Design of artificial tRNAs

DISSERTATION

zur Erlangung des akademischen Grades
Doctor rerum naturalium

Vorgelegt der
Formal- und Naturwissenschaftlichen Fakultät
der Universität Wien

von
Mag. Dagmar Friede

am Institut für Theoretische Chemie und molekulare Strukturbiologie

im März 2001

Abstract

RNA molecules are involved in numerous key cellular processes. Recent findings consolidated the current view that RNA is no longer regarded solely as a passive transporter of the genetic code, but as an extremely versatile class of molecules actively participating in all steps of gene expression. Major objectives are the design and preparation of RNA molecules with predicted structure and function in order to tackle hitherto unsolvable medical and pharmaceutical problems. Transfer RNAs provide excellent models for the determination of new properties, which mostly can be interpreted as features of RNA molecules in general.

A thorough investigation of eubacterial tRNA sequences revealed a considerable influence of base modifications on both the thermodynamic stability and the kinetic folding behaviour. An overwhelming amount of natural modified sequences indeed possess the established cloverleaf fold as minimum free energy structure in thermodynamic calculations. In kinetic folding simulations, however, the natural conformation is not necessarily formed directly by the majority of folding trajectories which are usually distributed among several conformations of substantial structural diversity partitioning the conformation space into basins separated by high energy barriers. The frequency of cloverleaf formation is not related to its appearance as mfe structure. Only sporadically, the cloverleaf presents the only structure that comes into question.

The fact is even more pronounced for the unmodified sequences. Apart from a decrease of the free energy difference between the ground state and the first suboptimal conformation, the lack of modified nucleotides leads to an enlarged amount of structures otherwise prevented to form, several of which offer alternative and more efficient pathways.

Insights concerning characteristic behaviour of natural tRNAs were subsequently applied to the design of artificial tRNA molecules. Using the complete sets of identity elements of some E. coli tRNAs as sequence constraints in inverse folding, a large amount of thermodynamically very stable sequences was obtained and subsequently sorted out due to inefficient folding behaviour. The yeast $tRNA^{asp}$ system was chosen as the first test object in collaboration with the team of Richard Giegé in Strasbourg because many data were available. Four designed tRNA sequences were synthesized by means of in vitro transcription. The test for aminoacylation capacity indicated that none of the sequences could be charged by the cognate synthetase. This result can be explained either in terms of the high mutation rate (Hamming distance of at least 30 to the natural $tRNA^{asp}$) or the lack of nucleotides known to participate in tertiary interactions with the enzyme. Another reason for the lack of aspartylation activity could lie in the rigidity of the system imposed by the strong stability constraint. Without a certain amount of flexibility to position the necessary nucleotides in close contact with the synthetase, no interaction can take place.

Zusammenfassung

RNA Moleküle sind an zahlreichen maßgebenden zellulären Prozessen beteiligt. Jüngste Erkenntnisse untermauern die geläufige Ansicht, daß RNA nicht mehr ausschließlich als passiver Übermittler des genetischen Codes, sondern als eine äußerst vielseitige Klasse an Molekülen mit aktiver Teilnahme an allen Schritten der Genexpression betrachtet werden sollte. Zu wichtigen Zielen gehören das Design und die Synthese von RNA Molekülen mit vorgebbaren Strukturen und Funktionen zur potentiellen Lösung medizinischer und pharmazeutischer Probleme. Transfer RNAs stellen exzellente Modelle für die Ermittlung neuer Eigenschaften dar, die sich meist als allgemeine Charakteristika von RNA Molekülen herausstellen.

Eine gründliche Untersuchung von eubakteriellen tRNA Sequenzen offenbarte einen beträchtlichen Einfluß von Modifikationen auf sowohl die thermodynamische Stabilität als auch die kinetische Faltbarkeit. Der überwiegende Anteil an modifizierten natürlichen Sequenzen besitzt tatsächlich die klassische Kleeblattstruktur als Struktur mit minimaler freier Energie in thermodynamischen Berechnungen. In kinetischen Faltungssimulationen jedoch wird diese natürliche Konformation nicht notwendigerweise direkt von dem Hauptanteil der Faltungstrajektorien gebildet; letztere sind normalerweise auf mehrere unterschiedlichste Strukturen verteilt, die den Konformationsraum in Becken mit hohen Energiebarrieren aufteilen. Die Häufigkeit der Kleeblattbildung steht nicht in Korrelation zu ihrem Auftreten als Struktur mit der niedrigsten Gesamtenergie. Nur selten stellt das Kleeblatt die einzige wahrscheinliche Strukturmöglichkeit dar.

Diese Tendenz ist noch ausgeprägter für unmodifizierte Sequenzen. Abgesehen von einer Abnahme der freien Energiedieferenz zwischen dem Grundzustand und der ersten suboptimalen Konformation führt die Abwesenheit von modifizierten Nukleotiden zu einer größeren Menge ansonsten verbotener Strukturen, von welchen einige alternative und effizientere Faltungswege ermöglichen.

Erkenntnisse über das charakteristische Verhalten natürlicher tRNAs wurden anschließend zum Design von künstlichen tRNA Molekülen herangezogen. Der vollständige Satz an "identity elements" einiger E. coli tRNAs wurde als Sequenzvorgabe bei der Umkehr der konventionellen Vorhersage von RNA Sekundärstrukturen benutzt. Die resultierende, umfangreiche Menge thermodynamisch sehr stabiler Sequenzen wurde in der Folge angesichts schlechter Faltungseigenschaften reduziert. Das gut untersuchte und mit vielen Daten belegte Hefe $tRNA^{asp}$ System wurde in Kollaboration mit der Forschungsgruppe von Richard Giegé in Strasbourg als erstes Testobjekt gewählt. Vier nicht natürliche tRNA Sequenzen wurden mittels in vitro Transkription synthetisiert. Der Test auf Aminoacylierungsaktivitaet zeigte allerdings an, daß keine der Sequenzen von der entsprechenden Synthetase beladen werden konnte. Dieses Resutat läßt sich entweder auf die hohe Mutationsrate (Hamming-Distanz von zumindest 30 zu der natürlichen $tRNA^{asp}$) oder auf den Mangel an tertiären Kontakten mit der Syn-

thetase beteiligter Nukleotide zurückführen. Eine andere Erklärung für die negative Aspartylierungsaktivität könnte die in der Forderung nach hoher Stabilität begründete Starrheit des Systems bieten. Ohne ein gewisses Maß an Flexibilität können die notwendigen Nukleotide nicht in engen Kontakt mit der Synthetase treten und folglicherweise kann keine Wechselwirkung erfolgen.

Contents

1 Introduction

Given their broad range of functions, the most essential biopolymers, namely proteins and nucleic acids (DNA and RNA), are constructed from an apparently moderate number of building blocks. As opposed to proteins which contain 20 different amino acid subunits, nucleic acids are built up by only four monomers with similar chemical properties. Whereas DNA is the carrier of genetic information, proteins are responsible for directing protein synthesis. The role of RNA molecules as intermediate stages in the translation of the genetic code and the subsequent synthesis of important biological molecules has been a well-established fact all along. A more recently acknowledged finding concerns RNA ability to exhibit catalytic activity. Naturally occurring "ribozymes" have been shown to efficiently catalyze the formation and cleavage of nucleic acid phosphodiester bonds. This narrow range of catalyzed reactions has soon been expanded by directed *in vitro* evolution methods to yield a wide variety of aminoacyl transferase ribozymes, including self-aminoacylating RNAs [55], amide and peptide synthetases [148] as well as 3' to 2' or 5' acyl transferases [57, 83]. The discovery of such diverse catalytic functions led to a reconsideration of the origin of life and the order of appearance of DNA, RNA and proteins during early biological evolution. The 'RNA World' hypothesis [59–62] assumes that the chemical processes resulting in the appearance of life were carried out by RNA molecules which implies the ability to catalyze RNA replication and govern peptide synthesis. A most recent result supporting the idea reports the isolation of a ribozyme by *in vitro* evolution that can specifically aminoacylate a tRNA [74].

Furthermore, RNA seems to provide a system simple enough to study genotype-phenotype relationships. Spiegelman [69] identifies the nucleotide sequence as the genotype and the molecular structure as its phenotype. Thus they represent two different expressions within the same molecule. Such sequence-structure mappings can be modeled by means of a sophisticated algorithm [36].

While direct computation of full three-dimensional structures from sequences remains a challenging task in bioinformatics and structural biology, secondary structures as coarse-grained versions of the spatial structure are accessible for dynamic programming algorithms [51] which yield either the minimum free energy (mfe) structure [150] alone or together with Boltzmann weighted suboptimal conformations in the sense of a partition function [90]. As an alternative, computer programs designed to determine kinetic structures [43, 87, 96, 97] take the

time limit for RNA folding into account. RNA folding is commonly accepted as proceeding in a hierarchical fashion [11]. Folding of the polynucleotide backbone leads to next-neighbour interactions which give rise first to secondary structure elements and then to 3D architectural motifs which finally assemble to form the tertiary structure. Among the molecules for which experimental data is available on both structure levels, tRNAs present a most interesting group.

Transfer RNAs are the interfaces between DNA storing the genetic information and proteins in which this information is unfolded. Apart from presenting key elements in the translation apparatus, they are also assumed to take part in transcription. In fact, tRNAs are involved in retroviral genome replication [79,85,142] and tRNA-like domains participate in plant viral RNA replication [86,103]. RNA function is closely related to its spatial structure, therefore theoretical predictions and modeling of tRNA structure may help to shed light on the molecular reasons for the different functioning of "tRNA" in viruses, procaryotes and eucaryotes. Our present understanding of RNA folding is still largely based on classical studies of tRNAs. The early determination of $tRNA^{phe}$ crystal structure [65, 116] lay the foundation for speculation of a compact, globular fold which has since been confirmed by the analysis of 3D structures [32] of the hammerhead ribozyme, the hepatitis delta virus ribozyme and the P4-P6 domain of *Tetrahymena* group I intron in recent years. Thermal unfolding experiments [21, 22, 115] and kinetic measurements [14, 129, 130] of tRNA helped to elucidate the role of ion binding. Since most of the knowledge gained through studies of tRNA holds true for various RNA molecules, transfer RNAs are obviously well suited as models aimed to discover novel properties applicable to the RNA field in general.
The design and subsequent synthesis of artificial sequences with predetermined properties and functions is one of the most interesting and challenging objectives as it opens up new perspectives. For example, the first steps towards a site-specific incorporation of non-natural amino acids into proteins *in vivo* have been successfully carried out [80]. The most important requirement for such an expansion of the genetic code is the creation of a new synthetase-tRNA pair. Aminoacyl-tRNA-synthetases catalyze the linkage of tRNAs to amino acids. Specific recognition of the correct tRNA and repulsion of all others is ensured by the occurrence of 'identity elements'. Protein engineering in the way mentioned above implies mutation of both synthetases and tRNAs in order to guarantee interaction with an artificial amino acid. An alternative goal is the synthesis of new

molecules meant specifically to inhibit tRNA-recognizing enzymes.
With this work we present reasonable sequence proposals for both approaches. First, a thorough analysis of naturally occurring tRNAs is conducted in order to gain information concerning their thermodynamic and kinetic properties. Special regard is given to base modifications and the discrepancy in the behaviour of unmodified sequences. This knowledge is subsequently utilized to exclude artificial sequences from a very large pool of variants. Application of two criteria essential for the reliability of structure prediction, thermodynamic stability and kinetic "foldability", leads to a considerable reduction in the number of candidates. The introduction of coaxial stacking as an important interaction in the folding algorithm resulted in a strong stabilization of the cloverleaf fold. For systems of which the complete sets of identity nucleotides are determined, their consideration further rules out some of the sequences. The remaining sequences fulfilling all these vital constraints are regarded as promising variants for synthesis.

2 General features of tRNAs

Transfer RNAs play a central role in gene expression as adaptor molecules that translate the codons in mRNA to amino acids in a protein. Exploration of the role of tRNAs in protein synthesis, determination of their primary sequences and clarification of their tertiary structure have posed an attractive challenge to researchers since their discovery about four decades ago. As is the case for proteins, the structural knowledge of RNAs is imperative in order to understand their function. One of the first techniques for the prediction of RNA secondary structure was the comparative analysis of an aligned set of sequences, which was successfully demonstrated for tRNAs in as early as 1965 by Holley et al. [54]. In just a few tRNA sequences (the first verified one being the one of $tRNA^{ala}$), they observed regions in the sequence showing covariation according to a base-pairing scheme from which they were able to deduce the now commonly accepted cloverleaf-model. The significance of their suggestion (using the fact that the probability of the base-pairing covariations arises randomly) was further validated by finding confirming patterns in subsequently elucidated tRNA sequences [37, 38]. Levitt [76] correctly predicted several isolated base pairs and a base triplet a few years later. With the determination of the yeast $tRNA^{phe}$ crystal structure [117], many of Holley's and Levitt's assumptions were confirmed. Since then, over 4250 sequences have been found, the majority by means of sequencing the corresponding genes.

Complete sets of tRNAs from one organism, including at least one isoacceptor species for each of the twenty amino acids, are known for several eubacteria (*Mycoplasma capricolum, Bacillus subtilis, Escherichia coli*), yeast (*Saccharomyces cerevisiae*) and chloroplasts (*Euglena gracilis, Marchantia polymorpha, Nicotiana tabacum*) or mitochondria (*Torulopsis glabra, ratus ratus*). The number of genes for a particular isoaccepting tRNA varies depending on the organism. Although these genes might have the same primary structure, it is more common that isoacceptor tRNAs feature the same anticodon but slightly differing sequences. In yeast, for example, the two $tRNA^{phe}_{GAA}$ [63] and the two $tRNA^{thr}_{IGU}$ [141] are identical except two nucleotides. Compensatory mutations frequently occur in the case when the difference between two isoacceptors is located in a stem. Again in yeast $tRNA^{phe}$, an A-U base pair in the amino acid acceptor stem is exchanged for a G-C pair. The same replacement is found in yeast $tRNA^{thr}$, albeit in the Tψ stem.

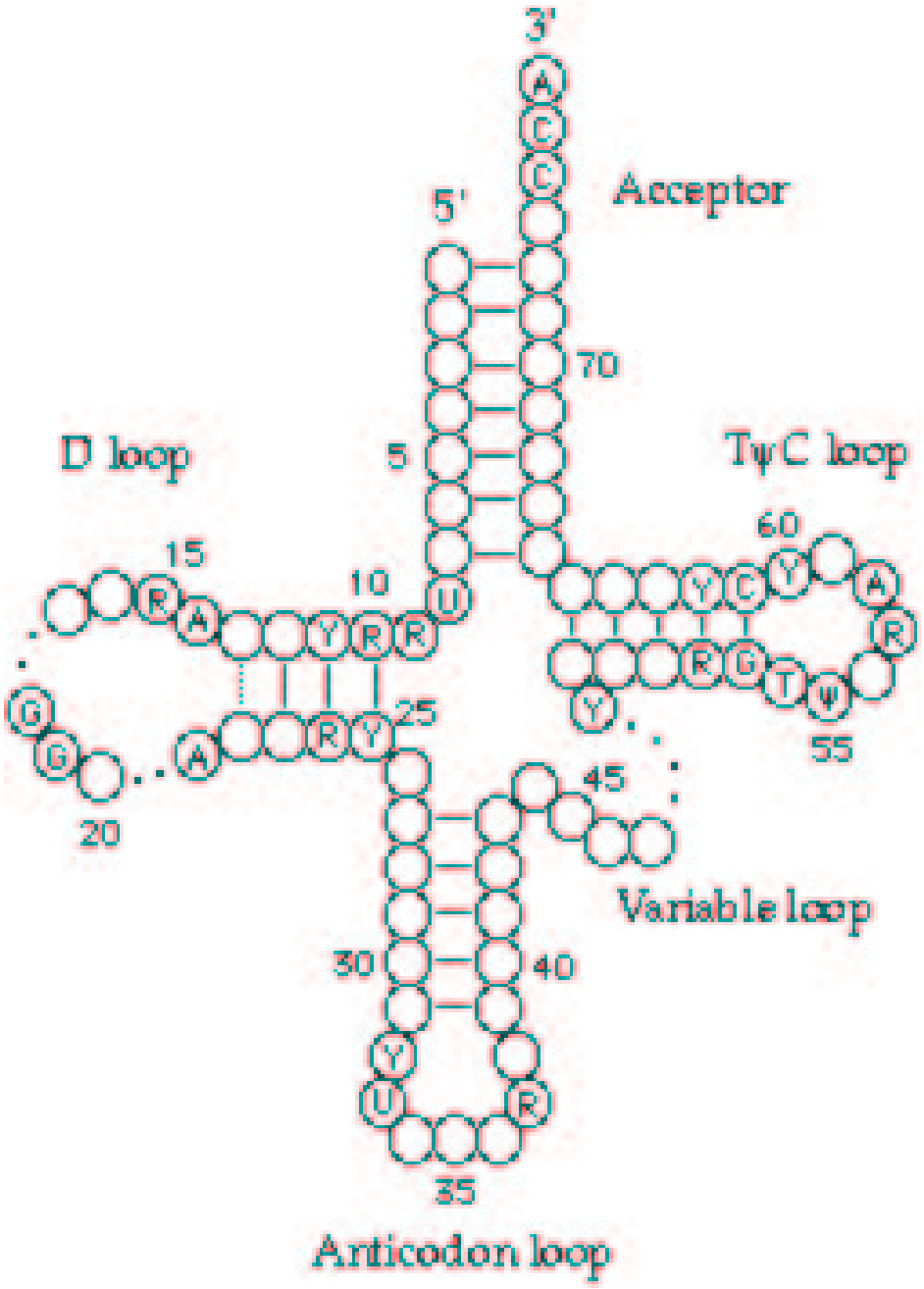


Figure 1: The characteristic secondary structure of tRNAs presenting a cloverleaf. Conserved nucleotides are emphasized (R = purine, Y = pyrimidine).

The canonical cloverleaf model first proposed by Holley [54] and illustrated in figure 1 consists of three hairpins, a variable region, a terminal stack and a 3' single-stranded $NCAA_{OH}$ end to which the amino acids become attached. Stems and loops can be related to their different domains according to their size. Acceptor stems are composed of seven base pairs, and thus represent the longest stems. There are five base pairs to be found in both anticodon and Tψ stems, and three or four pairs in D stems, depending on the class. The former D stem length is characteristic for class II tRNAs. This family of tRNAs is differentiated from class I by the length of the variable region which in the former case is extended to 10 to 24 nucleotides. Class II only comprises leu, ser and eubacterial and organellar tyrosine-specific tRNAs.

The frequent occurence of non-canonical G-U base pairs [88] is a noticeable feature of stem regions. Since their first discovery in $tRNA^{ala}$ [54], other possible non-canonical pairs (for example A-A, C-C, C-U, G-A, U-U, U-Y) have been detected in the stems of various tRNAs [75]. G-U "wobble" pairs, however, occur with the highest frequency. As to stems, a frequently occuring length can be attributed to loops as well. Anticodon and T loops contain seven nucleotides, whereas D loops and variable regions are areas of various lengths.

An important discovery regarding the primary structure was made in the early 1970s. Certain positions in tRNAs are occupied by invariant or semi-invariant nucleotides. Twenty-seven residues show this characteristic (U_8, R_9, R_{10}, Y_{11}, A_{14}, R_{15}, G_{18}, G_{19}, A_{21}, R_{24}, Y_{25}, Y_{32}, U_{33}, R_{37}, Y_{48}, R_{52}, G_{53}, T_{54}, ψ_{55}, R_{57}, A_{58}, Y_{60}, C_{61}, Y_{62}, C_{74}, C_{75}, A_{76}: R = purine, Y = pyrimidine). Their positions are indicated in Figure 1.

Transfer RNA is the most extensively modified nucleic acid in the cell. Modified nucleotides are contained in tRNAs from all three phylogenetic domains (archaea, bacteria, eucarya [127, 143]). The modifications are not introduced during transcription, but are formed after the synthesis of the polynucleotide chain, serving for an improvement of the specificitiy and efficiency of tRNA biological functions. To date, more than eighty modified residues have been discovered and their chemical structures revealed [10]. Modified nucleotides are located at 61 different positions in tRNAs, mainly in loop regions. A large variety is present in the anticodon area, especially in the first position of the anticodon (position 34), and one base 3' to the anticodon (position 37). Apart from one exception (archaeosine at position 15 in archael tRNAs [28], all hypermodified residues are

found in this region. Minor modifications like methylated or thiolated derivatives are usually situated outside the anticodon, with only one or two kinds of modified nucleosides present at each position. Some are common to almost all species, such as Dihydrouridine in D loops and Ribothymine in Tψ loops, whereas others are characteristic of specific tRNAs. Examples are found in the hypermodified wybutosine residue (a guanosine derivative) at position 37 in almost all eukaryotic $tRNA^{phe}$ (except that from *Bombyx mori* and *Drosophila melanogaster*) and queueosine (another complicated post-transcriptional modification of guanosine) at the first anticodon position of certain tRNAs specific for tyr, his, asn and asp from eubacteria and eukaryotes. In both domains, modification takes place at different stages during the processing of precursor tRNA, depending strongly on the concentration of the substrate as well as on both the amount and the activity of tRNA-modifying enzymes. Several studies have been carried out on precursor $tRNA^{tyr}$. The biosynthesis in *Xenopus laevis* oocytes initiated by injection of the yeast $tRNA^{tyr}$ gene into either the nucles or the cytoplasma revealed that most base modifications occur in a sequential fashion in the nucleus before splicing [94, 100]. In particular, ψ35 is introduced in the intron-containing pre-tRNA [58], a fact which has been confirmed by respective findings in *Drosophila melanogaster* and plants [128]. Apart from this nucleotide, though, modifications in the anticodon region, especially in positions 34 and 37, are primarily synthesized in the mature tRNA, and seem to mark the end of the complicated tRNA maturation process [124].

While modified nucleosides in several positions (8, 26, 32, 38, 39, 40, 46, 47, 54, 55) have proven to exert no significant influence on the aminoacylation efficiency, certain modifications in the anticodon effect an altered conformation and therefore play an important role in codon recognition [78]. For example, the modification of cytidine to lysidine at position 34 of *E. coli* $tRNA_2^{ile}$ [95] leads to a direct involvement in the aminoacylation process. The aminoacylation identity is switched from methionine to isoleucine by means of this single modification.
Correct expression of the genetic code at translation is directly correlated with tRNA identity. Recognition and specific aminoacylation of tRNAs by aminoacyl-tRNA-synthetases (aaRS) is therefore a key step in protein synthesis. Codon-anticodon interactions between messenger RNA and tRNA that lead to selection of amino acids during translation are independent of the nature of the amino acid esterified to the tRNA. Misaminoacylation of a tRNA would therefore un-

avoidably result in the incorporation of an erroneous amino acid in the protein. To synthesize the proper protein, the tRNA must exhibit high specificity in both translation and aminoacylation. The latter depends on the interaction with the cognate synthetase.

There are twenty different aminoacylation systems, one for each amino acid and tRNA type. Accurate discrimination between tRNAs presents a difficult challenge to aaRS enzymes because of the relative structural conformity among the various tRNAs. In contrast to tRNAs, the synthetases vary considerably in size and oligomeric state. Based on sequence similarities and the universal conservation of two mutually exclusive sets of sequence motifs, aaRS were split into two classes [30], with entirely different ATP-binding domains [24]. Class I aaRS's show a dinucleotide fold previously reported for dehydrogenase enzymes and aminoacylate the 2'-OH group of the terminal adenosine of the tRNA, whereas class II enzymes attach the amino acid to the 3'-OH group and contain an ATP-binding domain found in very few proteins.

The differentiation between systems is mediated by sets of certain structural elements, called "identity determinants" [40, 91, 92, 101, 104, 121], in both the tRNA and synthetase. They are not to be confused with the recognition elements, the term of which applies to nucleotides in contact or close proximity with synthetases. Other residues can still cause interaction with the protein by putting other nucleotides in a correct position for contact. The concept of identity determinants implies the probability of complex formation between chemical groups of both tRNA and synthetase.

With the before-mentioned exception of L34 in *E. coli* $\mathrm{tRNA^{ile}}_2$ and a thiolated uridine derivative at position 34 in *E. coli* $\mathrm{tRNA^{glu}}$ [119] which both help to govern specific recognition and aminoacylation, post-transcriptionally modified nucleosides do not constitute major identity elements. Neither do conserved or semi-conserved residues. As the anticodon specifies the decoding capacity of a tRNA, it would seem to be a potential candidate for tRNA acceptor identity. The assumption was soon verified in many cases (for a review of results dating before 1985, see [66]). The most convincing way to demonstrate that an anticodon dictates the aminoacylation of a tRNA is to replace the anticodon of one tRNA with that of another, thereby showing that the amino acid acceptor identity coincides with the anticodon [64].

However, the anticodon per se cannot possibly be a basis for discrimination due to the degeneracy of the genetic code. Nucleotides in the distal part of the ac-

ceptor arm constitute another group playing an important role in recognition, especially the first three base pairs and residue 73. The latter contributes to the identity of virtually every tRNA species, even when a family of tRNA isoacceptors contain different nucleotides at this site; it is appropriately called the "discriminator" base [93]. It is known to avert interactions with non-cognate aminoacyl-tRNA synthetases [136], and to stabilize the transition state of the acylation reaction [31, 34, 72, 126].

A compilation of identity determinants in both *E.coli* and *S.cerevisiae* is listed in Table 1 [41], presented according to the partition of synthetases in two classes. For *E. coli*, identity elements are known for all aminoacylation systems, while 14 sets have been compiled for *S. cerevisiae*. Mostly standard nucleotides are concerned, modified residues only play an important role as determinants in *E. coli* tRNAs specific for ile, glu and lys as well as ile in yeast. In these cases, the modifications involved in recognition are restricted to the anticodon loop. There is some flexibility with respect to the spatial distributions of identity elements in different tRNAs. For both *E. coli* and *S. cerevisiae* aminoacylation systems, the identity nucleotides are predominantly located in the anticodon and acceptor arm regions. Position 37 in the anticodon loop only contributes to the identity of tRNAs charged by class II enzymes. Nucleotides in other domains (nt 8-31 and 39-65) are involved in six class I (ile, leu, cys, glu, gln, arg) and four class II (ala, ser, pro, phe) identities. Altogether, 40 positions have been discovered to serve as identity determinants, including all seven positions of the anticodon loop, the discriminator base N73 and the last five base pairs of the acceptor stem. The remaining base pairs of the acceptor arm as well as the entire T stem never participate in tRNA identity. Taking a closer look at the distribution of identity elements in *E. coli*, it is remarkable that out of 16 systems where tRNA identity depends on anticodon residues, the middle position 35 is always used. Positions 34 and 36 appear less often and with variable nucleotides in ile, gln and arg. Only leu, ser and ala are devoid of anticodon identity determinants. The discriminator base N73 makes a contribution to 18 *E. coli* systems (glu [125] and thr [48] being the exceptions). The situation is similar for yeast. However, the ala and arg identity sets do not include the discriminator position, and anticodon residue 35 does indeed determine leu identity.
Since identity elements are a prerequisite for proper tRNA function, their inclusion poses a relevant constraint in sequence design.

E.coli	**S.cerevisiae**	**A**
A73;G3:C70;U4:A69 A35;C36 -	A73 A35 -	val
A73;C4:G69 G34;A35;U36;t6A37;A38 U12:A23;C29:G41	- I34;A35;U36 -	ile
A73 - U8·A14	A73 A35;G37 -	leu
A73;U4:A69;A5:U68 C34;A35;U36 -	A73 C34;A35;U36 + 4 other AC loop nts D arm	met
U73;G2:C71;C3:G70 G34;C35;A36 G15·G48;A13·A22	U73 - -	cys
A73 U35 -	A73;C1:G72 G34;ψ35 -	tyr
G73;A1:U72;G2:C71;G3:C70 C34;C35;A36 -	- C34;C35 -	trp
G1:C72;U2:A71 s4U34;U35;A37 U11:A24;U13:G22-A46;d47	- - -	glu
G73;U1:A72;G2:C71;G3:C70 Y34;U35;G36;A37;U38 G10	- - -	gln
A/G73 C35;U/G36 A20	- C35;U/G36 -	arg

Table 1: Identity elements in tRNAs aminoacylated by class I (A) and class II (B) synthetases.

E.coli	**S.cerevisiae**	**B**
G73;C72;G2:C71;A3:U70;C11:G24;R4:Y69 - C11:G24; variable loop	- - variable loop	ser
G1:C72;C2:G71 G34;G35;U36 -	G1:C72 G35;U36 -	thr
A73;G72 G35;G36 G15·C48	- - -	pro
U73;G1:C72;C2:G71;G3:C70 C35;C36 -	A73;C2:G71;G3:C70 C35;C36 -	gly
G73:G1 anticodon -	A73;G1 G34;U35 -	his
G73;G2:C71 G34;U35;C36;C38 G10	G73 G34;U35;C36;C38 G10·U25	asp
A73 U34;U35;U36;mnm^{5}s^{2}U34 -	- - -	lys
G73 G34;U35;U36 -	- - -	asn
A73 G34;A35;A36;G27:C43;G28:C42 U20;G44;U45;U59;U60	A73 G34;A35;A36;i^{6}A37 G20	phe
A73;G2:C71;G3·U70;C4:C69 - G20	G3·U70 - -	ala

Table 1 continued.

2.1 Processing of tRNAs

Transfer RNAs are synthesized in the form of larger transcripts (pre-tRNAs), which undergo cleavage at both ends, en route to becoming mature tRNAs. Aside from tRNAs embedded in pre-rRNA transcripts, the other tRNAs are synthesized in transcripts that contain one to seven tRNAs each, all surrounded by long flanking sequences. Since the discovery of tRNA precursors over two decades ago, a great amount of research has dealt with the decoding of the processing pathway from initial transcript to functional tRNA. The number and possible compulsory order of the individual steps, identity and function of the nucleases and the structure of intermediate products are only a few questions that have been posed and partly answered. Extensive studies on *E. coli* and bacteriophage-infected *E. coli* led to the elucidation of the processing pathway in bacteria.

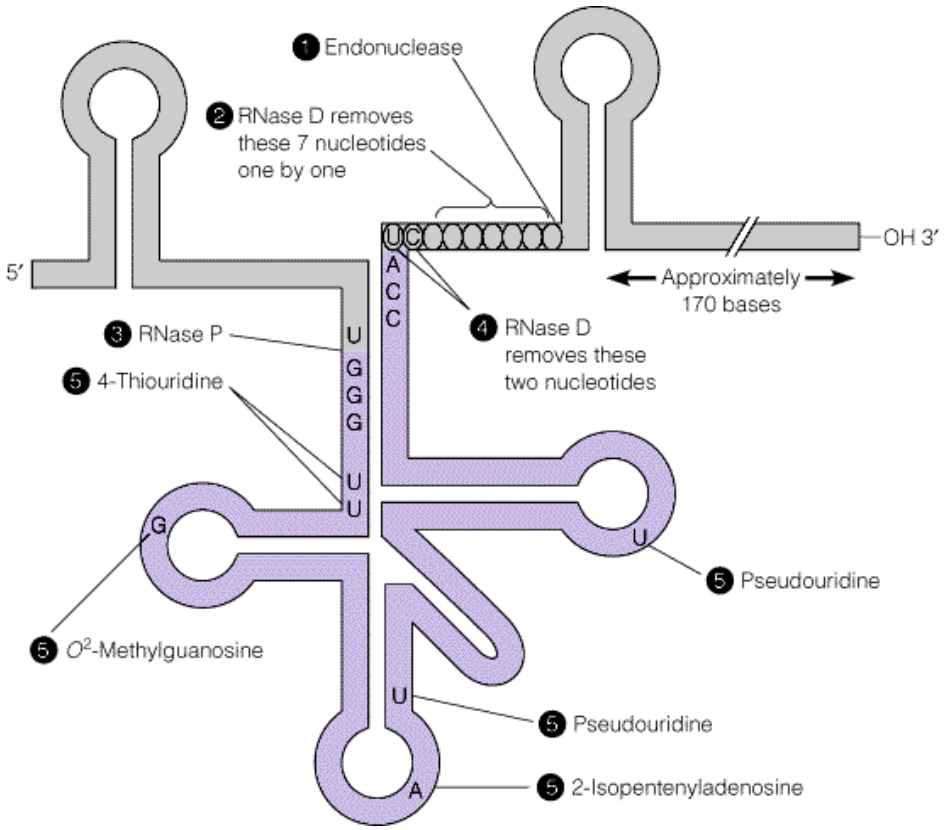


Figure 2: Post-transcriptional processing of bacterial tRNA (Example: *E. coli* $tRNA^{tyr}$).

The maturation steps are summarized in figure 2, taking the well-studied case of *E. coli* $tRNA^{tyr}$ as an example.
The process is subdivided into five distinctive steps:

1. Maturation starts with an endonuclease that cleaves at a stem/loop structure on the 3' end of the tRNA sequence.

2. Ribonuclease D carries out exonucleolytic cleavage to a point two neucleotides removed from the CCA sequence at the 3' end.

3. The 5' end is created by Ribonuclease P which cleaves to leave a phosphate on the 5' terminal guanosine. This enzyme provides the 5' terminus for **all** tRNA molecules. The structural features recognized by RNase P have not yet been explicitly determined, for different sequences are found at the cleavage sites. Ribonuclease P consists of one RNA molecule of 377 nucleotides and one protein molecule with a moleculuar weight of about 20.000. Both components are essential for full catalytic activity, but under nonphysical conditions the RNA part alone can catalyze accurate cleavage. RNase P therefore belongs to the class of ribozymes whose characteristic is catalytic function.

4. After the proper 5' terminus has been created, RNase D removes the remaining two nucleotides from the 3' end. Some tRNA genes do not code for the 3' terminal CCA. In these cases, the last three nucleotides are added by an enzyme that specifically recognizes the 3' end of tRNAs lacking the amino acid binding site.

5. Creation of the modified bases common to tRNAs occurs at the final stage, including methylations, thiolations, reduction of uridine to dihydrouridine, and so forth.

In contrast to eubacteria, no uniform mechanism has yet been determined for eucaryotes. Different precursors show different maturation even in the same cell extract. Since most eucaryotic substrates lack a proper amino acid binding site, an additional step in the processing pathway described above is required. Nucleases creating final 3' maturation can only remove trailer sequences in a way that allows tRNA nucleotidyltransferase to add the CCA end.
The continuity of tRNA genes is frequently interrupted by intervening sequences

that are not present in the mature tRNAs. Transcripts of these genes also contain the extra nucleotides (introns). Their excision (splicing) constitutes an additional maturation step before the functional tRNA can be exported into the cytoplasm. Introns are found in the tRNA genes of organisms in all three of the great lines of descent : the bacteria, the archaea, and the eucarya. However, the splicing mechanism of their precursors is distinctly different. The first introns in tRNA genes were discovered in yeast *S. cerevisiae* [138]. Experiments by Hopper and co-workers revealed a source of precursor substrates which led to the development of the first *in vitro* RNA splicing system [67, 102]. Based on this system, the tRNA splicing pathway could be deduced [68, 105].

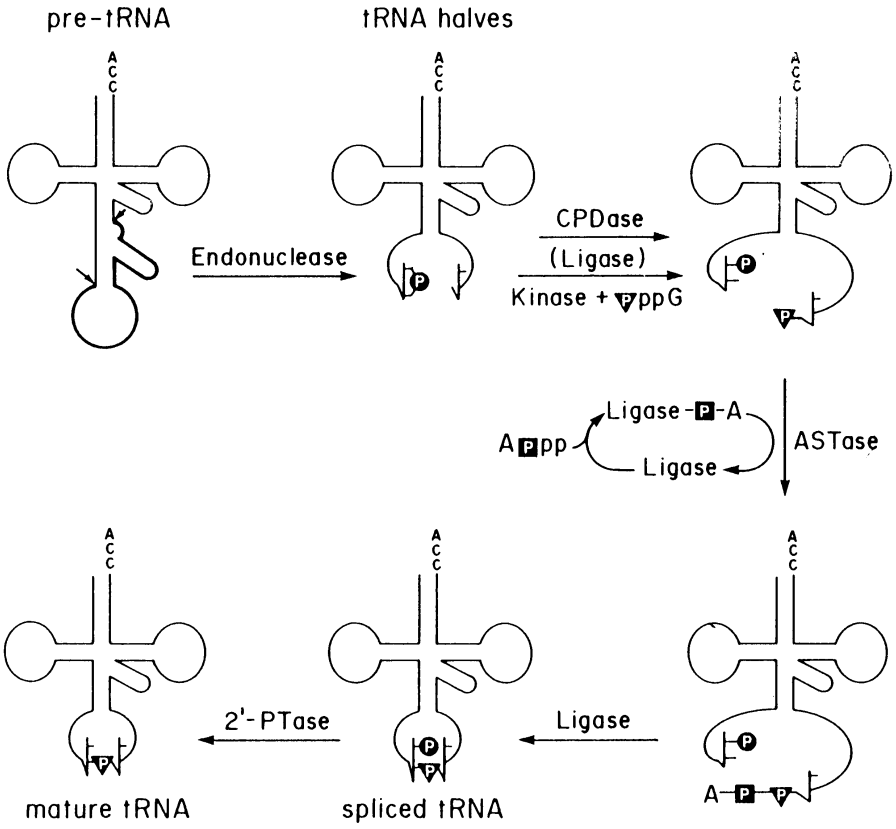


Figure 3: Splicing pathway in yeast.

