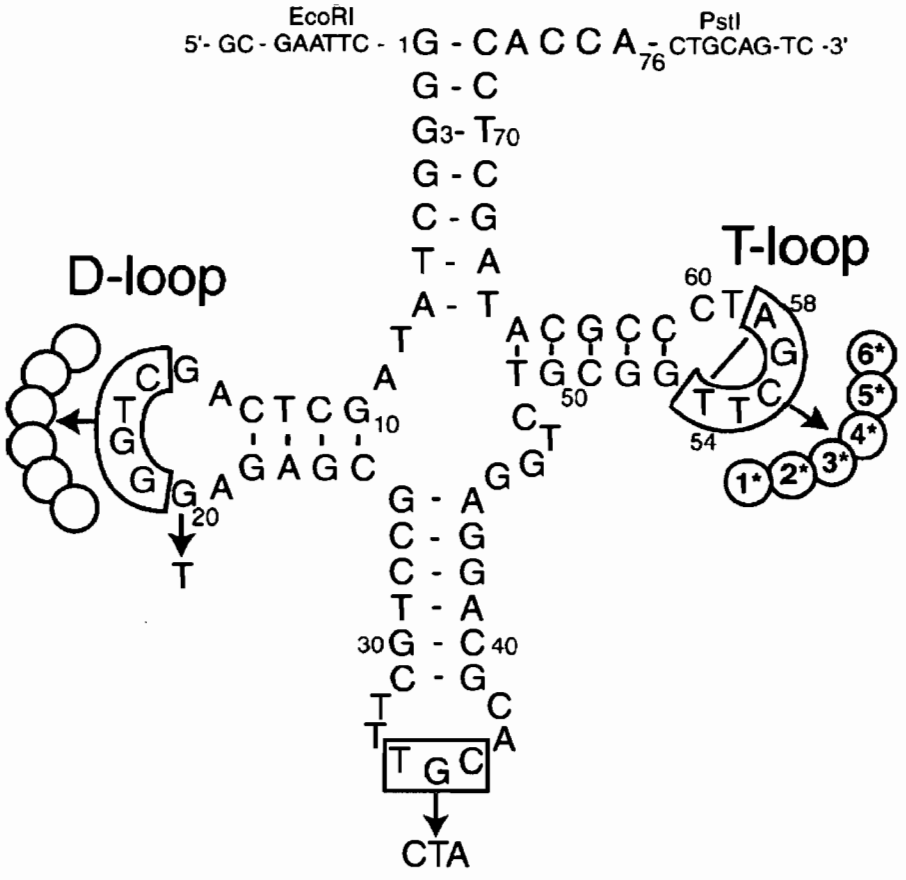

Figure 4.2 The design of the tRNA gene library (K-library) [ZDS2003]. The design of the K-library was based on the sequence of $tRNA^{Ala}$ UGC. Six randomized nucleotide positions (shown by circles) were introduced in each of the D and T-loops. In the T-loop, these positions are numbered from 1* to 6*. Compared to $tRNA^{Ala}$ UGC, the D and T-loops were extended by two and one nucleotide, respectively. Additional changes in the D- and anticodon loops are indicated by arrows. The *EcoRI* and *PstI* restriction sites flanking the tRNA sequence were used for cloning the library.

Unlike the reverse-Hoogsteen base-pair and the T-loop bulge, the top region of the T-loop closed by the reverse-Hoogsteen base-pair demonstrated considerable sequence variability in the selected clones and did not reveal any common pattern. This allowed us to conclude that the particular conformation of

this region was not crucially important for the function of suppressor tRNAs. It was not possible at the time to say whether this region contained any important elements, mainly because of an insufficient number of sequences. Since then, we have performed an additional selection of suppressor tRNAs from the same K-library, whose total number has now reached 50. Of these clones, 29 contain nucleotides U1* and A6* in the T-loop, which suggests the formation of the reverse-Hoogsteen base-pair between them, while all other clones contain G1* and either A6* or G6*, indicating the existence of an alternative pattern. Because the presence of the reverse-Hoogsteen base-pair UA provides an important structural constraint that simplifies the problem of elucidating the structure of the T-loop and of the whole DT region, we decided, at the first step, to limit our analysis to the sequences containing the U1*-A6* combination.

Here we present the theoretical analysis of the sequences of these tRNAs and molecular modeling of their tertiary structures, which has allowed us to identify two major types of nucleotide arrangement at the top of the T-loop. We call these types the specific and non-specific purine traps. The two types have very different sequence requirements and are easily recognizable in the nucleotide sequences. Both types allow the T-loop to interact with a part of the D-loop, although each type does it in its own way. The canonical structure of the T-loop constitutes a particular case of the specific purine trap, while the non-specific purine trap is described here

for the first time. Comparison of the two identified types has helped us to formulate the most general requirements imposed on the structure of the DT region in functional tRNAs.

4.2 RESULTS

4.2.1 The selected clones

The design of the suppressor tRNA combinatorial library was based on the sequence of the *Escherichia coli* $tRNA^{Ala}_{UGC}$ (Figure 4.2). In this design, the D and T-loop contained, respectively, two and one additional nucleotide. The extension of the loops aimed at providing them with additional conformational flexibility, which could increase the probability of selection of new as yet unknown structural forms. For the D-loop, whose length in different cytosolic tRNAs is variable, this extension did not exceed the limits of the standard pattern, while for the T-loop, whose length is highly conserved, the addition of one more nucleotide singles out, the designed suppressor tRNAs out of all other tRNAs. Six positions in the D-loop and six positions in the T-loop were randomized. The randomized positions in the T-loop are numbered from 1* to 6*. After screening of about 50,000 library clones, totally, 50 tRNA clones were selected, of which 29 contained combination U1*-A6*. This combination allowed the formation of the reverse-Hoogsteen base-pair analogous

to T54-A58 in the normal tRNAs. The nucleotide sequences of all U1*-A6*-containing clones are shown in Table 4.1. Although the suppressor activities of the selected clones are not very high, they are well above the background level, which was shown to be below 0.01% [ZKS2004].

Clone	*D-loop*		*T-loop*	*Activity*
$tRNA^{Ala}_{CUA}$	AG CUGG	GA	UU CGA	UC 100%
Type I				
T7-tRNAs				
K25[a]	AG GAACGC	UA	UG AAA	AC 17.7%
K15[c]	AG GCNUNU	UA	UG AAA	UC 11.0%
K29[c]	AG GAAAAA	UA	UG GGA	UC 5.1%
K3[a]	AG AACGAA	UA	UG AAA	UC 4.1%
K48	AG ACRGAC	UA	UG GAA	AC 2.5%
T8-tRNAs				
K2[c]	AG AAAGAC	UA	UGAGCA	UC 7.9%
K23[a]	AG UAAGGU	UA	UGCCAA	UC 5.9%
K42	AG UAGACA	UA	UGACAA	UC 3.8%
K39	AG GANGAA	UA	UGACGA	UC 3.4%
K49	AG AAAAGA	UA	UGCAAA	UC 2.2%
K12	AG CUACAU	UA	UGAAAA	UC 0.4%
K34	AG N	GA	UGGCGA	UC 1.4%
Type II				
K27[c]	AG UGAAAU	UA	UAGCCA	UC 9.9%
K24[a]	AG AAAAAC	UA	UAGCCA	UC 6.0%
K26[c]	AG AACAAC	UA	UAAACA	UC 3.9%
K47	AG AGAGCA	UA	UAGCCA	UC 3.3%
K20[a]	AG GAGAUC	UA	UAGCCA	UC 3.2%
K40	AG AAACAC	UA	UAGCCA	UC 2.8%
K36	AG AAAAAA	UA	UAGCCA	UC 2.7%
K18[a]	AG AACAAA	UA	UAAACA	UC 2.5%
K33	AG GAAGAA	UA	UAGUCA	UC 2.0%
K19[a]	AG ACAAC	UA	UAUACA	UC 2.0%
K5[a]	AG CGAAGA	UA	UAGCCA	UC 1.7%
K38	AG AAAUAC	UA	UAAACA	UC 1.5%
K51	AG ACCAAA	UA	UAACCA	UC 1.4%
K7[a]	AG GACAAA	UA	UAACCA	UC 1.3%
K1[a]	AG GAGAAC	UA	UAACCA	UC 1.3%
K41	AG N	CA	UAAACA	UC 4.4%
Non-identified type				
K30[a]	AG UGAGGA	UA	UC CAA	AU 10.8%

Table 4.1 Sequences of the D-and T-loops and suppressor activities of the selected tRNA clones. The tRNA sequences outside the D and T-loops are identical in all clones and are not shown. The randomized regions in the D and T-loops correspond,

respectively, to regions 16–19 and 54–58 of the standard tRNA structure. The nucleotides composing the reverse- Hoogsteen base-pair U54(1*)-A58(6*) are shown in magenta. In $tRNA^{Ala}$ CUA, U55 and G18, which form a base-pair, are green. In the Type I tRNAs, G55(2*), which corresponds to the standard U55, is also green. A nucleotide equivalent to the standard U55 does not exist in the Type II tRNAs. In the D-loop of both types, any purine suitable for the role of purine 18 is shown in green as well. The nucleotides that stack on top of purine 18 are yellow. For $tRNA^{Ala}_{CUA}$ and for the Type I tRNAs it is purine 57(5*), while for the Type II tRNAs it is base-pair A2*-C5*. The nucleotides forming the standard base-pair 19–56 or its equivalent 19–4* are red. For the Type II tRNAs, nucleotides 3* and 4* fitting to the common pattern (purine for 3* and A/C for 4*) are blue. The same colors are used for the same nucleotides in Figures 4.4 and 4.5. The nucleotide sequences that have been published earlier [ZDS2003] are marked by a. The suppressor activity of $tRNA^{Ala}_{CUA}$ was taken for 100%.

4.2.2 Type I: a quasi normal pattern

For the reason discussed below, none of the selected clones had the canonical tRNA sequence pattern. On the other hand, the initial inspection of the sequences revealed that at least some clones contained elements identical or similar to those existing in the normal tRNAs. We started our analysis with the five "quasi-normal" clones that had the standard seven nucleotides in the T-loop due to a spontaneous deletion. Henceforth, we will call such molecules T7-tRNAs, in contrast to T8-tRNAs, which have eight nucleotides in the T-loop, as presumed by the library design (Figure 4.2). For the T7 clones we will use the standard nucleotide numbering, while in the T8 clones, the randomized nucleotides of the T-loop are numbered from 1* to 6*.

