Copyright
by
Jung Chull Lee
2003
The Dissertation Committee for Jung Chull Lee
certifies that this is the approved version of the following dissertation:

**Structural Studies of Ribosomal RNA Based on Cross-Analysis of
Comparative Models and Three-Dimensional Crystal Structures**

Committee:

_________________________________
Christian P. Whitman, Supervisor

_________________________________
Robin R. Gutell, Co-Supervisor

_________________________________
Patrick J. Davis

_________________________________
Sean M. Kerwin

_________________________________
David W. Hoffman
Structural Studies of Ribosomal RNA Based on Cross-Analysis of
Comparative Models and Three-Dimensional Crystal Structures

by

Jung Chull Lee, B.S., M.S.

Dissertation
Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

The University of Texas at Austin
December, 2003


**Dedication**

To my caring and patient wife, Hee Young Jung, and my lovely son and daughter,

Young Lee and Gene Lee.
Acknowledgements

I have been very fortunate in having had the opportunity to learn bioinformatics and computational biology to derive RNA secondary structures from sequence alignment of homologous RNA molecules. I would like to thank my supervisor, Dr. Robin Gutell, for his thoughtful encouragement, guidance and advice, for being a fair and consistent critic, and for his tireless patience and support during the past three years. I would also like to thank my co-supervisor, Dr. Christian Whitman, who helps and guides me throughout my graduate study. In addition, I would like to thank my committee of Dr. Patrick Davis, Dr. Sean Kerwin, and Dr. David Hoffman.

I wish to thank Jamie Cannone for his wonderful assistance. I have enjoyed great discussions on RNA structure with Kishore Doshi. I also wish to thank all former and current members in the Bioinformatics and Computational Biology Lab at the University of Texas at Austin.

Finally, I would like to thank my wife, Hee Young Jung, for her endless love and encouragement during the difficult times of my graduate study, and my lovely two children, Young Lee and Gene Lee, for the happy days they have shared with me. I would also like to thank my families in Korea, who have supported me during the course of my PhD.

Jung Chull Lee

The University of Texas at Austin

December, 2003
Structural Studies of Ribosomal RNA Based on Cross-Analysis of Comparative Models and Three-Dimensional Crystal Structures

Publication No. __________

Jung Chull Lee, Ph.D.
The University of Texas at Austin, 2003

Supervisors: Christian Whitman, Robin Gutell

Basepair mapping of the rRNAs in the high-resolution crystal structures of the Thermus thermophilus 30S and Haloarcula marismortui 50S ribosomal subunits not only proved the authenticity of covariation analysis in predicting RNA secondary structures with high accuracy, but also provided a great wealth of information on RNA structure, ranging from diverse basepair conformations to structure motifs. Based on cross-analysis of comparative structure models and crystal structures of rRNAs, a systematic and unambiguous classification for basepair conformations is established, and the structural features and biological implications for a wide variety of RNA sequence and structure motifs, including the AA.AG@helix.ends motif, the sticky motif consisting of AGUA/GAA, GUA/GAA, and GGA/GAA motifs, and the lonepair triloop motif, are discussed.
# Table of Contents

List of Tables.......................................................................................................................... xii

List of Figures .......................................................................................................................... xiii

Chapter 1: Introduction ......................................................................................................... 1
  1.1 Bioinformatics in the Post-Genomic Era ................................................................. 1
  1.2 Comparative Sequence Analysis of RNA Molecules ........................................... 2

Chapter 2: Evaluation of rRNA Comparative Structure Models .............................. 6
  2.1 Comparative Structure Models of Ribosomal RNAs ........................................... 6
  2.2 Accuracy of Comparative rRNA Structure Models ........................................... 11

Chapter 3: Diversity of Basepair Conformations in RNA Structure ...................... 27
  3.1 Background ................................................................................................................. 27
  3.2 Results ......................................................................................................................... 29
    3.2.1 Variety of basepair conformations ....................................................................... 29
    3.2.2 Topological relationships of basepair conformations ........................................ 37
    3.2.3 Structural parameters of basepair conformations ............................................. 39
    3.2.4 Sugar puckering and directionality in basepairs .................................................. 53
    3.2.5 Tautomers diversifying basepair conformations ............................................... 54
    3.2.6 Protonated Basepair Conformations .................................................................. 56
    3.2.7 Basepair conformations with bifurcated hydrogen bonds .................................. 58
    3.2.8 Isostericity of basepair conformations ............................................................... 60
    3.2.9 Dynamic changes of basepair conformations ...................................................... 63
    3.2.10 Unusual Basepair conformations in other crystal structures ....................... 65
    3.2.11 Evaluation of the existing naming systems ....................................................... 66
  3.3 Discussion ....................................................................................................................... 68
    3.3.1 AA.AG@helix.ends basepairs.............................................................................. 68
3.3.2 A-mediated tertiary interactions .............................................. 71
3.3.3 AGUA/GAA motif ...................................................................... 72
3.4 Materials and Methods ....................................................................... 74

Chapter 4: The Sticky Motif: AGUA/GAA, GUA/GAA, and GGA/GAA Motifs in RNA Structure ................................................................. 76
4.1 Background ....................................................................................... 76
4.2 Results ............................................................................................... 78
4.2.1 Analysis of the rRNAs in the 30S and 50S crystal structures .... 78
4.2.2 General features of AGUA/GAA, GUA/GAA, and GGA/GAA motifs ......................................................................................... 84
4.2.2.1 AGUA/GAA (E) motif .......................................................... 84
4.2.2.2 GUA/GAA (EL) motif .......................................................... 90
4.2.2.3 GGA/GAA (ELL) motif ...................................................... 91
4.2.3 Interactions within AGUA/GAA, GUA/GAA, and GAA/GAA motifs ..................................................................................... 92
4.2.4 Interactions of sticky motifs with RNA and proteins .......... 94
4.2.4.1 AGUA/GAA motifs in asymmetric internal loops .... 95
4.2.4.2 An AGUA/GAA motif at the core of 16S rRNA .......... 97
4.2.4.3 An AGUA/GAA motif at the core of 23S rRNA .......... 99
4.2.4.4 An AGUA/GAA motif at position 952 (Ec: 858) in 23S rRNA ................................................................. 100
4.2.4.5 An AGUA/GAA motif at position 1327 (Ec: 1346) in 16S rRNA ................................................................. 101
4.2.4.6 An AGUA/GAA motif at the multistem loop in domain II of 23S rRNA ................................................................. 102
4.2.4.7 Other AGUA/GAA motifs inducing helical stacking in multistem loops ................................................................. 103
4.2.4.8 An AGUA/GAA motif at position 1219 (Ec: 1238) in 16S rRNA ................................................................. 104
4.2.4.9 An AGUA/GAA motif at position 1971 (Ec: 1930) in 23S rRNA ................................................................. 105
4.2.4.10 GUA/GAA motifs in symmetric internal loops ....... 106
4.2.4.11 A GUA/GAA motif at position 668 (Ec: 685) in 16S rRNA ................................................................. 107
4.2.4.12 GGA/GAA motifs in symmetric internal loops ....... 108
4.2.4.13 GGA/GAA motifs in asymmetric internal loops ...... 110
4.2.5 Sticky motifs interacting with metal ions ....................... 112
4.2.6 Phylogenetic distribution of sticky motifs in rRNAs .......... 113
4.2.7 Exchanges between sticky motifs ....................................... 115
4.3 Discussion ............................................................................... 117
4.3.1 K-turns versus sticky motifs ............................................... 117
4.3.2 GUA/GAA motif candidates not formed in the crystal structures ............................................................... 119
4.3.3 A base triple in the L11-binding domain ......................... 120
4.3.4 Sticky motifs in small catalytic RNAs .............................. 121
4.3.5 Possible roles of sticky motifs in RNA structure ............... 122
4.4 Materials and Methods ............................................................ 123

Chapter 5: The Lonepair Triloop: A New Motif in RNA Structure .......... 125
5.1 Background ........................................................................... 125
5.2 Results .................................................................................. 126
5.2.1 Initial identification of lonepair triloops with covariation analysis ................................................................. 126
5.2.2 Analyses of the tRNA and 30S and 50S ribosomal crystal structures ......................................................... 127
5.2.3 Classification of lonepair triloops ...................................... 128
5.2.4 Overall three-dimensional architecture of the lonepair triloop motif ........................................................ 137
5.2.5 Coaxial stacking of lonepairs .......................................... 140
5.2.6 Lonepair triloop-RNA interaction .................................... 143
5.2.6.1 R1 LPTLs (UGNRA) ...................................................... 143
5.2.6.2 R2 LPTLs (UYYRA) .......................................................... 145
5.2.6.3 R3 LPTLs (NUAAN or NRWAN) ........................................ 146
5.2.6.4 R4 LPTLs (NRYAN) .......................................................... 148
5.2.6.5 R5 LPTLs (NCNUN) .......................................................... 150

5.2.7 Lonepair triloop-protein interactions ...................................... 150
5.2.8 Sugar puckering, bases facing, and directionality of local chains ............................................................ 152
5.2.9 Lonepair types, exchanges, and conformations ...................... 154
5.2.10 Conservation of ribosomal LPTLs ........................................ 156
5.2.11 Distribution of lonepair triloops on the rRNA three-dimensional structures ................................................. 158
5.2.12 Solvent accessibility of lonepair triloops ............................. 160

5.3 Discussion ............................................................................. 162
5.3.1 U-turns and the structural integrity of a lonepair triloop ...... 162
5.3.2 RNA packing and the structural integrity of lonepair triloops 164
5.3.3 Conformations of lonepairs ................................................... 166
5.3.4 Molecular mimicry ............................................................... 167
5.3.5 Functional implications of lonepair triloop motifs ................. 168

5.4 Materials and Methods ......................................................... 170
5.4.1 Structural analysis of rRNAs ................................................. 170
5.4.2 Coaxial stacking ................................................................. 171
5.4.3 Solvent accessibility of LPTLs .............................................. 172

Chapter 6: Conclusions and Discussion........................................... 174
7.1 Conclusions and Discussion .................................................. 174
7.2 Perspective ............................................................................. 180
Appendix A List of Basepairs in the *T. thermophilus* 16S rRNA Comparative Structure Model and in the Crystal Structure of the 30S Ribosomal Subunit ................................................................................................... 182

Appendix B List of Basepairs in the *H. marismortui* 23S rRNA Comparative Structure Model and in the Crystal Structure of the 50S Ribosomal Subunit ................................................................................................... 198

Appendix C List of Basepairs in the *H. marismortui* 5S rRNA Comparative Structure Model and in the Crystal Structure of the 50S Ribosomal Subunit ................................................................................................... 229

Appendix D Comparative Model of the *T. thermophilus* 16S rRNA ..................... 231

Appendix E Comparative Model of the *H. marismortui* 23S rRNA: 5’-half ..... 232

Appendix F Comparative Model of the *H. marismortui* 23S rRNA: 3’-half ..... 233

Appendix G Correspondence of the Lee-Gutell and Three Other Naming Systems for Classifying Basepair Conformations ..................................................... 234

Appendix H AA.AG@helix.ends Basepairs in 16S and 23S rRNAs ................. 237

Appendix I Identified Sticky Motifs in rRNAs and their Sequence Variations in Aligned Sequences ........................................................................................................ 240

Appendix J Gallery of Sticky Motifs in the Variable Regions of rRNAs .......... 247

Appendix K Lonepair Triloop Motifs Occurring in Three Phylogenetic Domains and Two Organelles .................................................................................................................. 253

Bibliography .................................................................................................... 256

Vita ............................................................................................................... 269
**List of Tables**

Table 2.1: Accuracy of the rRNA comparative structure models ................. 22
Table 3.1: Basepairs and their conformations in the rRNAs in the *T. thermophilus* 30S and *H. marismortui* 50S crystal structures ........ 31
Table 3.2: Distribution of AA.AG@helix.ends basepairs.......................... 70
Table 4.1: Consensus sequence context and location of the sticky motif....... 79
Table 4.2: Sticky motifs and their molecular interactions ....................... 81
Table 4.3: Single nucleotide frequencies for conserved sticky motifs ........ 82
Table 4.4: Frequencies and phylogenetic distribution of rRNA sticky motifs.. 83
Table 5.1: Analyses of LPTLs in rRNA and tRNA sequences and structures 129
Table 5.2: Distribution of lonepair loops ............................................. 131
Table 5.3: Single nucleotide frequencies of LPTL nucleotides.................. 132
Table 5.4: Further grouping of class I and II LPTLs ............................... 133
Table 5.5: Solvent accessible surface areas for LPTL nucleotides............. 161
List of Figures

Fig. 2.1: The Noller-Woese-Gutell comparative model for 16S rRNA ...........7
Fig. 2.2: The Noller-Woese-Gutell comparative model for 23S rRNA: the
5'-half. .............................................................................................8
Fig. 2.3: The Noller-Woese-Gutell comparative model for 23S rRNA: the
3'-half. .............................................................................................9
Fig. 2.4: The Noller-Woese-Gutell comparative model for 5S rRNA. .......... 10
Fig. 2.5: Tentative interactions in 16S rRNA .............................................. 12
Fig. 2.6: Tentative interactions in the 5'-half of 23S rRNA......................... 13
Fig. 2.7: Tentative interactions in the 3'-half of 23S rRNA......................... 14
Fig. 2.8: Basepair mapping of the T. thermophilus 16S rRNA. ............... 16
Fig. 2.9: Basepair mapping of the H. marismortui 23S rRNA: the 5'-half..... 17
Fig. 2.10: Basepair mapping of the H. marismortui 23S rRNA: the 3'-half.....18
Fig. 2.11: Basepair mapping of the H. marismortui 5S rRNA................... 19
Fig. 2.12: Basepair mapping of the D. radiodurans 23S rRNA: the 5'-half.....23
Fig. 2.13: Basepair mapping of the D. radiodurans 23S rRNA: the 3'-half.....24
Fig. 2.14: Basepair mapping of the D. radiodurans 5S rRNA....................25
Fig. 3.1: Distribution of basepair conformations in the T. thermophilus
16S rRNA .......................................................................................... 33
Fig. 3.2: Distribution of basepair conformations in the H. marismortui 23S
rRNA: the 5'-half ............................................................................. 34
Fig. 3.3: Distribution of basepair conformations in the H. marismortui 23S
rRNA: the 3'-half ............................................................................. 35
Fig. 3.4: Distribution of basepair conformations in the *H. marismortui* 5S rRNA. ................................................................. 36

Fig. 3.5: Schematic representation of the topological relationships between basepair conformations for a given basepair group........ 38

Fig. 3.6: Basepair conformations for the C:G basepair group. ................. 40

Fig. 3.7: Basepair conformations for the U:A basepair group............... 41

Fig. 3.8: Basepair conformations for the U:G basepair group............... 42

Fig. 3.9: Basepair conformations for the G:A basepair group............... 43

Fig. 3.10: Basepair conformations for the C:A basepair group............... 44

Fig. 3.11: Basepair conformations for the U:C basepair group............... 45

Fig. 3.12: Basepair conformations for the A:A basepair group............... 46

Fig. 3.13: Basepair conformations for the C:C basepair group............... 47

Fig. 3.14: Basepair conformations for the G:G basepair group............... 48

Fig. 3.15: Basepair conformations for the U:U basepair group............... 49

Fig. 3.16: Structural parameters for the Watson-Crick, wobble, slipped Watson-Crick, and slipped wobble conformations. .............. 51

Fig. 3.17: Structural parameters for the sheared, reversed Hoogsteen, Hoogsteen, and reversed sheared conformations. ................. 52

Fig. 4.1: Structure gallery of AGUA/GAA motifs in the crystal structures... 85

Fig. 4.2: Structure gallery of GUA/GAA motifs in the crystal structures..... 86

Fig. 4.3: Structure gallery of GGA/GAA motifs in the crystal structures..... 87

Fig. 4.4: Representative AGUA/GAA, GUA/GAA, and GGA/GAA motifs and their stereoviews. ...................................................... 93
Fig. 4.5: All identified AGUA/GAA, GUA/GAA, and GGA/GAA motifs mapped onto comparative structure models. .......................... 114
Fig. 5.1: Lonepair triloops in the crystal structures................................. 130
Fig. 5.2: Schematic representation of LPTLs and their classification........ 135
Fig. 5.3: Gallery of LPTL-containing RNA fragments. .......................... 138
Fig. 5.4: Stereoviews for representative IA, IB and IIA LPTLs............... 139
Chapter 1: Introduction

1.1 Bioinformatics in the Post-Genomic Era

Bioinformatics and Computational Biology refers to a scientific discipline integrating the insights of mathematics and computer science with those of biology for understanding intricate modern biological problems from determining the function of novel genes or proteins in complex biological processes to analyzing and interpreting genomic and proteomic data that are being uncovered in large volumes. Findings made with bioinformatics are expected to have applications in health care for treating diseases, pharmaceutics for developing new drugs, agriculture for developing and improving agriculturally important plants, as well as phylogeny for investigating evolutionary relationships between organisms and species.

During the past few years, we have witnessed an overwhelming amount of new biological data, including the complete sequence of human genome or genetic blueprint of life,\textsuperscript{1,3} whereby opening the door to a new age in our analysis of biological systems. The completed sequence will help identify the 25,000 to 30,000 genes, including those involved in complicated diseases such as cancer and diabetes, and provide an enormous opportunity to better understand diseases
and accelerate the development of new therapeutics. Thus, bioinformatics is on great demand in the post-genomic era in searching for various hidden information in the genomic sequence, including data for expressed, mutated, and deleted genes.

Bioinformatics can be largely divided into two main areas: functional bioinformatics and structural bioinformatics. The latter mainly focuses on the prediction of the secondary and tertiary structures of biological macromolecules such as RNA and proteins, while the former is largely interested in understanding the functions of genes and proteins associated with a wide variety of biological processes.

1.2 Comparative Sequence Analysis of RNA Molecules

One of the greatest challenges in bioinformatics and computational biology is to determine from any given RNA sequence its secondary structure and its biologically active three-dimensional structure. As the number of nucleotides in an RNA sequence increases, the number of possible secondary structure models increases exponentially (greater than $1.8^n$, where $n$ is the number of nucleotides). For example, the number of possible secondary structures is approximately $2.5 \times 10^{19}$ for a tRNA sequence with 76 nucleotides, $4.3 \times 10^{393}$ for a 16S rRNA sequence with 1,500 nucleotides, and $6.3 \times 10^{740}$ for a 23S rRNA sequence with 3,000 nucleotides. Thus, identifying one correct secondary structure from these
astronomical numbers of possible secondary structures for a given RNA sequence requires *a priori* knowledge of the rules on RNA structure, including thermodynamic and kinetic properties of RNA structural elements, metal and protein interactions and binding that affect RNA folding, and relationships between sets of sequences and unique RNA structural motifs. In this respect, even the most significantly improved RNA folding algorithms based only on thermodynamic calculations\textsuperscript{5-6} are not able to accurately and reliably predict the correct secondary structure from such a larger number of theoretically possible secondary structures.\textsuperscript{7-8} Alternatively, however, it is currently possible with "comparative sequence analysis" to accurately identify the correct secondary structure common in all of the functionally homologous RNA molecules simply by aligning their sequences; comparative sequence analysis does not require any *a priori* knowledge or principles about RNA structure. Covariation analysis, a special application of comparative sequence analysis, specifically searches for the common structure by identifying all positions that covary with one another (e.g. C:G ↔ G:C ↔ U:A ↔ A:U, U:G ↔ C:A, U:U ↔ C:C, A:A ↔ G:G) in an alignment of homologous RNA sequences.\textsuperscript{9}

Shortly after the first few tRNA sequences were determined in the 1960s, it was rationalized that different tRNA sequences will fold into a similar secondary and tertiary structure, leading to the proposal of the tRNA structure,\textsuperscript{10-14}
which were subsequently validated with tRNA crystal structures. The comparative rationale employing multiple RNA sequences has been the foundation for the prediction of the secondary structure models for 5S, 16S, and 23S rRNAs. Initially, the canonical basepairs consisting of the standard Watson-Crick (C:G, G:C, U:A, and A:U) and wobble (U:G and G:U) basepairs which occur within potential helices in 5S, 16S, and 23S rRNAs were identified. With more advanced covariation algorithms and a significant increase in the number of rRNA sequences, it is now possible to search for all positions with the same patterns of variation (i.e., covariations), regardless of the types of basepairs and the proximity of those basepairs to other paired or unpaired nucleotides. The majority of the basepairs identified with this unrestricted search for positional covariations are canonical, which are arranged into standard secondary helices.

Although this form of covariation analysis only searches for positional covariations, the results independently determined the two most fundamental principles in RNA structure: the Watson-Crick basepair relationship and the antiparallel arrangement of consecutive basepairs. Given this success, we question if comparative sequence analysis will reveal other structural motifs and the rules for their prediction. A growing number of novel structural motifs have been identified with covariation analysis, which include the following non-canonical basepair exchanges U:U ↔ C:C, A:A ↔ G:G, A:G ↔ G:A, G:U ↔ A:C,
pseudoknots, base triples, U-turns, tetraloops (GNRA, UUCG, and CUUG), tandem GA basepairs, E and E-like loops, G:U basepairs, AA.AG@helix.ends basepairs, lonepair triloops, tetraloop receptors, short- and long-range tertiary basepairs arranged antiparallel or parallel with one another, K-turns, and H-turns.

Beyond the comparative studies of tRNAs and rRNAs, comparative approach has been applied with great success to a wide variety of other RNA molecules including group I and II introns, RNase P RNA, telomerase RNA, tmRNA, U RNA, and SRP RNA.
Chapter 2: Evaluation of rRNA Comparative Structure Models

2.1 COMPARATIVE STRUCTURE MODELS OF RIBOSOMAL RNAs

Currently, more than 400 covariation-based 16S, 23S, and 5S rRNA secondary structure models are available online at the Comparative RNA Web (CRW) Site (http://www.rna.icmb.utexas.edu). All are based on the *Escherichia coli* 16S, 23S, and 5S rRNA structure models. The most recent rRNA secondary structure models for the *E. coli* 16S, 23S, and 5S rRNAs, which have been used as the reference structure diagrams, are illustrated in Figures 2.1, 2.2, 2.3, and 2.4, where basepairs with a red identifier ("—" for the standard Watson-Crick basepairs (C:G, G:C, U:A, and A:U), •" for the canonical wobble basepairs (U:G and G:U), "○" for G:A and A:G, and "●" for other non-canonical basepairs) are predicted with the greatest confidence. We have less confidence in the basepairs with a green, black, grey, and blue identifiers, due to their lower covariation scores. While these latter basepairs have less positional covariation than the “red” basepairs, they are predicted due to their high frequencies of the standard Watson-Crick and/or wobble basepairs that are adjacent and antiparallel to the red basepairs with a strong covariation signal. These comparative secondary structure models have evolved over the past 20 years with significant increases in the
Fig. 2.1 The Noller-Woese-Gutell comparative model for 16S rRNA
**Fig. 2.2** The Noller-Woese-Gutell comparative model for 23S rRNA: the 5'-half
Fig. 2.3 The Noller-Woese-Gutell comparative model for 23S rRNA: the 3’-half
Escherichia coli
(V00336)

Fig. 2.4 The Noller-Woese-Gutell comparative model for 5S rRNA
number of complete rRNA sequences and more advanced covariation algorithms. The most current Noller-Woese-Gutell comparative structure models are the culmination of the comparative analysis of approximately 7,000 16S, 1,050 23S, 700 5S rRNA sequences from all branches of the phylogenetic tree, which have been aligned into one of the seven alignments: T, Three Phylogenetic Domains + Two Organelles (3P+2O); 3, Three Phylogenetic Domains (3P); B, Bacteria; A, Archaea; C, Chloroplasts; E, Eukaryotes; M, Mitochondria.61-62

In addition to the secondary and tertiary basepairs with a significant amount of positional covariations, covariation analysis has also identified some basepairs and base triples with a lower degree of positional covariations (called tentative interactions), which are shown separately on the reference *E. coli* 16S and 23S rRNA structure diagrams (Figures 2.5, 2.6, and 2.7).61

2.2 ACCURACY OF COMPARATIVE rRNA STRUCTURE MODELS

More recently, the atomic-resolution crystal structures for the bacterial *Thermus thermophilus* 30S (PDB 1FJF 63 at 3.05 Å, pH 6.5), the archaeal *Haloarcula marismortui* 50S (PDB 1FFK 64 and 1JJ2 45 at 2.40 Å, pH 5.8) ribosomal subunits prompted us to map all the basepairs in their rRNAs, allowing the quantitative evaluation of the rRNA comparative structure models.62 The authenticity of all the basepairs predicted with covariation analysis, including the canonical basepairs within a regular helix and the large variety of
Fig. 2.5 Tentative interactions in 16S rRNA

Escherichia coli (J01695)
Fig. 2.6 Tentative interactions in the 5'-half of 23S rRNA
Fig. 2.7 Tentative interactions in the 3'-half of 23S rRNA
non-canonical basepairs, was determined from mapping of basepairs in the rRNAs in these high-resolution crystal structures. To evaluate the current Noller-Woese-Gutell structure models against the crystal structures, the mapped basepairs are overlaid onto the *T. thermophilus* 16S (Fig. 2.8) and the *H. marismortui* 23S (Figures 2.9 and 2.10) and 5S rRNA (Fig. 2.11) comparative structure models: red, basepairs predicted with covariation analysis and present in the crystal structures (called +/- basepairs); green, basepairs predicted but not present in the crystal structures (called +/- basepairs); blue, basepairs not predicted but present in the crystal structures (called -/+ basepairs); magenta, tentative interactions or basepairs predicted with a motif-based analysis and present in the crystal structures; brown, interactions involving sugar phosphate backbones; purple, basepairs or unpaired nucleotides not resolved in the crystal structures; colored open square, used for clarity only; colored open circle, the third nucleotide in base triples; colored open rectangle, basepairs in base triples; black tick marks, crystal structure numbering (*T. thermophilus* numbering for the 16S and *H. marismortui* numbering for the 23S and 5S rRNA); red tick marks, *E. coli*-equivalent numberings for the 16S, 23S, and 5S rRNAs.

In addition to the covariation-based or comparative basepairs, the basepair mapping has identified many canonical and non-canonical basepairs, which are not predictable with covariation analysis due to the lack of positional covariations.
**Fig. 2.8** Basepair mapping of the *T. thermophilus* 16S rRNA
Fig. 2.9 Basepair mapping of the *H. marismortui* 23S rRNA: the 5'-half
Fig. 2.10 Basepair mapping of the *H. marismortui* 23S rRNA: the 3'-half
**Haloarcula marismortui**
(AF034620)

**Fig. 2.11** Basepair mapping of the *H. marismortui* 5S rRNA
All of the basepairs in the appropriate comparative structure models and the mapped basepairs in the 16S, 23S, and 5S rRNAs in the crystal structures of the *T. thermophilus* 30S and *H. marismortui* 50S ribosomal subunits are tabulated into Appendices A, B, and C, respectively. The fields of information with a brief description are: C, basepairs present (+) or absent (-) in the comparative models; X, basepairs present (+) or absent (-) in the crystal structures; CC, color-coded covariation ratings for basepairs (red for confident basepairs; green, black, grey, and blue for less confident basepairs); bpC, basepair conformations (WC for Watson-Crick, Wb for wobble, sWC for slipped Watson-Crick, sWb for slipped wobble, rWC for reversed Watson-Crick, rWb reversed wobble, H for Hoogsteen, rH for reversed Hoogsteen, S for sheared, rS for reversed sheared, fS for flipped sheared, pfS parallel flipped sheared, pS for parallel sheared, and rpS for reversed parallel sheared); RS, rasmol scripts for visualizing basepairs; Motifs, motifs that basepairs are involved in (A for AA.AG@helix.ends motifs, GA for tandem GA motifs, E for AGUA/GAA motifs, EL for GUA/GAA motifs, ELL for GGA/GAA motifs, and LPTL for lonepair triloop motifs); BT, basepairs involved in base triples; BQ, basepairs involved in base quadruples; CA, comparative information (CA for covariation-based basepairs and TT for tentatives). The comparative structure models for the *T. thermophilus* 16S and the *H. marismortui* 23S rRNAs.
are illustrated for reference in Appendices D, E, and F, with the color-coded covariation ratings and the helix numbers.65-66

The basepair mapping results are summarized in Table 2.1. Approximately 97–98% of the comparative basepairs in the 16S and 23S rRNAs are indeed present in the two ribosomal crystal structures,62 validating the underlying fundamental principle of comparative analysis that RNA molecules with different sequences and an equivalent biological function will adopt a similar secondary and tertiary structure. As well, our implementation of this principle, including the alignment of sequences, the development and implementation of the covariation algorithms, and the interpretation of the output from these programs and alignments, has also been validated.

Shortly after the *T. thermophilus* 30S and the *H. marismortui* 50S ribosomal crystal structures were determined, the crystal structure for the *Deinococcus radiodurans* 50S ribosomal subunit (PDB 1KC9 and 1LNR at 3.10 Å, pH 7.8)67 was resolved, prompting us to compare these two 50S ribosomal crystal structures. The basepair mapping results for the 23S and 5S rRNAs in the *D. radiodurans* 50S crystal structure are superimposed onto their respective comparative structure models (Figures 2.12, 2.13, and 2.14), using the same color-codes and symbols as in the basepair-mapped structure diagrams for the *T. thermophilus* 16S and the *H. marismortui* 23S and 5S rRNA structures. Although
Table 2.1 Accuracy of the rRNA comparative structure models

<table>
<thead>
<tr>
<th>rRNA</th>
<th>CA†</th>
<th>Xtal‡</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>red (+/+)</td>
<td>green (+/-)</td>
<td>magenta</td>
</tr>
<tr>
<td>16S (Tr)</td>
<td>473</td>
<td>458</td>
<td>15</td>
</tr>
<tr>
<td>23S (Hm)</td>
<td>812</td>
<td>797</td>
<td>15</td>
</tr>
<tr>
<td>5S (Hm)</td>
<td>36</td>
<td>36</td>
<td>0</td>
</tr>
</tbody>
</table>

† Number of basepairs in the comparative models
‡ Number of basepairs in the crystal structures
red (+/+): basepairs predicted with covariation analysis and present in the crystal structures
green (+/-): basepairs predicted with covariation analysis but not present in the crystal structures
magenta: basepairs predicted with covariation-based motif analysis and present in the crystal structures
blue (-/+): basepairs not predicted with covariation analysis and present in the crystal structures

Table 2.1 Accuracy of the rRNA comparative structure models
Fig. 2.12 Basepair mapping of the *D. radiodurans* 23S rRNA: the 5'-half
Fig. 2.13 Basepair mapping of the D. radiodurans 23S rRNA: the 3’-half
Fig. 2.14 Basepair mapping of the *D. radiodurans* 5S rRNA
the *H. marismortui* is in the Archaea phylogenetic domain and *D. radiodurans* is in the Bacteria phylogenetic domain, the functionally equivalent 23S rRNAs in the two 50S ribosomal subunits indeed fold into an almost identical secondary and tertiary structure. However, the *D. radiodurans* 23S rRNA crystal structure has less tertiary interactions than the *H. marismortui* 23S rRNA crystal structure, which is probably due to the lower resolution of the *D. radiodurans* crystal structure. As well, the sugar phosphodiester backbones in these two 23S rRNAs are in a good agreement with that in the 23S rRNA in the 5.5 Å resolution crystal structure for the *T. thermophilus* 70S ribosomal subunit (PDB 1GIX and 1GIY), which contains A-, P-, and E-site tRNAs and an mRNA fragment, confirming the fundamental premise of comparative sequence analysis.
Chapter 3: Diversity of Basepair Conformations in RNA Structure

3.1 BACKGROUND

Recently, the high-resolution crystal structures of the bacterial Thermus thermophilus 30S (PDB 1FJF\textsuperscript{63}) and archaeal Haloarcula marismortui 50S (PDB 1FFK\textsuperscript{64} and 1JJ2\textsuperscript{45}) ribosomal subunits were determined; the former includes the 16S rRNA and the latter the 23S and 5S rRNAs. An analysis of the basepairs present in the rRNAs in the two crystal structures not only validated the authenticity of the covariation-based rRNA structures models,\textsuperscript{62} but also provides a wealth of RNA structural folds, conformations and motifs to identify and relate to nucleotide sequences and basepairs. In addition to the canonical basepairs with canonical conformations consisting of the standard Watson-Crick (C:G, G:C, U:A, and A:U)\textsuperscript{69} and the canonical wobble (U:G and G:U)\textsuperscript{70} basepair types and conformations, these crystal structures contain many canonical and non-canonical basepairs with non-canonical conformations. The non-canonical conformations are frequently involved in a variety of motifs, including U-turns,\textsuperscript{22-23} GNRA, UUCG, and CUUG tetraloops,\textsuperscript{24-28} G·U basepair motifs,\textsuperscript{37} AA.AG@helix.ends motifs,\textsuperscript{38} tandem GA motifs,\textsuperscript{29-30} lonepair triloop motifs,\textsuperscript{39-40} K-turns,\textsuperscript{45} H-turns.\textsuperscript{46}
A-minor motifs, \textsuperscript{71,72} and \textit{sticky motifs} consisting of AGUA/GAA, GUA/GAA, and GGA/GAA motifs.\textsuperscript{73}

Many of the previously identified non-canonical basepairs have been organized onto a web site (http://prion.bchs.uh.edu/bp_type/),\textsuperscript{74} and a nomenclature has been proposed to classify basepair conformations by introducing the interacting edges,\textsuperscript{75} on which a computational approach was developed to automatically identify basepair conformations from RNA crystal structures.\textsuperscript{76} While this proposed naming system is not analogous to the widely accepted and traditionally established system (\textit{e.g.}, Watson-Crick, wobble, reversed Watson-Crick and wobble, Hoogsteen, reversed Hoogsteen, and sheared conformations) which includes the original basepair conformations,\textsuperscript{69-70,77-78} it primarily contains basepairs with two hydrogen bonds; consequently, it is not applicable to many basepairs with a single hydrogen bond that frequently occur in the crystal structures. Thus, regardless of the number of involved hydrogen bond(s), a simple and widely applicable nomenclature for basepair conformations is needed to unambiguously describe all the observed basepairs and to study RNA structures systematically and consistently. By analyzing topological arrangements of the bases and glycosidic bonds in a given basepair and by expanding the traditional classification, we propose a new simple and systematic nomenclature to classify all observed basepairs.
Beyond this classification, statistical analysis of structural parameters for basepair conformations and their sugar puckering are established. In addition, abnormal basepair conformations involving either keto-enol and amino-imino tautomerism or bifurcated hydrogen bonds are described. Furthermore, isosteric relationships and dynamic changes between basepair conformations are discussed along with their functional implications. Finally, the existing systems for naming basepair conformations are evaluated based on this classification.

3.2 RESULTS

3.2.1 Variety of basepair conformations

The basepair conformation refers to the spatial arrangement of the two bases in a given basepair, which are normally hydrogen bonded to one another. In addition to the standard Watson-Crick C:G and U:A and wobble U:G basepair types with canonical basepair conformations, an analysis of the basepair conformations in the rRNAs in the crystal structures of the *T. thermophilus* 30S (PDB 1FJF) and *H. marismortui* 50S (PDB 1FFK and 1JJ2) ribosomal subunits reveals many canonical and non-canonical basepair types with non-canonical conformations. All of these basepairs and their conformations were topologically analyzed; almost all of the basepairs directly flanked on both sides by two basepairs within regular helices (called middle or internal basepairs) have
the canonical conformations, while the basepairs at the helical termini (called ending or terminal basepairs directly flanked on a single side by an internal basepair) frequently have the non-canonical conformations. However, the vast majority of the basepairs with non-canonical conformations either occur in the unpaired regions in the RNA secondary structure models based on covariation analysis or are the basepairs involved in base triple interactions.

While all 16 possible basepairs are divided into ten basepair groups (i.e., C:G, U:A, U:G, G:A, C:A, U:C, A:A, C:C, G:G, and U:U), their basepair conformations are classified into 14 conformational families (Table 3.1): Watson-Crick (WC), wobble (Wb), slipped Watson-Crick (sWC), slipped wobble (sWb), reversed Watson-Crick (rWC), reversed wobble (rWb), Hoogsteen (H), reversed Hoogsteen (rH), sheared (S), reversed sheared (rS), flipped sheared (fS), parallel flipped sheared (pfS), parallel sheared (pS), and reversed parallel sheared (rpS).

Four significant trends are identified in Table 3.1. (1) The most frequent basepair groups are C:G (51%) > U:A (20%) > G:A (11%) > U:G (10%) > remaining six basepair groups (9%). (2) The most frequent basepair conformations are WC (66%) > S (11%) > Wb (10%) > rH (5%) > remaining ten basepair conformations (8%). (3) 98% of the C:G basepairs have the WC conformation; 71% and 17% of the U:A basepairs form the WC and rH
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>BP</td>
<td>WC</td>
<td>869</td>
<td>252</td>
<td>1</td>
<td>21</td>
<td>2</td>
<td>8</td>
<td>-</td>
<td>-</td>
<td>#</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Wb</td>
<td>1</td>
<td>1</td>
<td>124(3)</td>
<td>6</td>
<td>9</td>
<td>3</td>
<td>17</td>
<td>168</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sWC</td>
<td>2</td>
<td>4(2)</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>2</td>
<td>#</td>
<td>#</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>sWb</td>
<td>3</td>
<td>#</td>
<td>(1)</td>
<td>#</td>
<td>4</td>
<td>#</td>
<td>#</td>
<td>3</td>
<td>#</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>rWC</td>
<td>4</td>
<td>11</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>rWb</td>
<td>-</td>
<td>2</td>
<td>-</td>
<td>3</td>
<td>-</td>
<td>5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>1(1)</td>
<td>8</td>
<td>-</td>
<td>3(2)</td>
<td>-</td>
<td>(1)</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>rH</td>
<td>-</td>
<td>58(1)</td>
<td>-</td>
<td>1</td>
<td>13(1)</td>
<td>1</td>
<td>(1)</td>
<td>9</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>-</td>
<td>7(2)</td>
<td>1(7)</td>
<td>142</td>
<td>8</td>
<td>(3)</td>
<td>5</td>
<td>10</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>rS</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>9</td>
<td>2</td>
<td>1</td>
<td>19</td>
<td>-</td>
<td>#</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>rS</td>
<td>1</td>
<td>(1)</td>
<td>(2)</td>
<td>6</td>
<td>5</td>
<td>1(1)</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>pS</td>
<td>3</td>
<td>2</td>
<td>(1)</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>pS</td>
<td>-</td>
<td>1</td>
<td>1(2)</td>
<td>1</td>
<td>2</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>rpS</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>#</td>
<td>-</td>
<td>#</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>884</td>
<td>355</td>
<td>152</td>
<td>186</td>
<td>45</td>
<td>18</td>
<td>51</td>
<td>17</td>
<td>22</td>
<td>1745</td>
</tr>
</tbody>
</table>

| TQ  | WC  | 3   | 4   | -   | 1   | -   | 2   | -   | -   | #   | -    | 10    |
|     | Wb  | -   | -   | -   | -   | 1   | -   | 1   | -   | 1   | -    | 4     |
|     | sWC | (1) | -   | #   | -   | #   | -   | #   | #   | #   | #    | 1     |
|     | sWb | #   | #   | -   | -   | #   | -   | #   | -   | -   | -    | 0     |
|     | rWC | -   | 1   | -   | -   | -   | #   | -   | -   | -   | -    | 1     |
|     | rWb | -   | -   | 1   | -   | -   | -   | 9   | -   | -   | -    | 10    |
|     | H   | -   | 4   | 2   | -   | 1   | -   | 2   | 6   | 1   | -    | 16    |
|     | rH  | -   | 2(5)| 5   | (1) | 3   | 8(3)| -   | 2   | -   | 1    | 33    |
|     | S   | -   | 3   | (1) | 5   | 11  | 1   | 2   | 1   | 7   | -    | 31    |
|     | rS  | 2(1)| 1   | -   | 56  | 1   | 3   | 2   | #   | 4   | -    | 70    |
|     | rS  | (1) | (1) | (6) | 18  | -   | (1) | -   | -   | -   | -    | 29    |
|     | pS  | 1   | (1)| 1(2)| 3(2)| 9   | 3(5)| 1(1) | 4   | 1   | -    | 34    |
|     | pS  | 1   | (1)| 4   | 1(16)| 3  | (2) | 5   | 2   | 1   | 2    | 39    |
|     | rpS | (1) | (2)| 4(2)| 19  | #   | -   | #   | -   | #   | 3    | 31    |
|     | Total| 22  | 27  | 39  | 114 | 35  | 18  | 8   | 25  | 7   | -    | 309   |

The pound sign (#) indicates the basepair conformations that are unlikely to form.

**Table 3.1** Basepairs and their conformations observed in the rRNAs from the *T. thermophilus* 30S and *H. marismortui* 50S crystal structures.
conformations, respectively; 84% of the U:G basepairs have the Wb conformation; 76% and 11% of the G:A basepairs have the S and WC conformations, respectively; more than half of the C:A basepairs adopt the rH (31%) and S (24%) conformations, while 37% and 20% of the A:A basepairs assume the rS and S conformations, respectively. (4) Of the ten basepair groups and 14 theoretically possible conformations for each basepair type, 72 or 60% of a total of 121 possible nucleotide and conformational arrangements occur in simple basepairs (BP in **Table 3.1**) and 69 occur in the higher-order interactions consisting of base-basepair and basepair-basepair interactions (TQ in **Table 3.1**) in the rRNAs in the *T. thermophilus* 30S and *H. marismortui* 50S crystal structures. In contrast, the higher-order interactions have a wide variety of non-canonical basepair conformations; the most commonly observed conformations are rS (23%) > pS (13%) > pfS (11%) = rH (11%) > S (10%) = rpS (10%) > fS (9%), while the most common basepair groups involved in the higher-order interactions are G:A (37%) > U:G (13%) > C:A (11%) > U:A (9%). The distribution of basepair conformations is shown on the basepair-mapped 16S, 23S and 5S rRNAs (**Figures 3.1, 3.2, 3.3, and 3.4**): red, WC or Wb; magenta, S; cyan, rH; brown, sWC or sWb; black, rWC or rWb; blue, other non-canonicals.
Fig. 3.1 Distribution of basepair conformations in the *T. thermophilus* 16S rRNA
Fig. 3.2 Distribution of basepair conformations in the *H. marismortui* 23S rRNA: the 5'-half
Fig. 3.3 Distribution of basepair conformations in the *H. marismortui* 23S rRNA: the 3'-half
Fig. 3.4 Distribution of basepair conformations in the *H. marismortui* 5S rRNA
3.2.2 Topological relationships of basepair conformations

Basepair conformations can be systematically and unambiguously named based on the topological arrangements of the two bases and the two glycosidic bonds in a given basepair. All of the non-canonical basepair conformations are simply derived by manipulating the Y base of the starting Watson-Crick (WC) or wobble (Wb) conformation in each basepair group as depicted in Fig. 3.5: Bases are represented as triangles and glycosidic bonds as thick lines attached to triangles. Each conformation is obtained by simply manipulating Y or X:  s, shearing; f, flipping; r, reversing; p, paralleling; sl, slipping. For example, sWC/sWb is generated by slipping (translating base Y either along the negative y-axis (sWC/sWb) or along the positive y-axis (sWC*/sWb*) to form only a single hydrogen bond); rWC/rWb, by reversing (rotating nucleotide Y 180° about the x-axis); H, by flipping (rotating either base Y (H) or base X (H*) 180° about its glycosidic bond); rH, by flipping and then reversing base Y (rH) or base X (rH*); S, by shearing (either translating base Y along the negative y-axis and then along the negative x-axis (S) or translating base X along the negative y-axis and then along the positive x-axis (S*)); rS, by shearing and then reversing base Y (rS) or base X (rS*); fS, by flipping and then shearing base Y (fS) or base X (fS*); pfS, by flipping, shearing, and then paralleling nucleotide Y (pfS) or X (pfS*) (rotating either nucleotide Y or X about the y-axis to have the glycosidic bonds that run
**Fig. 3.5** Schematic representation of the topological relationships between basepair conformations for a given basepair group, X:Y. For simplicity, the starting Watson-Crick and wobble conformations are represented as X:Y WC/Wb in a box with a shaded border in the center. The alternative conformations are shown with an asterisk (*) in the parentheses. The dotted arrow shows other conformations that are not simply derived by manipulating either of Y and X.
parallel into the same direction); pS, by paralleling and then shearing either Y (pS) or X (pS*); rpS, by paralleling, shearing, and then reversing either X or Y. The order of successive manipulations of the base Y (or X) does not matter. The topological relationships of the observed and theoretically possible conformations for the ten basepair groups are shown in Figures 3.6-3.15: C:G, Fig. 3.6; U:A, Fig. 3.7; U:G, Fig. 3.8; G:A, Fig. 3.9; C:A, Fig. 3.10; U:C, Fig. 3.11; A:A, Fig. 3.12; C:C Fig. 3.13; G:G, Fig. 3.14; U:U, Fig. 3.15. Interestingly, some basepair conformations within the same conformation family have multiple hydrogen bonded interactions (e.g., H, rH*, fS* for U:G in Fig. 3.8 and rH for G:G in Fig. 3.14), while others have the potential to reverse the bases X or Y (e.g., rS for G:A in Fig. 3.9). Overall, each of the ten basepair groups can adopt at least ten major basepair conformations.

