

**COMPARATIVE ANALYSES AND STRUCTURAL MODELING OF RNAS
PARTICIPATING IN NON-CANONICAL INITIATION OF PROTEIN
SYNTHESIS**

by

Jody M. Burks

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
December 13, 2010

Keywords: Computational biology, RNA structure, tmRNA,
IRES RNA, bioinformatics, molecular modeling

Copyright 2010 by Jody M. Burks

Approved by

Jacek Wower, Co-chair, Professor of Biochemistry
Sang-Jin Suh, Co-chair, Associate Professor of Microbiology
Werner G. Bergen, University Reader, Professor of Biochemistry
Doug Goodwin, Associate Professor of Biochemistry
Christian Zwieb, Professor of Molecular Biology

ABSTRACT

Transfer-messenger RNA and the Internal Ribosomal Entry Site (IRES) RNAs of Foot and Mouth Disease Virus (FMDV) and Bovine Viral Diarrhea Virus (BVDV) play important roles in *trans*-translation and cap-independent protein synthesis. TmRNA sequences were examined using covariation analysis. Its secondary structure diagram was improved and three-dimensional molecular models were produced. These models were essential for a detailed interpretation of cryo-EM images of tmRNA as it binds to the ribosome. Subsequently, FMDV and BVDV IRES RNAs were investigated *in silico* and their secondary structures were determined. The modeling method for investigating tertiary structures of RNA molecules was expanded to include RNA-protein interactions captured in high resolution and then used to build biologically relevant structural models of FMDV and BVDV IRES RNAs both off and on the ribosome. Modeling of tmRNA and the IRES RNA structures provided important insights into non-canonical initiation of protein synthesis in prokaryotes and eukaryotes. Moreover, this project generated data essential for developing novel antiviral drugs that will be able to inhibit binding of IRES RNAs to the ribosome.

ACKNOWLEDGEMENTS

So many have encouraged me on this journey! First of all I would like to thank my loving parents, Larry and Janet Burks, for their support and patience with me during the course of my research and dissertation preparation.

This dissertation is dedicated in loving memory of Ms. Sushma Rao, M.S., P.A., who was an inspiration to many and sincerely missed by all her friends and family, my late grandparents, Mrs. Elsie T. Baker Diebler, an amazing, ahead-of-her-time woman who insisted on my first computer when I was six and heavily influenced my decision to study science, Mr. Melvin H. Diebler, for his advice and encouragement through the years, and Mr. Joe J. and Mrs. Ruth M. Burks, who supported and encouraged me in my pursuits of undergraduate and graduate studies, and in honor of my grandfather, Mr. Arvel “Art” L. Baker, from whom I inherited my multi-processing and three-dimensional thinking abilities.

I would also like to acknowledge many of my teachers. Mrs. Willene Lilly, retired teacher of computer science at Lindale Junior High School, Lindale, TX for turning me loose with computers and programming in 8th grade. I would also like to thank Dr. Manus Donahue, retired Professor of Biological Sciences and Assistant Dean of Texas Academy of Mathematics and Science (TAMS), and Dr. Richard Sinclair, Dean of TAMS at the University of North Texas for the opportunity to begin my undergraduate science education two years early through the TAMS program. Moreover, I would like to thank

Dr. Beverly Clement for her introduction to research, patience and encouragement during my undergraduate studies.

I would like to gratefully acknowledge that this work would absolutely not be possible without the unwavering support of my mentors, Drs. Jacek Wower and Christian Zwieb, who patiently helped me to learn how to carry out my research and how to publish the results of my work. I would also like to thank Drs. Sang-Jin Suh and Douglas Goodwin for their advice, humor and perspective. Thank you very much also to Dr. Werner G. Bergen, who served as my University Reader and critiqued my dissertation.

I would like to thank Auburn University and the Auburn University Graduate School for bestowing on me the “Top Ten Graduate Students Award” in 2005 and for the opportunities to join The Honor Society of Phi Kappa Phi, Gamma Sigma Delta Honor Society of Agriculture, Golden Key International Honour Society and Delta Epsilon Iota Academic Honor Society.

I would like to thank all of the people who have worked in the Wower and Zwieb labs for their scientific help, discussions and for making the labs enjoyable places to work and learn, and fun places to be.

My research was supported by grants from the Alabama Agricultural Experiment Station Foundation and the Alabama Cattlemen's Association to Dr. Iwona K. Wower and Dr. Jacek Wower, an Auburn University Biogrant to Dr. Jacek Wower, and was supported the National Science Foundation Grants No. 0091853 and NSF-EPS 0447675 to Dr. Frank F. “Skip” Bartol, the Director of the Auburn University Cellular and Molecular Biology Program. Publication costs were covered by the Upchurch Fund for Excellence.

TABLE OF CONTENTS

Abstract	ii
Acknowledgements	iii
List of Tables	x
List of Figures	xi
Chapter 1: Introduction	1
RNA: A point of interest	2
Studies of RNA Structures	3
Structural Modeling and RNA Structure Prediction	5
Protein Synthesis: Translation of mRNA Into Protein	7
The Ribosome: A Molecular Machine	8
Transfer-messenger RNA and <i>Trans</i> -Translation	10
Internal Ribosomal Entry Site-mediated Translation Initiation	11
Foot and Mouth Disease Virus IRES RNA	12
Bovine Viral Diarrhea Virus IRES RNA	14
Concluding Remarks	16
References	17
Chapter 2: The tmRDB and SRPDB Resources	37
Abstract	38
Introduction	39
Results and Discussion	40

tmRNA Genes	40
tmRNA-encoded Tag-peptides	41
tmRNA-associated Proteins	41
Ribosomal protein S1	42
Alanyl-tRNA Synthetase	42
EF-Tu	43
Phylogeny of tmRNP	43
Phylogeny of SRP RNA genes	44
SRP RNA features	44
SRP Proteins	46
SRP9, SRP14 and SRP21	46
SRP19	46
SRP54	46
SRP68 and SRP72	47
cpSRP43	47
SRP-associated proteins	48
SRP Receptor (alpha) (FtsY)	48
SRP Receptor (beta)	48
FlhF	48
Phylogeny of SRP	48
Outlook	49
Access	49
Materials and Methods	50
Comparative Sequence Analysis of RNA	50
Protein Alignments	50

Alignment Browser	51
Acknowledgements	51
References	52
Chapter 3: Comparative 3-D Modeling of tmRNA	60
Abstract	61
Introduction	62
Results	63
Identification of tmRNA sequences	63
Selection of tmRNA Sequences	64
Comparative Sequence Analysis	65
Quality Control of Sequence Information	66
tmRNA Alignment	66
Secondary Structure of tmRNA	67
TLD (helices 1, 2a and 12)	67
Helical sections 2b, 2c and 2d	68
Pseudoknot 1 (helices 3 and 4)	69
The mRNA-like region (MLR)	70
Pseudoknot 2 (helices 6 and 7)	70
Pseudoknot 3 (helices 8 and 9)	71
Pseudoknot 4 (helices 10 and 11)	71
Secondary Structure Prediction of the MLR	71
Secondary Structures of <i>E. coli</i> tmRNA	72
Tertiary Structure Modeling and Visualization of tmRNA	72
3-D Model of <i>E. coli</i> tmRNA	75
Discussion	78

Original Conclusions	81
Future Directions and Evolution of the <i>E. coli</i> tmRNA Model	81
Methods	82
Comparative Sequence Analysis	82
3-D Model Building	83
References	85
CHAPTER 4: In Silico Analysis of IRES RNAs of Foot-and-Mouth Disease Virus and Related Picornaviruses	97
Abstract	98
Introduction	99
Results and Discussion	101
Comparative Sequence Analysis	101
Secondary Structure of the FMDV IRES RNA	102
Domain 2	103
Domain 3	103
Domain 4	105
Domain 5 and the 3' region of the IRES RNA	106
Distribution of Conserved Elements In FMDV IRES RNA Secondary Structure	107
Modeling the Three-Dimensional Structure of FMDV IRES RNA	107
Modeling Constraints	108
Three-Dimensional FMDV IRES RNA Model	109
Topography of the IRES Ribonucleoprotein Complex	111
FMDV IRES RNA on the 40S Ribosomal Subunit	112
Conclusions	114
Methods	115