In all T7 clones, nucleotide 57 is a purine, as in the normal tRNAs, while the identities of nucleotides 55 and 56 always differ from the standard U(ψ)55 and C56 known to be involved in specific interactions, respectively; with nucleotides G18 and G19 of the D-loop. Interestingly, in all five sequences, position 55 is occupied by G. Although none of the cytosolic tRNA has G55, the existence of this nucleotide in the T-loop of some mitochondrial tRNAs [DKDW1995], viral tRNA-like structures [FG1995] and in the T-loop-like structures of the RNase P [KM2003] and of the ribosomal RNA [LCG2003, NF2002] is well-documented. In almost all such cases, replacement U55G coexists with replacement G18A, which infers the formation of base-pair A18-G55 instead of the standard G18- U(ψ)55. The G18-U555A18-G55 co-variation was also observed by Peterson *et al.* in the in vitro selection of $tRNA^{Phe}$-derived variants able to be aminoacylated by phenylalanine-tRNA synthetase and to bind to the elongation factor Tu [PBSU1993]. In the two cases of the T-loop-like structure existing in the RNase P [KYPM2003], the guanine staying in the position equivalent to position 55 of tRNA forms a base-pair with adenine according to the scheme shown in Figure 4.3. This scheme of base-pairing, identified as [A·G *trans* W.C./Sugar edge] [LW2001], makes the juxtaposition of the glycosidic bonds of G and A sufficiently close to that of U(ψ) and G in the standard base-pair G19-U(ψ)55 to allow the replacement of one base-pair by the other with only minimal distortions of the backbone conformation. As one can see in Table 4.1, each of the five T7 clones has several adenine bases in

the D-loop able to form a base-pair with G55 equivalent to G18-U(ψ)55. Henceforth, the particular adenine forming this base-pair is referred to as A18. The purine present in position 57 of all T7 clones can stack on top of this base-pair exactly as in the standard tRNA structure.

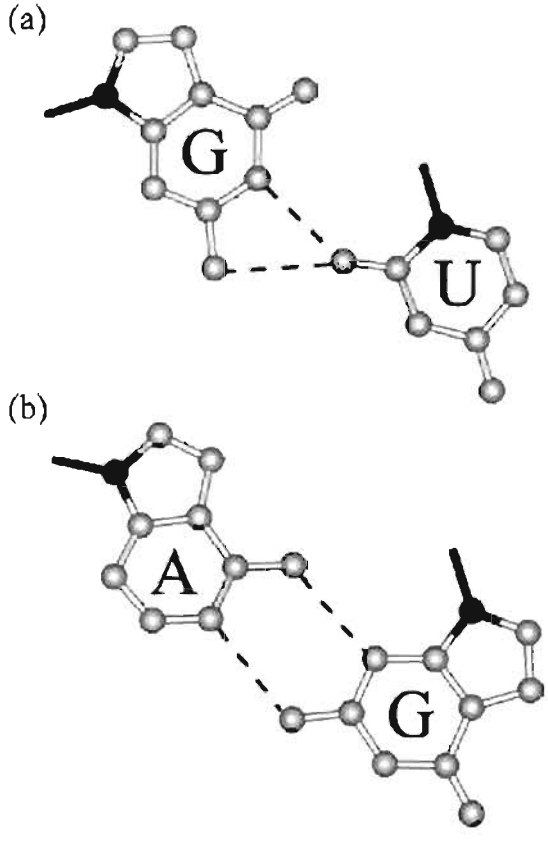

Figure 4.3 The nucleotide arrangements in two inter-loop base-pairs G18-U(ψ)55 (a) and A18-G55 (b). The structure of base-pair G18-U(ψ)55 is taken from the work done by Shi & Moore [SM2000]. In the currently used nomenclature [LW2001] it is named as a bifurcated G·U intermediate of [trans W.C./W.C.] and [*trans* W.C./Sugar edge]. The structure of base-pair A18-G55, identified as [A·G *trans* W.C./Sugar edge] [LW2001], is taken from the work done by Krasilnikov *et al.* [KYPM2003] (PDB entry 1NBS). Although the two base-pairs are not completely isosteric, the juxtapositions of the glycosidic bonds of the two nucleotides (shown in black) in these base-pairs are sufficiently close to each other to guarantee the interchangeability of these base-pairs in the DT region.

The top stacked layer of the DT region in the canonical tRNA structure is occupied by the Watson–Crick base-pair G19-C56 (Figure 4.1). In the T7 clones, the D-loop partner of nucleotide 56 in this base-pair should be a nucleotide following an adenine. Only in two sequences, K15 and K48, there is a possibility to make a Watson–Crick basepair 19–56, respectively UA and CG, while in the three other sequences this base-pair could be AA or CA (in clones K25 and K3) or AG (K29). It is, however, doubtful that the Watson–Crick basepair 19–56 exists even in clones K15 and K48, where, in principle, it can be formed. As one can judge from Table 4.1, the possibility to form this base-pair in clones K15 and K48 is not reflected in a higher suppressor activity among T7 clones. Based on this, we think that as a rule, base-pair 19–56 does not exist in the T7 clones, and even if it forms in some clones, this would not improve their activity. In the absence of this base-pair, the stacking position of purine 56 on top of purine 57 seems to be the most stable conformation (Figure 4.4(a)).

4.2.3 Extension of the quasi normal pattern for T8-tRNAs

Analysis of the T8-tRNA sequences shown in Table 4.1 revealed six of them (Type I T8-tRNA clones in Table 4.1) that could be aligned well with the T7 sequences if one assigns the extra eighth nucleotide in the T8 T-loop to position 3* between nucleotides 55(2*) and 56(4*). Indeed, all these T8 clones contain G2*

equivalent to G55 in the T7-clones and purine 5* equivalent to purine 57. In addition, they have at least one adenine in the D-loop able to form a basepair with G2* equivalent to A18-G55. This similarity allows us to combine all these T7 and T8 sequences into Type I.

As to nucleotide 4*, we can argue that it is equivalent to nucleotide 56 in the canonical tRNA structure. Indeed, among the T8-tRNAs, this nucleotide is cytidine or adenine, respectively, in four and two cases. Coincidently, in all four sequences containing C4*, one can find dinucleotide AG in the D-loop, which would play the role of dinucleotide 18–19 in the canonical tRNA structure and would allow the formation of base-pair G19-C4* equivalent to base-pair G19-C56. As one can judge from Table 4.1, the suppressor activity of all four clones having C4* is higher than that of the two clones having A4*. This demonstrates that in the T8-tRNAs of Type I, unlike in the T7-tRNAs, the existence of base-pair G19-C4* correlates with higher suppressor activity, and therefore, position 4* in T8 clones corresponds to position 56 in the normal tRNAs. The extra nucleotide 3* could stack on top of nucleotide 4* (Figure 4.4(b)). In some T8 clones, this nucleotide could also make a Watson–Crick basepair with the nucleotide following G19, although the necessity of such base-pair is not supported by most of the T8 sequences.

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

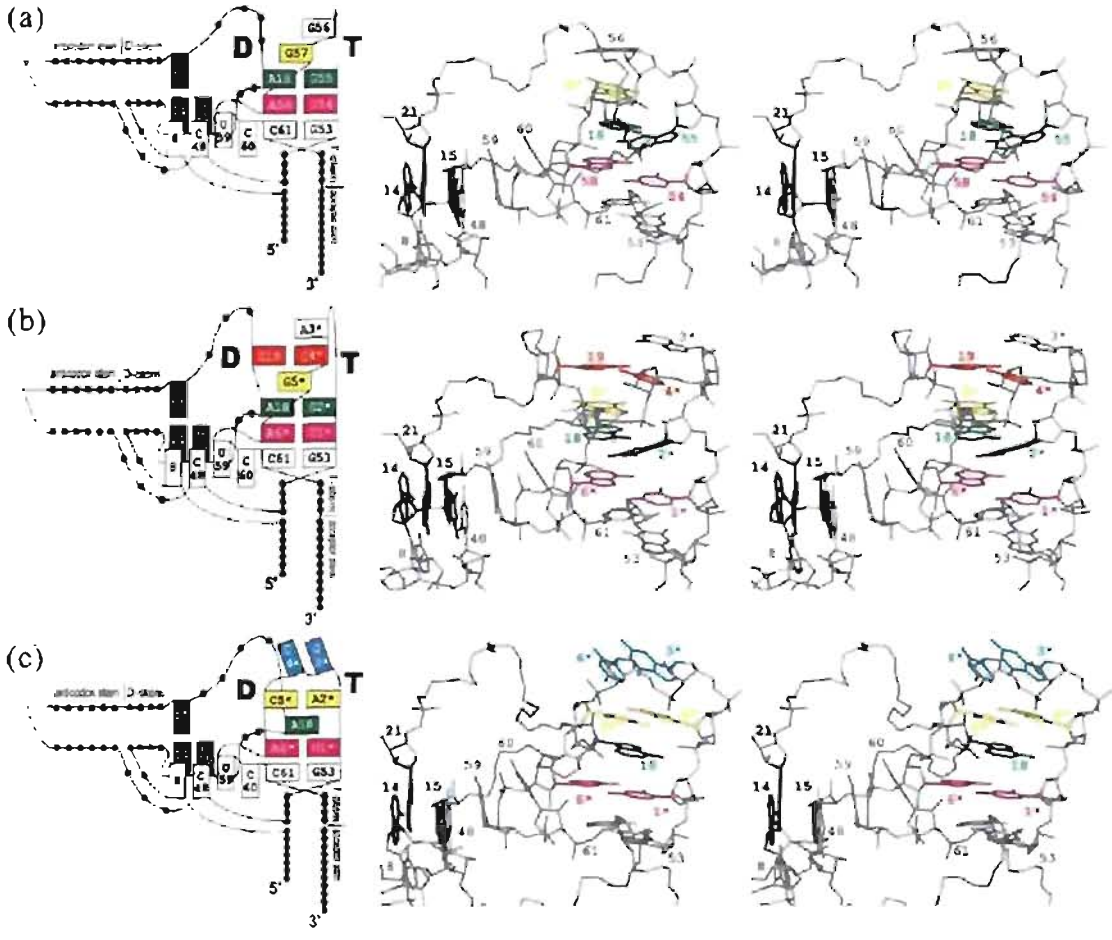

Figure 4.4 The general view of the DT region in the Type I and II tRNAs. On the left: schematic representation of the DT region in the context of the whole tRNA L-shape. On the right: stereo-view of the DT region in the corresponding three-dimensional arrangement. Those nucleotides that are discussed here are represented by rectangles on the left and shown explicitly on the right. For the same nucleotide, the same color is used in both the right and left figures as well as in Figure 4.5 and Table 4.1. **(a)** Type I T7-tRNA. The model was built for the nucleotide sequence of clone K29. U54 forms a reverse-Hoogsteen base-pair with A58. G55 intercalates between base-pair U54-A58 and purine G57 and forms a basepair with purine A18 according to the scheme presented in Figure 4.3(b). G56 stacks on top of G57 without making a basepair with a nucleotide of the D-loop. **(b)** Type I T8-tRNA. The model was built for the nucleotide sequence of clone K2. U1* forms a reverse-Hoogsteen base-pair with A6* analogous to the standard base-pair U54-A58. G2* intercalates between base-pair U1*-A6* and G5* and forms a base-pair with A18. C4* forms a Watson–Crick base-pair with G19 of the D-loop analogous to the standard base-pair G19-C56. A3* stacks on top of base-pair 4*-19 without making a base-pair with a nucleotide of the D-loop. **(c)** Type II tRNA. The model was built for the

nucleotide sequence of clone K24. Two reverse-Hoogsteen base-pairs, U1*-A6* and A2*-C5* are formed at a double distance from each other compared to the distance between two stacked nucleotides (rise of 6.2 Å compared to 3.1 Å). Purine A18 of the D-loop intercalates between these two base-pairs. G3* and C4* close the loop and are not involved in any inter-loop interactions.