The splicing reaction in yeast is a stepwise process where each step is catalyzed by a distinct enzyme (see figure 3).
The first step consists of the site-specific cleavage by an endonuclease, producing two tRNA half-molecules and the linear intron with 5'-OH and 2',3'-cyclic phosphate ends. The following ligation of the splicing intermediates is in itself a three-step process. The cyclic PO_4^{--} is opened to a 2'-phosphate by phosphodiesterase action. The second step arranges the phosphorylation of the 5'-OH by polynucleotide kinase. After the ligase is adenylated, transfer of its AMP to the 5'-phosphate end occurs. The ligase joins the two halves under the release of AMP. Finally, the 2'-phosphate having remained at the splice junction has to be removed by (NAD)-dependent 2'-phosphotransferase action. Since there is no conservation of sequence either in the introns between families or at the splice junction among the 10 different tRNA precursors in yeast, it was previously assumed that the intron itself plays no major role in the recognition by the splicing enzyme. The secondary structures of pre-tRNAs can be predicted from their mature counterparts and were indeed verified in a few cases [73, 132]. As only the conformation of the mature domain is retained in pre-tRNAs, it was proposed that endonuclease recognizes the splice sites by measuring the distance from the conserved domain to the splice sites. Changing features of the mature domain by means of insertion or deletion mutations strongly affected splicing specificity. Recent results have indicated that the intron actually takes an active part in the recognition process. According to Tocchini-Valentini and co-workers, recognition in *Xenopus laevis* depends on a tertiary base pair interaction between a conserved purine residue in the intron (located three nucleotides upstream of the 3' splice site) and a pyrimidine at position 32 in the anticodon loop [5].
Similar as in eucarya, archael introns are also small and often interrupt the anticodon loop immediately 3' to the anticodon (although other sites have been detected as well, including the D stem). However, the ligation step must differ from the one observed in eucarya since no homologue of the eucaryal tRNA splicing ligase has been found in the complete genome sequence of several members of the Archaea. Furthermore, recognition of the splice sites depends solely on a conserved structural motif consisting of two loops of three bases connected by a four base pair helix, the bulge-helix-bulge motif [135]. Characteristics of the mature domain do obviously not constitute a decisive factor in the recognition mechanism. Despite the differences in both substrate and recognition process between eucaryal and archael systems, it seems that the arrangement of the active

sites is conserved between the respective enzymes [77].

A controversial issue which remains unsolved, is whether introns are ancient features of gene structure ("introns early") or whether they are more recently acquired elements obtained through horizontal gene transfer ("introns late"). In support of the more recent acquisition of introns is evidence that some introns are mobile elements. However, the intron early view suggests that introns were present in the earliest progenitors of modern cells and there has been a trend in some organisms to lose introns, thus streamlining their genome and and allowing for more rapid growth and response to environmental factors. The class of self-splicing introns, called Group I introns, which occur in the tRNA genes of chloroplasts of higher plants, bacteriophages and viruses, archaebacteria, and (recently discovered) in certain eubacteria (a- and b-proteobacteria, cyanobacteria, [114]) lends credibility to the ancient origin of these sequences.

The first discovery of eubacterial group I introns occured in the $\mathrm{tRNA}^{\mathrm{Leu}}_{\mathrm{UAA}}$ genes of five cyanobacterial species. A homologous intron was observed at the same position in several plastid genomes, suggesting an inheritance from a common ancestor. The $\mathrm{tRNA}^{\mathrm{Leu}}_{\mathrm{UAA}}$ therefore represents the most likely example of an ancient intron, its origin predating the cyanobacterial endosymbiosis giving rise to plastids. A challenge to the introns early view was however presented by the identification of $\mathrm{tRNA}^{\mathrm{Leu}}_{\mathrm{UAA}}$ introns in the cyanobacterium Microcystis aeruginosa, which seem to have developed independently of the previously mentioned cyanobacterial ones through horizontal transfer. Phylogenetic analysis and comparison with two other eubacterial self-splicing introns [113] in the $\mathrm{tRNA}^{\mathrm{Ile}}_{\mathrm{CAU}}$ (Azoarcus sp. BH72, a β-purple bacterium) and the $\mathrm{tRNA}^{\mathrm{Arg}}_{\mathrm{CCU}}$ (Agrobacterium tumefaciens, an α-purple bacterium) have led to the conclusion that the Mycrocystis introns share a more recent common ancestor with the latter than either does with the other cyanobacterial $\mathrm{tRNA}^{\mathrm{Leu}}_{\mathrm{UAA}}$ introns.

Many attempts have been made to bring the introns early/late debate to a satisfying resolution in the last few years. However, arguments for either hypothesis are abundant. Recently, Rzhetsky and co-workers reasoned that the pattern of intron distribution in a large gene family resembles more one of addition or movement than loss [120], and Cho and Doolittle derived from intron position alignment that the distribution in genes encoding proteins is due to random insertion instead of loss [18]. De Souza et al. on the other hand claim that about 35% of present day intron positions in ancient proteins have been identified as ancient introns which supports the introns late theory [26].

3 RNA structure

The structure formation of RNA can be viewed as a two-stage process. First, the string of bases, called the **sequence** or **primary structure**, is converted into a set of complementary base pairs denoted as the **secondary structure**. In the second phase, the planar graph folds into a three-dimensional object which is known as the **spatial structure**.

3.1 Base interactions and other molecular forces

RNA molecules are composed of only four monomers or ribonucleotides which are linked together by covalent binding. Each nucleotide consists of three molecular fragments: sugar, heterocycle and phosphate. The sugar (β-D-ribose) is phosphorylated in 5' position and forms a β-glycosyl-C1'-N bond with one of the four possible heterocycles. These include the purine bases adenine (A) and guanine (G) as well as the pyrimidine derivatives cytosine (C) and uracil (U). Figure 4 shows a short strand of RNA. The phosphate groups link the 5' end of one ribose to the 3' end of the next, thereby imposing directionality on the backbone. The two free ends at the top and the bottom of figure 4 are consequently referred to as 5' and 3' ends, respectively.

RNA usually occurs single-stranded, but complementary base pairing via hydrogen bonding leads to helix formation. Although the standard four bases can be arranged in 28 different ways, helices exist almost exclusively of Watson-Crick type G-C and A-U pairs. Such canonical pairings have anti base-sugar conformations, locally anti-parallel strands and glycosidic bonds oriented cis with respect to each other (for a detailed review on geometric parameters for classifying nucleic acid pairs see [75]. Isostericity of all four pairs is the remarkable characteristic of Watson-Crick base pairing geometry, implying possible substitution for each other without distortion or even disruption of the double helices they are part of. The degeneracy of the genetic code (caused by the fact that 20 amino acids are coded by base triplets of the four nucleotides) gives rise to the so-called "wobble" hypothesis which states that while the first two base pairs formed between mRNA codon and tRNA anticodon are of the standard Watson-Crick type, the

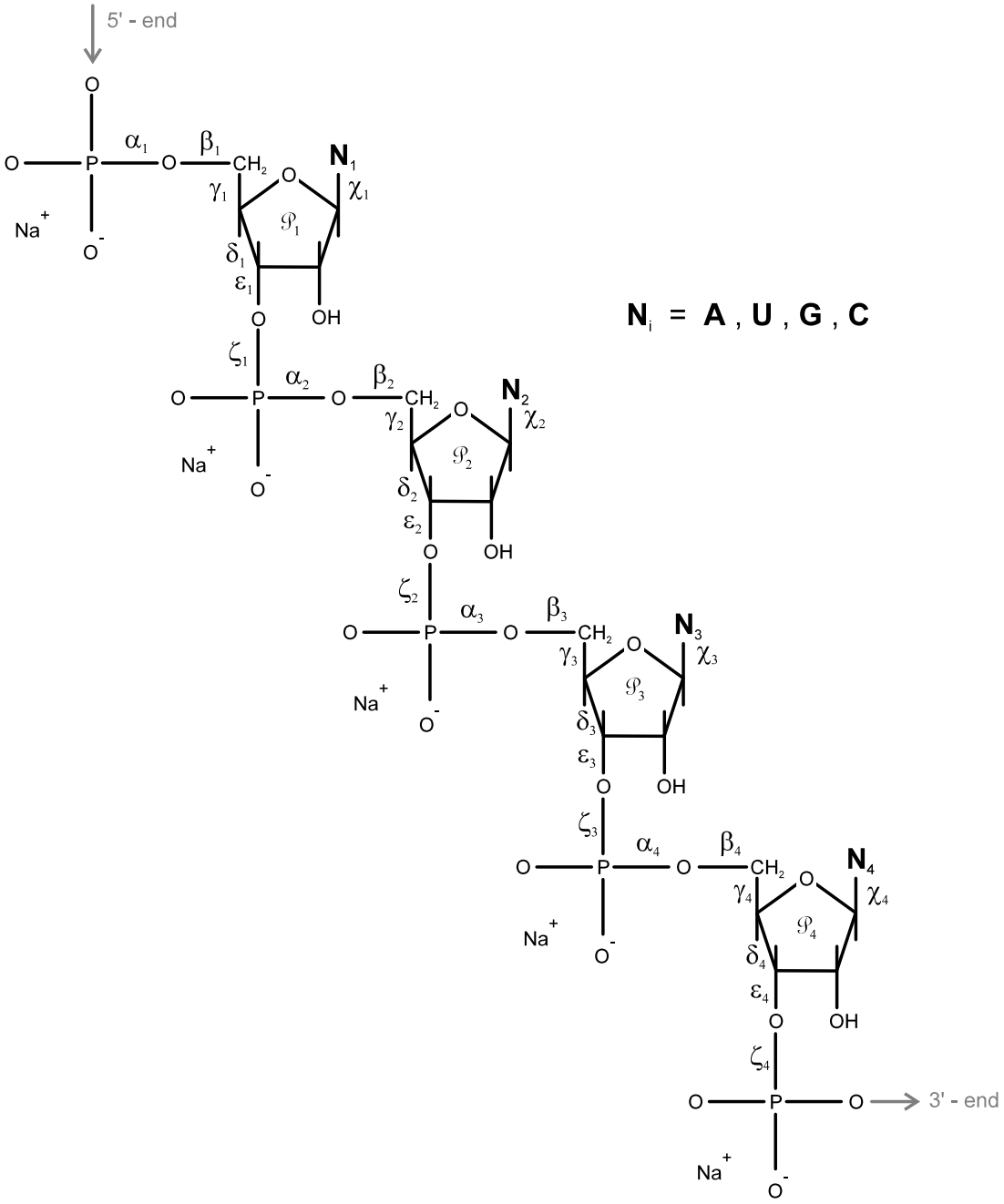


Figure 4: A short strand of RNA.

third may be an unusual pair (for example G-U, I-U, I-A, etc.), due to steric misalignment. The G-U pair which is slightly less stable than Watson-Crick pairs frequently substitutes for canonical pairs without distorting RNA double helices. Base pairing is mediated by mainly electrostatic hydrogen bonding between differently polarized atoms. One has to differentiate between five hydrogen-bonding modes [75]. Apart from standard interactions of the type N-H—N and N-H—O, bifurcated systems where two acceptors compete for the same hydrogen atom have been observed. Furthermore, water molecules may open up and take active part in a base pairing. Another possibility is presented by hydrogen atoms covalently attached to aromatic carbon atoms [3]. Finally, hydrogen bonds involving sugar OH-groups have been found.

Hydrogen bonding dominates in nonpolar solvents, whereas base stacking - the second important base-base interaction - is most pronounced in water where solvent molecules compete for the binding sites and hydrogen bonding is therefore suppressed. Dipol-induced dipole interactions, with the permanent dipole (C=O and C-H–N2 groups) superposed over the conjugated π electron system of the adjacent base, strongly favour vertical base pair stacking.

In addition, the hydrophobic effect plays an important role in aqueous solution. In order to hide the mostly hydrophobic bases from the solvent, base pairs tend to aggregate which leads to helix formation. Dipolar and London dispersion forces also contribute significantly to the stacking process, accounting for the fact that stacking is more pronounced in purines than in pyrimidines, and is additionally enhanced by alkylation.

3.2 Tertiary contacts

In the last few years, a growing number of three-dimensional RNA structures determined by NMR or X-ray cristallography became available (reviewed by [33]), leading to a more profound understanding of the principles governing the architecture of RNA folds. Several of the base interactions discussed in the previous section participate in tertiary motifs and stabilize RNA 3D structures. Such molecular modules are:

1. **Mismatches**

   Base covariation analysis ascertained the probability of non-canonical base pairs [38], a large number of which have actually been discovered in RNA molecules (figure 5). Scattered among Watson-Crick pairings within an

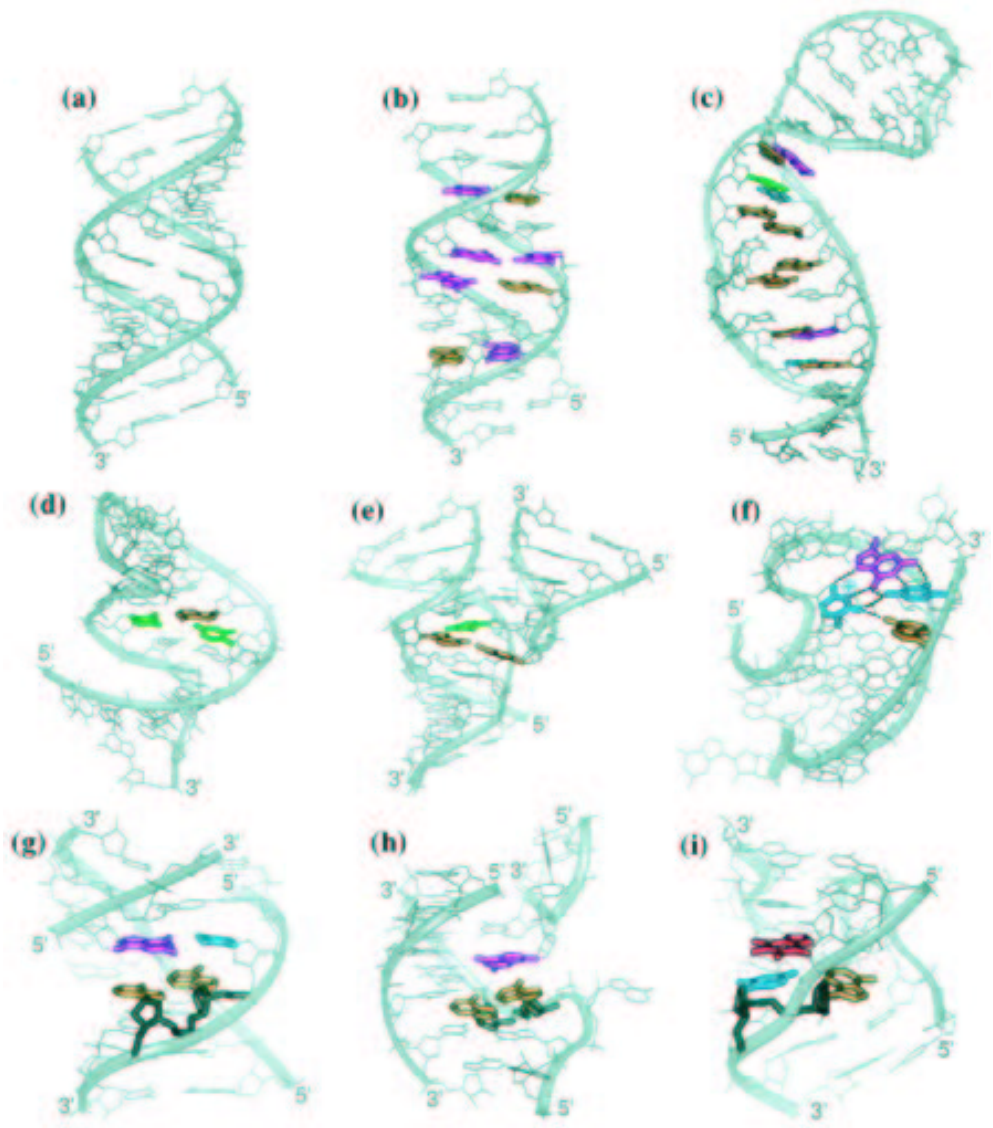


Figure 5: Interactions in the base pair plane [50]. a.) Hydrogen bonding between coplanar bases gives rise to A-form RNA helices. b.) Mismatches alterating with Watson-Crick pairs lead to a distortion of the grooves [20], and c.) participate in stacking interactions as in the loop B of the hairpin ribozyme [12]. d.) Base triples extend grooves for protein interaction in BIV TAR RNA [146] and e.) mediate the tertiary contact between tetraloops and helical loop-receptors, as in the group I ribozymes [15]. f.) The loop of an RNA pseudoknot [131] is stabilized by a base quadruple. g.) Platforms arise through the interaction of consecutive bases within one strand, first discovered in adenosine platforms in group I ribozymes [16], h.) another adenosine platform in the L11-binding domain of 23 S rRNA [19], and i.) consecutive A and C residues bring about an A-C platform in a theophylline aptamer RNA [149].

RNA helix, mismatch pairs may be involved in stacking interactions (figure 5c) and bring about a distortion of the grooves (figure 5b). As the most common among non-canonical pairs, the homopurine G-A pairs are mostly found at the end of helices (examples include the R26-R44 base pair flanking the anticodon stem of some tRNAs).

2. **Base triples and quadruples**
   Both Watson-Crick and non-canonical pairs take part in such motifs. Interaction of a nucleotide in a loop with a base pair in a helix gives rise to a base triplet [17]. In group I ribozymes (figure 5e, [15, 133], a base triple is responsible for the tertiary contact between a tetraloop and its receptor. The same motif also serves as an example of a G·A·C:G quadruple where an additional guanosine is associated by hydrogen bonding.

3. **Platforms**
   Consecutive base pairing within one strand accounts for the formation of platforms, the first of which has been observed for side-by-side paired adenosines in the P4-P6 domain of group I ribozymes (figure 5g, [16]). Monovalent metal ions provide stabilization.

4. **Cross-strand stacking**
   Base stacking, the major contribution to a molecule's overall free energy, usually refers to interactions between consecutive nucleotides within one strand. In several molecules, however, stacking of (mostly purine) bases belonging to different strands has also been discovered. The term "base zipper" motif denotes this form of interdigitation which occurs for example in the junction of D and T loops in tRNA [65, 116] and the loop B of the hairpin ribozyme [13].

5. **Ribose zippers**
   Apart from associating bases on complementary strands, hydrogen bonds can also connect the 2' OH group of a sugar in one strand with both the 2' OH group on the ribose and either the pyrimidine O2 or the purine N3 atom of a nucleotide on the other strand (noticed for example in the hammerhead ribozyme [108]).

All of these described helix motifs are arranged to result in a compact three-dimensional fold. The most common way to reach a closely packed architecture

is by means of end-to-end stacking of double helices (which will be described in detail in section 4.3). The single-stranded stretches connecting the helices can themselves form structural elements. Base pairing between two complementary loop sequences ("kissing loops") is a recurring tertiary interaction in tRNA (between the D and T loops) [111]. A pseudoknot is characterized by base pairing between nucleotides in the loop of a conventional hairpin with a complementary sequence outside that loop. A compact fold results from the coaxial stacking of both double helical segments (for a description of various possible stacking arrangements see [1].
An essential component without which complex RNA folds form secondary structure motifs, but little, if any, tertiary structure, is the presence of divalent cations, especially magnesium. As counter-ions for the negatively charged backbone, they are crucial for the structural integrity and biological function of RNA. While the predominant amount of cations is delocalized and involved in non-specific, electrostatic interactions, a number of them participates in site-specific binding to anionic ligands [71, 98]. For example, the junction connecting the acceptor helix and the D loop in tRNA, where folding of the RNA backbone arranges phosphate groups in close proximity, is sufficiently screened by magnesium ions [53, 112].

3.3 RNA Secondary Structures

A secondary structure consists of a set of vertices
$V = \{1, 2, ..., i, ..., N\}$ and a set of edges $S = \{i \cdot j, 1 \leq i < j \leq N\}$ fulfilling

(1) For $1 \leq i < n$, $i \cdot (i+1) \in S$.

(2) For each i there is at most one $h \neq i-1, i+1$ such that $i \cdot h \in S$.

(3) If $i \cdot j \in S$ and $h \cdot l \in S$ and $i < h < j$, then $i < l < j$.

The first statement simply implies that RNA is a linear polymer, the second condition restricts each base to at most a single pairing partner, and the last condition states that base pairs must not cross, and thus, that knots and pseudoknots are not allowed.
The exclusion of pseudoknots from the secondary structure definition results more from a practical point of view (as they are incompatible with most dynamic programming routines) than from knowledge-based considerations. While pseudoknots have been proven to play an essential role in some molecule's function

(for example in RNase P [25]), little information is yet available on their energy parameters and their integration into the mathematical model could not be accomplished without considerable computational effort. They should therefore rather be viewed as a first step towards prediction of the tertiary structure.
In addition to the three conditions, steric constraints can be specified, allowing for example for a minimum loop size of three nucleotides or excluding isolated base pairs.

A computation of full three-dimensional RNA structures from sequences cannot yet be achieved with desirable reliability. Prediction of RNA structures, however, is feasible on the level of the secondary structure for which a broad range of thermodynamic parameters, most of them having been directly measured in experiments, are available. The principle of hierarchical folding [2, 11] implies that the formation of secondary structure elements is seldom disrupted by the subsequent development of tertiary interactions (an interesting counter-example is discussed in [144]). The concept of the secondary structure as a coarse graining of the full three-dimensional structure provides a convenient theoretical construct and is justified by the following observations:

- The secondary structure of RNA molecules is well defined. In that aspect, RNAs differ from proteins.
- A secondary structure represents a relevant and experimentally verified intermediate in kinetic folding [6], and its formation constitutes the dominant part of the free energy of folding.
- RNA secondary structures are often well conserved in evolutionary phylogeny [44] and have been successfully relied upon for a correct interpretation of RNA function and reactivity.
- Secondary structures do not only provide the major set of distance constraints guiding the formation of the spatial structure, but they are also discrete and therefore easy to handle and compare computationally.

Secondary structures are composed of five distinct loop types as shown in figure 6.
Stacked base pairs lead to the elongation of helices and are one of the major driving forces for secondary structure formation. The hairpin, composed of a double-stranded region (helix or stem) and a connecting single-stranded part

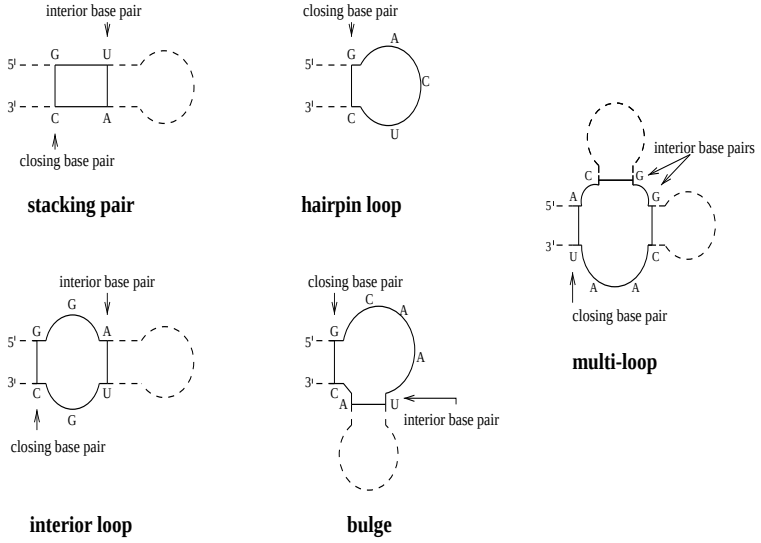


Figure 6: Secondary structure motifs

(loop) is the most prominent of these structural motifs. Other possibilities are the bulge (which has unpaired bases on just one side of the stem), the internal loop (unpaired nucleotides on both sides of the stem), and the multiloop (several stems connected by single-stranded junctions).

Any given position $i \leq k \leq j$ of a given pair (i, j) is called **interior** of (i, j). The notation **directly interior** of (i, j) refers to position k if there is no base pair (p, q) such that $i < p < k < q < j$. The loop closed by (i, j) is then the set of all positions directly interior of (i, j) including (i, j).

A classification of loops can be achieved according to their degree which comprises the number of base pairs within the loop. Hairpin loops therefore belong to the group of degree 1, bulges and interior loops of degree 2, and the category with degree ≥ 3 is composed of multi-loops. The number of unpaired nucleotides in the loop is defined as the loop size. Thus a stacked base pair is considered a loop of size zero.

Secondary structures can be uniquely decomposed in the described loops. With an additional virtual base pair enclosing the whole structure, the graph is equivalent to an ordered rooted tree (see left side of figure 7). An internal node (black) indicates a base pair, a leaf node (white) corresponds to an unpaired nucleotide, and several leafs branching off the same node constitute a loop.

Secondary structure representation

Secondary structures are most commonly displayed as planar graphs with each vertex representing a nucleotide and edges connecting consecutive nucleotides and base pairs (see left side of figure 8).
In case of larger structures, the composition of the molecule is however easier to survey in the so-called "mountain" representation [52]. In a two-dimensional graph, the x-coordinate presents the position k of a nucleotide in the sequence, while the y-coordinate depicts the number m(k) of base pairs that enclose nucleotide k. In other words, peaks are equivalent to hairpins, plateaus point to unpaired bases, and valleys denote unpaired regions separating secondary structure elements.
Finally, secondary structures can be stored in a string consisting of the symbols "(", ")" and ".", representing nucleotides that are paired with a partner towards the 3' end, the 5' end or unpaired, respectively. Matching brackets therefore constitute a base pair.

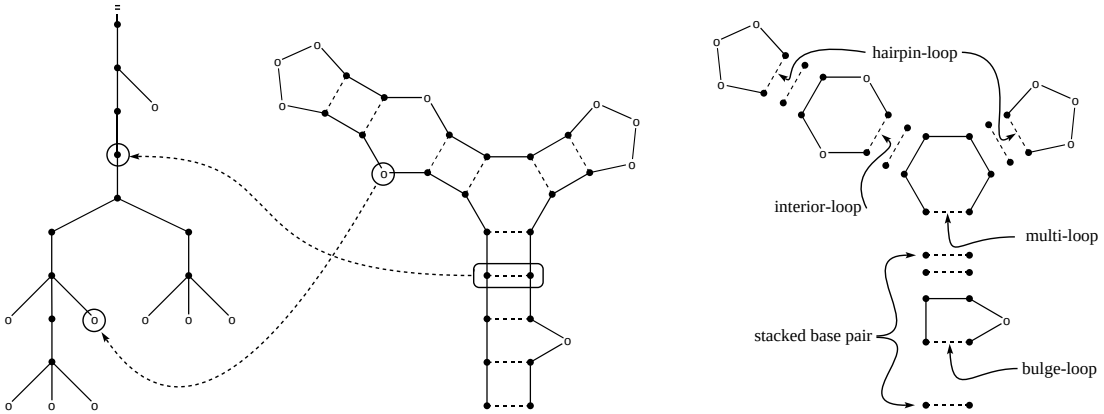


Figure 7: Loop decomposition of secondary structures: the tree representation (l.h.s.) presented as a secondary structure graph (middle) which can be decomposed into the various loop types (closing base pairs are drawn as dotted lines).

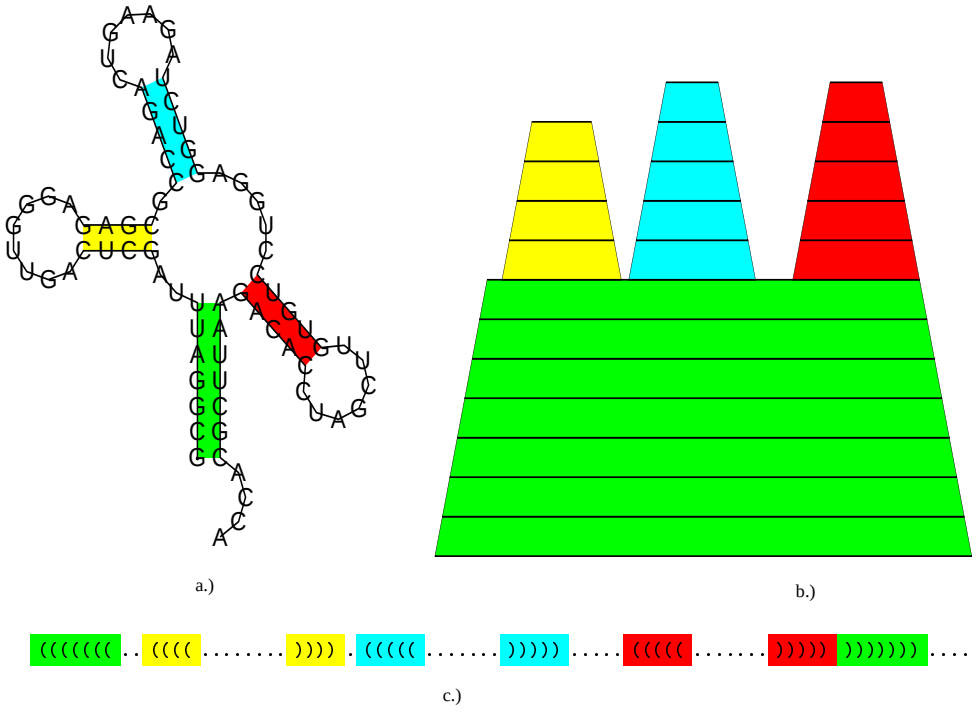


Figure 8: Secondary structure representations: a.) a planar graph, b.) a mountain plot, c.) in the bracket notation.

4 Methods

4.1 Vienna RNA Package

The package contains computer codes based on dynamic programming algorithms to allow and facilitate the correct prediction of secondary structures. The most simple approach to RNA structure prediction yields a single optimal solution, the minimum free energy structure [150]. While it is not too intricate to program and undeniably easy to interpret, a single result cannot come up to the truth, for at room temperature, an RNA molecule fluctuates between different states. A single structure may not be sufficient to characterize the thermodynamic equilibrium ensemble which, while often dominated by the mfe structure, may include widely differing conformations of similar energy. A remedy is presented for example by Zuker's algorithm (1989) which sends additional suboptimal structures to output. For each possible pair the best structure containing that pair will be computed.

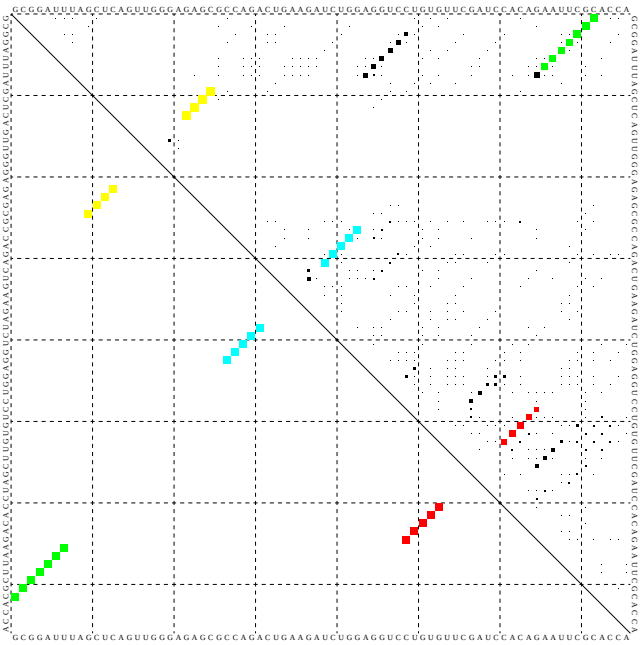


Figure 9: Dot plot of unmodified $tRNA^{phe}$: the lower left triangle displays the minimum free energy structure, while the upper right triangle shows significant base pairing from suboptimal conformations.

An alternative method, McCaskill's algorithm, procures the partition function $Q = \Sigma exp(-\Delta G(S)/kT)$ and equilibrium probabilities for all possible pairs. One obtains the frequency with which each base pair occurs in the Boltzmann weighted ensemble of all structures which can be conveniently represented in a so-called "dot plot". For each base pair (i, j), the equilibrium frequency p is plotted by a square of area p_{ij} in position i, j on a two-dimensional grid (illustrated in figure 9).

As opposed to just maximizing the number of base pairs (concept of earlier model), the standard energy model calculates the energy of a structure as the sum over the energy contributions of its constituent loops. It is thus based on the loop decomposition explained above.

$$E(S) = \sum_{l \in s} E(l) \tag{1}$$

The term "nearest-neighbour" rules has been established, indicating that the energy contributions of a base pair in the midst of a helix depend only on the previous and the following pair. Whereas hairpin and interior loop energies are well tabulated, thermodynamic measurements on multiloops are scarce. The energy function reads:

$$\Sigma G_{\mathrm{ML}} = a + b \cdot n + c \cdot k + \Delta G_{\mathrm{dangle}}, \tag{2}$$

where n denotes the loop size and k the loop degree. $\Delta G_{\mathrm{dangle}}$ is an energy bonus describing the stacking interactions between a pair and one adjacent unpaired base. Dangling ends are thus similar to mismatch energies except that the latter are composed of two terms derived from the unpaired nucletide 5' and 3' of the pair. In case of larger loops, the linear approach is frequently replaced by the logarithmic Jacobson-Stockmayer function so as not to overestimate the energy penalty.
Since a secondary structure consists of many micro states, the result is a free energy that can be split into an enthalpic and an entropic term.

$$\Delta G = \Delta H - T\Delta S \tag{3}$$

The standard set of energy parameters is measured for $T = 37°C$, but can be extrapolated using the temperature dependence of the free energy. The most recent compilation can be found in [89].

The `Vienna RNA Package` consists of several algorithms. As the most basic, `RNAfold` takes an input sequence and calculates its minimum free energy structure, on request also the partition function and the base pair probability matrix. It returns the mfe structure in bracket notation, together with its energy, the free energy of the thermodynamic ensemble and the frequency of this particular structure in the ensemble. In additon, it generates two Postscript files showing the respective secondary structure graph as well as the dot plot of the base pairing matrix.
True to its name, `RNAinverse` represents an inversion of the folding algorithm; it thus searches for sequences folding into specified structures. It also yields the Hamming distance to the start sequence (if none is specified, a random sequence will be used as starting point). The Hamming distance corresponds to the minimal number of point mutations required to convert two sequences into each other and serves as a metric in the abstract sequence space. A very useful application is the design of sequences whose thermodynamic ensemble is dominated by the target structure. The possible input of sequence constraints is an additional feature which may shift the result in the desired direction.
`RNAsubopt` generates **a**ll suboptimal conformations within a user-defined energy range above the mfe. It therefore gives a practical overview of structure possibilities in case of short sequences. For larger molecules, the information becomes difficult to survey, as a tRNA sequence for example gives rise to more than a million structures within 15 kcal of the mfe. The output may, however, be further utilized for a thorough exploration of the energy landscape (see following section).
These algorithms and other programs for the computation and comparison of RNA secondary structures are included in the `Vienna RNA Package` which is freely available from `http://www.tbi.univie.ac.at/RNA`.