3.2.3 Structural parameters of basepair conformations

Hydrogen bonds are weak and largely electrostatic in nature because of the partial positive hydrogen atom from the donor and the partial negative acceptor atom. Thus, two bases involved in many basepairs are not co-planar in the crystal structures; they are frequently propeller twisted, sometimes buckled, staggered, stretched, or sometimes open, suggesting that basepair conformations are in constant motion with respect to each other. The structural parameters in basepair conformations are likely to vary. The average structural parameters of the
Fig. 3.6 Basepair conformations for the C:G basepair group
Fig. 3.7 Basepair conformations for the U:A basepair group
Fig. 3.8 Basepair conformations for the U:G basepair group
Fig. 3.9 Basepair conformations for the G:A basepair group
Fig. 3.10 Basepair conformations for the C:A basepair group
Fig. 3.11 Basepair conformations for the U:C basepair group
Fig. 3.12 Basepair conformations for the A:A basepair group
Fig. 3.13 Basepair conformations for the C:C basepair group
Fig. 3.14 Basepair conformations for the G:G basepair group
Fig. 3.15 Basepair conformations for the U:U basepair group
representative basepair conformations, including dDAs (the donor-acceptor distances involved in hydrogen bonding interactions), dCC (the C1'-C1' distance), and <X and <Y values (the N1-C1'-C1' and N9-C1'-C1' angles) of the two basepairing nucleotides, are illustrated in Fig. 3.16 and 3.17.

In Fig. 3.16 and 3.17, the dCCs for the standard WC conformations (C:G WC and U:A WC) are 10.6 Å and 10.5 Å, respectively, while those of the U:G Wb and G:A WC conformations are 10.4 Å and 12.6 Å, respectively. Interestingly, the dDAs of the WC and Wb conformations with the potential to form multiple hydrogen bonds usually get shorter from the major to the minor groove. However, U:C WC and C:C Wb conformations have much longer dDAs toward the minor groove; they are mostly embedded within regular helices and have openings toward the minor groove with the dCCs of 11.1 Å and 10.1 Å, respectively. Compared to the dCCs for the standard WC conformations, however, the U:U Wb basepair has the much shorter dCC of 8.7 Å but are usually embedded within helices. The G:A S and A:A S conformations also have the shorter dCCs of 9.5 Å and 9.8 Å and are dominantly formed immediately outside covariation-derived regular helices; the A and G are directly 5' and 3' to the regular helices, respectively.38,73 While the U:A rH conformation has the shorter dCC of 9.7 Å, the C:A rH conformation possesses the dCC of 10.9 Å.

While the difference between the two angles, <X and <Y, for canonical
Fig. 3.16 Structural parameters for the Watson-Crick, wobble, slipped Watson-Crick, and slipped wobble conformations
Fig. 3.17 Structural parameters for the sheared, reversed Hoogsteen, Hoogsteen, and reversed sheared conformations
and non-canonical basepairs with the WC conformation is less than 10°, it is 20-40° for those with the Wb conformation, around 50-60° for those with the sWC or sWb conformation, and more than 70° for those with the S conformation; <X is usually larger than <Y. However, the asterisked basepair conformations including the Wb*, sWC*, sWb*, S*, fS* conformations have the opposite trend; <X is less than <Y. In this regard, many basepair conformations can be unambiguously determined based on the <X and <Y values.

3.2.4 Sugar puckering and directionality in basepairs

11% (185 out of 1745) of the total number of simple basepairs in the rRNAs in the *T. thermophilus* 30S and the *H. marismortui* 50S ribosomal crystal structures do not have the expected C3'-endo sugar puckering for at least one of the two participating nucleotides in a basepair; 26 of them have the C2'-endo puckering for both nucleotides. These include 25 C:G WC, 12 U:A WC, and three U:G Wb canonical conformations and 21 U:A rH, 20 G:A S, 16 A:A rS non-canonical conformations. Strikingly, four basepairs, including 410:423 (*Ec: 415:428*) and 670:686 (*Ec: 687:702*) in the 16S and 2485:2536 (*Ec: 2450:2501*) in the 23S rRNA, have the unexpected O4'-endo puckering, while 181 basepairs have the C2'-endo puckering.

All of the comparative internal basepairs but only three C:G WC basepairs at positions 1555:1566 (*Ec: 1448:1463*), 1827:2021 (*Ec: 1771:1980*), and
1853:1878 (*Ec: 1797:1822*) in the 23S rRNA, have the C3'-endo puckering for both basepairing nucleotides. The remaining 178 basepairs containing the C2'-endo puckering occur in terminal basepairs, in lonepairs, or in basepairs involved in motifs. Nonetheless, no correlation between basepair conformations and sugar puckering is observed. Thus, such regions containing basepairs with the C2'-endo or O4'-endo puckering will be highly dynamic, probably causing several different basepair conformations for a specific basepair. Previously, the C2'-endo puckering was considered as a signature for the instability in RNA structure.39

When it comes to directionality of the two local chains in a basepair, almost all of the simple basepairs are antiparallel. However, two local chains in 24 exceptional basepairs run parallel: three C:G WC, three U:A WC, five G:A fS, three A:A rS, three C:G rS, two G:A rS, three U:U rWb and rWb*, and two C:G fS and fS*. All of the Wb, S, sWC, and sWb conformations run antiparallel. Interestingly, four G:A rS conformations have neither anti-parallel nor parallel orientations, but their two local chains are aligned perpendicularly or have vertical orientations to each other. Taken together, RNA molecules will be highly flexible.

3.2.5 Tautomers diversifying basepair conformations

Keto-enol tautomerism frequently occurs in such carbonyl-containing compounds as ketones and aldehydes. Previously, ¹H NMR studies, employing
the imino proton resonances in the basepairs of $^{15}$N-labeled *E. coli* tRNAs and pH 7.0, suggested that, when forming C:G and U:A basepairs, 5-15% of the bases are present as unusual tautomers.\(^{80}\) The dynamic and tautomeric shifting either between keto and enol forms of G and U (keto-enol tautomerism) or between amino and imino forms of A and C (amino-imino tautomerism) contributes to the diversity of basepair conformations.

For example, U1068:G1081 (*E. coli* numbering (*Ec*): U1089:G1099) in the 16S rRNA adopts the U:G Wb\(^{*}\) conformation ([Fig. 3.8](#)). Strikingly, however, it is clear that the both bases form their enol tautomers because: (1) the dDA between NH of U and $-\text{NH}_2$ of G is only 2.75 Å, so that the both bases are not allowed to simultaneously have keto forms due to steric clash; (2) both the N3-C4 bond in U, d(N3-C4), and the N1-C6 bond of G, d(N1-C6), are shortened by 0.02 Å than usual to have double bond character; (3) both carbonyl groups facing toward the major groove, have unusually longer bond distances by 0.02 Å than normal to have single bond character. Similarly, G647:G724 (*Ec*: G664:G741) in the 16S rRNA forms the enolized G:G Wb conformation with the separation of 2.84 Å between $-\text{NH}_2$ of one G and NH of the next ([Fig. 3.14](#)), of which both G’s are almost co-planar and should have their enol tautomers.

Six more examples of keto-enol tautomerism are observed in the crystal structures: (1) U1003:G1018 (*Ec*: U1025:A1036) in the 16S rRNA and U9:G16
in the crystal structure of domain A of *Thermus flavus* 5 S rRNA (PDB 353D81) adopts the enolized U:G WC (Fig. 3.8); (2) U660:G696 (Ec: U677:G713) in the 16S rRNA forms the enolized U:G Wb* (Fig. 3.8); (3) U1108:U1130 (Ec: U1126:U1148) in the 16S rRNA assumes the enolized U:U sWb (Fig. 3.15) (4) U137:G140 (Ec: not homologous basepair) in the 23S rRNA adopts the enolized U:G S (Fig. 3.8); (5) G644:U903 (Ec: 587:810) in the 23S rRNA forms the U:G pS (Fig. 3.8); (6) U1169:A1177 (Ec: U1065:A1073) in the 23S rRNA has the enolized U:A Wb (Fig. 3.7).

Compared to the examples undergoing the keto-enol tautomerization, only two cases experiencing the amino-imino tautomerization are present in the 23S rRNA: A288:C364 (Ec: A282:U358) and A2488:C2534 (Ec: A2453:C2499). Both bases in the two cases are spatially arranged to form the unusual C:A WC conformation (Fig. 3.10), suggesting that either C or A should have the imino tautomer. However, amino-imino tautomerism in basepairs seems to be not as frequent as keto-enol tautomerism.

### 3.2.6 Protonated basepair conformations

C:A Wb and C:C Wb conformations can have two hydrogen bonds, one of which results from protonation of A at N1 and of C at N3, respectively. A spectroscopic study of the *E. coli* tRNA\(^{\text{Ala}}\) acceptor stem showed that N1 of the C:A Wb conformation is protonated at pH 5.0-5.5 and unprotonated at pH 7.0-
A $^1$H NMR study indicated that, upon forming DNA triplexes, the C:C Wb conformation is protonated up to pH 7.0 but completely unprotonated at pH 7.6.\(^{83}\)

Interestingly, C1384:A1477 (Ec: 1402:1500) in the *T. thermophilus* 16S rRNAs in the native 30S and substrate-bound 30S crystal structures exists in a very similar conformation with the putatively protonated C:A Wb conformation even at pH 7.8; the distances from C=O of C to N1 of A, d(O2-N1), is 2.41 Å in the native 16S rRNA and 2.24 Å in the ligand-bound 16S rRNA. C963:A1005 (Ec: U868:A909) in the *H. marismortui* 23S rRNA adopts a similar C:A Wb conformation with d(O2-N1) of 2.58 Å. In addition, C240:C278 in the 16S rRNAs in the *T. thermophilus* 30S crystal structures with and without substrates, albeit with the different pH, forms the supposedly protonated C:C Wb conformation; the distances from C=O of one C to N3 of the other, d(O2-N3), are 2.47 Å and 2.65 Å, respectively. However, all of the eight remaining C:C Wb conformations formed in the *H. marismortui* 23S rRNAs have 1.0-2.0 Å longer d(O2-N3) than 2.65 Å, leading to an opening toward the minor groove (Figures 3.16 and 3.17); as described above, all of them are directly flanked by two internal basepairs in comparative regular helices. Moreover, most of the water molecules interacting with basepairs in the *H. marismortui* 50S crystal structure are located in the major groove, not at the minor groove, thereby not allowing the protonation of C at N3 and A at N1 from the minor groove. In this respect, C:A
Wb and C:C Wb conformations without protonation are likely to form. Thus, protonation may not be required to form the C:A Wb and C:C Wb conformations.

3.2.7 Basepair conformations involving bifurcated hydrogen bonds

Bifurcated hydrogen bonding interactions usually occur: (1) when one hydrogen atom simultaneously interacts with two acceptor atoms (type I); (2) when one acceptor atom simultaneously contacts with two hydrogen atoms (type II); (3) when two hydrogen atoms from the donor make contacts with two different acceptor atoms (type III); (4) when one hydrogen atom interacts with one acceptor atom while the donor interacts with another hydrogen atom (type IV). For example, the type II bifurcated hydrogen bonds systematically and commonly occurs in β-sheets. Some bifurcated hydrogen bonds in RNA have been previously reported.

Based on our analysis of basepair conformations, the bifurcated hydrogen bonding interactions are not frequently observed upon forming basepairs in the rRNAs in the T. thermophilus 30S and the H. marismortui 50S ribosomal crystal structures. Nonetheless, the C:A sWb conformation adopted by C930:A1040 (Ec: C837:A941) in the 23S rRNA forms a type I bifurcated hydrogen bond; one hydrogen atom of –NH₂ of A interacts with both N3 and C=O of C. In contrast, two of the three G:G Wb basepairs (Table 3.1) are distorted to contain two hydrogen bonds that result from the simultaneous interaction of NH and –NH₂ of
one G with C=O of the next (type II), as observed in G104:G326 (Ec: G111:G331) in the 16S rRNA and G504:G512 (Ec: G498:G506) in the 23S rRNA. Such a very similar G:G Wb conformation is observed with G76:G100 in the crystal structure of the *E. coli* 5S rRNA fragment in complex with L25 (PDB 1DFU85). Interestingly, the glycosidic bonds of these three G:G Wb conformations involving bifurcated hydrogen bonds are almost reversed, probably being the intermediate for transition from G:G Wb to G:G rH conformation.

The U:G fS and U:G pfS conformations are also likely to contain bifurcated hydrogen bonds, when C=O of U simultaneously interacts with both NH and –NH₂ of G (type II). Surprisingly, the U:G fS conformation with bifurcated hydrogen bonds features the five UUCG and UNCG tetraloops in the 16S and 23S rRNAs; these include U338:G341 (Ec: U343:G346), U415:G418 (Ec: U420:G423), U1117:G1120 (Ec: U1135:G1138), and U1430:G1433 (Ec: U1450:G1453) in the 16S and U1770:G1773 (Ec: U1692:G1695) in the 23S rRNA. The bifurcated hydrogen bonds, together with another hydrogen bond between 2'-OH of U and C=O of G, probably stabilize the formation of the UUCG or UNCG tetraloops.

Type III bifurcated hydrogen bonds do not occur in simple basepairs, but frequently occur in higher-order interactions involving unpaired A’s. For example, when the two basepairs, G1247:A1250 (Ec: G1266:A1269) and
G1293:C1306 (Ec: G1312:C1325), interact with each other, the two hydrogen atoms of –NH₂ of the G1293 simultaneously interact with N3 of A1250 and C=O of C1306. Type IV bifurcated hydrogen bonds also occur only in higher-order interactions forming base triples and mostly involve 2'-OH, as observed in the (G106:C309)A348 base triple in the 16S rRNA; the hydrogen and oxygen atoms of 2'-OH of A348 interact with N3 and 2'-OH of G106, respectively. Interestingly, when the two basepairs, C2833:G2847 (Ec: G2816:C2830) and G2851:A2906 (Ec: G2834:A2883), interact with each other, they form both of the type III and IV bifurcated interactions: while –NH₂ of G2847 forms a bifurcated hydrogen bond with C=O of C2833 and N3 of A2906, 2'-OH of A2906 makes a bifurcated hydrogen bond with C=O and 2'-OH of C2833. In this respect, the type III and IV bifurcated hydrogen bonding interactions may play a role in stabilizing globular RNA structures by increasing hydrogen bonds in long-range tertiary interactions.

3.2.8 Isostericity of basepair conformations

Two or more basepair types with the same conformation may be structurally equivalent or isosteric. These isosteric basepairs covary or exchange with one another in an alignment of homologous RNA molecules, without necessarily affecting the overall three-dimensional structures. The best-known isosteric basepairs, the standard Watson-Crick C:G and U:A basepair types, are
also isosteric with G:A and U:C basepairs in the Watson-Crick conformation. In contrast, the U:G Wb and C:A Wb conformations are equivalent to each other but are not strictly isosteric to their alternative conformations, G:U Wb and A:C Wb, respectively. Nonetheless, both the U:G WC (enol) conformation (Fig. 3.8) and the C:A WC (imino) conformation (Fig. 3.10) are indeed equivalent to the standard Watson-Crick conformation: The former undergoes keto-enol tautomerism as observed in U9:G16 in the crystal structure of domain A of T. flavus 5S rRNA (PDB 353D81); the latter undergoes amino-imino tautomerism as observed in A288:C364 (Ec: A282:U358) and A2488:C2534 (Ec: A2453:C2499) in the 23S rRNA. Thus, the keto-enol or amino-imino tautomerism could be responsible for the exchange between the standard Watson-Crick basepairs and either of the wobble U:G or C:A basepairs.

Additionally, the C:G, U:A, G:A, and U:C basepairs also have the potential to form their corresponding wobble conformations (Figures 3.6, 3.7, 3.9, and 3.11). For example, G940:C1026 (Ec: C848:G930) and U1169:A1177 (Ec: U1065:A1073) in the H. marismortui 23S rRNA assume their respective wobble conformations, C:G Wb and U:A Wb, which are isosteric to the canonical wobble conformation, U:G Wb. Moreover, the G:G Wb (enol) conformation (Fig. 3.14), adopted by G647:G724 (Ec: G664:G741) in the 16S rRNA, is equivalent to C:G, U:A, U:G, G:A, C:A, U:C, A:A, C:C, and U:U basepairs in the
wobble conformation. However, the U:G Wb* conformation (Fig. 3.8), formed by U660:G696 (Ec: U677:G713) and U1068:G1081 (Ec: U1086:G1099) in the 16S and U2586:G2592 (Ec: C2551:G2557) in the 23S rRNA, is isosteric not to the canonical U:G Wb but to the canonical G:U Wb conformation. Thus, all the basepair types in the ten basepair groups can form their respective wobble conformations. Unlike all other basepair types that can form the Watson-Crick conformation through either keto-enol or amino-imino tautomerism, however, the G:G basepair is unlikely to adopt the Watson-Crick conformation.

When two consecutive nucleotides on an RNA strand are basepaired, they form the characteristic, non-canonical parallel sheared (pS) conformation; the first set of examples were observed in the adenosine platform motif in the group I intron. Several examples of this type of basepairing occur in the rRNAs at positions G175:U176 (Ec: A181:A182) and U624:A625 (Ec: U641:A642) in the 16S, C1105:A1106 (Ec: A1008:A1009), G1119:U1120 (Ec: G1022:U1023), and G1235:A1236 (Ec: G1131:U1132) in the 23S, and A51:A52 in the 5S rRNAs. In these examples, the leading nucleotide with the lower position number always moves into the major groove and the nucleotide with the higher number into the minor groove. A similar but not identical example involves the two consecutive bases A1193 and A1194 (Ec: A1089 and A1090) in the L11-binding region of the 23S rRNA that exchange with G and U respectively. These two positions form a
basepair with the parallel sheared (pS) conformation, while the second position A1194 (Ec: A1090) forms a regular basepair with position U1205 (Ec: U1101) to form a base triple. Moreover, the consecutive GU bases in the AGUA/GAA motif also form the parallel sheared conformation, U:G pS* (Fig. 3.8).31

The Hoogsteen (H) and reversed Hoogsteen (rH) conformations in the U:A and C:A basepair groups can also form in the basepair types C:G, U:G, G:A, A:A, and G:G. For example, two tertiary basepairs, G294:G549 (Ec: G299:G566) in the 16S and G604:G607 (no homologous basepair in E. coli) in the 23S rRNAs, adopt the Hoogsteen (H) conformation (Fig. 3.14), which is the same as that observed in the NMR and crystal structures of G-quadruplex DNAs formed by telomeric DNA sequences (PDB 139D90 and 1JPQ91). Furthermore, the exchange between G:A and A:A (or sometimes C:A) at the ends of helices38 can be explained by the isosteric sheared (S) conformation of those basepair groups. Therefore, isosteric basepairs are simply identified from the comparison of basepair conformations and can rationalize basepair exchanges in an alignment of homologous RNA sequences.

3.2.9 Dynamic changes of basepair conformations

When bases on a single RNA chain are vertically projected from the backbone to minimize steric clashes between bases and sugar rings and,
simultaneously, when any two consecutive sugar rings are helically twisted to minimize steric collisions in the backbone, then the helical base stacking is an intrinsic property of DNA and RNA molecules. For these situations, basepairs within regular helices will usually adopt the standard Watson-Crick and wobble conformations for structural integrity. In contrast, basepairs outside of or at the termini of regular helices can adopt a variety of other conformations and can potentially undergo conformational changes depending on structural perturbation(s). For example, positions 1390:1470 (Ec: 1408:1493) in the 16S rRNA form a G:A or A:A basepair with a wobble (Wb) conformation at one stage of protein synthesis and are unpaired at another stage to participate in the recognition of the first codon-anticodon basepair.92

Beyond this example, six other sets of different basepair conformations at the same positions were observed in the 23S rRNA from the two crystal structures of the H. marismortui 50S ribosomal subunit: one without the L11-binding domain (PDB 1FFK), and the second with the L11-binding domain region (PDB 1JJ2). The different basepair conformations and their locations are: (1) S and fS (Fig. 3.9) at a tertiary basepair, G314:A337 (Ec: G307:A330), (2) H* and WC (Fig. 3.6) at a lonepair in a lonepair triloop motif, C326:G330 (Ec: G319:C323); (3) H and pfS (Fig. 3.8) at a tertiary basepair in a multistem loop, G644:G903 (Ec: C587:U810); (4) H and WC (Fig. 3.6) at a basepair at the end of a helix and
adjacent to a three-way junction, C789:G856 (Ec: C698:G763); (5) S and fS (Fig. 3.9) at a basepair in a hairpin loop, G873:A876 (Ec: G780:A783); and (6) rH and rS (Fig. 3.12) at a tertiary basepair in the peptidyl transferase loop, A2103:A2538 (Ec: A2062:A2503). The structural and possible functional dynamics associated with these different conformations at the same basepair is currently not known.

3.2.10 Unusual basepair conformations in other crystal structures

As noted earlier in Table 3.1, approximately half of the total number of theoretically possible basepair types and conformational arrangements were observed in the rRNAs in the crystal structures of the T. thermophilus 30S (PDB 1FJF 63) and the H. marismortui 50S subunits (PDB 1FFK, 64 and 1JJ2 45). Six of the arrangements not present in the rRNAs in the T. thermophilus 30S and H. marismortui 50S crystal structures analyzed here occur in other crystal structures. (1) C16:C59 in the E. coli Cys-tRNA crystal structure (PDB 1B23 93) adopts the reversed Watson-Crick (rWC) conformation (Fig. 3.13); this same conformation was observed in the telomeric C-rich sequences that form an unusual intercalated DNA structure known as an i-motif (PDB 105D 94). The i-motif DNA is stabilized by TMPyP4, a DNA-binding cationic porphyrin causing chromosomal destabilization.95 (2) U:G H* (Fig. 3.8) is formed for the G80:U96 basepair in the crystal structure of the domain E of the T. flavus 5S rRNA (PDB 361D 96). (3) U:G rH* (Fig. 3.8) is assumed for the U168:G188 basepair in the crystal
structures of the P4-P6 domain of the *Tetrahymena thermophila* group I intron (PDB 1GID<sup>97</sup> and 1HR2<sup>98</sup>). (4) G:G sWb (*Fig. 3.14*) is formed for the G28:G40 basepair in the crystal structure of the UUCG tetraloop (PDB 1F7Y<sup>99</sup>). (5) G:A rH* is formed for the G622:A646 in the (C613:G622)A646 base triple in the crystal structure of the *Saccharomyces cerevisiae* Asp-tRNA complexed with Asp-tRNA synthetase (PDB 1ASY<sup>100</sup>), which can be enolized to have two hydrogen bonds (left of the two G:A rH* structures in *Fig. 3.9*) and is equivalent to the G22:G26 rH* conformation of the (C13:G22)G26 base triple in the *Saccharomyces cerevisiae* Phe-tRNA crystal structure (PDB 6TNA<sup>101</sup>). (6) U:U rWC (*Fig. 3.15*) is adopted for the U1301:U1339 basepair (*Ec*: G1288:U1326) in the 23S rRNA in the crystal structure of the *D. radiodurans* 50S crystal structure (PDB 1LNR<sup>67</sup>), which is equivalent to U:C rWC (*Fig. 3.11*) for the C1394:U1432 basepair (*Ec*: G1288:U1326) in the *H. marismortui* 23S rRNA. Thus, we anticipate that more of the theoretically possible arrangements of basepair types and conformations in *Table 3.1* that have not already been observed will be present in other RNA crystal and NMR structures.

### 3.2.11 Evaluation of the existing naming systems

This new classification system describes all possible arrangements and orientations of two bases that are hydrogen bonded with one another, without involving any interaction from the 2'-OH groups in the sugar phosphate backbone.
In contrast, the Leontis-Westhof (L-W) system that is based on three interacting edges (Watson-Crick (WC), Hoogsteen (H), and Sugar (S)) does consider the interactions from the 2'-OH groups. The currently available naming systems are compared in Appendix G. While the GC and AU cis WC/WC (and trans WC/WC) conformations in the L-W system correspond to WC (and rWC) conformations in the new system introduced here, GU cis WC/WC (and trans WC/WC) conformations to Wb (and rWb) conformations, and AG trans H/S and AG trans S/S conformations to the G:A S and G:A rS conformations, respectively, the new system describes more basepair conformations compared to the existing naming systems. In particular, basepair conformations with a single hydrogen bond are now included and have a formal name. In addition, the new system considers keto-enol and amino-imino tautomerisms of bases that increase the number of possible basepair conformations; for example, all 16 basepair types can adopt the sheared basepair conformations that are isosteric with one another. Moreover, the new system does not specifically distinguish the protonation-stabilized basepair conformations, while the previous system requires the protonation in some basepairs including C:G H in Fig. 3.6 (C′G cis WC/H), C:G rH in Fig. 3.6 (C′G trans WC/H), G:A H* in Fig. 3.9 (A′G cis WC/H), C:A Wb in Fig. 3.10 (A′C cis WC/WC), and C:C Wb in Fig. 3.13 (C′C cis WC/WC). These protonated basepairs may prevent the further tautomerization of basepairs.
For example, the protonated C:A Wb (A+C cis WC/WC) and C:C Wb (C+C cis WC/WC) basepairs are unlikely to convert into their tautomeric forms, C:A WC (Fig. 3.10) and C:C WC (Fig. 3.13), respectively. Furthermore, the previous L-W system does not distinguish between WC and Wb or between rWC and rWb conformations. For example, C:C rWb and C:C rWC conformations (Fig. 3.13) use the same name, CC trans WC/WC (Appendix G). Finally, the new system may topologically trace the variance of the basepair conformations occurring in structural regions that are not static. This new system is easily applied to the identification of the isosteric basepair conformations as well as to the description of the higher-order interactions, including base triples and quadruples, which have been observed in RNA structures.

3.3 DISCUSSION

3.3.1 AA.AG@helix.ends basepairs

The G:A basepairs are the third most common basepair group in the rRNAs in the *T. thermophilus* 30S and *H. marismortui* 50S crystal structures (Table 3.1). The G:A basepairs frequently occur at the ends of helices with the A 5’ and G 3’ to a flanking regular helix; they commonly exchange with A:A (or C:A) (Appendix H). This orientation of G:A and A:A basepairs, described earlier in the AA.AG@helix.ends motif family, usually adopt the G:A S and A:A S (G:A S and
A:A A3, our previous nomenclature at the CRW AA.AG site, http://www.icmb.utexas.edu/ANALYSIS/AAAG/) conformations, respectively. However, some other exceptional conformations are also formed.\textsuperscript{15} The quantitative ranking of the frequencies of the basepair conformations for the 76 AA.AG@helix.ends basepairs formed in the \textit{T. thermophilus} 16S and \textit{H. marismortui} 23S rRNA crystal structures is: S (80%; G:A S and A:A A3) > WC and Wb (8%; G:A I and A:A A1) = rS (8%; G:A G3 and A:A A7) > remaining three conformations (4%) including the G:A H* (G:A G7), A:A rH (A:A A7-1), and another unusual conformation (G:A G1) (Table 3.2). Instead of G:A and A:A, the two locations, 218:222 (\textit{Ec}: 247:251) and 1133:1229 (\textit{Ec}: 1029:1125) in the \textit{H. marismortui} 23S rRNA contain C:A and A:C with the C:A S conformation, respectively; the C:A S is isosteric to the G:A S. The sheared basepairs have a shorter dCC (8.5–10.0 Å) compared to the standard Watson-Crick basepairs (10.5 Å) in the A-form helix. This structural characteristic of the sheared A:A and G:A basepairs at the ends of helices may either stabilize the helix ends by preventing any potential structural perturbation from being further propagated into helical stems or may be recognized by proteins or other regions of rRNA.

The AA.AG@helix.ends basepairs with the WC or Wb conformation are involved in helical stacking.\textsuperscript{38} Although the dCC’s of these A:A Wb and G:A WC conformations are about 2 Å longer than those of the standard Watson-Crick
<table>
<thead>
<tr>
<th>Loop Type</th>
<th>Hairpin</th>
<th>Internal</th>
<th>Multi-Stem</th>
<th>T</th>
<th>Xtal</th>
</tr>
</thead>
<tbody>
<tr>
<td>16S rRNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Invariant</td>
<td>7[7,0,0]</td>
<td>[7,0,0]</td>
<td>9[6,0,3]</td>
<td>[3,2,1]</td>
<td>5[4,0,1]</td>
</tr>
<tr>
<td>G:A</td>
<td>7[7,0,0]</td>
<td>[7,0,0]</td>
<td>4[4,0,0]</td>
<td>[3,1,0]</td>
<td>4[3,0,1]</td>
</tr>
<tr>
<td>A:A</td>
<td>0[0,0,0]</td>
<td>[0,0,0]</td>
<td>5[2,0,3]</td>
<td>[0,1,1]</td>
<td>1[1,0,0]</td>
</tr>
<tr>
<td>Exchange</td>
<td>2[2,0,0]</td>
<td>[2,0,0]</td>
<td>10[9,1,0]</td>
<td>[7,1,1]</td>
<td>9[4,0,5]</td>
</tr>
<tr>
<td>Total</td>
<td>9[9,0,0]</td>
<td>[9,0,0]</td>
<td>19[15,1,3]</td>
<td>[10,3,2]</td>
<td>14[8,0,6]</td>
</tr>
<tr>
<td>%Xtal</td>
<td>9/9=100%</td>
<td></td>
<td>15/18=90%</td>
<td></td>
<td>8/14=57%</td>
</tr>
<tr>
<td>23S rRNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Invariant</td>
<td>11[9,2,0]</td>
<td>[9,0,0]</td>
<td>13[10,2,1]</td>
<td>[9,0,1]</td>
<td>13[6,1,6]</td>
</tr>
<tr>
<td>G:A</td>
<td>11[9,2,0]</td>
<td>[9,0,0]</td>
<td>9[9,0,0]</td>
<td>[9,0,0]</td>
<td>9[6,1,2]</td>
</tr>
<tr>
<td>A:A</td>
<td>0[0,0,0]</td>
<td>[0,0,0]</td>
<td>4[1,2,1]</td>
<td>[0,0,1]</td>
<td>4[0,0,4]</td>
</tr>
<tr>
<td>Exchange</td>
<td>4[2,1,1]</td>
<td>[2,0,0]</td>
<td>12[8,3,1]</td>
<td>[8,0,0]</td>
<td>20[9,6,5]</td>
</tr>
<tr>
<td>Total</td>
<td>15[11,3,1]</td>
<td>[11,0,0]</td>
<td>25[18,5,2]</td>
<td>[17,0,1]</td>
<td>33[15,7,11]</td>
</tr>
<tr>
<td>%Xtal</td>
<td>11/12=92%</td>
<td></td>
<td>18/20=90%</td>
<td></td>
<td>15/26=58%</td>
</tr>
<tr>
<td>rRNA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>24[20,3,1]</td>
<td>[20,0,0]</td>
<td>44[33,6,5]</td>
<td>[27,3,3]</td>
<td>47[23,7,17]</td>
</tr>
<tr>
<td>%Xtal</td>
<td>20/21=95%</td>
<td></td>
<td>33/38=87%</td>
<td></td>
<td>23/40=58%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>S: 100%</th>
<th>S: 82%</th>
<th>S: 61%</th>
<th>S: 80%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>W: 9%</td>
<td>W: 13%</td>
<td>W: 8%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>O: 9%</td>
<td>O: 26%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* C[+,NH,–]: C, number of predicted basepairs based on the bacterial alignment; +, number of predicted basepairs in the crystal structure; NH, number of predicted basepairs not homologous in the crystal structure; –, number of predicted basepairs not present in the crystal structure.

*S,W,O* represents basepair conformations: S, sheared; W, Watson-Crick or Wb; O, others.

T, the total number of homologous basepairs in each category in the comparative structure model.

Xtal, the total number of basepairs present in the crystal structure.

The percentage of comparatively predicted basepairs that are present in the crystal structure: + / (C-NH).

Table 3.2 Distribution of AA.AG@helix.ends basepairs

In contrast to the G:A basepairs at helix ends that have their A 5' and G 3' to a helix, the G:A basepairs with their A 3' and G 5' to a helix always adopt the WC conformation (previously known as G:A imino) and are usually involved in coaxial stacking. These include G105:A310 (Ec: G112:A315), A137:G215 (Ec: A143:G220), A316:G327 (Ec: A321:G332), G922:A1217 (Ec: G945:A1236), G1142:A1157 (Ec: G1160:A1176), and A1338:G1347 (Ec: A1357:G1365) in the T. thermophilus 16S and G71:A106 (Ec: G75:A111) and G2582:A2596 (Ec: A2547:U2561) in the H. marismortui 23S rRNA. These basepairs and their characteristics were recently redescribed as the cis Watson-Crick A/G basepairs.

3.3.2 A-mediated tertiary interactions

The N1 and N3 positions of A are nucleophilic due to the electron-donating –NH₂ group, allowing many unpaired A’s in the secondary structure to
be involved in tertiary interactions with other sections of the RNA chain. A’s frequently interact in the minor groove with C:G (or sometimes U:G and U:A) within helical stems. The most common A-mediated tertiary interactions employ the N3 position as observed for G:A basepairs with the conformations rS (known as type I A-minor motif\textsuperscript{71-72}) and rpS (Fig. 3.9). However, the A-mediated tertiary interactions utilizing the N1 position occur frequently, forming G:A fS and G:A pfS basepair arrangements. When an unpaired A interacts with C or U, its –NH\textsubscript{2} hydrogen bonds to the pyrimidine C=O in the minor groove as observed in C:A pfS and fS (Fig. 3.10) and U:A fS basepair arrangements (Fig. 3.7).

3.3.3 AGUA/GAA motif

The AGUA/GAA sequence motif, also known as the E loop and S-turn motif,\textsuperscript{31-32,35-36} occurs at many locations in the 16S and 23S rRNAs.\textsuperscript{73} Analysis of the rRNAs in the crystal structures of the \textit{T. thermophilus} 30S and the \textit{H. marismortui} 50S ribosomal subunits revealed that the AGUA/GAA motif consistently forms three consecutive non-canonical basepairs in the 5’ to 3’ direction, A:A rS, G(U:A) base triple with U:A rH, and A:G S. The U in the G(U:A) base triple is hydrogen bonded to A to form the U:A rH, while the two consecutive nucleotides G and U form the G:U pS (Fig. 3.8). Most of the AGUA/GAA motifs occurring in internal loops contain an additional non-canonical basepair directly next to the A:A basepair,\textsuperscript{35} which adopts either the
sheared (S) or slipped wobble (sWb) conformation depending on the nucleotide sequence of the AGUA/GAA motif; the sheared (S) conformation occurs with the 5'-AGUA/GAA-3' sequence, while the slipped wobble (sWb) conformation occurs with the 5'-AGUA/GAC-3' sequence. The sheared (S) conformation is observed in G865:A886 (Ec: G888:A909) in the 16S, C162:C173 (Ec: C192:A203), U211:C228 (Ec: C240:C257), U1368:C2056 (Ec: A1264:A2015) in the 23S, and G76:A105 (no homologous positions in E. coli) in the 5S rRNAs; the slipped wobble (sWb) conformation occur in C295:C356 (no homologous positions in E. coli) and C571:C586 (no homologous positions in E. coli) in the 23S rRNA. Surprisingly, the A:C oppositions in the AGUA/GAC motifs, which are homologous to the A:A basepair of the AGUA/GAA motif, do not form any hydrogen bonds, although they are spatially arranged in the same fashion to form a pseudo-basepair, C:A rS; instead, they form a hydrogen bond between –NH₂ of A and the pro-R phosphate oxygen atom. The G of the AGUA sequence does not basepair with the C of the GAC sequence, instead it hydrogen bonds to the U that is 3' of the G, although the G and C are immediately adjacent to the U:A rH basepair. A similar situation occurs in U2690:C2704 (Ec: U2653:C2667) in the sarcin/ricin loop in the 23S rRNA; the U:C S conformation forms with the AGUA/GAA motif in the crystal structures of the H. marismortui 23S rRNA (PDB 1FFK ⁶⁴ and 1JJ2 ⁴⁵), while the U:C sWb conformation forms with the
AGUA/GAC motif in the sarcin/ricin loop crystal structure of the Escherichia coli 23S rRNA (PDB 480D\textsuperscript{103}). In contrast, U1326:U1358 (Ec: U1345:U1376) in the 16S rRNA, which is contiguous to the AGUA/GAA motif occurring in a multistem loop, forms the U:U rWb conformation. This indicates that basepair conformations also depend on the sequence and structural context.

### 3.4 MATERIALS AND METHODS

Basepairs in the rRNAs in the crystal structures of the *T. thermophilus* 30S (PDB 1FJF\textsuperscript{63}) and *H. marismortui* 50S (PDB 1FFK\textsuperscript{64} and 1JJ2\textsuperscript{45}) ribosomal subunits were divided into ten basepair groups, C:G, U:A, U:G, G:A, C:A, U:C, A:A, C:C, G:G, and U:U. All of the basepair conformations observed in the crystal structures were then visually displayed using the RasMol program\textsuperscript{104-105} to measure their structural parameters and classified into 14 major families (Table 3.1) based on the relative arrangement of the two bases and two glycosidic bonds of a given basepair as described above. Hydrogen bonds were considered typically when the distance between the hydrogen bond donor and acceptor, dDA, is less than 3.5 Å.

The two 16S rRNAs in the native 30S (PDB FJF\textsuperscript{63} at 3.05 Å, pH 6.50) and the substrate-bound 30S (PDB 1I94\textsuperscript{106} at 3.20 Å, pH 7.80) crystal structures for the same bacterial *T. thermophilus* organism, regardless of the pH values used during crystallization, contain substantially similar basepair conformations to
each other, except in a single region with remarkable differences in domain III; nonetheless, the originally resolved native structure was used for the analysis of basepair conformations in the 16S rRNA. Furthermore, while the two 23S rRNAs in the *H. marismortui* 50S (PDB 1FFK and 1JJ2, pH 5.80) and the *D. radiodurans* 50S crystal structures (PDB 1LNR, pH 7.80) are in overall agreement, the latter was not utilized in the detailed analysis of the 23S rRNA due to its lower resolution. Additionally, the protonated and bifurcated hydrogen bonds in RNA were not considered for simplifying our classification of basepair conformations.
Chapter 4: The Sticky Motif: AGUA/GAA, GUA/GAA, and GGA/GAA Motifs in RNA Structure

4.1 BACKGROUND

Covariation analysis, a subset of comparative sequence analysis, has been the best tool to predict RNA secondary structures. The covariation-based RNA secondary structure models contain helices interrupted by hairpin, internal, and multistem loops that are left unpaired and frequently share certain sequence biases or motifs. The identification of such sequence motifs on the RNA secondary structure models and their corresponding structure motifs in the experimentally determined three-dimensional structures are essential to better understand RNA structure. The sequence-based structure motifs usually involve non-canonical basepairs and are not predictable by covariation analysis alone due to the lack of covariation. Such examples are GNRA tetraloops, tandem GA motifs, and AA.AG@helix.ends motifs.

The E loop motif is another common, well-established RNA sequence motif. The basepairs in the E loop of the PSTV and HeLa 5S rRNAs were first implicated by the UV-induced crosslinking experiment, and investigated by the two-dimensional NMR experiment for the asymmetric E loop of the
eukaryotic *Xenopus laevis* 5S rRNA.\textsuperscript{31} The E loop of the 5S rRNA was suggested as the recognition site for transcription factor TFIIIA and ribosomal protein L5.\textsuperscript{108-109} The ten E loop sequence motifs in 16S and 23S rRNAs were subsequently predicted.\textsuperscript{35} The high-resolution crystal structure for the *E. coli* 23S rRNA sarcin/ricin loop (SRL) domain (PDB 483D\textsuperscript{103}), which has been known as the binding site of the elongation factors EF-G and EF-Tu,\textsuperscript{110} further validated the formation of the E loop motif.

An analysis of the secondary structure models and the high-resolution crystal structures for the rRNAs in the *T. thermophilus* 30S (PDB 1FJF\textsuperscript{63}), *H. marismortui* 50S (PDB 1FFK\textsuperscript{64} and 1JJ2\textsuperscript{45}) ribosomal subunits not only revealed the structural details of the AGUA/GAA motif (previously known as the E loop motif\textsuperscript{30-32,35}), but also identified two additional families of “E loop-related” RNA motifs, GUA/GAA and GGA/GAA motifs. The GUA/GAA motif (referred to as E-like loop motif) is derived from the E loop of the 5S rRNA, while the GGA/GAA motif (referred to as EL-like loop motif) has the similar structure to the GUA/GAA motif formed in the crystal structure of the *E. coli* 5S rRNA (PDB 1DFU\textsuperscript{111}). Together, these three E-loop related sequence and structure motifs form a superfamily of RNA motifs, the sticky motif. Here, we present the frequent occurrence in all phylogenetic groups and structural characteristics of the sticky motif and address its potential roles in structure and function of RNA.
4.2 RESULTS

The sticky motif is the superfamily of AGUA/GAA (E), GUA/GAA (EL), and GGA/GAA (ELL) motifs and commonly occurring in internal and multistem loops of the comparative RNA secondary structure models. The AGUA/GAA (E), GUA/GAA (EL), and GGA/GAA (ELL) motifs possess the consensus sequences with a strict directionality, 5’-AGUA/GAA-3’, 5’-GUA/GAA-3’, and 5’-GGA/GAA-3’, respectively (Table 4.1). For consistent and unambiguous description, the 5’ and 3’ to the motif are defined by the 5’ and 3’ to the AGUA, GUA, or GGA sequence. Throughout the text, the nucleotide position numberings for the rRNAs in the crystal structures of the *T. thermophilus* 30S (PDB 1FJF \(^{63}\)) and *H. marismortui* 50S (PDB 1FFK \(^{64}\) and 1JJ2 \(^{45}\)) ribosomal subunits (called crystal structure numberings or Xtal) are used together with the *E. coli* numberings (Ec) in parenthesis and the basepair conformations are based on the recently proposed Lee-Gutell classification system.\(^{112}\) Additional information is available at (http://www.rna.icmb.utexas.edu/ANALYSIS/EELL/).