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

According to the analysis of the Type I sequences presented in the previous section, different Type I clones can form either one or two base-pairs between the D and T-loops corresponding to base-pairs 18–55 and 19–56 of the standard tRNA structure. However, only the first base-pair has been found in all clones, while the second one exists only in less than half of the Type I clones. This indicates a relative importance of base-pair 18–55 for the tRNA structure and function. Intuitively, such difference between the two base-pairs seems obvious, because base-pair 18–55 makes a part of the core of the T-loop, and therefore, its disruption will affect the whole structure of the DT region. Base-pair 19–56, on the other hand, is positioned on the periphery of the T-loop, so that its disruption would have only a mild, local effect. Only for the T8 clones the presence of base-pair 19–56(4*) may be essential, because here, the stacking of nucleotide 3*, predominantly adenine, on top of this base-pair makes it a part of the internal structure of the DT region. The higher relative importance of base-pair 18–55 compared to base-pair 19–56 fits also to the fact that in the crystal structure of the yeast

$tRNA^{Asp}$[MCFWTEG1980] the disruption of base-pair 19–56 does not affect the structure of the rest of the DT region. The formation of base-pair 18–55(2*) can be seen from another perspective. One can say that in the T-loop, there is a special trap made of base-pair U54(1*)-A58(6*) and purine 57(5*) arranged at a double distance from each other compared to the distance between two stacked bases. Such distance corresponds to a rise [LS1988] of about 6.2 Å between purine 57(5*) and A58(6*) of the T-loop compared to 3.1 Å between two stacked nucleotides. The T-loop trap allows purine 18 from the D-loop to intercalate and to form hydrogen bonds with nucleotide 55(2*) at the bottom of the trap. The presence of the latter nucleotide makes the trap specific: in the standard tRNA structure, nucleotide 55 is pseudouridine, and the trap "catches" guanine. In the Type I clones, nucleotide 55(2*) is guanine, and correspondingly, the trap "catches" adenine. Because this specific purine trap exists in all Type I clones, it seems to be indispensable for the function of these tRNAs.

4.2.5 Purine trap and reverse-Hoogsteen base-pair U54-A58

If the locking of purine 18 into the T-loop purine trap is so critical for the tRNA structure and function, it would be important to know how the trap is formed in the first place. In particular, we would like to know whether the extended distance between purines 57(5*) and 58(6*), which allows purine 18 to get in-

between, is arranged independently of or only upon the intercalation of the latter. The existing data support the idea that the gap between nucleotides 57 and 58 forms independently of the intercalation of nucleotide 18.

Thus, in the recently published crystal structure of the modifying enzyme archaeosine tRNAguanine transglycosylase bound to $tRNA^{Val}$ [INNONY2003] the D-domain of the tRNA was found to be strongly deformed. In particular, no nucleotides of the D-loop were involved in interaction with the T-loop. The only element of the T-loop that was in contact with the rearranged D-domain was nucleotide 59 of the bulge region. This nucleotide stacked on base-pair 23–48, thus forming the contact analogous to that between nucleotide 59 and basepair 15–48 in the standard tRNA structure. Despite this, the conformation of the T-loop was characterized by the double distance between nucleotides 57 and 58 with rise of 6.6 Å, which is close to that in the standard tRNA structure, even without any intercalating nucleotide at the place normally occupied by G18.

Our analysis showed that the double distance between nucleotides 57 and 58 originates from the particular conformation of the sugar–phosphate backbone of nucleotides 58 and 59 when the first of them forms the reverse-Hoogsteen base-pair with U54, while the second one is involved in the T-loop bulge. In our *in silico* modeling system, we took the structure of the T-stem-and-loop from the yeast

tRNAPhe and deleted nucleotides 55 and 56, leaving the other nucleotides untouched. This made the position of nucleotides 54, 57 and 58 independent of the conformation at the top of the T-loop. Then we tried to change the position of purine 57 to make it stacked on base-pair 54–58. We found, however, that as long as nucleotides 54 and 58 were involved in the standard reverse-Hoogsteen base-pair, while nucleotides 59 and 60 formed the T-loop bulge, the stacking of nucleotide 57 on base-pair 54–58 was not sterically possible. Such stacking could only be achieved at the expense of the integrity of region 58–60, for example, when nucleotides 59–60 were allowed to move from their standard positions. In other words, the standard conformation of region 58–60 does not allow nucleotide 57 to stack on basepair 54–58. Although the position of nucleotide 57 at the double distance from this base-pair, characteristic of the canonical tRNA structure, is not the only one possible, it has an advantage of allowing a nucleotide from the D-loop to intercalate into the created gap, thus optimizing the stacking interactions. This aspect of the T-loop structure is important for the Type II structures, which are discussed in the next section.

4.2.6 Type II as a non-specific purine trap

Among the clones presented in Table 4.1 that do not fit to Type I, all but one clone K30 have A in position 2*. We combine these A2*-containing clones into

Type II. Two aspects of the Type II clones do not allow their accommodation to the standard pattern. First, the standard pattern implies the formation of a base-pair between nucleotide 55(2*) and 18 similar to base-pairs G18-U(ψ)55 or A18-G55 seen in the Type I clones. For Type II tRNAs, however, our analysis failed to find a reasonable scheme of base-pairing for A2* with another nucleotide that would satisfy this requirement. Second, in the standard pattern, nucleotide 57(5*) should be a purine. As one can see in Table 4.1, none of the Type II clones contains a purine in position 5*; instead, in all these sequences, position 5* is exclusively occupied by C. The presence of a pyrimidine in position 5* would destabilize the stacking between nucleotides 5* and 18, thus jeopardizing the integrity of the structure of the whole region. On the other hand, neither A2* nor C5* has ever been seen among the standard or Type I tRNAs, and the fact that they systematically coincide in the Type II clones suggests the existence of an alternative structural pattern in which both these nucleotides play crucial roles.

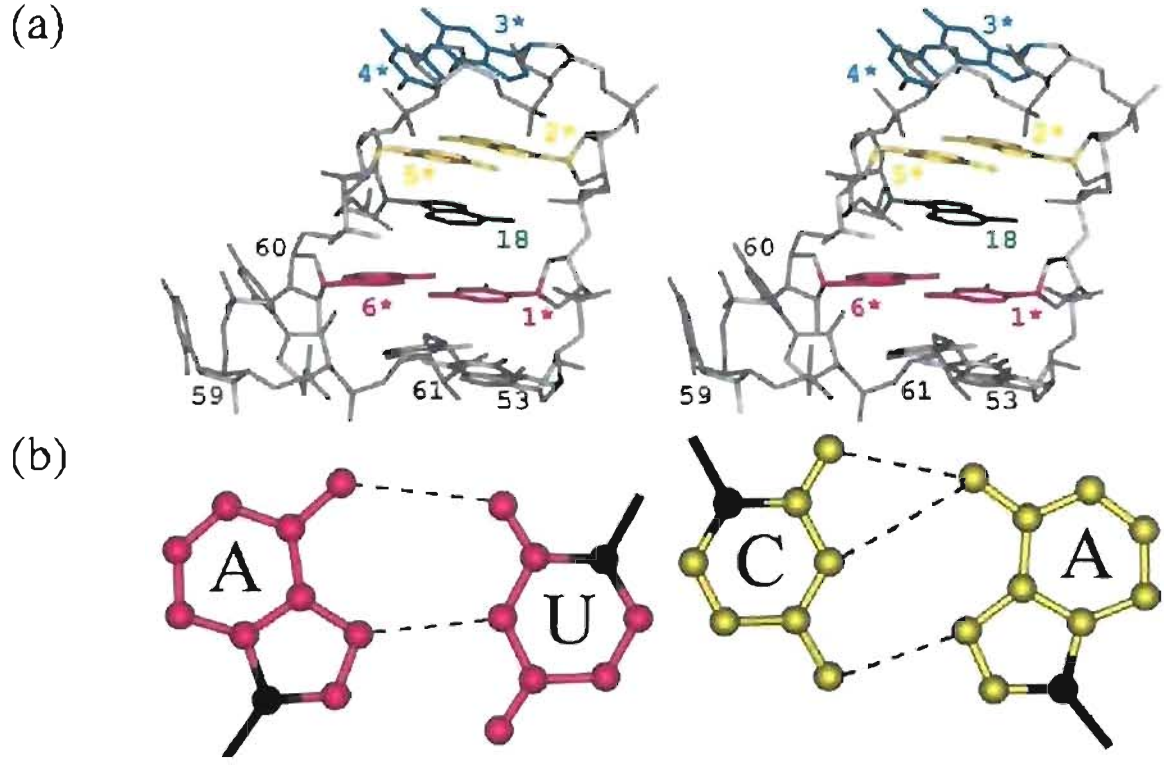

Figure 4.5 The detailed structure of the non-specific purine trap. (a) Stereoview of the top of the T-loop in Type II tRNAs. Individual nucleotides are colored as in Table 4.1 and Figure 4.4. **(b)** The nucleotide arrangement in the two reverse-Hoogsteen base-pairs, U1*-A6* [U-A *trans* W.C./Hoogsteen] [LW2001], and A2*-C5* [C·A *trans* W.C./Hoogsteen] [LW2001], forming the non-specific purine trap. The two base-pairs have a similar although different juxtaposition of glycosidic bonds of the two constituting nucleotides. The base-pairs in (b) are oriented in about the same way as in (a).