4.2 Kinfold and Barriers

Structure prediction based on thermodynamic parameters has been described in the previous section. This chapter, on the other hand, focuses on the dynamics of RNA folding on an energy landscape. A molecule's energy landscape can be described as a hyper-dimensional plot of the free energy versus the positions of its atoms. Due to the considerable number of degrees of freedom in a non-linear molecule (3N-6, with N stating the number of atoms), this concept gives rise to

highly complex landscapes in case of nucleic acids. A basic requirement for a detailed analysis of an RNA energy landscape is a compilation of all suboptimal structures within a predefined energy range as well as a measure of structural similarity between two conformations. A set of rules determining which operation is allowed in conformation space is called a *move set*. It lays down the topology of the energy landscape by defining which structures are neighbours of each other. The resulting folding trajectories get more realistic the smaller structural changes in a single step are allowed.
The most elementary move set on the level of RNA secondary structure consists of the addition or removal of a single base pair. These moves can be easily followed by means of the circular representation of RNA structures (see figure 10 moves B and C). The closing of another base pair is displayed by an additional chord in the diagram. With the help of just these two moves, a folding path between every arbitrarily chosen pair of structures can be constructed. The move set induces a metric onto conformation space. To permit and promote rapid chain sliding along a helix, a third move was included in the move set. The "shift" move combines base pair insertion and deletion in a way that one of the two positions of a given base pair is converted into a new one. This behaviour is demonstrated in figure 10 moves D and E with a displacement of a chord while one end stays fixed.
Complete suboptimal folding can be accomplished with the RNAsubopt program described in the previous section.

With the two requirements - a move set and a list of suboptimal conformations - available, a more thorough investigation of the energy landscape is within reach. Due to pairing rules and high stability of RNA secondary structures, the folding landscape is rugged, composed of many deep local minima in which the folding process may come to a temporary standstill.

In a short summary, the used algorithm simulates folding kinetics while based on a continuous time Monte Carlo method and passes repeatedly through the following phases:

1. Generation of all neighbours using the move set
2. Calculation of transition probabilities for each step ($P_i = \exp(-\Delta E/2kT)$)
3. Selection of a move with probability to its rate

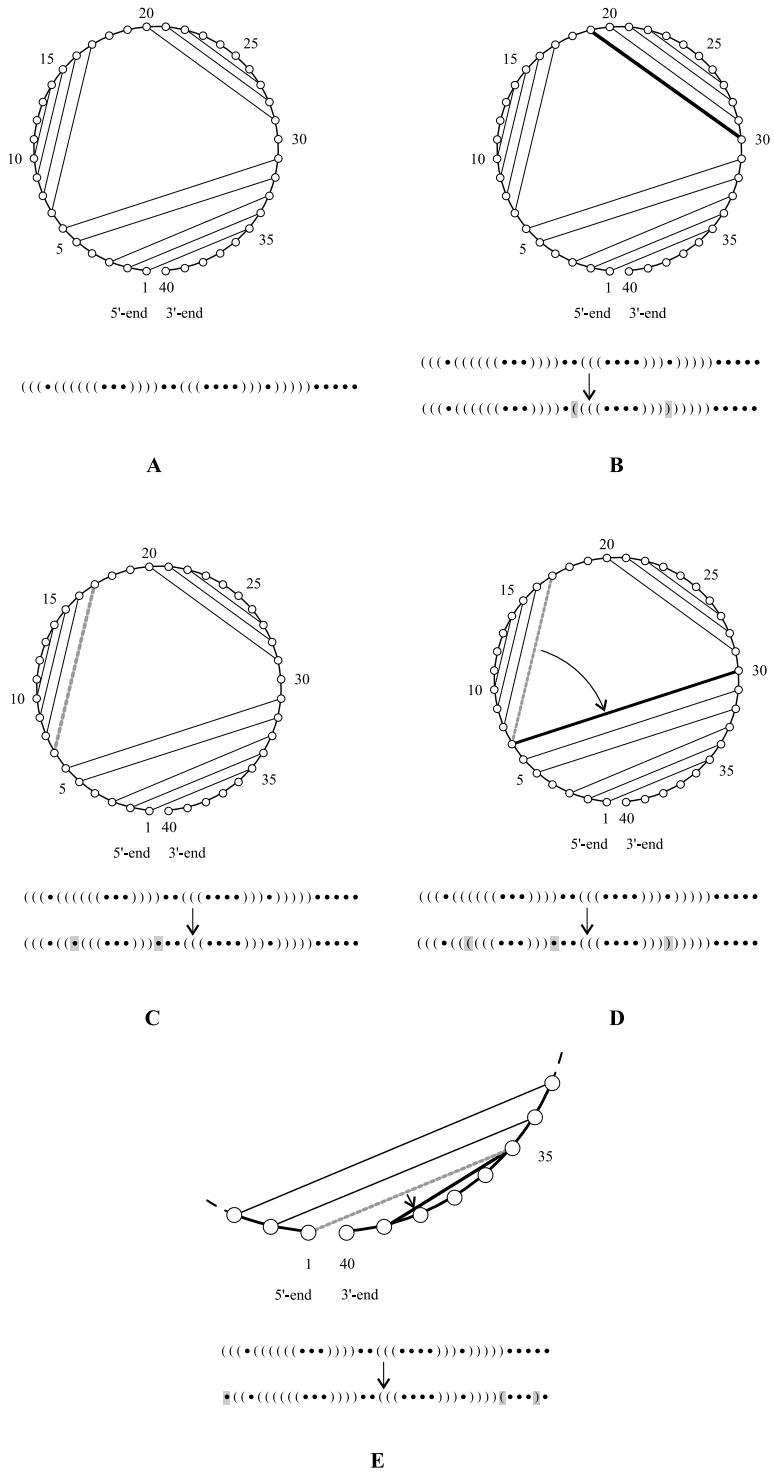


Figure 10: Circle representation of secondary structure and explanation of the move set : a.) secopndary structure, b.) insertion of a base pair, c.) deletion of a base pair, d.) shift move upstream, e.) shift move downstream.

Although founded upon the same model, the two folding algorithms Kinfold and Barriers tackle different tasks.
Kinfold makes a statement about the folding dynamics, most importantly about the folding time, which in this context refers to the first passage time from some initial state to the ground state. It takes an input sequence and simulates the folding starting from the open chain or a user-defined start structure. In addition, the stop structure can be specified if it differs from the minimum free energy structure. This way, the algorithm presents the possibility of refolding, for example for computing the folding time for a molecule to reach the mfe structure after it got trapped in a meta-stable state.

To gather further information on the influence of such kinetic traps on the folding pathway, Barriers is used. It requires the complete range of suboptimal conformations sorted after energy (an RNAsubopt output file) as input and generates a list of all local optima, together with their energies, barrier heights and basin widths. The decisive factor for the importance of such a basin of attraction is the barrier height. If a molecule encounters a local minimum with a high barrier on its way to the mfe structure, the chance of getting trapped is increased, resulting in at least a longer folding time. Sometimes the escape and the continuation to the ground state do not lie within the time-scale of the simulation. The distribution of energy barriers is highly sequence dependent.
In order to help visualize the folding path, a tree of local minima is computed which contains the global minimum, all local minima as well as the saddle points connecting them on the folding landscape. Energy barriers divide the resulting tree of local minima into several subtrees, each of them dominated by the respective conformation with the lowest energy.
Figure 11 illustrates the partition of the conformation space for an unmodified sequence: four different folding funnels are visible, leading to conformations S1, S7, S2 and S5, respectively. Although each of these subtrees in turn split up into several basins, their conformations are of minor significance in the folding pathway. With a barrier below or around five kcal/mole, an energy valley is not deep enough to hinder a molecule that has fallen into it from reaching the next saddle point within a reasonable period of time. From there, a descent into another local minimum and, finally, to the structure lying topmost of the particular subtree, is possible. It is those structures that represent (partly long-lived) intermediates in the folding process.

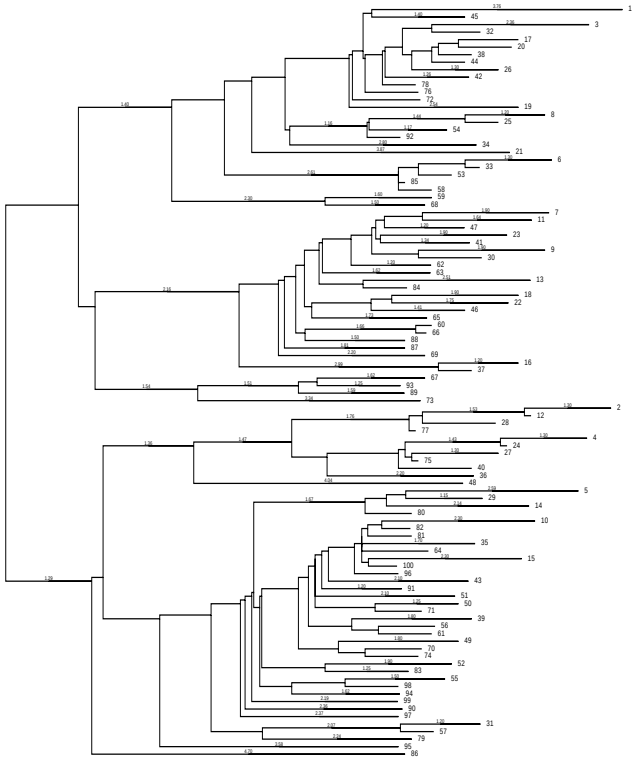


Figure 11: The tree of local minima as leaves and the transition states as internal nodes; the height of the energy barriers is represented by the branch lengths.

4.3 Kinetic folding with coaxial stacking

Multi-branched loops are a commonly distributed feature of RNA secondary structures. Their energy contributions have been described above, but no consideration has yet been given to another important characteristic of these motifs, and that is the possibility of coaxial stacking. It concerns a favourable interaction between two helices that are either directly adjacent or separated by at most one intervening mismatch. As a result of the strength of base pairing and stacking interactions, such two helices tend to stack end on end, forming a continuous helix. Coaxial stacking usually induces an enhancement of stability. For adjacent helices without a mismatch the energy bonus corresponds to the free energy parameter of a normal stacked pair. Coaxial stacking can, however, also take place in the case of one intervening base between the helices if there is another nucleotide to form a mismatch with (either 5' to the 5' helix or 3' to the 3' helix). The energy is then composed of the parameter for terminal mismatches combined

with a sequence independent value (-2.1 kcal/mole). The used program takes into account coaxial stacking by reevaluating the energy of structures by means of a more complex multiloop energy function (for $n \geq 6$)

$$\Delta G_{ML} = a + b \cdot 6 + 1.75RTln(n/6) + c \cdot k + \Delta G_{Stacking} \tag{4}$$

(a = 10.1;b = -0.3; c = -0.3 kcal/mole; parameters compiled in [89]).
$\Delta G_{stacking}$ includes the favourable free energy of coaxial stacking and terminal mismatch or dangling end stacking as described above.
Coaxial stacking leads to a relevant improvement of the accuracy of secondary structure prediction. For molecules with multi-branched loops, like tRNAs, the effect of increased stability will be most apparent. As shown in figure 12, $tRNA^{phe}$ folds neatly into two coaxial helical stacks meeting at right angles and characterizing the well-known L-shape.

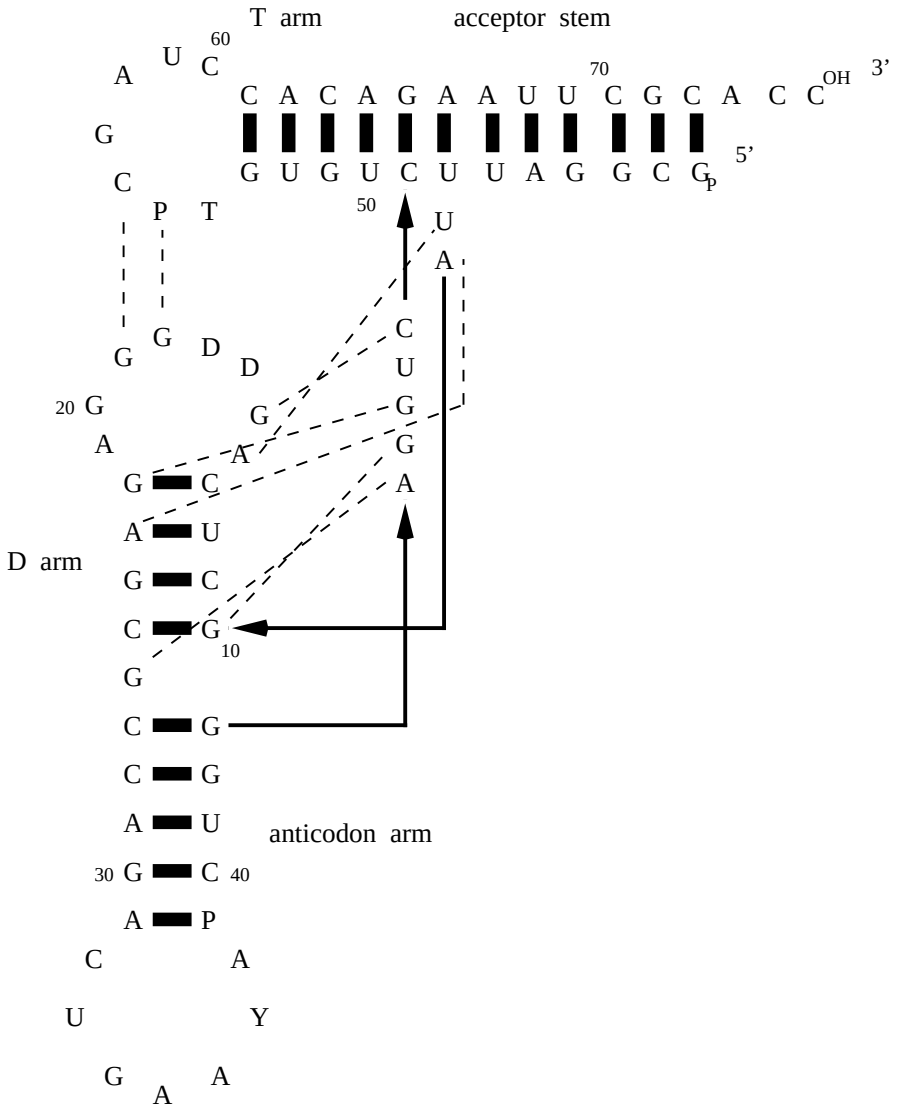


Figure 12: Secondary structure of $tRNA^{phe}$ building two coaxial stacks. Main-chain crossovers are displayed by thick lines, tertiary interactions by thin dotted lines.

5 Comparison of natural tRNAs

All the natural tRNAs investigated in the course of this thesis were downloaded from the aligned sequence database *COMPILATION OF tRNA SEQUENCES AND SEQUENCES OF tRNA GENES* located in Bayreuth [127]. The major amount of tRNA sequences were isolated from genes (3704), only a small part of the available ones were directly sequenced (550). Out of the latter, only Eubacteria were selected for analysis, the restriction necessarily arising from the rather time-consuming calculations. Apart from that, previous results have shown that most eucaryotes, especially mitochondria, show a peculiar folding behaviour and do not possess the typical cloverleaf structure as ground state, therefore the examination was confined to eubacterial sequences (list in the appendix). They were classified according to the twenty amino acids. Results in the following section are discussed and compared within the different amino acid families as well as in a certain organism in case of remarkable findings. All sequences contain several modified nucleotides (especially methylations) , one of the characteristics of tRNAs in contrast to other types of RNA. Since up until now, no experimental data for the energy contributions of the different kinds of modified bases have been available, all algorithms constituting the Vienna RNA Package treat a nucleotide other than A/C/G/U as a base that is not allowed to pair, although the stacking energy remains the same as for the unmodified base. Methylations were marked with an M, other or unknown modifications with an N. The essence of base modification never becomes more apparent then when one tries to draw comparison to the respective unmodified sequences. The influence of modified nucleosides on both structure and function has been facilitated by means of in vitro transcription methodology and solid-state synthesis which permit the synthesis of tRNAs lacking base modifications [46, 122]. Primarily employed to investigate its interaction within the translation machinery, unmodified tRNAs often retain their ability to bind to their cognate synthetase [4, 8, 56, 99, 107, 123] and react with both elongation factor TU [47] and tRNA-modifying enzymes [23]. The ease with which they fulfill their roles in the described processes might indicate a similarity in the gross strcuctures of unmodified and native tRNAs. To study the different folding behaviour and the effect of base modifications, a second set of these tRNAs were employed where the modified nucleotides were substituted for their standard counterparts.

5.1 Natural modified sequences

In order to make an assertion about the thermodynamic stability of a particular sequence, it is essential to gain knowledge about the suboptimal conformations in a certain energy range. The RNAsubopt program generates all suboptimal structures in a user-defined energy range above the minimum free energy. For small molecules, it can be rewarding to study the complete list of structure probabilities. However, for a molecule the size of a tRNA, the information (quickly) becomes difficult to survey, as it can comprise more than a million structures within about 10 kcal/mole of the mfe structure.

The following calculations were generally carried out using an energy range of $\Delta\epsilon = 15$ kcal/mole above the mfe. Only in the case of class II tRNAs (leu, ser, tyr), the limit was reduced to 13 kcal/mole in order to account for their increased length and number of structure possibilities. As most natural tRNAs have a minimum free energy around -25 kcal/mole, the open chain does not lie within the predefined energy interval. Table 2 presents a list of the mfe and the energy difference between mfe and first suboptimal conformation. The value of the minimum free energy ranges from about -17 kcal/mole to -29 kcal/mole, reaching up to -37 kcal/mole for class II tRNAs, depending more on the length of the sequence and the ratio of G-C to A-U or G-U pairs than the number of modifications. The serine sequence RS1661 turns out to be the thermodynamically most stable one (-37.28 kcal/mole), which does not come as a surprise considering its exceptional length of 93 nucleotides. In contrast to most other tRNAs which tend in the direction of 76 bases (forming around 22 base pairs in the cloverleaf structure), this particular sequence displays 28 pairs due to increased pairing probability in the variable region. In addition to the mere number of base pairs, their composition is of great importance. Drawing for example comparison between the two isoleucine sequences RI1180 and RI1580 reveals that, regardless of their identical length (77) and number of base pairs (21) as well as modifications (7), the difference between their minimum free energies amounts to almost 10 kcal/mole. In absence of other decisive factors, this result can only be explained in terms of base pair composition. As has been described in the section 4, G-C pairs provide a larger contribution to thermodynamic stability than A-U or even G-U pairs. Five more G-C and the lack of any G-U pairs are responsible for the low energy of the ground state of RI1580.

The energy gap between the mfe structure and the first suboptimal state varies

tRNA	mfe [kcal/mol]	ΔE [kcal/mol]	tRNA	mfe [kcal/mol]	ΔE [kcal/mol]
RA1140	-27.56	-0.96	RG1661	-28.26	-0.90
RA1180	-28.90	-0.70	RG1662	-26.21	-0.01
RA1540	-28.76	-0.84	RH1140	-21.70	-0.20
RA1660	-26.16	-0.40	RH1660	-21.86	-1.14
RA1661	-27.20	-0.18	RH1700	-21.86	-1.14
RA1662	-26.86	-0.06	RI1140	-23.56	-1.24
RR1140	-20.90	-0.13	RI1141	-18.16	-0.74
RR1141	-23.36	-0.55	RI1180	-18.06	-0.74
RR1540	-25.56	-0.84	RI1580	-27.96	-0.64
RR1660	-22.96	-1.04	RI1660	-20.72	-0.79
RR1661	-23.96	-1.04	RI1661	-20.72	-0.79
RR1662	-22.80	-0.40	RI1662	-28.67	-0.82
RR1663	-24.48	-0.14	RI1540	-24.26	-1.14
RR1664	-25.12	-0.90	RL1140	-21.66	-0.14
RN1140	-24.27	-1.21	RL2020	-21.60	-0.05
RN1660	-17.86	-1.42	RL2100	-27.78	-0.78
RN1720	-20.57	-0.50	RL2101	-27.46	-0.58
RN1721	-19.50	-0.50	RL1141	-23.75	-0.07
RD1140	-27.46	-0.76	RL1142	-23.41	-0.32
RD1580	-21.60	-0.20	RL1460	-28.28	-0.41
RD1660	-20.32	-0.22	RL1540	-27.54	-0.00
RC1140	-20.96	-0.50	RL1660	-30.38	-0.93
RC1660	-20.42	-0.22	RL1661	-24.90	-0.41
RQ1140	-21.06	-0.36	RL1662	-30.48	-1.60
RQ1660	-21.16	-0.86	RL1700	-24.90	-0.41
RQ1660	-23.36	-0.90	RK1140	-17.46	-0.02
RE1140	-22.89	-0.29	RK1141	-26.96	-0.26
RE1660	-29.54	-0.33	RK1540	-20.36	-0.64
RE1661	-27.74	-0.34	RK1541	-20.36	-0.64
RE1662	-27.85	-0.35	RK1660	-22.27	-0.91
RE2140	-28.40	-0.50	RM1140	-25.86	-0.70
RG1140	-21.76	-0.76	RM1540	-25.42	-0.25
RG1700	-30.30	-0.60	RM1580	-28.36	-0.10
RG1701	-26.21	-0.01	RM1660	-18.59	-1.48
RG1180	-21.76	-0.30	RF1140	-23.36	-1.06
RG1310	-27.20	-0.90	RF1460	-22.26	-1.36
RG1380	-20.03	-0.13	RF1540	-23.99	-0.13
RG1381	-22.40	-0.57	RF1580	-20.22	-0.27
RG1540	-19.10	-0.64	RF1660	-25.76	-1.60
RG1660	-27.80	-0.54	RF2020	-26.06	-1.04

Table 2: Modified natural tRNAs: minimum free energy and energy gap between the mfe and the first suboptimal conformation.

tRNA	mfe [kcal/mol]	ΔE [kcal/mol]	tRNA	mfe [kcal/mol]	ΔE [kcal/mol]
RF2060	-18.36	-1.01	RT1660	-26.96	-0.90
RP1140	-24.76	-1.10	RT1661	-25.70	-0.40
RP1180	-28.90	-0.60	RW1140	-20.23	-0.86
RP1540	-21.06	-0.66	RW1141	-21.56	-1.00
RP1700	-22.06	-0.30	RW1250	-20.06	-1.00
RP1701	-24.06	-0.84	RW1251	-18.80	-0.05
RP1702	-22.71	-0.59	RW1540	-18.60	-0.34
RS1140	-27.28	-0.03	RW1660	-22.86	-0.54
RS1663	-31.64	-0.06	RY1140	-30.76	-0.21
RS1664	-29.14	-0.86	RY1460	-28.77	-0.70
RS1141	-27.70	-0.01	RY1540	-28.30	-0.00
RS1180	-32.09	-0.39	RY1541	-28.30	-0.00
RS1540	-28.80	-0.64	RY1660	-25.54	-1.42
RS1541	-34.40	-0.10	RY1661	-25.54	-1.42
RS1542	-27.68	-0.70	RV1140	-23.17	-0.80
RS1660	-35.64	-0.06	RV1180	-25.10	-0.60
RS1661	-37.28	-0.06	RV1460	-25.16	-0.94
RS1662	-31.74	-0.06	RV1540	-25.06	-0.64
RT1140	-19.07	-0.57	RV1660	-27.26	-0.84
RT1141	-20.20	-0.84	RV1661	-23.26	-1.04
RT1180	-19.80	-0.84	RV1662	-26.66	-1.04
RT1540	-20.32	-0.01			

table 2 continued

considerably, within a certain amino acid as well as in a specific organism. Neither is there a correlation between the value and the ability to form the cloverleaf as minimum free energy structure.

Only three sequences show degenerate minimum free energies (RL1540, RY1540, RY1541). They all belong to class II tRNAs whose lengths give rise to an extraordinarily high amount of suboptimal conformations. The sequence with the largest gap (1.60 kcal/mole for RL1662) also falls into this class, while the smallest possible difference (0.01 kcal/mole for RG1701, RG1662, RT1540) occurs several times in the other category.

The cloverleaf secondary structure model first established by Holley et al. [54] over 35 years ago, has proven to be valid and common to all tRNAs discovered since [127]. The thermodynamic calculations carried out are compatible with the proposed motif. The results are presented in table 3. Out of 123 test sequences,

tRNA	Acceptor	D	anticodon	VR	T	Acceptor	rank
RA1140	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	mfe
RA1180	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	7.sub
RA1540	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	mfe
RA1660	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	mfe
RA1661	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	1.sub
RA1662	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	mfe
RR1140	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	mfe
RR1141	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	mfe
RR1540	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	mfe
RR1660	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	mfe
RR1661	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	mfe
RR1662	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	mfe
RR1663	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	mfe
RR1664	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	mfe
RN1140	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	mfe
RN1660	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	mfe
RN1720	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	mfe
RN1721	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	3.sub
RD1140	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	mfe
RD1580	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	3.sub
RD1660	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	mfe
RC1140	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	9.sub
RC1660	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	mfe
RQ1140	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	32.sub
RQ1660	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	mfe
RQ1661	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	mfe
RE1140	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	8.sub
RE1660	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	mfe
RE1661	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	mfe
RE1662	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	mfe
RE2140	(((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	7.sub

Table 3: Cloverleaves of modified tRNAs

tRNA	Acceptor	D	anticodon	VR	T	Acceptor	rank
RG1140	((((((.	((((.....)))).	((((.....))))	....	((((.....))))	)))))))....	mfe
RG1700	((((((.	(((.....))).	((((.....))))	....	((((.....))))	)))))))....	11.sub
RG1701	((((((.	(.....)).....	((((.....))))	....	((((.....))))	)))))))....	28.sub
RG1180	((((((.	((((.....)))).	((((.....))))	....	((((.....))))	)))))))....	mfe
RG1310	((((((.	(((.....))).	((((.....))))	....	((((.....))))	)))))))....	3.sub
RG1380	((((((..	((.....)))..	((((.....))))	....	((((.....))))	)))))))....	16.sub
RG1381	((((((.	(((.....))).	((((.....))))	....	((((.....))))	)))))))....	4.sub
RG1540	((((((.	(((.....))).	((((.....))))	....	((((.....))))	)))))))....	mfe
RG1660	((((((.	(((.....))).	((((.....))))	....	((((.....))))	)))))))....	mfe
RG1661	((((((.	(((.....))).	((((.....))))	....	((((.....))))	)))))))....	mfe
RG1662	((((((.	(.....)).....	((((.....))))	....	((((.....))))	)))))))....	28.sub
RH1140	((((((((	((((.....)))).	(((.....)))	....	((((.....))))	)))))))....	29.sub
RH1660	(((((((.	(((.....))).	(((.....)))	....	((((.....))))	)))))))....	mfe
RH1700	(((((((.	(((.....))).	(((.....)))	....	((((.....))))	)))))))....	mfe
RI1140	((((((.	(((.....))).	((((.....))))	....	((((.....))))	)))))))....	mfe
RI1141	((((((.	(((.....))).	((((.....))))	....	((((.....))))	)))))))....	mfe
RI1180	((((((.	(((.....))).	((((.....))))	....	((((.....))))	)))))))....	mfe
RI1540	((((((.	(((.....))).	((((.....))))	....	((((.....))))	)))))))....	mfe
RI1580	((((((.	(((.....))).	((((.....))))	....	(((.....))).	)))))))....	mfe
RI1660	((((((.	(((.....))))	(((((.....)))))	....	((((.....))))	)))))))....	mfe
RI1661	((((((.	(((.....))).	((((.....))))	....	((((.....))))	)))))))....	mfe
RI1662	((((((.	(.....)).....	((((.....))))	....	((((((.....))))))	)))))))....	mfe
RK1140	(((((....	((.....)))).	((((.....))))	....	(((.....))))	)))))))....	6.sub
RK1141	(((((((	((((.....)))).	(((.....)))	....	((((.....))))	)))))))....	mfe
RK1540	((((((.	(((.....))).	(((.....)))	....	((((.....))))	)))))))....	mfe
RK1541	((((((.	(((.....))))	(((((.....)))))	....	((((.....))))	)))))))....	mfe
RK1660	((((((.	(((.....))).	((((.....))))	....	((((.....))))	)))))))....	mfe
RM1140	((((((.	(((.....))).	((((.....))))	....	((((....))))))..	)))))))....	mfe
RM1540	((((((.	(((.....))).	((((.....))))	....	((((.....))))	)))))).....	mfe
RM1580	.(((((..	(((.....))))..	(.....))	...	((((((..............))))))	)))))))....	1.sub
RM1660	(((((..	(((.....))).	(((.....)))	....	((((.....))))	)))))))....	1.sub
RF1140	((((((.						mfe

table 3 continued

tRNA	Acceptor	D	anticodon	VR	T	Acceptor	rank
RF1460	((((((.·	(((.·····.)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RF1540	((((((.·	(((.·····.)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RF1580	((((.···	(((.·····.)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RF1660	((((((.·	(((.·····.)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RF2020	((((((.·	(((.·····.)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RP1140	((((((.·	(((.·····.)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RP1180	((((((.·	(((.·····.)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RP1540	((((((.·	(((.·····.)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RP1700	((((((.·	(((.·····.)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RP1701	((((((.·	(((.·····.)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RP1702	((((((.·	(((.·····.)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RT1140	((((((.·	(((.·····.)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RT1141	((((((.·	(((.·····.)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RT1180	((((((.·	(((.·····.)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RT1540	(((((.··	(((.·····.)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RT1660	((((((.·	(((.·····.)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RT1661	((((((.·	(((.·····.)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RW1140	((((((.·	(((.·····.)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RW1141	((((((.·	(((.·····.)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RW1250	((((((.·	(((.·····.)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RW1251	((((((.·	(((.·····.)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RW1540	((((((.·	(((.·····.)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RW1660	((((((.·	(((.·····.)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RV1140	((((((.	((((.·····)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RV1180	((((((.	(((.·····.)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RV1460	((((((.·	(((.·····.)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RV1540	((((((.·	(((.·····.)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RV1660	((((((.·	(((.·····.)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RV1661	((((((.·	(((.·····.)))).	(((.·····)))	·····	((((.·····))))	)))))))....	mfe
RV1662	((((((.·						mfe

table 3 continued

tRNA	Acceptor	D	anticodon	VR	T	Acceptor	rank
RL1140	((((((	((.....)))..	((...)))	((((...))))	(((((.....)))))	)))))))	68.sub
RL2020	((((((	((.....)))	(((.....)))))	((((...))))..	(((((.....)))))	)))))))	14.sub
RL2100	((((((	((.....)))).	(((.....))))	((((...)))).	(((((.....)))))	)))))))	mfe
RL2101	((((((	((.....)))	(((.....)))))	((((.....))))	(((((.....)))))	)))))))	1.sub
RL1141	((((((	((.....))))..	((.....)))	((((..))))	(((((.....)))))	)))))))	1.sub
RL1142	((((((	((.....)))	(((.....)))))	(((..))))..	(((((.....)))))	)))))))	12.sub
RL1460	((((((	((.....)))).	(((.....))))	((((...)))).	(((((.....)))))	)))))))	mfe
RL1540	((((((	((.....)))).	(((.....))))	((((...)))).	(((((.....)))))	)))))))	mfe
RL1660	((((((	((.....)))	((.....)))	(((...))))..	((((.....))))	)))))))	18.sub
RL1661	(((	(((((.....)))))	(((.....)))))	(((...))))..	(((((.....)))))	)))))))	mfe
RL1662	((((((	((.....)))	((.....)))	(((...))))..	((((.....))))	)))))))	240.sub
RL1700	(((	(((((.....)))))	(((((.....)))))..	(((..))))..	(((((.....)))))	)))))))	mfe
RS1140	((((((	((.....)))	(((.....))))	(((...)))).	(((((.....)))))	)))))))	240.sub
RS1663	((((((	((.....))).	((((.....)))).	(((...)))).	(((((.....)))))	)))))))	1.sub
RS1664	((((((	((.....))).	((((.....))))...	(((...))))..	(((((.....)))))	)))))))	53.sub
RS1141	((((((	((.....))).	((((.....))))...	(((...))))..	(((((.....)))))	)))))))	mfe
RS1180	((((((	((.....))).	((((.....))))..	(((.....))))..	(((((.....)))))	)))))))	3.sub
RS1540	((((((	((.....))))..	(((.....))))	((((((...))))))	(((((.....)))))	)))))))	4.sub
RS1541	((((((	((.....)))	(((.....)))))	(((((.....)))))).	(((((.....)))))	)))))))	mfe
RS1542	((((((	((.....))).	((((.....)))).	(((((...)))))).	(((((.....)))))	)))))))	21.sub
RS1660	((((((	((.....)))	(((((.....))))))	((((((....)))))).	(((((.....)))))	)))))))	1.sub
RS1661	((((((	((.....)))	(((.....))))	((((...)))).	(((((.....)))))	)))))))	mfe
RS1662	((((((	((.....)))	(((((.....)))))	(...)...	(((((.....)))))	)))))))	mfe
RY1140	((((((	((.....)))	(((.....)))))	((..)))...	(((((.....)))))	)))))))	85.sub
RY1460	((((((	((.....)))	(((.....)))))	((..)))...	(((((.....)))))	)))))))	1108.sub
RY1540	((((((	((.....)))	(((.....)))))	((..)))...	(((((.....)))))	)))))))	26.sub
RY1541	((((((	((.....)))	(((.....)))))	((..)))...	(((((.....)))))	)))))))	2.sub
RY1660	((((((	(((.....))))	(((.....)))).	((..)))...	(((((.....)))))	)))))))	2.sub
RY1661	((((((	(((.....))))	(((.....)))).	((..)))...			mfe
							mfe

table 3 continued

78 show the predicted cloverleaf fold as mimimum free energy structure. Furthermore, the first suboptimal structure also presents a slightly different cloverleaf for as much as 59 of them. In almost all of these cases, the main difference between the two cloverleaves lies in the closing of an additional base pair which accounts for a stabilizing contribution. While the Tψ and the anticodon helices are also among the possible areas of change, the preferred elongation takes place at the acceptor stem. The insertion of a base pair closest to the multiloop is favoured, whereas the first base pair is only sporadically involved. This finding is in agreement with the assumption that this stem forms last in kinetic simulations. Results concerning the structure with an opened terminal stack show that its energy is usually about one half to one third of the cloverleaf minimum free energy. This is consistent with the opinion that the inner hairpins form first and a barrier height of such magnitude has to be overcome to close the decisive multiloop [35].

Among the amino acid families, only ile and val show uniform behaviour. All of their representative sequences have cloverleaves as ground state and first suboptimal conformation. The free energy gap reaches exceptional values (1.24 kcal/mole for RI1140), its minimum still being 0.64 kcal/mole (RI1580) and 0.60 kcal/mole (RV1180), respectively. Out of the seven sequences of $tRNA^{phe}$ which are often cited as examples folding exceptionally well into the cloverleaf, two (RF1540 and RF2060) do not possess the native state as minimum free energy structure. The latter is in fact incapable of forming the cloverleaf due to a C-A mismatch interrupting the acceptor stem. Comparing results within a particular organism turns out to be difficult as in very few cases, the primary sequence has been determined for each amino acid family. *E. coli* present the only organism of which an inspection is reasonable since it has representatives in each amino acid group. A thorough examination of all 43 *E. coli* sequences revealed that apart from a handful of sequences, the cloverleaf characterizes the thermodynamically most stable conformation. Only seven exceptions can be found (RA1661, RM1660, RR1663, RG1662, RL1661, RS1662, RS1663), the first two of which have the natural fold as first suboptimal conformation. RM1660 is remarkable for the fact that its minimum free energy structure lacks the entire anticodon stem and shows an extensive Tψ loop (consisting of 15 nucleotides) in its place. The anticodon arm is, however, established in the lowest-lying suboptimal structure, albeit composed of only two base pairs. The energy difference between those two

states is accordingly rather high (1.48 kcal/mole). Mind though, that while *E. coli* sequences preferably fold into **a** cloverleaf structure, the variations in stem and loop sizes are manifold.