Collection and Cataloguing of Available Type II Picornavirus IRES Sequences	115
Comparative Sequence Analysis	115
Three-Dimensional Molecular Modeling	116
Acknowledgements	116
References	117
Chapter 5: Comparative Structural Studies of Bovine Viral Diarrhea Virus IRES RNA .	135
Abstract	136
Introduction	137
Results and Discussion	139
Identification of BVDV IRES RNA sequences	139
Comparative Sequence Analysis (CSA) of BVDV IRES RNAs	140
BVDV IRES RNA Secondary Structure	142
Helix 2	143
Helix 3	144
Helix 4	149
BVDV IRES RNA pseudoknot	149
A Three-Dimensional Model of the BVDV IRES	150
BVDV IRES RNA on the Human 40S Ribosomal Subunit	152
Conclusions	153
Methods	154
Comparative Analysis of BVDV IRES RNA Sequences	154
Molecular Modeling of IRES RNAs	154
Acknowledgements	155
References	157

LIST OF TABLES

Table 2.1 tmRNA Features and Representatives	56
Table 2.2 SRP RNA features and SRP components ordered by phylogeny	58
Table 3.1 Phylogenetic distribution of tmRNA features	95
Table 3.2 Structural motifs used for the <i>Escherichia coli</i> tmRNA model	96
Table 4.1 Summary of biochemical data used as modeling constraints	131
Table 4.2 Features used to model the FMDV IRES RNA	133
Table 5.1 Mutations and compensatory changes of the proposed base pairs of the BVDV or CSFV IRES RNA pseudoknots and their effects on IRES-mediated translation relative to wild-type	168
Table 5.2 Expected modifications in BVDV IRES RNA based on enzymatic modifications of wild-type CSFV IRES RNA due to binding of the rabbit reticulocyte 40S ribosomal subunit	169
Table 5.3 Table of the features used for generating a three-dimensional model of the BVDV IRES RNA	170
Table 5.4 Table of the features used for generating a three-dimensional model of the HCV IRES RNA	171

LIST OF FIGURES

Figure 1.1 Central Dogma of Molecular Biology, then and now	29
Figure 1.2 Ribosomal subunit and ribosome structures	30
Figure 1.3 Flow of information from sequence to tertiary structure model in this dissertation	32
Figure 1.4 tmRNA features and <i>trans</i> -translation mechanism	34
Figure 1.5 Secondary Structures of IRES RNAs	36
Figure 2.1 Schematic representation of the secondary structures of <i>Escherichia coli</i> tmRNA and SRP RNA	55
Figure 3.1 Secondary structure of <i>E. coli</i> tmRNA	90
Figure 3.2 Motif modeling procedure	91
Figure 3.3 3-D model of <i>Escherichia coli</i> tmRNA	92
Figure 3.4 Conformational changes in <i>Escherichia coli</i> tmRNA	94
Figure 4.1 Secondary structure diagram of FMDV O1K IRES RNA supported by CSA and biochemical data	126
Figure 4.2 Three-dimensional model of FMDV O1K IRES RNA	127
Figure 4.3 Protein protections in FMDV	128
Figure 4.4 Map of interactions between FMDV IRES RNA domains, initiation factors and the 40S ribosomal subunit	129
Figure 4.5 Placement of FMDV IRES RNA on the cryo-EM surface representation of the human 40S subunit	130
Figure 5.1 Secondary structure diagram of BVDV-1B strain Osloss IRES RNA	164
Figure 5.2 Secondary structure of helix 2 of BVDV-1b IRES RNA in comparison with those derived from NMR analysis of HCV and CSFV IRES RNAs	165

Figure 5.3 Three-dimensional IRES RNA Models166
Figure 5.4 Placement of the IRES RNA models on the 40S ribosomal subunit167

“Life is simply a matter of chemistry.” – James Watson (Nobel Prize, 1962)

CHAPTER 1: INTRODUCTION

RNA: A point of interest

The “Central Dogma of Molecular Biology” proposed by Francis Crick in 1958 (Crick, 1970; Crick, 1958) described the one-way flow of genetic information from DNA to RNA to protein in cellular systems (Figure 1.1A). This deterministic view was modified by later advancements in molecular biology (see Figure 1.1B). First, Baltimore and Temin independently demonstrated that genetic information could be transcribed from RNA into DNA by reverse transcriptase, an RNA-dependent DNA polymerase (Baltimore, 1970; Temin & Mizutani, 1970). Second, a number of studies revealed that many positive-sense, single-stranded RNA viruses (e.g. Poliovirus and Mengovirus) are able to replicate the RNA genome without the need of a DNA intermediate (Baltimore *et al.*, 1963; Baltimore & Franklin, 1963). This process is facilitated by RNA-dependent RNA polymerases, which have also recently been implicated in RNA silencing (Ahluquist, 2002). Third, the discovery of catalytic RNAs or “ribozymes” in the self-splicing of the Tetrahymena group I intron by Tom Cech and in maturation of tRNAs by Sidney Altman modified our understanding of the flow of genetic information and our assumptions of the beginning of life on Earth (Cech *et al.*, 1981; Guerrier-Takada *et al.*, 1983). Finally, the finding of a vast universe of “noncoding RNAs” that regulate every aspect of transmission and expression of genetic information led to reevaluation of the genetic determinism doctrine (Szymanski & Barciszewski, 2002). The noncoding RNAs database (ncRNADB; Szymanski *et al.*, 2007) contains over 30,000 eukaryotic, bacterial and archeal sequences ranging in size from tens to thousands of nucleotides (excluding siRNAs and microRNAs). They are involved in almost every biological process including responses to oxidative stress in bacteria (Altuvia *et al.*, 1998; Zhang *et al.*, 1997; Zhang

et al., 1998) and eukaryotes (Crawford *et al.*, 1996a; Crawford *et al.*, 1996b), X chromosome inactivation in mammals (Brockdorff *et al.*, 1992; Brown *et al.*, 1992; Penny *et al.*, 1996), post-transcriptional modifications (Bartel, 2004; 2009; Elbashir *et al.*, 2001; Hamilton & Baulcombe, 1999) and RNA-protein interactions (Brownlee, 1971; Wassarman & Storz, 2000).

Studies of RNA Structures

Structures of biological macromolecules are investigated using X-ray crystallography, Nuclear Magnetic Resonance (NMR) Spectroscopy and cryogenic electron microscopy (cryo-EM) and deposited in the RCSB Protein Data Bank (PDB) or Electron Microscopy Data Bank (EMDB), respectively (Berman *et al.*, 2002; Heymann *et al.*, 2005). As of January 2010, over 66,000 structures have been deposited in the PDB. Only 800 free RNA structures are available presently.

Free RNAs are notoriously difficult to crystallize due to their flexible elongated shapes and the consistent negative charges along their phosphate backbones (Ke & Doudna, 2004). To increase the chances of RNA crystallization, researchers carefully consider the design of investigated molecules, choose appropriate synthesis methods and develop homogenous purification protocols (Holbrook & Kim, 1997).

Crystallography is more successful for nucleic-acid-protein complexes than for nucleic acids in their free form as evidenced by the large number of nucleic acid-protein complex structures in the PDB. As of January 2010 the number of structures of nucleic-acid-protein complexes was greater than 2.5 times that for free nucleic acid structures (most of which are of DNA). Crystallization of RNA-protein complexes has several

advantages over that of free RNAs including maintenance of the biologically active RNA conformation (Holbrook *et al.*, 2001).

The crystallization of the bacterial 70S ribosome, a macromolecular protein-synthesizing machine consisting of RNA and protein assembled into two subunits 30S and 50S, constituted a major breakthrough in studying RNA structures (Ban *et al.*, 2000; Schluenzen *et al.*, 2000; Wimberly *et al.*, 2000) (Figure 1.2 A-B). As of October 2010, over 200 ribosome-related structures were available at the PDB. Of particular value are high-resolution structures of bacterial ribosomes complexed with antibiotics such as erythromycin (Dunkle *et al.*, 2010), CEM-101 (Llano-Sotelo *et al.*, 2010), viomycin and capreomyvin (Stanley *et al.*, 2010). These structures have allowed for the elucidation of how antibiotics inhibit bacterial translation and provided information important for the synthesis of new antibiotics. Moreover, these structures provided an important framework for studying both canonical and noncanonical mechanisms of translation.