4.2.1 Analysis of the rRNAs in the 30S and 50S crystal structures

An analysis of the secondary structure models (searching for sequence motifs) and the high-resolution crystal structures (searching for structure motifs) of the rRNAs in the *T. thermophilus* 30S (PDB 1FJF \(^{63}\)) and *H. marismortui* 50S.
<table>
<thead>
<tr>
<th>Sticky Motif</th>
<th>Consensus Sequence</th>
<th>Usually Present in</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGUA/GAA (E)</td>
<td>5'-AGUA-3'</td>
<td>Asymmetric internal loop</td>
</tr>
<tr>
<td></td>
<td>3'-A  AG-5'</td>
<td>Multistem loop</td>
</tr>
<tr>
<td>GUA/GAA (EL)</td>
<td>5'-GUA-3'</td>
<td>Symmetric internal loop</td>
</tr>
<tr>
<td></td>
<td>3'-AAG-5'</td>
<td></td>
</tr>
<tr>
<td>GGA/GAA (ELL)</td>
<td>5'-GGA-3'</td>
<td>Symmetric internal loop</td>
</tr>
<tr>
<td></td>
<td>3'-AAG-5'</td>
<td>Asymmetric internal loop</td>
</tr>
</tbody>
</table>

*Table 4.1* Consensus sequence context and location of the sticky motif
(PDB 1FFK\textsuperscript{64} and 1JJ2\textsuperscript{45}) ribosomal subunits has identified 33 sticky motifs: 15 AGUA/GAA (E), ten GUA/GAA (EL), and eight GGA/GAA (ELL) motifs; ten in 16S, 22 in 23S, and one in 5S rRNAs; 25 in internal and eight in multistem loops (Table 4.2). Each motif is represented using the single, lowest nucleotide position number within the motif. Of the 33 sequence motifs, 29 are articulated in the crystal structures, and 23 occur in the highly conserved regions in all three phylogenetic domains.\textsuperscript{61} The single nucleotide frequencies of the conserved motifs are shown in Table 4.3. The phylogenetic distribution and frequencies for the 33 AGUA/GAA, GUA/GAA, and GGA/GAA motifs identified in rRNAs in the crystal structures, along with those of the 34 additional motifs identified in other rRNA comparative structure models whose high-resolution crystal structures are not yet available, are shown in Table 4.4, using the percentages for the frequencies of each motif in each phylogenetic alignment of Archaea (A), Bacteria (B), Eukaryotes (E), Chloroplasts (C), and Mitochondria (M): The long dashes (—) represent the absence of the motifs in the current set of alignments; the motifs marked with the capital “X” are the motifs whose frequencies are not currently available. Their sequence variations are available in Appendix I.

The secondary structure fragments containing the 33 motifs (16 AGUA/GAA (E), 9 GUA/GAA (EL), and 8 GGA/GAA (ELL) motifs) identified in the \textit{T. thermophilus} 16S and the \textit{H. marismortui} 23S and 5S rRNAs are
<table>
<thead>
<tr>
<th>Ec.</th>
<th>Xtal</th>
<th>Seq. (Xtal)</th>
<th>Sequence Motif</th>
<th>Structure Motif</th>
<th>Interactions with RNA</th>
<th>Protein</th>
<th>Metal</th>
<th>Loop</th>
<th>Misc.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Conserved</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16S</td>
<td>581</td>
<td>564</td>
<td>GUA/GAG</td>
<td>EL</td>
<td>EL</td>
<td>Yes</td>
<td>S15+S17</td>
<td>—</td>
<td>sl</td>
</tr>
<tr>
<td></td>
<td>685</td>
<td>668</td>
<td>GUA/GAU</td>
<td>EL</td>
<td>EL</td>
<td>Yes</td>
<td>S11</td>
<td>M&lt;sup&gt;r&lt;/sup&gt;</td>
<td>asl</td>
</tr>
<tr>
<td></td>
<td>765</td>
<td>748</td>
<td>GAA/CUA</td>
<td>EL</td>
<td>—</td>
<td>Yes</td>
<td>—</td>
<td>M&lt;sup&gt;r&lt;/sup&gt;</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td>780</td>
<td>763</td>
<td>AAA/GUA</td>
<td>EL</td>
<td>EL</td>
<td>Yes</td>
<td>S11</td>
<td>M&lt;sup&gt;r&lt;/sup&gt;</td>
<td>sl</td>
</tr>
<tr>
<td></td>
<td>889</td>
<td>866</td>
<td>AGUA/GAA</td>
<td>E</td>
<td>E</td>
<td>Yes</td>
<td>—</td>
<td>M&lt;sup&gt;r&lt;/sup&gt;</td>
<td>asl</td>
</tr>
<tr>
<td></td>
<td>1304</td>
<td>1285</td>
<td>GGA/GAA</td>
<td>ELL</td>
<td>ELL</td>
<td>Yes</td>
<td>S13+THX</td>
<td>M&lt;sup&gt;r&lt;/sup&gt;</td>
<td>sl</td>
</tr>
<tr>
<td></td>
<td>1346</td>
<td>1327</td>
<td>AGUA/GAA</td>
<td>E</td>
<td>E</td>
<td>Yes</td>
<td>S7+S9</td>
<td>—</td>
<td>M</td>
</tr>
<tr>
<td>23S</td>
<td>26</td>
<td>23</td>
<td>GGA/GAA</td>
<td>ELL</td>
<td>ELL</td>
<td>Yes</td>
<td>—</td>
<td>Mg&lt;sup&gt;2+&lt;/sup&gt;</td>
<td>sl</td>
</tr>
<tr>
<td></td>
<td>189</td>
<td>159</td>
<td>GAA/GUA</td>
<td>E</td>
<td>E</td>
<td>Yes</td>
<td>L15e</td>
<td>Mg&lt;sup&gt;2+&lt;/sup&gt;</td>
<td>asl</td>
</tr>
<tr>
<td></td>
<td>241</td>
<td>212</td>
<td>AGUA/GAA</td>
<td>E</td>
<td>E</td>
<td>Yes</td>
<td>L15+L15e</td>
<td>Mg&lt;sup&gt;2+&lt;/sup&gt;</td>
<td>asl</td>
</tr>
<tr>
<td></td>
<td>298</td>
<td>304</td>
<td>GAA/GUA</td>
<td>EL</td>
<td>—</td>
<td>Yes</td>
<td>L4+L24</td>
<td>—</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td>371</td>
<td>380</td>
<td>AGUA/GAA</td>
<td>E</td>
<td>E</td>
<td>Yes</td>
<td>L15e</td>
<td>—</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td>457</td>
<td>463</td>
<td>AGUA/GAA</td>
<td>E</td>
<td>E</td>
<td>Yes</td>
<td>L4+L39e</td>
<td>Na&lt;sup&gt;r&lt;/sup&gt;</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td>704</td>
<td>795</td>
<td>GAA/GGA</td>
<td>ELL</td>
<td>ELL</td>
<td>Yes</td>
<td>L19e+L37ae</td>
<td>Mg&lt;sup&gt;2+&lt;/sup&gt;</td>
<td>sl</td>
</tr>
<tr>
<td></td>
<td>818</td>
<td>911</td>
<td>GAA/GAU</td>
<td>E</td>
<td>E&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Yes</td>
<td>L32e</td>
<td>Mg&lt;sup&gt;2+&lt;/sup&gt;, Na&lt;sup&gt;r&lt;/sup&gt;</td>
<td>sl</td>
</tr>
<tr>
<td></td>
<td>858</td>
<td>952</td>
<td>GGUA/AAA</td>
<td>E</td>
<td>E&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Yes</td>
<td>L21e</td>
<td>—</td>
<td>asl</td>
</tr>
<tr>
<td></td>
<td>1212</td>
<td>1316</td>
<td>GAA/GAU</td>
<td>ELL</td>
<td>ELL</td>
<td>Yes</td>
<td>L32e</td>
<td>—</td>
<td>asl</td>
</tr>
<tr>
<td></td>
<td>1265</td>
<td>1369</td>
<td>AGUA/GAA</td>
<td>E</td>
<td>E</td>
<td>Yes</td>
<td>L22</td>
<td>Mg&lt;sup&gt;2+&lt;/sup&gt;</td>
<td>asl</td>
</tr>
<tr>
<td></td>
<td>1687</td>
<td>1765</td>
<td>GUA/GAA</td>
<td>EL</td>
<td>—</td>
<td>Yes</td>
<td>—</td>
<td>—</td>
<td>asl</td>
</tr>
<tr>
<td></td>
<td>1930</td>
<td>1971</td>
<td>GAA/GUA</td>
<td>EL</td>
<td>E&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Yes</td>
<td>—</td>
<td>Na&lt;sup&gt;r&lt;/sup&gt;</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td>2467</td>
<td>2502</td>
<td>GAA/GAC</td>
<td>ELL</td>
<td>ELL</td>
<td>—</td>
<td>L10e</td>
<td>—</td>
<td>sl</td>
</tr>
<tr>
<td></td>
<td>2654</td>
<td>2691</td>
<td>AGUA/GAA</td>
<td>E</td>
<td>E</td>
<td>Yes</td>
<td>L6</td>
<td>—</td>
<td>asl</td>
</tr>
<tr>
<td>5S</td>
<td>76</td>
<td>77</td>
<td>AGUA/GAA</td>
<td>E</td>
<td>E&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Yes</td>
<td>—</td>
<td>—</td>
<td>asl</td>
</tr>
</tbody>
</table>

| **Variable:** | | | | | | | | | |
| 16S | 413 | 408 | GAA/GUA | EL | — | Yes | S4 | — | asl | |
| | 447 | 441 | GAC/GGUA | E | E | — | S16 | — | asl | |
| | 1238 | 1219 | AA/GGU | E | E | Yes | S7 | — | M | |
| 23S | 82 | 78 | GGA/GAA | ELL | ELL | Yes | L24 | — | asl | K<sup>33</sup> |
| | 286 | 292 | GAC/GUA | E | E | — | — | — | asl | |
| | 544.2 | 568 | GAC/GUA | E | E | — | L30 | — | asl | |
| | 635i | 706 | GAA/GAG | EL | EL | Yes | L18e | — | sl | |
| | 844 | 937 | CGA/GAC | ELL | ELL | Yes | — | — | asl | K<sup>35</sup> |
| | 1438 | 1543 | GU/CAA | EL | EL | Yes | — | — | sl | |
| | 1484 | 1588 | AGUA/GAA | ELL | ELL | Yes | — | — | asl | K<sup>35</sup> |

<sup>a</sup> The Escherichia coli-equivalent position numbers for the first nucleotides in the motifs and their corresponding position numbers in the rRNAs of the two ribosomal crystal structures, respectively.  
<sup>b</sup> Nucleotide compositions in the current crystal structures being used, with the AGUA, GUA, and GGA sequences highlighted.  
<sup>c</sup> Sequence and structure motifs, respectively; E, AGUA/GAA motif; EL, GUA/GAA motif; ELL, GGA/GAA motif. The dagger marks (†) are used for the “transformed” motifs or the motifs which are formed differently from the predicted ones.  
<sup>d</sup> Each M<sup>r</sup> represents a metal ion bound to the motifs but its identity was not determined in the crystal structure.  
<sup>e</sup> Loop type: sl, symmetric internal loop; asl, asymmetric internal loop; M, multistem loop.  
<sup>f</sup> Miscellaneous information: K, K-turn motif; SIF, present at the subunit interfaces; EFs, elongation factors.

Table 4.2 Sticky motifs and their molecular interactions
<table>
<thead>
<tr>
<th>Ec</th>
<th>Xtal</th>
<th>Frequency*</th>
</tr>
</thead>
<tbody>
<tr>
<td>16S</td>
<td></td>
<td></td>
</tr>
<tr>
<td>581</td>
<td>564</td>
<td>G (99.1)</td>
</tr>
<tr>
<td>592</td>
<td>565</td>
<td>Uc (77.6/22.1)</td>
</tr>
<tr>
<td>583</td>
<td>566</td>
<td>A (99.6)</td>
</tr>
<tr>
<td>756</td>
<td>741</td>
<td>Gc (66.5/30.1)</td>
</tr>
<tr>
<td>752</td>
<td>747</td>
<td>A (99.4)</td>
</tr>
<tr>
<td>760</td>
<td>743</td>
<td>G/a (69.1/18.6/12.1)</td>
</tr>
<tr>
<td>685</td>
<td>668</td>
<td>Gc/a (83.3/13.1/3.1)</td>
</tr>
<tr>
<td>686</td>
<td>669</td>
<td>Uc/g (83.9/11.0/3.8)</td>
</tr>
<tr>
<td>687</td>
<td>670</td>
<td>A (99.8)</td>
</tr>
<tr>
<td>703</td>
<td>686</td>
<td>G (95.4)</td>
</tr>
<tr>
<td>704</td>
<td>687</td>
<td>A (99.9)</td>
</tr>
<tr>
<td>705</td>
<td>688</td>
<td>Uc/g (73.4/23.7)</td>
</tr>
<tr>
<td>765</td>
<td>748</td>
<td>G (99.1)</td>
</tr>
<tr>
<td>766</td>
<td>749</td>
<td>A (99.2)</td>
</tr>
<tr>
<td>767</td>
<td>750</td>
<td>A (99.2)</td>
</tr>
<tr>
<td>812</td>
<td>795</td>
<td>Gc/a (74.9/14.6/10.2)</td>
</tr>
<tr>
<td>813</td>
<td>796</td>
<td>U (99.2)</td>
</tr>
<tr>
<td>814</td>
<td>797</td>
<td>A (99.8)</td>
</tr>
<tr>
<td>780</td>
<td>763</td>
<td>A/G (57.6/42.2)</td>
</tr>
<tr>
<td>781</td>
<td>764</td>
<td>A (99.6)</td>
</tr>
<tr>
<td>782</td>
<td>765</td>
<td>A (99.8)</td>
</tr>
<tr>
<td>800</td>
<td>783</td>
<td>G/a (95.5/4.1)</td>
</tr>
<tr>
<td>801</td>
<td>784</td>
<td>U (99.9)</td>
</tr>
<tr>
<td>802</td>
<td>785</td>
<td>A (99.8)</td>
</tr>
<tr>
<td>889</td>
<td>866</td>
<td>A (99.5)</td>
</tr>
<tr>
<td>890</td>
<td>867</td>
<td>G (97.2)</td>
</tr>
<tr>
<td>891</td>
<td>868</td>
<td>U (99.5)</td>
</tr>
<tr>
<td>892</td>
<td>869</td>
<td>G/a (95.5/4.1)</td>
</tr>
<tr>
<td>906</td>
<td>883</td>
<td>G/a (51.0/48.6)</td>
</tr>
<tr>
<td>907</td>
<td>884</td>
<td>A (99.7)</td>
</tr>
<tr>
<td>908</td>
<td>885</td>
<td>A (99.8)</td>
</tr>
<tr>
<td>1304</td>
<td>1285</td>
<td>G/a (96.0/3.6)</td>
</tr>
<tr>
<td>1305</td>
<td>1286</td>
<td>G (99.1)</td>
</tr>
<tr>
<td>1306</td>
<td>1287</td>
<td>A (99.5)</td>
</tr>
<tr>
<td>1331</td>
<td>1312</td>
<td>Gc/a (92.4/6.4)</td>
</tr>
<tr>
<td>1332</td>
<td>1313</td>
<td>A (99.8)</td>
</tr>
<tr>
<td>1333</td>
<td>1314</td>
<td>A (99.8)</td>
</tr>
<tr>
<td>1346</td>
<td>1327</td>
<td>A (94.1/5.2)</td>
</tr>
<tr>
<td>1347</td>
<td>1328</td>
<td>G (99.9)</td>
</tr>
<tr>
<td>1348</td>
<td>1329</td>
<td>U (99.9)</td>
</tr>
<tr>
<td>1349</td>
<td>1330</td>
<td>A (99.9)</td>
</tr>
<tr>
<td>1373</td>
<td>1355</td>
<td>G (99.4)</td>
</tr>
<tr>
<td>1374</td>
<td>1356</td>
<td>A (99.3)</td>
</tr>
<tr>
<td>1375</td>
<td>1357</td>
<td>A (98.1/10.6)</td>
</tr>
<tr>
<td>23S</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>23</td>
<td>G (99.8)</td>
</tr>
<tr>
<td>27</td>
<td>24</td>
<td>G (95.3)</td>
</tr>
</tbody>
</table>

*The nucleotides that are more than 50% conserved in the nuclear-encoded archaeal, bacterial, and eukaryotic 16S and 23S rRNA sequences (3P) are shown in upper case, while the nucleotides with 3-50% conservation are shown in lowercases. The nucleotide frequencies in the sticky motif occurring in the 5S rRNAs are not shown since the positions in Bacteria, Archaea, and Eukaryotes are not homologous with each other.

Table 4.3 Single nucleotide frequencies for conserved sticky motifs
**| Ec *a* | Motif *b* | A | B | E | C | M |
---|---|---|---|---|---|---|---|
Conserved: | 16S 581 | EL | 99.1 | 75.4 | 96.6 | 88.0 | 20.4 |
| 685 | E,ELL | 89.2 | 99.2 | 90.9 | 92.9 | 8.6 |
| 765 | EL | 99.5 | 97.3 | 93.0 | 94.1 | 31.9 |
| 780 | EL | 99.1 | 98.9 | 98.0 | 94.1 | 91.3 |
| 889 | E | 94.6 | 98.8 | 99.1 | 97.0 | 97.9 |
| 1304 | E,ELL | 94.7 | 99.4 | 95.7 | 95.4 | 79.0 |
| 1346 | E | 100 | 99.4 | 96.6 | 100 | 95.6 |
| 23S 26 | E,L | 94.4 | 94.4 | 97.4 | 95.7 | 13.6 |
| 189 | E | 100 | 98.9 | 77.5 | 100 | 15.5 |
| 241 | E | 100 | 98.9 | 97.5 | 96.0 | 88.4 |
| 298 | EL | 100 | 95.3 | 46.2 | 95.0 | 5.1 |
| 371 | E | 100 | 99.6 | 97.3 | 98.0 | 11.3 |
| 457 | E | 100 | 96.5 | 90.4 | 98.0 | — |
| 704 | E,ELL | 100 | 100 | 95.1 | 99.0 | 6.5 |
| 818 | E | 100 | 97.0 | 99.9 | 100 | 28.2 |
| 858 | E | 100 | 100 | 96.2 | 96.0 | 14.5 |
| 1212 | E,ELL | 100 | 98.9 | 92.6 | 96.0 | 6.5 |
| 1265 | E | 100 | 100 | 99.3 | 100 | 92.0 |
| 1687 | EL | 78.1 | 74.1 | 62.6 | 91.1 | 7.6 |
| 1930 | EL | 100 | 97.9 | 99.4 | 99.0 | 96.8 |
| 2467 | E,ELL | 100 | 99.6 | 99.3 | 100 | 66.2 |
| 2654 | E | 100 | 100 | 100 | 100 | 98.8 |
| 5S 76 | E,ELL | 86.7 | 94.4 | 99.7 | n/a | n/a |

Variable (Xtal): | 16S 413 | EL | — | 76.3 | 0.2 | 56.4 | 3.8 |
| 447 | E | — | 92.6 | 0.1 | 87.2 | 1.0 |
| 661 | E,ELL | 97.8 | 3.4 | 0.1 | 1.2 | — |
| 1238 | E | 98.6 | 98.6 | 0.4 | 97.7 | 3.2 |
| 1260 | E,ELL | 1.9 | 32.7 | 0.2 | 35.3 | 1.8 |
| 1417 | E | — | 0.2 | — | 0.5 |
| 23S 82 | E,ELL | 100 | 93.4 | 0.8 | 95.4 | 1.3 |
| 286 | E,ELL | 7.7 | 3.9 | 2.0 | — | — |
| 635 | E | 91.9 | — | — | 0.1 | — |

| Ec *a* | Motif *b* | A | B | E | C | M |
---|---|---|---|---|---|---|---|
| 1438 | E,ELL | X | X |
| 1484 | E,ELL | 100 | — | — | — |
| 1710 | E,ELL | 0.4 | — | — | — |

Variable (Others): | 16S 68i | E,ELL | X |
| 69i | E,ELL | X |
| 70i | E,ELL | X |
| 73i | E,ELL | X |
| 74i | E,ELL | X |
| 76i | E,ELL | X |
| 82i | E,ELL | X |
| 141i | E,ELL | — | 85.1 | — | — |

| Ec *a* | Motif *b* | A | B | E | C | M |
---|---|---|---|---|---|---|---|
| 544i | E,ELL | X |
| 641i | E,ELL | X |
| 1204i | E,ELL | X |
| 1415i | E,ELL | X |
| 1415j | E,ELL | X |
| 1418 | E,ELL | X |
| 1475i | E,ELL | 39.4 | 2.8 | — | — |
| 1526i | E,ELL | 57.0 | 9.0 | — | — |
| 1529i | E,ELL | X |
| 1530i | E,ELL | X |
| 1531i | E,ELL | X |
| 1713 | E,ELL | X |
| 1721 | E,ELL | 9.5 |
| 1727i | E,ELL | X |
| 1727j | E,ELL | X |
| 1727k | E,ELL | X |
| 1857 | E,ELL | 93.1 | — | 4.2 |
| 1865 | E,ELL | — | — | 3.9 |
| 2133 | E,ELL | 53.7 | 93.5 | 57.8 | 0.7 |

| 2209i | E | X |

The *Escherichia coli*-equivalent position numbers for the first nucleotides in the sequence motifs. The motifs present in the used crystal structures are shown in bold. The asterisk (*) marks are used for the motifs which exchange with either tandem GA motif or canonical basepairs in the rest of the sequences in the same or different phylogenetic groups.

* Sequence motifs: E, AGUA/GAA motif; EL, GUA/GAA motif; ELL, GGA/GAA motif.

‡ While the bacterial 5S rRNA has a GUA/GAA motif, the archaeal and eukaryotic 5S rRNAs have a AGUA/GAA motif.

Table 4.4 Frequencies and phylogenetic distribution of rRNA sticky motifs

83
represented in Figures 4.1, 4.2, and 4.3, with nucleotide positions in black for crystal structure numberings (T. thermophilus numbering for 16S rRNA, H. marismortui numbering for 23S and 5S rRNAs) and red for E. coli-equivalent numbering (16S, 23S, and 5S rRNAs). The nucleotides and their interactions are shown in different colors: red, nucleotides within sticky motifs; cyan, the nearest canonical basepairs in the flanking helices; green, other nucleotides between the two flanking helices and metal binding sites; black, all other nucleotides.

All of the 29 structure motifs but one at position 1219 (Ec: 1238) in 16S rRNA have three consecutive basepairs (the leading, middle, and trailing basepairs in the 5' to 3' direction, respectively) with the non-canonical basepair conformations. These three consecutive basepairs in the sticky motifs will not only stabilize the internal and multistem loops, but also mediate the helical stacking of the two flanking helices in the internal loops or induce the coaxial stacking of the two neighboring helices in the multistem loops.

4.2.2 General features of AGUA/GAA, GUAA/GAA, and GGA/GAA motifs

4.2.2.1 AGUA/GAA (E) motif

Of the 16 AGUA/GAA motifs articulated in rRNAs in the T. thermophilus 30S and H. marismortui 50S crystal structures, ten occur in the internal loops and six in the multistem loops; of the ten motifs in the internal loops, all but one at position 952 (Ec: 858) in 23S rRNA occur in the asymmetric internal loops.
Fig. 4.1 Structure gallery of AGUA/GAA motifs in the crystal structures. The coaxial stacking of a sticky motif onto the flanking helix or helices, along with the induced coaxial stacking of the neighboring or distal helices, are shaded in yellow.
Fig. 4.2 Structure gallery of GUA/GAA motifs in the crystal structures. The coaxial stacking of a sticky motif onto the flanking helix or helices, along with the induced coaxial stacking of the neighboring or distal helices, are shaded in light blue.
Fig. 4.3 Structure gallery of GGA/GAA motifs in the crystal structures. The coaxial stacking of a sticky motif onto the flanking helix or helices, along with the induced coaxial stacking of the neighboring or distal helices, are shaded in light magenta.
Additionally, all of the 16 motifs, but one at position 1219 (Ec: 1238) in 16S rRNA containing the partial AGU/AA sequence, have three consecutive basepairs with the non-canonical basepair conformations: the leading A:A rS, the middle U:A rH as part of the G(U:A) base triple, and the trailing A:G S.\textsuperscript{112}

The leading A:A basepair, which was initially proposed based on the NMR experiment,\textsuperscript{31} exchanges with A:C at positions 292 (Ec: 286) and 568 (Ec: 544.2) in 23S rRNA, with G:A at positions 911 (Ec: 818) in 23S and 77 (Ec: 76) in 5S rRNAs, or even with G:C at position 441 (Ec: 447) in 16S rRNA; all of those exchanging basepairs have the isosteric reversed sheared conformations such as C:A rS, G:A rS, and C:G rS. Interestingly, however, the sheared A:C basepair (C:A rS) has no hydrogen bonds despite its spatial arrangement isosteric to A:A rS. The G(U:A) base triple was first proposed based on the NMR experiment,\textsuperscript{31} and validated by the crystal structure for the sarcin/ricin loop (SRL) domain in the \textit{E. coli} 23S rRNA (PDB 483D\textsuperscript{103}). To form the G(U:A) base triple, the G of the AGUA sequence is protruded into the major groove and its –NH\textsubscript{2} is hydrogen bonded to C=O of the U in the U:A rH basepair, while creating the sharp S-turn in the backbone. As well, the protruded G causes the reversal of the leading A in the AGUA sequence, leading to the A:A rS. The trailing sheared A:G basepair (A:G S) usually has A 5’ and G 3’ to helix and exchanges with the
sheared A:A basepair (A:A S) at position 952 (Ec: 858) in 23S rRNA; it frequently overlaps with the AA.AG@helix.ends motif.38

Eight of the nine AGUA/GAA motifs occurring in the asymmetric internal loops and two occurring in the multistem loops contain one or two additional non-canonical basepairs, Y:Y or R:R (where Y and R is a pyrimidine and purine, respectively), directly 5' to the seven-nucleotide AGUA/GAA motif (Fig. 4.1). These additional non-canonical basepairs form either the sheared (S) or slipped wobble (sWb) conformation depending on the actual sequence of the AGUA/GAA motif: the sheared (S) conformation with the 5'-AGUA/GAA-3' sequence and the slipped wobble (sWb) conformation with the 5'-AGUA/GAC-3' sequence.112 For example, the former conformation is observed at positions 866 (Ec: 889) in 16S, 159 (Ec: 189), 212 (Ec: 241), 1369 (Ec: 1265), and 2691 (Ec: 2654) in 23S, and 77 (Ec: 76) in 5S rRNAs; the latter at positions 292 (Ec: 286) and 568 (Ec: 544i.2) in 23S rRNA.

Interestingly, of the ten AGUA/GAA (E) motifs previously predicted in 16S and 23S rRNAs,35 the seven at positions 1327 (Ec: 1346) in 16S and 159 (Ec: 189), 212 (Ec: 241), 380 (Ec: 371), 463 (Ec: 457), 1369 (Ec: 1265), and 2691 (Ec: 2654) in 23S rRNAs form as predicted, and the two at positions 866 (Ec: 889) in 16S and 911 (Ec: 818) in 23S rRNAs are articulated in a strikingly different way.
The remaining one, 766 (Ec: 675) in 23S rRNA, was not correctly predicted; its sequence variations are not possible to form the motif.61

4.2.2.2 GUA/GAA (EL) motif

The GUA/GAA motif has the similar nucleotide sequence to the AGUA/GAA motif and is also present in the E loop of 5S rRNAs similarly to the AGUA/GAA motif; the E loop of the archaeal and eukaryotic 5S rRNAs has the asymmetric internal loop with the AGUA/GAA motif, while that of the bacterial 5S rRNA has the symmetric internal loop with the GUA/GAA motif.61 As well, the E loop in the *E. coli* 5S rRNA (PDB 1DFU111) revealed the base stack in the minor groove that is similar to that in the AGUA/GAA motif. Of the five GUA/GAA motifs formed in the rRNAs in the *T. thermophilus* 30S and *H. marismortui* 50S crystal structures, four occur in the 3x3 symmetric internal loops and one occurs in an asymmetric internal loop (Fig. 4.2). Similarly to the AGUA/GAA motif, all of the five GUA/GAA motifs have three consecutive basepairs with the non-canonical basepair conformations: the leading G:A S), the middle U:A rH, and the trailing A:G S.112

Both the leading sheared G:A (G:A S) and the trailing sheared A:G (A:G S) basepairs have their A’s 5’ and G’s 3’ to helix and are often overlapped with the AA.AG@helix.ends motif.38 The leading G:A S basepair exchanges with G:G at positions 564 (Ec: 581) in the 16S and 706 (Ec: 635i) in the 23S rRNA, with C:C
at position 1543 \((Ec: 1438)\) in the 23S rRNA, or with G:U at position 668 \((Ec: 685)\) in the 16S rRNA with the respective isosteric sheared conformation. The trailing A:G S basepair exchanges with A:A S at position 763 \((Ec: 780)\) in the 16S rRNA. The middle U:A rH basepair is distinguishable from the G(U:A) base triple for the AGUA/GAA motif.

Four GUA/GAA motifs at positions 408 \((Ec: 413)\) and 748 \((Ec: 765)\) in the 16S and 304 \((Ec: 298)\) and 1765 \((Ec: 1687)\) in the 23S rRNAs are not formed in the crystal structures. Interestingly, two of them occur in the asymmetric internal loops and the remaining two occur in the multistem loops, suggesting that GUA/GAA motifs are destabilized in the asymmetric internal or multistem loops.

4.2.2.3 GGA/GAA (ELL) motif

The GGA/GAA motif is another six-nucleotide and three-basepair structure motif and, albeit the different nucleotide sequences, reveals the base stack similar to that for the AGUA/GAA or GUA/GAA motif. All of the eight GGA/GAA motifs identified in the \textit{T. thermophilus} 16S and \textit{H. marismortui} 23S rRNAs occur in the internal loops and are formed in the rRNAs in the \textit{T. thermophilus} 30S and \textit{H. marismortui} 50S crystal structures (Fig. 4.3): three in the symmetric and five in the asymmetric internal loops. The three consecutive basepairs with the non-canonical basepair conformations are the leading G:A S, the middle G:A S, and the trailing A:G S, respectively.\textsuperscript{112}
The leading G:A S basepair always has its A 5’ and G 3’ to a regular helix immediately 5’ to the motif and is sometimes overlapped with the AA.AG@helix.ends motif.38 Interestingly, the leading G:A S basepair exchanges with the C:C basepair with the C:C Wb conformation at positions 937 (Ec: 844) and 2502 (Ec: 2467) in 23S rRNA. Thus, when the leading G:A S basepair exchanges with standard Watson-Crick basepair, the GGA/GAA motif shortens to the tandem GA motif.61 The characteristic middle G:A S basepair corresponds to the G(U:A) base triple for the AGUA/GAA motif and the simple U:A basepair for the GUA/GAA motif. The trailing A:G S basepair can also be an AA.AG@helix.ends basepair, when it occurs in the 3x3 symmetric internal loop, e.g., at position 1285 (Ec: 1304) in the 16S rRNA.

4.2.3 Interactions within AGUA/GAA, GUA/GAA, and GGA/GAA motifs

AGUA/GAA, GUA/GAA, and GGA/GAA motifs are highly stabilized by the extensive network of hydrogen bonding and base stacking interactions within themselves (Fig. 4.4). A canonical AGUA/GAA motif has 12 hydrogen bonds and three base stacks (AAA in the minor groove, AU near the helical axis, and GG in the major groove). A canonical GUA/GAA motif has nine hydrogen bonds and one base stack (A(U:A)A in the minor groove); the two G’s in the major groove are two basepairs away and are not directly stacked, narrowing the major groove. A canonical GGA/GAA motif has 12 hydrogen bonds and two base
Fig. 4.4 Representative AGUA/GAA, GUA/GAA, and GGA/GAA motifs and their stereoviews down the helical axis with the 5' to the motif on the top.
stacks (AAA in the minor groove and GGG in the major groove). In this vein, the AGUA/GAA motif would be the most stable, while the GUA/GAA motif would be the least stable and vulnerable to any structural perturbation. When the trailing A:G basepair exchanges with A:A, as observed at position 763 (Ec: 780) in the 16S rRNA, the number of hydrogen bonds are reduced by one.

Interestingly, two GUA/GAA motifs, one containing the GUA/GAG sequence in the E loop of the *E. coli* 5S rRNA (PDB 1DFU) and the other containing the GUA/GAU sequence at position 668 (Ec: 685) in 16S rRNA, have the unexpected GAA and UAA base stacks respectively, which resemble the AAA stack for the AGUA/GAA and GGA/GAA motifs. Thus, AGUA/GAA, GUA/GAA, and GGA/GAA motifs are structurally similar with each other.

### 4.2.4 Interactions of sticky motifs with RNA and proteins

All of the 33 AGUA/GAA, GUA/GAA, and GGA/GAA motifs but one at position 286 (Ec: 292) in the 23S rRNA are contacted by either other regions of rRNAs or proteins (Table 4.2 and Figures 4.1, 4.2, and 4.3). Their minor groove containing the AAA or A(U:A)A stack is usually contacted by other section(s) of the rRNA; although nucleophilic N1 and N3 atoms of the A’s are frequently involved in the A-minor interactions, the 2'-OH groups and sugar phosphate oxygen atoms are also frequently employed. On the contrary, their major groove is preferentially interacted by protein(s), probably because the G and U bases in
the major groove have partial negative charges and attract the positively charged residues of proteins. Thus, there appears to be no difference in binding affinity of the conserved and variable motifs with RNA and proteins, suggesting that AGUA/GAA, GUA/GAA, and GGA/GAA motifs are “sticky” and work as “molecular magnets” by making contacts with other sections of RNA to help RNA folding or serve as the recognition elements for proteins. Furthermore, nine conserved sticky motifs are at or near the intersubunit bridges (Table 4.2), implying their important participation in the ribosomal association. As an example, the region containing the AGUA/GAA motif at 952 (Ec: 858) in the 23S rRNA makes ten hydrogen bonds with the region containing the AGUA/GAA motif at 77 (Ec: 76) in the 5S rRNA.

4.2.4.1 AGUA/GAA motifs in asymmetric internal loops

Seven AGUA/GAA motifs occur in the simple, asymmetric internal loops: three conserved ones at positions 159 (Ec: 189), 212 (Ec: 241), and 2691 (Ec: 2654) in the 23S rRNA, three variable ones at positions 441 (Ec: 447) in the 16S and (Ec: 292286) and 568 (Ec: 544i.2) in the 23S rRNA, and one at position 77 (Ec: 76) in the 5S rRNA (Table 4.2).

Of the three conserved motifs, the first one at position 159 (Ec: 189) in the 23S rRNA involves two additional non-canonical basepairs directly 5' to the motif and its minor groove is completely blocked by the helix, 769-771/890-893 (Ec:
A161 (Ec: A191) and A160 (Ec: A190) interact with C769:G892 (Ec: C678:G799) and C770:G891 (Ec: C679:G798) to respectively form type I and II A-minor motifs. The adjacent C162 (Ec: C192) interacts with C768:U893 (Ec: U677:C800) (Fig. 4.1). Its major groove is almost completely blocked by the 395 (Ec: 388) hairpin loop and L15e; the backbones are strongly interacted by the Arg and Lys residues of L15e. Interestingly, its flanking 5′-helix hairpin loop interacts strongly with the 5′-helix hairpin loop of the AGUA/GAA motif at position 212 (Ec: 241) in the 23S rRNA, whose minor groove is also blocked by the 395 (Ec: 388) hairpin loop and two ribosomal proteins, L15 and L15e. The second one at position 2691 (Ec: 2654) in the SRL loop in the 23S rRNA has the minor groove completely blocked by L6 and has strong hydrophobic interactions with such residues as Phe, Leu, and His. However, its major groove, the putative binding site of elongation factors EF-G and EF-Tu, is completely left open, implying that a sticky motif could be the binding site of any external molecule. The third one at position 77 (Ec: 76) in the 5S rRNA, as described above, strongly interacts in the minor groove with the AGUA/GAA motif at position 952 (Ec: 858) in the 23S rRNA.

Of the three variable motifs, the first one at position 441 (Ec: 447) in 16S rRNA is highly conserved in bacterial and chloroplastic sequences (Table 4.4) but is not involved in any interaction with other parts of rRNA. Its minor groove is
partly blocked by S16 and forms only a single hydrogen bond between C=O of C443 (Ec: C449) and –NH₂ of Arg42 (3.24 Å). The second one at position 292 (Ec: 286) in 23S rRNA occurs in one of the hypervariable regions in 23S rRNA and interacts with neither RNA nor proteins. The third one at position 568 (Ec: 544i.2) in 23S rRNA is highly conserved only in archaeal sequences (Table 4.4). Its minor and major grooves are almost completely left open, while C=O of the bulged G is hydrogen bonded to –NH₂ of Arg154 in L30 nearby. Thus, the variable motifs tend not to interact with other parts of rRNA and proteins. In this respect, the G(U:A) base triple for the AGUA/GAA motif is an intrinsic characteristic.

4.2.4.2 An AGUA/GAA motif at the core of 16S rRNA

The AGUA/GAA motif at position 866 (Ec: 889) in the 16S rRNA occurs in the 6x4 asymmetric internal loop directly 3’ to the helix, 862-864/887-889 (Ec: 885-887/910-912), which was initially proposed based on a covariation analysis of 16S rRNA sequences. Unlike the previous prediction, however, the three consecutive basepairs are created differently (Fig. 4.1), so that the sheared G865:A886 (Ec: G888:A909) basepair immediately 5’ to the motif serves as an AA.AG@helix.ends motif. Surprisingly, U239 (Ec: U244), which is more than 98% conserved in the three phylogenetic domains, is recruited from afar and basepaired to C870 (Ec: C893), the remaining unpaired nucleotide in the internal
loop that is C in Archaea and Bacteria or U in Eukaryotes. This long-range interaction agrees with the cross-linking data between U239 (Ec: U244) and G871 (Ec: G894). The resulting tertiary basepair, U239:C870 (Ec: U244:C893), adopts the Watson-Crick conformation (U:C WC) and is stacked between the two basepairs flanking C893 (Ec C870). The minor groove is partly blocked by the helix, 1391-1408/1461-1468 (Ec: 1409-1416/1484-1491), directly downstream to the A-site.

The covariation-based helix, 862-864/887-889 (Ec: 885-887/910-912), has been proposed to switch to the alternative helix, 865-867/887-889 (Ec: 888-890/910-912), in tRNA selection during translation. Interestingly, however, the same covariation-supported helix, 862-864/887-889 (Ec: 885-887/910-912), was observed in both the open and closed forms during tRNA selection. Theoretically, the conversion to the alternative helix should involve a dramatic conformational change in the 30S subunit by destroying eight basepairs: three in the covariation-based helix, three in the AGUA/GAA motif, G865:A886 (Ec: G888:A909), and U239:C870 (Ec: U244:C893). An analysis of the 5.5 Å resolution T. thermophilus 70S crystal structure containing the three tRNAs and mRNA (PDB 1GIX and 1GIY) reveals the sharp S-turn in the backbone containing the AGUA sequence, suggesting the AGUA/GAA motif and U239:C870 (Ec: U244:C893) and indicating that, if true, the conformational
switch should be reversible. However, the reversible breaking and reforming of the eight basepairs would be energetically unfavorable. Moreover, the highly stabilized AGUA/GAA motif will not be easily broken. Furthermore, the conformational switch under dramatic conformational change may not be topologically feasible. Taken together, it is not likely that the conformational switch could take place.

4.2.4.3 An AGUA/GAA motif at the core of 23S rRNA

The AGUA/GAA motif at position 1369 (Ec: 1265) in the 23S rRNA occurs in the internal loop flanked by the two helices with only two basepairs, 1366-1367/2057-2058 (Ec: 1262-1263/2016-2017) and 1373-1374/2051-2052 (Ec: 1269-1270/2010-2011),\(^{116}\) which comprise the core of 23S rRNA (Fig. 4.1). The former helix is stacked onto its direct 5'-helix, 636-641/1360-1365 (Ec: 579-584/1256-1261), and the latter helix is stacked onto its immediate 3'-helix, 1725-1728/2047-2050 (Ec: 1648-1651/2006-2009). These two helical stacking interactions not only link domain II and IV, but also induce an additional helical stacking of the two nested helices, 1382-1387/1395-1400 (Ec: 1276-1281/1289-1294) and 1401-1404/1718-1721 (Ec: 1295-1298/1642-1645).

This motif makes two hydrogen bonds in the minor groove with two distal regions: one between N1 of A2054 (Ec: A2013) and \(-\text{NH}\) of U2648 (Ec: U2613) (2.91 Å) in domain V, and the other between N3 of A2055 (Ec: A2014) and NH
of U840 (\textit{Ec}: U746) (2.89 Å) in domain II. Parts of the minor groove and the entire major groove are blocked by L22, forming several hydrogen bonding interactions: one between N7 of the bulged G1370 (\textit{Ec}: G1266) and –OH of Ser24 (2.77 Å), and the other between N3 of A2054 (\textit{Ec}: A2013) and –NH$_2$ of Arg128 (2.36 Å). Together, this motif plays a significant role in structuring 23S rRNA. For example, a mutagenesis study indicated that this core region is involved in protein synthesis by modulating the interaction with the 30S ribosomal subunit.$^{117}$

4.2.4.4 An AGUA/GAA motif at position 952 (\textit{Ec}: 858) in 23S rRNA

In contrast to other AGUA/GAA motifs in the asymmetric internal loops, the one at positions 952 (\textit{Ec}: 858) in 23S rRNA occurs in the highly conserved 4x4 symmetric internal loop in domain II (Fig. 4.1). G953 (\textit{Ec}: G859) is bulged out into the major groove and forms the G(U:A) base triple as usual. Surprisingly, however, the expected G952:A1014 (\textit{Ec}: G868:A918) was not formed; instead, G952 (\textit{Ec}: G868) is basepaired to C1015 (\textit{Ec}: U919) and A2302 (\textit{Ec}: A2268) in domain V is recruited into the minor groove to be basepaired to A1014 (\textit{Ec}: A918). The resultant A:A rS basepair, A1014:A2302 (\textit{Ec}: A918:A2268), is then stacked between the G(U:A) base triple and G952:C1015 (\textit{Ec}: G868: U919), achieving the AGUA/GAA motif.

As described above, this motif extensively interacts in the minor groove with the AGUA/GAA motif at position 77 (\textit{Ec}: 76) in the 5S rRNA, probably
playing a significant role in ribosomal association. In addition, the major groove is partly blocked by L21e, which electrostatically interacts with the backbone containing the AGUA sequence.

4.2.4.5 An AGUA/GAA motif at position 1327 (Ec: 1346) in 16S rRNA

The AGUA/GAA motif at position 1327 (Ec: 1346) in 16S rRNA occurs in the three-stem loop in domain III (Fig. 4.1). The AAA stack in the minor groove is completely blocked by the helix, 916-920/1321-1325 (Ec: 939-943/1340-1344), and S7. A1357 (Ec: A1375) of the leading A:A rS basepair make an interaction with G916:C1344 (Ec: G939:C1344) to form a type II-like A-minor motif interaction. The backbone containing the GAA sequence is strongly contacted by the positively charged Arg and Lys residues of S7. The side chain of Arg10 stretches into the major groove and is stacked directly onto A1327 (Ec: A1346); its ε-NH is hydrogen bonded to C=O of U1358 (Ec: U1376) in the reversed wobble basepair, U:U rWb, immediately 5' to the motif, mimicking a base triple and being stacked onto the A:A rS basepair of the motif. In addition, the major groove is blocked by S9; both C=O and N7 of G1355 (Ec: G1373) in the trailing A:G S basepair are hydrogen bonded to –NH₂ of Lys11, and both C=O and N7 of the bulged G1328 (Ec: G1347) are hydrogen bonded to the side chain of Arg10.
4.2.4.6 An AGUA/GAA motif at the multistem loop in domain II of 23S rRNA

The AGUA/GAA motif at position 911 \((Ec: 818)\) in 23S rRNA occurs in the multistem loop with seven helical stems (Fig. 4.1). This motif is not achieved as previously predicted \(^{35}\) but in a way that is not simply imaginable: Two nucleotides across the multistem loop, A1070 \((Ec: A972)\) and G1071 \((Ec: G973)\), are recruited to create the two unexpected tertiary basepairs, A913:G1071 \((Ec: A820:G973)\) and A914:A1070 \((Ec: A821:A972)\). The former basepair adopts the reversed sheared conformation \((A:G \ rS)\) isosteric to \(A:A \ rS\),\(^{112}\) while the latter basepair assumes the sheared conformation \((A:A \ S)\) and mimics an AA.AG@helix.ends basepair with respect to the helix, 1044-1054/1060-1069 \((Ec: 946-955/962-971)\), in the three-dimensional structure. G1291 \((Ec: A1186)\) is displaced and stacked onto the G1292 \((Ec: G1187)\) of the GG stack in the major groove. This unusual structure motif mediates the helical stacking of the two flanking helices across the multistem loop, followed by inducing two other helical stackings of two sets of nested helices. Thus, this motif contributes to the compaction and folding of the 23S rRNA. When it comes to the interaction with protein(s), the phosphate backbone containing the AGUA sequence is interacted by L32e.
4.2.4.7 Other AGUA/GAA motifs inducing helical stacking in multistem loops

Two other AGUA/GAA motifs at positions 380 (Ec: 371) and 463 (Ec: 457) in 23S rRNA induce the helical stacking in the multistem loops (Fig. 4.1). In the former, two nucleotides A429 (Ec: A422) and A430 (Ec: A423) across the multistem loop are recruited into the minor groove and contact with A407 (Ec: A401) and A408 (Ec: A402), respectively. While the tertiary basepair, C240:G379 (Ec C269:G370), is stacked onto its immediate 5'-helix, its adjacent 3'-helix is pulled over across the multistem loop and coaxially stacked onto the helix, 412-416/424-428 (Ec: 406-410/417-421). The unpaired A410 (Ec: A404) is stacked between the two coaxially stacking helices, while A411 (Ec: A405) is flipped out. This “through-space” helical stacking in turn induces the second helical stacking of the inserted helix (designated as “a” in Fig. 4.1) and the helix, 236-239/431-434 (Ec: 265-268/424-427). The resulting two sets of the stacked helices sit side by side. An α-helix of L15e occupies along the major groove and its Lys13 is stacked onto A380 (Ec: A371), causing the buckling of A380:A408 (Ec: A371:A402). Other positively charged Arg and Lys residues from L15e make strong electrostatic interactions with the backbones.