The classical cloverleaf model as depicted in figure 1 has the following stem lengths: a seven base pair acceptor stem, five base pairs in the anticodon as well as the Tψ helix, and a D stem consisting of three or four pairs (depending on the class of tRNAs). The following table indicates the percentage of occurrence of these conventional stem lengths in the computation. Out of 122 test objects, the 7 base pair acceptor stem is formed by the vast majority of 107 natural sequences. Other possible stem lengths are 4, 5, 6 and 8 base pairs, which only occur in a few cases. The cloverleaf of 59 sequences consists of a 5 base pair anticodon, but there is also a considerable number with the stem shortened by one base pair (32-fold). On the other hand, an additional base pair is formed in 16 cases. Other possiblities constitute stem sizes of 2, 3 or 7 base pairs which are rare. The proposed Tψ stem of seven base pairs is obtained by the overwhelming amount of 109 sequences, with only a few exceptions each building a helix of 3, 4 and 6 pairs. The result is not as clear for the D stem: for 76 sequences, it is composed of 4 base pairs. However, a substantial amount of cloverleaves (35) lack one of these base pairs, a feature which appears with exceptional majority for class II tRNAs (25-fold). Other stem lengths in the D arm include 2, 5 and 6 base pairs, each with low probability. Concerning the loops, the proposed classical size of seven unpaired nucleotides in the Tψ loop is unmistakably predominant (113-fold). Other possible sizes contain 5, 6, 8, 10, 13 and 15 bases, each occurring with minimal probability. The anticodon loop is written down with seven residues in 63 cases which amounts to just the majority of sequences. The opening of a

number of bp	acceptor	D	anticodon	T
2	-	2.4%	0.8%	-
3	-	28.7%	6.6%	3.3%
4	1.6%	62.3%	26.2%	4.1%
5	0.8%	5.7%	48.4%	89.3%
6	5.7%	0.8%	13.1%	3.3%
7	87.7%	-	4.9%	-
8	4.1%	-	-	-

Table 4: An overview of the frequency of different helix lengths occurring in the cloverleaves listed in table 3.

base pair and the resulting elongation of the loop to nine nucleotides does also appear with high frequency (40-fold). This finding is in accordance with favoured anticodon lengths mentioned above. For class II tRNAs, the 9 base pair loop is the most frequent one (formed by 17 out of 29 sequences). Sizes 3, 5, 6, 11 and 13 have also been detected. The D loop presents considerable variation in size. Forty-one cloverleaves depict 8 unpaired nucleotides in this loop region, while 30 have an additional residue. An increase to 11 bases is also frequently found for 25 sequences, most prominently however for class II tRNAs. In combination with the preferred D stem size of 3 base pairs, it appears that class II tRNAs accomodate their increased length by enlarging the D arm area as well as, of course, the variable region. The latter serves as a criterium for partitioning tRNAs in the two known categories, after all. It may comprise 10 to 24 nucleotides for class II tRNAs. As the applied algorithms appoint a favourable energy contribution to every base pair that is formed, it does not come as a surprise that an additional stem is built in the extra arm. Calculated stem sizes range from 2 to 7 base pairs, with a pronounced occurrence of 5, while the loops contain predominantly 4 nucleotides. Triloops are also common in this area. In four cases (RL1440, RL1441, RL1442, RS1541), a seamless transition between the three stem regions (anticodon, variable and Tψ arm) takes place. For class I tRNAs on the other hand, the variable region is devoid of base pairs, the number of unpaired nucleotides stretching from 2 to 7 residues, with a clear dominance of 5.

As there is mostly a substantial majority of a certain stem or loop length, it can be expected that it is also maintained within an individual amino acid family. This is indeed the case, most noticeably where the Tψ stem is concerned. The sizes of both Tψ stem and loop are conserved in no less than nine families (asn, cys, gly, his, phe, ser, thr, tyr and val). In ala, asp, glu, ile and lys, at least the number of unpaired nucleotides in the Tψ loop is constant. Likewise, the acceptor stem is retained in 12 different families (asn, asp, cys, gln, gly, his, ile, pro, ser, trp, tyr, val), composed of seven residues except in the case of his where its size is enlarged by one base. The dimension of the D stem is invariable in arg, asp, ile, met, ser and trp, while in five families (ala, gln, phe, pro and thr), even the entire D arm is conserved. The largest variations of size are found in the anticodon area. Either the loop or the stem is maintained (in asp, glu and ile, respectively). The phe family stands out as all D, anticodon and Tψ arms are entirely conserved in all six representatives. If it was not for sequence RF1580 that has a modified residue on position 6 which is, according to the used algo-

rithms, unable to participate in a base pair, and therefore restricts the acceptor stem to only five base pairs, the whole cloverleaf would be conserved.
As the other extreme, the leu family shows the greatest diversity of stem and loop sizes which is no doubt at least partly due to its large number of representatives (12) in comparison to all other families.
Going into a few peculiarities, it can be pointed out that gly sequences show a tendency to form an extended anticodon stem consisting of seven base pairs, and a resulting triloop. Furthermore, two representatives have a D stem composed of the minimal number of two base pairs and an extended region before the beginning of the anticodon arm. The cloverleaf structures of two lys sequences are also remarkable in the D region, where they describe a hairpin with a three base pair helix and a tetraloop, the smallest D loop size to be found among all test sequences.

In the majority of cases, thermodynamic calculations based on suboptimal folding confirm the conception of the cloverleaf as the native conformation of tRNAs. As for the few exceptions, the free energy difference between the minimum free energy structure and the cloverleaf is mainly small.
The next interesting aspect is the study of the molecule's folding dynamics and the determination whether the modifications also exert an effect on the folding pathway leading to the native state. Far from taking the direct route, many RNAs pass through a series of metastable states before finding their natural conformation. Such local minima represent potential folding traps, due to the high thermodynamic stability of RNA double helices. The stability bonus provided by base pair stacking is usually sufficient (that is to say large compared to the thermal energy) to postpone or even prevent an opening of an existing helix. In the latter case, the molecule does not reach the ground state within the time scale of the simulation. The first 100 local minima were determined with the help of Barriers. Whether a certain metastable state represents a potential kinetic trap was decided by means of its barrier height.
Only local minima with a barrier equal or over 6 kcal/mole were chosen as possible stop structures for Kinfold simulations. However, the cloverleaf fold with the lowest energy (for there are frequently several conformations with this distinctive pattern of four helices to be found, only differing in their respective lengths) was picked out regardless of its other attributes or its position in the tree of local minima. In order to determine the distribution of trajectories, a 1000 Kinfold sim-

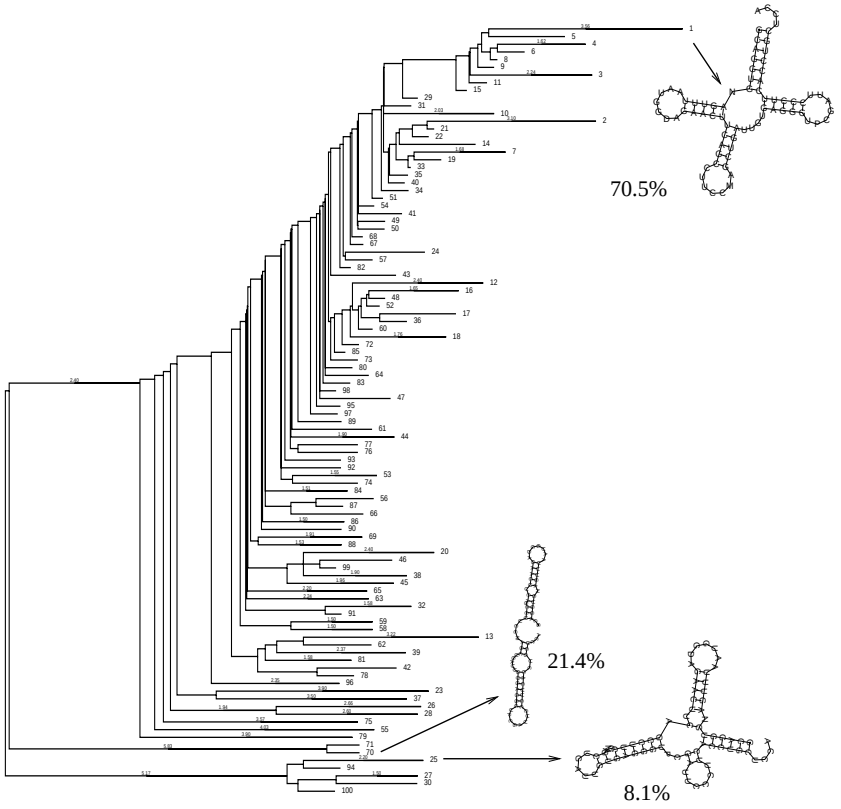


Figure 13: Tree of local minima for a modified $tRNA^{gly}$: Significant local minima and their stop structures are indicated by arrows, with the percentage of trajectories leading to them printed underneath. The cloverleaf appears as the ground state.

ulations were performed for each sequence. The glycine tRNA from *Mycoplasma capricolum* (RG1140) will be quoted as an example for illustration of the typical folding behaviour of modified sequences. Its tree of local minima is presented in figure 13. The cloverleaf does coincide with the mfe structure and is therefore displayed on top of the figure, ruling over a large basin composed of numerous tiny folding funnels whose barrier heights are not sufficient to prevent refolding into the lowest-lying conformation (S1). Apart from this extensive subtree, only two other basins are visible on the bottom of the figure, leading to conformations S70, a long hairpin-like structure, and S25 (closer related to the natural conformation in its composition of secondary structure motifs. By the looks of the tree, one might speculate that a molecule that has landed in a local minimum contained in the upper folding funnel will continue its way and will eventually

reach the correct ground state. Accordingly, the great majority of trajectories fold efficiently into the cloverleaf (70.5%). The remaining trajectories are divided among the other two structures, with a clear preference of S70 (21.4%). Only 8.1% of the simulations end in conformation S25.

This sequence was selected in order to draw comparison to the results of the respective unmodified sequence as well as to study the effect of coaxial stacking on the folding behaviour which are discussed in detail in the following two sections. Furthermore, it represents the typical properties of modified tRNAs. The folding characteristics defined by the number of potential folding traps and the percentage of trajectories forming the cloverleaf is diverse among the modified sequences as can be seen in table 5. As would be expected, there is a general correlation between the number of possible stop structures in Kinfold runs and the amount of simulations leading to the cloverleaf. With an increased variety of local minima, the frequency of cloverleaf formation usually declines in proportion. In half of the cases, the natural conformation is favoured against the other chosen meta-stable states by the major part of folding trajectories, just as in the quoted sequence. It can be checked in table 3 that only in a handful of sequences (RG1310, RK1140, RF1540, RW1251 and RL2101), the cloverleaf does not correspond to the ground state at the same time. On the other hand, its occurrence as mfe structure does not serve as a guarantee for excellent folding behaviour. Folding kinetics cannot necessarily be predicted from thermodynamic properties [35].

In order to further investigate this possible connection between thermodynamic and kinetic characteristics, a thorough cloverleaf analysis was conducted. As it turned out, several cloverleaf structures with differing helix lengths (up to sixteen for a single sequence) are usually located among the first hundred local minima calculated by Barriers. Selecting all these correctly folded states as input for Kinfold simulations, the calculations yielded a surprising result. The major part of the computer experiments ended in the natural folds listed in table 3. Nevertheless, in 1/4 of the cases, the cloverleaf found with the highest percentage of trajectories is not the thermodynamically most stable one. Free energy differences between the two respective conformations rise up to nearly 6 kcal/mole, although only a small energy gap usually turns the balance for a higher energy cloverleaf. A rearrangement in both the acceptor and D arms is the main structural difference, with a reduction of D loop size and seamless transition between these two helices obviously shifting the result in the direction of the less stable structure.

tRNA	stops	% clov	tRNA	stops	% clov	tRNA	stops	% clov
RA1140	7	12.6%	RG1661	6	46.4%	RP1140	5	68.0%
RA1180	6	16.9%	RG1662	3	20.0%	RP1180	7	14.0%
RA1540	7	24.8%	RH1140	2	10.1%	RP1540	3	16.4%
RA1660	11	13.2%	RH1660	1	100.0%	RP1700	2	72.5%
RA1661	13	12.1%	RH1700	1	100.0%	RP1701	2	86.1%
RA1662	12	3.3%	RI1140	4	84.0%	RP1702	4	13.2%
RR1140	3	84.9%	RI1141	5	32.9%	RS1140	8	0.6%
RR1141	6	16.3%	RI1180	6	35.4%	RS1663	n.v.	n.v.
RR1540	2	91.2%	RI1580	7	13.2%	RS1664	3	89.1%
RR1660	2	95.8%	RI1660	5	35.1%	RS1141	11	3.2%
RR1661	2	90.8%	RI1661	9	16.4%	RS1180	10	5.6%
RR1662	2	14.0%	RI1662	1	100.0%	RS1540	5	26.8%
RR1663	4	42.1%	RI1540	4	73.8%	RS1541	7	6.4%
RR1664	1	100.0%	RL1140	5	17.8%	RS1542	9	2.8%
RN1140	20	14.3%	RL2020	3	23.8%	RS1660	4	69.5%
RN1660	3	85.8%	RL2100	5	68.1%	RS1661	7	45.4%
RN1720	5	24.4%	RL2101	6	38.8%	RS1662	n.v.	n.v.
RN1721	6	14.6%	RL1141	9	5.7%	RT1140	2	14.4%
RD1140	4	33.0%	RL1142	5	27.5%	RT1141	7	10.9%
RD1580	2	57.1%	RL1460	4	14.3%	RT1180	12	8.6%
RD1660	3	46.3%	RL1540	7	3.3%	RT1540	2	77.1%
RC1140	6	3.0%	RL1660	5	33.8%	RT1660	2	73.5%
RC1660	2	60.6%	RL1661	n.v.	n.v.	RT1661	4	12.3%
RQ1140	5	7.1%	RL1662	3	77.7%	RW1140	1	100.0%
RQ1660	3	88.7%	RL1700	n.v.	n.v.	RW1141	1	100.0%
RQ1660	4	59.2%	RK1140	2	76.2%	RW1250	3	76.0%
RE1140	3	39.5%	RK1141	2	87.9%	RW1251	3	54.1%
RE1660	4	27.6%	RK1540	7	1.6%	RW1540	2	62.0%
RE1661	3	29.6%	RK1541	7	2.1%	RW1660	1	100.0%
RE1662	3	5.5%	RK1660	2	75.4%	RY1140	n.v.	n.v.
RE2140	4	25.5%	RM1140	1	100.0%	RY1460	7	21.6%
RG1140	3	70.5%	RM1540	6	34.4%	RY1540	9	7.2%
RG1700	4	25.6%	RM1580	3	1.4%	RY1541	9	7.2%
RG1701	3	21.4%	RM1660	2	4.6%	RY1660	2	81.4%
RG1180	4	43.2%	RF1140	4	46.5%	RY1661	1	100.0%
RG1310	3	55.0%	RF1460	2	98.5%	RV1140	8	11.0%
RG1380	3	5.7%	RF1540	3	74.0%	RV1180	6	21.5%
RG1381	4	5.5%	RF1580	7	9.7%	RV1460	3	83.1%
RG1540	2	94.5%	RF1660	1	100.0%	RV1540	5	53.1%
RG1660	2	69.4%	RF2020	1	100.0%	RV1660	4	40.9%

Table 5: The number of potential kinetic traps and the percentage of trajectories leading to the lowest-lying cloverleaf (compare table 3) for each sequence. "n.v." indicates that no cloverleaf was found among the first hundred local minima.

Before making a comparison to the respective unmodified sequences, a short insertion shall draw attention to an interesting feature frequently occurring in tRNAs.

5.1.1 RNA molecular switches

The determination of RNA switches is one of the most interesting features revealed by the calculation of local minima on the folding landscape. Its analysis on the level of the secondary structure has shown that non-native conformations are often energetically comparable to the ground state, albeit being separated from it by very high energy barriers. Experimental results of various RNA molecules point to the existence of stable alternative conformations [29, 49]. Within the same RNA, they are related to and responsible for entirely different functions [7, 106]. Replication of SV11 by Qβ replicase [9] for example strongly depends on the molecule's structure. There are two major conformations, a meta-stable multi-component structure and a rod-like conformation presenting the native state, explicitly separated by an immense energy barrier. The meta-stable conformation alone is an active template for the Qβ replication assay. The transformation from inactive, yet stable, and active, meta-stable state can be accomplished by melting and rapid quenching of the molecule [147]. Another switch has been reported to be responsible for codon-anticodon interaction and the specific recognition of tRNA at the ribosomal A site [82, 139].
The capability of RNA molecules to form multiple (meta-) stable conformations with varying function is used in nature to implement *molecular switches* adjusting and modulating the flow of a number of biological processes. The computation of low-energy local minima on the folding landscape marks the starting point for a determination of possibly recurring structural characteristics.
Based on the visual evidence presented by the tree of local minima, a program was developed to calculate the energy differences between successive conformations and establish a connection to individual secondary structure motives. The presumption was confirmed: a considerable number of structure pairs with the same energy gap and only differing by a single structural element usually appear throughout the tree.
A very convincing example is illustrated by the tree of local minima of a modified serine tRNA sequence in figure 14. The distinction between the mfe structure and the next following conformation, which happens to be the cloverleaf in this case, is attributed to a structural rearrangement of just seven nucleotides which is shown in figure 15. The correctly folded anticodon loop of the natural conformation is energetically not as favourable as the formation of two additional A-U base pairs which gives rise to an internal loop and restricts the anticodon loop to

three residues. The energy difference accounting for this conformational change amounts to -0.70 kcal/mole. The same value is repeatedly found and each time indicates the structural switch between the internal loop/triloop combination and the enlarged loop, while the remainder of the conformation naturally differs between arbitrary structure pairs. It follows from the markings in figure 14 that this switch does not only recur within one folding funnel, but instead pervades the entire tree, no less than 25 times. This feature of switching structural elements appears in most of the trees of both modified and unmodified tRNAs, with a wide variety regarding the nature and the number of occurrences. The question arises

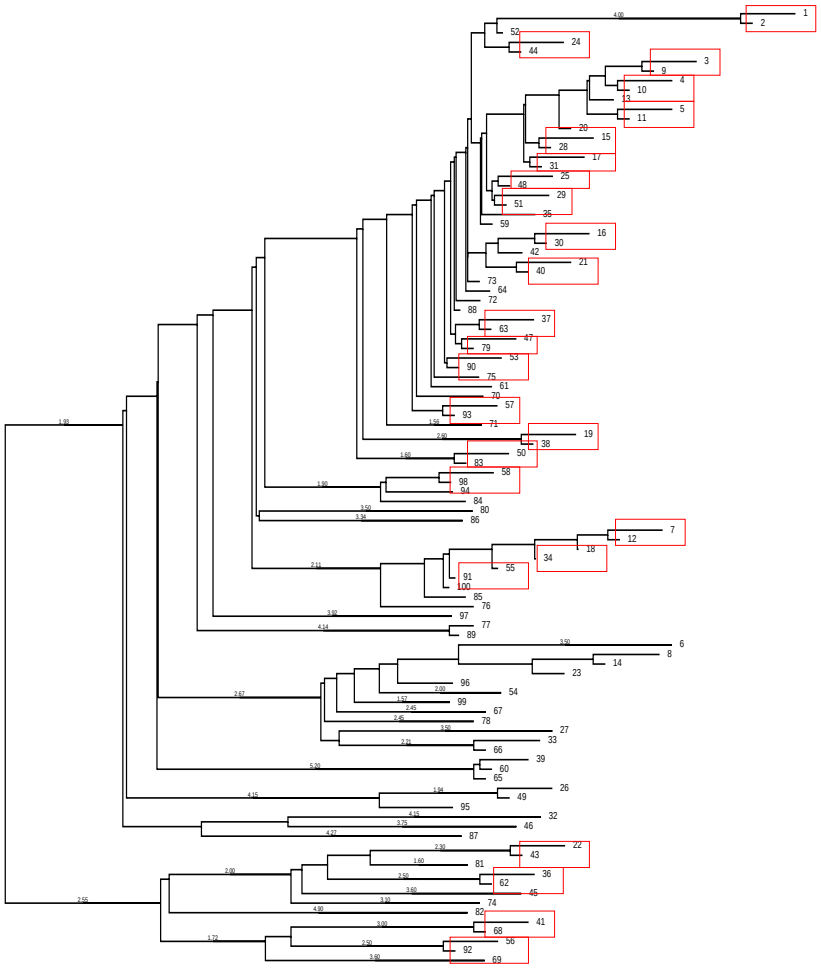


Figure 14: Tree of local minima for a modified serine tRNA; boxes indicate the existence of the same conformational switch between structure pairs.

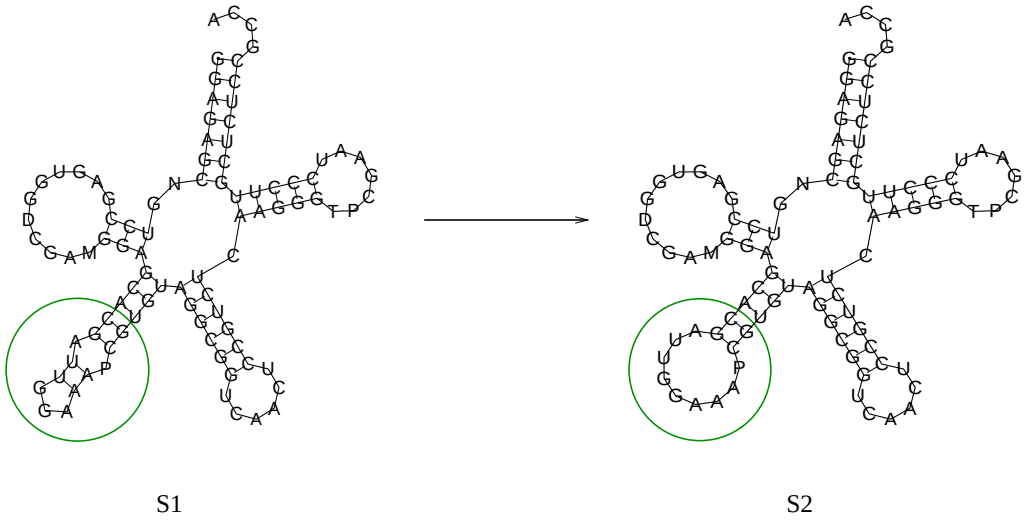


Figure 15: The recurring structural switch between an internal loop/triloop formation and the opening to a large loop, illustrated for mfe structure and first suboptimal conformation.

whether these modular units could actually be related to biological functions. In any case, the result suggests the feasibility of the design of RNA switches with predetermined alternative conformations, the first tests of which have already been successfully reported [42].

5.2 Unmodified sequences

The analysis of unmodified tRNAs was conducted in order to draw a parallel to the naturally occurring form of these sequences and to emphasize the effect of base modifications. Modified nucleotides prevent base pairing, but are allowed to contribute to the overall free energy in form of single base interactions, like terminal mismatches accounting for stacking on top of double helices. Conditions for suboptimal folding are as described in the previous section.
The minimum free energy and the energy difference between the mfe structure and the first suboptimal conformation for each sequence are given in table 6. The limits of the mfe are -19.00 kcal/mole (RW1251) and -37.81 kcal/mole (RS1660) for the lower and upper value, respectively. The free energy gap is as expected markedly smaller than for the modified sequences. Base modification effects tRNA stability by imposing a strong constraint on secondary structure formation.

tRNA	mfe [kcal/mol]	ΔE [kcal/mol]	tRNA	mfe [kcal/mol]	ΔE [kcal/mol]
RA1140	-32.10	-0.70	RG1661	-28.96	-0.46
RA1180	-32.10	-0.70	RG1662	-26.91	-0.01
RA1540	-29.56	-0.30	RH1140	-26.20	-0.40
RA1660	-27.06	-0.30	RH1660	-25.46	-0.60
RA1661	-29.30	-0.03	RH1700	-25.46	-0.60
RA1662	-30.30	-0.20	RI1140	-26.60	-0.60
RR1140	-29.70	-0.40	RI1141	-19.90	-0.30
RR1141	-26.30	-0.10	RI1180	-19.90	-0.30
RR1540	-27.70	-0.44	RI1580	-27.43	-0.60
RR1660	-27.50	0.00	RI1660	-30.37	-0.40
RR1661	-27.29	0.00	RI1661	-24.90	-0.20
RR1662	-24.50	-0.40	RI1662	-24.90	-0.20
RR1663	-30.50	-0.50	RI1540	-34.70	-0.60
RR1664	-30.16	-0.06	RL1140	-26.20	-0.60
RN1140	-23.87	-0.97	RL2020	-26.17	-0.27
RN1660	-22.40	-0.60	RL2100	-33.53	0.00
RN1720	-21.70	-0.43	RL2101	-32.80	-0.60
RN1721	-20.20	-0.5	RL1141	-28.60	-0.30
RD1140	-27.56	0.00	RL1142	-30.60	-0.60
RD1580	-28.66	-0.20	RL1460	-35.10	0.00
RD1660	-26.10	-0.60	RL1540	-37.10	0.80
RC1140	-26.00	-0.50	RL1660	-35.88	-0.18
RC1660	-28.50	-0.90	RL1661	-31.20	-0.24
RQ1140	-24.30	-0.34	RL1662	-32.50	-0.57
RQ1660	-26.90	-0.90	RL1700	-31.20	-0.24
RQ1661	-28.60	-0.90	RK1140	-22.27	-0.07
RE1140	-25.40	-0.31	RK1141	-28.80	-0.20
RE1660	-33.66	-0.10	RK1540	-25.97	-0.37
RE1661	-31.86	-0.10	RK1541	-25.97	-0.37
RE1662	-31.86	-0.50	RK1660	-26.80	-0.60
RE2140	-29.20	-0.50	RM1140	-27.30	-0.60
RG1140	-22.30	-0.24	RM1540	-29.60	-0.40
RG1700	-31.10	-0.70	RM1580	-31.16	-0.60
RG1701	-26.91	-0.01	RM1660	-22.70	0.00
RG1180	-22.06	-0.17	RF1140	-27.40	0.00
RG1310	-28.70	0.00	RF1460	-25.96	-0.20
RG1380	-22.60	-0.10	RF1540	-28.69	-0.49
RG1381	-23.70	-0.40	RF1580	-22.30	0.00
RG1540	-20.90	-0.30	RF1660	-26.80	-0.30
RG1660	-31.10	-0.70	RF2020	-24.30	-0.20

Table 6: Unmodified natural tRNAs: minimum free energy and energy gap between the mfe and the first suboptimal conformation.

tRNA	mfe [kcal/mol]	ΔE [kcal/mol]	tRNA	mfe [kcal/mol]	ΔE [kcal/mol]
RF2060	-24.60	-0.20	RT1660	-28.36	-0.60
RP1140	-29.80	-0.60	RT1661	-27.56	-0.40
RP1180	-29.8	-0.60	RW1140	-20.70	-0.21
RP1540	-28.90	-0.60	RW1141	-21.56	-0.60
RP1700	-28.00	-0.20	RW1250	-22.76	-0.10
RP1701	-28.80	-0.10	RW1251	-19.00	-0.20
RP1702	-30.46	-0.49	RW1540	-22.50	-0.20
RS1140	-28.23	-0.05	RW1660	-26.00	-0.30
RS1663	-34.48	0.00	RY1140	-32.00	-0.20
RS1664	-35.30	-1.60	RY1460	-35.10	-0.20
RS1141	-30.30	-0.34	RY1540	-33.80	-0.20
RS11180	-34.70	-0.34	RY1541	-33.80	-0.20
RS1540	-32.67	-0.07	RY1660	-29.44	-0.20
RS1541	-39.20	-0.60	RY1661	-29.44	-0.20
RS1542	-34.68	-0.38	RV1140	-24.90	-0.10
RS1660	-37.81	-0.40	Rv1180	-24.90	-0.10
RS1661	-45.30	-0.60	RV1460	-29.73	-0.10
RS1662	-34.48	0.00	RV1540	-28.40	-0.40
RT1140	-22.07	-0.17	RV1660	-28.60	-0.20
RT1141	-22.50	-0.03	RV1661	-24.70	-0.13
RT1180	-22.50	-0.03	RV1662	-31.20	-0.55
RT1540	-25.80	-0.20			

table 6 continued

Without this restriction, additional suboptimal conformations are allowed, leading to a decrease of the energy difference. Nine sequences have in fact degenerate ground states. With the exception of RS1664, which features the largest energy gap (-1.60 kcal/mole), its average lies around -0.60 kcal/mole. Energy differences vary among the amino acid groups (although not as drastically as in the modified case), the tyr family presenting the exception to the rule with a constant value of -0.20 kcal/mole.

Lacking modified nucleotides that preclude base pairs at a few essential positions and thus direct cloverleaf formation, unmodified sequences do not favour the predicted fold. Only 19 out of 123 tRNAs have the natural conformation as mfe structure (let us recall that the value amounted to 76 for the modified sequences).

tRNA	Acceptor	D	anticodon	VR	T	Acceptor	rank
RA1140	((((((.	(((.....))))..	(((.....)))	-	(((((((.....)))))))....	)))))))....	3.sub
RA1180	((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	3.sub
RA1540	((((((.	(((.....)))).	((((((.....))))))	.....	((((.....))))	)))))))....	mfe
RA1660	((((((.	(((.....)))).	((((.....))))	.....	((((.....)))).	)))))))....	mfe
RA1661	(((((..		((((.....))))	.....	(((.....))))).	)))))))....	84.sub
RA1662	(((((((.	(((.....))))....	(((.....)))))	-	((((.....))))	)))))))....	45.sub
RR1140	((((((.	(((.....)))))	(((.....)))))	-	((((..........))))))..	)))))))....	1579.sub
RR1141	((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	189.sub
RR1540	((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	11.sub
RR1660	((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	190.sub
RR1661	((((((.	(((.....)))).	((((.....))))	...	(((((.....))))	)))))))....	162.sub
RR1662	(((((..	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	mfe
RR1663	((((((.	(((.....)))).	((((.....))))	.....	(((.....)))).	)))))))....	4.sub
RR1664	((((((.	(((.....))))	(((((.....)))))	....	((((.....))))	)))))))....	215.sub
RN1140	((((((.	(((.....)))).	(((.....))))	.....	((((.....))))	)))))))....	mfe
RN1660	((((((.		(((.....))))	.....	((((.....))))	)))))))....	2.sub
RN1720	(((((((	(((((.....))))))	((((.....))))	.....	((((.....))))	)))))))....	1.sub
RN1721	((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	3.sub
RD1140	((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	mfe
RD1580	((((((.	(((.....)))).	((((.....))))	.....	((((.....))))	)))))))....	3.sub
RD1660	((((((.		(((.....))))	.....	((((.....))))	)))))))....	26.sub
RC1140	(((((((	((((...)))))....	((((((...))))))	....	((((.....))))	)))))))....	59.sub
RC1660	(((((((	(((.....))).	((((.....))))	...	((.....))))	)))))))....	4.sub
RQ1140	((((((.	((.....))).	(((.....))))	.....	((((.....))))	)))))))....	1142.sub
RQ1660	((((((.	((.....))).	((((.....))))	.....	((((.....))))	)))))))....	18.sub
RQ1661	((((((.	((.....))).	((((.....))))	.....	((((.....))))	)))))))....	56.sub
RE1140	((((((.	(((.....))))	(((((.....)))))	..	(((((.....)))))	)))))))....	345.sub
RE1660	(((((..	(((.....))))	(((((.....)))))	...	((((.....))))	)))))))....	1.sub
RE1661	((((((.	(((.....))))	(((((.....)))))	...	((((.....))))	)))))))....	1.sub
RE1662	((((((.	(((.....)))).	((((.....))))	....	((((.....))))	)))))))....	mfe
RE2140	((((((.						16.sub

Table 7: Cloverleaves of unmodified tRNAs

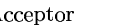
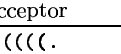
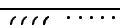
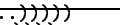
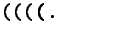
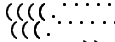
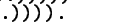
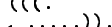
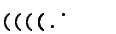
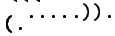
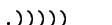
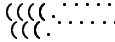
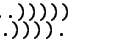
tRNA	Acceptor	D	anticodon	VR	T	Acceptor	rank
RG1140				....			1.sub
RG1700				....			15.sub
RG1701				....			28.sub
RG1180				-			mfe
RG1310				....			7.sub
RG1380				....			179.sub
RG1381				....			25.sub
RG1540				....			5.sub
RG1660				.....			15.sub
RG1661				....			mfe
RG1662				.....			28.sub
RH1140				.....			134.sub
RH1660				.....			mfe
RH1700				.....			mfe
RI1140				.....			76.sub
RI1141				.....			19.sub
RI1180				.....			19.sub
RI1580				.....			41.sub
RI1660				.....			19.sub
RI1661				.....			21.sub
RI1662				....			21.sub
RI1663				....			33.sub
RK1140				..			125.sub
RK1141				-			20.sub
RK1540				-			mfe
RK1541				-			mfe
RK1660				.....			4.sub
RM1140				.....			6.sub
RM1540				.....			149.sub
RM1580				.....			2.sub
RM1660				.....			9.sub
RF1140				....			10.sub

table 7 continued

tRNA	Acceptor	D	anticodon	VR	T	Acceptor	rank
RF1460	((((((.·	(((.·····.)))).	(((.··.)))))	·····	((((.·····))))	)))))))....	1.sub
RF1540	((((((.·	(((.·····.)))).	(((.·····.)))))	·····	((((.·····))))	)))))))....	1.sub
RF1580	((((.·	(((.·····.)))).	(((.·····.)))))	·····	((((.·····))))	))))).....	8.sub
RF1660	((((.·	(((.·····.)))).	((((((.··))))))	·····	((((.·····))))	)))))))....	6.sub
RF2020	((((.·	(((.·····.)))).	(((.·····.)))))	·····	((((.·····))))	)))))))....	51.sub
RP1140	((((((.·	(((.·····.)))).	((((((.··))))))	·····	((((.·····))))	)))))))....	28.sub
RP1180	((((((.·	(((.·····.)))).	((((((.··))))))	·····	((((.·····))))	)))))))....	99.sub
RP1540	((((((.·	(((.·····.)))).	((((((.··))))))	·····	((((.·····))))	)))))))....	99.sub
RP1700	((((((.·	(((.·····.))))....	((((.·····.))))	···	((((.·····))))	)))))))....	29.sub
RP1701	((((((.·	(((.·····.)))).	((((((.··))))))	·····	((((.·····))))	)))))))....	33.sub
RP1702	((((((.·	(((.·····.)))).	((((((.··))))))	·····	((((.·····))))	)))))))....	577.sub
RT1140	((((((.·	(((.·····.))))....	(((.·····.))))	-	((((.·····))))	)))))))....	10.sub
RT1141	((((((.·	(((.·····.))))....	((((.·····.))))))	-	((((.·····))))	)))))))....	6.sub
RT1180	((((((.·	(((.·····.))))....	((((.·····.))))))	-	((((.·····))))	)))))))....	1.sub
RT1540	((((((.·	(((.·····.)))).	((((.·····.))))	·····	((((.·····..))))..	)))))))....	1.sub
RT1660	((((((.·	(((.·····.)))).	((((((.··))))))	·····	((((.·····))))	)))))))....	47.sub
RT1661	((((((.·	(((.·····.)))).	((((((.··))))))	····	((((.·····))))	)))))))....	mfe
RW1140	((((((.·	(((.·····.)))).	((((.·····.))))	·····	((((.·····)))).	)))))))....	2.sub
RW1141	((((((.·	(((.·····.)))).	((((.·····.))))	·····	((((.·····))))	)))))))....	3.sub
RW1250	((((((.·	(((.·····.)))).	(((.·····.))))	·····	((((.·····))))	)))))))....	mfe
RW1251	((((((.·	(((.·····.)))).	((((.·····.))))	·····	((.·····)))..	)))))))....	mfe
RW1540	((((((.·	(((.·····.)))).	(((.·····.))))	····	((((.·····))))	)))))))....	22.sub
RW1660	((((((.·	(((.·····.)))).	((((.·····.))))	·····	((((.·····))))	)))))))....	1.sub
RV1140	((((((.	((((.·····.)))).	((((((.·····.))))))	..	((((.·····))))	)))))))....	17.sub
RV1180	((((((.	((((.·····.)))).	((((((.·····.))))))	..	((((.·····))))	)))))))....	mfe
RV1460	((((((.·	(((.·····.)))).	(((((.···.)))))	·····	((((.·····))))	)))))))....	mfe
RV1540	((((((.·	(((.·····.)))).	((((.·····.))))	·····	((((.·····))))	)))))))....	107.sub
RV1660	((((((.·	(((.·····.)))).	((((.·····.))))	·····	((((.·····))))	)))))))....	53.sub
RV1661	((((((.·	(((.·····.)))).	((((.·····.))))	·····	((((.·····))))	)))))))....	3.sub
RV1662	((((((.·	(((.·····.))))..	((((((.···.))))))	..	((((.·····))))	)))))))....	5.sub
							4.sub

table 7 continued

tRNA	Acceptor	D	anticodon	VR	T	Acceptor	rank
RL1140	(((((((.	((.....)))..	(((.....)))..	((((.....))))	((((.....))))	)....	45.sub
RL2020	(((((((.	((.....)))	(((.....)))..	(((.....))).	((((.....))))	)....	1330.sub
RL2100	(((((((.	((.....))).	(((.....)))..	((((.....))))	((((.....))))	)....	57.sub
RL2101	(((((((.	((.....)))	(((.....)))	(((.....)))..	((((.....))))	)....	35.sub
RL1141	(((((((.	((.....)))..	(((.....)))	((((.....))))	((((.....))))	)....	2.sub
RL1142	(((((((.	((.....)))..	(((.....)))	(((.....))).	((((.....))))	)....	472.sub
RL1460	(((((((.	((.....)))	(((.....)))..	(((.....))).	((((.....))))	)....	29.sub
RL1540	(((((((.	((.....))).	(((.....)))..	(((.....))).	((((.....))))	)....	1476.sub
RL1660	(((((((.	((.....))).	(((.....)))..	(((.....))).	((((.....))))	)....	mfe
RL1661	(((.....	((.....)))	(((.....)))..	(((.....))).	((((.....))))	)....	3428.sub
RL1662	(((((((.	((.....)))	(((.....)))..	(((.....))).	((((.....))))	)....	2.sub
RL1700	(((.....	((.....)))	(((.....)))..	(((.....))).	((((.....))))	)....	3428.sub
RS1140	(((((((.	((.....)))	(((.....)))..	(((.....))).	((((.....))))	)....	2.sub
RS1663	(((((((.	((.....))).	(((.....)))..	(((.....))).	((((.....))))	)....	8.sub
RS1664	(((((((.	((.....))).	(((.....)))..	(((.....))).	((((.....))))	)....	35.sub
RS1141	(((((((.	((.....))).	(((.....)))..	(((.....))).	((((.....))))	)....	257.sub
RS1180	(((((((.	((.....))).	(((.....)))..	(((.....))).	((((.....))))	)....	195.sub
RS1540	(((((((.	((....)))	(((.....)))	((((((...))))))	((((.....))))	)....	144.sub
RS1541	(((((((.	((.....)))	(((((....)))))))	(((((.....))))).	((((.....))))	)....	25.sub
RS1542	(((((((.	(((..)))).....	((((.....)))).	(((((...))))).	((((.....))))	)....	4.sub
RS1660	(((((((	(((...)))).....	(((((.....)))))	((((((...)))))).	((((.....))))	)....	mfe
RS1661	((((((((	(((.....)))	(((((.....)))))	(((((...))))).	((((.....))))	)....	5.sub
RS1662	(((((((.	((.....)))	(((.....)))	((..)))..	(((((...))))).	)....	8.sub
RY1140	(((((((.	(((.....))).....	(((((.....)))))	((..)))..	((((.....))))	)....	436.sub
RY1460	(((((((	(((.....))).....	(((((.....)))))	((..)))..	((((.....))))	)....	451.sub
RY1540	(((((((	(((.....))).....	(((((.....)))))	((..)))..	((((.....))))	)....	15.sub
RY1541	(((((((	((.....))).....	((((.....))))	((..)))..	((((.....))))	)....	15.sub
RY1660	(((((((.	(((.....))))	(((.....)))).	((..)))..	((((.....))))	)....	1.sub
RY1661	(((((((.	(((.....))))	(((.....)))).	((..)))..	((((.....))))	)....	1.sub

table 7 continued

The cloverleaf motives presented in table 7 mostly belong to high-energy suboptimal states. A remarkable example is noted by a glutamine tRNA from *Mycoplasma capricolum* (RQ1140) which retained the cloverleaf found for the modified case, however accompanied by a descent from 32^{nd} to 1142^{nd} suboptimal rank. With so few ccurrences as mfe structure, this feature is unlikely to be maintained within an amino acid group. All representatives of the ile family which showed uniform behaviour before have cloverleaves lying in the suboptimal range. In addition to a considerable shift towards less stable confomations, the detected cloverleaves often differ from their modified counterparts, due of course to the unconstrained possibility of base pairing.