NMR studies yielded the first structures of small RNA fragments in the early 1990s (Cheong *et al.*, 1990; Heus & Pardi, 1991). Only recently, NMR techniques were used to investigate larger free RNA molecules such as domain 2 of Hepatitis C Virus (HCV) Internal Ribosomal Entry Site (IRES) RNA (~ 25 kDa; Lukavsky *et al.*, 2003). To study the many RNA molecules that do not crystallize and are too large for NMR analysis, a piecemeal strategy was developed. It successfully discerned the structure of 5S ribosomal RNA (Betzel *et al.*, 1994; Lorenz *et al.*, 2000; Perbandt *et al.*, 1998). The structure of the whole 5S rRNA molecule became available upon crystallization of the 50S ribosomal subunit (Ban *et al.*, 2000). Alternatively, low-resolution approaches, such as electron microscopy, UV and chemical cross-linking, chemical and enzymatic methods

yield useful information about RNA structures (Frank, 2002; Harris *et al.*, 1994; Moazed *et al.*, 1986). If RNAs bind to the ribosome, their shape can be extracted from the electron density map of the complex and used as a guide for building a molecular model using available high-resolution structures of small RNA motifs. The latter approach provided important insights into the functions of transfer-messenger RNA in *trans*-translation (Kaur *et al.*, 2006; Valle *et al.*, 2003) and Hepatitis C Virus Internal Ribosomal Entry Site RNA in non-canonical translation (Boehringer *et al.*, 2005).

Structural modeling and RNA structure prediction

Structural models (referred to as “models”) can be constructed using available sequence data and computational techniques (see Chapters 3-5). In a recent survey by Laing and Schlick (2010), models predicted using a variety of programs, such as FARNA, iFoldRNA, MC-Fold/MC-Sym, and NAST were compared against experimentally-derived RNA structures in the PDB using as input either sequence alone or in combination with secondary structure information. The computed tertiary structure models were at best 6 Å Root Mean Square Deviation compared to a known structure for the same control molecule. This indicates that the development of structural modeling techniques for RNA molecules have lagged behind those for proteins.

We developed a semi-automated multiple sequence alignment procedure that take into account biological information, such as locations of open reading frames that can not presently be efficiently analyzed by a computer due to algorithm limitations (See Figure 1.3 for the overview of the procedure). First, unique sequences are extracted from databases such as GenBank (Benson *et al.*, 2009) into a local database or collection (See

Chapter 2). Second, the sequences are initially aligned using automated tools such as CLUSTAL and the alignment is improved manually by incorporating known biochemical or structural data. The aligned sequences are then compared (Comparative Sequence Analysis) to discern covarying nucleotides in two or more alignment column positions, especially those in which the nucleotides in two positions vary together but a Watson-Crick or G-U “wobble” pair is conserved, called a Compensatory Base Change (CBC) (Fox & Woese, 1975; Larsen & Zwieb, 1991) (See Chapter 3). The alignment is checked for errors by automated programs available in the RNAdbTools suite (Gorodkin *et al.*, 2001) and included in the Semi-Automated RNA Sequence Editor (SARSE) (Andersen *et al.*, 2007). For the existence of a base pair to be supported, the analyzed alignment positions must contain twice the number of CBCs than mismatched non-Watson Crick and G-U pairs (Larsen & Zwieb, 1991). While many forms of Comparative Sequence Analysis take advantage of a variety of complex mathematical comparative techniques, our calculations using the established “2:1 rule” are relatively simple and were proven effective for determining secondary structures of RNAs such as transfer-messenger RNA (Burks *et al.*, 2005; Zwieb *et al.*, 1999; See Chapter 3) and Signal Recognition Particle (SRP) RNA (Larsen & Zwieb, 1991).

Once the secondary structure is determined, it can be used as a blueprint to develop a phylogenetically-supported tertiary structure model. While other methods for tertiary structure prediction often use thermodynamic or physical calculations resulting in poor similarities to control structures, our method takes advantage of the wealth of biologically relevant partial or full RNA structures in the PDB. Given that structural features (or “motifs”) found in one RNA molecule are found in other RNA

macromolecules in nature, the three-dimensional coordinates of the structures of these motifs can be used to create more biologically relevant models expected to be close to the actual structure of the target molecule (Burks *et al.*, 2005). Due to the advances in studying the structure of the ribosome, crystallographic and NMR techniques, we have a wealth of three-dimensional macromolecular RNA structure motifs catalogued at the Structural Classification of RNA Database (Klosterman *et al.*, 2002). We also have at our disposal the tools to manually create structural models using this method, in which the program Editor for RNA in 3-D (ERNA-3D) is a cornerstone (Burks *et al.*, 2005; Mueller *et al.*, 1995) (See Chapter 3). ERNA-3D not only has the capability to generate A-form RNA for helical sections of RNA molecules and calculate the conformations of single strands in real time in a desktop computer, but also has the capability to copy coordinates of an available structural motif onto the working model. Additionally, ERNA-3D can incorporate available protein structures into models, a feature which was used in studies of Foot and Mouth Disease Virus Internal Ribosomal Entry Site RNA and Polypyrimidine Tract Binding Protein (see Chapter 4). Information about RNA in structures of RNA-protein complexes can be copied on the RNA model with the protein structure left in a biologically relevant position relative to the copied RNA.

Protein Synthesis: Translation of mRNA Into Protein

Translation is a three-stage biological process. All organisms use it to produce proteins from a genome-encoded message (Kapp & Lorsch, 2004; Simonetti *et al.*, 2009). The basic plan of protein synthesis in eukaryotes is similar to that in prokaryotes. However, there are significant differences between them at the initiation stage. During

this stage, initiation factors bind mRNA and the aminoacylated initiator tRNA and deliver them to the small ribosomal subunit to form an initiation complex. This complex in turn binds to the large ribosomal subunit. In the elongation phase, ribosomes read the mRNA-encoded messages and synthesize elongating chains of amino acids called polypeptides. During termination, stop codons signal the end point of the reading frame in the mRNA and protein termination factors coordinate disassembly of the ribosome. Canonical translation components in both prokaryotic and eukaryotic systems include a pool of transfer RNAs (tRNAs) which bring the activated amino acids to the ribosome. Translation is facilitated by protein initiation, elongation, and termination factors (Kapp & Lorsch, 2004; Pestova *et al.*, 2007; Simonetti *et al.*, 2009).

The Ribosome: A Molecular Machine

The ribosome is a dynamic molecular machine constructed of RNA and proteins and driven by GTP hydrolysis. In bacteria, the ribosome is composed of the small (30S) and large (50S) ribosomal subunits that form a 70S particle (Figure 1.2C). In eukaryotes, it consists of 40S and 60S subunits, which form 80S ribosomes (Figure 1.2D). Only high-resolution structures of bacterial and archeobacterial ribosomes are available (Ban *et al.*, 2000; Schluenzen *et al.*, 2000; Wimberly *et al.*, 2000) (Figure 1.2 A and B). Because ribosomes are highly conserved these structures are useful in interpretation of structural information regarding eukaryotic ribosomes (Figure 1.2D). The eukaryotic ribosome is currently investigated using homology modeling and cryo-EM (Schmeing & Ramakrishnan, 2009; Spahn *et al.*, 2001b).

The ribosome catalyzes the formation of peptide bonds using information encoded in mRNA (Noller, 2010). The small subunit binds mRNA and has three sites to bind tRNA molecules. An aminoacylated tRNA enters the ribosome at the Aminoacyl (A)-site, is bound to the growing peptide chain in the Peptidyl (P)-site and exits the ribosome through the E site. The mRNA is read in triplets of nucleotides called codons which are recognized by anticodons of the tRNAs. The growing peptide chain of the P site-bound tRNA is transferred to the amino acid of the incoming A site tRNA. The reaction takes place within the peptidyl transferase center composed primarily of rRNA (Ban *et al.*, 2000; Nissen *et al.*, 2000; Yusupov *et al.*, 2001). Translocation, a coordinated movement of the mRNA and tRNAs through the ribosome, is accomplished through a ratcheting motion of the ribosomal subunits and is catalyzed by elongation factors (Horan & Noller, 2007; Korostelev *et al.*, 2008).