In the latter, G458 (Ec: G452) interacts with A:A rS, forming a base triple that is stacked between A459 (Ec: A453) and the GG stack in the minor groove. In addition, C478 (Ec: A472) is stacked between C461 (Ec: C455) and the AAA
stack in the minor groove. The two flanking helices of U30:A453 (Ec: U33:A447) are coaxially stacked with the unpaired C31 (Ec: U34) flipped out, probably due to the two base stacking interactions (Fig. 4.1). Moreover, A462 (Ec: C456) is flipped out into the minor groove and stacked by His37 of L39e, and its N3 is hydrogen bonded to 2'-OH of U465 (Ec: U459). Furthermore, while the minor and the major groove is mostly left open, the backbones containing AGUA and GAA sequences make strong electrostatic interactions with L39e and L4, respectively.

4.2.4.8 An AGUA/GAA motif at position 1219 (Ec: 1238) in 16S rRNA

Surprisingly, the AGUA/GAA motif at position 1219 (Ec: 1238) in 16S rRNA lacks the trailing A:G basepair (Fig. 4.1). This motif was initially identified based on the analysis of the 16S rRNA in the T. thermophilus 30S crystal structure. Although highly conserved in Archaea and Bacteria (Table 4.4), this motif could have not been identified based on the secondary structure models alone; it occurs in the multistem loop with the “partial” AGU/AA sequence. This motif is disordered and only partly stacked onto its flanking lonepair 3'-helix, 1284/1315 (Ec: 1303/1334), which is also stabilized by the GGA/GAA motif at position 1285 (Ec: 1304) (see below).

The leading A:A rS basepair, A1220:A1280 (Ec: A1239:A1299), has an opening to one side. While A1220 (Ec: A1239) of the A:A basepair is stacked
onto G1222 (Ec: G1241) of the flanking basepair, G1222:C1277 (Ec: G1241:C1296), the bulged G1281 (Ec: G1300) and U1282 (Ec: U1301) of the U:A rH basepair are stacked onto G1315 (Ec: G1334) and C1284 (Ec: C1303) of the flanking lonepair 3’ to the motif, respectively. While U1283 (Ec: U1302) directly 3’ to the AGU sequence is flipped out, and C1279 (Ec: C1298) immediately 5’ to the AGU sequence is stacked onto the side chain of Arg114 from S7.

4.2.4.9 An AGUA/GAA motif at position 1971 (Ec: 1930) in 23S rRNA

The motif at position 1971 (Ec: 1930) in 23S rRNA was initially predicted as an GUA/GAA motif and occurs at the intersubunit bridge B2b in a multistem loop in domain IV (Fig. 4.1). Amazingly, however, it is “transformed” into an AGUA/GAA motif in rRNAs in the H. marismortui 50S crystal structures; G1971 (Ec: G1930) is retreated into the major groove and interacts with the U:A rH basepair, U1972:A2010 (Ec: U1931:A1969). The resultant G(U:A) base triple is directly stacked onto the helix, 1891-1893/1944-1946 (Ec: 1835-1837/1903-1905) by stacking the G:U pS* basepair onto G1891:C1946 (Ec: G1835:C1905). However, this transformed AGUA/GAA motif lacks the leading A:A rS basepair. A2010 (Ec: A1969) interacts with C1889:G2013 (Ec: C1833:G1972) to form a type I-like A-minor motif.71-72,112
This “through-space” helical stacking is probably induced by the lonepair stacking in the nearby lonepair triloop motif. In addition, its flanking 3’-helix made up of only two basepairs is pseudo-stacked onto the helix, 1986-1991/1997-2002 (Ec: 1945-1950/1956-1961), by the mediation of the base stacking of the two unpaired bases, G1976 (Ec: G1935) and U2003 (Ec: U1962). No ribosomal proteins interact with this motif. Thus, the structure motif should not be a complete reflection of its respective sequence motif.

4.2.4.10 GUA/GAA motifs in symmetric internal loops

Four GUA/GAA motifs occur in the symmetric 3x3 internal loops: two conserved ones at positions 564 (Ec: 581) and 763 (Ec: 780) in the 16S rRNA, and two variable one at positions 706 (Ec: 635i) and 1543 (Ec: 1438) in the 23S rRNA. All of the four motifs has the A(U:A)A stack in the minor groove. The first conserved one at position 564 (Ec: 581) in the 16S rRNA, which contains the GUA/GAG sequence, is strongly contacted by the C-terminal region of S17: While the α-helix near the C-terminal makes a strong electrostatic contact in the minor groove, the C-terminal loop stretches between G743 (Ec: G760) and G744 (Ec: G761), and then the –CO₂H of the C-terminal Ala105 is hydrogen bonded to N3 of G564 (Ec: G581). Another α-helix from S15 makes a strong electrostatic contact along the backbone containing the GUA sequence. The second conserved one at position 763 (Ec: 780) in the 16S rRNA makes a contact between 2’-OH of
A764 (Ec: A781) and C=O of U1499 (Ec: U1522) in the terminal helix of the 16S rRNA (2.53 Å) (Fig. 4.2). However, it has no interactions with proteins although its narrowed major groove is partly occupied by S11.

The first variable one at position 706 (Ec: 635i) in 23S rRNA is highly conserved only in archaeal sequences (Table 4.4). G722 (no homologous position in E. coli) in the minor groove is hydrogen bonded to the phosphate group of G938 (Ec: G845) in the GGA/GAA motif at position 937 (Ec: 844) in the 23S rRNA (Fig. 4.2). The minor groove is strongly contacted by L18e. The second variable one at position 1543 (Ec: 1438) in the 23S rRNA (Fig. 4.2), which contains the GUC/GAA sequence, occurs in a highly variable region; its trailing C:C S basepair, C1545:C1640 (Ec: A1439:A1552), makes a contact with U1702 (Ec: U1627) in the minor groove. Interestingly, A1616 (Ec: A1515) interacts with the nearest basepair, G1542:C1643 (Ec: G1436:C1556), in the flanking 5'-helix to form a type I A-minor motif.71-72,112

4.2.4.11 A GUA/GAA motif at position 668 (Ec: 685) in 16S rRNA

The motif at position 668 (Ec: 685) in the 16S rRNA is currently the only example of the GUA/GAA motifs that are formed in the asymmetric internal loops and contains the GUA/GAU sequence (Fig. 4.2). The leading G:U basepair adopts the sheared conformation, G:U S.112 Immediately following this motif, the flanking 3'-helix is sharply bent and interacts with the minor groove of this motif.
In addition, the minor groove is strongly contacted by a \_\_pleated sheet from S11. This motif happens to occur at the site of one of the K-turn motifs previously described. Interestingly, however, the UAA base stack in the minor groove is very similar to the AAA stack for the GGA/GAA motif, rationalizing the exchange between GUA/GAA and GGA/GAA motifs at this position (Table 4.4).

### 4.2.4.12 GGA/GAA motifs in symmetric internal loops

Four GGA/GAA motifs at positions 1285 (Ec: 1304) in the 16S and 23 (Ec: 26), (Ec: 795704), and 2502 (Ec: 2467) in the 23S rRNAs occur in the symmetric and highly conserved internal loops. The first one at position 1285 (Ec: 1304) in the 16S rRNA is flanked by the lonepair 5\'-helix, 1284/1315 (Ec: 1303/1334), and occurs in the 3x3 internal loop near the multistem loop that contains the AGUA/GAA motif at position 1219 (Ec: 1238) (Fig. 4.3). Here, the unstable lonepair helix is stabilized by the formation of this motif. The AAA stack in the minor groove is contacted both by the helix across the multistem loop and by part of S13; 2\'-OH of A1287 (Ec: A1306) is hydrogen bonded both to 2\'-OH of G924 (Ec: G947) (3.25 Å) and to –OH of Thr109 (3.21 Å). A1314 (Ec: A1333) interacts with A923:U1216 (Ec: A946:U1235) to form a type II-like A-minor motif. The major groove is blocked and electrostatically contacted by parts of S13 and THX.
The second one at position 23 (Ec: 26) in the 23S rRNA unusually contains two additional basepairs, one directly 5' and the other directly 3' to the motif (Fig. 4.3). This motif stabilizes the flanking lonepair 3'-helix, 27/516 (Ec: 30/510). The minor groove is contacted by the helix at the core of the 23S rRNA, 636-641/1360-1365 (Ec: 579-584/1256-1261); A520 (Ec: A514) interacts with C638:G1363 (Ec: C581:G1259) to form a type II A-minor motif. Additionally, A521 (Ec: A515) of the additional basepair, U22:A521 (Ec: U25:A515), interacts with C637:G1364 (Ec: C580:G1260) to form type I A-minor motif. However, no ribosomal proteins are not associated with this motif.

The third one at position 795 (Ec: 704) in the 23S rRNA was initially predicted in an asymmetric internal loop in domain II. The formation of this motif, however, creates the additional basepair, U794:A819 (Ec: U703:A728), directly 3' to the motif; thus, this motif occurs in a symmetric internal loop (Fig. 4.3). The highly conserved G820 (Ec: G729) is flipped out and basepaired to another highly conserved C1830 (Ec: C1774) in domain IV, forming a G:C basepair in the Watson-Crick conformation. The minor groove is occupied mostly by the N-terminal region of L37ae: The backbone NH of Arg10 is hydrogen bonded to both 2'-OH of A796 (Ec: A705) (3.09 Å) and O4' of A797 (Ec: A706) (2.37 Å); 2'-OH of A797 (Ec: A706) is hydrogen bonded to –NH₂ of
Lys34 (3.19 Å). The minor groove is also blocked by L19e; the pro-S phosphate oxygen of G816 (Ec: G725) is hydrogen bonded to –NH₂ of Lys91 (2.61 Å).

The last one at position 2502 (Ec: 2467) in the 23S rRNA contains the non-canonical CAA/GAC sequence and occurs the internal loop in a helix that is projected from the peptidyl transferase loop in domain V (Fig. 4.3). While this motif has no interactions with other sections of the 23S rRNA, the minor groove is in contact with two β-helices from L10e; C=O of C2518 (Ec: C2483) is hydrogen bonded to –NH₂ of Lys152 (3.09 Å). On the contrary, the major groove is not contacted by any ribosomal proteins.

4.2.4.13 GGA/GAA motifs in asymmetric internal loops

Four GGA/GAA motifs at positions 78 (Ec: 82), 937 (Ec: 844), 1316 (Ec: 1212), and 1588 (Ec: 1484) in the 23S rRNAs occur in the asymmetric internal loops; while the third one is present at a highly conserved region, the remaining three occur at the variable regions (Table 4.2). All of the four GGA/GAA motifs have their leading basepair directly 3' to their flanking 5’-helix (Fig. 4.3). Interestingly, however, the first three have their GGA sequence that are directly linked to their flanking 3’-helix, while the fourth at position 1588 (Ec: 1484) in the 23S rRNAs is interrupted by an additional unpaired nucleotide between the GGA sequence and the flanking 3’-helix.
All of the four GGA/GAA motifs are directly accompanied by the sharp bending in their sugar phosphate backbones; while maintaining the AAA stack in the minor groove, the trailing A:G S basepair is buckled, leading to the perturbed GGG stack in the major groove. In addition, the unpaired base directly 5' to the GAA sequence is characteristically flipped out into the major groove. In contrast, the minor groove is almost completely blocked by the flanking 3'-helix, while creating two hydrogen bonds between the motif and the flanking 3'-helix (Fig. 4.3). Interestingly, the first three share the same pattern in hydrogen bonding interactions: one between the A of the GGA sequence and the unpaired nucleotide immediately 3' to the flanking 3'-helix, and the other between the middle A of the GAA sequence and the paired nucleotide directly 3' to the GGA sequence. On the contrary, the fourth at position 1588 (Ec: 1484) in the 23S rRNAs has the different hydrogen bonding pattern: one between the A of the GGA sequence and the first paired nucleotide 5' to the GAA sequence, and the other between the middle A of the GAA sequence and the second paired nucleotide 5' to the GAA sequence. In particular, the unpaired base immediately 3' to the flanking 3'-helix, A1603 (Ec: A1501), is stacked between the unpaired base directly 3' to the GGA sequence, A1591 (Ec: A1487), and the AAA stack, thereby extending the AAA stack in the minor groove.
When it comes to the interaction with proteins, the first one at position 78 (Ec: 82) in the 23S rRNA has the GGA-containing backbone wrapped around by L24. The G97 (Ec: G101) of the trailing A:G S basepair is “basepaired” to –CO₂H of Asp105 by forming two hydrogen bonds, followed by stacking onto the side chain of Lys107. In addition, the phosphate oxygen atoms of A80 (Ec: A84) are hydrogen bonded to the side chains of Asn43 and Arg111. In contrast, the remaining three are not or only weakly contacted by ribosomal proteins, indicating that the sharp bending following the GGA/GAA motif in the asymmetric internal loops be not guided by ribosomal proteins.

4.2.5 Sticky motifs interacting with metal ions

14 sticky motifs interact with metal ions, usually in the major groove (Table 4.2): five in the 16S and nine in the 23S rRNA. Interestingly, all of the 14 sticky motifs interacting with metal ions are conserved and 12 are involved in a long-range tertiary interaction (Figures 4.1, 4.2, and 4.3). In particular, the AGUA/GAA motif at position 911 (Ec: 818) in 23S rRNA, which requires the recruitment of two nucleotides across the multistem loop and induces the two sets of helical stackings of the nested helices, contains five bound metal ions in proximity. This suggests that the binding of metal ions may play a role in structuring the local structures by either stabilizing the sticky motif or mediating tertiary interactions. While the metal ions bound to the H. marismortui 50S
crystal structures were determined including Mg$^{2+}$, K$^+$, and Na$^+$, the identity of the individual metal ions in the *T. thermophilus* 30S crystal structure was not resolved.

### 4.2.6 Phylogenetic distribution of sticky motifs in rRNAs

In addition to the 33 sticky motifs identified in the *T. thermophilus* 16S and the *H. marismortui* 23S and 5S rRNAs, additional 34 sequence motifs were identified in the variable regions of the rRNAs, based on the rRNA comparative secondary structure models currently available at the CRW Site (http://rna.icmb.utexas.edu/), amounting to a total of 67 sequence motifs in rRNAs: 21 in 16S, 45 in 23S, and one in 5S rRNAs (*Table 4.4*). Of the additional 34 sequence motifs, four at positions 644 (*Ec*: 661), 1241 (*Ec*: 1260), and 1399 (*Ec*: 1417) in the 16S and 1788 (*Ec*: 1710) in the 23S rRNAs have their homologous region in the rRNAs in the *T. thermophilus* 30S and *H. marismortui* 50S crystal structures; however, they are shortened to tandem GA sequences or exchange with canonical basepairs in one or more phylogenetic groups. The approximate locations of the 67 sequence motifs are mapped onto the *E. coli* rRNA comparative structure models (*Fig. 4.5*); the collection of the fragments containing the additional 34 sequence motifs from the representative rRNA comparative structure models is shown in *Appendix J*. 
**Fig. 4.5** All identified AGUA/GAA, GUA/GAA, and GGA/GAA motifs mapped onto comparative structure models: (a) tRNA; (b) 5S rRNA; (c) 16S rRNA; (d) 23S rRNA
4.2.7 Exchanges between sticky motifs

Interestingly, either two or all three of the AGUA/GAA, GUA/GAA, and GGA/GAA sequence motifs are observed for 13 of the total of 67 locations in different rRNA sequences or in different phylogenetic groups (Table 4.4 and Appendix I). Six of the 13 locations at 69i (Ec: 69i), 70i (Ec: 70i) and 71i (Ec: 73i) in 16S rRNA and 1521i.2 (Ec: 1415i.2), 1627i.3 (Ec: 1530i), and 1790i.1 (Ec: 1713) in 23S rRNA accommodate either AGUA/GAA or GGA/GAA motif. Five locations at 668 (Ec: 685) in 16S rRNA and 78 (Ec: 82), 1543 (Ec: 1438), 1579 (Ec: 1475), and 1790i.2 (Ec: 1721) in 23S rRNA exchange between the GUA/GAA and GGA/GAA motifs in different phylogenetic domains. The one at position 668 (Ec: 685) in 16S rRNA occurs in an asymmetric internal loop, but it has the AAA stack in the minor groove instead of the A(U:A)A stack, rationalizing the structural similarity of the AGUA/GAA, GUA/GAA, and GGA/GAA motifs. The one at position 78 (Ec: 82) in 23S rRNA usually occurs in archaeal and bacterial sequences. The one at position 1543 (Ec: 1438) in 23S rRNA occurs in archaeal and bacterial sequences. The one at position 1579 (Ec: 1475) in 23S rRNA contains the GUA/GAA motif in 40% of the bacterial sequences and the GGA/GAA motif in 3% of the eukaryotic sequences (Appendix I); In addition, this motif exchanges with the AA/GA or GA/GA
tandems in Bacteria, Eukaryotes, and Chloroplasts. The one at position 1790i.2 (Ec: 1721) in 23S rRNA has either the GUA/GAA or GGA/GAA motif in bacterial sequences, while its basepairs exchange with the canonical basepairs in archaeal and mitochondrial sequences.

The one at position 292 (Ec: 286) in 23S rRNA exchanges between all of the AGUA/GAA, GUA/GAA, and GGA/GAA motifs: the AGUA/GAA motif in 8% of the archaean sequences; the GUA/GAA motif in 4% of the bacterial sequences; the GGA/GAA motif in 1% of the eukaryotic sequences (Appendix I). The region containing this motif is left gapped in the chloroplastic and mitochondrial sequences. Interestingly, the motif at position 77 (Ec: 76) in 5S rRNA contain the similar E loop sequences but is not homologous across the three phylogenetic domains; while the archaean and eukaryotic sequences with the asymmetric E loop contain the AGUA/GAA motif, the bacterial sequences with the symmetric E loop have the GUA/GAA motif (Appendix I), leading to the sequence similarity between the AGUA/GAA and GUA/GAA motifs.

In addition, the GGA/GAA motif at positions 1241 (Ec: 1260) and 1399 (Ec: 1417) in 16S rRNA exchanges with a tandem GA motif by replacing the leading basepair with a canonical basepair: The former has the GGA/GAA motif in 33% of the bacterial sequences, while containing the tandem GA motif in the 66% of the archaean sequences (Appendix I). For example, the leading basepair
is replaced with G:C and adopts the Watson-Crick conformation in the 16S rRNA in the *T. thermophilus* 30S crystal structure. The latter usually contains the tandem GA motif, while one mitochondrial sequence, *Acanthamoeba castellanii*, has the GGA/GAA motif (*Appendix J*). Furthermore, all three basepairs in four motifs at positions 644 (*Ec*: 661) in 16S and 298 (*Ec*: 294), 1588 (*Ec*: 1484), and 1788 (*Ec*: 1710) in 23S rRNA are exchangeable with three canonical basepairs.

### 4.3 Discussion

#### 4.3.1 K-turns versus sticky motifs

The K-turn is an RNA structure motif whose asymmetric internal loop is flanked with a sheared G:A basepair on one side and G:C (or C:G) basepair on the other, causing the sharp turn in the RNA backbone. On the contrary to the AA.AG@helix.ends motif with the G directly 3' to a regular helix, the G of the sheared G:A basepair involved in the K-turn motif is always 5' to a regular helix. Based on the analysis of the rRNAs in the *T. thermophilus* 30S and *H. marismortui* 50S crystal structures, eight K-turns were previously identified at positions 241 (*Ec*: 246) and 670 (*Ec*: 687) in the 16S rRNA and 80 (*Ec*: 84), 247 (*Ec*: 269i), 939 (*Ec*: 846), 1151 (*Ec*: 1047), 1316 (*Ec*: 1212), 1590 (*Ec*: 1486) in the 23S rRNA. All of the eight K-turn motifs were not predicted but identified; they were not predictable without the three-dimensional structures. Interestingly,
however, five of them occur in the asymmetric internal loops containing either the GGA/GAA or GUA/GAA motif: four GGA/GAA motifs at positions 78 (Ec: 82), 937 (Ec: 844), 1216 (Ec: 1212), and 1588 (Ec: 1484) in the 23S rRNA, and one GUA/GAA motif at position 668 (Ec: 685) in the 16S rRNA.

The GGA/GAA and GUA/GAA motifs can be classified into two types, type I and II: Type I has the GGA (or GUA) sequence directly 5' to the flanking 3'-helix, while type II contains an additional unpaired nucleotide between the GGA (or GUA) sequence and the flanking 3'-helix. When articulated in the asymmetric internal loops, they create two characteristic hydrogen bonds in the minor groove (Figures 4.2 and 4.3). The five K-turn motifs overlapping with the GGA/GAA or GUA/GAA motifs can be similarly classified; four type I K-turns at positions 668 (Ec: 685) in the 16S rRNA and 78 (Ec: 82), 937 (Ec: 844), 1216 (Ec: 1212) in the 23S rRNA, and one type II K-turn at position 1588 (Ec: 1484) in the 23S rRNA. Thus, it is now possible to predict only by analyzing the RNA secondary structure models. As an example, the GGA/GAA motif at position 2176 (Ec: 2133) will form a type I K-turn motif (Fig. 1.3).

More interestingly, two K-turn motifs at positions 241 (Ec: 246) in the 16S and 1151 (Ec: 1047) in the 23S rRNA occur in the asymmetric internal loops that contain the three-dimensional structure similar to that in the GGA/GAA motif (Fig. 2.2). Furthermore, the remaining K-turn motif at position 247 (Ec: 269i) in
23S rRNA has the GA/GU sequence (Fig. 2.2). Nonetheless, both basepairs adopt their respective sheared conformations, thereby suggesting the shortening of a GGA/GAA motif to a tandem GA motif. In this respect, the K-turn motif is a subfamily of the GGA/GAA motif.

4.3.2 GUA/GAA motif candidates not formed in the crystal structures

All of the four sequence motifs that are not formed in the crystal structures are the GUA/GAA motif candidates and occur in the asymmetric internal or multistem loops; these occur at positions 408 (Ec: 413) and 748 (Ec: 765) in the 16S and 304 (Ec: 298) and 1765 (Ec: 1687) in the 23S rRNA (Table 4.2 and Fig. 4.2). This implies that the putatively least stable GUA/GAA motif would be destabilized in the asymmetric internal and multistem loops. Nonetheless, the nucleotides in all of the four GUA/GAA motif candidates interact with either other section(s) of rRNA or proteins.

The candidate at position 408 (Ec: 413) in the 16S rRNA occurs in an asymmetric internal loop and strongly interacts with S4 in the major groove, probably preventing the articulation of this candidate. The candidate at position 748 (Ec: 765) in the 16S rRNA, occurring in a multistem loop, unusually have an additional non-canonical A:C basepair directly 5’ to itself and sits near a site involved in ribosomal subunit association. In addition, this candidate maintains the AAA stack in the minor groove; A749 (Ec: A766) and A750 (Ec: A767) form
type I A-minor motifs with U1487:G1502 (Ec: C1510:G1525) and G1488:C1501 (Ec: G1511:C1524) in the terminal helix in the 16S rRNA, respectively. This strong RNA-RNA interaction may disturb the formation of this candidate.

The candidate at position 304 (Ec: 298) in the 23S rRNA occurs in a multistem loop containing a lonepair triloop motif.39 Interestingly, A329:U346 (Ec: A322:U339) forms the unusual A:U pfS* conformation 112 and is roughly stacked between the two basepairs with the sheared conformation, G304:A347 (Ec: G298:A340) and A331:G345 (Ec: A324:G338), thereby leading to the overall stacking of the two flanking helices. However, the helical stacking per se may be induced by the coaxial stacking of the lonepair in the lonepair triloop motif. The candidate at position 1765 (Ec: 1687) in the 23S rRNA occurs in an asymmetric internal loop in domain IV. G1777 (Ec: G1699) is flipped out and causes the reversal of A1776 (Ec: A1698), leading to the reversed sheared A:A basepair, A1767:A1776 (Ec: A1689:A1698). Interestingly, comparative basepair in the flanking 3'-helix, C1768:A1775 (Ec: C1690:A1697), adopts the sheared conformation, C:A S.112

4.3.3 A base triple in the L11-binding domain

A1193 (Ec: A1089) in the L11-binding domain of the 23S rRNA in the H. marismortui 50S crystal structure (PDB 1JJ2 45) is bulged out into the major groove and form an A(A:U) base triple, A1193(A1194:U1205) (Ec:
A1089(A1090:U1101), where A1194:U1205 (Ec: A1090:U1101) is in the Watson-Crick conformation. In one sense, this A(A:U) base triple is similar to the G(U:A) base triple for the AGUA/GAA motif. The same A(A:U) base triple was observed at the homologous position in the crystal structure of the E. coli L11-binding region (PDB 1QA6)\(^{118}\). Interestingly, the A(A:U) base triple exchanges with G(U:U) in the three phylogenetic domains. The crystal structure of the Thermotoga maritima L11-binding domain (PDB 1MMS)\(^{119}\) revealed the G(U:U) base triple with the wobble U:U basepair, U:U Wb, suggesting that the isosteric basepair conformation be maintained for the homologous basepairs in rRNAs.

4.3.4 Sticky motifs in small catalytic RNAs

AGUA/GAA, GUA/GAA, and GGA/GAA motifs are also observed in other catalytic RNAs. For example, the hairpin ribozyme has the AUUA/GAA sequence near the catalytic site. In fact, the AUUA/GAA sequence in the recent crystal structure (PDB 1M5V)\(^{120}\) revealed the similar structure to the AGUA/GAA motif by forming the three consecutive non-canonical basepairs: the reversed Hoogsteen A:A basepair (A:A rH), the U(U:A) base triple containing the reversed Hoogsteen U:A basepair (U:A rH), and the sheared A:G basepair (A:G S). The A:A rH is a substitute for the A:A rS in the AGUA/GAA motif. Interestingly, the bulged U runs parallel to the helical axis and its C=O is
hydrogen bonded to –NH\textsubscript{2} of the A of the U:A rH basepair in the U(U:A) base triple. Thus, this may explain the exchange between the AGUA/GAA and either AUUA/GAA or GUUA/GAG sequences, e.g. 441 (Ec: 447), 866 (Ec: 889) in the 16S and 159 (Ec: 189) in the 23S rRNA (Appendix I).

The hammerhead ribozyme contains the UGA/GAA sequence near the catalytic pocket, which forms three sheared basepairs in the crystal structures (PDB 1HMH\textsuperscript{121} and 1MME\textsuperscript{122}), leading to the GGA/GAA motif. The additional A:U basepair formed directly 5’ to the motif mediates the coaxial stacking of the two stems, II and III. Thus, this motif may play an important role in creating the catalytic pocket.

4.3.5 Possible roles of sticky motifs in RNA structure

The recurrent use of AGUA/GAA, GUA/GAA, and GGA/GAA motifs in RNA molecules implies their structural and functional roles in RNA structure. (1) Sticky motifs stabilize internal loops or helical ends in multistem loops by creating three consecutive basepairs in the unpaired regions of the rRNA secondary structure models. (2) Sticky motifs may play a role in RNA compaction not only by mediating the helical stacking of the two flanking helices directly in internal loops and indirectly (through space) in multistem loops, but also by inducing the helical stacking(s) of neighboring or distal helices. (3) Sticky motifs work as “molecular magnets” to involve other sections of the rRNA
in the tertiary interactions; the AAA and A(U:A)A stack in the minor groove interact with other sections of the RNA chain, thereby facilitating RNA folding as well as helping the association of ribosomal subunits. (4) Sticky motifs may be involved in fine-tuning the RNA function by organizing local or global structures. (4) Sticky motifs may serve as recognition sites for proteins or metal ions. The partial negative charges on the G’s and U’s in the major groove can be contacted by the positively charged residues of proteins or metal ions. (5) Sticky motifs add more basepairs to the covariation-based RNA secondary structure models, contributing to the improvement of the RNA comparative structure models.

4.4 MATERIALS AND METHOD

The structural details of the AGUA/GAA, GUA/GAA, and GGA/GAA motifs were investigated based on the analysis of the high-resolution crystal structures of the rRNAs in the *T. thermophilus* 30S (PDB 1FJF 63) and the *H. marismortui* 50S (PDB 1FFK 64 and 1JJ2 45) ribosomal subunits, the hairpin ribozyme (PDB 1M5V 120), and the hammerhead ribozyme (PDB 1HMH 121 and 1MME 122), using the RasMol program.104-105 The crystal structures of the *E. coli* 5S rRNA (PDB 1DFU 111) and the sarcin/ricin loop (PDB 483D 103) in 23S rRNA were also investigated and compared with their corresponding motifs in the *H. marismortui* 50S crystal structures. In addition, the crystal structures of the L11
binding domains from two other organisms, *E. coli* (PDB 1QA6\textsuperscript{118}) and *T. maritima* (PDB 1MMS\textsuperscript{119}), were investigated. Although the structure of the 23S rRNAs in the *D. radiodurans* 50S crystal structure (PDB 1LNR\textsuperscript{67}) is in overall agreement with that in the *H. marismortui* 50S crystal structure (PDB 1JJ2\textsuperscript{45}), the structural details of the sticky motifs in the former are not remained consistent, so that the *D. radiodurans* 50S crystal structure (PDB 1LNR\textsuperscript{67}) was not employed in our detailed structural analysis.

Additional information including RasMol scripts for the sticky motifs is available at http://www.rna.icmb.utexas.edu/ANALYSIS/EELL/. Each individual sticky motif can be displayed interactively with RasMol\textsuperscript{104-105} or Protein Explorer.\textsuperscript{123} In particular, all the structural details and interactions of the sticky motifs with other section(s) of rRNA and proteins can also be displayed using the rasmol scripts.
Chapter 5: The Lonepair Triloop: A New Motif in RNA Structure

5.1 BACKGROUND

Given the successful prediction of basepairs (approximately 97-98%) in the 16S and 23S rRNAs and sequence motifs described above, we now question if comparative analysis will reveal other structural motifs and the rules for their prediction. Our earlier covariation analysis of the 16S and 23S rRNA revealed seven examples of a lone basepair capped by three nucleotides in a hairpin loop.\textsuperscript{9,20,61,124-126} Four of these had a stronger pattern of covariation at the two basepaired positions, while three had a weaker pattern and were considered tentative interactions.\textsuperscript{61} Herein we describe, based on the analysis of the crystal structures of the \textit{T. thermophilus} 30S (PDB 1FJF\textsuperscript{63}), \textit{H. marismortui} 50S (PDB 1FFK\textsuperscript{64} and 1JJ2\textsuperscript{45}), and \textit{D. radiodurans} 50S (PDB 1KC9 and 1LNR)\textsuperscript{67} ribosomal subunits, the validation of these predicted lonepair triloop motifs, the identification of more lonepair triloop motifs in the rRNA crystal structures and structural features that are characteristic of this new RNA structural motif, and relate this information with comparative data describing the types of sequences present in this structural motif and their extent of sequence conservation. This analysis of the lonepair triloop motif reveals several dimensions of comparative
analysis and the additional contributions that comparative methods make to our understanding of RNA structure.

5.2 Results

The lonepair triloop (LPTL) is a RNA structural motif containing a lone basepair capped with a hairpin loop of three nucleotides and can be expressed as 5'-FXYZL-3'. The underlined nucleotides, F and L (the "first" and "last" nucleotides of this five-nucleotide motif, respectively), form a lonepair and the three nucleotides (X, Y, and Z) form a triloop. Throughout the text, the nucleotide positions for the rRNAs in the crystal structures of the *T. thermophilus* 30S (PDB 1FJF) and *H. marismortui* 50S (PDB 1FFK and 1JJ2) ribosomal subunits (called crystal structure numberings or Xtal) are used together with the *E. coli* numberings (Ec) in parenthesis and the basepair conformations are based on the recently proposed Lee-Gutell classification system.

5.2.1 Initial identification of lonepair triloops with covariation analysis

Three more "tentative" LPTLs were also predicted with covariation analysis at positions 911:915 (Ec: 934:937) and 1035:1039 (Ec: 1053:1057) in the 16S rRNA and 335:339 (Ec: 328:332) in the 23S rRNA. Interestingly, the two LPTLs in the 16S rRNA were unusual since they were found in helical regions. While the lonepairs in the first four LPTLs contain a significant amount of positional covariation, the three "tentative" LPTLs contain a minimal amount of covariation.

5.2.2 Analyses of the tRNA and 30S and 50S ribosomal crystal structures

Analyses of the atomic crystal structures of the T. thermophilus 30S (PDB 1FJF)\textsuperscript{63} and H. marismortui 50S (PDB 1FFK\textsuperscript{64} and 1JJ2\textsuperscript{45}) ribosomal subunits not only substantiated the seven (four confident and three tentative) LPTLs predicted with covariation analysis, but also identified 16 additional LPTLs that do not have positional covariations and cannot be predicted with covariation analysis. Of these 23 LPTLs in the rRNAs, seven occur in the 16S, 15 in the 23S, and one in the 5S rRNA. An analysis of the D. radiodurans 50S crystal structure (PDB 1KC9 and 1LNR)\textsuperscript{67} revealed LPTLs at all of the 23S rRNA positions that are strictly equivalent to the H. marismortui 23S rRNA. Two of the LPTLs in the H. marismortui 23S rRNA are not fully equivalent to those in the D. radiodurans 23S rRNA: The D. radiodurans equivalent of 125:129 (Ec: 131:148) is part of a longer helical stem, as it is in the majority of the bacteria, including E. coli; the 1651:1656 (Ec: 1565:1568) LPTL has the hairpin loop with two nucleotides
instead of three. While the *D. radiodurans* 50S crystal structure is in overall agreement with the *H. marismortui* 50S crystal structure, the former was not employed in our detailed structural analysis due to its lower resolution. An additional LPTL was found in the T loop of the tRNA crystal structures with and without tRNA synthetases [with tRNA synthetases (PDB 1ASY, 2FMT, 1QF6, 1F7U, 1IL2, and 1G59) and without tRNA synthetases (PDB 6TNA, 1B23, and 1EXD)], leading to the total of 24 LPTLs (*Table 5.1*).

All of the 24 LPTLs are highlighted on the covariation-based secondary structure models for the *T. thermophilus* 16S, the *H. marismortui* 23S and 5S rRNAs, and the *S. cerevisiae* Phe-tRNA (*Fig. 5.1*); 14 occur in multi-stem loops, seven in hairpin loops, and three in internal loops (*Table 5.2*). While C:G and G:C lonepairs tend to form Watson-Crick conformations, the majority of lonepairs adopt non-canonical basepair conformations, with several types of lonepair conformations in the multi-stem loops. Interestingly, all of the seven LPTLs occurring in hairpin loops in the 16S and 23S rRNA contain a U:A lonepair with the rH conformation. Six of these seven U:A basepairs are invariant in the three phylogenetic domains data sets (*Table 5.3*).

### 5.2.3 Classification of lonepair triloops

Lonepair triloops were classified based upon three sets of criteria: (1) proximity to the 5' adjacent helix and the presence of a U-turn in the triloop
<table>
<thead>
<tr>
<th></th>
<th>Ec’</th>
<th>Xtal’</th>
<th>LP (X)</th>
<th>LP (C)</th>
<th>TL (X)</th>
<th>TL (C)</th>
<th>dCC</th>
<th>lps</th>
<th>d</th>
<th>sp/bf</th>
<th>tf</th>
<th>L</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>16S</td>
<td>323:327</td>
<td>U:A (H)</td>
<td>U:A (99.6)</td>
<td>GAG</td>
<td>GAG (97.6)</td>
<td>9.58</td>
<td>5’</td>
<td>†</td>
<td>†</td>
<td>33332/3bmm</td>
<td>X/S20</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td>16S</td>
<td>1177:1181</td>
<td>G:G (H)</td>
<td>G:G (95.9)</td>
<td>GAA</td>
<td>GAA (95.4)</td>
<td>12.42</td>
<td>5’</td>
<td>†</td>
<td>†</td>
<td>33332/3bmm</td>
<td>X/Y/S10</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td>16S</td>
<td>1315:1319</td>
<td>G:G (H)</td>
<td>U:A (99.7)</td>
<td>GCA</td>
<td>GCA (87.9)</td>
<td>9.54</td>
<td>5’</td>
<td>†</td>
<td>†</td>
<td>33332/3bmm</td>
<td>X/Y/S14, S19</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td>23S</td>
<td>306:310</td>
<td>U:A (H)</td>
<td>U:A (96.2)</td>
<td>GGA</td>
<td>GGR (79.4)</td>
<td>9.67</td>
<td>5’</td>
<td>†</td>
<td>†</td>
<td>33332/3bmm</td>
<td>X/Y+Z/L24</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td>23S</td>
<td>475:479</td>
<td>U:A (H)</td>
<td>U:A (82.0)</td>
<td>GCA</td>
<td>GAA (85.4)</td>
<td>9.62</td>
<td>3’</td>
<td>†</td>
<td>†</td>
<td>33332/3bmm</td>
<td>X/L24</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td>23S</td>
<td>499:503</td>
<td>C:A (C)</td>
<td>U:A (92.9)</td>
<td>GGA</td>
<td>GAA (54.3)</td>
<td>9.76</td>
<td>5’</td>
<td>†</td>
<td>†</td>
<td>33332/3bmm</td>
<td>X/—</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td>23S</td>
<td>1282:1286</td>
<td>U:A (H)</td>
<td>U:A (83.8), c:a (10.0)</td>
<td>GAG</td>
<td>GAR (94.1)</td>
<td>9.78</td>
<td>5’</td>
<td>†</td>
<td>†</td>
<td>33332/3bmm</td>
<td>X/—</td>
<td>H</td>
</tr>
<tr>
<td></td>
<td>23S</td>
<td>328:332</td>
<td>C:G (L)</td>
<td>U:A (96.1)</td>
<td>GAC</td>
<td>GAB (66.3)</td>
<td>9.50</td>
<td>5’</td>
<td>†</td>
<td>†</td>
<td>32222/buuum</td>
<td>X/Y/L4, L24</td>
<td>H</td>
</tr>
</tbody>
</table>

| R2 | 23S | 567:571 | U:A (H) | c:a (43.2), u:u (31.7) | UUG | UUG (96.4) | 9.40 | 5’ | † | † | 33332/3bmm | Y/L15 | H | IA |
|    | URNA | 54:58 | U:A (H) | U:A (95.2) | UCG | UCR (99.1) | 9.77 | 5’ | † | † | 33332/3bmm | X/Y/— | H | IA |
| R3 | 16S | 956:960 | U:U (Wb) | U:U (99.6) | UAA | UAA (97.7) | 9.01 | 5’ | † | † | 33332/3bmm | Z-(G:C)/S19 | M | IB |
|    | 23S | 1082:1086 | C:G (rWC) | U:A (52.4), c:g (46.8) | UAA | UAA (83.4) | 11.14 | 5’ | † | † | 33333/3bmm | Z-(G:C)/— | M | IB |
|    | 23S | 1925:1929 | C:G (rWC) | U:A (98.9) | UAA | UAA (83.4) | 11.14 | 5’ | † | † | 33332/3bmm | Z-(G:C)/— | M | IB |
|    | 23S | 2562:2566 | U:A (H) | U:A (68.8), c:u (25.3) | UAA | UAA (83.4) | 9.53 | 5’ | † | † | 33333/3bmm | Z-(G:C)/L14 | M | IB |
|    | 23S | 319:323 | G:C (WC) | C:G (77.4), c:g (16.3) | AUA | RWA (68.0) | 10.95 | 5’ | † | † | 33322/MMmm | Z/L4 | M | IB |
|    | 23S | 476:480 | G:A (S) | G:A (88.5) | CAA | AAA (92.7) | 9.39 | 5’ | † | † | 33332/3bmm | Z/L24 | M | IB |
| 5S | 24:27 | G:C (rWb) | g.m (45.2), a.a (43.0) | UUG | — (98.5) | 9.22 | 5’ | † | † | 32333/MMbb | Z/— | I | IB |
| R4 | 23S | 1752:1756 | C:G (WC) | c:g (47.5), g:c (35.7) | GCA | GYA (65.3) | 10.69 | 5’ | † | † | 33333/3bmm | Z-(G:C)/L24/2e | M | IB |
|    | 23S | 2447:2451 | 2482:2486 | G:A (H) | G:A (97.4) | AUA | AUA (99.8) | 12.83 | 5’ | † | † | 32233/MMmmM | Z/Y/— | M | IB |
|    | 23S | 131:148 | 125:129 | U:A (WC) | a.u (44.7), g:c (41.6) | CUU | — (100) | 10.74 | 5’ | † | † | 37333///7??? | None/— | M | IB |
|    | 23S | 1565:1568 | 1651:1655 | C:G (WC) | G:C (94.6) | CAU | CDU (100)* | 10.52 | 5’ | † | † | 32223/MMmbb | None/L2, 37ae | M | IB |
|    | 23S | 64:68 | 64:68 | G:G (H) | G:G (78.1), u:u (14.6) | UGC | ARC (70.7) | 11.49 | 5’,3’ | † | † | 32333/MMmmM | Z+(Y+Z)/— | I | III |
|    | 23S | 934:938 | 911:915 | C:A (rWb) | C:A (96.0) | ACA | ACA (83.0) | 11.39 | 3’ | † | † | 23333/mmMmmM | X/Y+Z/— | M | III |
|    | 23S | 1053:1057 | 1035:1039 | G:G (H) | G:G (99.4) | CAU | CAU (96.8) | 11.31 | 5’,3’ | † | † | 32333/mmMmmM | Y+Z+/S3 | I | III |

---

a: Nucleotide position numbers for the lonepairs in E.coli and crystal structure organisms, respectively; b: Lonepairs and triloops in the crystal structures.

c: Comparative information for lonepairs and triloops in the nuclear-encoded tRNA genes in the three phylogenetic domains for rRNAs and Type 1 tRNAs for tRNAs. The dominant (more than 50% conserved) sequences are in uppercase and the minor (10-50% conserved) sequences in lowercase. The IUPAC-IBC nomenclature for nucleotides was utilized (Cornish-Bowden, A. (1985). *Nucleic Acids Res*. 13, 3021-3030). The asterisk (*) represents the consensus triloop sequence in Archaea only.
d: dCC: the C1'-C1' distance in the lonepairs in the crystal structures; lps: Lonepair stacking onto the closest 5'- and/or 3'-helices.

e: Directionality of the two local chains involved in a lonepair: ↑↓, antiparallel; †↑↓, parallel.
f: sp/bf: sugar puckering/bases facing for each of the five nucleotides in the lonepair triloop motif. Sugar puckering: 3, C3'-endo; 2, C2'-endo. Bases facing into: M, major groove; m, minor groove; h, boundary of major and minor grooves; u, straight up the backbone.
g: The tertiary interactions of LPTL: X, Y and Z indicate the first, second and third nucleotides in the triloop, respectively. Interactions with ribosomal proteins are indicated following the slash.
h: Loop types. H, hairpin; I, internal; M, multistem; C, Classification of lonepair triloops.