The suggested model of the structure of the DT region in the Type II tRNAs that addresses the above-mentioned problems is shown in Figures 4.4(c) and 4.5. In this model, nucleotide 5* stays at its standard place at the double distance from the reverse-Hoogsteen base-pair U1*-A6*, while nucleotide 18 still intercalates in-between. However, unlike in the Type I tRNAs, nucleotide 2* neither stacks to base-pair U1*-A6*, nor does it form a basepair with nucleotide 18. Instead, it is

displaced farther from base-pair U1*-A6* to the level of nucleotide 5* with which it forms a base-pair. As a result, nucleotide 18 becomes sandwiched between two base-pairs, U1*-A6* and A2*-C5*. The displacement of A2* resolves three potential problems of the Type II clones. First, it removes A2* from the standard position, where it is unable to find a base-pair partner from the D-loop, and places it at the position where it forms a base-pair. Second, the newly formed base-pair A2*-C5* stabilizes the position of C5* at the double distance from basepair U1*-A6*. Finally, base-pair A2*-C5* extends the area of contact for the intercalating nucleotide 18 compared to its stacking with C5* alone, thus compensating for the fact that nucleotide 5* is a pyrimidine. The scheme of base-pairing [C·A *trans* W.C./Hoogsteen] [LW2001] seen in Figure 4.5 seems to be a good choice fitting well to the structure of the DT region. Moreover, according to our further analysis, it represents the only possibility of a stable arrangement of nucleotides 2* and 5* in the given structural context, thus justifying the conservation of the identities of these two nucleotides among Type II tRNAs. For convenience and due to the similarity with the U1*-A6* base-pair, we call A2*-C5* base-pair reverse-Hoogsteen as well.

The intercalation of nucleotide 18 between the two reverse-Hoogsteen base-pairs provides an important stabilizing effect on the structure. To ensure that this interaction is sufficiently strong to keep the integrity of the arrangement, the

intercalating nucleotide must be a purine. On the other hand, it does not seem to matter whether the intervening nucleotide is A or G, because both purines fit equally well into the gap between the two reverse-Hoogsteen base-pairs. As one can see in Table 4.1, in each Type II clone, there are several purines in the D-loop, and there is no reason to prefer any one of them to another as a candidate for the intercalation. An additional stabilization of the structure would come from the interaction of the amino group of the intercalating purine with the nearest phosphate group between the two reverse- Hoogsteen base-pairs.

A minimum of two unpaired nucleotides, 3* and 4*, are needed to close the T-loop, which fits well to the fact that among Type II clones, unlike among Type I clones, no T7-tRNAs were found. None of nucleotides 3* and 4* is conserved among the Type II clones. We noticed, however, that nucleotide 3* is almost exclusively a purine, while nucleotide 4*, with only one exception, is either C or A. This observation fits well to the model of the DT region in Type II tRNAs shown in Figure 4.5. A purine in position 3* allows an extensive stacking interaction with nucleotide 4* as well as a strong contact with A2*. Depending on whether nucleotide 3* is G or A, the hydrogen bond can be formed between its extracyclic atom O6 or N6 and either N1 or N6 of A2*. The presence of either C or A in position 4* allows the formation of the hydrogen bond between the amino group of 4* and atom O2 of C5*. Another hydrogen bond is possible between the amino

group of C5* and phosphate 4*. All these interactions make the top-most region of the T-loop sufficiently stable on its own, without any additional interactions with nucleotides of the D-loop. The absence of such inter-loop interactions is also supported by the fact that we have not found any significant nucleotide co-variation between nucleotides of the two loops.

The structural block consisting of purine 18 sandwiched between the two T-loop reverse-Hoogsteen base-pairs is shared by all Type II clones. By analogy with the Type I clones, we can look at the gap between the two reverse-Hoogsteen basepairs positioned at the double distance from each other as a trap for a purine from the D-loop. However, in the Type II clones, unlike in the Type I clones, this trap is non-specific, mostly due to the absence of a nucleotide analogous to nucleotide 55(2*) in Type I tRNAs, which would form hydrogen bonds with the intercalating nucleotide and restrict its identity. This allows us to name the structural block characteristic of the Type II clones as the non-specific purine trap, which thus represents another structural solution of the D and T-loop arrangement alternative to the specific purine trap.

4.2.7 Stability of the non-specific purine trap and its dependence on the identity of certain nucleotides

To evaluate the stability of the proposed structure of the DT region in the Type II tRNAs and its dependence on the identity of the participating nucleotides, we performed energy minimization of the model of this structure and of some of its variants in the context of the whole tRNA. Thus, the energy minimization of the proposed arrangement in clone K24 optimized some atom–atom contacts and hydrogen bonds, but did not move substantially any nucleotide, which indicated a relative stability of this arrangement. The replacement of A18 by G did not result in any essential movement of nucleotides, which can be seen as an indication that indeed, both purines fit the gap between the two reverse-Hoogsteen base-pairs reasonably well. However, when this purine was replaced by either C or U, the same energy minimization test either completely destroyed the arrangement (C) or strongly deformed its stacking and H-bond pattern (U), demonstrating that such arrangement is not stable and that a pyrimidine in position 18 is not acceptable. Other tests showed that the stability of the arrangement is very sensitive to the structure of base-pair 2*-5*. Thus, when we replaced base-pair A2*-C5* by a relatively close to it reverse-Hoogsteen base-pair A2*-U5*, by the inverted reverse-Hoogsteen base-pair C2*-A5*, by Watson–Crick base-pair C2*-G5* or by sheared base-pair A2*-G5*, in all cases, the same energy minimization test broke the

structure. In other words, the particular geometry of the reverse-Hoogsteen base-pair A2*- C5* is essential for the stability of the arrangement, which can explain the absolute conservation of these nucleotides among the Type II clones.

Finally, among the Type II clones, like in Type I, there is a clone, K41, characterized by a deletion of five nucleotides in the D-loop. As in the Type I clone K34, this deletion in K41 makes the D-loop so short that it is able to provide only one nucleotide, a purine, for the intercalation between the two reverse-Hoogsteen base-pairs. However, unlike in the case of K34, the energy minimization of the model of clone K41 did not reveal any steric problem for structure of the DT region, indicating that even such abbreviated D-loop seemed to be sufficient for keeping the integrity of the tRNA and for making it functional. The difference of K41 from K34 comes from the fact that in the Type II structure, purine 18 does not form specific H-bonds. This provides some flexibility to its position, which allows a better accommodation to a shorter D-loop and can explain the higher suppressor activity of K41 compared to that of K34.

4.3 DISCUSSION

This work represents a new step in our study of the general structural constraints imposed on functional tRNAs. Previously, we showed that the standard

juxtaposition of the two tRNA helical domains required the presence of exactly two unpaired nucleotides 59–60 in the T-loop bulge, which was achieved by the formation of the reverse-Hoogsteen base-pair between positions 54 and 58 [ZDS2003, ZKS2004]. However, the fact that the top of the T-loop did not seem to have a common motif in all selected clones meant that a particular conformation of this region was not critical for the function of the suppressor tRNAs. On the other hand, due to an insufficient number of the sequences, it was not possible to say whether this region played any important role in the tRNA structure and function and whether any sequence constraints were imposed on it. After the additional selection of more than 20 new suppressor tRNA clones and the analysis of their nucleotide sequences, we now can conclude that the top of the T-loop is involved in specific interactions with the D-loop. The analysis shows that the constraints imposed on the nucleotide sequences of the two loops in functional tRNAs are strict, so that for those tRNAs having the reverse-Hoogsteen basepair U-A equivalent to base-pair 54–58 in the canonical tRNA structure, only two distinct families of sequences are allowed. Each family corresponds to a particular type of the T-loop conformation. The first type, the so-called specific purine trap, is a generalized version of the standard T-loop conformation, while the second, the non-specific purine trap, is suggested here for the first time.

The specific purine trap is based on two key elements, base-pair 18–55(2*) and purine 57(5*). In the tertiary structure, it consists of a gap in the T-loop between the reverse-Hoogsteen base-pair 54(1*)–58(6*) and nucleotide 57(5*), which is filled by purine 18 from the D-loop. The requirement for a purine in position 57(5*) comes from the fact that the stacking interactions play the most important role in the formation of this arrangement. The Type I purine trap is characterized by the specificity with respect to the identity of purine 18, which is restricted via formation of base-pair 18–55(2*). Only two combinations are allowed for this basepair, GU and AG, which are interchangeable due to a similar juxtaposition of the glycosidic bonds. Interestingly, in the selected Type I tRNAs we found only the AG base-pair, and no GU. We attribute this to the fact that in the randomized positions of the commercially purchased K-library, the four standard nucleotides were not present in equal proportions. In fact, when we determined the nucleotide sequences of 40 randomly chosen negative clones (data not shown), which characterize better the distribution of the identities of each nucleotide due to the fact that negative clones constitute about 99.9% of the library, we found that at each randomized position, adenine appeared more often than the other three nucleotides. This disproportion would increase the chance to be selected for those combinations that contain more adenine bases, like base-pair AG versus GU in our case. Compared to the K-library, our next library [ZKS2004] was randomized more evenly, and correspondingly, the selected tRNAs predominantly contained the

G18-U55 base-pair. The other inter-loop base-pair 19–56(4*), which exists in the canonical tRNA structure, is not present in all selected clones. This base-pair is not directly involved in the formation of the purine trap and is positioned on the periphery of the structure. This allows the disruption of basepair 19–56 without jeopardizing the integrity of the whole DT region. The stability of this base-pair becomes essential for the tRNA function only when the additional eighth nucleotide 3* in the T-loop stacks on top of this base-pair. From these data, however, it is not clear why base-pair 19–56 exists in the normal tRNAs. One of the hypotheses is that while this base-pair does not affect the overall activity of the tRNA, it may ensure more accurate codon–anticodon recognition due to the additional stabilization of the structure of the DT-region and, indirectly, of the whole tRNA L-shape. Although some existing data support this point of view [HAG1999], additional studies are needed to show whether it is really true.