Nevertheless, the proposed helix lengths are still formed by the majority of sequences as pointed out in table 8. The acceptor stem consists of seven nucleotides in 102 cases, considerably out-weighing the possibility of 4, 5, 6 or 8 base pairs. Five pairs is the favoured length of of the anticodon stem (52-fold), with the size resulting from either insertion or deletion of a single base pair occurring with almost equal frequency (25-fold and 24-fold, respectively). An extension of the anticodon helix to seven pairs is much more common as in the modified case (19-fold as opposed to just 6-fold), a fact which is obviously attributed to the lack of the clustering of modifications in the anticodon area. A reduction to only three base pairs, on the other hand, only occurs twice. The conventional Tψ stem length of five base pairs is met by the vast majority of 106 sequences. Other stem sizes of 3, 4 or 6 pairs are accordingly rare. In three cases, an eight base pair helix is built, unseen for the modified sequences and due to the reasons mentioned above. While not favoured to the same extent as the Tψ stem, the proposed D helix of four base pairs is indeed formed by 81 sequences. As before,

number of bp	acceptor	D	anticodon	T
2	-	2.5%	-	-
3	-	23.8%	1.6%	2.8%
4	2.5%	66.4%	19.7%	4.1%
5	2.5%	5.7%	42.6%	86.9%
6	5.7%	1.6%	20.5%	4.1%
7	83.6%	-	15.6%	-
8	5.7%	-	-	2.5%

Table 8: An overview of the frequency of different helix lengths occurring in the cloverleaves listed in table 7.

class II tRNAs usually present one base pair less in this region.
From all the components of the cloverleaf fold, the Tψ arm is the one best fulfilling the predictions of its size. Accordingly, 108 cloverleaves contain the seven nucleotide loop. The same loop size is predominant in the anticodon area (82-fold). The high frequency of nine residues found for the modified tRNAs is drastically reduced (12-fold), in favour of a triloop (15-fold) whose occurrence is related to the seven base pair helix. The situation is not as obvious for the D loop. Whereas forty sequences form a loop composed of eight nucleotides, the probability of an additional residue is almost as high (38-fold). As mentioned before, class II tRNAs mostly have an extended D arm, caused primarily by a rise to eleven nucleotides in the loop, a feature which is found exclusively in this category. The same is true for base pairing in the variable region. Four to five pairs are accomodated in this area, with a preferred loop length of four residues. In class tRNAs, the variable region is composed of 2 to 7 nucleotides, with size five prevailing. Surprisingly often (11-fold), and not detected in the modified case, the variable region disappears altogether, for the benefit of either an enlarged Tψ arm (RA1140, RA1180 and RR1441), or, more often, an extended anticodon region (RR1140, RG1310, RK1540, RK1541, RK1660, RT1140, RT1141 and TR1180). This feature accumulates in both the lys and tyr families. In all cases, a stretch of four unpaired nucleotides between D and anticodon stems implies a shift of the variable region. Unmodified tRNAs show a wide variety of stem and loop sizes. Nevertheless, some arms are conserved within an amino acid family, above all the Tψ arm in asn, asp, cys, his, ile, phe, pro and val. At least the length of the Tψ loop is constant in arg, glu and lys. The acceptor stem is maintained in ten groups (asn, asp, cys, gln, his, ile, pro, trp, tyr and val), only in the his family composed of eight instead of the usual seven base pairs. The entire D arm remains unchanged in ala, gln, phe and thr. The anticodon arm is responsible for the clear distinction between the cloverleaves. Only in ile, at least the loop size is uniform. The other exception is presented by the asp family which has the entire cloverleaf conserved, a result that does not occur in the table of modified tRNAs. Both the pro and ile families come close, with just one representative preventing the perfect outcome.
In general, a tendency to close additional base pairs and thus extend the helices constituting the cloverleaf fold, is observed for most of the unmodified sequences. Suboptimal folding shows that the lack of modifications impairs thermodynamic properties, but they were also expected to exert an effect on the folding dynamics.

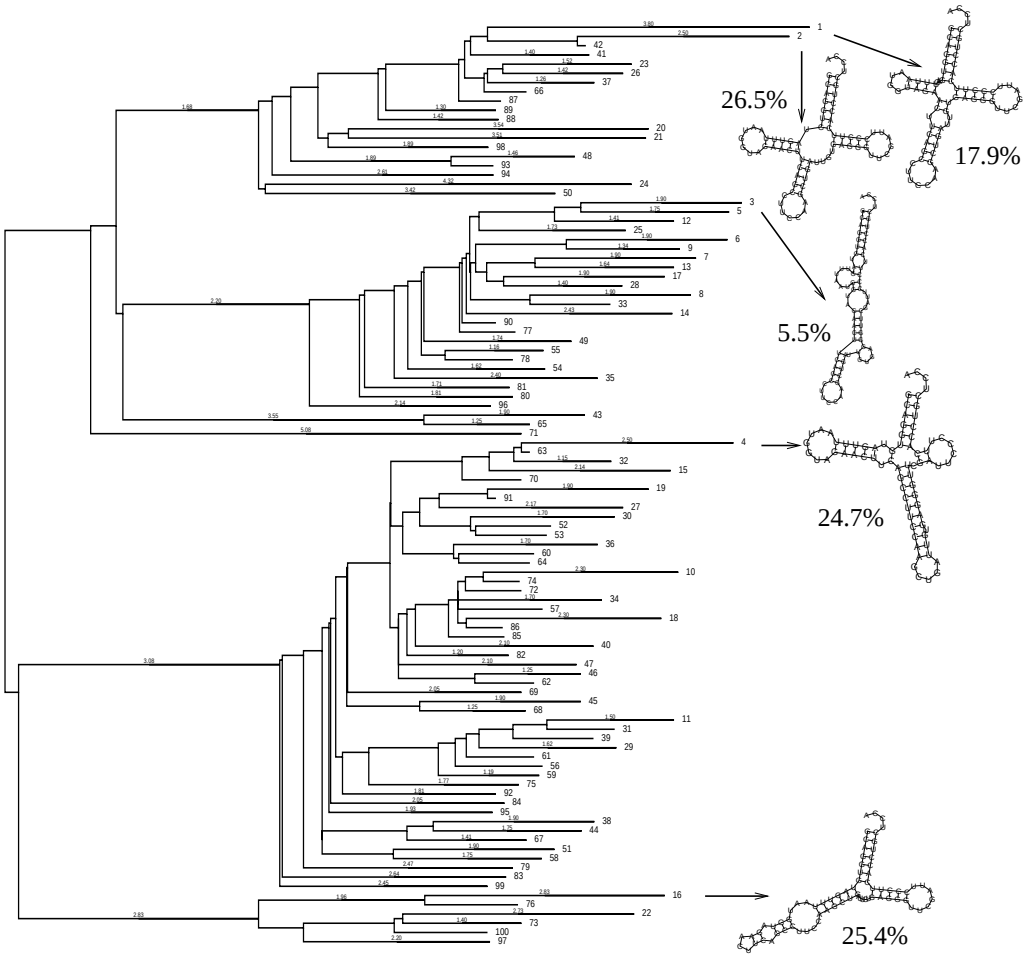


Figure 16: Tree of local minima for the unmodified glycine tRNA. Significant local minima and their stop structures are indicated by arrows.

The glycine tRNA from *Mycoplasma capricolum* serves again as an example, this time however with the four modifications replaced by their standard nucleotides. Figure 16 illustrates the partition of the conformation space for the unmodified sequence: four different folding funnels separated by large energy barriers are visible, leading to conformations S1, S3, S4 and S16, respectively. The natural conformation does not coincide with the minimum free energy structure as in the modified case. The cloverleaf appears right underneath as structure S2. The structural similarity between these two is thus obvious. The Tψ and acceptor stem are correctly folded, but a shortened D arm and the introduction of an internal loop in the anticodon stem are the distinguishing features of structure

S1. In order to determine the distribution of trajectories, the native state was chosen as a possible stop structure in addition to the four previously mentioned conformations for Kinfold calculations. Although the cloverleaf itself does not possess a barrier high enough to qualify as a relevant local minimum, still 26.5% of the folding trajectories of 1000 Kinfold simulations fold into it. In comparison, only 17.9% of the runs end in the minimum free energy structure. Nevertheless, the mfe structure and the two conformations S4 and S16, with the frequency of trajectories leading to them being 24.7% and 25.4%, respectively, represent true long-lived meta-stable states, since molecules that have reached one of these local minima are not likely to escape from them without considerable increase of the folding time because of the high barriers separating them. The only exception to this approximately equal distribution of folding trajectories among the specified stop structures is S3, a long hairpin-like structure which is formed by a rather insignificant amount of 5.5% of the trajectories. While in case of the unmodified tRNA, the folding path to the natural conformation is not kinetically favoured against alternative pathways leading to other basins of attraction, this sort of competition is substantially reduced by the introduction of modified nucleotides. As was pointed out in the previos section, base modification changes the sequence of conformations and turns the cloverleaf into the ground state (70% of the folding trajectories).

Table 9 presents the equivalent to table 4, only concerning the unmodified sequences. As a consequence of several local minima being only reachable via very high energy barriers, the number of specified stop structures for Kinfold simulations often amounted to twelve or more, while the average value was limited to six in the modified case. With such a variety of structure probabilities, it comes to no surprise that the percentage of trajectories forming the cloverleaf is usually one-figured or at least not much higher. In less than 10% of the cases, the majority of simulations end in the natural conformation.

A comparison between different cloverleaf folds could not be carried out because, as opposed to the behavour of modified tRNAs, the exclusion of easily attainable meta-stable states lead to unacceptably long computation times.

The presented results indicate that just a few modified nucleotides at crucial positions both strongly affect thermodynamic properties and considerably improve the folding kinetics.

tRNA	stops	% clov	tRNA	stops	% clov	tRNA	stops	% clov
RA1140	11	4.3%	RG1661	9	15.2%	RP1140	12	7.1%
RA1180	11	3.5%	RG1662	3	22.4%	RP1180	12	6.2%
RA1540	23	12.6%	RH1140	4	33.7%	RP1540	16	0.4%
RA1660	16	12.5%	RH1660	3	69.7%	RP1700	10	4.2%
RA1661	13	27.7%	RH1700	4	55.9%	RP1701	8	13.1%
RA1662	19	2.5%	RI1140	8	31.9%	RP1702	8	2.3%
RR1140	4	70.9%	RI1141	17	3.0%	RS1140	8	0.6%
RR1141	7	2.4%	RI1180	17	2.6%	RS1663	n.v.	n.v.
RR1540	8	17.7%	RI1580	10	14.6%	RS1664	3	89.1%
RR1660	11	6.3%	RI1660	17	3.4%	RS1141	11	3.2%
RR1661	12	6.1%	RI1661	11	4.7%	RS1180	10	5.6%
RR1662	4	12.2%	RI1662	11	4.8%	RS1540	5	26.8%
RR1663	8	26.8%	RI1540	6	12.5%	RS1541	7	6.4%
RR1664	11	13.2%	RL1140	5	17.8%	RS1542	9	2.8%
RN1140	20	9.1%	RL2020	3	23.8%	RS1660	4	69.5%
RN1660	11	17.7%	RL2100	5	68.1%	RS1661	7	45.4%
RN1720	n.v.	n.v.	RL2101	6	38.8%	RS1662	n.v.	n.v.
RN1721	n.v.	n.v.	RL1141	9	5.7%	RT1140	5	7.8%
RD1140	7	0.3%	RL1142	5	27.5%	RT1141	17	1.8%
RD1580	5	33.4%	RL1460	4	14.3%	RT1180	17	2.4%
RD1660	13	0.2%	RL1540	7	3.3%	RT1540	11	10.2%
RC1140	8	4.3%	RL1660	5	33.8%	RT1660	9	25.6%
RC1660	7	17.9%	RL1661	n.v.	n.v.	RT1661	5	38.0%
RQ1140	n.v.	n.v.%	RL1662	3	77.7%	RW1140	6	23.0%
RQ1660	7	8.8%	RL1700	n.v.	n.v.	RW1141	3	80.4%
RQ1660	9	11.5%	RK1140	8	1.6%	RW1250	5	34.7%
RE1140	6	16.0%	RK1141	8	19.2%	RW1251	6	20.8%
RE1660	7	8.3%	RK1540	17	1.5%	RW1540	7	27.2%
RE1661	7	1.8%	RK1541	17	2.0%	RW1660	5	57.0%
RE1662	6	7.4%	RK1660	7	8.3%	RY1140	n.v.	n.v.
RE2140	5	16.2%	RM1140	6	35.2%	RY1460	7	21.6%
RG1140	5	26.5%	RM1540	13	1.8%	RY1540	9	7.2%
RG1700	3	33.5%	RM1580	14	5.3%	RY1541	9	7.2%
RG1701	3	21.4%	RM1660	6	n.v.	RY1660	2	81.4%
RG1180	4	42.7%	RF1140	n.v.	n.v.	RY1661	1	100.0%
RG1310	4	20.0%	RF1460	7	32.3%	RV1140	11	7.1%
RG1380	4	1.9%	RF1540	7	73.6%	RV1180	11	6.7%
RG1381	4	3.7%	RF1580	18	1.9%	RV1460	13	21.3%
RG1540	3	41.3%	RF1660	n.v.	n.v.%	RV1540	5	53.1%
RG1660	3	32.8%	RF2020	13	9.8%	RV1660	4	40.9%

Table 9: The number of potential kinetic traps and the percentage of trajectories leading to the lowest-lying cloverleaf (compare table 7) for each sequence.

5.3 Comparison and the influence of coaxial stacking

In RNA molecules, four-way junctions are folded by pairwise coaxial stacking of helical arms [140], even in absence of added metal ions (in contrast to DNA [27]). The compact fold of a few RNAs determined by X-ray crystallography ($tRNA^{phe}$, hammerhead ribozyme, the P4-P6 domain of the Tetrahymena group I intron, hepatitis delta virus ribozyme [32]) is based on the packing of coaxial helical stacks.

The computation of tRNA multiloops without consideration of this characteristic will remain a crude approximation. Since tRNAs are well-known for their cloverleaf fold, this tertiary interaction has definitely be taken into account to improve the accuracy of prediction. As has been previously mentioned, this is accomplished by re-evaluating the energy using an alternative parameter set compiled in [89].
The consideration of coaxial stacking results in an increased stability of the mfe structure in both modified and unmodified sequences. The ground state is stabilized by one to four kcal/mole, with larger values applying to modified tRNAs. However, the frequency of cloverleaf folds as native states is not importantly affected. A slight improvement in the unmodified case (23 as to 19 occurrences) is noted. A few times, the former mfe structure is moved to a slightly higher suboptimal rank. Accordingly, coaxial stacking changes a suboptimal cloverleaf into the mfe structure. It is interesting to note that this behaviour is mainly found for class II tRNAs, a fact which is obviously attributed to the additional helix formed in the variable region and the augmented development of coaxial stacks. As described in the section 4, stacking contributions are computed and compared for directly adjacent helices as well as for stems separated by one intervening nucleotide. It turned out that the latter case often results in increased stability. Consequently, the cloverleaf folds seldom remain the same as predicted by the normal parameter set. It frequently comes to an opening of a single base pair in the acceptor or Tψ stem to introduce this crucial mismatch.

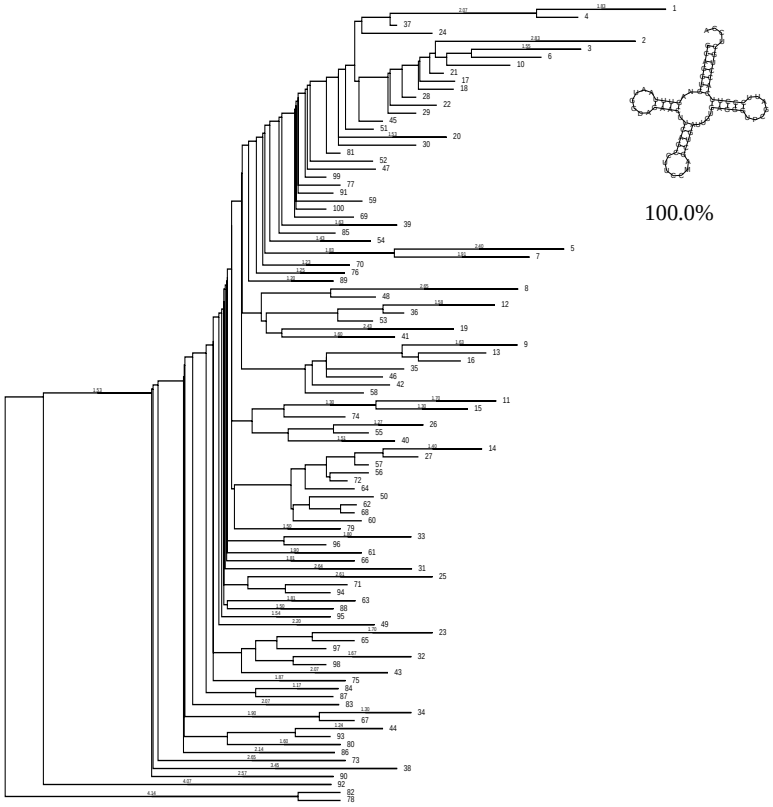


Figure 17: Tree of local minima for the modified glycine tRNA sequence computed with coaxial stacking considered. All trajectories lead to the cloverleaf which dominates the folding tree as minimum free energy structure.

Whereas base modification leads to an obvious enhancement of the folding ability, the incorporation of coaxial stacking should help to stress the dominance of the natural fold still further. As can be seen for the example of glycine tRNA in figure 17, the implementation of this important tertiary interaction allows for a straightforward folding into the cloverleaf. It clearly controls the entire folding tree, and, as a consequence of presenting the only structure with a high barrier, is formed by **all** trajectories. It also slightly differs from the one displayed in figures 13 and 16. Since in this case, a single mismatch between stacking helices results in a higher stability bonus than that provided for directly adjacent helices, a base pair in the D stem is opened to accomodate this situation. A summarizing overview of the folding kinetics of both the unmodified and the modified glycine tRNA as well as the influence of coaxial stacking on the latter's folding behaviour is presented in figure 18. Solid lines denote the fraction of folded sequences as a

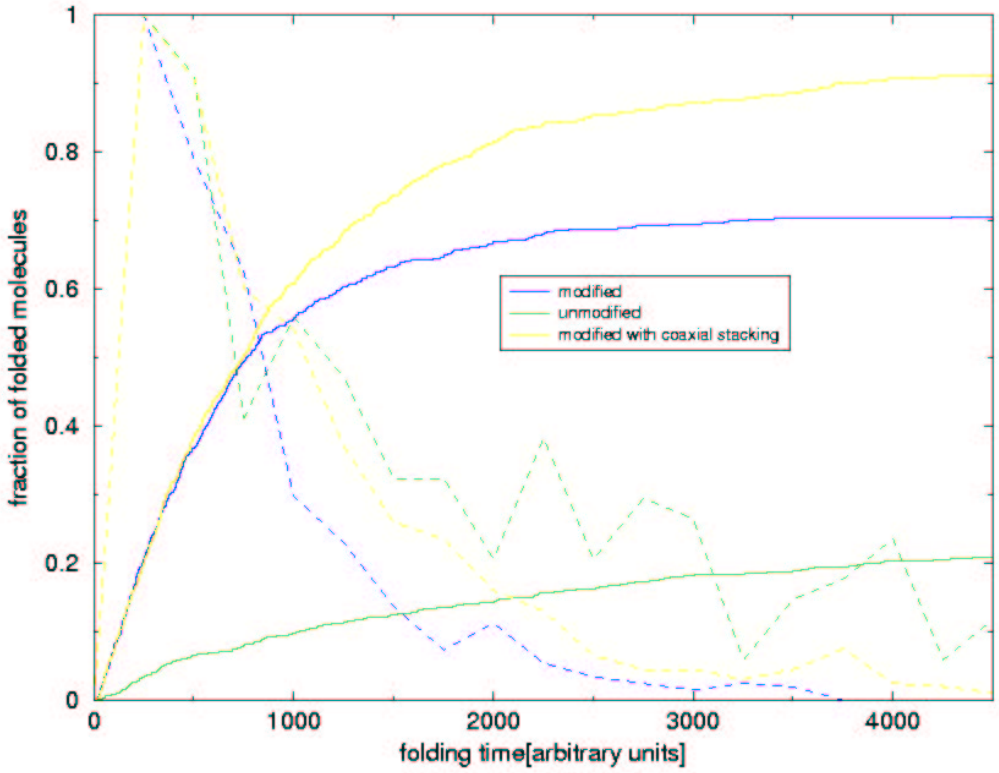


Figure 18: Folding kinetics of both modified and unmodified glycine tRNA, as well as of the modified sequence computed with consideration of coaxial stacking. Solid lines display the fraction of simulations that have found the ground state as a function of time. Dashed lines indicate the density of folding times, scaled in a way that the maximum has height one.

function of time, whereas dashed lines show the density of folding times (which in this context refers to the first passage time from some initial state to the ground state).

Recapitulating the last chapters, the percentage of folded molecules rises from 26.5% to over 70% through the introduction of just four modified nucleotides. Allowing the helical arms of the multiloop to stack coaxially further increases the folding efficiency of the modified tRNA which subsequently found the cloverleaf ground state in all of the 1000 simulations. The respective curve in the diagram has not saturated yet, a fact which results from very differing folding times and the wish to point out their distribution. A very high value for the time limit has been set in order to ensure that all sequences reach a ground state within the simulation time. The modified sequence is an exceptionally fast folder, whereas the unmodified and the coaxially stacked tRNAs have comparable time factors

and fold much more slowly. Regarding the density of folding times, the curves are almost equally smooth for both cases of modified tRNAs. The unmodified sequence, on the other hand, presents a very irregular curve, with several peaks instead of a continuous descent, even at high folding times.
This observation holds true for all unmodified tRNAs. In general, it can be said that the consideration of modified nucleotides entails increased thermodynamic stability as expressed by the free energy gap between mfe structure and first suboptimal conformation as well as the lack of additional suboptimal structures. While this result was more or less taken for granted, the situation for the folding behaviour was not as predictable. It will be shown in the section 7 that it is fairly easy to design sequences to be extraordinarily stable, but without appropriate folding properties. The latter cannot be predicted by the knowledge of thermodynamic qualities.
In the case of natural tRNAs, however, it seems allowed to infer kinetic folding behaviour from thermodynamic calculations, a result which is probably attributed to the optimization by evolutionary processes.
Table 10 provides a comparison between the folding efficiency of modified and unmodified sequences, computed without and with coaxial stacking. Only 15 sequences out of 122 show a higher percentage of trajectories leading to the cloverleaf in the unmodified case, the majority of which encompass class II tRNAs. They present a special case as additional base pairing in the variable region offers alternative cloverleaf possibilities.
The results obtained from the introduction of coaxial stacking are not as unequivocal. Considering only the modified tRNAs, the amount of trajectories folding into the cloverleaf is increased by the tertiary interaction for the preponderant quantity of sequences. However, about 1/3 of the simulations show converse behaviour. A similar distribution of the results describes the influence of coaxial stacking on the unmodified sequences. Since the fraction of trajectories forming the cloverleaf is mostly very low to begin with, the tertiary interaction seldom brings about a significant improvement. Futhermore, no cloverleaf structure is located among the first hundred local minima in Barriers for a number of unmodified tRNAs, thus preventing a calculation and rendering a comparison impossible. The maximum value of 100% of the trajectories is not found once, in contrast to the modified sequences for which this result occurs 13 times after all.
Summing up, it can be said that the behaviour of the different sets of sequences usually met the expectations: modest thermodynamic stability and moderate ki-

Percent cloverleaf				
	normal		coaxial	
tRNA	mod	unmod	mod	unmod
RA1140	12.6%	2.1%	41.7%	4.3%
RA1180	16.9%	2.2%	18.7%	4.7%
RA1540	24.8%	12.6%	26.3%	7.7%
RA1660	13.2%	12.5%	34.6%	6.6%
RA1661	12.1%	n.v.%	18.0%	8.5%
RA1662	3.3%	2.5%	16.3%	n.v.
RR1140	84.9%	n.v.	92.1%	6.7%
RR1141	16.3%	0.6%	51.1%	2.8%
RR1540	91.2%	17.7%	54.4%	4.1%
RR1660	95.8%	6.3%	82.2%	6.3%
RR1661	90.8%	6.1%	100.0%	9.5%
RR1662	14.0%	12.2%	62.9%	21.9%
RR1663	42.1%	26.8%	38.4%	29.0%
RR1664	100.0%	n.v.	100.0%	7.5%
RN1140	14.3%	0.6%	50.3%	20.2%
RN1660	85.8%	17.7%	77.1%	31.1%
RD1140	33.0%	0.3%	35.3%	n.v.
RD1580	57.1%	33.4%	65.2%	7.8%
RD1660	46.3%	0.2%	70.1%	4.6%
RC1140	3.0%	1.8%	10.5%	1.0%
RC1660	60.0%	17.9%	53.4%	6.4%
RQ1140	7.1%	n.v.	43.1%	7.5%
RQ1660	88.7%	8.8%	96.3%	6.1%
RQ1661	69.2%	11.5%	85.6%	7.3%
RE1140	39.5%	16.0%	11.1%	5.5%
RE1660	27.6%	8.3%	10.3%	19.4%
RE1661	29.3%	1.8%	40.7%	13.7%
RE1662	5.5%	7.4%	50.0%	45.1%
RE2140	25.5%	16.2%	16.1%	6.8%
RG1140	70.5%	26.5%	100.0%	37.7%
RG1180	43.2%	42.7%	44.8%	34.6%
RG1310	55.0%	19.6%	70.8%	19.9%
RG1380	5.7%	1.9%	87.4%	n.v.
RG1381	5.5%	3.7%	3.3%	n.v.
RG1540	94.5%	41.3%	100.0%	32.2%
RG1660	69.4%	32.8%	76.1%	67.4%
RG1661	46.4%	15.2%	67.8%	15.2%
RG1662	20.0%	22.4%	25.5%	21.3%
RG1700	25.6%	33.5%	56.7%	69.0%

Table 10: Effects of modifications and coaxial stacking on the amount of trajectories leading to the cloverleaf. "n.v." indicates the absence of a cloverleaf structure.

Percent cloverleaf				
	normal		coaxial	
tRNA	mod	unmod	mod	unmod
RG1701	21.4%	21.4%	23.2%	21.4%
RH1140	10.1%	33.7%	9.8%	12.2%
RH1660	100.0%	69.7%	100.0%	69.4%
RH1700	100.0%	55.9%	100.0%	64.4%
RI1140	84.0%	31.9%	92.8%	24.1%
RI1141	32.9%	3.0%	33.4%	3.3%
RI1180	35.4%	2.6%	28.9%	3.6%
RI1540	13.2%	14.6%	17.1%	15.1%
RI1580	35.1%	3.4%	45.9%	4.6%
RI1660	16.4%	4.7%	16.2%	6.4%
RI1661	100.0%	4.8%	35.4%	6.6%
RK1140	76.2%	1.6%	100.0%	12.0%
RK1141	87.9%	19.2%	73.1%	15.7%
RK1540	1.6%	1.1%	18.7%	3.5%
RK1541	2.1%	1.8%	20.3%	1.8%
RK1660	75.4%	8.3%	80.4%	16.6%
RL1140	2.7%	7.5%	6.1%	6.0%
RL2020	23.8%	n.v.%	27.7%	n.v.
RL2100	68.1%	23.4%	92.7%	49.6%
RL2101	38.8%	32.7%	33.8%	n.v.
RL1141	5.7%	19.7%	56.5%	25.6%
RL1142	27.5%	39.4%	n.v.	45.0%
RL1460	14.3%	17.1%	32.8%	35.6%
RL1540	3.3%	9.1%	11.7%	n.v.
RL1660	33.8%	54.4%	80.6%	46.6%
RL1661	36.5%	n.v.	n.v.	n.v.
RL1662	77.7%	16.3%	69.5%	20.7%
RL1700	n.v.	n.v.	n.v.	n.v.
RM1140	100.0%	35.2%	99.8%	48.6%
RM1540	34.4%	1.8%	27.4%	4.0%
RM1580	1.4%	5.3%	n.v.	n.v.
RM1660	4.6%	n.v.	100.0%	n.v.
RF1140	46.5%	32.3%	81.1%	24.2%
RF1460	98.5%	no bar	99.0%	no bar
RF1540	74.0%	n.v.	68.5%	n.v.
RF1580	9.7%	1.9%	21.9%	4.1%
RF1660	100.0%	no bar	100.0%	56.6%
RF2020	100.0%	n.v.	98.0%	n.v.
RP1140	68.0%	7.1%	85.3%	7.9%
RP1180	14.0%	6.2%	100.0%	9.5%

table 7 continued

Percent cloverleaf				
	normal		coaxial	
tRNA	mod	unmod	mod	unmod
RP1540	16.4%	0.4%	21.5%	0.4%
RP1700	72.5%	4.2%	80.2%	2.6%
RP1701	86.1%	13.1%	84.7%	n.v.
RP1702	13.2%	2.3%	21.7%	1.6%
RS1140	0.6%	5.1%	19.5%	15.2%
RS1663	100.0%	60.9%	n.v.	2.2%
RS1664	89.1%	35.6%	62.1%	19.6%
RS1141	3.2%	n.v.	n.v.	3.3%
RS1180	5.6%	n.v.	n.v.	2.4%
RS1540	26.8%	5.2%	35.6%	4.8%
RS1541	6.4%	7.2%	42.9%	20.9%
RS1542	2.8%	27.8%	6.6%	26.2%
RS1660	69.5%	65.1%	61.7%	45.5%
RS1661	45.4%	8.0%	3.3%	53.6%
RS1662	n.v.	n.v.	n.v.	0.5%
RT1140	14.4%	8.4%	21.8%	11.8%
RT1141	10.9%	3.6%	10.9%	6.1%
RT1180	8.6%	5.4%	12.2%	6.7%
RT1540	77.1%	10.2%	66.1%	15.6%
RT1660	73.5%	25.6%	71.4%	25.9%
RT1661	12.3%	38.0%	65.3%	16.1%
RW1140	100.0%	23.0%	100.0%	44.0%
RW1141	100.0%	80.4%	64.0%	46.5%
RW1250	76.0%	34.7%	72.8%	27.5%
RW1251	54.1%	20.8%	46.9%	15.8%
RW1540	62.0%	27.2%	60.9%	49.1%
RW1660	100.0%	57.0%	76.6%	n.v.
RY1140	n.v.	9.3%	n.v.	5.8%
RY1460	21.6%	n.v.	8.3%	n.v.
RY1540	7.2%	n.v.	6.9%	3.0%
RY1541	7.2%	n.v.	7.7%	2.5%
RY1660	81.4%	n.v.	100.0%	n.v.
RY1661	100.0%	1.7%	100.0%	3.2%
RV1140	11.0%	7.1%	23.2%	17.4%
RV1180	21.5%	6.7%	36.2%	18.2%
RV1460	83.1%	21.3%	86.4%	23.6%
RV1540	53.1%	5.8%	50.7%	11.1%
RV1660	40.9%	26.3%	47.5%	22.1%
RV1661	99.7%	9.2%	97.0%	28.7%
RV1662	13.5%	3.4%	35.3%	7.5%

table 7 continued

netic folding properties of unmodified sequences are considerably improved by the insertion of modifications, and the intromission of coaxial stacking generally led to a higher frequency of occurrence of the natural conformation. The anticipated enhancement of percentage from unmodified to modified to stacked sequence is demonstrated several times, if not always with such clarity as in the example glycine sequence.