Table 5.1 Analyses of LPTLs in rRNA and tRNA Sequences and Structures

129
Fig. 5.1 Lonepair triloops in the crystal structures: (a) *S. cerevisiae* (Sc) Phe-tRNA (PDB 6TNA<sup>101</sup>), (b) *H. marismortui* (Hm) 5S rRNA (PDB 1JJ2<sup>45</sup>), (c) *T. thermophilus* (Tt) 16S rRNA (PDB 1FJF<sup>63</sup>) and (d) *H. marismortui* 23S rRNA (PDB 1JJ2<sup>45</sup>). Asterisks (*) represent LPTLs occurring at the interface and double asterisks (**) mark the LPTLs that are at the interface and within 10 Å of functional sites.
<table>
<thead>
<tr>
<th>Lonepair</th>
<th>Hairpin Loop (H)</th>
<th>Multi-stem Loop (M)</th>
<th>Internal Loop (I)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>U:A†</td>
<td>7 (rH)</td>
<td>3 (2 rH, 1 WC)</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>C:G</td>
<td>0</td>
<td>3 (2 WC, 1 rWC)</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>G:G</td>
<td>0</td>
<td>1 (rH)</td>
<td>2 (H)</td>
<td>3</td>
</tr>
<tr>
<td>C:A</td>
<td>0</td>
<td>2 (1 rH, 1 rWb)</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>G:A</td>
<td>0</td>
<td>2 (1 S, 1 H)</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>G:C</td>
<td>0</td>
<td>1 (WC)</td>
<td>1 (fS*)</td>
<td>2</td>
</tr>
<tr>
<td>U:G</td>
<td>0</td>
<td>1 (rWb)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>U:U</td>
<td>0</td>
<td>1 (rWb)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>7</strong></td>
<td><strong>14</strong></td>
<td><strong>3</strong></td>
<td><strong>24</strong></td>
</tr>
</tbody>
</table>

†The lonepair T:A in the T loop of tRNAs is included here.

**Table 5.2** Distribution of lonepair loops
<table>
<thead>
<tr>
<th></th>
<th>Xta</th>
<th>Conservation</th>
<th></th>
<th>Xta</th>
<th>Conservation</th>
<th></th>
<th>Xta</th>
<th>Conservation</th>
</tr>
</thead>
<tbody>
<tr>
<td>5S rRNA (3P)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>22</td>
<td>G/a (50.4/42.3)</td>
<td>131</td>
<td>125</td>
<td>a (33.1)</td>
<td>1082</td>
<td>1186</td>
<td>U/c (52.3/46.9)</td>
</tr>
<tr>
<td>25</td>
<td>23</td>
<td>A/u (65.3/20.8)</td>
<td>-</td>
<td>-</td>
<td>126</td>
<td>1084</td>
<td>1188</td>
<td>A (99.2)</td>
</tr>
<tr>
<td>-</td>
<td>24</td>
<td>a/u (48.6/22.0)</td>
<td>-</td>
<td>127</td>
<td>1085</td>
<td>1189</td>
<td>A (99.0)</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>25</td>
<td>a/g (47.9/25.6)</td>
<td>-</td>
<td>128</td>
<td>1086</td>
<td>1190</td>
<td>A/g (52.9/46.9)</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>26</td>
<td>C (88.6)</td>
<td>148</td>
<td>129</td>
<td>u/c (49.9/41.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16S rRNA (3P)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>64</td>
<td>G (79.0)</td>
<td>306</td>
<td>313</td>
<td>U (78.1)</td>
<td>1282</td>
<td>1388</td>
<td>U (84.6)</td>
</tr>
<tr>
<td>65</td>
<td>65</td>
<td>A (88.9)</td>
<td>307</td>
<td>314</td>
<td>G (83.1)</td>
<td>1284</td>
<td>1390</td>
<td>A (95.1)</td>
</tr>
<tr>
<td>66</td>
<td>66</td>
<td>A/g (61.5/36.8)</td>
<td>309</td>
<td>316</td>
<td>A (82.6)</td>
<td>1285</td>
<td>1391</td>
<td>G (78.7)</td>
</tr>
<tr>
<td>67</td>
<td>67</td>
<td>C (77.5)</td>
<td>310</td>
<td>317</td>
<td>A (87.2)</td>
<td>1286</td>
<td>1392</td>
<td>A (95.7)</td>
</tr>
<tr>
<td>68</td>
<td>68</td>
<td>G (81.3)</td>
<td>311</td>
<td>318</td>
<td>C (77.7)</td>
<td>1287</td>
<td>1393</td>
<td>C (95.5)</td>
</tr>
<tr>
<td>323</td>
<td>318</td>
<td>U (99.9)</td>
<td>319</td>
<td>326</td>
<td>C (77.7)</td>
<td>1565</td>
<td>1651</td>
<td>C (95.5)</td>
</tr>
<tr>
<td>324</td>
<td>319</td>
<td>G (99.4)</td>
<td>320</td>
<td>327</td>
<td>A (72.4)</td>
<td>1566</td>
<td>1653</td>
<td>A (95.5)</td>
</tr>
<tr>
<td>325</td>
<td>320</td>
<td>A (98.9)</td>
<td>321</td>
<td>328</td>
<td>U (65.1)</td>
<td>1567</td>
<td>1654</td>
<td>A (94.7)</td>
</tr>
<tr>
<td>326</td>
<td>321</td>
<td>G (98.4)</td>
<td>322</td>
<td>329</td>
<td>A (90.2)</td>
<td>1568</td>
<td>1655</td>
<td>G (95.5)</td>
</tr>
<tr>
<td>109</td>
<td>102</td>
<td>A (99.6)</td>
<td>323</td>
<td>330</td>
<td>G (81.6)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>327</td>
<td>322</td>
<td>A (99.7)</td>
<td>328</td>
<td>335</td>
<td>U (99.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>934</td>
<td>911</td>
<td>C (96.3)</td>
<td>329</td>
<td>336</td>
<td>G (81.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>935</td>
<td>912</td>
<td>A (99.4)</td>
<td>330</td>
<td>337</td>
<td>A (87.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>936</td>
<td>913</td>
<td>C (99.5)</td>
<td>331</td>
<td>338</td>
<td>A (43.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>937</td>
<td>914</td>
<td>A (83.7)</td>
<td>332</td>
<td>339</td>
<td>A (98.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>938</td>
<td>915</td>
<td>A (96.3)</td>
<td>475</td>
<td>481</td>
<td>U (82.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>956</td>
<td>933</td>
<td>U (99.7)</td>
<td>476</td>
<td>482</td>
<td>G (88.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>957</td>
<td>934</td>
<td>U (98.1)</td>
<td>477</td>
<td>483</td>
<td>A (93.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>958</td>
<td>935</td>
<td>A (99.8)</td>
<td>478</td>
<td>484</td>
<td>A (95.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>959</td>
<td>936</td>
<td>A (99.6)</td>
<td>480</td>
<td>486</td>
<td>A (95.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>960</td>
<td>937</td>
<td>U (96.8)</td>
<td>479</td>
<td>485</td>
<td>A (95.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1053</td>
<td>1035</td>
<td>G (99.5)</td>
<td>476</td>
<td>482</td>
<td>G (88.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1054</td>
<td>1036</td>
<td>C (99.8)</td>
<td>477</td>
<td>483</td>
<td>A (93.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1055</td>
<td>1037</td>
<td>G (99.5)</td>
<td>478</td>
<td>484</td>
<td>A (95.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1056</td>
<td>1038</td>
<td>U (97.0)</td>
<td>479</td>
<td>485</td>
<td>A (95.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1057</td>
<td>1039</td>
<td>G (99.7)</td>
<td>480</td>
<td>486</td>
<td>A (95.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1177</td>
<td>1158</td>
<td>G (96.2)</td>
<td>499</td>
<td>505</td>
<td>U (93.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1178</td>
<td>1159</td>
<td>G (97.0)</td>
<td>500</td>
<td>506</td>
<td>G (67.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1179</td>
<td>1160</td>
<td>A (96.9)</td>
<td>501</td>
<td>507</td>
<td>A (71.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1180</td>
<td>1161</td>
<td>A (98.5)</td>
<td>502</td>
<td>508</td>
<td>A (91.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1157</td>
<td>1139</td>
<td>A (99.6)</td>
<td>504</td>
<td>511</td>
<td>A (95.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1181</td>
<td>1162</td>
<td>G (98.5)</td>
<td>503</td>
<td>509</td>
<td>A (94.6)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1115</td>
<td>1296</td>
<td>U (99.9)</td>
<td>567</td>
<td>624</td>
<td>U/c (54.6/43.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1136</td>
<td>1297</td>
<td>G (99.8)</td>
<td>568</td>
<td>625</td>
<td>U (99.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1117</td>
<td>1298</td>
<td>C (88.4)</td>
<td>569</td>
<td>626</td>
<td>U (97.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1138</td>
<td>1299</td>
<td>A (99.9)</td>
<td>570</td>
<td>627</td>
<td>G (100)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>978</td>
<td>953</td>
<td>A (99.5)</td>
<td>2030</td>
<td>2071</td>
<td>A (90.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.3 Single nucleotide frequencies of LPTL nucleotides

---

8 The nucleotides that are conserved in more than 50% of the nuclear-encoded archael, bacterial and eucaryotic rRNA sequences (3P) and all type 1 tRNAs (Z) are shown in capital letters; nucleotides with 30-50% conservation are in lowercase letters. The fourth position that is recruited between positions Z and L in type A motifs is italicized.

9 The nucleotide numbers of RNAs present in the *T. thermophilus* 30S (PDB 1FJF), *H. marismortui* 50S (PDB 1JJ2), and the *S. cerevisiae* Phe-tRNA (PDB 6TNA) crystal structure. The dash (-) mark represents rRNA positions that are not equivalent to *E. coli* (Ec) positions.

10 The nucleotides of rRNAs present in less than 50% of the nuclear-encoded archael, bacterial and eucaryotic rRNA sequences (3P) and all type 1 tRNAs (Z) are shown in capital letters; nucleotides with 30-50% conservation are in lowercase letters. The fourth position that is recruited between positions Z and L in type A motifs is italicized.
<table>
<thead>
<tr>
<th>LPTL Groups</th>
<th>LPTL-RNA</th>
<th>Class I</th>
<th>Class II</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Type A</td>
<td>Type B</td>
<td></td>
</tr>
<tr>
<td>R1 (UGNRA)</td>
<td>X or X+Y</td>
<td>5</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>R2 (UYYRA)</td>
<td>Y</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>R3 (NUAAN, NRWAN)</td>
<td>Z</td>
<td>0</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>R4 (NRYAN)</td>
<td>Y+Z</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>R5 (NCUN)</td>
<td>—</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>7</td>
<td>10</td>
<td>21</td>
</tr>
</tbody>
</table>

* Triloop sequences were based on the more than 50% conservation for the single nucleotide frequencies in the three phylogenetic domains, R = purines (A or G) and Y = pyrimidines (C or U).

Table 5.4 Further grouping of class I and II LPTLs
(Classes I, II and III); (2) recruitment of a tertiary base from another region of the RNA (Type A and B); and (3) triloop nucleotide involved in tertiary interactions with other regions of the RNA (Groups R1–R5). Each classification system has some degree of overlap with the others. The three systems will be described individually below.

Three major conformations are observed for the lonepair triloops in the rRNAs and tRNAs. As depicted in Fig. 5.2, the "canonical" LPTLs are directly (Class I) or indirectly (Class II) appended to the helix that is upstream of the lonepairs. A U-turn \(^{22-23}\) forms between the first and second positions in the triloop (positions X and Y) in these two categories. The last category (Class III) occurs within a helical region and is unusual due to (1) at least two of the triloop bases basepairing to form part of a regular helical stem and (2) the missing U-turn in the triloop.

The 21 Class I and II LPTLs are subdivided into nine Type A and 12 Type B LPTLs. The characteristic feature of the Type A motifs is the recruitment of a tertiary base (tB) from another region of the RNA molecule (Table 5.3). This recruited base pairs to the first base of the triloop (X) and stacks between the third base (Z) and the 3' base in the lonepair (L) to mediate the stacking of three consecutive bases (Y, Z, and L). The net result is a structural conformation that resembles the tetraloop motif.\(^{26-27}\) The tertiary base is usually an A, although
Fig. 5.2 Schematic representation of LPTLs and their classification: Arrows indicate triloop nucleotides involved in tertiary interactions (solid, required; hollow, optional).
other nucleotides are observed (Table 5.3). In contrast to Type A motifs, Type B motifs do not weave a tertiary base between positions Z and L. The three consecutive nucleotides, Y, Z, and L, are stacked in some but not all of these Type B LPTLs. Five of the 12 Type B motifs have a strong contact with a G:C basepair within a helical stem.

Class I and II LPTLs can be further divided into five groups (R1–R5), based on which nucleotide(s) in the triloop are primarily involved in a tertiary interaction with another region of the RNA. The number (1-5) associated with the R designates which (if any) of the nucleotides in the triloop is involved in a tertiary interaction (Table 5.1 and Fig. 5.1); the "1" designates the first (or the first and second), "2" designates the second, "3" designates the third, "4" designates the second and third, and "5" designates no nucleotides. While seven of the ten R1 and R2 LPTLs occur in the hairpin loops, ten of the eleven R3, R4, and R5 LPTLs are in the multi-stem loops. Compared to the former classification, nine of the ten R1 and R2 LPTLs are Type A motifs that mimic the tetraloop conformation. One exception occurs at positions 335:339 (Ec: 328:332) in 23S rRNA: Although an interaction occurs with the first two nucleotides (thus it is an R1), the Type A conformation is prevented from forming because the two nucleotides in this triloop are basepaired to the first two nucleotides in the triloop of another LPTL, 313:317 (Ec: 306:310). All eleven
R3, R4, and R5 LPTLs are Type B motifs. The gallery of LPTL-containing RNA fragments (Fig. 5.3) is grouped according to this classification system (Groups R1–R5).

5.2.4 Overall three-dimensional architecture of the lonepair triloop motif

The majority of the lonepairs in Class I and II LPTLs adopt non-canonical basepair conformations (Table 5.1). With few exceptions, all contain the same set of three-dimensional architectural characteristics underlying the integrity of the lonepair triloop: (1) coaxial stacking of the LPTL lonepair onto the adjacent or nearest helical stem upstream of the LPTL; (2) stacking of X (the first base in the triloop) onto F (the 5' base of the lonepair) at the boundary of the major and minor grooves; (3) U-turn between positions X and Y; (4) two triloop bases Y and Z facing into the minor groove; (5) mediated (Type A) or direct (most Type B) stacking of the three consecutive bases of Y, Z, and L in the minor groove, (6) hydrogen bonding interactions between X and Z. The overall three-dimensional architecture of the LPTL can be viewed in the stereo images depicted in Fig. 5.4 and the interactive RasMol images available online at the CRW-LPTL Page (http://www.rna.icmb.utexas.edu/ANALYSIS/LPTL).

The tertiary base (tB) that is basepaired to X and stacked between the bases Z and L is an A in nearly all of the sequences in eight of the nine Type A motifs, and is a U in the other. Moreover, when Class I and II LPTLs contain a
**Fig. 5.3** Gallery of LPTL-containing RNA fragments with the nucleotides in different colors: red, ones involved in the LPTL; cyan, the direct or nearest 5'-basepair to the lonepair; green, nucleotides interacting with the triloop bases; purple, unpaired nucleotides between the 3' end of the LPTL and the 5'-helix.

138
Fig. 5.4 Stereoviews for representative IA, IB and IIA LPTLs with nucleotides in different colors: red, nucleotides in an LPTL; green, tertiary bases interacting with LPTLs; cyan, the closest 5'-basepair to a lonepair; purple, unpaired nucleotides between the 3' end of a LPTL and the 5'-helix onto which the lonepair is stacked; black, nucleotides forming the intervening basepair between the LPTL and the 5'-helix.
purine at the Z position, the 2'-OH of X is hydrogen bonded to the N7 of Z, while
X interacts with the phosphodiester backbone of L. However, when Z is a
pyrimidine (for example, LPTLs 335:339 (Ec: 328:332) and 1651:1655 (Ec:
1565:1568) in 23S rRNA), the LPTL is destabilized, has a strong interaction with
ribosomal proteins, and does not maintain the last five architectural characteristics
mentioned above.

5.2.5 Coaxial stacking of lonepairs

Earlier, it was observed that the first four covariation-based LPTLs were
all immediately adjacent and 3' to a helix. Since the lonepair of the LPTL motif
was considered energetically unstable, it was deduced that the lonepair needs to
form a coaxial stack with the 5'-helix to be stabilized. An analysis of the 16S,
and 23S, 5S rRNAs and tRNA crystal structures revealed that all but one of the 21
Class I and II LPTLs are indeed coaxially stacked onto the 5'-helix. The only
exception occurs in the 23S rRNA LPTL at 481:485 (Ec: 475:479). Here, the
lonepair, U481:A485 (Ec: U475:A479), is stacked between G482:A486 (Ec:
G476:A480) and G487:C515 (Ec: G481:C509), forming a short helix and thereby
preventing the U481:A485 (Ec: U475:A479) lonepair from being stacked onto its
5'-helix, 28-29/479-480 (Ec: 31-32/473-474), which is stacked against the helix,
12-27/516-531 (Ec: 15-30/510-525). The coaxial stackings for all of the rRNA
and tRNA lonepair triloops are shown in Fig. 5.3.
Two distinct patterns of coaxial stacking in lonepair triloops were observed in seven LPTLs occurring in hairpin loops: 318:322 (Ec: 323:327) and 1296:1300 (Ec: 1315:1319) in 16S rRNA, 313:317 (Ec: 306:310), 335:339 (Ec: 328:332), 624:628 (Ec: 567:571), and 1388:1392 (Ec: 1282:1286) in 23S rRNA, and 54:58 in tRNA.

The first pattern occurs in four of these LPTLs at positions 318:322 (Ec: 323:327) in 16S rRNA and 335:339 (Ec: 328:332), 624:628 (Ec: 567:571), and 1388:1392 (Ec: 1282:1286) in 23S rRNA. The coaxial stacking in this set of structures forces the unpaired bases between the 3' side of the lonepair and the 5'-helix out of the compound helix, and some of the "flipped-out" bases (shown in purple in Fig. 5.3) are involved in other interactions. For example, positions A340 (Ec: A333) and C341 (Ec: C334) in the 23S rRNA, which are at the 3' side of the U335:A339 (Ec: U328:A332) lonepair, are flipped out and basepaired to the two flipped-out bases, U325 (Ec: U318) and G324 (Ec: G317) respectively, beyond the 326:330 (Ec: 319:323) LPTL. In all four examples, the tertiary basepairs formed by the extruded bases are stabilized by stacking onto either basepair(s) or base(s).

The flipped-out bases in the remaining three LPTLs, 1296:1300 (Ec: 1315:1319) in 16S rRNA, 313:317 (Ec: 306:310) in 23S rRNA, and 54:58 in tRNA, are not involved in any interactions; instead, they are stacked onto another
basepair or base in the current set of crystal structures. However, in contrast to the *S. cerevisiae* Phe-tRNA crystal structure, the flipped-out base U59 is basepaired to U16 in the Asp-tRNA (PDB 1ASY and 1IL2) and Cys-tRNA (PDB 1B23) crystal structures. Unlike the Class I and II LPTLs, the lonepairs in Class III are stacked onto their 5'- and 3'-helices at 64:68 (*Ec*: 64:68) and 1035:1039 (*Ec*: 1053:1057) in 16S rRNA, while the lonepair at 911:915 (*Ec*: 934:937) in 16S rRNA is stacked onto the 3'-helix.

Several LPTLs in four-way junctions and multi-stem loops are nested within or adjacent to another set of helices that are coaxially stacked onto one another. Two LPTLs 125:129 (*Ec*: 131:148) and 1966:1970 (*Ec*: 1925:1927) in 23S rRNA have nested coaxial helices, while the two sets of coaxial helices in two LPTLs 1186:1190 (*Ec*: 1082:1086), 2597:2601 (*Ec*: 2562:2566) in 23S rRNA and the LPTL 54:58 in tRNA are adjacent to one another (Fig. 5.3). In three of these examples including 1186:1190 (*Ec*: 1082:1086), 1966:1970 (*Ec*: 1925:1929), and 2597:2601 (*Ec*: 2562:2566) in 23S rRNA, the two sets of stacked helices sit side-by-side with tertiary interactions between the LPTL triloop and another helix that is stacked in this compound complex. Tertiary interactions with the triloop of LPTL 125:129 (*Ec*: 131:148) in 23S rRNA could not be determined since this LPTL was not completely resolved in the crystal structure.
Nearly all of the LPTL sites in rRNAs maintain a LPTL in the rRNA sequence alignments. LPTL 125:129 (Ec: 131:148) in the *H. marismortui* 23S rRNA is an exception because most organisms have more than one basepair in this region. The variation in size is very dramatic, ranging from no basepairs to more than 50 in some of the organisms in the alpha proteobacteria. The typical number of basepairs is between three and ten; *E. coli* has seven basepairs and *D. radiodurans* has five basepairs. The 125:129 (Ec: 131:148) lonepair in the *H. marismortui* 23S rRNA is stacked onto its 5'-helix, 116-118/122-124 (Ec: 121-123/128-130), forming a compound helix with four basepairs. As well, a coaxial stack forms between the two equivalent helices in the *D. radiodurans* 23S rRNA, forming a compound helix with eight basepairs. The outer sets of helices are also coaxially stacked in the *D. radiodurans* 23S rRNA as they are in the *H. marismortui* 23S rRNA. This result suggests that one of the helices involved in a coaxial stack could have a single basepair.

5.2.6 Lonepair triloop-RNA interaction

5.2.6.1 R1 LPTLs (UGNRA)

The main unifying features that are present in nearly all of these R1 LPTLs are outlined here (Table 5.1): (1) The first (and second) nucleotides in the triloop form a tertiary interaction with another region of the RNA; (2) the lonepair is a U:A in at least 69% of the sequences; (3) all of the lonepairs form a reversed
Hoogsteen conformation; (4) the triloop sequence is GNR; (5) the tertiary interaction that forms between G, the first nucleotide in the triloop, and an A establishes a GNR'A' tetraloop sequence\textsuperscript{24} and conformation.\textsuperscript{26-27} In the eight R1 LPTLs, a few exceptions to these unifying features appeared. (1) The lonepair in LPTL 1158:1162 (Ec: 1177:1181) in 16S rRNA is a G:G (instead of U:A) in 96% of the sequences. (2) LPTL 335:339 (Ec: 328:332) in 23S rRNA has a GAC (instead of GNR) triloop sequence. (3) Only one of the R1 LPTLs, LPTL 335:339 (Ec: 328:332) in 23S rRNA, does not form a tetraloop conformation. Here, the triloop base A337 (Ec: A330, Y) is intercalated between the two bases A316 (Ec: A309, Z) and A317 (Ec: A310, L) to mediate the stacking of G315 (Ec: G308, Y), A316 (Ec: A309, Z), and A317 (Ec: A310, L) in LPTL 313:317 (Ec: 306:310). This LPTL-LPTL interaction is augmented with a second base stacking between the flipped-out base C318 (Ec: C311) in the 313:317 (Ec: 306:310) LPTL and the U335:A339 (Ec: U328:A332) lonepair in LPTL 335:339 (Ec: 328:332). This alternative conformation for LPTL 335:339 (Ec: 328:332) in 23S rRNA suggests that the pyrimidine at position Z destabilize the LPTL motif.

LPTL 505:509 (Ec: 499:503) in the 23S rRNA was included in this group since its lonepair is a U:A in the majority of the 23S (or 23S-like) rRNA sequences in the three phylogenetic domains (Table 5.1),\textsuperscript{61} although the \textit{H. marismortui} 23S sequence has a C:A lonepair. This C:A lonepair adopts the C:A
rH conformation equivalent to the U:A rH conformation. Additionally, this LPTL recruits a tertiary base to mimic the tetraloop sequence and structure.

5.2.6.2 R2 LPTLs (UUYRA)

The two R2 LPTLs, 624:628 (Ec: 567:571) in 23S rRNA and 54:58 in tRNA, are similar to the R1 LPTLs. First, the lonepairs are U:A in these two LPTLs, and both form the rH conformation (Table 5.1). While LPTL 54:58 in tRNA has a U:A lonepair in nearly all of the tRNA sequences, and LPTL 624:628 (Ec: 567:571) in the 23S rRNA has a C:A or U:U in 75% of the archaeal, bacterial and eucaryotic nuclear sequences (Table 5.1). Second, both LPTLs are Type A.

However, in contrast with the R1 LPTLs, these two R2 LPTLs have a tertiary interaction only with the second nucleotide in the triloop (Y) and have a UYR triloop sequence, rather than the GNR sequence in the majority of the R1 LPTLs (Table 5.1). LPTL 624:628 (Ec: 567:571) in the 23S rRNA recruits the tertiary base C2071 (Ec: A2030) between Z and L to mediate the stacking of the three consecutive bases (Y, Z, and L), while C2071 (Ec: A2030) is not basepaired to U625 (Ec: U568), the first nucleotide in the triloop (X), due in part to the smaller size of the U base. U626 (Ec: U569), the second nucleotide in the triloops (Y), forms two hydrogen bonds with A1081 (Ec: A983): one between C=O of U626 (Ec: U569) and –NH2 of A1081 (Ec: A983) and the other between 2‘-OH of U626 (Ec: U569) and N1 of A1081 (Ec: A983). However, considering that
C2071 (Ec: A2030) is mostly an A in the three phylogenetic domains (Table 5.3), the recruited base may be basepaired to U625 (Ec: U568, X). In LPTL 54:58 in the T loop of tRNAs, the tertiary base G18 (tB) inserts between G57 (Z) and A58 (L) and mediates the stacking of the three consecutive bases (C56, G57 and A58). However, the recruited base G18 makes a weak interaction with Ψ55 (X), while C56 (Y) and G19 strongly interact to form a standard Watson-Crick basepair.

5.2.6.3 R3 LPTLs (NUAAN or NRWAN)

Seven LPTLs form an interaction with Z, the third nucleotide in the triloop. Four of these LPTLs, 933:937 (Ec: 956:960) in 16S rRNA and 1186:1190 (Ec: 1082:1086), 1966:1970 (Ec: 1925:1929), and 2597:2601 (Ec: 2562:2566) in 23S rRNA, make an interaction in the minor groove between Z and the G of a G:C basepair to form a base triple in a nearby helix.71-72 These R3 LPTLs have a UAA triloop sequence in the vast majority of rRNA sequences (Table 5.1) and are further stabilized by the direct stacking of the three consecutive bases (Y, Z, and L) and by utilizing the 2′-OH groups in the vicinity of the G:C basepair as a source of hydrogen bonding. Their lonepairs are not restricted to a specific type of conformation, although they all have the reversed conformation (Table 5.1).

LPTL 1186:1190 (Ec: 1082:1086) was initially substantiated in two crystal structures for the L11-binding domains from E. coli (PDB 1QA618) and T.
*maritima* (PDB 1MMS\textsuperscript{119}) and was subsequently resolved for the *H. marismortui* 50S crystal structure.\textsuperscript{45} The sequence for this LPTL, UUAAA, is conserved in these three crystal structures, and the three-dimensional structures are identical with one another. Despite differences in base identity, the overall three-dimensional structures for the entire L11-binding domain from the three crystal structures are remarkably similar.

The remaining three R3 LPTLs do not form a base triple, do not have the conserved UAA triloop and lack the U-turn between X and Y. Instead, they have more heterogeneity in the triloop sequence and structural variations in the tertiary interactions with the third position of the triloop (Z). LPTL 326:330 (*Ec*: 319:323) 23S in the rRNA has a AUA triloop sequence in the crystal structure and a RWA sequence in 68% of the comparative sequences (Table 5.1). LPTL 482:286 (*Ec*: 476:480) in the 23S rRNA has a CAA triloop sequence in the crystal structure but an AAA triloop sequence in 93% of the comparative sequences. While LPTL 22:26 (*Ec*: 24:27) in the 5S rRNA has a UUG sequence in the *H. marismortui* 5S rRNA, nearly all of the other archaeal, bacterial, and eucaryotic 5S rRNAs have only two (instead of three) nucleotides between positions 22 (*Ec*: 24) and 26 (*Ec*: 27). Thus, the consensus of this 5S rRNA LPTL is shown as a dash in Table 5.1. The third base in the triloop, A329 (*Ec*: A322) in LPTL 326:330 (*Ec*: 319:323) is stacked between A305 (*Ec*: A299) and A347 (*Ec*: A340)
and is basepaired to U346 (Ec: U339), forcing the three bases G326 (Ec: G319, F), A327 (Ec: A320, X) and C330 (Ec: C323, L) into the major groove. As discussed above, all nucleotides in LPTL 482:286 (Ec: 476:480) are involved in the formation of the cross-LPTL A-stack. In LPTL 22:26 (Ec: 24:27) in the 5S rRNA, the third base in the triloop G25 (Ec: G26) is basepaired to A3 (no equivalent position in E. coli). This basepairing interaction pulls the 5' end of the A3-helix to the lonepair triloop, leading to the coaxial stacking of the two helices, 14-21/59-67 (Ec: 16-23/60-68) and 69-90/93-112 (Ec: 70-86/90-106).

5.2.6.4 R4 LPTLs (NRYAN)

Two LPTLs form tertiary interactions with the second and third nucleotides in the triloop (Y and Z). The sequences of these triloops are constrained to RYA. These R4 LPTLs occur in the 23S rRNA at positions 1808:1812 (Ec: 1752:1756) and 2482:2486 (Ec: 2447:2451). The second and third nucleotides in the triloop of LPTL 1808:1812 (Ec: 1752:1756) in the domain IV of 23S rRNA interact in the minor groove with two consecutive G:C basepairs in one helix of domain VI to form two A-minor motif base triples.71-72 Multiple 2'-OH groups in the vicinity of the G:C basepairs are hydrogen bonded to increase the stability of these base triples, analogous to the base triples in the R3 LPTLs. However, in contrast to R3 LPTL base triples, this tertiary interaction forms two consecutive base triples at the minor groove, with the base of L tucked under the
base of X at the boundary of the major and minor grooves. As a consequence, only the second and third bases in the triloop are stacked onto each other in the minor groove, instead of stacking the three consecutive bases Y, Z, and L.

LPTL 2482:2486 (Ec: 2447:2451) in the 23S rRNA is at the site of protein synthesis \(^{134}\) and reveals an unusual LPTL architecture. Although the G2482:A2486 (Ec: G2447:A2451) lonepair is immediately 3' to the 5'-helix 2105-2111/2476-2481 (Ec: 2064-2070/2441-2446) on the 23S rRNA secondary structure diagram (Fig. 5.3), a tertiary basepair A2485:C2536 (Ec: A2450:C2501) intercalates between the lonepair, G2482:A2486 (Ec: G2447:A2451), and the 5'-helix in the 23S rRNA in the \(H.\) marismortui 50S crystal structures,\(^{45,64}\) resembling the basepair between the lonepair and the 5'-helix in Class II LPTLs. The intervention of the A2485:C2536 (Ec: A2450:C2501) basepair brings the ends of the 2105-2111/2476-2481 (Ec: 2064-2070/2441-2446) and 2487-2492/2529-2535 (Ec: 2452-2457/2494-2500) helices into proximity so that they can coaxially stack onto one another in the peptidyl transferase loop. This conformation in the catalytic site for peptide bond formation is cooperatively stabilized by three more sets of base stacking interactions: one between A2059 and A2538 (Ec: A2503), one between A2101 (Ec: A2060) and G2537 (Ec: G2502) and the other C1044 (Ec: A945) and A2483 (Ec: A2448).
5.2.6.5 R5 LPTLs (NCUN)

Two LPTLs, 125:129 (Ec: 131:148) and 1651:1655 (Ec: 1565:1568) in the 23S rRNA, have no interactions with nucleotides in the triloop; the latter interacts with the ribosomal proteins L2 and L37ae (see below), while the former does not appear to interact with ribosomal proteins. Although the three bases in the triloop were not fully resolved in the H. marismortui 50S crystal structure (PDB 1FFK and 1JJ2), we doubt that this lonepair triloop interacts with ribosomal proteins since no ribosomal proteins are in proximity to this LPTL in the current crystal structures.

5.2.7 Lonepair triloop-protein interactions

Fourteen of the 23 rRNA LPTLs interact with ribosomal proteins (Table 5.1). The 318:322 (Ec: 323:327), 933:937 (Ec: 956:960) and 1158:1162 (Ec: 1177:1181) LPTLs in 16S rRNA make contacts with the side chains of S20, S19, and S10, respectively. In contrast, LPTL 1296:1300 (Ec: 1315:1319) in the 16S rRNA has a stacking interaction with S14 by stacking C1298 (Ec: C1317) directly against Phe16 of S14 and interacts electrostatically with S19 in the major groove. LPTLs 313:317 (Ec: 306:310) and 335:339 (Ec: 328:332) in the 23S rRNA interact with L4 and L24, which probably facilitate the LPTL-LPTL interaction between these two LPTLs. The ribosomal protein L24 also interacts with two more intimately clustered LPTLs, 481:485 (Ec: 475:479) and 482:486 (Ec:
476:480) in 23S rRNA. The 624:628 (Ec: 567:571), 1808:1812 (Ec: 1752:1756), and 2597:2601 (Ec: 2562:2566) LPTLs in 23S rRNA make contact with L15, L24e, and L14, respectively. The majority of these interactions between the lonepair triloop and a ribosomal protein do not affect the integrity of this RNA motif. The 54:58 LPTL in tRNA does not interact with tRNA synthetases in complexed crystal structures.

In contrast, a few LPTLs have an altered conformation due to their interaction with proteins. LPTL 326:330 (Ec: 319:323) has a different conformation due to a series of contacts with L4 in the minor groove, altering its architecture. As a consequence, the position A327 (Ec: A320) forms two H-bonds with L4: one between N7 of A327 (Ec: A320) and –NH₂ of Gln151 and the other between N3 of A327 (Ec: A320) and –NH₂ of Asn206. LPTL 1651:1655 (Ec: 1565:1568) is surrounded by L2 and L37ae, with the first position in the triloop, C1652 (no equivalent position in E. coli), penetrating deeply into L2 to stack against Phe169 in L2 and forming two H-bonds: one between –NH₂ of C1652 and the backbone C=O of Lys167 in L2 and the other between the phosphate group of C1652 and –NH₂ of Arg49 in L37ae. In addition, the second nucleotide in the triloop, A1653 (Ec: A1566), drops below the lonepair C1651:G1655 (Ec: C1565:G1568) to stack against His177 in L2, and its phosphate group is hydrogen bonded to the -OH group of Thr52 in L37ae.
Moreover, the third nucleotide in the triloop, U1654 (Ec: U1567), retreats into the major groove and intercalates between U1654 (Ec: C1565) and His47 in L2. These two lonepair triloops contain pyrimidines at the third position of the triloop, suggesting that pyrimidines in this position might alter the regularity of the lonepair triloop conformation.

5.2.8 Sugar puckering, bases facing, and directionality of local chains

The three-dimensional architecture of the lonepair triloop usually maintains the same conformation in the presence or absence of interactions with distal regions of the RNA chain or ribosomal proteins, with a few dramatic exceptions. The overall conformation can be described in terms of the changes in sugar puckering or bases facing of each nucleotide involved in the LPTL motif. As seen in Table 5.1, the most frequent pattern of sugar puckering/bases facing (sp/bf) is 33332/bbmmm, observed when a LPTL maintains its architectural integrity without significant structural perturbations caused by a tertiary interaction. Here, the "2" and "3" designate C2'-endo and C3'-endo puckering, respectively, and the "b" and "m" designate the bases facing into the major-minor groove boundary and the minor groove, respectively.

C2'-endo puckering is observed in B-DNA with an intrastrand interphosphate distance (dPP) of 7.0 Å, while C3'-endo puckering is observed in A-DNA with a shorter dPP of 5.9 Å. A-RNA shows a dPP similar to that of A-
DNA, suggesting that A-RNA should have C3'-endo puckering. In this respect, the presence of C2'-endo puckering in RNA structure indicates structural anomaly and perturbation. Thus, increases in the C2'-endo puckering are associated with strong tertiary interactions, as in the LPTLs at positions 326:330 (Ec: 319:323), 335:339 (Ec: 328:332), 1651:1655 (Ec: 1565:1568), and 2482:2486 (Ec: 2447:2451) in 23S rRNA.

The 33332/bbmmm pattern is in sharp contrast to the 333333/bbmmmb pattern that is observed in typical GNRA tetraloops24,26-27 which contain no C2'-endo puckering. In addition, while the 3' side base of the triloop-closing lonepair in a lonepair triloop faces into the minor groove, that of the tetraloop-closing basepair in a tetraloop stays at the boundary of the major and minor grooves. Moreover, the 2'-OH of the 3' side nucleotide (L) of the lonepair in a lonepair triloop is parallel with the helical axis, while the corresponding group of the basepair closing a tetraloop is perpendicular to the axis.

When an RNA chain folds back to form a hairpin loop, it reverses its direction so that the two local chains of the loop-closing basepair run antiparallel. Based on the relative orientations of the two O4' atoms of each lonepair,135 the directionality of the local chains involved in each LPTL is shown in Table 5.1 and, as expected, the majority of Class I and II LPTLs exhibit the antiparallel directionality. However, two LPTLs, 326:330 (Ec: 319:323) and 2482:2486 (Ec:
2447:2451) in 23S rRNA, have parallel directionality, because of the dramatic conformational distortions caused by tertiary interactions with these lonepair triloop motifs. All three Class III LPTLs, 64:68 (Ec: 64:68), 911:915 (Ec: 934:938), and 1035:1039 (Ec: 1053:1057) in 16S rRNA, also have parallel directionality (Table 5.1).

5.2.9 Lonepair types, exchanges, and conformations

Several types of lonepairs occur in the LPTL motifs. The most frequent lonepair in the rRNA and tRNA crystal structures analyzed here is U:A (ten occurrences), followed by C:G (3), G:G (3), G:C (2), G:A (2), C:A (2), U:G (1), and U:U (1) (Table 5.2). The majority of these lonepairs form non-canonical conformations (Tables 5.1 and 5.2). Seven types of lonepair conformations were observed in the current set of LPTLs: 11 reversed Hoogsteen (9 U:A rH, 1 C:A rH, and 1 G:G rH), four Watson-Crick (2 C:G WC, 1 G:C WC, and 1 U:A WC), three Hoogsteen (2 G:G H and 1 G:A H), three reversed wobble (1 U:G rWb, 1 C:A rWb, and 1 U:U rWb), one G:C rWC, one G:A S and one G:C fS* conformation. The most frequent U:A lonepair adopts the reversed Hoogsteen conformation in nine of the ten U:A examples; the tenth forms the Watson-Crick conformation. The distance between the two C1' atoms (dCC) in U:A rH lonepairs is 9.4-9.8 Å, which is 0.6-1.5 Å shorter than the 10.4-10.9 Å observed for U:A basepairs with the Watson-Crick conformation in A-RNA. In contrast,
three of the five G:C and C:G lonepairs adopt the standard Watson-Crick conformation with three hydrogen bonds. It is interesting to note that when a LPTL is in a strong contact with another part of the RNA chain or protein(s), the lonepair does not form the reversed conformation, as observed at positions 326:330 (Ec: 319:323), 482:486 (Ec: 476:480), 1651:1655 (Ec: 1565:1568), 1808:1812 (Ec: 1752:1756), and 2482:2486 (Ec: 2447:2451) in 23S rRNA and 22:26 (Ec: 24:27) in 5S rRNA. Thus, the conformation of the lonepair can be influenced by the presence or absence of tertiary interactions.

Of the 24 LPTLs present in the 16S, 23S and 5S rRNA and tRNA crystal structures studied here, 14 have a single lonepair type that is conserved in more than 80% of the nuclear-encoded rRNA genes in the three phylogenetic domains or Type I tRNA genes (Table 5.1): Seven have a U:A lonepair (318:322 (Ec: 323:327) and 1296:1300 (Ec: 1315:1319) in 16S rRNA, 335:339 (Ec: 328:332), 481:485 (Ec: 475:479), 505:509 (Ec: 499:503), and 1388:1392 (Ec: 1282:1286) in 23S rRNA and 54:58 in tRNA); two have a G:A lonepair (482:486 (Ec: 476:480) and 2482:2486 (Ec: 2447:2451) in 23S rRNA); two have a G:G lonepair (1035:1039 (Ec: 1053:1057) and 1158:1162 (Ec: 1177:1181) in 16S rRNA); and three have single examples of C:A (911:915 (Ec: 934:938) in 16S rRNA), C:G (1651:1655 (Ec: 1565:1568) in 23S rRNA), and U:U (933:937 (Ec: 956:960) in 16S rRNA) lonepairs. The remaining ten have more sequence variation at the two

5.2.10 Conservation of ribosomal LPTLs

All 23 of the lonepair triloops in the *T. thermophilus* 16S and *H. marismortui* 23S and 5S rRNAs occur in regions that are present in the nuclear-encoded rRNA sequences in the three primary phylogenetic branches, the archaea, bacteria and eucarya. The five positions in these LPTLs are present in nearly all of these rRNA sequences. In addition, the majority of the LPTLs in the rRNAs occur at positions with nucleotides that are conserved in at least 90% of
these sequences (Table 5.3). In contrast, only nine of the 23 LPTLs occur in regions of the rRNAs that are present in all of the mitochondrial-encoded rRNAs (Appendix K).61

However, some of the LPTLs have less nucleotides in the three nucleotide hairpin loop, while others have more than one basepair and/or more than three nucleotides in the triloop. The 131:148 LPTL in 23S rRNA is extremely variable in size, ranging from zero nucleotides to more than 50 basepairs, as noted earlier. While the basepair in the 22:26 (Ec: 24:27) LPTL in the 5S rRNA occurs in all nuclear-encoded archaea, bacteria and eucarya, the hairpin loop varies in size. Only a few of these sequences, including H. marismortui, have three nucleotides in this loop; the remainder have only two. The lonepair in LPTL 1651:1655 (Ec: 1565:1568) in the 23S rRNA also occurs in all nuclear-encoded archaea, bacteria and eucarya. The hairpin loop for this LPTL is a triloop in all of the archaea sequences and a diloop in nearly all of the bacteria and eucarya sequences. Last, LPTL 1388:1392 (Ec: 1282:1286) in the 23S rRNA has a few insertions and deletions in the eucarya. For example, the Microsporidia have deleted this region of the rRNA, while other lower eucaryotes such as Euglenozoa have extended this helix beyond the LPTL motif.
5.2.11 Distribution of lonepair triloops on the rRNA three-dimensional structures

While the rRNA LPTLs are distributed across the covariation-based rRNA secondary structure diagrams, our analysis of the *T. thermophilus* 30S and *H. marismortui* 50S crystal structures reveals that many of these rRNA LPTLs are clustered at functionally or structurally important regions in the three-dimensional space, including (1) 11 of the 22 16S and 23S rRNA LPTLs at the interface of the 30S and 50S subunits (Fig. 5.1), (2) three 16S rRNA LPTLs (911:915 (Ec: 934:937), 933:937 (Ec: 956:960) and 1035:1039 (Ec: 1053:1057)) in proximity to the substrate binding sites in the 30S subunit, (3) when viewed from the interface side, seven 23S rRNA LPTLs (313:317 (Ec: 306:310), 326:330 (Ec: 319:323), 335:339 (Ec: 328:332), 481:485 (Ec: 475:479), 482:486 (Ec: 476:480), 505:509 (Ec: 499:503), and 1388:1392 (Ec: 1282:1286)) that are clustered at the lower back of the 23S rRNA and are directly engaged in the formation of the polypeptide exit tunnel (PET), and (4) three 23S rRNA LPTLs (624:628 (Ec: 567:571), 1966:1970 (Ec: 1925:1929) and 2482:2486 (Ec: 2447:2451) close to the substrate binding sites in the 50S subunit. The positioning of the 2482:2486 (Ec: 2447:2451) LPTL at the putative site of peptidyl transferase activity, A2486 (Ec: A2451), is intriguing and suggests that this LPTL is intimately associated with protein synthesis.
The seven LPTLs at the lower back of the large ribosomal subunit show this propensity to cluster with one another. The 481:485 (Ec: 475:479) and 482:486 (Ec: 476:480) LPTLs in 23S rRNA intimately interact with each other and share two triloop nucleotides, C483 (Ec: C477) and A484 (Ec: A478). As discussed earlier, U481:A485 (Ec: U475:A479) is stacked between two basepairs (G482:A486 (Ec: G476:A480) and G487:C515 (Ec: G481:C509), creating a short helix composed of three basepairs and thereby blocking the coaxial stacking of the lonepair U481:A485 (Ec: U475:A479) onto the 5'-helix, 28-29/479-480 (Ec: 31-32/473-474). Because of the overlapping nucleotides between these two lonepair triloops, LPTL 482:486 (Ec: 476:480) has the U-turn between F and X rather than X and Y. The lonepair G482:A486 (Ec: G476:A480) adopts the sheared conformation with the distance between the two C1' atoms in the lonepair (dCC) equal to 9.39 Å, which is shorter than the distances in the imino and Hoogsteen counterparts (Table 5.1). This LPTL-LPTL cluster undergoes an extensive lateral association with the third LPTL 505:509 (Ec: 499:503) in the immediate vicinity, forming a cross-LPTL A-stack of the four consecutive base-base interactions A485:A509 (Ec: A479:A503), A486:A511 (Ec: A480:A505), A484:A508 (Ec: A478:A502) and C483:A507 (Ec: C477:A501), among which the last two do not form actual basepairs but “pseudopairs.” This cluster of three LPTLs is in close contact with the two more upstream LPTLs (313:317 (Ec:
306:310) and 335:339 (Ec: 328:332)) through the base stacking interactions of A507 (Ec: A501) onto G315 (Ec: G308) and C483 (Ec: C477) onto G336 (Ec: G329), respectively. Moreover, LPTL 1388:1392 (Ec: 1282:1286) is also in proximity to LPTL 505:509 (Ec: 499:503) through the base stacking of A1390 (Ec: A1284) onto A495 (Ec: A489). Consequently, seven of the 15 23S rRNA LPTLs, including the neighboring LPTL 326:330 (Ec: 319:323), are clustered in three-dimensional space.