The structural type of the DT region that is alternative to the specific purine trap is the nonspecific purine trap. It represents a new structural motif never seen before in tRNAs or other RNA structures. The key sequence requirements that allow the formation of this motif consist of the presence of A2*, C5* and of at least two nucleotides between them. Nucleotides A2* and C5* form the specific reverse-Hoogsteen base-pair positioned at the double distance from base-pair U1*-A6*, thus allowing purine 18 from the D-loop to intercalate between them. The

conservation of A2* and C5* among the Type II tRNAs correlates with the fact that in the proposed structure these nucleotides play a crucial role. Indeed, all our attempts to model in the given structural context of a stable alternative to the proposed reverse-Hoogsteen base-pair A2*- C5* failed. The additional eighth nucleotide in the T-loop is critical for the integrity of the motif: without it, the T-loop cannot be closed. Correspondingly, all Type II tRNAs belong to the T8 class, which contrasts with the Type I tRNAs, where a seven nucleotide T-loop is acceptable and where almost a half of the clones are T7-tRNAs.

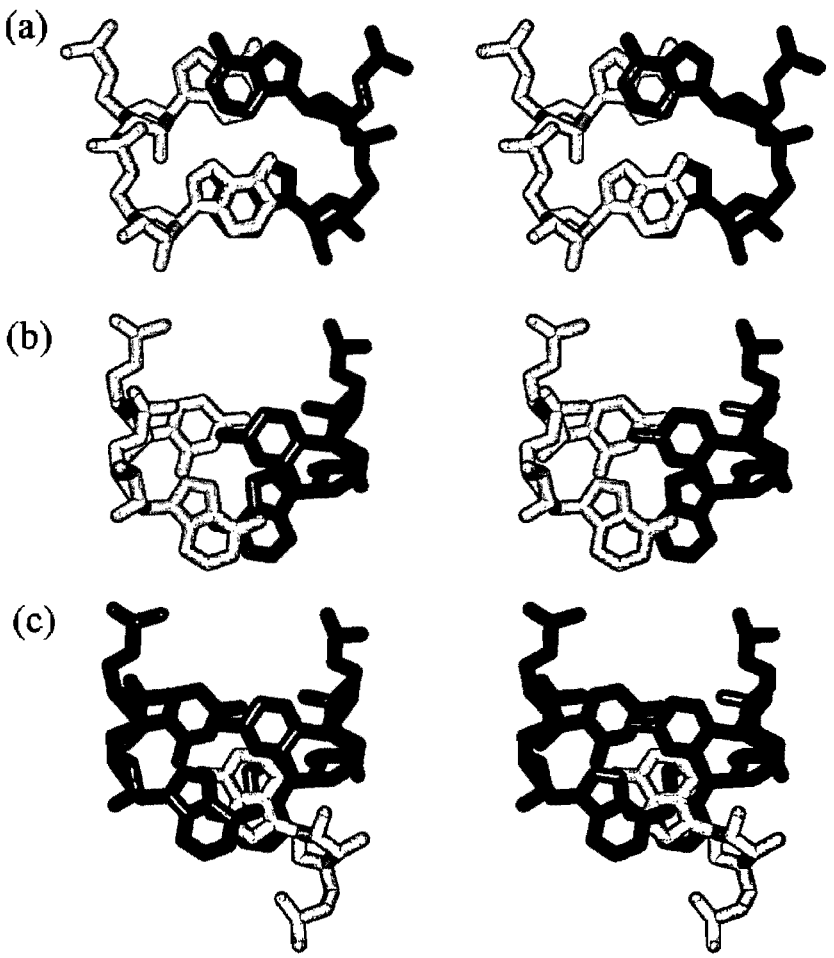

Figure 4.6 Comparison of the model of the non-specific purine trap and the AA-AA tandem motif. (a) Stereoview of the tandem AA-AA from the crystal structure of group I intron RNA [CGPZGKCD1996] (PDB entry 1GID, nucleotides A113-A114 (grey) and A206-A207 (black)). The two sheared base-pairs, A113-A207 and A114-A206 are positioned, respectively, farther and closer to the reader. In each fragment, 113-114 and 206-207, the first and the second nucleotide are shifted, respectively, in the direction of the major and minor groove. **(b)** Stereoview of dinucleotides U1*-A2* (grey) and C5*-A6* (black) forming the non-specific purine trap modeled for Type II tRNAs. The two reverse-Hoogsteen base-pairs, U1*-A6* and A2*-C5* are positioned, respectively, farther and closer to the reader. Like in structure (a), in each fragment, 1*-2* and 5*-6*, the first and the second nucleotide are shifted, respectively, in the direction of the major and minor groove. **(c)** The same nonspecific purine trap as in (b) (black) filled with purine 18 (grey).

The non-specific purine trap does not exist in natural tRNAs. Neither has it been found in other known RNA structures: our search of the database of three-dimensional structures of nucleic acids [BOBWGDHSS1992] for structures consisted of a nucleotide sandwiched between two reverse-Hoogsteen base-pairs UA and AC has not provided any hit. On the other hand, we found a number of elements that resemble some important aspects of the non-specific purine trap. As one can see in Figure 4.6(b), in both fragments U1*-A2* and C5*-A6* forming the non-specific purine trap, the first nucleotide is shifted in the direction of the major groove, while the following adenine is shifted in the direction of the minor groove. Such a situation is not uncommon in non- Watson–Crick RNA duplexes and can also be seen in the quite frequent arrangement of four purines, GA versus GA or AA versus AA, which form tandem purine–purine base-pairs within an antiparallel duplex [HH2003] (Figure 4.6(a)). Here as well, the first nucleotide in each chain is shifted in the direction of the major groove, while the following adenine is shifted to the minor groove. In this motif, as in the non-specific purine trap, specific base–base hydrogen bonds are able to stabilize the structure.

Another motif having a similarity to the nonspecific purine trap is I-DNA, built of four poly(C) chains [GLG1993] (Figure 4.7). In this structure, two consecutive nucleotides from each of the two parallel chains are positioned at the distance corresponding to rise of 6.2Å and form two specific CCC base-pairs. Between

these two base-pairs, nucleotides from the two other chains intercalate and form another CCC base-pair. The juxtaposition of the glycosidic bonds in the reverse-Hoogsteen base-pairs UA and AC is remarkably close to that in the CCC base-pairs 64 Purine Trap in the T-loop of tRNA composing the I-DNA(Figure 4.7(c)).Also, like I-DNA, the non-specific purine trap is very sensitive to the geometry of the base-pairs, so that a replacement of base-pair A2*-C5* by a similar although different base-pair AU is very damaging for the structure. The similarity of the non-specific purine trap to these two well-known nucleic acid motifs serves as an additional argument in support of the proposed structure.

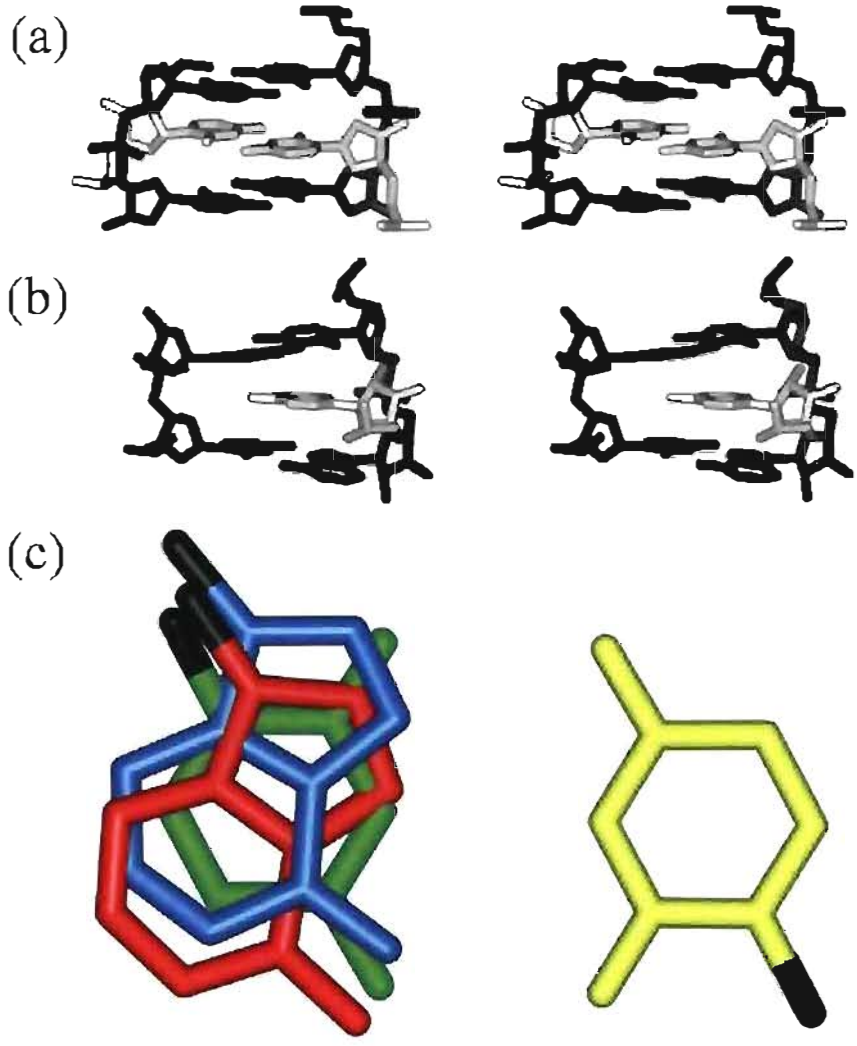

Figure 4.7 Comparison of the model of the non-specific purine trap and I-DNA. (a) Stereoview of a fragment of I-DNA [GLG1993] (PDB entry 225D). Two black parallel strands form two CCC base-pairs, which are positioned at the double distance from each other (rise 6.2 Å). Between these two base-pairs, intercalate two grey cytosine bases from the other two strands and form another CCC base-pair. A replacement of any cytosine in this structure by another nucleotide destroys the arrangement. **(b)** Stereoview of the non-specific purine trap modeled for Type II tRNAs. Purine 18 of the D-loop (grey) intercalates between the two reverse-Hoogsteen base-pairs (black) U1*-A6* (bottom) and A1*-C5* (top). A replacement of purine 18 by a pyrimidine or replacement of the A1*-C5* base-pair by another one destroys the arrangement. **(c)** Comparison of the structure of the reverse-Hoogsteen base-pairs AU and AC and of base-pair CCC. The pyrimidine in all three base-pairs is yellow. The other nucleotide in base-pair red (AU), blue (AC), or green (CCC). The glycosidic bonds in all nucleotides are black.