6 Folding properties of tRNAs composed of a restricted alphabet

Its composure of only four building blocks of similar chemical properties presents a debatable counter-argument against the idea of RNA being the carrier of genetic information and possessor of catalytic activity as proposed in the RNA world hypothesis [59–62]. However, RNA has been proven to fulfill a wide variety of catalytic functions [45,55,57,70,74,83,84,134,137,148]. The most recent development of an RNA ligase ribozyme obtained by in vitro evolution demonstrates that RNA catalysis can even be achieved by a three-nucleotide alphabet [118]. The functionality of this ribozyme despite its lack of cytidine might be explained in terms of this base presenting the least stable of the four nucleosides, with a high probability of deamination to uridine. Furthermore, the possibility of forming A-U Watson-Crick and G-U wobble base pairs is preserved. Since folding into a stable secondary and tertiary structure could be accomplished in this case, the question was whether tRNA structure could be successfully predicted by computer experiments, with the sequence constraint of the restricted alphabet adenosine, guanosine and uridine. With the help of RNAinverse, a pool of 1000 AGU sequences folding into the native structure of $tRNA^{phe}$ was generated. Considering the high stability contribution of G-C pairs as opposed to A-U ones, a second set of sequences was created, this time with an adenosine deficiency. Flanked by C-G pairs on both sides, G-U pairs in the middle of a helix hardly lead to any distortion [75], resulting in a thermodynamically more stable structure than in the case of adjacent A-U pairs. The proposed GUC sequences were in fact exceptionally stable, with an average mfe of 37 kcal/mole, several kcal/mole more than for the natural tRNAs (compare with results in section 5.1. The number of sequences was reduced to 746 after exclusion of the ones that had been found several times during RNAinverse search. Out of these unique sequences, nearly 2/3 form the predetermined natural conformation as mfe structure. It is drawn as S1 in the middle of figure 19, surrounded by eight other frequently occurring metastable states acting as potential folding traps. Altogether, 12 such local minima were detected, 6 of them showing cloverleaf structures, only differing in their helix lenghts. They were numbered according to decreasing frequency.
Comparison of sequences able to form a particular structure led to the detection of nucleotides occurring with over 90% frequency, which are printed out in fig-

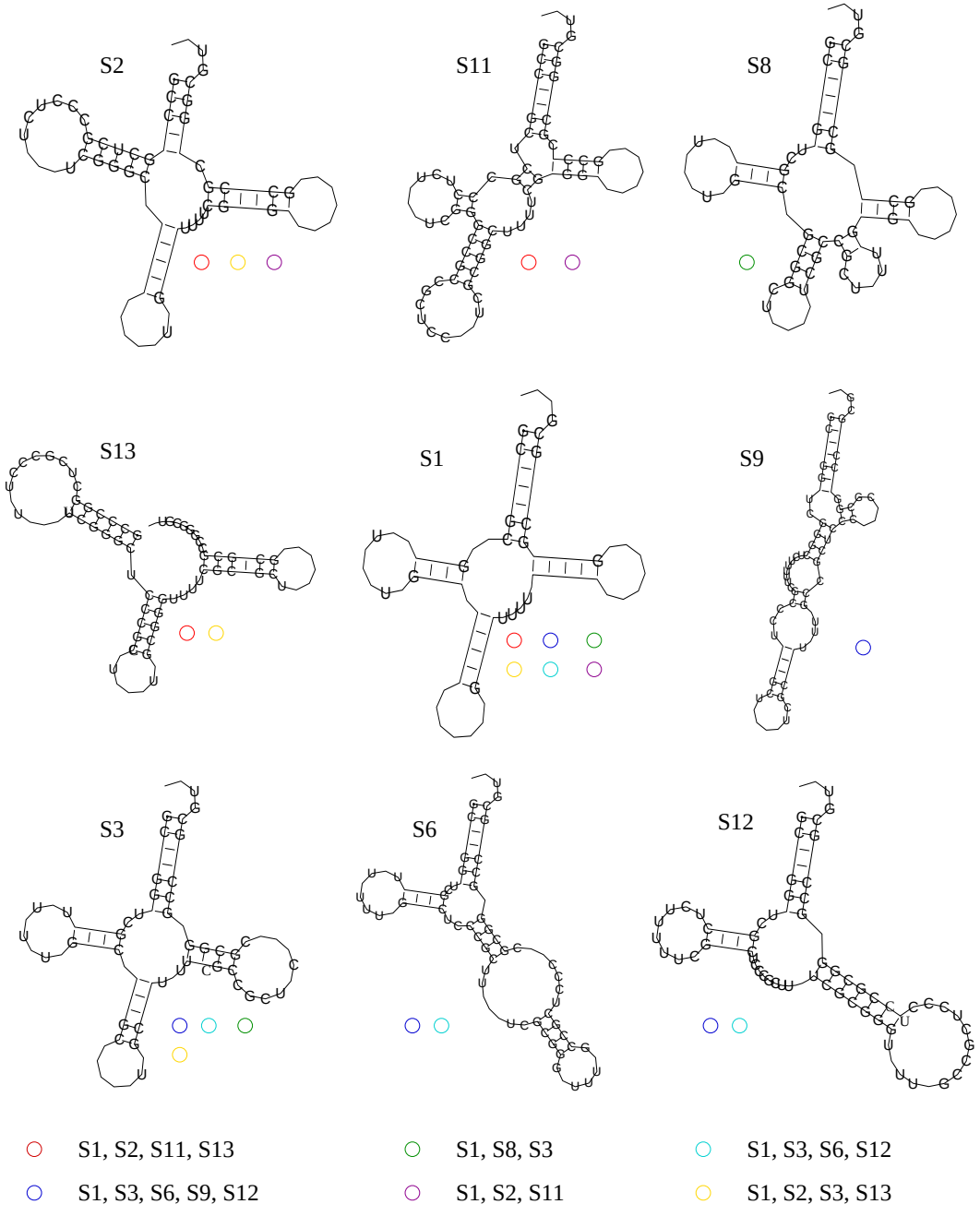


Figure 19: Nine frequently occurring stop structures for CGU sequences. Structures acting as competitors for a particular sequence are arranged into classes with different colour codes.

ure 19 . A surprising number of such "conserved" residues was found, not only in single-stranded stretches, but in helices as well. Maximum conservation is visible in S12 (57-fold), S11 and S13 (55-fold) and S6 (53-fold). Naturally, as the number of fixed bases correlates with the amount of sequences folding into the respective structure, it does not reach a large value for the mfe structure S1 (only 19-fold). Nevertheless, four base pairs of the seven pairs containing acceptor helix are retained (all G-C and C-G pairs), in addition to both closing pairs of the D stem and the innermost pairs of anticodon and Tψ stem, which are all composed of a conserved G and varying probabilities of U and C, with a clear dominance of the latter. Whereas G-U pairs are not frequent, single uridine residues accumulate in loops and junctions, especially in the variable region. This feature is common to all nine structures displayed in figure 19, as is the conservation of the first two pairs in the acceptor stem, with the sole exception of structure S13, which does not form a multiloop and therefore lacks a closing stem altogether. Still, two of three helix regions of this conformation, accounting for a total of 11 base pairs, are entirely conserved. While this presents an exceptional result, at least one stem of a structure is usually fixed. It concerns the D stem in S2, the anticodon stem in S8 and S11, and the Tψ stem in S6, S3 and S12. The last two are also remarkable for the fact that their outermost base pair constitutes a 100% conserved G-U pair. Suggested by the described discoveries, sequence comparison made it clear that several of these stop structures can be organized into groups. Structures within one group act with high probability as competitors for the same sequence. Each of the resulting six groups has a different colour code in figure 19. S1 as central structure participates in all groups which are as follows:

S13, S1, S2, S3
S1, S2, S11, S13
S9, S1, S3, S6, S9
S1, S3, S8

Underlined structures indicate the presence of a subgroup. For example, sequences able to form S13 or S11 always fold into S1 and S2 as well, but the former two conformations never occur together as structure possibilities. Structural similarities justify the division into groups. Among the first three structures, a cloverleaf fold is maintained, with an elongation of both D and Tψ stem on the expense of the acceptor helix in S2, and an enlargement of the Tψ loop in S3. As the remaining structure in this group, S13 combines both of these structural

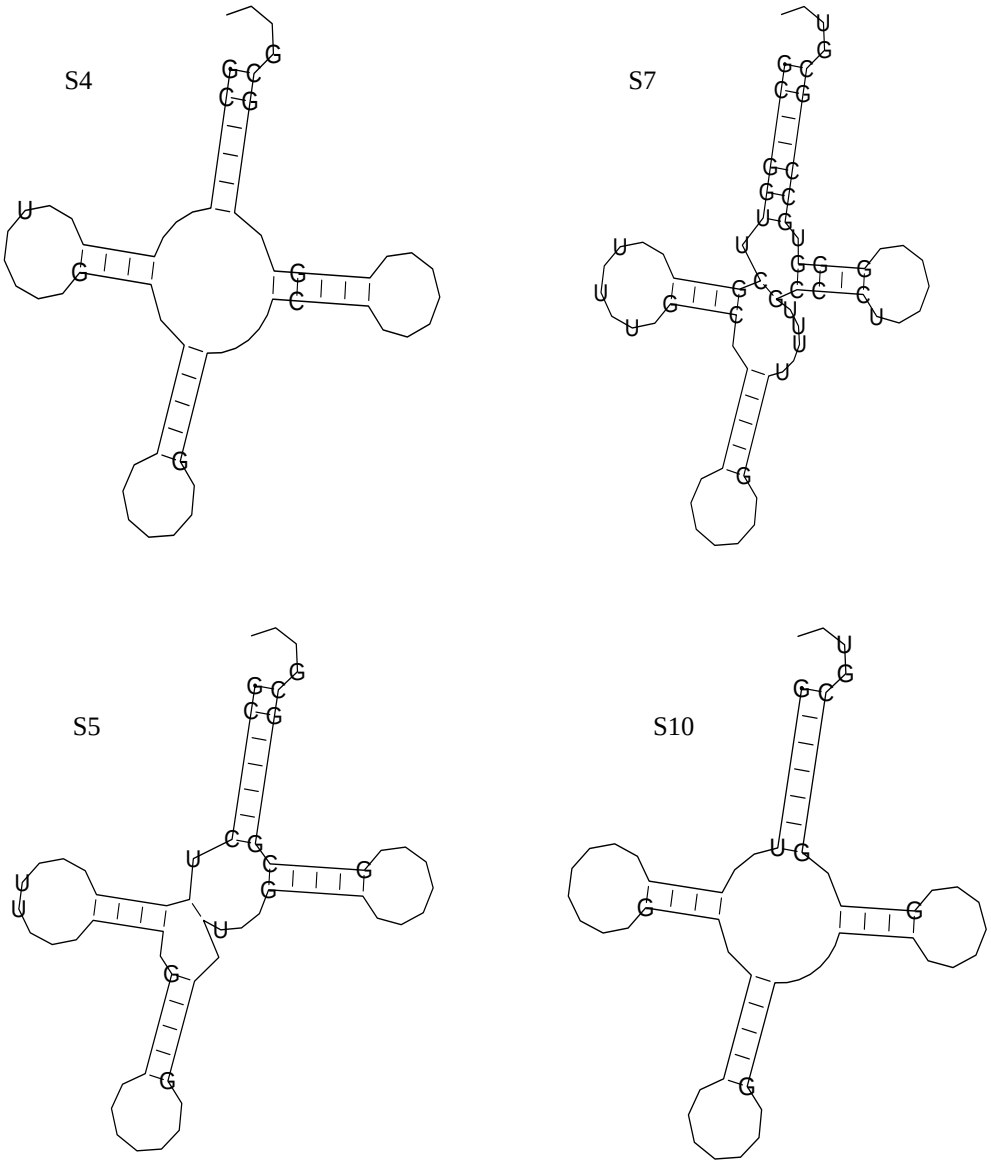


Figure 20: Four frequently occurring stop structures for CGU sequences.

structure	percentage	mft [arb. units]
S1	54.1%	25007.00
S2	10.3%	28912.03
S3	5.7%	9024.12
S4	36.9%	28238.36
S5	62.5%	54181.61
S6	6.2%	2577.84
S7	40.5%	22982.84
S8	12.6%	9239.32
S9	1.5%	1012.49
S10	53.7%	51320.74
S11	10.1%	3622.39
S12	5.4%	7119.24
S13	25.9%	8752.67

Table 11: Folding kinetics of GUC sequences in terms of percentage of trajectories leading to a specified structure and the mean folding time.

features - extension of stems and loop - , which, while leading to an opening of the multiloop, leaves three arms of the cloverleaf intact. As another example, S12 shows considerable conformational resemblance to S6, with which it always appears in combination. In both structures, a fusion of anticodon and Tψ arm takes place, the continuing helix in S12 interrupted only by an additional large internal loop in S6.

For all 13 conformations, table 11 lists the results of kinetic folding simulations, given in terms of the fraction of molecules forming the respective structure and the mean folding time. There is no case for which the trajectories are distributed among all 13 conformations, therefore the given percentages denote the average result for the respective structure. Of all the competitive structures illustrated in figure 19, the mfe structure is actually found with the highest percentage of trajectories, with an average value of 54.1%. The only other structure being formed by a mentionable amount of trajectories is S13 (25.9%). None of the remaining conformations presents a serios rival to the designated ground state.

However, four frequently occurring stop structures still need to be considered. Printed in figure 20, the first two usually appear in combination. Thereby, the cloverleaf fold of S4 which only differs from the one in S1 by the opening of a base pair in the acceptor stem, is formed by a smaller fraction of molecules (namely 37%) than the non-native structure of S7 (40.5%) in which an unusual base pair-

ing between the junction of acceptor and D stem and the variable region divides the multiloop.
The last two conformations are remarkable for the fact that they only appear individually. As still another cloverleaf possibility, a junction is introduced between acceptor and Tψ stem by the latter's opening of a base pair in S10. The presence of a mismatch between helices (also in S4) is favoured by coaxial stacking, which was also pointed out in the previous chapter. On average, 53.7% of the simulations end in this particular cloverleaf conformation. Finally, S5 which bears a strong resemblance to S7, turns out to be the structure dominating the distribution of the folding trajectories. Almost exclusively formed by over 50%, the fraction of folded molecules forming it amounts to 62.5% on average. It also presents the longest mean folding time, though. Taken together, both values indicate that once a molecule has reached this state, refolding into another local minimum is unlikely. A similar situation arises for S10 and S1, which have nearly identical percentages of trajectories leading to them, although the latter conformation is formed with half the folding time. Structures S4 and S7 show similar results for both values of fraction of molecules and mean folding time, the latter of which lies in the range of S1. Surprisingly, all the conformations found by only a minor amount of trajectories (S2, S3, S6, S8, S9, S11, S12) turn out to be formed exceptionally fast which implies a high probability of refolding. In general, a rough correlation between rising quantity of trajectories and mean folding time is observed.
While it seems possible to design CGU sequences with high thermodynamic stabilty and folding efficiently into the tRNA cloverleaf, a number of alternative conformations occurring with similar probability prevent an unequivocal result.

The most striking difference between the adenosine- and the cytosine-deficient set of sequences lies of course in the thermodynamic properties. The difference of the minimum free energy between CGU and AGU sequences is enormous, due to the exchange of G-C pairs for less stable A-U and G-U pairs. The mfe of AGU sequences reaches average values of around only -7 kcal/mole, while the energy gap to the first suboptimal conformation lies around -0.75 kcal/mole . It does not come as a surprise that the predefined cloverleaf (S1 in figure 19) does never even appear as mfe structure. In its place, an alternative natural fold with a shortened acceptor stem presents the ground state for 1/6 of the sequences. Much more frequent, however, is the conformation with an opened terminal stack. Structure

optimization experiments in a flow reactor have confirmed that a single point mutation introducing a mismatch in one base pair is sufficient to open this stem of marginal stability. In our case, an accumulation of less stable A-U base pairs accounts for the frequent absence of this helix. All in all, only six structures occur with remarkable frequency as mfe structures, four of them presenting varying forms of cloverleaves. They are illustrated in figure 21.
As opposed to the CGU sequences, conserved nucleotides are very rare. Only for structures S5 and S6, which occur with minor frequency, a considerable number of positions is fixed (26 and 27, respectively). These are also the only conformations in which adenosine residues, even an A-U base pair, are conserved. In general, entire helices are built of G-U pairs. Consequently and in contrast with the results of GUC sequences, purine nucleotides, especially adenosines, cluster in junctions and loops. Both structures always pose as stop structures in combination with S1, although never together. According to table 12, the amount of

structure	percentage	mft [arb. units]
S1	39.9%	754.22
S2	79.2%	287.58
S3	68.3%	1097.01
S4	82.0%	1211.04
S5	43.6%	1409.61
S6	37.4%	1175.64

Table 12: Folding kinetics of AUG sequences in terms of percentage of trajectories leading to a specified structure and the mean folding time.

trajectories folding into them is moderate (37.4% and 43.6% for S6 and S5, respectively). For sequences able to form S4, on the other hand, this conformation presents the only structure with a reasonably high barrier in most cases, leading to an average percentage of 82.0% of the trajectories. The same is often true for S3 (68.3%). However, set against S2, the structure with the missing acceptor helix clearly dominates the folding distribution (79.2%). For a considerable amount of sequences, both S2 and S1 are suitable structure candidates. The distribution of folding trajectories, however, shows incontestably that S2 weighs more heavily as folding end point. In each case, S2 constitutes the mfe structure, and three times as many molecules fold directly in the conformation lacking the terminal stem as compared to the tRNA native state. This finding is rather unexpected as no previous result indicated a correlation between the occurrence of a confor-

mation as mfe structure and the percentage of trajectories folding into it. The average amount of computer experiments which is composed of all appearances of a particular structure is still twice as much for S2 (79.2%) than for S1 (39.9%) which is thus degraded to a less meaningful conformation. In addition, the maen folding time argues for the preferential formation of S2.

Comparing the two sets of restricted alphabets, the assumption was confirmed that AUG-only sequences can mostly not maintain the predetermined tRNA cloverleaf in respects of either thermodynamic stability or kinetic folding properties. The energy contributions of A-U and G-U pairs do generally not suffice to close the multiloop, and the resulting structure also dominates the distribution of folding trajectories. Commonly, the majority of simulations does not lead to the natural conformation. The chances are somewhat increased by exchanging adenosine for cytosine as the third nucleotide in a restricted alphabet. A line-up of stable G-C pairs in helix regions accounts for excellent thermodynamic properties. The conformation with an opened terminal stack has moved up to a high suboptimal rank and does in no case represent a relevant metastable state. Despite having several potential folding traps as competition, better folding results in terms of both percentage of trajectories and mean folding time were determined for the designated mfe structure.

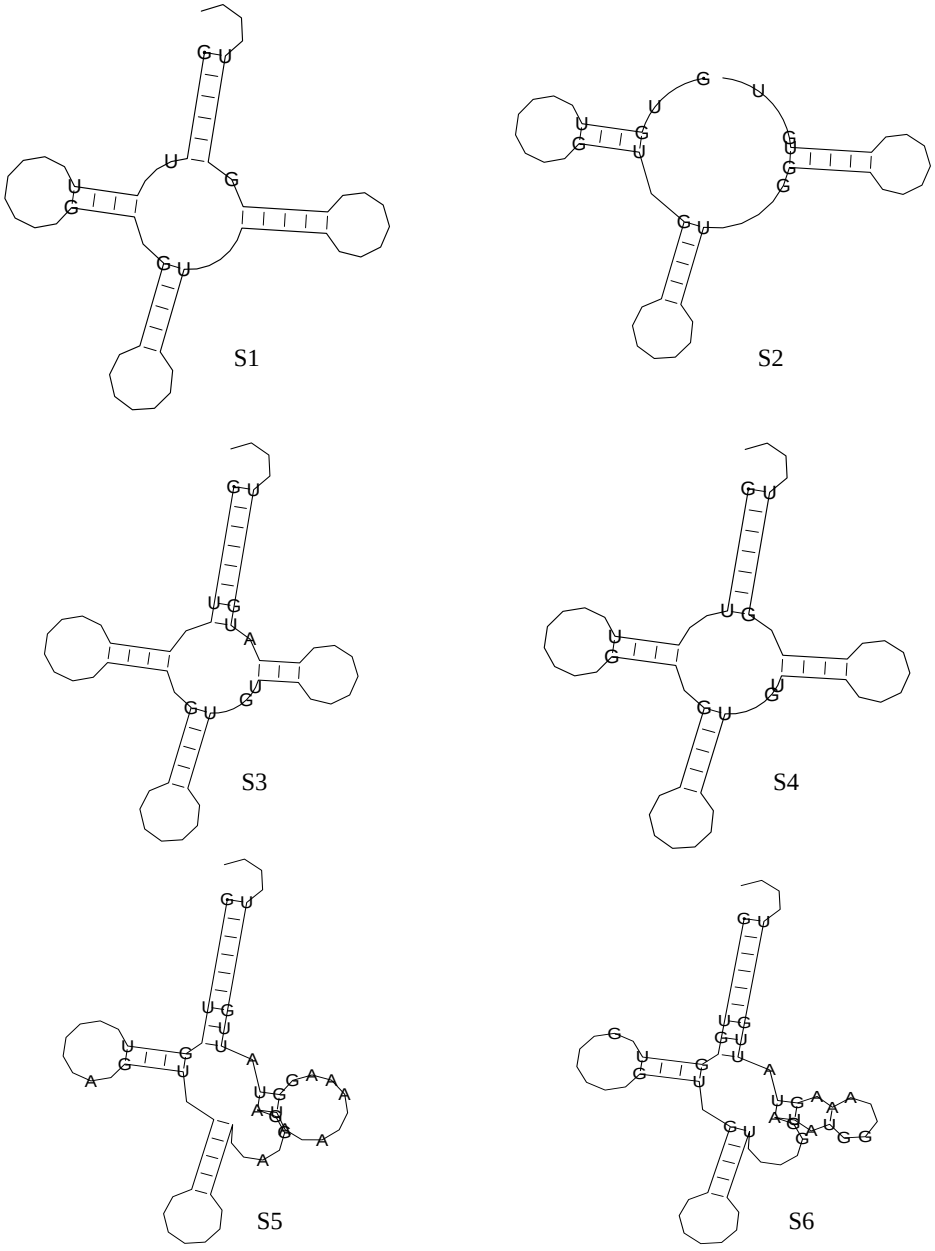


Figure 21: Six frequently occurring stop structures for AGU sequences.

7 Design of non-natural tRNAs

Nowadays, many medical and pharmaceutical problems are solved with the help of molecular biology, requiring the synthesis of new biopolymers, especially RNAs and proteins, with programmable functions.
Molecular recognition related to tRNAs is one of the most interesting themes in molecular biology and has been approached from various directions. Aminoacyl-tRNA synthetases (aaRSs) catalyze the linkage of tRNAs to amino acids and are known to recognize sites that consist only of a few bases of their cognate tRNAs. These sites, which are not necessarily restricted to the anticodon loop and the acceptor stem, are called 'identities'. To fully understand the mechanism of this recognition, it is indispensable to detect identities in various kinds of tRNAs. Through modification of tRNAs and aminoacyl tRNA synthetases, a reprogramming of the ribosomal protein synthesis shall come into reach. While some of these new molecules will be charged with non-natural amino acids by the respective aaRS, others are aimed to inhibit tRNA-recognising enzymes. The reasoning of this proposition is supported by the fact that the first steps towards a site-specific incorporation of non-natural amino acids into proteins *in vivo* have already been successfullly performed [80, 81]. The specific combination of rational design based on findings in structural biology with evolutionary biotechnology seems to be a promising approach for this purpose.
In the course of this work, we aim to make suggestions for sequences that have been subjected to several essential constraints.

7.1 Strategy for sequence proposals

Reasonable sequence proposals can only be made by chosing from a very large pool of possible variants and carefully excluding those not fulfilling the imposed constraints. With the purpose of reducing an initially determined amount of RNA sequences to a reasonable number of potential candidates for synthesis, the realization of this work can be divided into several steps which are as follows:

1. **Generation of sequences**
   At the beginning, a very great deal of sequences folding into given structures are produced with the help of the RNAinverse program.

2. **Selection of sequences with regard to stability and foldability**
   The large number of sequences is reduced with the help of these two cri-

teria which are essential for the reliability of structure prediction. Using the RNAsubopt program, **all** suboptimal conformations lying within a predefined range above the ground state are computed and rated. Sequences forming thermally unstable structures because of suboptimal states with low energies are therefore being ruled out.
Foldability is the second vital criterion for structure prediction. A variety of sequences forming thermodynamically stable structures are sorted out due to inefficient folding behaviour.

3. **Consideration of tertiary interactions and introduction of special functions**
Specific nucleotides at certain positions will be indispensable for special functions and cannot be switched without losing those. The complete sets of identity elements are known for several aminoacylation systems in both *E. coli* and *S. cerevisiae* [41] and will be used as additional constraints.

7.2 Selection of sequences

Cloverleaf structures of several *E. coli* tRNAs whose identity elements have been experimentally determined were used as targets for RNAinverse search, with these nucleotides specified as sequence constraints. Good results were obtained by a test of the $tRNA^{phe}$ system, which will be cited as an example in the following. As shown in figure 22, 13 fixed determinants shall ensure interaction with the cognate synthetase. The anticodon area presents a major recognition site; apart from the anticodon triplet, two base pairs in the stem and the two adjacent residues in the variable region serve as identity elements. The set is complemented by uridines in the D and Tψ loop, and the mostly crucial discriminator base. They obviously represent a strong constraint, because the search for 1000 sequences yielded only 424 unique ones. These potential tRNAs, though, readily adopt the predefined cloverleaf from a thermodynamic point of view. No less than 387 form the wanted mfe structure which is clearly separated from suboptimal states in terms of energy (average $\Delta G = 3$ kcal/mole). This result is indicative of the feasibilty of designing thermodynamically very stable sequences (in comparison to the findings on natural tRNAs, for which the energy gap amounted to -1 kcal/mole, see section 5).

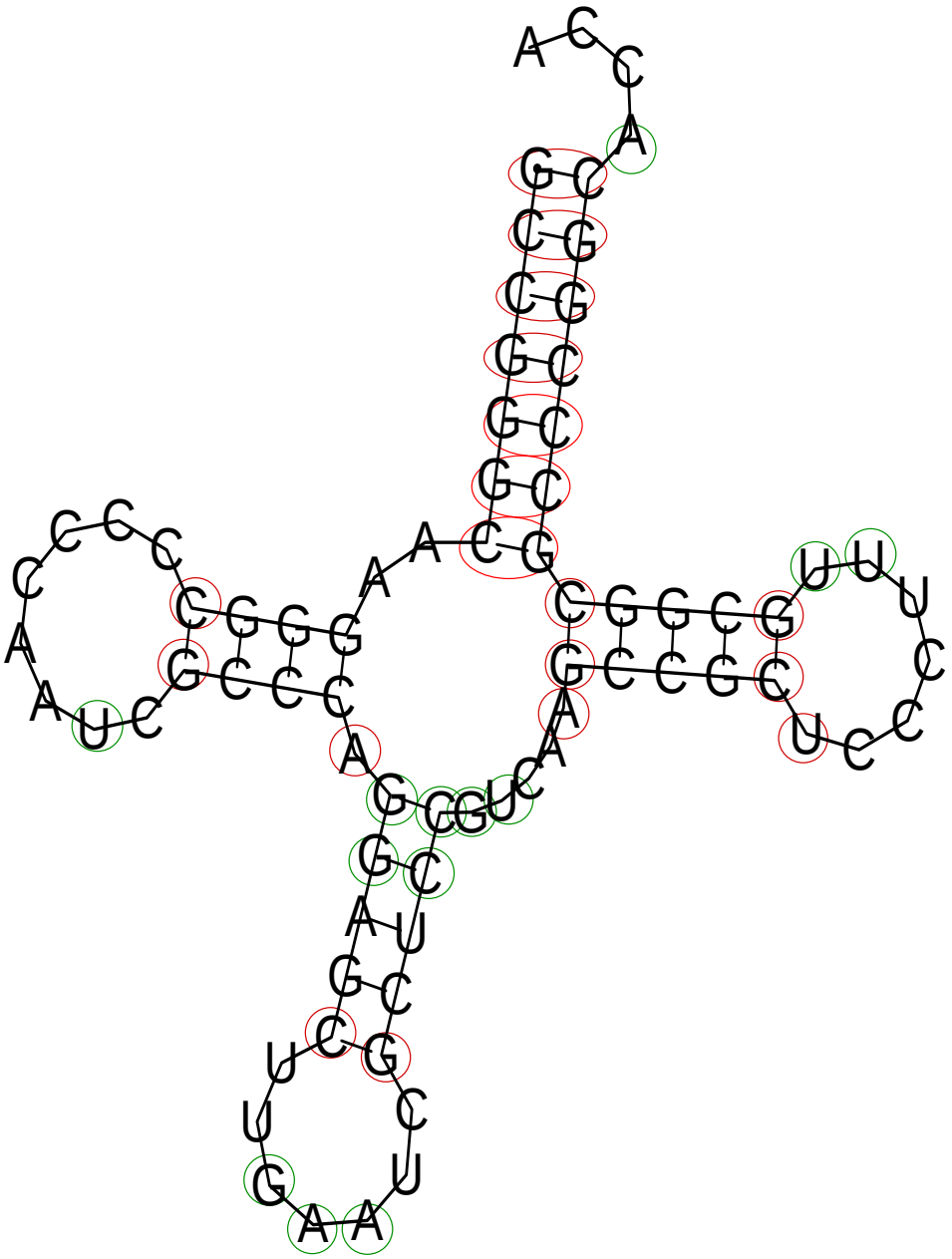


Figure 22: Sequence proposal for an artificial $tRNA^{phe}$ sequence: identity determinants are circled in green. Additional conserved nucleotides essential for excellent kinetic folding properties are marked in red.

S	structure	perc.	occ.
S1	(((((((..((((........)))).(((((.......))))).....(((((.......))))))))))))....	58.33%	387
S2	....((...((((........)))).(((((.......))))).(((((.(((((.......))))).))))...)))).)).	15.60%	63
S3	...(((((.((((........)))).(((((.......))))).....(((((.......))))).....))).)).	9.52%	47
S4	.....(((((((.........)))).(((((.......))))).....(((((.......))))).....))))....	6.03%	35
S5	(((((((..((((........)))).(((((.......)))))))...)))).....(((((....)))))...	35.27%	29
S6	(((((((..((((........)))).((.(((((....((.((.....)).)).....))))).)))))))))....	7.19%	24
S7	((((((...((((........)))).(((((.......))))).....(((((.......))))).)))))))....	54.20%	23

Table 13: Average percentage of trajectories leading to the mfe structure S1 and six minor relevant meta-stable states.

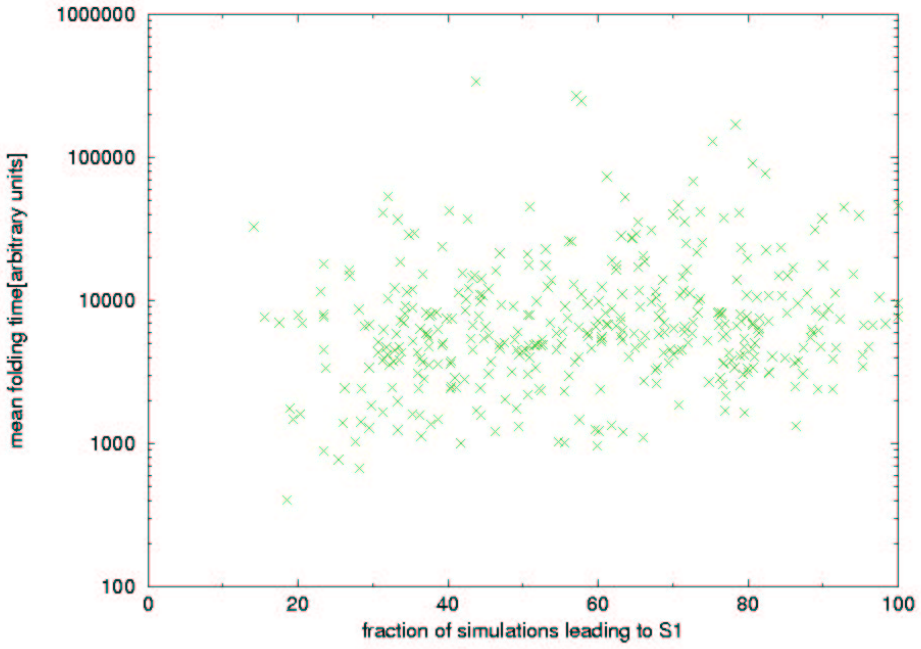


Figure 23: Folding kinetics of artificial $tRNA^{phe}$ sequences. No correlation is found between the percentage of trajectories leading to the cloverleaf and the mean folding time.

Kinetic simulations revealed six other meta-stable states (see table 13) on the energy landscape which might act as a folding trap, although they never all together present structure possibilities for the sequence under consideration. Table 13 lists their respective number of occurrences (for 424 sequences) and the resultant average percentage of trajectories folding into them. They appear much less frequent than the predetermined conformation, therefore the good average percentage of 54.20% for the alternative cloverleaf S7 does not render it a serious competition for S1. The results prove that they are all of minor importance. The cloverleaf structure S1 notes down the major amount of trajectories in most cases. A plot of the fraction of folded molecules forming S1 against the mean folding time (figure 23) points out that no correlation exists between these two kinetic properties. A diagram contrasting the minimum free energy and the percentage of trajectories is similarly uncorrelated in appearance. For excellent folding ability (over 90% of the folding trajectories), certain nucleotides at specific positions were discovered which obviously help to stabilize the fold. Nucleotides with over 90% frequency are circled in red in figure 22. The entire acceptor stem is conserved, in addition to the closing pairs of the D, the antocdon and the Tψ loop. The first base pair of the Tψ helix is also fixed. In each case, a G/C combination is concerned. Further essential residues include an adenosine in the D and anticodon arm junction and the variable region, respectively, as well as a uridine in the Tψ loop. The number of potential candidates was reduced to just 19 variants according to suitable folding kinetics. The respective sequences present promising proposals for synthesis.

7.3 Conversion of *in silico* sequences to *in vitro* transcripts

With the assurance of a collaboration with the Strasbourg team in mind which concentrates on the yeast $tRNA^{asp}$ system and is responsible for almost all of the data on this system [39, 109], we took the $tRNA^{asp}$ cloverleaf as input structure and specified the set of identity elements as sequence constraints. After having subjected the resulting sequences to the foldability criterium, the number of possible variants was drastically reduced. Figure 24 shows the four candidates for synthesis. All bases are randomly chosen except for the seven identity elements which are conserved in all sequences. They comprise the first base pair G10-U25 in the D stem, the anticodon GUC, in addition to C38 and the discriminator base G73. Tertiary interactions play of course a significant role in the stabilization of the folded structure and contribute to specific recognition and aminoacylation by the synthetase. However, we first wanted to test the charging without the influence of tertiary interactions, therefore all tertiary contacts present in these sequences were not demanded, but arose by chance. Variant inv13 for example shows a possible trans Hoogsteen pair between U8 and A14 (also in inv196), and a trans Watson-Crick pair between A15 and U48 could form in inv114. The rather high amount of G-C base pairs is due to the strong stability constraint. There are almost no A-U pairs found in the helix regions. The loops, on the other hand, are very A rich. Each of the sequences has a rather impressive amount of percentage of trajectories finding the cloverleaf, ranging from 81% to 93%.

With these proposals, I went to Strasbourg where I had been kindly invited to in order to experience the experimental side and be a part of the transformation of in silico sequences to in vitro transcripts. We went through the following steps during synthesis. First, we ordered DNA oligos for our sequences which have a promotor for T7 RNA polymerase and for the tRNA gene to be transcribed. After several rounds of amplification by means of PCR, we had to exclude one of our sequences (inv138) from further experiments because other side products apart from our 95 nucleotide DNA (75 for $tRNA^{asp}$ and 20 for the 5 end primer) with different lengths were visible. Next, we started the in vitro transcription with T7 RNA polymerase on our PCR samples, followed by a 12% PAGE (polyacrylamid gel electrophoresis) for purification and separation of our transcripts from non-incorporated nucleotides and other products. The final step concerned the interesting question whether these tRNAs could actually be aminoacylated. The charging can be followed by adding radioactive aspartate (^3H) and observing

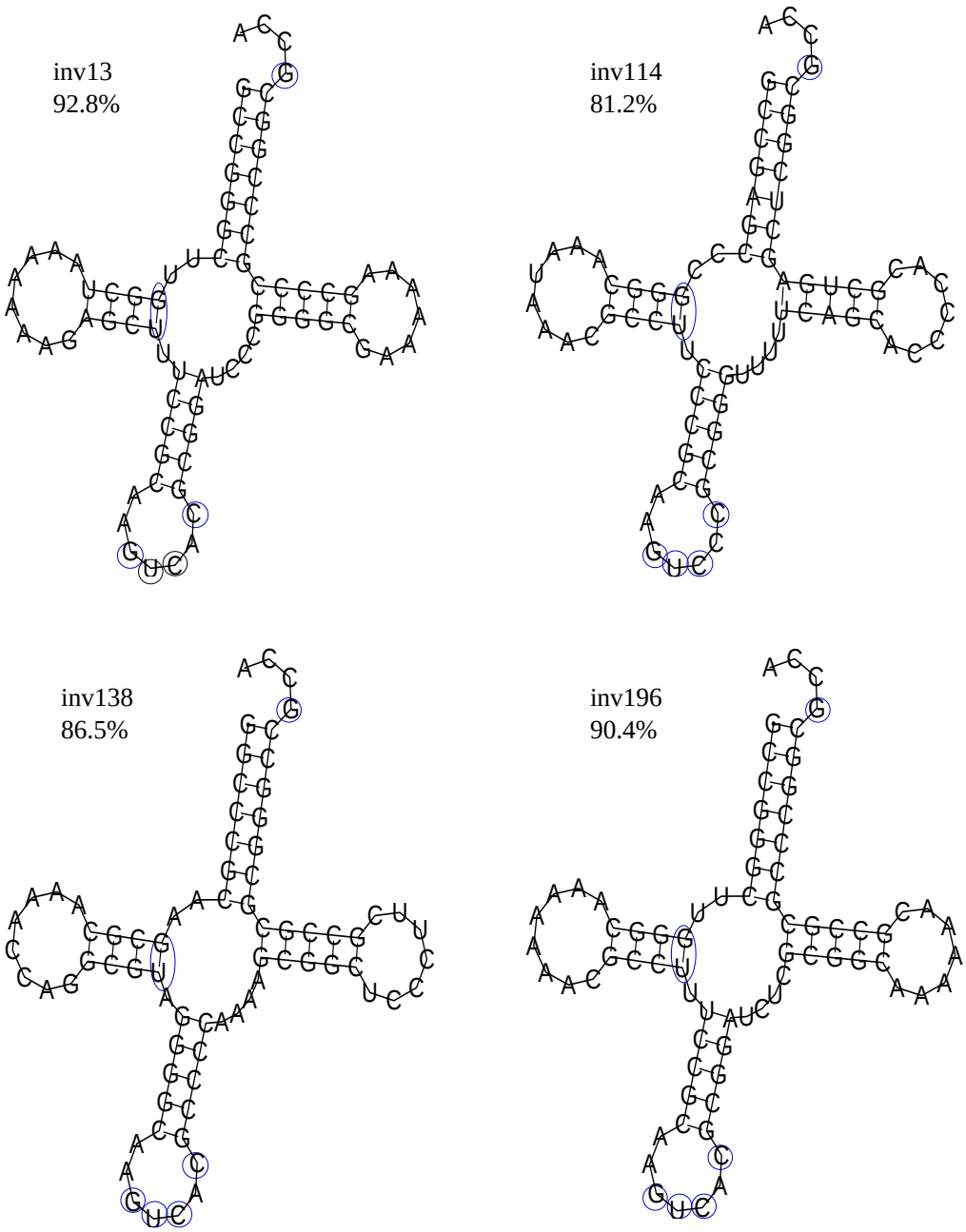


Figure 24: Four sequences folding efficiently into the $tRNA^{asp}$ cloverleaf. Identity elements are circled in blue.