Throughout this dissertation, a number of components of the ribosome or its subunits are referenced. Figure 1.2E-F provides a guidemap for the 70S ribosome. It shows the three tRNA binding sites (A, P and E sites), and the peptidyl transferase center. The eukaryotic 40S ribosomal subunit plays an essential role in viral IRES-mediated translation initiation. The viral RNA genome binds to the 40S subunit with or without the assistance of canonical eukaryotic translation initiation factors (Kieft, 2008). Many ribosomal proteins are thought to play a role in allowing viral IRES RNAs to bind to the cellular 40S subunit, including but not limited to ribosomal proteins S5, S25, and S14 (Figure 1.2G) (Babaylova *et al.*, 2009; Fukushi *et al.*, 2001; Landry *et al.*, 2009).

Transfer-messenger RNA and *Trans*-Translation

Trans-translation is a quality control process for the elongation step of prokaryotic translation (Keiler *et al.*, 1996; Tu *et al.*, 1995). This process involves 70S ribosomes stalled on truncated mRNAs or those lacking stop codons (Hayes & Keiler ; Keiler, 2008). Transfer-messenger RNA (tmRNA) (Figure 1.4A), the key player in *trans*-translation, is encoded by the *ssrA* gene (Keiler *et al.*, 1996; Oh *et al.*, 1990; Tu *et al.*, 1995). It contains a tRNA-like domain (TLD) and an mRNA-like region (MLR) containing a short open reading frame and a series of pseudoknots (Felden *et al.*, 1997; Hou & Schimmel, 1988; Keiler *et al.*, 1996; Komine *et al.*, 1994; Roche & Sauer, 1999; Tu *et al.*, 1995; Ushida *et al.*, 1994; Zwieb *et al.*, 1999). TmRNA promotes recycling of the stalled ribosomes and tags the defective polypeptides for degradation by cellular proteases (Keiler *et al.*, 2000).

In *trans*-translation (Figure 1.4B), tmRNA forms a complex with small protein B (SmpB) and Elongation Factor Tu (EF-Tu) and binds to the A site of stalled 70S ribosomes (Keiler *et al.*, 1996). While the details of the next steps are still poorly understood, the ribosome switches from the truncated mRNA to the MLR of the tmRNA, EF-Tu is released and the peptide encoded by the truncated mRNA is transferred to the alanine of the tmRNA. Translation of the MLR open reading frame proceeds until the ribosome reaches the stop codon (or codons).

tmRNA and *trans*-translation are prospective targets for pharmaceutical intervention because 1) they are found in all species of bacteria analyzed to date (Keiler *et al.*, 2000; Mao *et al.*, 2009), 2) they have no eukaryotic analogs, 3) tmRNA is essential for survival of some bacteria (Huang *et al.*, 2000), and 4) the *ssrA* gene is required for

virulence (Julio *et al.*, 2000) or survival of stressed cells (Nakano *et al.*, 2001). In order to effectively disrupt or inhibit *trans*-translation, one must understand the structure of tmRNA on the ribosome. To achieve this goal, the structure of tmRNA was investigated using comparative sequence analysis and structural modeling approaches (see Chapter 3). Sequences used in the tmRNA analysis were collected into the tmRDB (Andersen *et al.*, 2006) as described in Chapter 2. Our modeling approach incorporated available sequence, biochemical and biophysical data and produced a model that fits into the shape of ribosome-bound tmRNA observed in cryo-EM studies (Valle *et al.*, 2003). The proposed model can be used to facilitate rational design of novel therapeutics, to interpret cryo-EM images of tmRNA on the ribosome or to investigate tmRNA-protein interactions that may facilitate tmRNA in transit throughout the ribosome.

Internal Ribosomal Entry Site-mediated Translation Initiation

Cap-independent translation initiation was discovered only two decades ago (Jang & Wimmer, 1990; Pelletier & Sonenberg, 1988). Some viruses with positive-sense RNA genomes can use a highly structured segment of the 5'-Untranslated Region (5'-UTR) called the Internal Ribosomal Entry Site (IRES) RNA to hijack the eukaryotic protein synthesis apparatus. This process is called "Internal Ribosomal Entry" and is aided by various collections of host cellular factors depending on the IRES RNA involved (Kieft, 2008). In the simplest IRES-mediated viral translation initiation, the intergenic region (IGR) IRES RNAs of Cricket Paralysis Virus (CrPV) (Figure 1.5A) and other viruses in the family *Dicistroviridae* bind to the 40S ribosomal subunit without the help of host protein factors (Pestova & Hellen, 2003; Sasaki & Nakashima, 1999; 2000; Wilson *et al.*,

2000). The Cryo-EM studies of IGR IRES RNAs of CrPV and its relative *Plautia stali* Intestine Virus (PSIV) are able to mimic a mRNA codon:tRNA anticodon interaction when they bind directly to the mRNA decoding region in the 40S ribosomal subunit with the bulk of the IRES structure protruding from the ribosome (Costantino *et al.*, 2008; Pestova *et al.*, 2004; Pflugsten *et al.*, 2006; Schuler *et al.*, 2006). The larger, more elongated IRES RNA of Hepatitis C Virus (Figure 1.5B) also binds to the neck and platform of the 40S subunit *without* the aid of host protein factors but its translation is assisted by eIF2 and eIF3 (Filbin & Kieft, 2009; Kieft, 2008; Spahn *et al.*, 2001b). On the other hand, viruses in the *Picornaviridae* family make extensive use of a variety of IRES RNA secondary structures (such as in Figure 1.5C) and host initiation factors, with the exception of cap-binding protein eIF4E (Belsham & Jackson, 2000; Kieft, 2008). An example of a complex network of interactions utilized by Foot and Mouth Disease Virus (FMDV) IRES RNA is illustrated in Figure 4.4.

IRES RNA tertiary structural data are only available for representatives of the virus families *Dicistroviridae* (CrPV, PSIV) and *Flaviviridae* (HCV) (Kieft *et al.*, 2008). To fill the gaps, this project focused on IRES RNAs from FMDV (family *Picornaviridae*) and Bovine Viral Diarrhea Virus (BVDV), a less studied member of *Flaviviridae* (see Chapter 5).

Foot and Mouth Disease Virus IRES RNA

Foot and Mouth Disease Virus (FMDV) is a positive-sense, single-stranded RNA (ssRNA) virus of the *Aphthovirus* genus in the *Picornaviridae* family (Belsham, 2005). It infects cloven-hoofed livestock and wild animals such as cattle, sheep, pigs, goats, and

deer, and remains one of the most prolific viruses in written history (Grubman & Baxt, 2004; Mahy, 2005). FMDV constitutes a significant threat to food animal industries in the United States of America and elsewhere in the world because currently there are no antiviral treatments or vaccines effective against all serotypes of this virus. Its infectious genome contains a poly(A) tail but no canonical cap structure (Belsham & Bostock, 1988). Translation of the FMDV genome by ribosomes of the animal host constitutes an essential step in the viral life cycle (Belsham, 2005). Thus, translation of the FMDV genome is an attractive target for therapeutic intervention.