Although the differences between the structures of the two purine traps are obvious, they are not dramatic and can, generally, be reduced to the reorientation of nucleotide 2* from base-pairing with 18 to base-pairing with 5* and to the subsequent adjustment of the identities of nucleotides 2*, 5* and 18. Many other important elements are shared by both types. Thus, in both types, the structure of the lower part of the T-loop, including bulge 59–60, its interaction with the D-domain, and the reverse-Hoogsteen base-pair 54(1*)–58(6*) is essentially the same. Also, as we showed, the formation of the reverse-Hoogsteen base-pair 54(1*)–58(6*) forces nucleotide 57(5*) to be positioned at the double distance from A58(6*), which opens the possibility for nucleotide 18 from the D-loop to intercalate between them. This intercalation, specific in the Type I tRNAs and non-specific in the Type II tRNAs, is shared by all clones of both groups and thus seems to be indispensable for the tRNA function.

The role of this intercalation for the whole tRNA is of a special interest. Modeling the tertiary structures of the selected tRNAs and their comparison with the standard tRNA structure allows us to suggest that this intercalation is crucially important for the proper juxtaposition of the two helical domains. From our previous experiments [ZDS2003, ZKS2004] we know that the key element in the fixation of the L-shape is the T-loop dinucleotide bulge 59–60. This bulge fills the space between the D-domain and the T-stem and thus mediates the interaction between

them (Figure 4.1). When in some of our mutant tRNAs the D-domain became shorter, correspondingly, the T-loop bulge extended in the way that the total number of the stacked layers in the DT region remained unchanged. This compensation provides the condition for the standard arrangement of the two helical domains. However, because the interaction between the T-loop bulge and the D-domain is based solely on the stacking contact of base 59 and base-pair 15–48, it is able to keep the standard juxtaposition of the helical domains only if the domains are already arranged relatively closely to the standard L-shape. The locking of nucleotide 18 into the purine trap would serve exactly this purpose of giving the two loops a chance to associate with each other in the orientation close to the standard.

It is important, however, that although the interaction of nucleotide 18 with the purine trap brings the two loops together and orients them similar to the standard, it is unable to fix the juxtaposition of the two helical domains. This inability is mostly due to the fact that neither in the canonical tRNA structure, nor in the selected clones the position of the intervening nucleotide 18 is fixed with respect to the D-domain. In the canonical tRNA structure, the G18-G19 dinucleotide is connected to the rest of the D-domain via two regions, 16–17 and 20–20b, which are variable in length and are conformationally flexible. In most selected clones, on the other hand, there are several purines in the D-loop that can

play the role of nucleotide 18, and there is no reason to prefer any one to another. Therefore, the fixation of the tRNA L-shape can be achieved only via cooperation between the interaction of purine 18 with the purine trap and the stacking of the T-loop bulge on the D-domain. In this cooperation, the first element arranges the two domains relatively closely to the standard L-shape, while the second one provides the final fixation.

The interaction of nucleotide 18 with the purine trap can affect the tRNA L-shape in one more way. The proper number of nucleotides in the T-loop bulge, which is critical for the standard juxtaposition of the helical domains, is guaranteed via the formation of the reverse-Hoogsteen base-pair 54–58. On the other hand, this base-pair is shown here to force nucleotide 57 to take the position at the double distance from base-pair 54–58, compared to that between two stacked nucleotides, thus creating a gap between the bases. The filling of this gap by nucleotide 18 fixes the double distance between nucleotides 57 and 58, thus providing the additional stability for base-pair 54–58. This, in turn, stabilizes the structure of bulge 59–60 and indirectly, the whole tRNA L-shape.

Although both structural types demonstrate the ability to provide for a functional tRNA, only Type I is found in the naturally selected tRNA species. The obvious advantage of the Type I structure deals with the fact that it restricts both

the identity of the intercalating purine and its orientation in the complex with the T-loop, which would help to avoid misfoldings and kinetic traps during the formation of the tRNA tertiary structure.

We have analyzed here only two types of the structure of the DT region in the suppressor tRNAs selected from the K-library. There are serious reasons to think that these two types do not exhaust all possibilities of the structural organization of the DT region. Indeed, at least one clone in Table 4.1, K30, has combination U54-A58, but does not fit to any of the two major structural types. In addition, as mentioned in the Introduction, 21 K-clones out of 50 selected from the K-library did not have combination U54-A58, but were still functional. Analysis of the nucleotide sequences of these clones could reveal new types of the structure of the DT region. Although these types are as yet unknown, we can anticipate that they will also, like Types I and II, provide mechanisms for the correct mutual fixation of the two helical domains. The elucidation of these mechanisms will constitute an important step in understanding of the particular role played by irregular unpaired regions in RNA structure and function.

4.4 METHODS

4.4.1 Cloning, selection and measuring the suppressor tRNA activity

The activity of selected tRNAs was evaluated by their ability to suppress the nonsense mutation in the *lacZ* gene of XAC-1 cells. The details of the cloning and screening of the tRNA library as well as of the measuring the suppressor activity of the selected tRNAs were described earlier [ZDS2003, ZKS2004]. Only clones that provided for blue colonies after overnight growth on X-Gal plates were taken for further consideration.

4.4.2 Computer modeling

Preliminary modeling was done interactively, using the InsightII/Discover package (Version 2000, Accelrys Inc., San Diego, CA). The X-ray structure of yeast $tRNA^{Phe}$ [SM2000] was used as a starting conformation, to which the elements different from the standard tRNA structure were appended. Each model was submitted to energy minimization using the AMBER forcefield [P1995]. The

elements of the modeled structures identical with the corresponding elements of the yeast $tRNA^{Phe}$ were fixed during the minimization. Only those structures for which the energy minimization showed a clear convergence toward an energy minimum were taken for further consideration. Visualizations were done on a Silicon Graphics O2 computer.

4.5 ACKNOWLEDGEMENTS

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

CHAPITRE 5 : Apport scientifique de l'article

Le principal apport scientifique de cet article est l'identification du motif du piège à purine (« purine trap ») et de son rôle dans le repliement de la molécule d'ARNt. Ce motif est présenté dans le contexte de la région DT et les multiples contraintes spécifiques de séquence et de structure y sont également discutées.

Les résultats d'études précédentes ont montré que l'élément le plus conservé de la région DT est sans équivoque la paire de base reverse-Hoogsteen U-A (ou son équivalent G-A ou G-G) [ZDS2003]. Le reste de la région semblait, à l'origine, indistinct; ce qui semblait vouloir dire qu'elle était plus accessoire au maintient de la fonction de la molécule. Par contre, cet aspect ne semblait pas coller avec ce que l'on connaissait à propos de la structure et de la fonction des ARNts.

En effet, c'est par les interactions entre les boucles D et T que l'ARNt adopte sa structure tertiaire que l'on croit si critique à sa fonction. En suivant cette logique, la structure de la région DT devrait donc promouvoir l'association entre les deux boucles. C'est en générant des clones additionnels que l'on a pu éclaircir ce mystère et pousser notre analyse à un niveau supérieur.

Nos travaux ont permis un examen plus approfondie des nouvelles séquences, nous permettant d'identifier des éléments de conservation à même la structure de la région DT. Ces éléments diffèrent grandement selon l'identité de la paire de baise reverse-Hoogsteen.

Nous avons décrit les différentes façons dont la région DT peut s'agencer. Il est intéressant d'observer la forte propension des séquences comprenant 7 nucléotides dans la boucle T (T7) à former une structure presqu'identique à la structure standard du type sauvage. Il fut aussi intéressant de constater une plus grande variabilité chez les nucléotides formant le haut de la boucle T, leur permettant d'accommoder une plus grande diversité dans la structure des paires de bases dans lesquelles ils sont impliqués. D'ailleurs, cette plus grande variabilité s'apparente au phénomène des ARNt mitochondriaux, qui arborent une structure parfois très inusitée.

Les séquences ayant 8 nucléotides dans la boucle T (T8) peuvent aussi adopter la structure canonique. Ils possèdent effectivement tous les éléments nécessaires, de la guanine 55 à la purine 57. De plus, des deux nucléotides complétant la boucle, le deuxième (celui précédant la purine 57) possède une forte affinité Watson-Crick pour le nucléotide qui joue le rôle du nucléotide 19 de la boucle dans la structure standard. On remarque aussi une correspondance entre

les nucléotides 55 et 18 pour les séquences se conformant à la structure standard. Deux possibilités peuvent être observées : U55-G18 ou G55-A18.

La comparaison et l'analyse de la structure canonique ainsi que de la structures des séquences T7 et T8 consistant en une variante de la structure canonique nous a donc permis de conclure que différents éléments structurels doivent être présents pour que la structure de la molécule d'ARNt conserve son intégrité. Ces éléments sont : la présence de la paire de base reverse-Hoogsteen UA; une purine en position 57 intercalée entre les paires de base 18-55 et 19-56; et une correspondance UG ou GA entre les nucléotides 18 et 55 (ou leurs équivalents). C'est l'agencement entre ces éléments formés des nucléotides 54, 55, 57, 58 qu'on a baptisé « piège à purine spécifique », puisque l'identité du nucléotide 18 est contrôlé par l'identité du nucléotide 55.

En poussant l'analyse un peu plus loin et en impliquant les autres séquences T8, on découvre que le remplacement de 2 nucléotides dans la boucle T peut rendre le piège à purine non-spécifique et accommoder n'importe quelle purine de la boucle D. Il est clair d'emblée que les séquences restantes ne peuvent pas se conformer aux exigences prescrites par la structure de type sauvage. Elles ne comprennent ni U ni G en position 55 et, qui plus est, n'arborent pas de purine en position 57. Bien que l'absence de U ou G en position 55 puisse signifier une différente correspondance pour la spécificité du piège à purine, la présence d'une

purine en position 57 est primordiale. De plus, on peut remarquer qu'un nouvel élément de conservation émerge de cette analyse : toutes les séquences restantes possèdent une adénine en position 55 et une cytosine en position 57.

Nous avons effectué *in silico* la modélisation de ce nouvel élément de structure et conclut que les nucléotides A55 et C57 s'appariaient pour former une deuxième paire de base reverse-Hoogsteen. Nous avons aussi démontré que cette paire de base se positionne à une double distance de la paire de base reverse-Hoogsteen U54-A58, de façon à ce qu'une purine de la boucle D puisse venir s'intercaler entre les deux. C'est de cette manière que fut identifié le piège à purine non-spécifique. L'identification de ce nouveau motif structural, qui possède certaines caractéristiques observées dans d'autres motifs déjà connu (tel GA-AG et le I-DNA), fut un accomplissement.

CHAPITRE 6 : Conclusion

Nous avons entreprit ici d'élucider les règles et contraintes gouvernant la structure de l'ARN. Plus spécifiquement, nous nous sommes servi de la région conservée formée des interactions entre les D et T de l'ARNt en tant que modèle.