ID:H3 1 MIN

USER: 4 COMMENT:

PRESET TIME : 1.00

DATA CALC : CPM H# : NO SAMPLE REPEATS:

COUNT BLANK : NO IC# : NO REPLICATES :

TWO PHASE : NO AQC : NO CYCLE REPEATS :

SCINTILLATOR: LIQUID LUMEX: NO LOW SAMPLE REJ:

LOW LEVEL : NO HALF LIFE CORRECTION DATE:

ISOTOPE 1: 3H %ERROR: 0.00 FACTOR: 1.00000

SAM NO	POS	TIME MIN	3H CPM	3H %ERROR	LUMEX %	ELAPSED TIME
1	**-1	1.00	103778.0	0.62	0.01	1.31
2	**-2	1.00	89518.02	0.67	0.01	2.70
3	**-3	1.00	152.00	16.22	0.61	4.04
4	**-4	1.00	124.00	17.96	0.32	5.42
5	**-5	1.00	183.00	14.78	0.22	6.77
6	**-6	1.00	186.00	14.66	0.13	8.12
7	**-7	1.00	9436.01	2.06	0.01	9.49
8	**-8	1.00	9240.01	2.08	0.01	10.84
9	**-9	1.00	8817.01	2.13	0.01	12.21
10	**-10	1.00	8815.01	2.13	0.01	13.59
11	**-11	1.00	8669.01	2.15	0.02	14.95
12	**-12	1.00	9104.02	2.10	0.02	16.31
13	**-1	1.00	9872.02	2.01	0.01	17.79
14	**-2	1.00	9250.02	2.08	0.01	19.16
15	**-3	1.00	4984.01	2.83	0.02	20.53
16	**-4	1.00	5830.01	2.62	0.01	21.91
17	**-5	1.00	6243.01	2.53	0.01	23.28
18	**-6	1.00	6120.02	2.56	0.01	24.64
19	**-7	1.00	168.00	15.43	0.13	26.02
20	**-8	1.00	174.00	15.16	0.20	27.39
21	**-9	1.00	211.00	13.77	0.20	28.74
22	**-10	1.00	210.00	13.80	0.13	30.12
23	**-11	1.00	214.00	13.67	0.18	31.47
24	**-12	1.00	240.00	12.91	0.24	32.84
25	**-1	1.00	235.00	13.05	0.14	34.32
26	**-2	1.00	235.00	13.05	0.25	35.69
39	**-3	1.00	182.00	14.82	0.23	44.19
40	**-4	1.00	262.00	12.36	0.21	45.57
41	**-5	1.00	195.00	14.32	0.34	46.94
42	**-6	1.00	222.00	13.42	0.24	48.29
43	**-7	1.00	166.00	15.52	0.30	49.67
44	**-8	1.00	208.00	13.87	0.73	51.04
45	**-9	1.00	200.00	14.14	0.29	52.41
46	**-10	1.00	257.00	12.48	0.39	53.79
47	**-11	1.00	167.00	15.48	0.28	55.16
48	**-12	1.00	237.00	12.99	0.46	56.53
49	**-1	1.00	226.00	13.30	0.12	58.01
50	**-2	1.00	173.00	15.21	0.41	59.37

Figure 25: Result of aminoacylation assay.

the incorporation. Figure 25 shows the result. Each sample was measured four times with the timespan of one minute. Apart from my inverse sequences, we took the modified $tRNA^{asp}$ as well as a PCR sample of it which had been subjected to the same conditions as my sequences as control. The samples marked with K contained no tRNA. We received an average of 9000 cpm for the wild-type as well as for the PCR sample which is the normal value for the natural tRNA. Unfortunately, cpm values for all inverse sequences are in the order of magnitude of the control K, indicating that these tRNAs could not be aminoacylated at all. The threshold value for aspartylation lies around 3500 cpm.
This result is regrettable, but does not come as a real surprise if one considers that we were not dealing with transcripts with only two or three mutations (of which a number have already been analyzed), but with sequences which have at least a hamming distance of about 30 to the wild-type $tRNA^{asp}$.

8 Conclusion and outlook

RNA performs a remarkable range of functions in all cells. Its versatility is intimately connected to its capacity to form elaborate spatial structures. Secondary structures present a convenient model of coarse-graining which allows for application of efficient dynamic programming algorithms. Apart from thermodynamic predictions, information on folding pathways and kinetics of secondary structure formation can be derived by means of computer programs restricting the observation time to a finite value. The necessity to fold sufficiently fast leads to a clear distinction of such kinetically controlled structures from thermodynamically most stable conformations.

The kinetic folding algorithm developed in our research group [35] simulates the folding process in dependence of the time coordinate and yields information about the folding time (which in this context refers to the first passage time from some initial state to the ground state). In addition, it presents the possibility of refolding, for example for computing the folding time for a molecule to reach the minimum free energy (mfe) structure after it got trapped in a metastable state. A compilation of such potential "kinetic traps" is provided by the "barriers" algorithm which makes use of complete suboptimal folding [145] in order to calculate local minima of a given sequence and shows their connection in a practical representation of the energy landscape, the barrier tree.

The cloverleaf of tRNAs is a commonly accepted and experimentally verified fact. A detailed analysis of eubacterial natural tRNAs confirmed the predominance of this conformation as mfe structure in thermodynamic calculations. Kinetic simulations, however, demonstrated that natural folding pathways seldom end in the mfe state. Instead, the conformation space is usually divided into basins of attraction for several dominant structures in addition to the mfe structure, which is sometimes reached by only a small fraction of folding molecules. Refolding from metastable states into the ground state involved crossing of large energy barriers due to breaking and reforming of entire helices, and was seldom accomplished within the time-scale of the simulation. Implementation of coaxial stacking into the folding algorithm resulted in an expected additional stabilization of the multiloop and an increased percentage of trajectories folding into the natural conformation for modified sequences.

The effect is much less pronounced for unmodified tRNAs. Apart from a decrease of the free energy difference between the ground state and the first suboptimal

conformation, the lack of base modifications led to an enlarged amount of structures otherwise prevented to form, several of which offered alternative and more efficient pathways.
While the comparison between modified and unmodified sequences illustrated that the regard of modified nucleotides brings on a definite improvement of both thermodynamic and kinetic folding properties, a fact which arose most probably from complex evolutionary processes, it is not permissable to infer kinetic folding behaviour from thermodynamic calculations for the sequence under consideration.
As a direct consequence, the design of artificial tRNA sequences with predetermined properties presents a challenging attempt. Transfer RNAs play an essential role in the accurate decoding and translation of genetic information into amino acid sequences. The correct aminoacylation of a given tRNA with its cognate amino acid is accomplished by the aminoacyl-tRNA-synthetases which have to discriminate between the different tRNAs despite their similar secondary and tertiary structures. This distinction is achievd by means of molecular signals, so-called identity determinants, in tRNAs which trigger specific aminoacylation. Thermodynamic stability and kinetic foldability are with certainty two basic requirements for non-natural sequences ever hoping to fulfill their destined functions. Application of the first criterion to a large pool of possible candidates predetermined to fold into the tRNA cloverleaf hardly ruled out any of the sequences. The condition of excellent kinetic folding ability, on the other hand, led to a considerable reduction of potential variants. While it seems therefore rather uncomplicated to generate extraordinarily stable sequences, the kinetic folding constraint is not that easily satisfied - at least not simultaneously. Individual results indicate that an improvement of one quality can be reached by relaxing the constraint on the other. However, in order to draw definite conclusions regarding their correlation, further investigations need to be carried out.
Apart from fulfilling these two criteria essential for the reliabilty of structure prediction, the remaining sequences were analysed with regards to specific nucleotides at certain positions which may be indispensable for special functions.
For several *E. coli* systems of which the principles of molecular recognition are unravelled and the complete sets of identity elements are determined, the latter were considered as additional constraints.
The decisive test for *in silico* sequence proposals, the question of aminoacylation capability, was performed on the *S. cerevisiae* $tRNA^{asp}$ system in collaboration

with the group of Richard Giegé in Strasbourg. The synthesis of four artificial sequences folding efficiently into the $tRNA^{asp}$ cloverleaf was successfully achieved by means of *in vitro* transcription with T7 RNA polymerase. The aminoacylation assay, however, clearly indicated that no charging of the non-natural tRNAs took place.

This result can be explained by the considerable number of mutations introduced in comparison to native $tRNA^{asp}$ or the absence of nucleotides participating in tertiary contacts with the synthetase. Tertiary interactions do play a significant role in recognition and aminoacylation. The Strasbourg team has already synthesized and tested a large number of mutants specifically designed to disrupt the tertiary structure and has observed negative effects on aspartylation activity in most cases [110]. Consequently, we have analyzed another set of inverse sequences, this time containing all nucleotides involved in tertiary contacts as search constraints, in addition to the complete set of identity elements (which accounts for a total of 19 fixed positions during RNAinverse search). Under these new conditions, we expect to continue our experiments in the near future.

Another possible reason for the lack of aspartylation activity could be due to the little conformational flexibilty of the tRNA, caused by the strong stability constraint. A certain amount of deformability is obviously required to allow tRNA to fit into the active site of the synthetase. Reduction of the stabilty criterion should lessen the structural rigidity and will also be considered in future experiments.

From a different perspective, the observed result can be interpreted as a starting point in the design of RNA domains specifically meant to block tRNA-recognising enzymes. The development of inhibition strategies of aminoacylation systems could provide access to a way for the synthesis of new RNA antibiotics and antiviral strategies.

Appendix

All tRNA sequences were taken from the Compilation of tRNA sequences and tRNA genes (Bayreuth). The sequence number codes as follows: first letter is D or R for DNA or RNA, respectively. Second letter gives the one-letter symbol of the amino acid. The four-digit number stands for organism and isoacceptor. Modified nucleotides apart from dihydrouridine (D), pseudouridine (P), ribosylthymine (T) or inosine (I) are replaced with M in case of methylations and N in any other.

```
> RA1140 [MYCOPLASMA CAPRIC.] [EUBACT] [76]
GGGCCCUNAGCUCAGCDGGGAGAGCACCUGCCUUGCMCGCAGGGGMUCGACGGUPCGAUCCCGUUAGGGUCCACCA
> RA1180 [MYCOPLASMA MYCOID.] [EUBACT] [76]
GGGCCCUUAGCUCAGCDGGGAGAGCACCUGCCUUGCMCGCAGGGGMUCGACGGUUCGAUCCCGUUAGGGUCCACCA
> RA1540 [BACILLUS SUBTILIS] [EUBACT] [76]
GGAGCCUUAGCUCAGCDGGGAGAGCGCCUGCUUMGCMCGCAGGAGMUCAGCGGTPCGAUCCCGCUAGGCUCCACCA
> RA1660 [E. COLI] [EUBACT] [76]
GGGGCUANAGCUCAGCDGGGAGAGCGCUUGCAUGGCAUGCAAGAGMUCAGCGGTPCGAUCCCGCUUAGCUCCACCA
> RA1661 [E. COLI] [EUBACT] [76]
GGGGGCANAGCUCAGCDGGGAGAGCGCCUGCUUNGCACGCAGGAGMUCUGCGGTPCGAUCCCGCGCGCUCCCACCA
> RA1662 [E. COLI] [EUBACT] [76]
GGGGCUAUAGCUCAGCDGGGAGAGCGCCUGCUUNGCACGCAGGAGMUCUGCGGTPCGAUCCCGCAUAGCUCCACCA
> RR1140 [MYCOPLASMA CAPRIC.] [EUBACT] [77]
GCGCCCGNAGAUCAAUDGGADAGAUCGCUUGACUICGMAPCAAAAGMUUGGGGGUPCGAGUCCCUCCGGGCGCACCA
> RR1141 [MYCOPLASMA CAPRIC.] [EUBACT] [77]
GCCCAUGUAGCUCAGUAGGADAGAGCACGCGCCUMCUNAGCGUGAGMUCGGAAGUPCGAGCCUUCUCGUGGGCACCA
> RR1540 [BACILLUS SUBTILIS] [EUBACT] [76]
GCGCCCGUAGCUCAAUGGADAGAGCGUUUGACUICGMAUCAAAAGMUUAGGGGTPCGACUCCCCUCGGGCGCGCCA
> RR1660 [E. COLI] [EUBACT] [76]
GCAUCCGNAGCUCAGCDGGDAGAGUACUCGGTUICGMACCGAGCGMNCGGAGGTPCGAAUCCUCCCGGAUGCACCA
> RR1661 [E. COLI] [EUBACT] [77]
GCAUCCGNAGCUCAGCDGGADAGAGUACUCGGCUICGMACCGAGCGMNCGGAGGTPCGAAUCCUCCCGGAUGCACCA
> RR1662 [E. COLI] [EUBACT] [75]
GUCCUCUUAGUUAAAUGGADAUAACGAGCCCTUMCUNAGGGCUAAUUGCAGGTPCGAUUCCUGCAGGGGACACCA
> RR1663 [E. COLI] [EUBACT] [77]
GCGCCCUUAGCUCAGUUGGAUAGAGCAACGACTUMCUNAGPCGUGGGCCGCAGGTPCGAAUCCUGCAGGGCGCGCCA
> RR1664 [E. COLI] [EUBACT] [77]
GCGCCCGUAGCUCAGCDGGADAGAGCGCUGCCTUCCGMAGGCAGAGMUCUCAGGTPCGAAUCCUGUCGGGCGCGCCA
> RN1140 [MYCOPLASMA CAPRIC.] [EUBACT] [76]
GGCUUUUNAGCUCAGCAGGDAGAGCAACCGGCUGUUNACCGGUUUMUCACAGGUPCGAGCCCUGUAAAAGCCGCCA
> RN1660 [E. COLI] [EUBACT] [76]
UCCUCUGNAGUUCAGDCGGDAGAACGGCGGACUQUUNAPCCGUAUMUCACUGGTPCGAGUCCAGUCAGAGGAGCCA
> RN1720 [AZOSPIR. LIPO.] [EUBACT] [75]
UUCACAGUAGCUCAGUGGDAGAGCUAUCGGCUQUUNACCGAUCGAUCGUAGGTPCGAGUCCUACCUGUGAAUCCA
> RN1721 [AZOSPIR. LIPO.] [EUBACT] [75]
UUCACAGUCGCUCAGUGGDAGAGCUAUCGGCUQUUNACCGAUCGAUCGUAGGTPCGAGUCCUACCUGUGAAUCCA
> RD1140 [MYCOPLASMA CAPRIC.] [EUBACT] [77]
GGCCCCANAGCGAAGUDGGDDAUCGCGCCUCCCUGUCACGGAGGAGAUCACGGGUPCGAGUCCCGUUGGGGUCGCCA
> RD1580 [THERMUS THERMOPHI.] [EUBACT] [77]
GGCCCCGNGGUGPAGUUMGDDAACACACCCGCCUGUCACGPGGGAGAUCGCGGGMPCGMGUCCCGUCGGGGCCGCCA
> RD1660 [E. COLI] [EUBACT] [77]
GGAGCGGNAGUUCAGDCGGDDAGAAUACCUGCCUQUCMCGCAGGGGMUCGCGGGTPCGAGUCCCGPCCGUUCCGCCA
> RC1140 [MYCOPLASMA CAPRIC.] [EUBACT] [75]
GGCAACANGGCCAAGCGGCDAMGGCAUGGGUCUGCAMCACCCUGAUCAUCGGUPCGAAUCCGAUUGUUGCCUCCA
> RC1660 [E. COLI] [EUBACT] [74]
```

GGCGCGUNAACAAAGCGGDDAUGUAGCGGAPUGCAMAPCCGUCUAGUCCGGTPCGACUCCGGAACGCGCCUCCA
> RQ1140 [MYCOPLASMA CAPRIC.] [EUBACT] [75]
UGGGCUANAGCCAAGCGGDAMGGCAAGGGACUMUGMCUCCCUCAUGCGCCGGUPCGAAUCCUGCUAGCCCAACCA
> RQ1660 [E. COLI] [EUBACT] [75]
UGGGGUANCGCCAAGCMGDAAGGCACCGGAMUCUGMPPCCGGCAUUCCGAGGTPCGAAUCCUCGUACCCCAGCCA
> RQ1661 [E. COLI] [EUBACT] [75]
UGGGGUANCGCCAAGCMGDAAGGCACCGGUMUNUGMPACCGGCAUUCCCUGGTPCGAAUCCAGGUACCCCAGCCA
> RE1140 [MYCOPLASMA CAPRIC.] [EUBACT] [76]
GGCCUGUUGGUGAAGCGGDDAMCACACACGGUUMUCAUCCGUGGACACACGGGUPCGAACCCCGUACAGGCUACCA
> RE1660 [E. COLI] [EUBACT] [76]
GUCCCCUUCGUCPAGAGGCCCAGGACACCGCCCUMUCMCGGCGGUAACAGGGGTPCGAAUCCCCUGGGGGACGCCA
> RE1661 [E. COLI] [EUBACT] [76]
GUCCCCUUCGUCPAGAGGCCCAGGACACCGCCCUMUCMCGGCGGUAACAGGGGTPCGAAUCCCCUAGGGGACGCCA
> RE1662 [E. COLI] [EUBACT] [75]
GUCCCCUUCGUCPAGAGGCCAGGACACCGCCCUMUCMCGGCGGUAACAGGGGTPCGAAUCCCCUAGGGGACGCCA
> RE2140 [SYNECHOCYSTIS SP.] [EUBACT] [76]
GCCCCCAUCGUCUAGAGGCCDAGGACACCUCCCCUNUCACGGAGGCGACAGGGATPCGAAUUCCCUUGGGGGUACCA
> RG1140 [MYCOPLASMA CAPRIC.] [EUBACT] [74]
GCAGGUGNAGUUUAAUGGDAGAACUUCAGCCUUCCMAGCUGAUUGUGAGGGUPCGAUUCCCUUCACCUGCUCCA
> RG1180 [MYCOPLASMA MYCOID.] [EUBACT] [74]
GCAGGUGNAGUUUAAUGGCAGAACUUCAGCCUUCCMAGCUGAUUGUGAGGGUPCGAUUCCCUUCACCUGCUCCA
> RG1310 [STREPTOMYCES COEL.] [EUBACT] [74]
GCGGGUGUAGUUCAAUGGDAGAACAUCAGCUUCCCAAGCUGAGAGCGCGAGTPCGAUUCUCGUCACCCGCUCCA
> RG1380 [STAPHYLOCOC. EPID.] [EUBACT] [75]
GCGGGAGNAUUUCAACUUUDAGAAUACGUUCCUUCCCGGAACGAGAUAUAGGUGCAAAUCCUAUCUUCCGCUCCA
> RG1381 [STAPHYLOCOC. EPID.] [EUBACT] [74]
GCGGGAGNAGUUCAAUUUDAGAACACAUUCCUUCCCGGAAUGAGGUAUAGGUGCAAGUCCUAUCUUCCGCUCCA
> RG1540 [BACILLUS SUBTILIS] [EUBACT] [74]
GCGGGUGUAGUUUAGUGGDAAAACCUCAGCCUMCCAAGCUGAUGUCGUGAGTPCGAUUCUCAUCACCCGCUCCA
> RG1660 [E. COLI] [EUBACT] [74]
GCGGGCGNAGUUCAAUGGDAGAACGAGAGCUUCCCAAGCUCUAUACGAGGGTPCGAUUCCCUUCGCCCGCUCCA
> RG1661 [E. COLI] [EUBACT] [76]
GCGGGAAUAGCUCAGDDGGDAGAGCACGACCUUGCCAAGGUCGGGMUCGCGAGTPCGAGUCUCGUUUCCCGCUCCA
> RG1662 [E. COLI] [EUBACT] [75]
GCGGGCAUCGUAUAAUGGCUAUUACCUCAGCCUNCCAAGCUGAUGAUGCGGGTPCGAUUCCCGCUGCCCGCUCCA
> RG1700 [SALMONELLA TYPHI.] [EUBACT] [74]
GCGGGCGUAGUUCAAUMGDAGAACGAGAGCUUCCCAAGCUCUAUACGAGGGTPCGAUUCCCUUCGCCCGCUCCA
> RG1701 [SALMONELLA TYPHI.] [EUBACT] [75]
GCGGGCAUCGUAUAAUGGCUAUUACCUCAGCCUNCCAAGCUGAUGAUGCGGGTPCGAUUCCCGCUGCCCGCUCCA
> RH1140 [MYCOPLASMA CAPRIC.] [EUBACT] [77]
GGCGUAGGUGGUGAAGUGGDDAMCACAUCAGGUUGUMMCPCUGACAUACGCGGGUPCGAUCCCCGUUCUACGCCCCA
> RH1660 [E. COLI] [EUBACT] [77]
GGUGGCUANAGCUCAGDDGGDAGAGCCCUGGAUUQUGMPPCCAGUUMUCGUGGGTPCGAAUCCCAUUAGCCACCCCA
> RH1700 [SALMONELLA TYPHI.] [EUBACT] [77]
GGUGGCUANAGCUCAGDDGGDAGAGCCCUGGAUUQUGMPPCCAGUUMUCGUGGGTPCGAAUCCCAUUAGCCACCCCA
> RI1140 [MYCOPLASMA CAPRIC.] [EUBACT] [77]
GGACCUUNAGCUCAGUDGGDDAGAGCAUCCGGCUNAUMACCGGACGMUCAUUGGUPCAAGUCCAAUAAGGUCCACCA
> RI1141 [MYCOPLASMA CAPRIC.] [EUBACT] [77]
CGGAAUAUAGCUCAGCDGGDDAGAGCAUUCCGCUGAUNACGGAGAGMUCGUUGGUPCAAGUCCAAUUAUUCCGACCA
> RI1180 [MYCOPLASMA MYCOID.] [EUBACT] [77]
CGGAAUANAGCUCAGCDGGDDAGAGCAUUCCGCUGAUNACGGAGAGMUCGUUGGUPCAAGUCCAAUUAUUCCGACCA
> RI1580 [THERMUS THERMOPHI.] [EUBACT] [77]
GGGCGAUUAGCUCAGCUMGUDAGAGCGCACGCCUGAUNAGCGUGAGMUCGGUGGMPCAMGUCCACCAUCGCCCACCA
> RI1660 [E. COLI] [EUBACT] [77]

AGGCUUGUAGCUCAGGDGGDDAGAGCGCACCCCUGAUNAGGGUGAGMNCGGUGGTPCAAGUCCACPCAGGCCUACCA
> RI1661 [E. COLI] [EUBACT] [77]
AGGCUUGUAGCUCAGGUGGDDAGAGCGCACCCCUGAUNAGGGUGAGMNCGGUGGTPCAAGUCCACPCAGGCCUACCA
> RI1662 [E. COLI] [EUBACT] [76]
GGCCCCUNAGCUCAGUMGDDAGAGCAGGCGACUNAUNAPCGCUUGMNCGCUGGTPCAAGUCCAGCAGGGGCCACCA
> RI1540 [BACILLUS SUBTILIS] [77]
GGACCUUUAGCUCAGUDGGDDAGAGCAGACGGCULAUMACCGUCCGMUCGUAGGTPCGAGUCCUACAAGGUCCACCA
> RL1140 [MYCOPLASMA CAPRIC.] [EUBACT] [85]
GCCUUUUUGGCGGAAUDGGCAGMCGCAUUAGACUMAAMAPCUAACGAAGAAAUUCGUAUCGGUPCGAAUCCGAUAAAGGGCACCA
> RL1141 [MYCOPLASMA CAPRIC.] [EUBACT] [89]
CCCCAAGNGGCGGAAUAGGDAGMCGCAUUGGACUMAAMAPCCAACGGGCUUAAUAUCCUGUGCCGGUPCAAGUCCGGCCUUGGGGACCA
> RL1142 [MYCOPLASMA CAPRIC.] [EUBACT] [84]
GGGGGAUNGGCGGAAUDGGCAGMCGCACUAGACUUAGMAPCUAGCGUCUUUGACGUAAGGGUPCAAGUCCCUUAUCCCCCACCA
> RL1460 [BACILLUS STEARO.] [EUBACT] [86]
GCCGAUGNGGCGGAAUDGGCAGMCGCGCACGACUMAAMAPCGUGUGGGCUUUGCCCGUGUGGGTPCGACUCCCACCAUCGGCACCA
> RL1540 [BACILLUS SUBTILIS] [EUBACT] [87]
GCGGGUGUGCGGGAAUDGGDAGACCGGCUAGAUUCAGMAPCUAGGGUCUUUAUGGACCUGAGGGTPCAMGUCCCUUCACCCGCACCA
> RL1660 [E. COLI] [EUBACT] [87]
GCCCGGANGGUGGAADCMGDAGACACAAGGGAPUNAAMAPCCCUCGGCGUUCGCGCUGUGCGGGTPCAAGUCCCGCUCCGGGUACCA
> RL1661 [E. COLI] [EUBACT] [87]
GCGAAGGUGGCGGAADDMGDAGACGCGCUAGCUUCAGNPGPUAGUGUCCUUACGGACGUGGGGGTPCAAGUCCCCCCCCUCGCACCA
> RL1662 [E. COLI] [EUBACT] [87]
GCCGAGGUGGUGGAADDMGDAGACACGCUACCUUGAGNPGGUAGUGCCCAAUAGGGCUUACGGGTPCAAGUCCCGUCCUCGGUACCA
> RL1700 [SALMONELLA TYPHI.] [EUBACT] [87]
GCGAAGGUGGCGGAADDMGDAGACGCGCUAGCUUCAGNPGPUAGUGUCCUUACGGACGUGGGGGTPCAAGUCCCCCCCCUCGCACCA
> RL2020 [RHODOSPIRIL. RUB.] [EUBACT] [85]
GCCUUUGUAGCGGAADGGDAACGCGGCAGACUCAANAPCUGCUUUGGUAACCCAGGUGGUAGTPCGACUCUCCCCAAAGGCACCA
> RL2100 [ANACYSTIS NIDULANS] [EUBACT] [87]
GGGCAAGUGGCGGAAUDGGDAGACGCAGCAGACUCAANAPCUGCCGCUAGCGAUAGUGUGUGGGTPCGAGUCCCACCUUGCCCACCA
> RL2101 [ANACYSTIS NIDULANS] [EUBACT] [87]
GCGGAACUGGCGGAAUDGGDAGACGCGCUAGAUUCAGMPPCUAGUGGUUUCACGACUGUCCGGGTPCAAGUCCCGGGUUCCGCACCA
> RK1140 [MYCOPLASMA CAPRIC.] [EUBACT] [76]
GUCUGAUUAGCGCAACDGGCAGAGCAACUGACUCUUNAPCAGUGGMUUGUGGGUPCGAUUCCCACAUCAGGCACCA
> RK1141 [MYCOPLASMA CAPRIC.] [EUBACT] [76]
GACUCGUUAGCUCAGCCGGDAGAGCAACUGGCUMUUNACCAGUGGMUCCGGGGUPCGAAUCCCCGACGAGUCACCA
> RK1540 [BACILLUS SUBTILIS] [EUBACT] [76]
GAGCCAUUAGCUCAGUDGGDAGAGCAUCUGACUMUUNAPCAGAGGMUCGAAGGTPCGAGUCCUUCAUGGCUCACCA
> RK1541 [BACILLUS SUBTILIS] [EUBACT] [76]
GAGCCAUUAGCUCAGUDGGDAGAGCAUCUGANUMUUMAPCAGAGGMUCGAAGGTPCGAGUCCUUCAUGGCUCACCA
> RK1660 [E. COLI] [EUBACT] [76]
GGGUCGUUAGCUCAGDDGGDAGAGCAGUUGACUMUUNAPCAAUUGMNCGCAGGTPCGAAUCCUGCACGACCCACCA
> RM1140 [MYCOPLASMA CAPRIC.] [EUBACT] [77]
GGCGGGGNAGCUCAGUDGGDDAGAGCGUUCGGUUCAUACCCGAAAGMUCGAGAGUPCAAAUCUCUCCCCCGCUACCA
> RM1540 [BACILLUS SUBTILIS] [EUBACT] [76]
GGCGGUGUAGCUCAGCGGCDAGAGCGUACGGUUCAUMCCCGUGAGMDCGGGGGTPCGAUCCCCUCCGCCGCUACCA
> RM1580 [THERMUS THERMOPHI.] [EUBACT] [77]
CGCGGGGNGGAGCAGCCUMGDAGCUCGUCGGGMUCAUAACCCGAAGMUCGCGGGMPCAMAUCCCGCCCCCGCAACCA
> RM1660 [E. COLI] [EUBACT] [77]
GGCUACGNAGCUCAGDDMGDDAGAGCACAUCACUMAUNAPGAUGGGMNCACAGGTPCGAAUCCCGUCGUAGCCACCA
> RF1140 [MYCOPLASMA CAPRIC.] [EUBACT] [76]
GGUCGUGUAGCUCAGUCGGDAGAGCAGCAGACUGAAMCPCUGCGUMUCGGCGGUPCAAUUCCGUCCACGACCACCA
> RF1460 [BACILLUS STEARO.] [EUBACT] [76]
GGCUCGGNAGCUCAGUCGGDAGAGCAAAGGACUMAAMAPCCUUGUMUCGGCGGTPCGAUUCCGUCCCGAGCCACCA
> RF1540 [BACILLUS SUBTILIS] [EUBACT] [76]

GGCUCGGUAGCUCAGUDGGDAGAGCAACGGACUMAAMAPCCGUGUMUCGGCGGTPCGAUUCCGUCCCGAGCCACCA
> RF1580 [THERMUS THERMOPHI.] [EUBACT] [76]
GCCGAMGNAGCUCAGUUMGDAGAGCAUGCGACUGAANAPCGCAGUMUCGGCGGTPCGAUUCCGCUCCUCGGCACCA
> RF1660 [E. COLI] [EUBACT] [76]
GCCCGGANAGCUCAGDCGGDAGAGCAGGGGAPUGAAMAPCCCCGUMNCCUUGGTPCGAUUCCGAGUCCGGGCACCA
> RF2020 [RHODOSPIRIL. RUB.] [EUBACT] [76]
GCCCGGGUAGCUCAGCDGGDAGAGCACGUGACUGAAMAPCACGGUMUCGGUGGTPCGACUCCGCCCCCGGGCACCA
> RF2060 [AGMENELLUM QUADR.] [EUBACT] [76]
GCCAGGAUAGCNCAGUDMGDAGAGCAGAGGACUGAAMAPCCUCGUMUCGGCGGTPCAAUUCCGCCUCCCGGCACCA
> RP1140 [MYCOPLASMA CAPRIC.] [EUBACT] [77]
CGGGAAGUGGCUCAGUUUGGDAGAGCAUUCGGUUUGGMACCGAAGGMNCGCAGGUPCAAAUCCUGUCUUCCCGACCA
> RP1180 [MYCOPLASMA MYCOID.] [EUBACT] [77]
CGGGAAGUGGCUCAGUUUGGDAGAGCAUUCGGUUUGGMACCGAAGGGUCGCAGGUUCAAAUCCUGUCUUCCCGACCA
> RP1540 [BACILLUS SUBTILIS] [EUBACT] [77]
CGGGAAGUAGCUCAGCUUGGDAGAGCACAUGGPUMGGMACCAUGGGMUCGCAGGTPCGAAUCCUGUCUUCCCGACCA
> RP1700 [SALMONELLA TYPHI.] [EUBACT] [77]
CGGUGAUNGGCGCAGCCUGGDAGCGCACUUCGMUCGGMACGAAGGGMUCGGAGGTPCGAAUCCUCUAUCACCGACCA
> RP1701 [SALMONELLA TYPHI.] [EUBACT] [77]
CGGCACGNAGCGCAGCCUGGDAGCGCACCGUCMUGGGMUPGCGGGGMUCGGAGGTPCAAAUCCUCUCGUGCCGACCA
> RP1702 [SALMONELLA TYPHI.] [EUBACT] [77]
CGGCGAGNAGCGCAGCUUGGDAGCGCAACUGGMUNGGMACCAGUGGMUCGGAGGTPCGAAUCCUCPCUCGCCGACCA
> RS1140 [MYCOPLASMA CAPRIC.] [EUBACT] [90]
GGGUUAANACUCAAGUDGGDGAMGAGGACACCCUGCUNAGGUGUUAGGUCGGUCUCCGGCGCGAGGGUPCGAGUCCCUCUUAACCCGCCA
> RS1141 [MYCOPLASMA CAPRIC.] [EUBACT] [92]
GGAAGAUNACCCAAGUCCGGCDGAMGGGAUCGGUCUUGAMAACCGAGAGUCGGGGAAACCGAGCGGGGGUPCGAAUCCCUCAUCUUCCGCCA
> RS1180 [MYCOPLASMA MYCOID.] [EUBACT] [93]
GGAAGAUUACCCAAGUCCGGCDGAHGGGAUCGGUCUUGAMAACCGAGAGUCGGGGAAACCCGAGCGGGGGUUCGAAUCCCUCAUCUUCCGCCA
> RS1540 [BACILLUS SUBTILIS] [EUBACT] [92]
GGAGGAANACCCAAGUCUGGCDGAMGGGAUCGGUCUMGANAACCGACAGGGUGUCAAAGCCCGCGGGGGTPCGAAUCCCUCUUCCUCCGCCA
> RS1541 [BACILLUS SUBTILIS] [EUBACT] [89]
GGAGAAGUACUCAAGUGGCDGAMGAGGCGCCCCUGCUNAGGGUGUGUCGCGUAAGCGGCGCGAGGGTPCAAAUCCCUCCUUCUCCGCCA
> RS1542 [BACILLUS SUBTILIS] [EUBACT] [92]
GGAGAGCNGUCCGAGUGGDCGAMGGAGCACGAUUGGAAAPCGUGUAGGCGGUCAACUCCGUCUCAAGGGTPCGAAUCCCUUGCUCUCCGCCA
> RS1660 [E. COLI] [EUBACT] [90]
GGAGAGAUGCCGGAGCMGCDGAACGGACCGGUCUCGAMAACCGGAGUAGGGGCAACUCUACCGGGGGTPCAAAUCCCCCUCUCUCCGCCA
> RS1661 [E. COLI] [EUBACT] [93]
GGUGAGGNGGCCGAGAGGCDGAAGGCGCUCCCTUGCUNAGGGAGUAUGCGGUCAAAAGCUGCAUCCGGGGTPCGAAUCCCCGCCUCACCGCCA
> RS1662 [E. COLI] [EUBACT] [88]
GGUGAGGUGUCCGAGUMGCDGAAGGAGCACGCCUGGAAAGPGUGUAUACGGCAACGUAUCGGGGGTPCGAAUCCCCCCCUCACCGCCA
> RS1663 [E. COLI] [EUBACT] [88]
GGUGAGGNGUCCGAGUMGDDGAAGGAGCACGCCUGGAAAGPGUGUAUACGGCAACGUAUCGGGGGTPCGAAUCCCCCCCUCACCGCCA
> RS1664 [E. COLI] [EUBACT] [88]
GGAAGUGNGGCCGAGCMGDDGAAGGCACCGGUMUNGAMAACCGGCGACCCGAAAGGGUUCCAGAGTPCGAAUCUCUGCGCUUCCGCCA
> RT1140 [MYCOPLASMA CAPRIC.] [EUBACT] [76]
GCUGACUNAGCUCAGUDGGDAGAGCAAUUGACUAGUNAPCAAUAGMUCGAAGGUPCAAAUCCUUUAGUCAGCACCA
> RT1141 [MYCOPLASMA CAPRIC.] [EUBACT] [76]
GCUGACUNAGCUCAGCAGGCAGAGCAACUGACUUGUNAPCAGUAGMUCGUAGGUPCGAUUCCUAUAGUCAGCACCA
> RT1180 [MYCOPLASMA MYCOID.] [EUBACT] [76]
GCUGACUUAGCUCAGCAGGCAGAGCAACUGACUUGUNAPCAGUAGMUCGUAGGUUCGAUUCCUAUAGUCAGCACCA
> RT1540 [BACILLUS SUBTILIS] [EUBACT] [76]
GCCGGUGUAGCUCAAUDGGDAGAGCAACUGACUMGUNAPCAGUAGMUUGGGGGTPCAAGUCCUCUUGCCGGCACCA
> RT1660 [E. COLI] [EUBACT] [76]
GCUGAUAUAGCUCAGDDGGDAGAGCGCACCCUUGGUMAGGGUGAGMUCGGCAGTPCGAAUCUGCCUAUCAGCACCA
> RT1661 [E. COLI] [EUBACT] [75]