Protein synthesis is initiated downstream of the IRES at the region encoding an autoproteolytic polyprotein (Belsham & Brangwyn, 1990; Kuhn *et al.*, 1990). FMDV IRES recruits host translation initiation factors and cellular proteins including but not limited to polypyrimidine tract binding protein (PTB) (Belsham & Bostock, 1988; Belsham & Brangwyn, 1990; Belsham, 2005; Jackson, 2002; Kuhn *et al.*, 1990; Pestova *et al.*, 2001; Pilipenko *et al.*, 2000). The latter has no function in canonical translation but is thought to act as a RNA chaperone in IRES-mediated translation initiation (Song *et al.*, 2005). Most information about the secondary and tertiary structure of the 460-nucleotide long FMDV IRES RNA has been inferred from studies of the IRES RNA of Encephalomyocarditis virus (EMCV), a virus that belongs to genus *Cardiovirus* in the *Picornaviridae* family (Pilipenko *et al.*, 1989; Pilipenko *et al.*, 2000). Only segments of FMDV IRES RNA were investigated using chemical and enzymatic footprinting approaches (Fernandez-Miragall & Martinez-Salas, 2003; 2007). The historical secondary structures, originally derived from comparative sequence analysis of three sequences of FMDV, EMCV and Theiler's murine encephalomyelitis (TMEV) IRES

RNAs and structural modification data for EMCV IRES, show five finger-like domains, of which domains 2-5 are absolutely required for IRES functions (Duke *et al.*, 1992; Jang & Wimmer, 1990; Niepmann *et al.*, 1997; Pilipenko *et al.*, 1989). The project described in Chapter 4 utilizes information available for the IRES RNAs of EMCV and the related Theiler's Murine Encephalomyelitis virus (TMEV), and the tools and methods described for tmRNA (see Chapters 2 and 3) to investigate the structure of FMDV IRES RNA and its roles in the translation of the viral genome. Because of the absence of high-resolution structural data for FMDV IRES RNA and the availability of the structures of PTB-RNA complexes (Oberstrass *et al.*, 2005), the modeling approach was modified to incorporate biochemical data for protein-RNA interactions as additional three-dimensional constraints.

Bovine Viral Diarrhea Virus IRES RNA

Bovine Viral Diarrhea Virus (BVDV) is an approximately 12.5 kb uncapped, positive-sense ssRNA virus of the family *Flaviviridae* and genus *Pestivirus* (Brown *et al.*, 1992b; Brock *et al.*, 1992; Collett *et al.*, 1988a; Collett *et al.*, 1988b; Lindenbach *et al.*, 2007; Meyers & Thiel, 1996; Renard *et al.*, 1987; Thiel *et al.*, 2005). It causes acute diarrhea, fetal infections and mucosal disease in cattle and can result in persistent infections (Goens, 2002). Infected animals are culled rather than treated because no effective treatment is yet available (Moennig *et al.*, 2005).

Interest has arisen recently in using BVDV infection as a surrogate model system for studying the infection cycle of Hepatitis C Virus (HCV), another member of the *Flaviviridae* family (genus *Hepacivirus*) and a worldwide cause of liver disease in

humans (Buckwold *et al.*, 2003). As in FMDV, the genomic RNA of BVDV is polycistronic and contains an open reading frame encoding an autoproteolytic polypeptide flanked at the 5' end by an untranslated region (5' UTR) containing an internal ribosomal entry site (IRES) RNA (Brown *et al.*, 1992b; Collett *et al.*, 1988b). The IRES-mediated translation of the viral ORF by host ribosomes is expected to be an essential step in the viral life cycle, as in as HCV and chimeric HCV-Poliioviruses (Friebe *et al.*, 2001; Jang, 2006). The historical secondary structures for the IRES RNAs of BVDV and other pestiviruses are similar to that observed for HCV IRES RNA, with an additional helix in the third domain (Brown *et al.*, 1992; Deng & Brock, 1993; Moes & Wirth, 2007) and little of the secondary structure surrounding the start codon which was seen in HCV IRES RNA (Honda *et al.*, 1996). Unlike FMDV IRES RNA, the HCV-like IRES RNA secondary structures contain a compact pseudoknot (Fletcher *et al.*, 2002; Moes & Wirth, 2007; Rijnbrand *et al.*, 1997; Wang *et al.*, 1995; Wang *et al.*, 1994). This pseudoknot and helices 2-3 are required for translation (Chon *et al.*, 1998; Honda *et al.*, 1996; Moes & Wirth, 2007; Poole *et al.*, 1995). The tertiary structure of the pseudoknot is unknown and presents an opportunity for structural modeling using tools developed for pseudoknot-containing tmRNA (Burks *et al.*, 2005; see also Chapter 3). Structural information for HCV and CSFV IRES RNA along with the comparative sequence and modeling approaches were utilized to predict the structural features of BVDV IRES RNA.

Cryo-EM, NMR and X-ray crystallography data are only available for HCV IRES RNA (Boehringer *et al.*, 2005; Collier *et al.*, 2002; Dibrov *et al.*, 2007; Kieft *et al.*, 2002; Klinck *et al.*, 2000; Locker *et al.*, 2007; Lukavsky *et al.*, 2003; Lukavsky *et al.*, 2000;

Rijnbrand *et al.*, 2004; Siridechadilok *et al.*, 2005) and the intergenic IRES RNAs of the unrelated Dicistroviruses (Costantino *et al.*, 2008; Schuler *et al.*, 2006; Spahn *et al.*, 2004). However, the complete structure of HCV IRES RNA remains unavailable. This created another opportunity to explore the structure of HCV IRES RNA through modeling and comparison with BVDV IRES RNA (see Chapter 5).

Concluding Remarks

My structural models will be important for further studies of tmRNA in *trans*-translation and of the IRES RNAs of FMDV, BVDV and HCV and their functions in IRES-mediated translation initiation and viral life cycles, for identifying similarities between widely varying ribosome-binding IRES RNA structures, and for designing a unique RNA-protein modeling procedure which can be used to bridge the gap between chemical modification data and structure-based drug design.

References

- Ahlquist, P. (2002). RNA-dependent RNA polymerases, viruses, and RNA silencing. *Science* 296, 1270-1273.
- Altuvia, S., Zhang, A., Argaman, L., Tiwari, A. & Storz, G. (1998). The Escherichia coli OxyS regulatory RNA represses fhlA translation by blocking ribosome binding. *EMBO J* 17, 6069-6075.
- Andersen, E., Lind-Thomsen, A., Knudsen, B., Kristensen, S., Havgaard, J., Torarinsson, E., Larsen, N., Zwieb, C., Sestoft, P., Kjems, J. & Gorodkin, J. (2007). Semiautomated improvement of RNA alignments. *RNA* 13, 1850-1859.
- Andersen, E. S., Rosenblad, M. A., Larsen, N., Westergaard, J. C., Burks, J., Wower, I. K., Wower, J., Gorodkin, J., Samuelsson, T. & Zwieb, C. (2006). The tmRDB and SRPDB resources. *Nucleic Acids Res* 34, D163-168.
- Babaylova, E., Graifer, D., Malygin, A., Stahl, J., Shatsky, I. & Karpova, G. (2009). Positioning of subdomain III_d and apical loop of domain II of the hepatitis C IRES on the human 40S ribosome. *Nucleic Acids Res* 37, 1141-1151.
- Baltimore, D. (1970). RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* 226, 1209-1211.
- Baltimore, D., Eggers, H. J., Franklin, R. M. & Tamm, I. (1963). Poliovirus-induced RNA polymerase and the effects of virus-specific inhibitors on its production. *Proc Natl Acad Sci U S A* 49, 843-849.
- Baltimore, D. & Franklin, R. M. (1963). Properties of the mengovirus and poliovirus RNA polymerases. *Cold Spring Harb Symp Quant Biol* 28, 105-108.
- Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* 289, 905-920.
- Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281-297.
- Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215-233.
- Belsham, G. J. (2005). Translation and replication of FMDV RNA. *Curr Top Microbiol Immunol* 288, 43-70.
- Belsham, G. J. & Bostock, C. J. (1988). Studies on the infectivity of foot-and-mouth disease virus RNA using microinjection. *J Gen Virol* 69 (Pt 2), 265-274.
- Belsham, G. J. & Brangwyn, J. K. (1990). A region of the 5' noncoding region of foot-and-mouth disease virus RNA directs efficient internal initiation of protein