5.2.12 Solvent accessibility of lonepair triloops

The solvent accessible surface areas (SASAs) were calculated using the GETAREA program to determine if rRNA LPTLs are exposed to the surface or are buried in the ribosome (Table 5.5). The SASAs for all nucleotides in the rRNAs were obtained in the presence or absence of the ribosomal proteins, and the SASAs for the nucleotides involved in lonepair triloops are available at the CRW-LPTL. In the absence of proteins, all of the lonepair triloops in the rRNAs are solvent accessible, except for the 624:628 (Ec: 567:571) and 2482:2486 (Ec: 2447:2451) LPTLs in 23S rRNA; while the entire 2482:2486 (Ec: 2447:2451) LPTL is not accessible to solvent, nucleotide A2486 (Ec: A2451), the putative site of the peptidyl transferase activity, is accessible to solvent. In contrast to the solvent accessibility of the RNA by itself, nine lonepair triloops (318:322 (Ec: 323:327), 1035:1039 (Ec: 1053:1057), and 1296:1300 (Ec: 1315:1319) in 16S...
Table 5.5 Solvent accessible surface areas for LPTL nucleotides
rRNA and 326:330 (Ec: 319:323), 335:339 (Ec: 328:332), 481:485 (Ec: 475:479), 482:486 (Ec: 476:480), 1651:1655 (Ec: 1565:1568) and 2597:2601 (Ec: 2562:2566) in 23S rRNA) are not accessible to solvent due to the presence of the ribosomal proteins. Three LPTLs (64:68 (Ec: 64:68) in 16S rRNA and 1966:1970 (Ec: 1925:1929) and 2482:2486 (Ec: 2447:2451) in 23S rRNA) have the same SASA values for each nucleotide in the presence or absence of proteins, suggesting that they have no interactions with proteins. The difference in SASA values in the presence and absence of proteins is proportional to the strength of RNA-protein interaction; the larger the difference, the stronger the interaction.

5.3 DISCUSSION

5.3.1 U-turns and the structural integrity of a lonepair triloop

The U-turn is a common and fundamental RNA structural motif that is characterized by a sharp turn in the RNA chain.22-23 Thirteen of the 21 Class I and II LPTLs studied here have a U-turn between the first (X) and second (Y) nucleotides in the triloop. Specifically, the U-turn is defined as a sharp change in the torsion angle (O3'-P-O5'-C4) which ranges from -60 to -70° between F (the 5' position in the lonepair) and X and 160 to 170° between X and Y, leading to the net change of 220-230° in the direction of the RNA backbone. Due to this sharp directional change, the local chains involved in a lonepair are antiparallel with
one another while the phosphodiester backbone forms a S-shape, when viewed from the top of a LPTL.

Within the U-turn, the sugar rings (and their 2'-OH groups) flip over to invert the direction of the backbone. Accordingly, while the base of X is retreated to the boundary of the major and minor grooves and located right under the phosphate group of Z, the two bases of Y and Z are forced to face into the minor groove, giving rise to the characteristic 33332/bbmmm pattern in sugar puckering and bases facing. The retreated base of X is directly stacked onto the base of F, and the base of Y is stacked onto the base of Z. This base stacking continues onto the base of Z either directly with no intervening bases (R3 LPTLs) or indirectly with intervening bases (R1 and R2 LPTLs). Thus, the U-turn is an important component in maintaining the architectural integrity of a LPTL.

contrast, when a lonepair triloop has an A at position X (326:330 (Ec: 319:323) and 2482:2486 (Ec: 2447:2451) in 23S rRNA), a pyrimidine (C or U) at position Z (335:339 (Ec: 328:332) and 1651:1655 (Ec: 1565:1568) in 23S rRNA), or a G at position F (326:330 (Ec: 319:323) and 2482:2486 (Ec: 2447:2451) in 23S rRNA and 22:26 (Ec: 24:27) in 5S rRNA), the lonepair triloop is highly destabilized, is involved in strong tertiary interactions with other parts of the RNA chain or protein(s) and does not form the U-turn between positions X and Y, thus perturbing the structural integrity of the LPTL.

5.3.2 RNA packing and the structural integrity of lonepair triloops

Previously, it was stated that the lonepair of the lonepair triloop motif is not stable by itself and thus needs to be associated with other structural elements to maintain its structure and conformation. Given that the seven original LPTLs predicted with covariation analysis were all 3' to an existing helix, we proposed that the lonepair is stacked onto this 5'-helix. Our analysis of the 30S and 50S ribosomal crystal structures substantiated these seven LPTLs while identifying 16 more LPTLs that were also stacked onto the 5'-helix, with one exception. A compound helix is formed from the coaxial stacking of a lonepair onto the direct (Class I) or nearest (Class II) 5'-helix. The stacking of the lonepair flips the unpaired bases between the 3' side of the lonepair and the 5'-helix out of the compound helix. In four of the LPTLs, including two that occur in four-way
junctions (1186:1190 (Ec: 1082:1086) and 2597:2601 (Ec: 2562:2566) in 23S rRNA) and two that occur in multi-stem loops (125:129 (Ec: 131:148) and 1966:1970 (Ec: 1925:1929) in 23S rRNA), the stacking of the lonepair induces the coaxial stacking of two neighboring or remote helices (Fig. 5.3). Moreover, the lonepair stacking in the 23S rRNA LPTL at positions 2482:2486 (Ec: 2447:2451), along with the tertiary interaction of its triloop, gives rise to the coaxial stacking of the two flanking helices, 2105-2111/2476-2481 (Ec: 2064-2070/2441-2446) and 2487-2492/2529-2535 (Ec: 2452-2457/2494-2500). Therefore, the stacking of the lonepair in this motif contributes to RNA packing. While we currently understand neither the full biological significance of these coaxial stackings in the lonepair triloops nor the coaxial stackings for other sets of helices in tRNA and rRNAs, it is apparent that these compound helices are important for the three-dimensional arrangement and orientation of the helical elements in RNA molecules.

In addition to the lonepair stacking, nearly all of the Class I and II LPTLs are associated with tertiary interactions to other regions of the RNA. The triloops in R1 and R2 LPTLs recruit an unpaired tertiary base from another section of the RNA chain into their triloops to mimic the tetraloop sequence and structure, while nucleotides in some R3 and R4 LPTLs form base triples with G:C basepairs in a helix. These LPTL-RNA tertiary interactions also contribute to RNA packing.
5.3.3 Conformations of lonepairs

Comparative analysis identifies the secondary and tertiary structure that is conserved in a set of equivalent sequences. Given the success of this method, we now wonder if the same comparative paradigm can be used to predict the conformation of a basepair that is conserved for a set of different basepair types. For one lonepair, we can compare the conformations from four different crystal structures. The lonepair at positions 1186:1190 (Ec: 1082:1086) in 23S rRNA covaries between U:A (52.4%) and C:G (46.8%) in the nuclear-encoded rRNA genes from the three phylogenetic domains. The C:G basepair in the H. marismortui 50S crystal structure \(^{45,64}\) adopts the reversed Watson-Crick conformation, while the identical conformation is formed by U:A lonepairs in each of the two crystal structures for the L11-binding domains from E. coli\(^{118}\) and T. maritima.\(^{119}\) In addition, every lonepair in the H. marismortui 23S rRNA LPTLs has the same or similar basepair conformation with the homologous lonepair in the crystal structure for the D. radiodurans 23S rRNA,\(^{67}\) although the basepair types between the two crystal structures can be different. This suggests that the conformation for each homologous basepair is conserved between different organisms to maintain the same or similar three-dimensional structure that is necessary to perform the same biological function. Thus, the basepair
conformation may be inferred at least for some examples from the basepair exchange information obtained with comparative sequence analysis.\textsuperscript{137}

5.3.4 Molecular mimicry

Eight of the nine R1 and R2 LPTLs recruit a tertiary base from a different section of RNA chain and position it between Z and \( L \) to mediate the stacking of three consecutive bases (\( Y, Z, \) and \( L \)). The consensus triloop sequence in the six R1 LPTLs that mimic the tetraloop conformation is GNR (Tables 5.1 and 5.4). The recruited base in five of the R1’s is a highly conserved A (Table 5.3), forming the most common of the tetraloop sequences, GNRA. The sixth R1 recruits a conserved U, forming a GAGU sequence at positions 1388:1392 (\( Ec \): 1282:1286) in the 23S rRNA; the conformation of this loop is the same as that found in GNRA tetraloops. In contrast, the consensus of the triloop sequences in the two R2 LPTLs is UYR (Tables 5.3 and 5.4). A conserved A is recruited for both of these R2 LPTLs, forming UYRA loop sequences. These latter two examples have the same conformation as GNRA tetraloops. As a result, their overall three-dimensional conformation usually mimics the stable tetraloop conformation.\textsuperscript{24-28}

To the best of our knowledge, this is the first example of a small RNA motif that is composed of consecutive nucleotides that can form similar sequence and structural conformations from nucleotides that are disconnected on the RNA sequence. This example of tetraloop mimicry is analogous to examples involving
both larger regions of RNA and RNA/protein complexes. The 3' ends of plant viral mRNAs do not form the traditional tRNA cloverleaf secondary structure; they form a different tRNA secondary structure that folds into a tRNA-like three-dimensional structure that is aminoacylated. Some proteins mimic the overall three-dimensional structures of nucleic acids. The protein complex of EF-G:GDP resembles the conformation of the ternary complex of Phe-tRNA:EF-Tu:GDPNP in translation. Interestingly, the phenomenon of molecular mimicry has been implicated in both immune-mediated diseases and viral pathogenesis.

5.3.5 Functional implications of lonepair triloop motifs

The presence of many lonepair triloop motifs in highly conserved regions of 16S and 23S rRNAs suggests that these motifs are functionally important. Consistent with this premise, three of the 16S rRNA LPTLs (911:915 (Ec: 934:937), 933:937 (Ec: 956:960), and 1035:1039 (Ec: 1053:1057)) are within 7 Å of the binding sites of A-, P-, and E-tRNAs and mRNA (APEM sites) in the 30S crystal structure (PDB 1JGO). Of these, the lonepair triloop at positions 1035:1039 (Ec: 1053:1057) occurs in the S3-binding domain and at one of the primary binding sites of tetracycline. The first nucleotide in the triloop, C1036 (Ec: C1054), hangs down from the upper part of the mRNA entrance tunnel (MET). These two observations suggest that this lonepair triloop might be engaged in substrate binding and mRNA movement.
The three 23S rRNA LPTLs at positions 624:628 (Ec: 567:571), 1966:1970 (Ec: 1925:1929), and 2482:2486 (Ec: 2447:2451) occur within 5 Å of the bound substrates in the 50S and 70S crystal structures (PDB 1KQS\textsuperscript{143} and 1GIY,\textsuperscript{68} respectively). The LPTL at 1966:1970 (Ec: 1925:1929), which is involved in the formation of the peptide exit tunnel (PET) and in close proximity to the tRNA bound to the P-site,\textsuperscript{134} is in direct contact with the 773 loop (Ec: 790 loop) in 16S rRNA to form the B2b intersubunit bridge.\textsuperscript{68} Most strikingly, however, the 2482:2486 (Ec: 2447:2451) LPTL is at the site of peptidyl transfer during protein synthesis and at the binding site of chloramphenicol.\textsuperscript{134} Consistent with its phylogenetic invariance in all known sequences, mutations at A2486 (Ec: A2451) had a deleterious effect on cell viability.\textsuperscript{144} In conjunction with the stacking effects discussed above, this suggests that the 2482:2486 (Ec: 2447:2451) LPTL specifically organizes the local structures to create the catalytic site of the transferase activity. Although the 624:528 (Ec: 567:571) LPTL is physically very close to the 2482:2486 (Ec: 2447:2451) LPTL, no experimental evidence currently associates it to ribosomal function.

A number of other LPTLs have potential roles in ribosome function. The 1808:1812 (Ec: 1752:1756) LPTL is near a region which forms another intersubunit bridge.\textsuperscript{68} The LPTL at position 2597:2601 (Ec: 2562:2566) in 23S rRNA is near the A-loop; position G2588 (Ec: G2553) is basepaired to C75 in the
A-site bound tRNA in the *H. marismortui* 50S crystal structure (PDB 1FFK, 1FFZ, and 1FG0). The 1186:1190 (*Ec*: 1082:1086) LPTL occurs at the L11-binding region of 23S rRNA, although it is not in direct contact with L11. Seven of the 23S rRNA LPTLs, at positions 313:317 (*Ec*: 306:310), 326:330 (*Ec*: 319:323), 335:339 (*Ec*: 328:332), 481:485 (*Ec*: 475:479), 482:486 (*Ec*: 476:480), 505:509 (*Ec*: 499:503), and 1388:1392 (*Ec*: 1282:1286), are clustered in three-dimensional space and near the polypeptide exit tunnel (PET) which the growing polypeptide chain travels through to exit the ribosome during protein synthesis. Of these clustered 23S rRNA LPTLs, the first five occur in the L4- or L24-binding domain. Finally, LPTL 54:58 in the T loop of tRNAs may play an essential role in folding tRNAs into the characteristic L-shaped functional form by making a tertiary contact with the D loop. Together, the lonepair triloop motifs may be essential for structural organization and ribosomal function.

5.4 **Materials and Method**

5.4.1 **Structural analysis of rRNAs**

The structural details of the lonepair triloop motifs were identified from a visual inspection of the high-resolution crystal structures of the 16S rRNA in the *T. thermophilus* 30S (PDB 1FJF) subunit, the 23S and 5S rRNAs in the *H. marismortui* 50S (PDB 1FFK and 1JJ2) subunit, and the *S. cerevisiae* Phe-
tRNA (PDB 6TNA) using the interactive RasMol program. In addition to the *S. cerevisiae* Phe-tRNA crystal structure, other tRNA crystal structures were investigated: *S. cerevisiae* Asp-tRNA (PDB 1ASY), *E. coli* fMet-tRNA (PDB 2FMT), *E. coli* Thr-tRNA (PDB 1QF6), *S. cerevisiae* Arg-tRNA (PDB 1F7U), *E. coli* Asp-tRNA (PDB 1IL2), and *T. thermophilus* Glu-tRNA (PDB 1G59) are complexed with tRNA synthetases, while *S. cerevisiae* Phe-tRNA (PDB 6TNA), *E. coli* Cys-tRNA (PDB 1B23), and *E. coli* Glu-tRNA (PDB 1EXD) are uncomplexed.

Additional information, including RasMol scripts and nucleotide frequency tables for lonepairs and triloops, is available from the CRW-LPTL Page (http://www.rna.icmb.utexas.edu/ANALYSIS/LPTL/). Single nucleotide frequencies are also available online at the main CRW Site (http://www.rna.icmb.utexas.edu/). Each LPTL can be displayed interactively with RasMol or Protein Explorer. In addition, the three-dimensional distribution of the lonepair triloop motifs in the entire rRNA structure can be viewed with the RasMol scripts available at the CRW-LPTL.

### 5.4.2 Coaxial stacking

The coaxial stacking of two helical stems in nucleic acids can be evaluated with *Curves* analysis. In general, the cutoff values for stacking angle and helix displacement are 45° and 5 Å, respectively. However, because one of the
two stacking helices in lonepair stacking contains only a single basepair, Curves analysis could not be applied to the evaluation of the coaxial stacking of a lonepair onto its 5'-helix. Therefore, the lonepair stacking was visually judged by highlighting the lonepair stacked onto 5'-helix using the RasMol program.\textsuperscript{104-105}

5.4.3 Solvent accessibility of LPTLs

The solvent accessibility of LPTLs was evaluated based on the solvent accessible surface areas (SASAs) obtained using a probe radius of 1.4 Å, employing the GETAREA program.\textsuperscript{136} The “standard” SASA (in Å\textsuperscript{2}) of a nucleotide X (A, C, G, and U) was defined as the average SASA for X in a set of model trinucleotide double-stranded RNAs (dsRNAs), one strand of which limits X in the middle (N-X-N, where N is any nucleotide), assuming that the middle X is fully solvent accessible. Depending on the nucleotide of interest, 21-27 model trinucleotide dsRNAs were pooled out of the crystal structures of the \textit{T. thermophilus} 30S (PDB 1FJF\textsuperscript{63}) and \textit{H. marismortui} 50S (PDB 1JJ2\textsuperscript{45}) subunits and then applied to SASA calculations. The calculated “standard” SASA values were 172.56 Å\textsuperscript{2} for A, 169.49 Å\textsuperscript{2} for C, 175.74 Å\textsuperscript{2} for G, and 169.31 Å\textsuperscript{2} for U. The SASA for each nucleotide in rRNA was calculated with and without the ribosomal proteins in the \textit{T. thermophilus} 30S and \textit{H. marismortui} 50S rRNA crystal structures: all other bound substrates such as mRNA and metal ions were not included in the calculation. A nucleotide was considered to be solvent
accessible or exposed to the surface (external) when its SASA is more than 50% of the standard (86.00 Å² for A, 84.50 Å² for C, 87.87 Å² for G, and 84.66 Å² for U); otherwise, it is considered buried (internal). The average of the standard SASAs for four nucleotides is 85.76 Å². Thus, a lonepair triloop was considered to be solvent accessible when the average of the SASA values for its nucleotides was more than 85.76 Å².
Chapter 6: Conclusion and Perspective

6.1 Conclusion

The cross-analysis of the comparative structure models and the three-dimensional structures for the rRNAs in the recently determined high-resolution crystal structures of the *T. thermophilus* 30S and *H. marismortui* 50S ribosomal subunits have revealed a tremendous amount of information regarding RNA structure. First, the accuracy of the 16S, 23S, and 5S rRNA comparative structure models was determined from the mapping of the basepairs in the rRNAs from the two ribosomal crystal structures. Approximately 97-98% of the basepairs predicted with covariation analysis of the 16S and 23S rRNAs are present in the crystal structures. This high accuracy score not only validated the authenticity of covariation-based rRNA structure models, but also proved that covariation analysis, a subset of comparative sequence analysis, is a powerful and reliable method for predicting RNA secondary and tertiary structures.

In addition to the canonical basepair types with the canonical basepair conformations, consisting of the standard Watson-Crick (C:G, G:C, U:A, and A:U) and wobble (U:G and G:U) basepairs, the analysis of the chemical structures in the mapped basepairs revealed a wide variety of canonical and non-canonical basepair types with the non-canonical basepair conformations. While the
basepairs occurring within regular helices have canonical conformations, the basepairs occurring at helical termini or in the unpaired regions in the comparative models usually have non-canonical conformations. These diverse basepair conformations observed in the rRNAs from the two ribosomal crystal structures prompted us to investigate the detailed topological relationships of two bases and their glycosidic bonds in a given basepair, leading to a new classification system that systematically and unambiguously describes all the basepair conformations, regardless of the number of hydrogen bonds involved in the basepair. In the new system, all 16 possible basepair types are divided into ten basepair groups (i.e., C:G, U:A, U:G, G:A, C:A, U:C, A:A, C:C, G:G, and U:U) and adopt one of the 14 major basepair conformations: Watson-Crick (WC), wobble (Wb), slipped Watson-Crick (sWC), slipped wobble (sWb), reversed Watson-Crick (rWC), reversed wobble (rWb), Hoogsteen (H), reversed Hoogsteen (rH), sheared (S), reversed sheared (rS), flipped sheared (fS), parallel flipped sheared (pfS), parallel sheared (pS), and reversed parallel sheared (rpS). In addition to the “conventional” basepairs involving two or three hydrogen bonds, the new system can describe a much wider variety of basepair conformations than the existing naming systems, including conformations with a single hydrogen bond and conformations undergoing keto-enol or amino-imino tautomerism. Moreover, higher-order interactions such as base triples and
quadruples can be easily described based on the new system. For example, many of the unpaired A’s in comparative structure models are involved in A-mediated tertiary interactions in the crystal structures; interactions with the N3 position form G:A rS (previously known as type I A-minor motif\textsuperscript{71-72}) or G:A rpS, while interaction with the N1 position form G:A fS or G:A pfS. This new classification system will provide the scientific community with a common method to describe and analyze basepair conformations.

As well, we have utilized the high-resolution ribosomal crystal structures to evaluate and validate the sequence motifs previously identified with comparative analysis as well as for identifying new RNA motifs. For example, the AA and AG oppositions frequently occur immediately outside of regular helices (the G and A for AG oppositions are immediately 3’ and 5’ to helix, respectively) and are highly conserved in rRNA sequences; they have been called as the AA.AG@helix.ends motif.\textsuperscript{38} The basepair mapping revealed that approximately 77% these oppositions are basepaired and 80% of and AA.AG@helix.ends basepairs adopt the sheared conformations (G:A S and A:A S). These A:A and A:G basepairs in the sheared conformations have shorter dCC’s (8.5–10.0 Å), compared to this distance in the standard Watson-Crick basepairs in the A-form helix (10.5 Å). The AA.AG@helix.ends basepairs may stabilize the helical ends and/or may interact with proteins or other regions of the
Interestingly, the G:A basepairs immediately flanking regular helices that have their G 5’ and A 3’ to helix are usually stacked onto another helix and always adopt the Watson-Crick conformation with the longer dCC (about 12.6 Å).

An analysis of the comparative structure models and the crystal structures for the rRNAs in the two ribosomal crystal structures validated another RNA sequence motif, the AGUA/GAA motif (previously known as the E loop motif\textsuperscript{31-36}), as well as identified its two new relatives, GUA/GAA and GGA/GAA motifs.\textsuperscript{73} In particular, the AGUA/GAA and GUA/GAA motifs are found in the E loop of 5S rRNAs; the former in Archaea and Eukaryotes, while the latter in Bacteria. These three motifs frequently occur in the highly conserved regions of 16S and 23S rRNAs in all three phylogenetic domains, and they are stabilized by the network of hydrogen bonds and base stacks within themselves, with the characteristic AAA or A(U:A)A stack in their minor groove. Most importantly, all three motifs are ‘sticky’ and tend to interact with other section of RNA, proteins, and/or metal ions, thereby forming a superfamily of RNA motifs called the sticky motif. Most of the sticky motifs formed in the crystal structures have three consecutive basepairs: A:A rS, G(U:A) base triple with U:A rH, and A:G S for the AGUA/GAA motif; G:A S, U:A rH, and A:G S for the GUA/GAA motif; G:A S, G:A S, and A:G S for the GGA/GAA motif. Beyond the 33 sticky motifs identified in the rRNAs from the ribosomal crystal structures, analysis of other
comparative structure models of rRNAs identified 34 more sticky motifs, for a total of 67 sticky motifs. Some of the positions in the rRNAs with the 67 AGUA/GAA, GUA/GAA, and GGA/GAA motifs exchange with other forms of the sticky motif, rationalizing their relatedness. Interestingly, the GGA/GAA and GUA/GAA motifs formed in the asymmetric internal loops are closely related to the K-turn motif; in fact, the K-turn motif is a subfamily of the GGA/GAA motif. Moreover, the highly conserved sticky motifs are present at or near the intersubunit bridges for ribosomal association. Furthermore, small catalytic RNAs, including the hairpin ribozyme and the hammerhead ribozyme have the sticky motif near their catalytic sites. Thus, the sticky motif has several structural and functional implications: (1) stabilizes internal loops or helical ends in multistem loops, (2) mediates helical stacking directly in internal loops or indirectly in multistem loops, leading to RNA compaction, (3) participates in RNA folding by serving molecular magnets to attract other section(s) of RNA, (4) may be recognized by proteins and metal ions, and (5) may modulate the catalytic activity of small RNA molecules.

In addition to the sequence motifs that are primarily implicated by their nucleotide sequences, RNA structure motifs can be also identified from the analysis of three-dimensional molecular structures. As an example, the lonepair triloop (LPTL) motif is a single independent basepair or lonepair capped with a
three-nucleotide hairpin loop; the two nucleotides immediately outside of this motif are not basepaired. Beyond the seven lonepair triloops predicted with covariation analysis of 16S and 23S rRNA sequences, the detailed analysis of the lonepair triloops in the 16S, 23S, and 5S rRNAs in the two ribosomal and tRNA crystal structures revealed 17 additional lonepair triloops. This analysis also revealed their classification, structural characteristics, and molecular interactions. These motifs are usually positioned directly 3' to a regular helix and are stabilized by coaxial stacking onto this flanking helix. They are divided into three classes (I, II, and III) based on the secondary structure context as well as into five groups (R1-R5) based on which nucleotide(s) in the triloop is involved in a tertiary interaction. The R1 and R2 LPTLs have the U:A (or C:A) lonepair with the reversed Hoogsteen conformation, contain a U-turn between the first and second positions in the triloop, have their first and/or second triloop nucleotides in a tertiary interaction, and with the sugar puckering/bases facing pattern of 33332/bbmmm. The R1 LPTLs that contain the consensus GNR triloop sequence recruit a tertiary base (mostly unpaired A) between the third base in the triloop and the 3' base of their lonepair, mimicking the sequence and structure of the GNRA tetraloop motif. R2 LPTLs contain the consensus UYR triloop sequence and recruit a tertiary base (unpaired A or U) between the third base in the triloop and the 3' base of their lonepair; their resulting new UYR'A' tetraloop
also resembles the structure of the GNRA tetraloop. R3 LPTLs with the UAA consensus sequence in the triloop are involved in the formation of base triples; the third nucleotide (A) forms an A-mediated tertiary interaction with the G in a C:G basepair in another region of RNA. The majority of the lonepair triloops in the rRNAs occur in the highly conserved regions in all three phylogenetic domains, and many of them are clustered at functionally or structurally important regions, including the three at the substrate-binding sites in the 30S subunit and the one at the proposed protein synthesis site in the 50S subunit. Moreover, the lonepair triloops in multistem loops frequently induce the coaxial stacking of neighboring helices, playing a role in RNA compaction. Besides, more than half of the rRNA LPTLs interacts with ribosomal proteins. Furthermore, the lonepair triloop in the T loop of tRNAs may play an important role in folding tRNAs into the functional L-shaped form. Thus, the lonepair triloop motif in RNA structure may play an essential role in organizing RNA structures as well as in RNA folding.

6.2 PERSPECTIVE

This study has established a set of structural principles of RNA by systematically classifying basepair conformations and analyzing the structural details of several structure motifs. Such obtained knowledge will not only help predict more basepairs in covariation-based structure models through motif analysis, but also contribute to our understanding of RNA structure including
RNA folding and compaction, RNA-RNA interaction, and RNA-protein interaction. Despite our efforts to identify new RNA motifs, there are still many basepairs with the non-canonical conformations, whose functional and biological roles are still not fully understood. Those basepairs with the non-canonical conformations may be associated with new RNA structure motifs that could be found through more detailed and exhaustive structural analysis of the 16S and 23S rRNAs. In this respect, the identification of new motifs in RNA structure is essential to establish the principles of RNA structure, which eventually makes it possible to achieve the ultimate goal, to accurately predict the secondary and tertiary structure from a single RNA sequence.
## Appendix A

List of Basepairs in the *T. thermophilus* 16S rRNA Comparative Structure Model and in the Crystal Structure of the 30S Ribosomal Subunit

<table>
<thead>
<tr>
<th>E. coli</th>
<th>T. therm</th>
<th>X</th>
<th>CCG</th>
<th>bpT</th>
<th>bpC</th>
<th>RS</th>
<th>Motifs</th>
<th>RT</th>
<th>Rq</th>
<th>CA</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:25</td>
<td>9:25</td>
<td>w</td>
<td>G</td>
<td>C</td>
<td>WC</td>
<td>83</td>
<td>11:25</td>
<td>542</td>
<td>CA</td>
<td></td>
</tr>
<tr>
<td>10:24</td>
<td>10:24</td>
<td>w</td>
<td>AU</td>
<td>C</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
</tr>
<tr>
<td>11:23</td>
<td>11:23</td>
<td>w</td>
<td>G</td>
<td>C</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
</tr>
<tr>
<td>12:22</td>
<td>12:22</td>
<td>w</td>
<td>G</td>
<td>UG</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
</tr>
<tr>
<td>13:21</td>
<td>13:21</td>
<td>w</td>
<td>G</td>
<td>UG</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
</tr>
<tr>
<td>13:920</td>
<td>13:920</td>
<td>w</td>
<td>G</td>
<td>UG</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
</tr>
<tr>
<td>14:1398</td>
<td>14:1398</td>
<td>w</td>
<td>G</td>
<td>UG</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
</tr>
<tr>
<td>15:920</td>
<td>15:920</td>
<td>w</td>
<td>G</td>
<td>UG</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
</tr>
<tr>
<td>16:919</td>
<td>16:919</td>
<td>w</td>
<td>AA</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16:920</td>
<td>16:920</td>
<td>w</td>
<td>AA</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16:1080</td>
<td>16:1080</td>
<td>w</td>
<td>AA</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17:918</td>
<td>17:918</td>
<td>w</td>
<td>UA</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18:917</td>
<td>18:917</td>
<td>w</td>
<td>UA</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19:916</td>
<td>19:916</td>
<td>w</td>
<td>UA</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20:915</td>
<td>20:915</td>
<td>w</td>
<td>UA</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21:914</td>
<td>21:914</td>
<td>w</td>
<td>UA</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26:557</td>
<td>26:557</td>
<td>w</td>
<td>UA</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26:557</td>
<td>26:557</td>
<td>w</td>
<td>UA</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27:556</td>
<td>27:556</td>
<td>w</td>
<td>UA</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28:555</td>
<td>28:555</td>
<td>w</td>
<td>UA</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29:554</td>
<td>29:554</td>
<td>w</td>
<td>UA</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30:553</td>
<td>30:553</td>
<td>w</td>
<td>UA</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32:552</td>
<td>32:552</td>
<td>w</td>
<td>UA</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33:551</td>
<td>33:551</td>
<td>w</td>
<td>UA</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34:550</td>
<td>34:550</td>
<td>w</td>
<td>UA</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35:549</td>
<td>35:549</td>
<td>w</td>
<td>UA</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36:548</td>
<td>36:548</td>
<td>w</td>
<td>UA</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37:547</td>
<td>37:547</td>
<td>w</td>
<td>UA</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39:403</td>
<td>39:403</td>
<td>w</td>
<td>UA</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40:402</td>
<td>40:402</td>
<td>w</td>
<td>UA</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41:401</td>
<td>41:401</td>
<td>w</td>
<td>UA</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42:400</td>
<td>42:400</td>
<td>w</td>
<td>UA</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42:622</td>
<td>42:622</td>
<td>w</td>
<td>UA</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43:399</td>
<td>43:399</td>
<td>w</td>
<td>UA</td>
<td>WC</td>
<td>83</td>
<td>11:23</td>
<td>544</td>
<td>CA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44:398</td>
<td>44:393</td>
<td>[GC]</td>
<td>WC</td>
<td>64</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45:396</td>
<td>45:391</td>
<td>[UG]</td>
<td>Wb</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46:395</td>
<td>46:390</td>
<td>[GC]</td>
<td>WC</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47:361</td>
<td>47:356</td>
<td>[GC]</td>
<td>WC</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50:366</td>
<td>50:355</td>
<td>[AR]</td>
<td></td>
<td>61</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52:359</td>
<td>52:354</td>
<td>[GC]</td>
<td>WC</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>53:358</td>
<td>53:353</td>
<td>[GC]</td>
<td>WC</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>54:357</td>
<td>54:352</td>
<td>[CG]</td>
<td>WC</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>56:356</td>
<td>56:351</td>
<td>[UA]</td>
<td>WC</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>57:355</td>
<td>57:350</td>
<td>[GC]</td>
<td>WC</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>58:354</td>
<td>58:349</td>
<td>[AC]</td>
<td>CG</td>
<td>WC</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>59:353</td>
<td>59:348</td>
<td>[AG]</td>
<td>iS</td>
<td>61</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60:107</td>
<td>60:102</td>
<td>[AG]</td>
<td>iS</td>
<td>61</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>61:106</td>
<td>61:101</td>
<td>[GC]</td>
<td>WC</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>62:105</td>
<td>62:100</td>
<td>[UG]</td>
<td>Wb</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>63:104</td>
<td>63:103</td>
<td>[GC]</td>
<td>WC</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>64:382</td>
<td>64:377</td>
<td>[GC]</td>
<td>WC</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>67:102</td>
<td>67:101</td>
<td>[CG]</td>
<td>WC</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>67:171</td>
<td>67:166</td>
<td>[CG]</td>
<td>WC</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>69:95</td>
<td>69:90</td>
<td>[GC]</td>
<td>WC</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>70:98</td>
<td>70:93</td>
<td>[GC]</td>
<td>WC</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>73:97</td>
<td>73:92</td>
<td>[ GC]</td>
<td>WC</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>74:96</td>
<td>74:91</td>
<td>[ GC]</td>
<td>WC</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75:95</td>
<td>75:90</td>
<td>[ GC]</td>
<td>WC</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>77:92</td>
<td>77:87</td>
<td>[ GC]</td>
<td>WC</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>79:90</td>
<td>79:85</td>
<td>[ CG]</td>
<td>WC</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80:89</td>
<td>80:84</td>
<td>[GU]</td>
<td>Wb</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>81:88</td>
<td>81:83</td>
<td>[GC]</td>
<td>WC</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>82:87</td>
<td>82:82</td>
<td>[UA]</td>
<td>S</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>101:152</td>
<td>101:147</td>
<td>[GC]</td>
<td>WC</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>102:151</td>
<td>102:146</td>
<td>[AG]</td>
<td>S</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>103:172</td>
<td>103:167</td>
<td>[GU]</td>
<td>Wb</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>105:379</td>
<td>105:374</td>
<td>[AG]</td>
<td>S</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>111:331</td>
<td>111:326</td>
<td>[GU]</td>
<td>Wb</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>112:315</td>
<td>112:310</td>
<td>[GC]</td>
<td>WC</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>113:314</td>
<td>113:309</td>
<td>[GC]</td>
<td>WC</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>114:313</td>
<td>114:308</td>
<td>[UA]</td>
<td>WC</td>
<td>63</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Position</td>
<td>Nucleotide</td>
<td>Type</td>
<td>Description</td>
<td>Sequence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>------------</td>
<td>------</td>
<td>-------------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>115:312</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>118:288</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>119:287</td>
<td>AU</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>122:239</td>
<td>GU</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>123:238</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>124:237</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>125:236</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>126:235</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>126:605</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>127:234</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>128:233</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>129:232</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130:229</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>131:228</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>132:227</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>133:226</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>134:225</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>135:224</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>136:223</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>137:222</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>138:221</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>139:220</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>140:219</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>141:218</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>142:217</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>143:216</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>144:215</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>145:214</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>146:213</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>147:212</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>148:211</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>149:210</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>150:171</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>151:170</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>152:169</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>153:168</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>154:167</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>155:166</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>156:155</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>157:154</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>158:153</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>159:152</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>160:151</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>161:150</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>162:149</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>163:148</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>164:147</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>165:146</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>166:145</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>167:144</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>168:143</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>169:142</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>170:141</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>171:140</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>172:139</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>173:138</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>174:137</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>175:136</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>176:135</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>177:134</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>178:133</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>179:132</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>180:131</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>181:130</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>182:129</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>183:128</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>184:127</td>
<td>CA</td>
<td>HH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**CA**

**PT**
<table>
<thead>
<tr>
<th>185</th>
<th></th>
<th>186</th>
<th></th>
<th>187</th>
<th></th>
<th>188</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-NH-</td>
<td>178</td>
<td>-NH-</td>
<td>179</td>
<td>-NH-</td>
<td>180</td>
<td>-NH-</td>
<td>181</td>
</tr>
<tr>
<td></td>
<td>182</td>
<td></td>
<td>183</td>
<td></td>
<td>184</td>
<td></td>
<td>185</td>
</tr>
<tr>
<td></td>
<td>186</td>
<td></td>
<td>187</td>
<td></td>
<td>188</td>
<td></td>
<td>189</td>
</tr>
<tr>
<td></td>
<td>190</td>
<td></td>
<td>191</td>
<td></td>
<td>192</td>
<td></td>
<td>193</td>
</tr>
<tr>
<td></td>
<td>194</td>
<td></td>
<td>195</td>
<td></td>
<td>196</td>
<td></td>
<td>197</td>
</tr>
<tr>
<td></td>
<td>198</td>
<td></td>
<td>199</td>
<td></td>
<td>200</td>
<td></td>
<td>201</td>
</tr>
<tr>
<td></td>
<td>202</td>
<td></td>
<td>203</td>
<td></td>
<td>204</td>
<td></td>
<td>205</td>
</tr>
<tr>
<td></td>
<td>206</td>
<td></td>
<td>207</td>
<td></td>
<td>208</td>
<td></td>
<td>209</td>
</tr>
<tr>
<td></td>
<td>210</td>
<td></td>
<td>211</td>
<td></td>
<td>212</td>
<td></td>
<td>213</td>
</tr>
<tr>
<td></td>
<td>214</td>
<td></td>
<td>215</td>
<td></td>
<td>216</td>
<td></td>
<td>217</td>
</tr>
<tr>
<td></td>
<td>218</td>
<td></td>
<td>219</td>
<td></td>
<td>220</td>
<td></td>
<td>221</td>
</tr>
<tr>
<td></td>
<td>222</td>
<td></td>
<td>223</td>
<td></td>
<td>224</td>
<td></td>
<td>225</td>
</tr>
<tr>
<td></td>
<td>226</td>
<td></td>
<td>227</td>
<td></td>
<td>228</td>
<td></td>
<td>229</td>
</tr>
<tr>
<td></td>
<td>230</td>
<td></td>
<td>231</td>
<td></td>
<td>232</td>
<td></td>
<td>233</td>
</tr>
<tr>
<td></td>
<td>234</td>
<td></td>
<td>235</td>
<td></td>
<td>236</td>
<td></td>
<td>237</td>
</tr>
<tr>
<td></td>
<td>238</td>
<td></td>
<td>239</td>
<td></td>
<td>240</td>
<td></td>
<td>241</td>
</tr>
<tr>
<td></td>
<td>242</td>
<td></td>
<td>243</td>
<td></td>
<td>244</td>
<td></td>
<td>245</td>
</tr>
<tr>
<td></td>
<td>246</td>
<td></td>
<td>247</td>
<td></td>
<td>248</td>
<td></td>
<td>249</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td></td>
<td>251</td>
<td></td>
<td>252</td>
<td></td>
<td>253</td>
</tr>
<tr>
<td></td>
<td>254</td>
<td></td>
<td>255</td>
<td></td>
<td>256</td>
<td></td>
<td>257</td>
</tr>
<tr>
<td></td>
<td>258</td>
<td></td>
<td>259</td>
<td></td>
<td>260</td>
<td></td>
<td>261</td>
</tr>
<tr>
<td></td>
<td>262</td>
<td></td>
<td>263</td>
<td></td>
<td>264</td>
<td></td>
<td>265</td>
</tr>
<tr>
<td></td>
<td>266</td>
<td></td>
<td>267</td>
<td></td>
<td>268</td>
<td></td>
<td>269</td>
</tr>
<tr>
<td></td>
<td>270</td>
<td></td>
<td>271</td>
<td></td>
<td>272</td>
<td></td>
<td>273</td>
</tr>
<tr>
<td></td>
<td>274</td>
<td></td>
<td>275</td>
<td></td>
<td>276</td>
<td></td>
<td>277</td>
</tr>
<tr>
<td></td>
<td>278</td>
<td></td>
<td>279</td>
<td></td>
<td>280</td>
<td></td>
<td>281</td>
</tr>
<tr>
<td></td>
<td>282</td>
<td></td>
<td>283</td>
<td></td>
<td>284</td>
<td></td>
<td>285</td>
</tr>
<tr>
<td></td>
<td>286</td>
<td></td>
<td>287</td>
<td></td>
<td>288</td>
<td></td>
<td>289</td>
</tr>
<tr>
<td></td>
<td>290</td>
<td></td>
<td>291</td>
<td></td>
<td>292</td>
<td></td>
<td>293</td>
</tr>
<tr>
<td></td>
<td>294</td>
<td></td>
<td>295</td>
<td></td>
<td>296</td>
<td></td>
<td>297</td>
</tr>
<tr>
<td></td>
<td>298</td>
<td></td>
<td>299</td>
<td></td>
<td>300</td>
<td></td>
<td>301</td>
</tr>
</tbody>
</table>