Nous avons démontré la paire de base reverse-Hoogsteen est l'élément central de la région DT. Nous avons aussi avancé qu'elle joue un rôle au sein de la fixation des deux domaines hélicoïdaux en allouant le bon nombre de nucléotides pour l'appendice. Elle est ainsi d'une grande aide pour conserver l'arrangement perpendiculaire nécessaire à la fonction de la molécule.

Nous avons aussi démontré qu'il existait un élément structural qui était toujours présent en même temps que la paire de base « reverse-Hoogsteen ». Cet élément a été nommé «piège à purine». Celui-ci consiste en l'intercalation de la boucle T d'une purine entre les nucléotides 57 et 58. Cette purine proviendrait de la boucle D. De plus, l'identité de la purine peut être contrôlée par le nucléotide 55 dans le cas où celui-ci ne forme pas une paire de base avec le nucléotide 57. Le piège à purine peut donc être spécifique ou non-spécifique.

La réflexion sur ces éléments de structure qui étaient à la base connus de tous depuis la cristallisation de la molécule, est applicable à d'autres structures d'ARN plus complexes. Ellepermettra aussi sans aucun doute de mieux

comprendre la manière fondamentale dont les ARNs se replient et, à plus long terme, à mieux comprendre leur fonction.

Bibliographie

[AKT1995] Altman, S., Kirsebom, L. and Talbot, S. (1995) Recent studies of RNase P. In Söll, D. and RajBhandary, U. (eds.), *tRNA: Structure, Biosynthesis, and Function.* American Society for Microbiology, Washington, DC, pp.67-78.

[AU2002] Asahara, H. and Uhlenbeck, O.C. (2002) The tRNA specificity of *Thermus thermophilus* EF-Tu *Proc. Natl Acad. Sci. USA*, **99**, 3499-3504.

[BA 2002] Bruce Alberts *et al.* (2002) The Molecular Biology of the Cell, *Garland Science*, fourth edition, 342.

[BNHMS2000] Ban N., Nissen P., Hansen J., Moore P., Steitz T. (2000) The complete atomic structure of the large ribosomal suburnit at 2.4 Å resolution *Science* **289**, 905-20.

[BOBWGDHSS1992] Berman, H.M., Olson, W.K., Beveridge, D.L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A.R., Schneider, B. (1992). The Nucleic Acid Database: A Comprehensive Relational Database of Three-dimensional Structures of Nucleic Acids. *Biophys. J.*, **63**, 751-759.

[BRD1999] Batey, R.T., Rambo, R.P. & Doudna, J.A. (1999). Tertiary motifs in RNA structure and folding. *Angew. Chem. Int. Ed.*, **38**, 2326-2343.

[BSFECC1998] Bourdeau, V., Steinberg, S.V., Ferbeyre, G., Emond, R., Cermakian, N. and Cedergren, R. (1998) Amber suppression in *Escherichia coli* by unusual mitochondria-like transfer RNAs. *Proc. Natl. Acad. Sci. USA*, **95**, 1375-1380.

[BSP2000] Bullock, T.L., Sherlin, L.D. and Perona, J.J. (2000) Tertiary core rearrangements in a tight binding transfer RNA aptamer. *Nat. Struct. Biol.*, **7**, 497-504.

[BYTC1994] Biou, V., Yaremchuk, A., Tukalo, M. and Cusack, S. (1994) The 2.9 Å crystal structure of *T. thermophilus* seryl-tRNA synthetase complexed with $tRNA^{Ser}$. *Science,* **263**, 1404-1410.

[CD1998] Conn, G.L. & Draper, D.E. (1998). RNA structure. *Curr. Opin. Struct. Biol.*, **8**, 278-285.

[CG2000] Charette M, Gray MW (2000). Pseudouridine in RNA: what, where, how, and why. *IUBMB Life* **49** (5) 341-51.

[CGPZGKCD1996] Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R. & Doudna, J.A. (1996).

Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science*, **273**, 1678-1685.

[CGPZGSKCD1996a] Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Szewczak, A.A., Kundrot, C.E., Cech, T.R. & Doudna, J.A. (1996). RNA tertiary structure mediation by adenosine platforms. *Science*, **273**, 1696-1699.

[dBK1983] deBruijn, M.H.L. and Klug, A. (1983) A model for the tertiary structure of mammalian mitochondrial transfer RNAs lacking the entire "dihydrouridine" loop and stem. *EMBO J.* **2**, 1309-1321.

[DBMD2001] Doherty, E.A., Batey, R.T., Masquida, B. & Doudna J.A. (2001). A universal mode of helix packing in RNA. *Nat.Struct. Biol.*, **8**, 339-343.

[DKDW1995] Dirheimer, G., Keith, G., Dumas, P. & Westhof, E. (1995). Primary, secondary, and tertiary structures of tRNAs. In *tRNA: Structure, Biosynthesis, and Function*, (Söll, D. & Rajbhandary, U.L., eds), pp. 93-126, ASM Press, Washington, DC.

[DW2003] Du, X. & Wang, E.D. (2003). Tertiary structure base pairs between D- and TΨC-loops of *Escherichia coli*

$tRNA^{Leu}$ play important roles in both aminoacylation and editing. *Nucleic Acids Res.*, **31**, 2865-2872.

[EHLYWT2001] Elbashir S, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T (2001). "Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells". Nature **411** (6836): 494-8

[FG1995] Florentz, C. & Giege, R. (1995). tRNA-like structures in plant viral RNAs. In *tRNA: Structure, Biosynthesis, and Function*, (Söll, D. & Rajbhandary, U.L., eds), pp. 141-163, ASM Press, Washington, DC.

[FPHT1983] Freier SM, Petersheim M, Hickey DR, Turner DH (1984) Thermodynamic studies of RNA stability, *J Biomol Struct Dyn.* **5**, 1229-42.

[GE2000] Gott JM, Emeson RB (2000) Functions and mechanisms of RNA editing, *Annu Rev Genet.*, **34**, 499-531.

[GLG1993] Gehring, K., Leroy, J.L. & Gueron, M. (1993). A tetrameric DNA structure with protonated cytosine.cytosine base pairs. *Nature*, **363**, 561-565.

[HAG1999] Herr, A.J., Atkins, J.F., Gesteland, R.F. (1999). Mutations which alter the elbow region of $tRNA_2^{Gly}$

reduce T4 gene *60* translational bypassing efficiency. *EMBO J.*, **18**, 2886-2896.

[HB1999] Hamilton A, Baulcombe D (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science **286** (5441): 950-2.

[HBH2005] Hendrix DK, Brenner SE, Holbrook SR (2005) RNA structural motifs: building blocks of a modular biomolecule *Q Rev Biophys.* **38**: 3, 221-43.

[HH2003] Heus, H.A. & Hilbers, C.W. (2003). Structure of non-canonical tandem base pairs in RNA helices: review. *Nucleosides, Nucleotides, Nucleic Acids*, **22**, 559-571.

[HM1999] Hermann, T. & Patel, D.J. (1999). Stitching together RNA tertiary architecture. *J. Mol. Biol.*, **294**, 829-849.

[HP2000] Hermann, T. & Patel, D.J. (2000). RNA bulges as architectural recognition motifs. *Structure Fold. Des.*, **8**, R47-54.

[HS1988] 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.

[IHGSC2004] International Human Genome Sequencing Consortium (2004). "Finishing the euchromatic sequence of the human genome." Nature 431 (7011): 931-45

[INNONY2003] Ishitani, R., Nureki, O., Nameki, N., Okada, N., Nishimura, S. & Yokoyama, S. (2003). Alternative tertiary structure of tRNA for recognition by a posttranscriptional modification enzyme. *Cell*, **113**, 383-394.

[IS1999] Ioudovitch, A. and Steinberg, S.V. (1999) Structural compensation in an archaeal selenocysteine transfer RNA. *J. Mol. Biol.*, **290**, 365-371.

[J1997] George A. Jeffrey (1997) An Introduction to Hydrogen Bonding - Topics in Physical Chemistry, *Oxford University Press*, USA March 13

[K1958] Kendrew, J.C. et al. (1958) A three-dimensional model of the myoglobin molecule obtained by X-ray analysis, *Nature* **181**, 662–666

[KM2003] Krasilnikov, A.S. & Mondragon, A. (2003). On the occurrence of the T-loop RNA folding motif in large RNA molecules. *RNA*, **9**, 640-643.

[KSMS2001] Klein, D.J., Schmeing, T.M., Moore, P.B. & Steitz, T.A. (2001). The kink-turn: a new RNA secondary structure motif. *EMBO J.*, **20**, 4214-4221.

[KYPM2003] Krasilnikov, A.S., Yang, X., Pan, T. & Mondragon, A. (2003). Crystal structure of the specificity domain of Ribonuclease P. *Nature*, **421**, 760-764.

[KTF] http://www.ktf-split.hr/glossary/en_index.html

[L2001] Leach A.R. (2001) Molecular Modelling: Principles and Applications, *Pearson Education EMA*, second edition.

[LBKCRWM1998] Levinger, L., Bourne, R., Kolla, S., Cylin, E., Russel, K., Wang, X. & Mohan, A. (1998). Matrices of paired substitutions show the effects of tRNA D/T loop sequence on Drosophila RNase P and 3'-tRNase processing. *J. Biol. Chem.*, **273**, 1015-1025.

[LCG2003] Lee, J.C., Cannone, J.J. & Gutell, R.R. (2003). The lonepair triloop: a new motif in RNA structure. *J. Mol. Biol.*, **325**, 65-83.

[LS1988] Lavery, R., Sklenar, H. (1988). The Definition of Generalized Helicoidal Parameters and of Axis Curvature for Irregular Nucleic Acids. *J. Biomol. Struct. Dyn.*, **6**, 63-91.

[LW2001] Leontis, N.B., Westhof, E. (2001). Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499-512.

[LWU2001] LaRiviere, F.J., Wolfson, A.D. and Uhlenbeck, O.C. (2001) Uniform binding of aminoacyl-tRNAs to elongation factor Tu by thermodynamic compensation. *Science*, **294**, 165-168.

[M1972] Miller, J.H. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor Lab. Press, Plainview, NY, pp.352-355.

[M1999] Moore, P.B. (1999). Structural motifs in RNA. *Annu. Rev. Biochem.*, **68**, 287-300.

[MCFWTEG1980] Moras, D., Comarmond, M.B., Fischer, J., Weiss, R., Thierry, J.C., Ebel, J.P. & Giege, R. (1980). Crystal structure of yeast $tRNA^{Asp}$. *Nature*, **288**, 669-674.