- synthesis within cells: involvement with the role of L protease in translational control. *J Virol* 64, 5389-5395.
- Belsham, G. J. & Jackson, R. (2000). Translation initiation on picornavirus RNA. In *Translational Control of Gene Expression*, pp. 869-900. Edited by N. Sonenberg, J. W. B. Hershey & M. B. Mathews. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. (2009). GenBank. *Nucleic Acids Res* 37, D26-31.
- Betzl, C., Lorenz, S., Furste, J. P., Bald, R., Zhang, M., Schneider, T. R., Wilson, K. S. & Erdmann, V. A. (1994). Crystal structure of domain A of *Thermus flavus* 5S rRNA and the contribution of water molecules to its structure. *FEBS Lett* 351, 159-164.
- Boehringer, D., Thermann, R., Ostareck-Lederer, A., Lewis, J. & Stark, H. (2005). Structure of the hepatitis C virus IRES bound to the human 80S ribosome: remodeling of the HCV IRES. *Structure* 13, 1695-1706.
- Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V. M., Norris, D. P., Cooper, P. J., Swift, S. & Rastan, S. (1992). The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* 71, 515-526.
- Brock, K. V., Deng, R. & Riblet, S. M. (1992). Nucleotide sequencing of 5' and 3' termini of bovine viral diarrhoea virus by RNA ligation and PCR. *J Virol Methods* 38, 39-46.
- Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafreniere, R. G., Xing, Y., Lawrence, J. & Willard, H. F. (1992a). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* 71, 527-542.
- Brown, E. A., Zhang, H., Ping, L. H. & Lemon, S. M. (1992b). Secondary structure of the 5' nontranslated regions of hepatitis C virus and pestivirus genomic RNAs. *Nucleic Acids Res* 20, 5041-5045.
- Brownlee, G. G. (1971). Sequence of 6S RNA of *E. coli*. *Nat New Biol* 229, 147-149.
- Buckwold, V. E., Beer, B. E. & Donis, R. O. (2003). Bovine viral diarrhoea virus as a surrogate model of hepatitis C virus for the evaluation of antiviral agents. *Antiviral Res* 60, 1-15.
- Burks, J., Zwieb, C., Muller, F., Wower, I. & Wower, J. (2005). Comparative 3-D modeling of tmRNA. *BMC Mol Biol* 6, 14.

- Carter, A. P., Clemons, W. M., Brodersen, D. E., Morgan-Warren, R. J., Wimberly, B. T. & Ramakrishnan, V. (2000). Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature* 407, 340-348.
- Cech, T. R., Zaug, A. J. & Grabowski, P. J. (1981). In vitro splicing of the ribosomal RNA precursor of Tetrahymena: involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell* 27, 487-496.
- Cheong, C., Varani, G. & Tinoco, I., Jr. (1990). Solution structure of an unusually stable RNA hairpin, 5'GGAC(UUCG)GUCC. *Nature* 346, 680-682.
- Chon, S. K., Perez, D. R. & Donis, R. O. (1998). Genetic analysis of the internal ribosome entry segment of bovine viral diarrhoea virus. *Virology* 251, 370-382.
- Collett, M. S., Anderson, D. K. & Retzel, E. (1988a). Comparisons of the pestivirus bovine viral diarrhoea virus with members of the flaviviridae. *J Gen Virol* 69 (Pt 10), 2637-2643.
- Collett, M. S., Larson, R., Gold, C., Strick, D., Anderson, D. K. & Purchio, A. F. (1988b). Molecular cloning and nucleotide sequence of the pestivirus bovine viral diarrhoea virus. *Virology* 165, 191-199.
- Collier, A., Gallego, J., Klinck, R., Cole, P., Harris, S., Harrison, G., Aboul-Ela, F., Varani, G. & Walker, S. (2002). A conserved RNA structure within the HCV IRES eIF3-binding site. *Nat Struct Biol* 9, 375-380.
- Costantino, D. A., Pflugsten, J. S., Rambo, R. P. & Kieft, J. S. (2008). tRNA-mRNA mimicry drives translation initiation from a viral IRES. *Nat Struct Mol Biol* 15, 57-64.
- Crawford, D. R., Schools, G. P. & Davies, K. J. (1996a). Oxidant-inducible adapt 15 RNA is associated with growth arrest- and DNA damage-inducible gadd153 and gadd45. *Arch Biochem Biophys* 329, 137-144.
- Crawford, D. R., Schools, G. P., Salmon, S. L. & Davies, K. J. (1996b). Hydrogen peroxide induces the expression of adapt15, a novel RNA associated with polysomes in hamster HA-1 cells. *Arch Biochem Biophys* 325, 256-264.
- Crick, F. (1970). Central dogma of molecular biology. *Nature* 227, 561-563.
- Crick, F. H. (1958). On protein synthesis. *Symp Soc Exp Biol* 12, 138-163.
- Deng, R. & Brock, K. V. (1993). 5' and 3' untranslated regions of pestivirus genome: primary and secondary structure analyses. *Nucleic Acids Res* 21, 1949-1957.
- Dibrov, S., Johnston-Cox, H., Weng, Y. & Hermann, T. (2007). Functional architecture of HCV IRES domain II stabilized by divalent metal ions in the crystal and in solution. *Angew Chem Int Ed Engl* 46, 226-229.

- Duke, G. M., Hoffman, M. A. & Palmenberg, A. C. (1992). Sequence and structural elements that contribute to efficient encephalomyocarditis virus RNA translation. *J Virol* 66, 1602-1609.
- Dunkle, J. A., Xiong, L., Mankin, A. S. & Cate, J. H. (2010). Structures of the *Escherichia coli* ribosome with antibiotics bound near the peptidyl transferase center explain spectra of drug action. *Proc Natl Acad Sci U S A* 107, 17152-17157.
- Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K. & Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* 411, 494-498.
- Felden, B., Himeno, H., Muto, A., McCutcheon, J. P., Atkins, J. F. & Gesteland, R. F. (1997). Probing the structure of the *Escherichia coli* 10Sa RNA (tmRNA). *Rna* 3, 89-103.
- Fernandez-Miragall, O. & Martinez-Salas, E. (2003). Structural organization of a viral IRES depends on the integrity of the GNRA motif. *RNA* 9, 1333-1344.
- Fernandez-Miragall, O. & Martinez-Salas, E. (2007). In vivo footprint of a picornavirus internal ribosome entry site reveals differences in accessibility to specific RNA structural elements. *J Gen Virol* 88, 3053-3062.
- Filbin, M. E. & Kieft, J. S. (2009). Toward a structural understanding of IRES RNA function. *Curr Opin Struct Biol* 19, 267-276.
- Fletcher, S. P., Ali, I. K., Kaminski, A., Digard, P. & Jackson, R. J. (2002). The influence of viral coding sequences on pestivirus IRES activity reveals further parallels with translation initiation in prokaryotes. *RNA* 8, 1558-1571.
- Fox, G. E. & C. R. Woese. (1975). 5S RNA secondary structure. *Nature* 256, 505-507.
- Frank, J. (2002). Single-particle imaging of macromolecules by cryo-electron microscopy. *Annu Rev Biophys Biomol Struct* 31, 303-319.
- Friebe, P., Lohmann, V., Krieger, N. & Bartenschlager, R. (2001). Sequences in the 5' nontranslated region of hepatitis C virus required for RNA replication. *J Virol* 75, 12047-12057.
- Fukushi, S., Okada, M., Stahl, J., Kageyama, T., Hoshino, F. B. & Katayama, K. (2001). Ribosomal protein S5 interacts with the internal ribosomal entry site of hepatitis C virus. *J Biol Chem* 276, 20824-20826.
- Gao, H., Ayub, M. J., Levin, M. J. & Frank, J. (2005). The structure of the 80S ribosome from *Trypanosoma cruzi* reveals unique rRNA components. *Proc Natl Acad Sci U S A* 102, 10206-10211.