**Translation:**

<table>
<thead>
<tr>
<th>Line</th>
<th>Numbers</th>
<th>Characters</th>
</tr>
</thead>
<tbody>
<tr>
<td>297</td>
<td>300</td>
<td>297-300</td>
</tr>
<tr>
<td>298</td>
<td>560</td>
<td>298-560</td>
</tr>
<tr>
<td>299</td>
<td>566</td>
<td>299-566</td>
</tr>
<tr>
<td>315</td>
<td>338</td>
<td>315-338</td>
</tr>
<tr>
<td>316</td>
<td>337</td>
<td>316-337</td>
</tr>
<tr>
<td>317</td>
<td>336</td>
<td>317-336</td>
</tr>
<tr>
<td>318</td>
<td>335</td>
<td>318-335</td>
</tr>
<tr>
<td>319</td>
<td>1468</td>
<td>319-1468</td>
</tr>
<tr>
<td>320</td>
<td>1435</td>
<td>320-1435</td>
</tr>
<tr>
<td>321</td>
<td>1432</td>
<td>321-1432</td>
</tr>
<tr>
<td>334</td>
<td>1434</td>
<td>334-1434</td>
</tr>
<tr>
<td>338</td>
<td>351</td>
<td>338-351</td>
</tr>
<tr>
<td>339</td>
<td>350</td>
<td>339-350</td>
</tr>
<tr>
<td>340</td>
<td>349</td>
<td>340-349</td>
</tr>
<tr>
<td>341</td>
<td>346</td>
<td>341-346</td>
</tr>
<tr>
<td>342</td>
<td>347</td>
<td>342-347</td>
</tr>
<tr>
<td>343</td>
<td>348</td>
<td>343-348</td>
</tr>
<tr>
<td>362</td>
<td>365</td>
<td>362-365</td>
</tr>
<tr>
<td>366</td>
<td>369</td>
<td>366-369</td>
</tr>
<tr>
<td>367</td>
<td>393</td>
<td>367-393</td>
</tr>
<tr>
<td>369</td>
<td>392</td>
<td>369-392</td>
</tr>
<tr>
<td>370</td>
<td>391</td>
<td>370-391</td>
</tr>
<tr>
<td>373</td>
<td>451</td>
<td>373-451</td>
</tr>
<tr>
<td>374</td>
<td>389</td>
<td>374-389</td>
</tr>
<tr>
<td>375</td>
<td>388</td>
<td>375-388</td>
</tr>
<tr>
<td>376</td>
<td>387</td>
<td>376-387</td>
</tr>
<tr>
<td>377</td>
<td>386</td>
<td>377-386</td>
</tr>
<tr>
<td>378</td>
<td>388</td>
<td>378-388</td>
</tr>
<tr>
<td>379</td>
<td>384</td>
<td>379-384</td>
</tr>
<tr>
<td>380</td>
<td>385</td>
<td>380-385</td>
</tr>
<tr>
<td>405</td>
<td>499</td>
<td>405-499</td>
</tr>
<tr>
<td>406</td>
<td>436</td>
<td>406-436</td>
</tr>
<tr>
<td>407</td>
<td>435</td>
<td>407-435</td>
</tr>
<tr>
<td>408</td>
<td>434</td>
<td>408-434</td>
</tr>
<tr>
<td>409</td>
<td>433</td>
<td>409-433</td>
</tr>
<tr>
<td>410</td>
<td>432</td>
<td>410-432</td>
</tr>
<tr>
<td>411</td>
<td>430</td>
<td>411-430</td>
</tr>
<tr>
<td>412</td>
<td>429</td>
<td>412-429</td>
</tr>
<tr>
<td>413</td>
<td>428</td>
<td>413-428</td>
</tr>
<tr>
<td>414</td>
<td>427</td>
<td>414-427</td>
</tr>
<tr>
<td>415</td>
<td>426</td>
<td>415-426</td>
</tr>
</tbody>
</table>

186
<p>| 416-427 | 411-422 | GU | WC | 41 | (411-422)(487-524) | CA |
| 417-426 | 412-421 | CG | WC | 42 | | CA |
| 418-425 | 413-420 | CG | WC | 43 | | CA |
| 419-424 | 414-419 | CG | WC | 44 | | CA |
| 420-425 | 415-419 | CG | RS | 45 | | CA |
| 427-541 | 417-532 | UA | H | 46 | bg.(411-422)(487-524) | |
| 429-437 | 422-426 | UA | H | 47 | | CA |
| 438-495 | 427-429 | UA | H | 48 | | CA |
| 440-497 | 432-481 | AU | H | 49 | 43(481-478) | TT |
| 441-493 | 435-477 | GA | 50 | | | CA |
| 442-492 | 436-476 | CG | WC | 51 | | CA |
| 444-491 | 437-475 | CG | WC | 52 | | CA |
| 444-490 | 438-474 | CG | WC | 53 | | CA |
| 445-489 | 439-473 | GC | WC | 54 | | CA |
| 446-488 | 440-472 | GC | WC | 55 | | CA |
| 447-487 | 441-471 | GA | S | 56 | A,GA | |
| 448-486 | 442-470 | AT | H | 57 | GA,E | 442.470x469 |
| 449-484 | 443-468 | CG | H | 58 | | CA |
| 450-483 | 444-467 | GC | WC | 59 | | CA |
| 451-482 | 445-466 | CG | WC | 60 | | GA |
| 452-481 | 446-465 | CG | WC | 61 | | GA |
| 454-477 | 448-461 | CG | WC | 63 | | CA |
| 455-476 | 450-469 | CG | WC | 64 | | CA |
| 456-475 | 451-459 | CG | WC | 65 | | CA |
| 457-474 | 452-458 | CG | WC | 66 | | CA |
| 458-473 | 453-472 | GC | WC | 67 | | CA |
| 459-472 | 454-471 | GC | WC | 68 | | CA |
| 460-471 | 455-470 | GC | WC | 69 | | CA |
| 462-469 | 456-468 | GC | WC | 70 | | CA |
| 464-468 | 453-467 | GC | WC | 71 | | CA |
| 497-547 | 481-520 | UA | H | 72 | (481-530)(499 | (37-392)(481-530) |
| 500-545 | 483-525 | GC | WC | 73 | | CA |
| 501-544 | 484-527 | GC | WC | 74 | (35,532)(484-527), (36-351)(484-527) | CA |
| 501-549 | 484-532 | GC | WC | 75 | bg.(35-532)(484-527) | |
| 502-543 | 485-526 | GC | WC | 76 | (485-526)(492 | | |
| 503-542 | 486-525 | GC | WC | 77 | (486-525)(491-493) | CA |
| 504-541 | 487-524 | GC | WC | 78 | (411-422)(487-524) | CA |
| 505-526 | 488-509 | GC | WC | 79 | | CA |
| 506-525 | 489-508 | GC | WC | 80 | | CA |
| 507-524 | 490-507 | GC | WC | 81 | | CA |
| 508-501 | 491-493 | CA | pS | 82 | (486-525)(491-493) | |
| 510-542 | 493-525 | AG | pS | 83 | 187 |</p>
<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Gene</th>
<th>Transcript</th>
<th>Exon</th>
<th>Intron</th>
<th>Splice Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>192</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Position</td>
<td>Sequence</td>
<td>Description</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>----------</td>
<td>-------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1010:1019</td>
<td>989-997</td>
<td>GC WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1011:1018</td>
<td>989-996</td>
<td>GC WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1012:1017</td>
<td>990-995</td>
<td>WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1013:1016</td>
<td>991-994</td>
<td>GA S 83 965.1198@991.994</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1016:1217</td>
<td>994.1198</td>
<td>HC 965.1198@991.994</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1025:1036</td>
<td>1003-1010</td>
<td>GC WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1026:1035</td>
<td>1004-1011</td>
<td>GA WC 83 1004-1017862</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1027:1034</td>
<td>1005-1012</td>
<td>GC WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1029:1039</td>
<td>1006-1016</td>
<td>GC WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1046:1211</td>
<td>1028-1192</td>
<td>GU WC 83 1028-1192</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1047:1210</td>
<td>1029-1191</td>
<td>GC WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1048:1209</td>
<td>1030-1190</td>
<td>GC WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1050:1208</td>
<td>1032-1188</td>
<td>GC WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1051:1207</td>
<td>1033-1189</td>
<td>GC WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1052:1206</td>
<td>1034-1187</td>
<td>UG WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1053:1205</td>
<td>1035-1186</td>
<td>GU 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1055:1205</td>
<td>1037-1186</td>
<td>AU WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1056:1204</td>
<td>1038-1186</td>
<td>UA WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1057:1203</td>
<td>1039-1181</td>
<td>GC WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1058:1199</td>
<td>1040-1182</td>
<td>GU WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1059:1199</td>
<td>1041-1179</td>
<td>GC WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1060:1199</td>
<td>1042-1178</td>
<td>GC WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1061:1198</td>
<td>1043-1177</td>
<td>GC WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1062:1194</td>
<td>1044-1174</td>
<td>UG WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1063:1193</td>
<td>1045-1172</td>
<td>GC WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1064:1192</td>
<td>1046-1171</td>
<td>GC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1066:1191</td>
<td>1047-1172</td>
<td>CA 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1067:1191</td>
<td>1049-1179</td>
<td>AG 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1068:1107</td>
<td>1050-1184</td>
<td>GC WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1069:1106</td>
<td>1051-1183</td>
<td>GC WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1070:1105</td>
<td>1052-1182</td>
<td>GC WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1071:1104</td>
<td>1053-1181</td>
<td>GC WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1072:1102</td>
<td>1054-1180</td>
<td>WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1073:1102</td>
<td>1055-1179</td>
<td>GC WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1074:1083</td>
<td>1056-1163</td>
<td>GC WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1075:1082</td>
<td>1057-1164</td>
<td>GC WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1076:1081</td>
<td>1058-1165</td>
<td>GC WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1077:1080</td>
<td>1059-1164</td>
<td>GA S 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1086:1099</td>
<td>1060-1163</td>
<td>UG WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1087:1098</td>
<td>1061-1164</td>
<td>GC WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1088:1097</td>
<td>1062-1165</td>
<td>GC WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1089:1096</td>
<td>1063-1166</td>
<td>GC WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1091:1179</td>
<td>1071-1165</td>
<td>GA 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1092:1079</td>
<td>1072-1166</td>
<td>GC WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1093:1169</td>
<td>1073-1167</td>
<td>UG WC 83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bpT</td>
<td>bpC</td>
<td>RS</td>
<td>Motifs</td>
<td>BT</td>
<td>BQ</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>----</td>
<td>--------</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>UG</td>
<td>WC</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AU</td>
<td>WC</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GC</td>
<td>WC</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG</td>
<td>WC</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UG</td>
<td>WC</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GC</td>
<td>WC</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG</td>
<td>WC</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GA</td>
<td>S</td>
<td>CA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C: basepairs present (+) or absent (-) in the comparative model
X: basepairs present (+) or absent (-) in the crystal structure
CC: color-coded covariation ratings for basepairs (red for confident basepairs; green, black, grey, or blue for less confident basepairs)
bpT: basepair type
bpC: basepair conformations, Watson-Crick (WC), wobble (Wb), slipped Watson-Crick (sWC), slipped wobble (sWb), reversed Watson-Crick (rWC), reversed wobble (rWb), Hoogsteen (H), reversed Hoogsteen (rH), sheared (S), reversed sheared (rS), flipped sheared (fS), parallel flipped sheared (pS), parallel sheared (pS), and reversed parallel sheared (rpS)
RS: rasmol scripts for visualizing basepairs
Motifs: motifs that basepairs are involved in (A for AA.AG@helix.ends motifs, GA for tandem GA motifs, E for AGUA/GAA motifs, EL for GUA/GAA motifs, and LPTL for lonepair triloop motifs)
BT: basepairs involved in base triples
BQ: basepairs involved in base quadruples
CA: comparative information (CA for covariation-based basepairs and TT for tentatives)
Appendix B

List of Basepairs in the *H. marismortui* 23S rRNA Comparative Structure Model and in the Crystal Structure of the 50S Ribosomal Subunit

<table>
<thead>
<tr>
<th>E. coli</th>
<th><em>H. maris.</em></th>
<th>CA</th>
<th>BQ</th>
<th>Motifs</th>
<th>BT</th>
<th>RS</th>
<th>bpT</th>
<th>bpC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:2902</td>
<td>-NH-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2:2901</td>
<td>-NH-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3:2900</td>
<td>-NH-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4:2899</td>
<td>-NH-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5:2898</td>
<td>-NH-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6:2897</td>
<td>-NH-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7:2896</td>
<td>-NH-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8:2895</td>
<td>-NH-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9:2894</td>
<td>-NH-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:2893</td>
<td>-NH-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11:2892</td>
<td>-NH-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

198
<p>| 237-260 | 208-231 | CG | WC | RS | (208-231) | (208-231) | (208-231) | (208-231) | CA |
| 237-269 | 208-231 | CG | WC | RS | (208-231) | (208-231) | (208-231) | (208-231) | CA |
| 239-258 | 210-231 | UG | Wb | RS | (210-231) | (210-231) | (210-231) | (210-231) | CA |
| 241-256 | 212-233 | AA | rs | RS | (212-233) | (212-233) | (212-233) | (212-233) | CA |
| 243-255 | 214-236 | UA | rH | RS | (214-236) | (214-236) | (214-236) | (214-236) | CA |
| 245-253 | 216-234 | AU | WC | RS | (216-234) | (216-234) | (216-234) | (216-234) | CA |
| 260-269 | 221-231 | GA | rs | RS | (221-231) | (221-231) | (221-231) | (221-231) | CA |
| 261-270 | 222-231 | CG | WC | RS | (222-231) | (222-231) | (222-231) | (222-231) | CA |
| 267-275 | 228-243 | CG | WC | RS | (228-243) | (228-243) | (228-243) | (228-243) | CA |
| 534:559 | 526:636 | I P | AU WC RS | CA |
| 535:558 | 524:615 | I P | GC WC RS | CA |
| 536:557 | 522:614 | I P | AU WC RS | CA |
| 537:556 | 531:653 | I P | GC WC RS, GA | CA |
| 538:555 | 534:552 | I P | GU Wb RS, GA | CA |
| 539:554 | 535:551 | I P | GU Wb RS | CA |
| 540:553 | 536:550 | I P | CG WC RS | CA |
| 541:552 | 537:550 | I P | AU WC RS | CA |
| 542:551 | 538:606 | I P | UA WC RS | CA |
| 543:550 | 539:559 | I P | 548.608(1324:1333) | CA |
| 544:549 | 540:548 | I P | NH | CA |
| 545:548 | 541:548 | I P | GU Wb RS | CA |
| 546:548 | 542:547 | I P | UC WC RS | CA |
| 547:547 | 543:546 | I P | UA WC RS | CA |
| 548:546 | 544:546 | I P | GC WC RS, GA | CA |
| 549:545 | 545:545 | I P | UA WC RS | CA |
| 550:544 | 546:544 | I P | GC WC RS, GA | CA |
| 551:543 | 547:543 | I P | AU WC RS | CA |
| 552:542 | 548:542 | I P | CC Wb RS | CA |
| 553:541 | 549:541 | I P | GA S RS, A, GA,E | CA |
| 554:540 | 550:540 | I P | AG S RS | CA |
| 555:540 | 551:539 | I P | AG rS RS, GA,E | CA |
| 556:539 | 552:538 | I P | UA WC RS | CA |
| 557:538 | 553:537 | I P | CA rS RS, aE | CA |
| 558:537 | 554:536 | I P | CC S, Wb RS | CA |
| 559:536 | 555:535 | I P | GC WC RS | CA |
| 560:535 | 556:534 | I P | AU WC RS | CA |
| 561:534 | 557:533 | I P | GC WC RS | CA |
| 562:533 | 558:532 | I P | CG WC RS | CA |
| 563:532 | 559:531 | I P | CG WC RS | CA |
| 564:531 | 560:530 | I P | GC WC RS, GA | CA |
| 565:530 | 561:529 | I P | GA S RS | CA |
| 566:529 | 562:528 | I P | AA Wb RS | CA |
| 567:528 | 563:527 | I P | UA rH RS, CPTL | CA |</p>
<table>
<thead>
<tr>
<th>Line</th>
<th>Symbol</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>569</td>
<td>UA</td>
<td>S</td>
<td>(626.1081)(1045.1069)</td>
</tr>
<tr>
<td>572</td>
<td>AG</td>
<td>S</td>
<td>(629.3070)(629.3073)</td>
</tr>
<tr>
<td>574</td>
<td>AG</td>
<td>S</td>
<td>(631.2075)(631.2075)</td>
</tr>
<tr>
<td>577</td>
<td>GC</td>
<td>WC</td>
<td>(636.1365)(2060.2070)</td>
</tr>
<tr>
<td>579</td>
<td>AG</td>
<td>S</td>
<td>(636.1365)(2060.2070)</td>
</tr>
<tr>
<td>580</td>
<td>CG</td>
<td>WC</td>
<td>(637.1364)(2060.2070)</td>
</tr>
<tr>
<td>581</td>
<td>GC</td>
<td>WC</td>
<td>(638.1363)(2060.2070)</td>
</tr>
<tr>
<td>582</td>
<td>AU</td>
<td>WC</td>
<td>(640.1362)(2060.2070)</td>
</tr>
<tr>
<td>583</td>
<td>GC</td>
<td>WC</td>
<td>(642.1358)(2060.2070)</td>
</tr>
<tr>
<td>584</td>
<td>GC</td>
<td>WC</td>
<td>(642.1358)(2060.2070)</td>
</tr>
<tr>
<td>585</td>
<td>GA</td>
<td>S</td>
<td>(643.1355)(643.1355)</td>
</tr>
<tr>
<td>586</td>
<td>AG</td>
<td>S</td>
<td>(643.1355)(643.1355)</td>
</tr>
<tr>
<td>587</td>
<td>AA</td>
<td>S</td>
<td>(643.1355)(643.1355)</td>
</tr>
<tr>
<td>588</td>
<td>UA</td>
<td>WC</td>
<td>(645.1354)(645.1354)</td>
</tr>
<tr>
<td>589</td>
<td>UA</td>
<td>WC</td>
<td>(645.1354)(645.1354)</td>
</tr>
<tr>
<td>590</td>
<td>UA</td>
<td>WC</td>
<td>(646.1353)(646.1353)</td>
</tr>
<tr>
<td>591</td>
<td>UA</td>
<td>WC</td>
<td>(646.1353)(646.1353)</td>
</tr>
<tr>
<td>592</td>
<td>UA</td>
<td>WC</td>
<td>(647.1352)(647.1352)</td>
</tr>
<tr>
<td>593</td>
<td>GC</td>
<td>WC</td>
<td>(647.1352)(647.1352)</td>
</tr>
<tr>
<td>594</td>
<td>CG</td>
<td>WC</td>
<td>(648.1351)(648.1351)</td>
</tr>
<tr>
<td>595</td>
<td>CG</td>
<td>WC</td>
<td>(649.1350)(649.1350)</td>
</tr>
<tr>
<td>596</td>
<td>CG</td>
<td>WC</td>
<td>(650.1349)(649.1350)</td>
</tr>
<tr>
<td>597</td>
<td>CG</td>
<td>WC</td>
<td>(651.1348)(649.1350)</td>
</tr>
<tr>
<td>598</td>
<td>CG</td>
<td>WC</td>
<td>(652.1347)(649.1350)</td>
</tr>
<tr>
<td>599</td>
<td>CG</td>
<td>WC</td>
<td>(653.1346)(649.1350)</td>
</tr>
<tr>
<td>600</td>
<td>CG</td>
<td>WC</td>
<td>(654.1345)(649.1350)</td>
</tr>
<tr>
<td>601</td>
<td>CG</td>
<td>WC</td>
<td>(655.1344)(649.1350)</td>
</tr>
<tr>
<td>602</td>
<td>AA</td>
<td>A,GA</td>
<td>(656.1343)(649.1350)</td>
</tr>
<tr>
<td>603</td>
<td>AA</td>
<td>A,GA</td>
<td>(657.1342)(649.1350)</td>
</tr>
<tr>
<td>604</td>
<td>GC</td>
<td>WC</td>
<td>(658.1341)(649.1350)</td>
</tr>
<tr>
<td>605</td>
<td>GC</td>
<td>WC</td>
<td>(659.1340)(649.1350)</td>
</tr>
<tr>
<td>606</td>
<td>CG</td>
<td>WC</td>
<td>(660.1339)(649.1350)</td>
</tr>
<tr>
<td>607</td>
<td>GC</td>
<td>WC</td>
<td>(661.1338)(649.1350)</td>
</tr>
<tr>
<td>608</td>
<td>GC</td>
<td>WC</td>
<td>(662.1337)(649.1350)</td>
</tr>
<tr>
<td>609</td>
<td>GC</td>
<td>WC</td>
<td>(663.1336)(649.1350)</td>
</tr>
<tr>
<td>610</td>
<td>CG</td>
<td>WC</td>
<td>(664.1335)(649.1350)</td>
</tr>
<tr>
<td>611</td>
<td>CG</td>
<td>WC</td>
<td>(665.1334)(649.1350)</td>
</tr>
<tr>
<td>612</td>
<td>GC</td>
<td>WC</td>
<td>(666.1333)(649.1350)</td>
</tr>
<tr>
<td>613</td>
<td>GC</td>
<td>WC</td>
<td>(667.1332)(649.1350)</td>
</tr>
<tr>
<td>614</td>
<td>GC</td>
<td>WC</td>
<td>(668.1331)(649.1350)</td>
</tr>
<tr>
<td>615</td>
<td>GC</td>
<td>WC</td>
<td>(669.1330)(649.1350)</td>
</tr>
<tr>
<td>616</td>
<td>GC</td>
<td>WC</td>
<td>(670.1329)(649.1350)</td>
</tr>
<tr>
<td>617</td>
<td>GC</td>
<td>WC</td>
<td>(671.1328)(649.1350)</td>
</tr>
<tr>
<td>618</td>
<td>GC</td>
<td>WC</td>
<td>(672.1327)(649.1350)</td>
</tr>
<tr>
<td>619</td>
<td>GC</td>
<td>WC</td>
<td>(673.1326)(649.1350)</td>
</tr>
<tr>
<td>620</td>
<td>GC</td>
<td>WC</td>
<td>(674.1325)(649.1350)</td>
</tr>
<tr>
<td>621</td>
<td>GC</td>
<td>WC</td>
<td>(675.1324)(649.1350)</td>
</tr>
<tr>
<td>622</td>
<td>GC</td>
<td>WC</td>
<td>(676.1323)(649.1350)</td>
</tr>
<tr>
<td>623</td>
<td>GC</td>
<td>WC</td>
<td>(677.1322)(649.1350)</td>
</tr>
<tr>
<td>624</td>
<td>GC</td>
<td>WC</td>
<td>(678.1321)(649.1350)</td>
</tr>
<tr>
<td>625</td>
<td>GC</td>
<td>WC</td>
<td>(679.1320)(649.1350)</td>
</tr>
<tr>
<td>626</td>
<td>GC</td>
<td>WC</td>
<td>(680.1319)(649.1350)</td>
</tr>
<tr>
<td>627</td>
<td>GC</td>
<td>WC</td>
<td>(681.1318)(649.1350)</td>
</tr>
<tr>
<td>628</td>
<td>GC</td>
<td>WC</td>
<td>(682.1317)(649.1350)</td>
</tr>
<tr>
<td>629</td>
<td>GC</td>
<td>WC</td>
<td>(683.1316)(649.1350)</td>
</tr>
<tr>
<td>630</td>
<td>GC</td>
<td>WC</td>
<td>(684.1315)(649.1350)</td>
</tr>
<tr>
<td>631</td>
<td>GC</td>
<td>WC</td>
<td>(685.1314)(649.1350)</td>
</tr>
<tr>
<td>632</td>
<td>GC</td>
<td>WC</td>
<td>(686.1313)(649.1350)</td>
</tr>
<tr>
<td>633</td>
<td>GC</td>
<td>WC</td>
<td>(687.1312)(649.1350)</td>
</tr>
<tr>
<td>634</td>
<td>GC</td>
<td>WC</td>
<td>(688.1311)(649.1350)</td>
</tr>
<tr>
<td>635</td>
<td>GC</td>
<td>WC</td>
<td>(689.1310)(649.1350)</td>
</tr>
<tr>
<td>636</td>
<td>GC</td>
<td>WC</td>
<td>(690.1309)(649.1350)</td>
</tr>
<tr>
<td>637</td>
<td>GC</td>
<td>WC</td>
<td>(691.1308)(649.1350)</td>
</tr>
<tr>
<td>638</td>
<td>GC</td>
<td>WC</td>
<td>(692.1307)(649.1350)</td>
</tr>
<tr>
<td>639</td>
<td>GC</td>
<td>WC</td>
<td>(693.1306)(649.1350)</td>
</tr>
<tr>
<td>640</td>
<td>GC</td>
<td>WC</td>
<td>(694.1305)(649.1350)</td>
</tr>
<tr>
<td>641</td>
<td>GC</td>
<td>WC</td>
<td>(695.1304)(649.1350)</td>
</tr>
<tr>
<td>642</td>
<td>GC</td>
<td>WC</td>
<td>(696.1303)(649.1350)</td>
</tr>
<tr>
<td>643</td>
<td>GC</td>
<td>WC</td>
<td>(697.1302)(649.1350)</td>
</tr>
<tr>
<td>644</td>
<td>GC</td>
<td>WC</td>
<td>(698.1301)(649.1350)</td>
</tr>
<tr>
<td>645</td>
<td>GC</td>
<td>WC</td>
<td>(699.1300)(649.1350)</td>
</tr>
<tr>
<td>646</td>
<td>GC</td>
<td>WC</td>
<td>(700.1299)(649.1350)</td>
</tr>
<tr>
<td>647</td>
<td>GC</td>
<td>WC</td>
<td>(701.1298)(649.1350)</td>
</tr>
<tr>
<td>648</td>
<td>GC</td>
<td>WC</td>
<td>(702.1297)(649.1350)</td>
</tr>
<tr>
<td>649</td>
<td>GC</td>
<td>WC</td>
<td>(703.1296)(649.1350)</td>
</tr>
<tr>
<td>650</td>
<td>GC</td>
<td>WC</td>
<td>(704.1295)(649.1350)</td>
</tr>
<tr>
<td>779:785</td>
<td>872:878</td>
<td></td>
<td>UG</td>
</tr>
<tr>
<td>780:783</td>
<td>873:878</td>
<td></td>
<td>GA</td>
</tr>
<tr>
<td>812:1195</td>
<td>905:1300</td>
<td></td>
<td>CG</td>
</tr>
<tr>
<td>812:1195</td>
<td>905:1300</td>
<td></td>
<td>CG</td>
</tr>
<tr>
<td>813:1194</td>
<td>906:1298</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>814:1193</td>
<td>907:1298</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>815:1192</td>
<td>908:1229</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>816:1191</td>
<td>909:1229</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>817:1190</td>
<td>910:1229</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>818:1189</td>
<td>911:1229</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>819:1188</td>
<td>912:1229</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>820:972</td>
<td>913:1075</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>820:972</td>
<td>913:1075</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>821:971</td>
<td>914:1070</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>822:970</td>
<td>915:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>823:969</td>
<td>916:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>824:968</td>
<td>917:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>825:967</td>
<td>918:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>826:966</td>
<td>919:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>827:965</td>
<td>920:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>828:964</td>
<td>921:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>829:963</td>
<td>922:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>830:962</td>
<td>923:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>831:961</td>
<td>924:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>832:960</td>
<td>925:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>833:959</td>
<td>926:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>834:958</td>
<td>927:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>835:957</td>
<td>928:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>836:956</td>
<td>929:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>837:955</td>
<td>930:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>838:954</td>
<td>931:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>839:953</td>
<td>932:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>840:952</td>
<td>933:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>841:951</td>
<td>934:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>842:950</td>
<td>935:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>843:949</td>
<td>936:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>844:948</td>
<td>937:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>845:947</td>
<td>938:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>846:946</td>
<td>939:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>847:945</td>
<td>940:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>848:944</td>
<td>941:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>849:943</td>
<td>942:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>850:942</td>
<td>943:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>851:941</td>
<td>944:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>852:940</td>
<td>945:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>853:939</td>
<td>946:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>854:938</td>
<td>947:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>855:937</td>
<td>948:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>856:936</td>
<td>949:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>857:935</td>
<td>950:970</td>
<td></td>
<td>CA</td>
</tr>
<tr>
<td>857:920</td>
<td>951-1016</td>
<td>GC WC</td>
<td>CA</td>
</tr>
<tr>
<td>858:918</td>
<td>952-1014</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>858:919</td>
<td>952-1014</td>
<td>GC WC</td>
<td>FT</td>
</tr>
<tr>
<td>859:918</td>
<td>953-1014</td>
<td>E</td>
<td>FT</td>
</tr>
<tr>
<td>860:917</td>
<td>954-1013</td>
<td>UA pH</td>
<td>A,G,R</td>
</tr>
<tr>
<td>861-916</td>
<td>955-1012</td>
<td>AA S</td>
<td>A,G,R</td>
</tr>
<tr>
<td>862:915</td>
<td>956-1012</td>
<td>GC WC</td>
<td>FT</td>
</tr>
<tr>
<td>863:913</td>
<td>957-1011</td>
<td>AU WC</td>
<td>CA</td>
</tr>
<tr>
<td>864:912</td>
<td>958-1009</td>
<td>GC WC</td>
<td>CA</td>
</tr>
<tr>
<td>868:909</td>
<td>959-1008</td>
<td>CA Wb</td>
<td>CA</td>
</tr>
<tr>
<td>869:908</td>
<td>960-1007</td>
<td>GC WC</td>
<td>CA</td>
</tr>
<tr>
<td>870:907</td>
<td>961-1006</td>
<td>AU WC</td>
<td>CA</td>
</tr>
<tr>
<td>871:906</td>
<td>962-1005</td>
<td>UG pH</td>
<td>CA</td>
</tr>
<tr>
<td>872:905</td>
<td>963-1004</td>
<td>IU Wb</td>
<td>CA</td>
</tr>
<tr>
<td>873:904</td>
<td>964-1003</td>
<td>GC WC</td>
<td>CA</td>
</tr>
<tr>
<td>874:903</td>
<td>965-999</td>
<td>AU</td>
<td>CA</td>
</tr>
<tr>
<td>875:902</td>
<td>966-998</td>
<td>UA</td>
<td>CA</td>
</tr>
<tr>
<td>876:901</td>
<td>967-997</td>
<td>GC</td>
<td>CA</td>
</tr>
<tr>
<td>877:900</td>
<td>968-996</td>
<td>UA</td>
<td>ACA</td>
</tr>
<tr>
<td>879:898</td>
<td>969-995</td>
<td>NH</td>
<td>CA</td>
</tr>
<tr>
<td>880:897</td>
<td>970-994</td>
<td>GC</td>
<td>CA</td>
</tr>
<tr>
<td>881:896</td>
<td>971-993</td>
<td>CU</td>
<td>CA</td>
</tr>
<tr>
<td>882:895</td>
<td>972-992</td>
<td>GC</td>
<td>CA</td>
</tr>
<tr>
<td>883:894</td>
<td>973-991</td>
<td>GC</td>
<td>CA</td>
</tr>
<tr>
<td>884:893</td>
<td>974-990</td>
<td>GC</td>
<td>CA</td>
</tr>
<tr>
<td>885:892</td>
<td>975-989</td>
<td>GC</td>
<td>CA</td>
</tr>
<tr>
<td>918:2268</td>
<td>1004-2302</td>
<td>AA sR</td>
<td>CA</td>
</tr>
<tr>
<td>945:972</td>
<td>1004-1010</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>946:971</td>
<td>1005-1009</td>
<td>GC WC</td>
<td>CA</td>
</tr>
<tr>
<td>947:970</td>
<td>1006-1008</td>
<td>GC WC</td>
<td>CA</td>
</tr>
<tr>
<td>948:969</td>
<td>1007-1006</td>
<td>GC WC</td>
<td>CA</td>
</tr>
<tr>
<td>949:968</td>
<td>1008-1006</td>
<td>GC WC</td>
<td>CA</td>
</tr>
<tr>
<td>950:967</td>
<td>1009-1005</td>
<td>GC WC</td>
<td>CA</td>
</tr>
<tr>
<td>951:966</td>
<td>1010-1004</td>
<td>GC WC</td>
<td>CA</td>
</tr>
<tr>
<td>952:965</td>
<td>1011-1002</td>
<td>GC WC</td>
<td>CA</td>
</tr>
<tr>
<td>953:964</td>
<td>1012-1002</td>
<td>GC WC</td>
<td>CA</td>
</tr>
<tr>
<td>953:2266</td>
<td>1013-2300</td>
<td>Bq (1052-1062) (2300-2307)</td>
<td>CA</td>
</tr>
<tr>
<td>954:963</td>
<td>1013-1065</td>
<td>GC WC</td>
<td>CA</td>
</tr>
<tr>
<td>955:962</td>
<td>1014-1060</td>
<td>GC WC</td>
<td>CA</td>
</tr>
<tr>
<td>956:960</td>
<td>1015-1058</td>
<td>GC WC</td>
<td>CA</td>
</tr>
<tr>
<td>960:2456</td>
<td>1018-2500</td>
<td>AG epS</td>
<td>BQ</td>
</tr>
<tr>
<td>960:2455</td>
<td>1018-2500</td>
<td>AG epS</td>
<td>BQ</td>
</tr>
<tr>
<td>961:2030</td>
<td>1019-2071</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>962:2496</td>
<td>1020-2531</td>
<td>Bq (1054-1060) (2490-2531)</td>
<td>CA</td>
</tr>
<tr>
<td>964:2273</td>
<td>1022-2300</td>
<td>Bq (1052-1062) (2300-2307)</td>
<td>CA</td>
</tr>
<tr>
<td>965:2267</td>
<td>1023-2302</td>
<td>GA S</td>
<td>CA</td>
</tr>
<tr>
<td>971:983</td>
<td>1026-2308</td>
<td>Bq (626-1081,1045-1069)</td>
<td>CA</td>
</tr>
</tbody>
</table>
210
<p>| 1433:1559 | 1539:1646 | UG | Wb | RS | CA |
| 1434:1558 | 1540:1645 | GU | Wb | RS | (1540:1645)(1581:1614) | CA |
| 1435:1557 | 1541:1644 | GC | WC | RS | CA |
| 1436:1551 | 1542:1618 | GA | rS | RS | BQ, (1542:1643)(1579:1616) | CA |
| 1436:1556 | 1542:1643 | GC | WC | RS | (1542:1643)(1579:1616) | CA |
| 1437:1555 | 1543:1642 | GA | S | RS | EL | CA |
| 1438:1553 | 1544:1641 | UA | Wb | RS | EL | CA |
| 1439:1552 | 1545:1639 | CC | S | RS | A.EL | (1525:1649)(1594) |
| 1460:1511 | 1546:1638 | UA | WC | RS | CA |
| 1461:1550 | 1547:1637 | UA | WC | RS | CA |
| 1462:1549 | 1548:1636 | CG | WC | RS | CA |
| 1463:1548 | 1549:1635 | AU | WC | RS | CA |
| 1464:1547 | 1550:1634 | -NH– | CG | WC | RS | CA |
| 1465:1546 | 1551:1633 | -NH– | CG | WC | RS | CA |
| 1466:1545 | 1552:1632 | GU | Wb | RS | CA |
| 1467:1544 | 1553:1631 | GU | Wb | RS | (1553:1568)(1629:1632) | CA |
| 1470:1521 | 1554:1567 | UA | WC | RS | (1554:1567)(1631) | CA |
| 1471:1520 | 1555:1566 | GC | WC | RS | CA |
| 1472:1519 | 1556:1565 | GC | WC | RS | CA |
| 1473:1518 | 1557:1564 | GC | WC | RS | CA |
| 1474:1517 | 1558:1563 | GC | WC | RS | CA |
| 1475:1516 | 1559:1562 | CC | S | RS | A.EL | (1497:1508)(1570:1627) | CA |
| 1476:1515 | 1560:1561 | CG | WC | RS | CA |
| 1477:1514 | 1561:1560 | CG | WC | RS | CA |
| 1478:1513 | 1562:1556 | UA | WC | RS | CA |
| 1479:1512 | 1563:1551 | UA | WC | RS | CA |
| 1480:1511 | 1564:1546 | CG | WC | RS | CA |
| 1481:1508 | 1565:1541 | CG | WC | RS | CA |
| 1482:1507 | 1566:1536 | CG | WC | RS | CA |
| 1483:1506 | 1567:1531 | GU | Wb | RS | CA |
| 1484:1505 | 1568:1526 | GA | S | RS | A.ELL | CA |
| 1485:1504 | 1569:1521 | GA | S | RS | ELL | (1589:1606)(1593:1601) | CA |
| 1486:1500 | 1570:1516 | CG | WC | RS | CA |
| 1487:1502 | 1571:1511 | UA | WC | RS | CA |
| 1488:1503 | 1572:1506 | UA | WC | RS | CA |</p>
<table>
<thead>
<tr>
<th>Start</th>
<th>End</th>
<th>Amino Acid</th>
<th>w.b</th>
<th>Start</th>
<th>End</th>
<th>Amino Acid</th>
<th>w.b</th>
<th>Start</th>
<th>End</th>
<th>Amino Acid</th>
<th>w.b</th>
</tr>
</thead>
<tbody>
<tr>
<td>1667</td>
<td>2042</td>
<td>GU</td>
<td>S*</td>
<td>1668</td>
<td>2043</td>
<td>AU</td>
<td>tH</td>
<td>1668</td>
<td>2053</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1669</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1669</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1670</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1671</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1672</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1673</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1674</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1675</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1676</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1677</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1678</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1679</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1680</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1681</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1682</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1683</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1684</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1685</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1686</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1687</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1688</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1689</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1690</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1691</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1692</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1693</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1694</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1695</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1696</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1697</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1698</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1699</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1700</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1701</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1702</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1703</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1704</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1705</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1706</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1707</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1708</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1709</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1710</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1711</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1712</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1713</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1714</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1715</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1716</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1717</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1718</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1719</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1720</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1721</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1722</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1723</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1724</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1725</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1726</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1727</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1728</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1729</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1730</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1731</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1732</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1733</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1734</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1735</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1736</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1737</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1738</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1739</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1740</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1741</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1742</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1743</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1744</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1745</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1746</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1747</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1748</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1749</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1750</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1751</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1752</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1753</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1754</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1755</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1756</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1757</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1758</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1759</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1760</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1761</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1762</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1763</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1764</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1765</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1766</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1767</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1768</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1769</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1770</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1771</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1772</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1773</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1774</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1775</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1776</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1777</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1778</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1779</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1780</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1781</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1782</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1783</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1784</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1785</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1786</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1787</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1788</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1789</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1790</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1791</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1792</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1793</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1794</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1795</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1796</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
<tr>
<td>1797</td>
<td>2034</td>
<td>GU</td>
<td>S*</td>
<td>1798</td>
<td>2034</td>
<td>AU</td>
<td>tH</td>
<td>1799</td>
<td>2034</td>
<td>GC</td>
<td>WC</td>
</tr>
</tbody>
</table>

217
| 1773:1798 | 1829:2019 | CA | AA | S | RS | 1830.1846-82 | FT |
| 1774:1790 | 1831:1848 | CA | UA | S? | RS | 1831.1845-82 | FT |
| 1776:1788 | 1832:1848 | CA | GC | WC | RS | 1832.1844-82 | FT |
| 1777:1787 | 1833:1848 | CA | UA | WC | RS | 1833.1842-82 | FT |
| 1778:1785 | 1834:1848 | CA | CC | HH | RS | 1835.1841-82 | FT |
| 1782:2586 | 1838:2621 | CA | AA | 82 | RS | 1838.2621-2111 | FT |
| 1783:2607 | 1839:2642 | CA | AG | 8S | RS | 1839.2642-2111 | FT |
| 1791:1826 | 1848:1883 | CA | GU | Wb | RS | 1848.1883-1884 | FT |
| 1793:1826 | 1849:1884 | CA | GC | WC | RS | 1849.1884-1885 | FT |
| 1794:1825 | 1850:1885 | CA | UA | WC | RS | 1850.1885-1886 | FT |
| 1795:1824 | 1851:1886 | CA | GC | WC | RS | 1851.1886-1887 | FT |
| 1796:1823 | 1852:1887 | CA | AU | WC | RS | 1852.1887-1888 | FT |
| 1797:1822 | 1853:1888 | CA | CG | WC | RS | 1853.1888-1889 | FT |
| 1798:1821 | 1854:1889 | CA | CG | WC | RS | 1854.1889-1890 | FT |
| 1799:1818 | 1855:1890 | CA | GU | rS | RS | 1855.1890-1891 | FT |
| 1800:1817 | 1856:1891 | CA | GC | WC | RS | 1856.1891-1892 | FT |
| 1802:1815 | 1858:1892 | CA | AU | rS | RS | 1858.1892-1893 | FT |
| 1803:1814 | 1859:1893 | CA | AC | Wb | RS | 1859.1893-1894 | FT |
| 1804:1813 | 1860:1894 | CA | UA | WC | RS | 1860.1894-1895 | FT |
| 1805:1812 | 1861:1895 | CA | GC | WC | RS | 1861.1895-1896 | FT |
| 1806:1811 | 1862:1896 | CA | GC | WC | RS | 1862.1896-1897 | FT |
| 1807:1810 | 1863:1897 | CA | GA | WC | RS | 1863.1897-1898 | FT |
| 1830:1975 | 1826:2026 | CA | AU | S | RS | 1830.2026-2027 | FT |
| 1831:1974 | 1827:2027 | CA | UA | WC | RS | 1831.2027-2028 | FT |
| 1832:1973 | 1828:2028 | CA | UA | WC | RS | 1832.2028-2029 | FT |
| 1833:1972 | 1829:2029 | CA | CG | WC | RS | 1833.2029-2030 | FT |
| 1834:1928 | 1830:2030 | CA | CG | WC | RS | 1834.2030-2031 | FT |
| 1834:1929 | 1831:2031 | CA | UA | H | RS | 1835.2031-2032 | FT |
| 1835:1930 | 1832:2032 | CA | UA | WC | RS | 1835.2032-2033 | FT |
| 1836:1931 | 1833:2033 | CA | GC | WC | RS | 1836.2033-2034 | FT |
| 1837:1932 | 1834:2034 | CA | GC | WC | RS | 1837.2034-2035 | FT |
| 1840:1902 | 1835:2035 | CA | GC | WC | RS | 1840.2035-2036 | FT |
| 1841:1901 | 1836:2036 | CA | UA | WC | RS | 1841.2036-2037 | FT |
| 1842:1898 | 1837:2037 | CA | GU | Wb | RS | 1842.2037-2038 | FT |
| 1843:1897 | 1838:2038 | CA | GC | WC | RS | 1843.2038-2039 | FT |
| 1844:1896 | 1839:2039 | CA | GC | WC | RS | 1844.2039-2040 | FT |
| 1845:1895 | 1840:2040 | CA | GC | WC | RS | 1845.2040-2041 | FT |
| 1846:1894 | 1841:2041 | CA | GC | WC | RS | 1846.2041-2042 | FT |
| 1849:1831 | 1842:2042 | CA | UA | WC | RS | 1849.2042-2043 | FT |
| 1850:1892 | 1843:2043 | CA | GC | WC | RS | 1850.2043-2044 | FT |
| 1851:1891 | 1844:2044 | CA | UC | Wb | RS | 1851.2044-2045 | FT |
| 1852:1890 | 1845:2045 | CA | GA | S | RS | 1852.2045-2046 | FT |