[MF1988] 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.

[MF1988a] McClain, W.H. and Foss, K. (1988) Changing the acceptor identity of a transfer RNA by altering

nucleotides in a "variable pocket". *Science*, **241**, 1804-1807.

[MFJS1991] McClain, W.H., Foss, K., Jenkins, R.A. and Schneider, J. (1991) Four sites in the acceptor helix and one site in the variable pocket of $tRNA^{Ala}$ determine the molecule's acceptor identity. *Proc. Natl Acad. Sci. USA*, **88**, 9272-9276.

[MM1986] Masson, J.M. and Miller, J.H. (1986) Expression of synthetic suppressor tRNA genes under the control of a synthetic promotor. *Gene*, **47**, 179-183.

[ND] http://nationaldiagnostics.com/article_info.php/articles_id/3

[NF2002] Nagaswamy, U. and Fox, G.E. (2002) Frequent occurrence of the T-loop RNA folding motif in ribosomal RNAs. *RNA*, **8**, 1112-1119.

[NHU1994] Nazarenko, I.A., Harrington, K.M. and Uhlenbeck, O.C. (1994) Many of the conserved nucleotides of $tRNA^{Phe}$ are not essential for ternary complex formation and peptide elongation. *EMBO J.*, **13**, 2464-2471.

[NIBMS2001] Nissen, P., Ippolito, J.A., Ban, N., Moore, P.B. & Steitz, T.A. (2001). RNA tertiary interactions in the large

ribosomal subunit: the A-minor motif. *Proc. Natl. Acad. Sci. USA*, **98**, 4899-4903.

[NMAVGS2001] Nomanbhoy, T., Morales, A.J., Abraham, A.T., Vortler, C.S., Giege, R. and Schimmel, P. (2001) Simultaneous binding of two proteins to opposite sides of a single transfer RNA. *Nat. Struct. Biol.*, **8**, 344-348.

[NMKAM1986] Normanly, J., Masson, J.M., Kleina, L.G., Abelson, J. and Miller, J.H. (1986) Construction of two Escherichia coli amber suppressor genes: tRNAPheCUA and tRNACysCUA. *Proc Natl Acad Sci USA*, **83**, 6548-6552.

[NTK1999] Nashimoto, M., Tamura, M. and Kaspar, R.L. (1999) Minimum requirements for substrates of mammalian tRNA 3` processing endoribonuclease. *Biochemistry*, **38**, 12089-12096.

[OAK1995] Ogata, H., Akiyama, Y., and Kanehisa, M. (1995) A genetic algorithm based molecular modeling technique for RNA stem-loop structures. *Nucl. Acids Res.* **23**, 419-426.

[P1995] Pearlman, D.I., *et al.* (1995) AMBER 4.1, University of California, San Francisco, CA.

[PBSU1993] Peterson, E.T., Blank, J., Sprinzl, M. and Uhlenbeck, O.C. (1993) Selection for active *E.coli* $tRNA^{Phe}$ variants from a randomized library using two proteins. *EMBO J.*, **12**, 2959-2967.

[PCCRCFSSWK1995] Pearlman, D.A., Case, D.A., Caldwell, J.W., Ross, W.R., Cheatham, T.E., Ferguson D.M., Seibel, G.L., Singh, U.C., Weiner, P. & Kollman, P.A. (1995). *AMBER Version 4.1*. University of California.

[PFM1994] Pley, H.W., Flaherty, K.M. & McKay, D.B. (1994). Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature*, **372**, 111-113.

[PPFG1993] Puglisi, J.D., Putz, J., Florentz, C. & Giege, R. (1993). Influence of tRNA tertiary structure and stability on aminoacylation by yeast aspartyl-tRNA synthetase. *Nucleic Acids Res.*, **21**, 41-49.

[PU1992] Peterson, E.T. and Uhlenbeck, O.C. (1992) Determination of recognition nucleotides for *Escherichia coli* phenylalanyl-tRNA synthetase. *Biochemistry*, **31**, 10380-10389.

[PWU2000] Pleiss, J.A, Wolfson, A.D. and Uhlenbeck, O.C. (2000) Mapping contacts between Escherichia coli alanyl tRNA synthetase and 2` hydroxyls using a complete tRNA molecule. *Biochemistry*, **39**, 8250-8258.

[RCWEEEG1987] Romby, P., Carbon, P., Westhof, E., Ehresmann, C., Ebel, J.P, Ehresmann, B. and Giege, R. (1987) Importance of conserved residues for the conformation of the T-loop in tRNAs. *J. Biomol. Struct. Dyn.*, **5**, 669-687.

[QR1976] Quigley, G.J. and Rich, A. (1976) Structural domains of transfer RNA molecules. *Science* **194**, 796-806.

[QSMKSR1974] Quigley G.J., Suddath F.L., McPhersen A., Kim J.J., Sneden D., Rich A. (1974) The Molecular Structure of Yeast Phenylalanine Transfer RNA in Monoclinic Crystals, *Proc Natl Acad Sci USA*. **71**, 2146-50.

[RLFRBCK1974] Robertus, J.D., Ladner, J.E., Finch, J.T., Rhodes, D., Brown, R.S., Clark, B.F. & Klug, A. (1974). Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature*, **250**, 546-551.

[SG2005] Steffen P., Giegerich, R. (2005) Versatile and declarative dynamic programming using pair algebras. *BMC Bioinformatics*, **6**, 224.

[SHBIS1998] Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A. and Steinberg, S. (1998) Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.*, **26**, 148-153.

[SKT1991] SantaLucia, J. Jr., Kierzek, R. & Turner, D.H. (1991). Stabilities of consecutive A.C, C.C, G.G, U.C and U.U mismatches in RNA internal loops: evidence for stable hydrogen-bonded U.U and C.C+ pairs. *Biochemistry*, **30**, 8242-8251.

[SLC1997] Steinberg, S., Leclerc, F. and Cedergren, R. (1997) Structural rules and conformational compensations in the tRNA L-form. *J. Mol. Biol.*, **266**, 269-282.

[SM2000] Shi, H. and Moore, P.B. (2000) The crystal structure of yeast phenylalanine tRNA at 1.93 A resolution: a classic structure revisited. RNA, **6**,1091-1105.

[SQMSKKSR1974] Suddath, F.L., Quigley, G.J., McPherson, A., Sneden, D., Kim, J.J., Kim, S.H. & Rich, A. (1974). Three-

dimensional structure of yeast phenylalanine transfer RNA at 3.0 angstroms resolution. *Nature*, **248**, 20-24.

[SSA1994] Saks, M.E., Sampson, J.R., Abelson J.N. (1994) The transfer RNA identity problem: a search for rules. Science, **263**, 191-7.

[ST1993] SantaLucia, J. Jr. & Turner, D.H. (1993). Structure of $(rGGCGAGGCC)_2$ in solution from NMR and restrained molecular dynamics. *Biochemistry*, **32**, 12612-12623.

[TLAD1994] Tuohy, T.M., Li, Z., Atkins, J.F. and Deutscher, M.P. (1994) A functional mutant of tRNA(2Arg) with ten extra nucleotides in its TΨC arm. *J. Mol. Biol.,* **235**, 1369-1376.

[W2002] Westhof, E. (2002). Group I introns and RNA folding. *Biochem. Soc. Trans.*, **30**, 1149-1152.

[WBCMWCVHR2000] Wimberly B.T., Brodersen D.E., Clemons W.M. Jr., Morgan-Warren R.J., Carter A.P., Vonrhein C., Hartsch T., Ramakrishnan V. (2000) Structure of the 30S ribosomal subunit *Nature* **407**, 327-39.

[WC1953] Watson J.D. and Crick F.H.C. (1953) A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738

[WC1966] Wobble F, Crick F (1966). "Codon--anticodon pairing: the wobble hypothesis". *J Mol Biol* **19** (2), 548-55.

[WF2000] Westhof, E. & Fritsch, V. (2000). RNA folding: beyond Watson-Crick pairs. *Structure Fold. Des.*, **8**, R55-65.

[YYBLECN2001] Yusupov M.M., Yusupova G.Z., Baucom A., Lieberman K., Earnest T.N., Cate J.H., Noller H.F. Crystal structure of the ribosome at 5.5 Å resolution (2001) *Science* **292**, 883-96.

[ZDS2003] Zagryadskaya, E.I., Doyon, F.R. & Steinberg, S.V. (2003). Importance of the reverse Hoogsteen base pair 54-58 for tRNA function. *Nucleic Acids Res.*, **31**, 3946-3953.

[ZKS2004] Zagryadskaia, E.I., Kotlova, N. & Steinberg S.V. (2004). Key elements in maintenance of the tRNA L-shape. *J. Mol. Biol.* **340**, 435-444.

Annexe I : Accord des coauteurs de l'article

Nom de l'étudiant : Félix Doyon

Titre du programme : Maîtrise en bio-informatique, 2-468-1-0

Titre de l'article : Importance of the reverse Hoogsteen base pair 54-58 for tRNA function

Liste des coauteurs : Ekaterina I. Zagryadskaya, Félix R. Doyon, Sergey V. Steinberg

Status : Publié; Nucleic Acids Research

Déclaration des coauteurs :

À titre de coauteur de l'article identifié ci-dessus, je suis d'accord pour que Félix Doyon inclue cet article dans son mémoire de maîtrise qui a pour titre :

Étude des interactions entre les boucles D et T chez les ARNs de transfert (ARNts).

Coauteur : Sergey V. Steinberg
Signature :
Date :
Coauteur : Ekaterina I. Zagryadskaya
Signature :
Date :

Annexe II : Accord des coauteurs de l'article

Nom de l'étudiant : Félix Doyon

Titre du programme : Maîtrise en bio-informatique, 2-468-1-0

Titre de l'article : Specific and Non-Specific Purine Trap in the T-Loop of Normal and Suppressor tRNAs

Liste des coauteurs : Félix R. Doyon, Ekaterina I. Zagryadskaya, Jianhong Chen, Sergey V. Steinberg

Status : Publié; Journal of Molecular Biology

Déclaration des coauteurs :

À titre de coauteur de l'article identifié ci-dessus, je suis d'accord pour que Félix Doyon inclue cet article dans son mémoire de maîtrise qui a pour titre :

Étude des interactions entre les boucles D et T chez les ARNs de transfert (ARNts).

Coauteur : Sergey V. Steinberg
Signature :
Date :
Coauteur : Ekaterina I. Zagryadskaya
Signature :
Date :
Coauteur : Jianhong Chen
Signature :
Date :