- Goens, S. D. (2002). The evolution of bovine viral diarrhea: a review. *Can Vet J* 43, 946-954.
- Gorodkin, J., Zwieb, C. & Knudsen, B. (2001). Semi-automated update and cleanup of structural RNA alignment databases. *Bioinformatics* 17, 642-645.
- Grubman, M. J. & Baxt, B. (2004). Foot-and-mouth disease. *Clin Microbiol Rev* 17, 465-493.
- Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* 35, 849-857.
- Hamilton, A. J. & Baulcombe, D. C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* 286, 950-952.
- Harris, M. E., Nolan, J. M., Malhotra, A., Brown, J. W., Harvey, S. C. & Pace, N. R. (1994). Use of photoaffinity crosslinking and molecular modeling to analyze the global architecture of ribonuclease P RNA. *EMBO J* 13, 3953-3963.
- Hayes, C. S. & Keiler, K. C. Beyond ribosome rescue: tmRNA and co-translational processes. *FEBS Lett* 584, 413-419.
- Heus, H. A. & Pardi, A. (1991). Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science* 253, 191-194.
- Heymann, J. B., Chagoyen, M. & Belnap, D. M. (2005). Common conventions for interchange and archiving of three-dimensional electron microscopy information in structural biology. *J Struct Biol* 151, 196-207.
- Higgins, D. G., Thompson, J. D. & Gibson, T. J. (1996). Using CLUSTAL for multiple sequence alignments. *Methods Enzymol* 266, 383-402.
- Holbrook, S. R., Holbrook, E. L. & Walukiewicz, H. E. (2001). Crystallization of RNA. *Cell Mol Life Sci* 58, 234-243.
- Holbrook, S. R. & Kim, S. H. (1997). RNA crystallography. *Biopolymers* 44, 3-21.
- Honda, M., Brown, E. A. & Lemon, S. M. (1996). Stability of a stem-loop involving the initiator AUG controls the efficiency of internal initiation of translation on hepatitis C virus RNA. *RNA* 2, 955-968.
- Horan, L. H. & Noller, H. F. (2007). Intersubunit movement is required for ribosomal translocation. *Proc Natl Acad Sci U S A* 104, 4881-4885.
- Hou, Y. M. & Schimmel, P. (1988). A simple structural feature is a major determinant of the identity of a transfer RNA. *Nature* 333, 140-145.

- Huang, C., Wolfgang, M. C., Withey, J., Koomey, M. & Friedman, D. I. (2000). Charged tmRNA but not tmRNA-mediated proteolysis is essential for *Neisseria gonorrhoeae* viability. *Embo J* 19, 1098-1107.
- Jackson, R. J. (2002). Proteins Involved in the Function of Picornavirus Internal Ribosomal Entry Sites. In *Molecular Biology of Picornaviruses*, pp. 171-183. Edited by B. Semler, L. & E. Wimmer. Washington, D.C. 20036-2904: ASM Press.
- Jang, S. & Wimmer, E. (1990). Cap-independent translation of encephalomyocarditis virus RNA: structural elements of the internal ribosomal entry site and involvement of a cellular 57-kD RNA-binding protein. *Genes Dev* 4, 1560-1572.
- Jang, S. K. (2006). Internal initiation: IRES elements of picornaviruses and hepatitis c virus. *Virus Res* 119, 2-15.
- Julio, S. M., Heithoff, D. M. & Mahan, M. J. (2000). ssrA (tmRNA) plays a role in *Salmonella enterica* serovar Typhimurium pathogenesis. *J Bacteriol* 182, 1558-1563.
- Kapp, L. D. & Lorsch, J. R. (2004). The molecular mechanics of eukaryotic translation. *Annu Rev Biochem* 73, 657-704.
- Kaur, S., Gillet, R., Li, W., Gursky, R. & Frank, J. (2006). Cryo-EM visualization of transfer messenger RNA with two SmpBs in a stalled ribosome. *Proc Natl Acad Sci U S A* 103, 16484-16489.
- Ke, A. & Doudna, J. A. (2004). Crystallization of RNA and RNA-protein complexes. *Methods* 34, 408-414.
- Keiler, K. C. (2008). Biology of *trans*-translation. *Annu Rev Microbiol* 62, 133-151.
- Keiler, K. C., Shapiro, L. & Williams, K. P. (2000). tmRNAs that encode proteolysis-inducing tags are found in all known bacterial genomes: A two-piece tmRNA functions in *Caulobacter*. *Proc Natl Acad Sci U S A* 97, 7778-7783.
- Keiler, K. C., Waller, P. R. & Sauer, R. T. (1996). Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* 271, 990-993.
- Kieft, J., Zhou, K., Grech, A., Jubin, R. & Doudna, J. (2002). Crystal structure of an RNA tertiary domain essential to HCV IRES-mediated translation initiation. *Nat Struct Biol* 9, 370-374.
- Kieft, J. S. (2008). Viral IRES RNA structures and ribosome interactions. *Trends Biochem Sci* 33, 274-283.

- Klinck, R., Westhof, E., Walker, S., Afshar, M., Collier, A. & Aboul-Ela, F. (2000). A potential RNA drug target in the hepatitis C virus internal ribosomal entry site. *RNA* 6, 1423-1431.
- Klosterman, P., Tamura, M., Holbrook, S. & Brenner, S. (2002). SCOR: a Structural Classification of RNA database. *Nucleic Acids Res* 30, 392-394.
- Komine, Y., Kitabatake, M., Yokogawa, T., Nishikawa, K. & Inokuchi, H. (1994). A tRNA-like structure is present in 10Sa RNA, a small stable RNA from *Escherichia coli*. *Proc Natl Acad Sci U S A* 91, 9223-9227.
- Korostelev, A., Ermolenko, D. N. & Noller, H. F. (2008). Structural dynamics of the ribosome. *Curr Opin Chem Biol* 12, 674-683.
- Kuhn, R., Luz, N. & Beck, E. (1990). Functional analysis of the internal translation initiation site of foot-and-mouth disease virus. *J Virol* 64, 4625-4631.
- Laing, C. & Schlick, T. (2010). Computational approaches to 3D modeling of RNA. *J Phys Condens Matter* 22.
- Landry, D. M., Hertz, M. I. & Thompson, S. R. (2009). RPS25 is essential for translation initiation by the Dicistroviridae and hepatitis C viral IRESs. *Genes Dev* 23, 2753-2764.
- Larsen, N. & Zwieb, C. (1991). SRP-RNA sequence alignment and secondary structure. *Nucleic Acids Res* 19, 209-215.
- Lindenbach, B. D., Thiel, H.-J. & Rice, C. M. (2007). Flaviviridae: The Viruses and Their Replication. In *Fields Virology*, 5 edn, pp. 1101-1152. Edited by P. H. Knipe, P. M. Howley, D. E. Griffin, R. A. Lamb, M. A. Martin, B. Roizman & S. E. Straus. Philadelphia, PA: Lohmann Williams & Wilkins.
- Llano-Sotelo, B., Dunkle, J., Klepacki, D., Zhang, W., Fernandes, P., Cate, J. H. & Mankin, A. S. (2010). Binding and action of CEM-101, a new fluoroketolide antibiotic that inhibits protein synthesis. *Antimicrob Agents Chemother*.
- Locker, N., Easton, L. & Lukavsky, P. (2007). HCV and CSFV IRES domain II mediate eIF2 release during 80S ribosome assembly. *EMBO J* 26, 795-805.
- Lorenz, S., Perbandt, M., Lippmann, C., Moore, K., DeLucas, L. J., Betzel, C. & Erdmann, V. A. (2000). Crystallization of engineered *Thermus flavus* 5S rRNA under earth and microgravity conditions. *Acta Crystallogr D Biol Crystallogr* 56, 498-500.
- Lukavsky, P., Kim, I., Otto, G. & Puglisi, J. (2003). Structure of HCV IRES domain II determined by NMR. *Nat Struct Biol* 10, 1033-1038.