218
<table>
<thead>
<tr>
<th>Year 1</th>
<th>Year 2</th>
<th>Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>1853-1889</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1854-1888</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1855-1887</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1856-1886</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1858-1884</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1860-1882</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1861-1881</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1862-1880</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1863-1879</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1864-1878</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1865-1876</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1866-1875</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1875-2411</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1867-1874</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1868-1873</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1899-2343</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1891-2225</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1896-1924</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1907-1923</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1908-1922</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1909-1921</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1910-1920</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1911-1919</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1925-1929</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1930-1970</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1931-1969</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1932-1968</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1933-1967</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1934-1964</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1935-1962</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1936-1943</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1940-1942</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1940-1945</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1945-1951</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1946-1950</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1947-1959</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1948-1958</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1949-1957</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1950-1956</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1969-1972</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1998-2724</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1999-2687</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>1999-2722</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>2019-2035</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>2020-2034</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>2022-2042</td>
<td>1859-1885</td>
<td>CA</td>
</tr>
<tr>
<td>2085-2234</td>
<td>2126-2227</td>
<td>CG</td>
</tr>
<tr>
<td>2086-2233</td>
<td>2127-2228</td>
<td>UA</td>
</tr>
<tr>
<td>2087-2232</td>
<td>2128-2229</td>
<td>GU</td>
</tr>
<tr>
<td>2088-2231</td>
<td>2129-2230</td>
<td>UA</td>
</tr>
<tr>
<td>2089-2230</td>
<td>2130-2231</td>
<td>CG</td>
</tr>
<tr>
<td>2090-2229</td>
<td>2131-2232</td>
<td>GC</td>
</tr>
<tr>
<td>2091-2228</td>
<td>2132-2233</td>
<td>CC</td>
</tr>
<tr>
<td>2092-2227</td>
<td>2133-2234</td>
<td>UA</td>
</tr>
<tr>
<td>2093-2196</td>
<td>2134-2235</td>
<td>GC</td>
</tr>
<tr>
<td>2094-2195</td>
<td>2135-2236</td>
<td>GU</td>
</tr>
<tr>
<td>2095-2194</td>
<td>2136-2237</td>
<td>GC</td>
</tr>
<tr>
<td>2096-2193</td>
<td>2137-2238</td>
<td>AA</td>
</tr>
<tr>
<td>2097-2192</td>
<td>2138-2239</td>
<td>CG</td>
</tr>
<tr>
<td>2098-2191</td>
<td>2139-2240</td>
<td>GC</td>
</tr>
<tr>
<td>2099-2190</td>
<td>2140-2241</td>
<td>GC</td>
</tr>
<tr>
<td>2100-2189</td>
<td>2141-2242</td>
<td>GU</td>
</tr>
<tr>
<td>2101-2188</td>
<td>2142-2243</td>
<td>GC</td>
</tr>
<tr>
<td>2102-2187</td>
<td>2143-2244</td>
<td>UA</td>
</tr>
<tr>
<td>2103-2186</td>
<td>2144-2245</td>
<td>GC</td>
</tr>
<tr>
<td>2104-2185</td>
<td>2145-2246</td>
<td>GU</td>
</tr>
<tr>
<td>2105-2184</td>
<td>2146-2247</td>
<td>CG</td>
</tr>
<tr>
<td>2106-2183</td>
<td>2147-2248</td>
<td>CG</td>
</tr>
<tr>
<td>2107-2182</td>
<td>2148-2249</td>
<td>GC</td>
</tr>
<tr>
<td>2108-2181</td>
<td>2149-2250</td>
<td>GC</td>
</tr>
<tr>
<td>2109-2180</td>
<td>2150-2251</td>
<td>GC</td>
</tr>
<tr>
<td>2110-2179</td>
<td>2151-2252</td>
<td>GC</td>
</tr>
<tr>
<td>2112-2169</td>
<td>2152-2253</td>
<td>GC</td>
</tr>
<tr>
<td>2113-2170</td>
<td>2153-2254</td>
<td>GC</td>
</tr>
<tr>
<td>2117-2172</td>
<td>2154-2255</td>
<td>GC</td>
</tr>
<tr>
<td>2120-2178</td>
<td>2155-2256</td>
<td>GC</td>
</tr>
<tr>
<td>2121-2177</td>
<td>2156-2257</td>
<td>GC</td>
</tr>
<tr>
<td>2122-2176</td>
<td>2157-2258</td>
<td>UA</td>
</tr>
<tr>
<td>2123-2175</td>
<td>2158-2259</td>
<td>UA</td>
</tr>
<tr>
<td>2124-2174</td>
<td>2159-2260</td>
<td>GC</td>
</tr>
<tr>
<td>2125-2173</td>
<td>2160-2261</td>
<td>GC</td>
</tr>
<tr>
<td>2126-2162</td>
<td>2161-2262</td>
<td>GC</td>
</tr>
<tr>
<td>2127-2161</td>
<td>2162-2263</td>
<td>GC</td>
</tr>
<tr>
<td>2128-2160</td>
<td>2163-2264</td>
<td>GC</td>
</tr>
<tr>
<td>2129-2159</td>
<td>2164-2265</td>
<td>GC</td>
</tr>
<tr>
<td>2136-2155</td>
<td>2165-2266</td>
<td>GC</td>
</tr>
<tr>
<td>2137-2154</td>
<td>2166-2267</td>
<td>GC</td>
</tr>
<tr>
<td>2138-2153</td>
<td>2167-2268</td>
<td>UA</td>
</tr>
<tr>
<td>2139-2152</td>
<td>2168-2269</td>
<td>UA</td>
</tr>
<tr>
<td>2140-2151</td>
<td>2169-2270</td>
<td>UA</td>
</tr>
<tr>
<td>2141-2150</td>
<td>2170-2271</td>
<td>UA</td>
</tr>
<tr>
<td>2142-2149</td>
<td>2171-2272</td>
<td>UA</td>
</tr>
<tr>
<td>2143-2148</td>
<td>2172-2273</td>
<td>UA</td>
</tr>
<tr>
<td>2144-2147</td>
<td>2173-2274</td>
<td>UA</td>
</tr>
<tr>
<td>Start</td>
<td>End</td>
<td>GA</td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
<td>----</td>
</tr>
<tr>
<td>2197</td>
<td>2225</td>
<td></td>
</tr>
<tr>
<td>2198</td>
<td>2225</td>
<td></td>
</tr>
<tr>
<td>2199</td>
<td>2224</td>
<td></td>
</tr>
<tr>
<td>2200</td>
<td>2223</td>
<td></td>
</tr>
<tr>
<td>2201</td>
<td>2227</td>
<td></td>
</tr>
<tr>
<td>2202</td>
<td>2221</td>
<td></td>
</tr>
<tr>
<td>2203</td>
<td>2220</td>
<td></td>
</tr>
<tr>
<td>2204</td>
<td>2221</td>
<td></td>
</tr>
<tr>
<td>2205</td>
<td>2221</td>
<td></td>
</tr>
<tr>
<td>2206</td>
<td>2221</td>
<td></td>
</tr>
<tr>
<td>2207</td>
<td>2221</td>
<td></td>
</tr>
<tr>
<td>2208</td>
<td>2221</td>
<td></td>
</tr>
<tr>
<td>2209</td>
<td>2221</td>
<td></td>
</tr>
<tr>
<td>2211</td>
<td>2221</td>
<td></td>
</tr>
<tr>
<td>2246</td>
<td>2258</td>
<td></td>
</tr>
<tr>
<td>2247</td>
<td>2257</td>
<td></td>
</tr>
<tr>
<td>2248</td>
<td>2256</td>
<td></td>
</tr>
<tr>
<td>2249</td>
<td>2255</td>
<td></td>
</tr>
<tr>
<td>2255</td>
<td>2275</td>
<td></td>
</tr>
<tr>
<td>2258</td>
<td>2272</td>
<td></td>
</tr>
<tr>
<td>2259</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td>2260</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td>2261</td>
<td>2279</td>
<td></td>
</tr>
<tr>
<td>2262</td>
<td>2278</td>
<td></td>
</tr>
<tr>
<td>2263</td>
<td>2277</td>
<td></td>
</tr>
<tr>
<td>2264</td>
<td>2276</td>
<td></td>
</tr>
<tr>
<td>2265</td>
<td>2274</td>
<td></td>
</tr>
<tr>
<td>2266</td>
<td>2273</td>
<td></td>
</tr>
<tr>
<td>2267</td>
<td>2272</td>
<td></td>
</tr>
<tr>
<td>2268</td>
<td>2276</td>
<td></td>
</tr>
<tr>
<td>2269</td>
<td>2277</td>
<td></td>
</tr>
<tr>
<td>2270</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td>2271</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td>2272</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td>2273</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td>2274</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td>2275</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td>2276</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td>2277</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td>2278</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td>2279</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td>2280</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td>2281</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td>2282</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td>2283</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td>2284</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td>2285</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td>2286</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td>2287</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td>2288</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td>2289</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td>2290</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td>2291</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td>2292</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td>2293</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td>2294</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td>2295</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td>2296</td>
<td>2281</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>2543.2850</td>
<td>L</td>
<td>UA</td>
</tr>
<tr>
<td>2544.2851</td>
<td>L</td>
<td>AU</td>
</tr>
<tr>
<td>2545.2852</td>
<td>L</td>
<td>GC</td>
</tr>
<tr>
<td>2546.2853</td>
<td>L</td>
<td>CG</td>
</tr>
<tr>
<td>2547.2854</td>
<td>L</td>
<td>UG</td>
</tr>
</tbody>
</table>

227
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>C: basepairs present (+) or absent (-) in the comparative model</td>
<td>X: basepairs present (+) or absent (-) in the crystal structure</td>
<td>CC: color-coded covariation ratings for basepairs (red for confident basepairs; green, black, grey, or blue for less confident basepairs)</td>
<td>Motifs: motifs that basepairs are involved in (A for <a href="mailto:AA@helix.ends">AA@helix.ends</a> motifs, GA for tandem GA motifs, E for AGUA/GAA motifs, EL for GUA/GAA motifs, ELL for GGA/GAA motifs, and LPTL for lonepair triloop motifs)</td>
<td>CA: comparative information (CA for covariation-based basepairs and TT for tentatives)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix C

List of Basepairs in the *H. marismortui* 5S rRNA Comparative Structure Model and in the Crystal Structure of the 50S Ribosomal Subunit

<table>
<thead>
<tr>
<th>E. coli</th>
<th>H. marismortui</th>
<th>S</th>
<th>C1</th>
<th>UCU</th>
<th>G1</th>
<th>S2</th>
<th>Motif</th>
<th>RT</th>
<th>BQ</th>
<th>CA</th>
</tr>
</thead>
<tbody>
<tr>
<td>20.63</td>
<td>14.67</td>
<td>G</td>
<td>G</td>
<td>WC</td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21.62</td>
<td>15.66</td>
<td></td>
<td></td>
<td></td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22.61</td>
<td>16.64</td>
<td></td>
<td></td>
<td></td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23.60</td>
<td>17.63</td>
<td></td>
<td></td>
<td></td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24.27</td>
<td>18.62</td>
<td></td>
<td></td>
<td></td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25.36</td>
<td>19.61</td>
<td></td>
<td></td>
<td></td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26.35</td>
<td>20.60</td>
<td></td>
<td></td>
<td></td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27.34</td>
<td>21.59</td>
<td></td>
<td></td>
<td></td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28.33</td>
<td>22.58</td>
<td></td>
<td></td>
<td></td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29.54</td>
<td>23.66</td>
<td></td>
<td></td>
<td></td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30.53</td>
<td>24.67</td>
<td></td>
<td></td>
<td></td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31.52</td>
<td>25.73</td>
<td></td>
<td></td>
<td></td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32.51</td>
<td>26.76</td>
<td></td>
<td></td>
<td></td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33.50</td>
<td>27.79</td>
<td></td>
<td></td>
<td></td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34.49</td>
<td>28.82</td>
<td></td>
<td></td>
<td></td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35.48</td>
<td>29.85</td>
<td></td>
<td></td>
<td></td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36.51</td>
<td>30.88</td>
<td></td>
<td></td>
<td></td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37.54</td>
<td>31.91</td>
<td></td>
<td></td>
<td></td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38.44</td>
<td>32.94</td>
<td></td>
<td></td>
<td></td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39.43</td>
<td>33.97</td>
<td></td>
<td></td>
<td></td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40.42</td>
<td>34.99</td>
<td></td>
<td></td>
<td></td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41.41</td>
<td>35.10</td>
<td></td>
<td></td>
<td></td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42.40</td>
<td>36.13</td>
<td></td>
<td></td>
<td></td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>43.39</td>
<td>37.16</td>
<td></td>
<td></td>
<td></td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44.38</td>
<td>38.19</td>
<td></td>
<td></td>
<td></td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45.37</td>
<td>39.21</td>
<td></td>
<td></td>
<td></td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46.36</td>
<td>40.24</td>
<td></td>
<td></td>
<td></td>
<td>GC</td>
<td>WC</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

229
<table>
<thead>
<tr>
<th>-NH-</th>
<th>72-104</th>
<th>X</th>
<th>AA</th>
<th>+S</th>
<th>CA</th>
<th>X</th>
<th>A</th>
<th>-NH-</th>
</tr>
</thead>
<tbody>
<tr>
<td>77-99</td>
<td>79-103</td>
<td>X</td>
<td>UA</td>
<td>+H</td>
<td>RS</td>
<td>X</td>
<td></td>
<td>79:103/78</td>
</tr>
<tr>
<td>78-98</td>
<td>80-102</td>
<td>X</td>
<td>AG</td>
<td>S</td>
<td>RS</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>79-97</td>
<td>81-101</td>
<td>X</td>
<td>CG</td>
<td>WC</td>
<td>CA</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>80-96</td>
<td>82-100</td>
<td>X</td>
<td>UE</td>
<td>WS</td>
<td>CA</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>81-95</td>
<td>83-99</td>
<td>X</td>
<td>GU</td>
<td>WS</td>
<td>CA</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>82-94</td>
<td>84-98</td>
<td>X</td>
<td>GC</td>
<td>WC</td>
<td>CA</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>83-93</td>
<td>85-97</td>
<td>X</td>
<td>AU</td>
<td>WC</td>
<td>CA</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>84-92</td>
<td>86-96</td>
<td>X</td>
<td>GC</td>
<td>WC</td>
<td>CA</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>85-91</td>
<td>87-95</td>
<td>X</td>
<td>GC</td>
<td>WC</td>
<td>CA</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>86-90</td>
<td>88-94</td>
<td>X</td>
<td>CG</td>
<td>WC</td>
<td>CA</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-NH-</td>
<td>89-93</td>
<td>X</td>
<td>GA</td>
<td>S</td>
<td>RS</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- NH: basepairs present (+) or absent (-) in the comparative model
X: basepairs present (+) or absent (-) in the crystal structure
CC: color-coded covariation ratings for basepairs (red for confident basepairs; green, black, grey, or blue for less confident basepairs)
bpT: basepair type
bpC: basepair conformations, Watson-Crick (WC), wobble (Wb), slipped Watson-Crick (sWC), slipped wobble (sWb), reversed Watson-Crick (rWC), reversed wobble (rWb), Hoogsteen (H), reversed Hoogsteen (rH), sheared (S), reversed sheared (rS), parallel slipped sheared (pS), parallel sheared (pS), and reversed parallel sheared (rpS)
RS: rasmol scripts for visualizing basepairs
Motifs: motifs that basepairs are involved in (A for AA.AG@helix.ends motifs, GA for tandem GA motifs, E for AGUA/GAA motifs, EL for GUA/GAA motifs, ELL for GGA/GAA motifs, and LPTL for lonepair triloop motifs)
BT: basepairs involved in base triples
BQ: basepairs involved in base quadruples
CA: comparative information (CA for covariation-based basepairs and TT for tentatives)
Appendix D

Comparative Structure Model of the *T. thermophilus* 16S rRNA
Appendix E

Comparative Model of the *H. marismortui* 23S rRNA: 5′-half

*Haloarcula marismortui* rRNA
(AF034620)
Appendix F

Comparative Model of the *H. marismortui* 23S rRNA: 3' -half
### Appendix G

**Correspondence of the Lee-Gutell and Three Other Naming Systems for Classifying Basepair Conformations**

<table>
<thead>
<tr>
<th></th>
<th>Lee-Gutell</th>
<th>Common designation</th>
<th>Leontis &amp; Westhof</th>
<th>Saenger</th>
</tr>
</thead>
<tbody>
<tr>
<td>C:G</td>
<td>WC</td>
<td>GC Watson-Crick</td>
<td>GC cis WC/WC</td>
<td>XIX</td>
</tr>
<tr>
<td></td>
<td>Wb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sWC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sWb</td>
<td>#</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rWC</td>
<td>CG reverse Watson-Crick</td>
<td>GC trans WC/WC</td>
<td>XXII</td>
</tr>
<tr>
<td></td>
<td>rWb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>GC carbonyl-amino</td>
<td>C’G cis WC/H</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rH</td>
<td></td>
<td>C’G trans WC/H</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>GC N3-amino, amino-N3</td>
<td>CG trans WC/S</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U:A</td>
<td>WC</td>
<td>AU Watson-Crick</td>
<td>AU cis WC/WC</td>
<td>XX</td>
</tr>
<tr>
<td></td>
<td>Wb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sWC</td>
<td>AU amino-2-carbonyl</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sWb</td>
<td>#</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rWC</td>
<td>AU reverse Watson-Crick</td>
<td>AU trans WC/WC</td>
<td>XXI</td>
</tr>
<tr>
<td></td>
<td>rWb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>AU Hoogsteen</td>
<td>UA cis WC/H</td>
<td>XXIII</td>
</tr>
<tr>
<td></td>
<td>rH</td>
<td>AU reverse Hoogsteen</td>
<td>UA trans WC/H</td>
<td>XXIV</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>AU amino-2-carbonyl</td>
<td>AU cis WC/S</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U:G</td>
<td>WC</td>
<td>GU wobble</td>
<td>GU cis WC/WC</td>
<td>XVIII</td>
</tr>
<tr>
<td></td>
<td>Wb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sWC</td>
<td>#</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sWb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rWC</td>
<td>GU reverse wobble</td>
<td>GU trans WC/WC</td>
<td>XVII</td>
</tr>
<tr>
<td></td>
<td>rWb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G:A</td>
<td>WC</td>
<td>GA imino</td>
<td>AG cis WC/WC</td>
<td>VIII</td>
</tr>
<tr>
<td></td>
<td>Wb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sWC</td>
<td>#</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sWb</td>
<td>#</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rWC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rWb</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

234
<table>
<thead>
<tr>
<th></th>
<th>GA N1-N7, carbonyl-amino</th>
<th>A’G cis WC/H</th>
<th>IX</th>
</tr>
</thead>
<tbody>
<tr>
<td>rH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>GA sheared</td>
<td>AG trans H/S</td>
<td>XI</td>
</tr>
<tr>
<td>rS</td>
<td></td>
<td>AG trans S/S</td>
<td></td>
</tr>
<tr>
<td>fS</td>
<td>GA N3-amino, amino-N1</td>
<td>AG trans WC/S</td>
<td>X</td>
</tr>
<tr>
<td>pFS</td>
<td>GA N3-amino (1 bond)</td>
<td>AG cis WC/S</td>
<td></td>
</tr>
<tr>
<td>pS</td>
<td></td>
<td>AG cis S/S</td>
<td></td>
</tr>
<tr>
<td>rpS</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C:A</th>
<th>WC</th>
<th>A’C cis WC/WC</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wb</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sWC</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>sWb</td>
<td>AC N1-amino (1 bond)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rWC</td>
<td>AC reverse wobble</td>
<td>AC trans WC/WC</td>
</tr>
<tr>
<td></td>
<td>rWb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>AC reverse Hoogsteen</td>
<td>CA trans WC/H</td>
<td>XXV</td>
</tr>
<tr>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rFS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rps</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>U:C</th>
<th>WC</th>
<th>UC cis WC/WC</th>
<th>XVIII</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wb</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sWC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sWb</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>rWC</td>
<td>UC 2-carbonyl-imino</td>
<td>CU trans WC/WC</td>
</tr>
<tr>
<td></td>
<td>rWb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>UC 2-carbonyl-amino (1 bond)</td>
<td>CU trans H/S</td>
<td></td>
</tr>
<tr>
<td>fS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rFS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rps</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A:A</th>
<th>WC</th>
<th>AA cis WC/WC</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wb</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sWC</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>sWb</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rWC</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>rWb</td>
<td>AA N1-amino, symmetric</td>
<td>AA trans WC/WC</td>
</tr>
<tr>
<td>H</td>
<td>AA N7-amino (1 bond)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rH</td>
<td></td>
<td>AA trans WC/H</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>AA N3-amino</td>
<td>AA trans H/S</td>
<td></td>
</tr>
<tr>
<td>fS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pFS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rpS</td>
<td></td>
<td>#</td>
<td>#</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C:C</th>
<th>WC</th>
<th>CC cis WC/WC</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Wb</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sWC</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>sWb</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rWC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rWb</td>
<td>CC N3-amino, symmetric</td>
<td>CC trans WC/WC</td>
</tr>
<tr>
<td>rWC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rWb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base Pair</td>
<td>WC</td>
<td>#</td>
<td>WC</td>
</tr>
<tr>
<td>-----------</td>
<td>----</td>
<td>---</td>
<td>----</td>
</tr>
<tr>
<td>G:G</td>
<td>Wb</td>
<td>GG carbonyl-imino (1 bond)</td>
<td>Gg cis WC/WC</td>
</tr>
<tr>
<td></td>
<td>sWb</td>
<td>#</td>
<td>#</td>
</tr>
<tr>
<td></td>
<td>rWC</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>rWb</td>
<td>GG N1-carbonyl, symmetric</td>
<td>GG trans WC/WC</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>GG N1-carbonyl, N7-amino</td>
<td>GG cis WC/H</td>
</tr>
<tr>
<td></td>
<td>rH</td>
<td>GG N7-imino</td>
<td>GG trans WC/H</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>rS</td>
<td>GG N3-amino, symmetric</td>
<td>GG trans S/S</td>
</tr>
<tr>
<td></td>
<td>fS</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>pfS</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>pS</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>rpS</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Base Pair</th>
<th>WC</th>
<th>#</th>
<th>WC</th>
<th>#</th>
<th>WC</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>U:U</td>
<td>Wb</td>
<td>UU imino-carbonyl</td>
<td>UU cis WC/WC</td>
<td>XVI</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sWb</td>
<td>#</td>
<td>#</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rWC</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rWb</td>
<td>UU 4-carbonyl-imino, symmetric</td>
<td>UU trans WC/WC</td>
<td>XII, XIII</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>-</td>
<td>UU cis WC/H</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rH</td>
<td>UU 4-carbonyl-C5H, imino-4-carbonyl</td>
<td>UU trans WC/H</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rS</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>fS</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pfS</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pS</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rpS</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

† The pound sign (#) represents the basepair conformations that are unlikely to form.
Appendix H

AA.AG@helix.ends Basepairs in 16S and 23S rRNAs

Escherichia coli

(310695)

Red nucleotides: base pairs and motif pairings present in crystal structure.
Blue nucleotides: base pairs and motifs predicted by comparative analysis.
Non-homologous or non-termini not present in crystal structure.
Present in crystal structure.
Absent from crystal structure.
Exchange patterns between AA AND AG are marked with:

* | Si gni li cant exchange i n all ggment wt th mi ni ccept one.
** | Si gni li cant exchange i n at least one al ggment wt th mi ni ccept one, but wt th more ccept one i n at least one other al ggment.
*** | Exchange i n at least one subal ggment, exclu si ng mi locontid i.

(none) | Invar ant AA OR AG.

Coaxal stacki ng possi ble li li es for crystal structure i nteracti ons adj acent to AA. Agi gnet i.

Adj acent i n crystal structure.
Blue nucleotides: base pairs and motif pairings present in crystal structure.

Exchange patterns between AA AND AG are marked with *.

Exchange patterns between AA AND AG are marked with *.

Exchange patterns between AA AND AG are marked with *.

Exchange patterns between AA AND AG are marked with *.

Coaxial stacking possibilities for crystal structure: (Haloarcula marismortui) (AF034620).

interactions adjacent to AA.AG@helix.ends

Coaxial stacking present in crystal structure.

Coaxial stacking absent from crystal structure.
## Appendix I

### Identified Sticky Motifs in rRNAs and their Sequence Variations in Aligned Sequences

<table>
<thead>
<tr>
<th>rRNA</th>
<th>Ec*</th>
<th>Xtal*</th>
<th>Motif*</th>
<th>A*</th>
<th>B*</th>
<th>C*</th>
<th>E*</th>
<th>M*</th>
</tr>
</thead>
<tbody>
<tr>
<td>16S</td>
<td>68i</td>
<td>68i</td>
<td>ELL</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>69i</td>
<td>69i</td>
<td>E.ELL</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>70i</td>
<td>70i</td>
<td>E.ELL</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>73i</td>
<td>73i</td>
<td>E.ELL</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>74i</td>
<td>74i</td>
<td>ELL</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>413</td>
<td>408</td>
<td>EL</td>
<td>gap</td>
<td>237</td>
<td>GAA/GUA</td>
<td>370</td>
<td>GAA/GUA</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GAA/GUA</td>
<td>4</td>
<td>GAA/GUA</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>gap</td>
<td>1029</td>
<td>Others</td>
<td>51</td>
</tr>
<tr>
<td>447</td>
<td>441</td>
<td>E</td>
<td>gap</td>
<td>237</td>
<td>GAA/GGUAA</td>
<td>184</td>
<td>GAA/GGUAA</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GAA/GGUAA</td>
<td>91</td>
<td>GAA/GGUAA</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AAA/GGUAA</td>
<td>139</td>
<td>AAA/GGUAA</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GAA/GGUAA</td>
<td>132</td>
<td>GAA/GGUAA</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AAA/GGUAA</td>
<td>122</td>
<td>AAA/GGUAA</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GAA/GGUAA</td>
<td>100</td>
<td>GAA/GGUAA</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GAA/GGUAA</td>
<td>70</td>
<td>GAA/GGUAA</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AAG/GGUAA</td>
<td>33</td>
<td>AAG/GGUAA</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GAG/GGUAA</td>
<td>22</td>
<td>GAG/GGUAA</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CAA/GGUAA</td>
<td>16</td>
<td>CAA/GGUAA</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AAG/GGUAA</td>
<td>16</td>
<td>AAG/GGUAA</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GAC/GGUAA</td>
<td>8</td>
<td>GAC/GGUAA</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AAA/GGUAA</td>
<td>7</td>
<td>AAA/GGUAA</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GAG/GGUAA</td>
<td>5</td>
<td>GAG/GGUAA</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Others</td>
<td>318</td>
<td>Others</td>
<td>46</td>
</tr>
<tr>
<td>581</td>
<td>564</td>
<td>EL</td>
<td>gap</td>
<td>205</td>
<td>GUA/GAG</td>
<td>278</td>
<td>GUA/GAG</td>
<td>114</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GUA/GAG</td>
<td>6</td>
<td>GUA/GAG</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GUA/GAG</td>
<td>4</td>
<td>GUA/GAG</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GUA/GAG</td>
<td>2</td>
<td>GUA/GAG</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GUA/GAG</td>
<td>1</td>
<td>GUA/GAG</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GUA/GAG</td>
<td>1</td>
<td>GUA/GAG</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GUA/GAG</td>
<td>1</td>
<td>GUA/GAG</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GUA/GAG</td>
<td>1</td>
<td>GUA/GAG</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Others</td>
<td>1138</td>
<td>Others</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>gap</td>
<td>40</td>
<td>gap</td>
<td>1</td>
</tr>
<tr>
<td>661*</td>
<td>644*</td>
<td>ELL,WC</td>
<td>GGA/GAA</td>
<td>177</td>
<td>GGA/GAA</td>
<td>157</td>
<td>GGA/GAA</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>GAA/GAG</td>
<td>133</td>
<td>GGA/GAA</td>
<td>7</td>
<td>GGA/GAA</td>
<td>167</td>
<td>GGA/GAA</td>
<td>171</td>
</tr>
<tr>
<td></td>
<td>GAA/GAG</td>
<td>111</td>
<td>GGA/GAA</td>
<td>2</td>
<td>GGA/GAA</td>
<td>1022</td>
<td>GGA/GAA</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>GAA/GAG</td>
<td>8</td>
<td>GGA/GAA</td>
<td>2</td>
<td>GGA/GAA</td>
<td>2</td>
<td>GGA/GAA</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>GAA/GAG</td>
<td>6</td>
<td>GGA/GAA</td>
<td>1</td>
<td>GGA/GAA</td>
<td>1</td>
<td>GGA/GAA</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Others*</td>
<td>535</td>
<td>Others*</td>
<td>4559</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>685 668</td>
<td>E.LLL</td>
<td>GUA/AAU</td>
<td>141 GUA/GAU</td>
<td>313 GGA/GAG</td>
<td>68 UCA/GAU</td>
<td>625 GUA/GAU</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>765 748</td>
<td>EL</td>
<td>GAA/GUA</td>
<td>196 GAA/GUA</td>
<td>369 GAA/GUA</td>
<td>157 GAA/AUA</td>
<td>858 GAA/CUA</td>
<td>141</td>
<td></td>
</tr>
<tr>
<td>780 763</td>
<td>EL</td>
<td>GAA/GUA</td>
<td>146 GAA/GUA</td>
<td>3272 GAA/GUA</td>
<td>85 GAA/GUA</td>
<td>734 AAG/GUA</td>
<td>709</td>
<td></td>
</tr>
<tr>
<td>889 866</td>
<td>E</td>
<td>AGU/GAA</td>
<td>206 AGU/GAA</td>
<td>2529 AGU/GAA</td>
<td>151 AGU/GAA</td>
<td>961 AGU/GAA</td>
<td>808</td>
<td></td>
</tr>
<tr>
<td>1238 1219</td>
<td>E</td>
<td>AA/AGU</td>
<td>200 AA/AGU</td>
<td>4450 AA/AGU</td>
<td>153 ACU/GCU</td>
<td>441 AA/AGU</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>1260 1241</td>
<td>ELL.GA</td>
<td>GAC/GGA*</td>
<td>62 GAC/GGA*</td>
<td>580 GAC/GGA*</td>
<td>28 GAC/GGA*</td>
<td>41 GAC/GGA*</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>1304 1285</td>
<td>ELL</td>
<td>GGA/GAA</td>
<td>185 GGA/GAA</td>
<td>4370 GGA/GAA</td>
<td>136 GGA/GAA</td>
<td>565 GGA/GAA</td>
<td>382</td>
<td></td>
</tr>
</tbody>
</table>

*GAP*
<table>
<thead>
<tr>
<th>GGA/GAA</th>
<th>GGA/AAA</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>16</td>
<td>17</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

1346 1327 E

<table>
<thead>
<tr>
<th>AGUA/GAA</th>
<th>AGUA/GGA</th>
<th>AGUA/GAG</th>
<th>AGUA/AAA</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>203</td>
<td>237</td>
<td>6</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>426</td>
<td>446</td>
<td>1</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>426</td>
<td>446</td>
<td>1</td>
<td>2</td>
<td>11</td>
</tr>
</tbody>
</table>

1417 1399 ELL,GAG

<table>
<thead>
<tr>
<th>GAG/GUA*</th>
<th>GAG/GCA*</th>
<th>GAG/CA*</th>
<th>GAA/AGUA*</th>
<th>GAA/GUA*</th>
</tr>
</thead>
<tbody>
<tr>
<td>79</td>
<td>36</td>
<td>20</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>121</td>
<td>551</td>
<td>477</td>
<td>311</td>
<td>6</td>
</tr>
<tr>
<td>121</td>
<td>551</td>
<td>477</td>
<td>311</td>
<td>6</td>
</tr>
</tbody>
</table>

1441 1423 ELL

<table>
<thead>
<tr>
<th>gap</th>
<th>gap</th>
<th>gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>GGA/GAA</td>
<td>GGA/AAA</td>
<td>GGA/GAA</td>
</tr>
<tr>
<td>34</td>
<td>237</td>
<td>44</td>
</tr>
<tr>
<td>gap</td>
<td>gap</td>
<td>gap</td>
</tr>
</tbody>
</table>

23S 26 23 ELL

<table>
<thead>
<tr>
<th>GGA/GAA</th>
<th>GGA/AAA</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>237</td>
<td>15</td>
</tr>
<tr>
<td>GGA/AAA</td>
<td>44</td>
<td>8</td>
</tr>
<tr>
<td>Others</td>
<td>2</td>
<td>10</td>
</tr>
</tbody>
</table>

82 78 ELL,GAG

<table>
<thead>
<tr>
<th>GGA/GAA</th>
<th>GGA/GAU</th>
<th>GUA/GAC</th>
<th>GUA/GAA</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>136</td>
<td>10</td>
<td>84</td>
<td>4</td>
</tr>
<tr>
<td>12</td>
<td>136</td>
<td>10</td>
<td>84</td>
<td>4</td>
</tr>
<tr>
<td>12</td>
<td>136</td>
<td>10</td>
<td>84</td>
<td>4</td>
</tr>
</tbody>
</table>

189 159 E

<table>
<thead>
<tr>
<th>GAA/GUA</th>
<th>GAA/GAU</th>
<th>CAA/GAA</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>266</td>
<td>76</td>
</tr>
<tr>
<td>90</td>
<td>90</td>
<td>90</td>
</tr>
</tbody>
</table>

241 212 E

<table>
<thead>
<tr>
<th>AGUA/GAA</th>
<th>AGUA/GAU</th>
<th>AGUA/GAC</th>
<th>AGUA/GAA</th>
<th>AGUA/GAA</th>
<th>AGUA/GAA</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>236</td>
<td>30</td>
<td>96</td>
<td>96</td>
<td>96</td>
</tr>
<tr>
<td>34</td>
<td>236</td>
<td>30</td>
<td>96</td>
<td>96</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>GAA/GCU</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>818</td>
<td>GAA/GUA</td>
<td>22</td>
<td>GAA/AGUA</td>
<td>223</td>
<td>GAA/AGUA</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td>1</td>
<td>gap</td>
<td>239</td>
<td>gap</td>
</tr>
<tr>
<td>844</td>
<td>GGA/AGA</td>
<td>13</td>
<td>Others</td>
<td>84</td>
<td>Others</td>
</tr>
<tr>
<td></td>
<td>gap</td>
<td>7</td>
<td>gap</td>
<td>41</td>
<td>gap</td>
</tr>
<tr>
<td>858</td>
<td>AGUA/GAA</td>
<td>32</td>
<td>AGUA/AGA</td>
<td>241</td>
<td>AGUA/AGA</td>
</tr>
<tr>
<td></td>
<td>AGUA/AAA</td>
<td>5</td>
<td>AGUA/AAA</td>
<td>3</td>
<td>AGUA/AAA</td>
</tr>
<tr>
<td></td>
<td>gap</td>
<td>1</td>
<td>gap</td>
<td>159</td>
<td>gap</td>
</tr>
<tr>
<td>1204</td>
<td>ELL</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1212</td>
<td>GAA/GGA</td>
<td>37</td>
<td>GAA/GGA</td>
<td>254</td>
<td>GAA/GGA</td>
</tr>
<tr>
<td></td>
<td>AAG/GGA</td>
<td>4</td>
<td>AAG/GGA</td>
<td>3</td>
<td>AAG/GGA</td>
</tr>
<tr>
<td></td>
<td>AAA/GGA</td>
<td>2</td>
<td>AAA/GGA</td>
<td>3</td>
<td>AAA/GGA</td>
</tr>
<tr>
<td></td>
<td>gap</td>
<td>1</td>
<td>gap</td>
<td>159</td>
<td>gap</td>
</tr>
<tr>
<td>1265</td>
<td>AGUA/GAA</td>
<td>41</td>
<td>AGUA/GAA</td>
<td>235</td>
<td>AGUA/GAA</td>
</tr>
<tr>
<td></td>
<td>AGUA/AAA</td>
<td>1</td>
<td>AGUA/AAA</td>
<td>1</td>
<td>AGUA/AAA</td>
</tr>
<tr>
<td></td>
<td>gap</td>
<td>2</td>
<td>gap</td>
<td>1</td>
<td>gap</td>
</tr>
<tr>
<td>1418</td>
<td>1526 ELL</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1438</td>
<td>1543 ELL</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1475</td>
<td>GAA/GAA</td>
<td>11</td>
<td>GAA/GAA</td>
<td>10</td>
<td>GAA/GAA</td>
</tr>
<tr>
<td></td>
<td>gap</td>
<td>94</td>
<td>gap</td>
<td>25</td>
<td>Others</td>
</tr>
<tr>
<td>1484</td>
<td>GGA/GAA</td>
<td>40</td>
<td>GGA/GAA</td>
<td>235</td>
<td>GGA/GAA</td>
</tr>
<tr>
<td></td>
<td>gap</td>
<td>2</td>
<td>gap</td>
<td>25</td>
<td>gap</td>
</tr>
<tr>
<td>1526</td>
<td>AGC/GAU</td>
<td>18</td>
<td>AGC/GAU</td>
<td>98</td>
<td>AGC/GAU</td>
</tr>
<tr>
<td></td>
<td>gap</td>
<td>4</td>
<td>gap</td>
<td>2</td>
<td>gap</td>
</tr>
<tr>
<td>1529</td>
<td>AGC/GAU</td>
<td>40</td>
<td>AGC/GAU</td>
<td>40</td>
<td>AGC/GAU</td>
</tr>
<tr>
<td>1530</td>
<td>AGC/GAU</td>
<td>7</td>
<td>AGC/GAU</td>
<td>7</td>
<td>AGC/GAU</td>
</tr>
<tr>
<td></td>
<td>gap</td>
<td>12</td>
<td>gap</td>
<td>72</td>
<td>gap</td>
</tr>
</tbody>
</table>

244
<table>
<thead>
<tr>
<th></th>
<th>AGUA/GAA</th>
<th>AGUA/GAC</th>
<th>AGUA/GAU</th>
<th>UGUA/GAU</th>
<th>UGUA/GAC</th>
<th>UGUA/GAA</th>
<th>Others</th>
<th>gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>2654</td>
<td>AGUA/GAA 4</td>
<td>AGUA/GAC 234</td>
<td>AGUA/GAU 4</td>
<td>UGUA/GAU 10</td>
<td>UGUA/GAC 2</td>
<td>UGUA/GAA 14</td>
<td>Others 87</td>
<td>gap 16</td>
</tr>
<tr>
<td>2691</td>
<td>AGUA/GAA 132</td>
<td>AGUA/GAC 159</td>
<td>AGUA/GAU 56</td>
<td>UGUA/GAU 15</td>
<td>UGUA/GAC 4</td>
<td>UGUA/GAA 16</td>
<td>Others 2</td>
<td>gap 1</td>
</tr>
<tr>
<td>5S</td>
<td>AGUA/GAA 76</td>
<td>AGUA/GAG 190</td>
<td>AGUA/GAA 297</td>
<td>Others 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>AGUA/GAA 15</td>
<td>AGUA/GAG 60</td>
<td>AGUA/GAA 70</td>
<td>Others 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>AGUA/GAA 15</td>
<td>AGUA/GAG 60</td>
<td>AGUA/GAA 70</td>
<td>Others 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E.EL</td>
<td>AGUA/GAA 5</td>
<td>AGUA/GAG 21</td>
<td>AGUA/GAA 9</td>
<td>Others 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*: The lowest E. coli-equivalent position numbers in the sequence motifs and their respective positions for the rRNAs in the T. thermophilus 30S and H. marismortui 50S crystal structures.

**: Sequence motifs: E, AGUA/GAA; EL, GUA/GAA; ELL, GGA/GAA motif.

***: Phylogenetic groups: A, Archaea; B, Bacteria; E, Eukaryotes; C, Chloroplasts; M, Mitochondria.

The asterisk (*) marks represent the sequences which form either tandem GA motifs or canonical basepairs.
Appendix J

Gallery of Sticky Motifs in the Variable Regions of rRNAs

Encephalotaeniun hellem
(AF118143)
1. Bacteria 2. Fungi/Metazoa
3. Fungi 4. Mollicutes
16S: 68i

Mycobacterium xanthus
(M34114)
1. Bacteria 2. Proteobacteria
3. Delta/Epsilon subdivisions
16S: 69i

Geodoga subterranea
(L10659)
1. Bacteria 2. Thermotogae
3. Thermotogae
16S: 70i

Thermotoga maritima
(AE031703)
1. Bacteria 2. Thermotogae
3. Thermotogae
16S: 73i

Erysipelothrix rhusiopathiae
(M22728)
1. Bacteria 2. Firmicutes
3. Mollicutes
16S: 74i

Helobacterium chlorum
(M11312)
1. Bacteria 2. Firmicutes
3. Clostridia
16S: 76i

Vibrio parahaemolyticus
(XC6530)
1. Bacteria 2. Proteobacteria
3. Gammaproteobacteria
16S: 82i

Thermus thermophilus
(M26923)
1. Bacteria 2. Deinococcus/Thermus
3. Thermus
16S: 413
Mycoplasma pneumoniae
(U00089)
1. Bacteria 2. Firmicutes
3. Bacillus/Clostridium

23S: 1530i

Campylobacter jejuni
(AL139074)
1. Bacteria 2. Proteobacteria
3. Delta/Epsilon subdivisions

23S: 1531i

Leptospira interrogans
(X14249)
1. Bacteria 2. Spirochaetes
3. Spirochaetes

23S: 1710

Pseudomonas aeruginosa
(Y00432)
1. Bacteria 2. Proteobacteria
3. Gammaproteobacteria

23S: 1721

Thermotoga maritima
(M67498)
1. Bacteria 2. Thermotogae
3. Thermotogae

23S: 1727i
Appendix K

Lonepair Triloop Motifs Occurring in Three Phylogenetic Domains and Two Organelles

Reference sequence and structure: Escherichia coli (J01695)

Three Phylogenetic Domains / Two Organelles

Positions with a nucleotide in more than 95% of the sequences are shown in one of four categories:

- ACGU - 98-100% conserved
- acgu - 90-98% conserved
- ac - less than 90% conserved

Otherwise, the regions are represented by arcs.

Arc labels indicate the upper and lower number of nucleotides known to exist within the associated variable region.

Blue tags indicate insertions relative to the reference sequence.

In insertions that appear:
1. length 1-4 in more than 10% of sequences
2. length 5 or greater in at least one sequence with format:
   (Max. Length of Insertion:Percentage of seqs with any length insertion)

Citation and related information available at http://www.rna.icmb.utexas.edu
Three Phylogenetic Domains / Two Organelles

Reference sequence and structure: Escherichia coli (J01695)

1. cellular organisms
2. virions
3. cell-free viruses
4. viroids

July 2021

Number of sequences: 930

Positions with a nucleotide in more than 90% of the sequences are shown in one of four categories:

- 99-100% conserved
- 80-90% conserved
- 60-80% conserved
- less than 60% conserved

Otherwise, the regions are represented by arcs.

Arc labels indicate the upper and lower number of nucleotides known to exist within the associated variable region.

Blue tags indicate insertions relative to the reference sequence.

Insertions that appear are:

1. length 1-4 in more than 10% of sequences
2. length 5 or greater in at least one sequence

With format:

(Max. Length of Insertion: Percentage of seqs with any length insertion)

Citations and related information available at http://www.rna.icmb.utexas.edu
Three Phylogenetic Domains / Two Organelles

Reference sequence and structure: Escherichia coli (J01695)

October 2001 - 98+% conserved
- 90-98% conserved
- 80-90% conserved
- less than 80% conserved

ACGU - 96% conserved
acgu - 90-96% conserved
acg - 80-90% conserved
a - less than 80% conserved

Otherwise, the regions are represented by arcs.
Arc labels indicate the upper and lower number of nucleotides known to exist within the associated variable region.

Positions with a nucleotide in more than 95% of the sequences are shown in one of four categories:

ACGU - 96% conserved
acgu - 90-96% conserved
acg - 80-90% conserved
a - less than 80% conserved

Other nucleotides are shown in one of four categories:

Citation and related information available at http://www.rna.icmb.utexas.edu

Number of sequences: 930

Blue tags indicate insertions relative to the reference sequence.
Insertions that appear are:
(1) length 1-4 in more than 10% of sequences
(2) length 5 or greater in at least one sequence

with format:
(Max. Length of Insertion:Percentage of seqs with any length insertion)

Insertions that appear are:
(1) length 1-4 in more than 10% of sequences
(2) length 5 or greater in at least one sequence

with format:
(Max. Length of Insertion:Percentage of seqs with any length insertion)


Features in the Ternary Complex and in tRNA. *Structure Fold. Des.* 7, 143-156.


268
Vita

Jung Chull Lee was born in Seoul, Korea on April 18th, 1965, as the son of Won Jae Lee and Hee Nam Sung, and graduated from a high school in 1985. He received his B.S. (*Summa Cum Laude*) in Chemistry from Kyung Hee University in 1989 (fully supported by the University Honor Student Fellowship for four years) and his M.S. in Quantum and Computational Chemistry from Korea Advanced Institute of Science and Technology (KAIST) in 1991 (fully supported by the Korean Government Honor Student Fellowship for two years), followed by working as a Research Scientist in Computer-Aided *de novo* Drug Design at Korea Institute of Science and Technology (KIST) until he was admitted to Medicinal Chemistry, College of Pharmacy, The University of Texas at Austin, in 1998. He has been married to Hee Young Jung since 1996, with one son, Young Lee, and one daughter, Gene Lee. He joined Dr. Robin Gutell’s lab in 2000 and studied Bioinformatics and Computational Biology.

Publications:


Permanent address: 3361 Lake Austin Blvd. #D
Austin, TX 78703

This dissertation was typed by the author.