GCUGAUAUGGCUCAGDDGGDAGAGCGCACCCUUGGUMAGGGUGAGMUCCCAGTPCGACUCUGGGUAUCAGCACCA
> RW1140 [MYCOPLASMA CAPRIC.] [EUBACT] [75]
AGGAGAGUAGUUCAAUGGDAGAACGUCGGUCUMCAMAACCGAGCMUUGAGGGUPCGAUUCCUUUCUCUCCUGCCA
> RW1141 [MYCOPLASMA CAPRIC.] [EUBACT] [76]
AGGGGCAUAGUUCAGUAGGDAGAACAUCGGUCUMCAMAACCGAGUMUCACGAGUPCGAGUCUUGUUGCCCCUGCCA
> RW1250 [SPIROPLASMA CITRI] [EUBACT] [76]
AGGGGUAUAGUUCAAUCGGUAGAACACCGGACUNCANAPCCGGGUMUUGUGGGUPCAAGUCCUGCUACCCCUGCCA
> RW1251 [SPIROPLASMA CITRI] [EUBACT] [75]
AGGGGUGUAGUUUAAUGGUAGAACAGCGGUCUCCANCACCGUACGUUGUGGGUPCAAGUCCUGUCACCCCUGCCA
> RW1540 [BACILLUS SUBTILIS] [EUBACT] [74]
AGGGGCAUAGUUUAACGGDAGAACAGAGGPCUCCANAACCUCCGGUGUGGGTPCGAUUCCUACUGCCCCUGCCA
> RW1660 [E. COLI] [EUBACT] [76]
AGGGGCGNAGUUCAADDGGDAGAGCACCGGUMUCCAMAACCGGGUMUUGGGAGTPCGAGUCUCUCCGCCCCUGCCA
> RY1140 [MYCOPLASMA CAPRIC.] [EUBACT] [84]
GGAGGGGUAGCGAAGUGGCDAAMCGCGGGUGGCUGUAMCCCACUUCCUUACGGUUCGGGGGUPCGAAUCCCUCCCCCCUCCACCA
> RY1460 [BACILLUS STEARO.] [EUBACT] [85]
GGAGGGGNAGCGAAGUMGCUAAMCGCGGCGGACUQUAMAPCCGCUCCCUUUGGGUUCGGCGGTPCGAAUCCGUCCCCCUCCACCA
> RY1540 [BACILLUS SUBTILIS] [EUBACT] [85]
GGAGGGGNAGCGAAGUGGCUAAMCGCGGCGGACUQUANAPCCGCUCCCUCAGGGUUCGGCAGTPCGAAUCUGCCCCCCUCCACCA
> RY1541 [BACILLUS SUBTILIS] [EUBACT] [85]
GGAGGGGNAGCGAAGUGGCUAAMCGCGGCGGACUQUAMAPCCGCUCCCUCAGGGUUCGGCAGTPCGAAUCUGCCCCCCUCCACCA
> RY1660 [E. COLI] [EUBACT] [85]
GGUGGGGNUCCCGAGCMGCCAAAGGGAGCAGACUQUAMAPCUGCCGUCAUCGACUUCGAAGGTPCGAAUCCUUCCCCCCACCACCA
> RY1661 [E. COLI] [EUBACT] [85]
GGUGGGGNUCCCGAGCMGCCAAAGGGAGCAGACUQUAMAPCUGCCGUCACAGACUUCGAAGGTPCGAAUCCUUCCCCCCACCACCA
> RV1140 [MYCOPLASMA CAPRIC.] [EUBACT] [76]
GGAGUGUUAGCUCAGCDGGGAGAGCUCCUGCCUUACMAGCAGGCGMUCAUAGGUPCAAGUCCUAUACACUCCACCA
> RV1180 [MYCOPLASMA MYCOID.] [EUBACT] [76]
GGAGUGUNAGCUCAGCDGGGAGAGCUCCUGCCUUACMAGCAGGCGGUCAUAGGUUCAAGUCCUAUACACUCCACCA
> RV1460 [BACILLUS STEARO.] [EUBACT] [76]
GAUUCCGUAGCUCAGCDGGGAGAGCGCCACCUUGACMGGGUGGAGMUCGCUGGTPCGAGCCCAGUCGGAAUCACCA
> RV1540 [BACILLUS SUBTILIS] [EUBACT] [76]
GGAGGAUUAGCUCAGCDGGGAGAGCAUCUGCCUMACMAGCAGAGGMUCGGCGGTPCGAGCCCGUCAUCCUCCACCA
> RV1660 [E. COLI] [EUBACT] [77]
GCGUCCGNAGCUCAGDDGGDDAGAGCACCACCUUGACAUGGUGGGGMNCGGUGGTPCGAGUCCACUCGGACGCACCA
> RV1661 [E. COLI] [EUBACT] [77]
GCGUUCANAGCUCAGDDGGDDAGAGCACCACCUUGACAUGGUGGGGMNCGUUGGTPCGAGUCCAAUUGAACGCACCA
> RV1662 [E. COLI] [EUBACT] [76]
GGGUGAUNAGCUCAGCDGGGAGAGCACCUCCCUNACMAGGAGGGGMUCGGCGGTPCGAUCCCGUCAUCACCCACCA

References

[1] J. P. Abrahams, M. van den Berg, E. van Batenburg, and C. Pleij. Prediction of RNA secondary structure, including pseudoknotting, by computer simulation. *Nucl. Acids Res.*, 18:3035–3044, 1990.

[2] I. Tinoco amd C. Bustamante. How RNA folds. *J. Mol. Biol.*, 293:271–281, 1999.

[3] P. Auffinger, S. Louise-May, and E. Westhof. Hydration of C-H groups in tRNA. *Faraday Discuss.*, 103:151–173, 1996.

[4] J. M. Avis, A. G. Day, G. A. Garcia, and A. R. Fersht. Reaction of modified and unmodified tRNA(tyr) substrates with tyrosyl-tRNA synthetase (bacillus stearothermophilus). *Biochemistry*, 32:5312–5320, 1993.

[5] M. I. Baldi, E. Mattoccia, E. Bufardeci, S. Fabbri, and G. E. Tocchini-Valentini. Participation of the intron in the reaction catalyzed by the Xenopus laevis tRNA splicing endonuclease. *Science*, 255:1404–1408, 1992.

[6] Robert T. Batey, Robert P. Rambo, and Jennifer A. Doudna. Tertiäre Motive bei Struktur und Faltung von RNA. *Angew. Chem.*, 111:2472–2491, 1999.

[7] T. Baumstark, A. R. Schroder, and D. Riesner. Viroid processing: Switch from cleavage to ligation is driven by a change from a tetraloop to a loop E conformation. *EMBO J.*, 16:599–610, 1997.

[8] S. Beresten, M. Jahn, and D. Soll. Aminoacyl-tRNA synthetase-induced cleavage of tRNA. *Nucl. Acids Res.*, 20:1523–1530, 1992.

[9] C. K. Biebricher and R. Luce. In vitro recombination and terminal elongation of RNA by Qβ replicase. *EMBO J.*, 11:5129–5135, 1992.

[10] G. R. Bjork, J. M. Durand, T. G. Hagervall, R. Leipuviene, H. K. Lundgren, K. Nilsson, P. Chen, Q. Qian, and J. Urbonavicius. Transfer RNA modification: influence on translational frameshifting and metabolism. *FEBS Lett*, 452:47–51, 1999.

[11] P. Brion and E. Westhof. Hierarchy and dynamics of RNA folding. *Annu. Rev. Biophys. Biomol. Struct.*, 26:113–137, 1997.

[12] S. E. Butcher, F. H. Allain, and J. Feigon. Solution structure of the loop B domain from the hairpin ribozyme. *Nat. Struct. Biol.*, 6:212–216, 1999.

[13] S. E. Butcher, T. Dieckmann, and J. Feigon. Solution structure of a GAAA tetraloop receptor. *EMBO*, 16:7490–7499, 1997.

[14] M. G. Caprara, G. Mohr, and A. M. Lambowitz. A tyrosyl-tRNA synthetase protein induces tertiary folding of the group i intron catalytic core. *J. Mol. Biol.*, 257:512–531, 1996.

[15] J. H. Cate, A. R. Gooding, E. Podell, K. Zhou, B . L. Golden, A. A. Szewczak, C. D. Kundrot, T. R. Cech, and J. A. Doudna. Crystal structure of a group I ribozyme domain: Principles of RNA packing. *Science*, 273:1678–1685, 1996.

[16] J. H. Cate, A. R. Gooding, E. Podell, K. Zhou, B . L. Golden, A. A. Szewczak, C. D. Kundrot, T. R. Cech, and J. A. Doudna. RNA tertiary structure mediation by adenosine platforms. *Science*, 273:1696–1699, 1996.

[17] M. Chastain and I. Tinoco. Nucleoside triples from the group I intron. *Biochemistry*, 32:14220–14228, 1993.

[18] G. Cho and R. F. Doolittle. Intron distribution in ancient paralogs supports random insertion and not random loss. *j. Mol. Evol.*, 44:573–548, 1997.

[19] G. L. Conn, D. E. Draper, E. E. Lattman, and A. G. Gittis. Crystal structure of a conserved ribosomal protein-RNA complex. *Science*, 284:1171–1174, 1999.

[20] C. C. Correll, B. Freeborn, P. B. Moore, and T. A. Steitz. Metals, motifs and recognition in the crystal structure of a 5s rRNA domain. *Cell*, 91:705–712, 1997.

[21] S. M. Coutts, J. Gangloff, and G. Dirheimer. Conformational transitions in tRNA asp (brewer's yeast). thermodynamic, kinetic, and enzymatic measurements on oligonucleotide fragments and the intact molecule. *Biochemistry*, 13:3938–3944, 1974.

[22] D. M. Crothers, P. E. Cole, C. W. Hilbers, and R. G. Shulman. The molecular mechanism of thermal unfolding of Escherichia coli formylmethionine transfer RNA. *J. Mol. Biol.*, 87:63–88, 1974.

[23] A. W. Curnow, F. L. Kung, K. A. Koch, and G. A. Garcia. tRNA-guanine transglycosylase from Escherichia coli: gross tRNA structural requirements for recognition. *Biochemistry*, 32:5239–5246, 1993.

[24] S. Cusack, C. Berthet-Colominas, M. Hartlein, N. Nassar, and R. Leberman. A second class of synthetase structure revealed by x-ray analysis of escherichia coli seryl-tRNA synthetase at 2.5 a. *Nature*, 347:249–255, 1990.

[25] S. C. Darr, K. Zito, D. Smith, and N. R. Pace. Contributions of phylogenetically variable strcutural elements to the function of the ribozyme ribonuclease p. *Biochemistry*, 31:328–333, 1992.

[26] S. J. desouza, M. Long, R. J. Klein, S. Roy, S. Lin, and W. Gilbert. Towards a resolution of the introns early/late debate: only phase zero introns are corelated with the structure of ancient proteins. *Proc. Natl. Acad. Sci. USA*, 95:5094–5099, 1998.

[27] D. E. Duckett, A. I. Murchie, M. J. Giraud-Panis, J. R. Pohler, and D. M. Lilley. Structure of the four-way DNA junction and its interaction with proteins. *Philos. Trans. R. Cos. Lond. B. Biol. Sci.*, 347:27–36, 1995.

[28] C. G. Edmonds, P. F. Crain, R. Gupta, T. Hashizume, C. H. Hocart, J. A. Kowalak, S. C. Pomerantz, K. O. Stetter, and J. A. McCloskey. Posttranscriptional modification of tRNA in thermophilic archae (archeabacteria). *J. Bacteriology*, 173:3138–3148, 1991.

[29] V. L. Emerick and S. A. Woodson. Selfsplicing of the Tetrahymena pre-rRNA is decreased by misfolding during transcription. *Biochemistry*, 32:14062–14067, 1993.

[30] G. Eriani, M. Delarue, O. Poch, J. Gangloff, and D. Moras. Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature*, 347:203–206, 1990.

[31] G. Eriani and J. Gangloff. Yeast aspartyl-tRNA synthetase residues interacting with tRNA(asp) identity bases connectively contribute to tRNA(asp) binding in the ground and transition-state complex and discriminate against non-cognate tRNAs. *J. Mol. Biol.*, 291:761–773, 1999.

[32] A. Ferre-D'Amare and J.A. Doudna. RNA folds: Insights from recentcrystal structures. *Annu. Rev. Biophys. Biomol. Struct*, 28:57–73, 1999.

[33] Adrian R. Ferre-D'Amare and Jennifer A. Doudna. RNA folds: Insights from recent crystal structures. *Annu. Rev. Biophys. Biomol. Struct.*, 28:57–73, 1999.

[34] A. E. Fischer, P. J. Beunig, and K. Musier-Forsyth. Identification of discriminator base atomic groups that modulate the alanine aminoacylation reaction. *J. Biol. Chem.*, 274:37093–37096, 1999.

[35] C. Flamm, I. L. Hofacker, and P. F. Stadler. RNA in silico: the computational biology of RNA secondary structures. *Adv. Complex Syst.*, 2:65–90, 1999.

[36] W. Fontana and P. Schuster. Shaping space: The possible and the attainable in RNA genotype-phenotype mapping. *J. Theor. Biol.*, 194:491–515, 1998.

[37] D. Gautheret, S. H. Damberger, and R. R. Gutell. Identification of base triples in RNA using comparative sequence analysis. *J. Mol. Biol.*, 248:27–43, 1995.

[38] Daniel Gautheret and Robin R. Gutell. Inferring the conformation of RNA base pairs and triples from patterns of sequence variation. *Nucleic Acids Research*, 25(8):1559–1564, 1997.

[39] R. Giege, C. Florentz, D. Kern, J. Gangloff, G. Eriani, and D. Moras. Aspartate identity of transfer RNAs. *Biochimie*, 78:605–623, 1996.

[40] R. Giege, J. D. Puglisi, and C. Florentz. tRNA structure and aminoacylation efficiency. *Proc. Nucl. Acid Res. Mol. Biol.*, 45:129–206, 1993.

[41] R. Giege, M. Sissler, and C. Florentz. Universal rules and idiosyncratic features in tRNA identity. *NAR*, 26:5017–5035, 1998.

[42] R. Giegerich, D. Haase, and M. Rehmsmeier. Prediction and visualization of structural switches in RNA. *Proceedings of the Pacific Symposium on Biocomputing*, 4:126–137, 1999.

[43] A. P. Gultyaev, F. H. D. van Batenburg, and C. W. A. Pleij. The computer simulation of RNA folding pathways using a genetic algorithm. *J. Mol. Biol.*, 250:37–51, 1995.

[44] R. R. Gutell. Evolutionary characteristics of RNA: Inferring higher-order structure from patterns of sequence variation. *Curr. Opin. Struct. Biol.*, 3:313–322, 1993.

[45] A. J. Hager, J. D. Pollard, and J. W. Szostak. Ribozymes: aiming at RNA replication and protein synthesis. *Chem. Biol.*, 3:717–725, 1996.

[46] K. B. Hall, J. R. Sampson, O. C. Uhlenbeck, and A. G. Redfield. Structure of an unmodified tRNA molecule. *Biochemistry*, 28:5794–5801, 1989.

[47] K. M. Harrington, I. A. Nazarenko, D. B. Dix, R. C. Thompson, and O. C. Uhlenbeck. In vitro analysis of translational rate and accuracy with an unmodified tRNA. *Biochemistry*, 32:7617–7622, 1993.

[48] T. Hasegawa, M. Miyano, H. Himeno, Y. Sano, K. Kimura, and M. Shimizu. Identity determinants of e. coli threonine tRNA. *Biochem. Biophys. Res. Commun.*, 184:478–484, 1992.

[49] E. R. Hawkins, S. H. Chang, and W. L. Mattice. Kinetics of the renaturation of yeast $tRNA_3^{leu}$. *Biopolymers*, 16:1557–1566, 1977.

[50] T. Hermann and D. J. Patel. Stitching together RNA tertiary motifs. *J. Mol. Biol.*, 294:829–849, 1999.

[51] I. L. Hofacker, W. Fontana, P. F. Stadler, S. Bonhoeffer, M. Tacker, and P. Schuster. Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.*, 125:167–188, 1994.

[52] P. Hogeweg and B. Hesper. Energy directed folding of RNA sequences. *Nucl. Acids Res.*, 12:67–74, 1984.

[53] S. R. Holbrook, J. L. Sussman, R. W. Warrant, G. M. Church, and S. H. Kim. RNA-ligant interactions. (i) Magnesium binding sites in yeast tRNAphe. *Nucl. Acids Res.*, 4:2811–2820, 1977.

[54] R. W. Holley. Structure of an alanine transfer ribonucleic acid. *JAMA*, 194:868–871, 1965.

[55] M. Illangasekare, G. Sanchez, T. Nickles, and M. Yarus. Aminoacyl-RNA synthesis catalyzed by an RNA. *Science*, 267:643–647, 1995.

[56] M. Jahn, M. J. Rogers, and D. Soll. Anticodon and acceptor stem nucleotides in tRNA(gln) are major recognition elements for E. coli glutaminyl-tRNA synthetase. *Nature*, 352:258–260, 1991.

[57] A. Jenne and M. Famulok. A novel ribozyme with ester transferase activity. *Chem. Biol.*, 5:23–34, 1998.

[58] P. F. Johnson and J. Abelson. The yeast tRNAtyr gene intron is essential for correct modification of its tRNA product. *Nature*, 302:681–687, 1983.

[59] G. F. Joyce. RNA evolution and the origins of life. *Nature*, 338:217–224, 1989.

[60] G. F. Joyce. The rise and fall of the RNA world. *The New Biologist*, 3:399–407, 1991.

[61] G. F. Joyce. Building the RNA world. Ribozymes. *Curr. Biol.*, 6:965–967, 1996.

[62] G.F. Joyce. Amplification, mutation, and selection of catalytic RNA. *Gene*, 82:85–87, 1989.

[63] G. Keith and G. Dirheimer. Evidence for the existence of an expressed minor variant tRNAphe in yeast. *Biochem. Biophys. Res. Commun.*, 142:183–187, 1987.

[64] H. S. Kim, I. Y. Kim, D. Soll, and S. Y. Lee. Transfer RNA identity change in anticodon variants of E. coli tRNA(phe) in vivo. *Mol Cells*, 10:76–82, 2000.

[65] S. H. Kim, J. L. Sussman, F. L. Suddath, G. J. Quigley, A. McPherson, A. H. Wang, N. C. Seeman, and A. Rich. The general structure of transfer RNA molecules. *Proc. Natl. Acad. Sci. U*, 71:4970–4974, 1974.

[66] L. L. Kisselev. The role of the anticodon in recognition of tRNA by aminoacyl-tRNA synthetases. *Prog. Nucleic. Acid Res. Mol. Biol.*, 32:237–266, 1985.

[67] G. Knapp, J. S. Beckmann, P. F. Johnson, S. A. Fuhrmann, and J. Abelson. Transcription and processing of intervening sequences in yeast tRNA genes. *Cell*, 14:221–236, 1978.

[68] G. Knapp, R. C. Ogden, C. L. Peebles, and J. Abelson. Splicing of yeast tRNA precursors: structure of the reaction intermediates. *Cell*, 18:37–45, 1979.

[69] F.R. Kramer, D. R. Mills, P. E. Cole, T. Nishihara, and S. Spiegelman. Evolution in vitro: sequence and phenotype of a mutant RNA resistant to ethidium bromide. *J. Mol. Biol.*, 89:719–736, 1974.

[70] K. Kruger, P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, and T. R. Cech. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell*, 31:147–157, 1982.

[71] L. G. Laing, T. C. Gluick, and D. E. Draper. Stabilization of RNA structure by mg ions. specific and non-specific effects. *J. Mol. Biol*, 237:577–587, 1994.

[72] C. P. Lee, N. Mandal, M. R. Dyson, and U. L. RajBhandary. The discriminator base influences tRNA structure at the end of the acceptor stem and possibly its interaction with proteins. *Proc. Natl. Acad. Sci. USA*, 90:7149–7152, 1993.

[73] M. C. lee and G. Knapp. Transfer RNA splicing in saccharomyces cerevisiae. secondary and tertiary structure of the substrates. *J. Biol. Chem*, 260:3108–3115, 1985.

[74] N. Lee, Y. Bessho, K. Wei, J. W. Szostak, and H. Suga. Ribozyme-catalyzed tRNA aminoacylation. *Nat. Struct. Biol.*, 7:28–33, 2000.

[75] N. B. Leontis and E. Westhof. Conserved geometrical base-pairing patterns in RNA. *Quart. Rev. Biophys.*, 31:399–455, 1998.

[76] M. Levitt. Detailed molecular model for transfer ribonucleic acid. *Nature*, 224:759–763, 1969.

[77] H. Li, C. R. Trotta, and J. Abelson. Crystal structure and evolution of a transfer RNA splicing enzyme. *SCIENCE*, 280:279–284, 1998.

[78] J. Li, B. Esberg, J. F. Curran, and G. R. Bjork. Three modified nucleotides present in the anticodon stem and loop influence the in vivo aa-tRNA selsction in a tRNA-dependent manner. *J. Mol. Biol*, 271:209–221, 1997.

[79] S. Litvak, L. Sarih-Cottin, M. Fournier, M. Andreola, and M. Tarrago-Litvak. Priming of HIV replication by tRNA(lys3): role of reverse transcriptase. *Trends Biochem. Sci.*, 19:114–118, 1994.

[80] D. R. Liu, T. J. Magliery, M. Pastrnak, and P. G. Schultz. Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. *Proc. Natl. Acad. Sci. USA*, 94:10092–10097, 1997.

[81] D. R. Liu and P. G. Schultz. Progress toward the evolution of an organism with an expanded genetic code. *Proc. Natl. Acad. Sci. USA*, 96:4780–4785, 1999.

[82] J. Stephen Lodmell and Albert E. Dahlberg. A conformational switch in Escherichia coli 16S ribosomal RNA during decoding of messenger RNA. *Science*, 277:1262–1267, 1997.

[83] P. A. Lohse and J. W. Szostak. Ribozyme-catalysed amino-acid transfer reactions. *Nature*, 381:442–444, 1996.

[84] J. R. Lorsch and J. W. Szostak. In vitro evolution of new ribozymes with polynucleotide kinase activity. *Nature*, 371:31–36, 1994.

[85] Y. Lu, V. Planelles, X. Li, C. Palaniappan, B. Day, P. Challita-Eid, R. Amado, D. Stephens, D. B. Kohn, A. Bakker, P.Fay, R. A. Bambara, and J. D. Rosenblatt. Inhibition of HIV-1 replication using a mutated tRNAlys-3 primer. *J. Biol. Chem.*, 272:14523–14531, 1997.

[86] R. Marquet, C. Isel, C. Ehresmann, and B. Ehresmann. tRNAs as primer of reverse transcriptases. *Biochimie*, 77:113–124, 1995.

[87] Hugo M. Martinez. An RNA folding rule. *Nucleic Acids Research*, 12(1):323–334, 1984.

[88] B. Masquida and E. Westhof. On the wobble G-U and related pairs. *RNA*, 6:9–15, 2000.

[89] D. H. Mathews, J. Sabina, M. Zuker, and D. Turner. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *JMB*, 288:911–940, 1999.

[90] J. S. McCaskill. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, 29:1105–1119, 1990.

[91] W. H. McClain. Rules that govern tRNA identity in protein synthesis. *J. Mol. Biol*, 234:257–280, 1993.

[92] W. H. McClain. Transfer RNA identity. *FASEB J*, 7:257–80, 1993.

[93] W. H. McClain, K. Foss, R. A. Jenkins, and J. Schneider. Nucleotides that determine Escherichia coli tRNA(arg) and tRNA(lys) acceptor identities revealed by analyses of mutant opal and amber suppressor tRNAs. *Proc. Natl. Acad. Sci. USA*, 87:9260–9264, 1990.

[94] D. A. Melton, E. M. deRobertis, and R. Cortese. Order and intracellular location of the events involved in the maturation of a spliced tRNA. *Nature*, 284:143–148, 1980.

[95] T Miramatsu, K. Nishikava, F. Nemoto, Y. Kuchino, S. Nishimura, T. Miyazawa, and S. Yokoyama. Codon and amino-acid specificities of a transfer RNA are both converted by a single post-transcriptional modification. *Nature*, 336:179–181, 1988.

[96] A. Mironov and A. Kister. A kinetic approach to the prediction of RNA secondary structures. *J. Biomol. Struct. Dyn.*, 2:953–962, 1985.

[97] A. Mironov and V. F. Lebedev. A kinetic modle of RNA folding. *BioSystems*, 30:49–56, 1993.

[98] V. K. Misra and D. E. Draper. On the role of magnesium ions in RNA stability. *Biopolymers*, 48:113–135, 1998.

[99] N. A. Moor, V. N. Ankilova, and O. L. Lavrik. Recognition of tRNAphe by phenylalanyl-tRNA synthetase of thermus thermophilus. *Eur. J. Biochem.*, 234:897–902, 1995.

[100] K. Nishikura and E. M. DeRobertis. RNA processing in microinjected xenopus oocytes. Sequential addition of base modifications in the spliced transfer RNA. *J. Mol. Biol.*, 145:405–420, 1981.

[101] J. Normanly and J. Abelson. tRNA identity. *Annu. Rev. Biochem*, 58:6309–6313, 1989.

[102] P. Z. O'Farrell, B. Cordell, P. Valenzuela, W. J. Rutter, and H. M. Goodman. Structure and processing of yeast precursor tRNAs containing intervening sequences. *Nature*, 274:438–445, 1978.

[103] T. A. Osman, C. L. Hemenway, and K. W. Buck. Role of the 3' tRNA-like structure in tobacco mosaic virus minus-strand RNA synthesis by the viral RNA-dependent RNA polymerase in vitro. *J. Virol.*, 74:11671–11680, 2000.

[104] L. Pallanck and L. H. Schulman. Anticodon-dependent aminoacylation of a noncognate tRNA with isoleucine, valine and phenylalanine in vivo. *Proc. Natl. Acad. Sci. USA*, 88:3872–3876, 1991.

[105] C. L. Peebles, R. C. Ogden, G. Knapp, and J. Abelson. Splicing of yeast tRNA precursors: a two-stage reaction. *Cell*, 18:27–35, 1979.

[106] A. T. Perrotta and M. D. Been. A toggle duplex in hepatitis delta virus self-cleaving RNA that stabilizes an inactive and a salt-dependent pro-active ribozyme conformation. *J. Mol. Biol.*, 279:361–373, 1998.

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

[108] H. W. Pley, K. M. Flaherty, and D. B. McKay. Three-dimensional structure of a hammerhead ribozyme. *Nature*, 372:68–74, 1994.

[109] J. Puetz, J. D. Puglisi, C. Florentz, and R. Giege. Identity elements for specific aminoacylation of yeast tRNA(asp) by cognate aspartyl-tRNA synthetase. *Science*, 252:1696–1699, 1991.

[110] J. D. Puglisi, J. Puetz, C. Florentz, and R. Giege. Influence of tRNA tertiary structure and stability on aminoacylation by yeast aspartyl-tRNA synthetase. *NAR*, 21:41–49, 1993.

[111] G. J. Quigley and A. Rich. Structural domains of transfer RNA molecules. *Science*, 194:796–806, 1976.

[112] G. J. Quigley, M. M. Teeter, and A. Rich. Structural analysis of spermine and magnesium ion binding to yeast phenylalanine transfer RNA. *Proc. Natl. Acad. Sci. USA*, 75:64–68, 1978.

[113] B. Reinhold-Hurek and D. A. Shub. Self-splicing introns in tRNA genes of widely divergent baxcteria. *Nature*, 357:173–176, 1992.

[114] B. Reinhold-Hurek and D. A. Shub. Experimental approaches for detecting self-splicing group I introns. *Methods Enzymol.*, 224:491–502, 1993.

[115] D. Rhodes. Initial stages of the thermal unfolding of yeast phenylalanine transfer RNA as studied by chemical modification: the effect of magnesium. *Eur. J. Biochem.*, 81:91–101, 1977.

[116] J. D. Robertus, J. E. Ladner, J. T. Finch, D. Rhodes, R. S. Brown, B. F. Clark, and A. Klug. Correlation between three-dimensional structure and chemical reactivity of transfer RNA. *Nucl. Acids Res.*, 1:927–932, 1974.

[117] G. T. Robillard, C. E. Tarr, F. Vosman, and J. L. Sussman. An NMR approach to tRNA tertiary structure in solution. *Biophys. Chem*, 6:291–298, 1977.

[118] J. Rogers and G. F. Joyce. A ribozyme that lacks cytidine. *Nature*, 402:323–325, 1999.

[119] M. A. Rould, J. J. Perona, and T. A. Steitz. Structural basis of anti-codon loop recognition by glutaminyl-tRNA synthetase. *Nature*, 352:213–218, 1991.

[120] A. Rzhetsky, F. J. Ayala, L. C. Hsu, C. Chang, and A. Yoshida. Exon/intron structure of aldehyde dehydrogenase genes supports the "introns late" theory. *Proc. Natl. Acad. Sci*, 94:6820–6825, 1997.

[121] M. E Saks, J. R Sampson, and J. N. Abelson. The transfer RNA identity problem: a search for rules. *Science*, 263:191–197, 1994.

[122] J. R. Sampson and O. C. Uhlenbeck. Biochemical and physical characterization of an unmodified yeast phenylalanine transfer RNA transcribed in vitro. *Proc. Natl. Acad. Sci USA*, 85:1033–1037, 1988.

[123] T. Samuelsson, T. Boren, T. I. Johansen, and F. Lustig. Properties of a transfer RNA lacking modified nucleosides. *J Biol. Chem*, 263:13692–13699, 1988.

[124] S. J. Sharp, J. Schaack, L. Cooley, D. J. Burke, and D. Soll. Structure and transcription of eukaryotic tRNA genes. *Crit. Rev. Biochem.*, 19:107–144, 1994.

[125] J. M. Sherman, K. Rogers, M. J. Rogers, and D. Soll. Synthetase competition and tRNA context determine the in vivo identify of tRNA discriminator mutants. *J. Mol. Biol.*, 228:1055–1062, 1992.

[126] J. P. Shi and P. Schimmel. Aminoacylation of alanine minihelices. "Discriminator" base modulates transition state of single turnover recaction. *J. Biol. Chem.*, 266:2705–2708, 1991.

[127] M. Sprinzl, C. Horn, M. Brown, A. Ioudovitch, and S. Steinberg. Compilation of tRNA sequences and sequences of tRNA genes. *NAR*, 26:148–153, 1998.

[128] N. Stange and H. Beier. A cell-free plant extract for accurate pre-tRNA processing, splicing and modification. *EMBO J.*, 6:2811–2818, 1987.

[129] A. Stein and D. M. Crothers. Conformational changes of transfer RNA. The role of magnesium (II). *Biochemistry*, 15:160–167, 1976.

[130] A. Stein and D. M. Crothers. Equilibrium binding of magnesium(II) by Escherichia coli tRNAfmet. *Biochemistry*, 15:157–160, 1976.

[131] L. Su, L. Chen, M. Egli, J. M. Berger, and A. Rich. Minor groove RNA triplex in the crystal structure of a ribosomal frameshifting viral pseudoknot. *Nat. Struct. Biol*, 6:285–292, 1999.

[132] H. Swerdlow and C. Guthrie. Structure of intron-containing tRNA precursors. analysis of solution conformation using chemical and enzymatic probes. *J. Biol. Chem.*, 259:5197–5207, 1984.

[133] M. A. Tanner, E. M. Anderson, R. R. Gutell, and T. R. Cech. Mutagenesis and comparative sequence analysis of a base triple joining the two domains of group I ribozymes. *RNA*, 3:1037–1051, 1997.

[134] T. M. Tarasow, S. L. Tarasow, and B. E. Eaton. RNA-catalysed carbon-carbon formation. *Nature*, 389:54–57, 1997.

[135] L. D. Thompson and C. J. Daniels. A tRNA(trp) intron endonuclease from halobacterium volcanii. Unique substrate recognition properties. *J. Biol. Chem.*, 263:17951–17959, 1988.

[136] H. Uemura, M. Imai, E. Ohtsuka, M. Ikehara, and D. Soll. E. coli initiator tRNA analogs with different nucleotides in the discriminator base position. *Nucleic Acids Res.*, 10:6531–6539, 1982.

[137] P. J. Unrau and D. P. Bartel. RNA-catalysed nucleotide synthesis. *Nature*, 395:260–264, 1998.

[138] P. Valenzuela, A. Venegas, F. Weinberg, R. Bishop, and W. J. Rutter. Structure of yeast phenylalanine-tRNA genes: an intervening DNA segment within the region coding for the tRNA. *Proc. Natl. Acad. Sci. USA*, 75:190–194, 1978.

[139] Uwe von Ahsen. Translational fidelity: Error-prone versus hyper-accurate ribosomes. *Chem. Biol.*, 5:R3–R6, 1998.

[140] A. E. Walter, A. I. Murchie, D. R. Duckett, and D. M. Lilley. Global stucture of four-way RNA junctions studied using fluorescence resonance energy transfer. *RNA*, 4:719–728, 1998.

[141] J. Weissenbach, I. Kiraly, and G. Dirheimer. Primary structure of tRNAthr 1a and b from brewer's yeast. *Biochimie*, 59:381–391, 1977.

[142] S. K. Westaway, L. Cagnon, Z. Chang, S. Li, G.P. Larson, J. A. Zaia, and J. J. Rossi. Virion encapsidation of tRNA(3lys)-ribozyme chimeric RNAs inhibits HIV infection. *Antisense Nucleic Acid Drug Dev.*, 8:185–197, 1997.

[143] C. R. Woese, O. Kandler, and M. L. Wheelis. Towards a natural system of organisms: proposal for the domains archaea, bacteria and eucarya. *Proc. Natl. Acad. Sci. USA*, 87:4576–4579, 1990.

[144] M. Wu and jr I. Tinoco. RNA folding causes secondary structure rearrangement. *Proc. Natl. Acad. Sci. USA*, 95:11555–11560, 1998.

[145] S. Wuchty, W. Fontana, I. L. Hofacker, and P. Schuster. Complete suboptimal folding of RNA and the stability of secondary structure. *Biopolymers*, 1998. *submitted.*

[146] X. Ye, R. A. Kumar, and D. J. Patel. Molecular recognition in the bovine immunodeficiency virus tat peptide-TAR RNA complex. *Chem. Biol.*, 2:827–840, 1995.

[147] H. Zamora, R. Luce, and C. K. Biebricher. Design of artificial short-chained RNA species that are replicated by Qβ replicase. *Biochemistry*, 34:1261–1266, 1995.

[148] B. Zhang and T. R. Cech. Peptidyl-transferase ribozymes: trans reactions, structural characterization and ribosomal RNA-like features. *Chem. Biol.*, 5:539–553, 1998.

[149] G. R. Zimmermann, R. D. jenison, C. L. Wick, J. P. Simorre, and A. Pardi. Interlocking structural motifs mediate molecular discrimination by a theophylline-binding RNA. *Nat. Struct. Biol.*, 4:644–649, 1997.

[150] M. Zuker and D. Sankoff. RNA secondary structures and their prediction. *Bull. Math. Biol.*, 46:591–621, 1984.