- Lukavsky, P., Otto, G., Lancaster, A., Sarnow, P. & Puglisi, J. (2000). Structures of two RNA domains essential for hepatitis C virus internal ribosome entry site function. *Nat Struct Biol* 7, 1105-1110.
- Mahy, B. W. (2005). Introduction and history of foot-and-mouth disease virus. *Curr Top Microbiol Immunol* 288, 1-8.
- Mao, C., Bhardwaj, K., Sharkady, S. M., Fish, R. I., Driscoll, T., Wower, J., Zwieb, C., Sobral, B. W. & Williams, K. P. (2009). Variations on the tmRNA gene. *RNA Biol* 6, 355-361.
- Mathews, D. H., Sabina, J., Zuker, M. & Turner, D. H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* 288, 911-940.
- Meyers, G. & Thiel, H. J. (1996). Molecular characterization of pestiviruses. *Adv Virus Res* 47, 53-118.
- Moazed, D., Stern, S. & Noller, H. F. (1986). Rapid chemical probing of conformation in 16 S ribosomal RNA and 30 S ribosomal subunits using primer extension. *J Mol Biol* 187, 399-416.
- Moennig, V., Houe, H. & Lindberg, A. (2005). BVD control in Europe: current status and perspectives. *Anim Health Res Rev* 6, 63-74.
- Moes, L. & Wirth, M. (2007). The internal initiation of translation in bovine viral diarrhea virus RNA depends on the presence of an RNA pseudoknot upstream of the initiation codon. *Virol J* 4, 124.
- Mueller, F., Doring, T., Erdemir, T., Greuer, B., Junke, N., Osswald, M., Rinke-Appel, J., Stade, K., Thamm, S. & Brimacombe, R. (1995). Getting closer to an understanding of the three-dimensional structure of ribosomal RNA. *Biochem Cell Biol* 73, 767-773.
- Nakano, H., Goto, S., Nakayashiki, T. & Inokuchi, H. (2001). Temperature-sensitive mutations in various genes of Escherichia coli K12 can be suppressed by the ssrA gene for 10Sa RNA (tmRNA). *Mol Genet Genomics* 265, 615-621.
- Niepmann, M., Petersen, A., Meyer, K. & Beck, E. (1997). Functional involvement of polypyrimidine tract-binding protein in translation initiation complexes with the internal ribosome entry site of foot-and-mouth disease virus. *J Virol* 71, 8330-8339.
- Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. (2000). The structural basis of ribosome activity in peptide bond synthesis. *Science* 289, 920-930.
- Noller, H. F. (2010). Evolution of Protein Synthesis from an RNA World. *Cold Spring Harb Perspect Biol*.

- Oberstrass, F. C., Auweter, S. D., Erat, M., Hargous, Y., Henning, A., Wenter, P., Reymond, L., Amir-Ahmady, B., Pitsch, S., Black, D. L. & Allain, F. H. (2005). Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science* 309, 2054-2057.
- Oh, B. K., Chauhan, A. K., Isono, K. & Apirion, D. (1990). Location of a gene (*ssrA*) for a small, stable RNA (10Sa RNA) in the *Escherichia coli* chromosome. *J Bacteriol* 172, 4708-4709.
- Pacheco, A., Reigadas, S. & Martinez-Salas, E. (2008). Riboproteomic analysis of polypeptides interacting with the internal ribosome-entry site element of foot-and-mouth disease viral RNA. *Proteomics* 8, 4782-4790.
- Pelletier, J. & Sonenberg, N. (1988). Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* 334, 320-325.
- Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. & Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. *Nature* 379, 131-137.
- Perbandt, M., Nolte, A., Lorenz, S., Bald, R., Betzel, C. & Erdmann, V. A. (1998). Crystal structure of domain E of *Thermus flavus* 5S rRNA: a helical RNA structure including a hairpin loop. *FEBS Lett* 429, 211-215.
- Pestova, T. V. & Hellen, C. U. (2003). Translation elongation after assembly of ribosomes on the Cricket paralysis virus internal ribosomal entry site without initiation factors or initiator tRNA. *Genes Dev* 17, 181-186.
- Pestova, T. V., Lomakin, I. B. & Hellen, C. U. (2004). Position of the CrPV IRES on the 40S subunit and factor dependence of IRES/80S ribosome assembly. *EMBO Rep* 5, 906-913.
- Pestova, T. V., Lorsch, J. R. & Hellen, C. U. T. (2007). The mechanism of translation initiation in eukaryotes. In *Translational Control in Biology and Medicine*, pp. 87-128. Edited by M. B. Mathews, N. Sonenberg & J. W. B. Hershey. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
- Pestova, T.V., Kolupaeva, V., Lomakin, I., Pilipenko, E., Shatsky, I., Agol, V. & Hellen, C. (2001). Molecular mechanisms of translation initiation in eukaryotes. *Proc Natl Acad Sci U S A* 98, 7029-7036.
- Pfingsten, J., Costantino, D. & Kieft, J. (2006). Structural basis for ribosome recruitment and manipulation by a viral IRES RNA. *Science* 314, 1450-1454.
- Pilipenko, E., Blinov, V., Chernov, B., Dmitrieva, T. & Agol, V. (1989). Conservation of the secondary structure elements of the 5'-untranslated region of cardio- and aphthovirus RNAs. *Nucleic Acids Res* 17, 5701-5711.

- Pilipenko, E. V., Pestova, T. V., Kolupaeva, V. G., Khitrina, E. V., Poperechnaya, A. N., Agol, V. I. & Hellen, C. U. (2000). A cell cycle-dependent protein serves as a template-specific translation initiation factor. *Genes Dev* 14, 2028-2045.
- Poole, T. L., Wang, C., Popp, R. A., Potgieter, L. N., Siddiqui, A. & Collett, M. S. (1995). Pestivirus translation initiation occurs by internal ribosome entry. *Virology* 206, 750-754.
- Renard, A., Schmetz, D., Guiot, C., Brown-Shimmer, S., Dagenais, L., Pastoret, P. P., Dina, D. & Martial, J. A. (1987). Molecular cloning of the bovine viral diarrhoea virus genomic RNA. *Ann Rech Vet* 18, 121-125.
- Rijnbrand, R., Thiviyanathan, V., Kaluarachchi, K., Lemon, S. & Gorenstein, D. (2004). Mutational and structural analysis of stem-loop III C of the hepatitis C virus and GB virus B internal ribosome entry sites. *J Mol Biol* 343, 805-817.
- Rijnbrand, R., van der Straaten, T., van Rijn, P. A., Spaan, W. J. & Bredenbeek, P. J. (1997). Internal entry of ribosomes is directed by the 5' noncoding region of classical swine fever virus and is dependent on the presence of an RNA pseudoknot upstream of the initiation codon. *J Virol* 71, 451-457.
- Roche, E. D. & Sauer, R. T. (1999). SsrA-mediated peptide tagging caused by rare codons and tRNA scarcity. *Embo J* 18, 4579-4589.
- Sasaki, J. & Nakashima, N. (1999). Translation initiation at the CUU codon is mediated by the internal ribosome entry site of an insect picorna-like virus in vitro. *J Virol* 73, 1219-1226.
- Sasaki, J. & Nakashima, N. (2000). Methionine-independent initiation of translation in the capsid protein of an insect RNA virus. *Proc Natl Acad Sci U S A* 97, 1512-1515.
- Schlueder, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F. & Yonath, A. (2000). Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell* 102, 615-623.
- Schmeing, T. M. & Ramakrishnan, V. (2009). What recent ribosome structures have revealed about the mechanism of translation. *Nature* 461, 1234-1242.
- Schuler, M., Connell, S. R., Lescoute, A., Giesebrecht, J., Dabrowski, M., Schroer, B., Mielke, T., Penczek, P. A., Westhof, E. & Spahn, C. M. (2006). Structure of the ribosome-bound cricket paralysis virus IRES RNA. *Nat Struct Mol Biol* 13, 1092-1096.
- Simonetti, A., Marzi, S., Jenner, L., Myasnikov, A., Romby, P., Yusupova, G., Klaholz, B. P. & Yusupov, M. (2009). A structural view of translation initiation in bacteria. *Cell Mol Life Sci* 66, 423-436.

- Siridechadilok, B., Fraser, C., Hall, R., Doudna, J. & Nogales, E. (2005). Structural roles for human translation factor eIF3 in initiation of protein synthesis. *Science* 310, 1513-1515.
- Song, Y., Tzima, E., Ochs, K., Bassili, G., Trusheim, H., Linder, M., Preissner, K. T. & Niepmann, M. (2005). Evidence for an RNA chaperone function of polypyrimidine tract-binding protein in picornavirus translation. *RNA* 11, 1809-1824.
- Spahn, C., Kieft, J., Grassucci, R., Penczek, P., Zhou, K., Doudna, J. & Frank, J. (2001a). Hepatitis C virus IRES RNA-induced changes in the conformation of the 40s ribosomal subunit. *Science* 291, 1959-1962.
- Spahn, C. M., Beckmann, R., Eswar, N., Penczek, P. A., Sali, A., Blobel, G. & Frank, J. (2001b). Structure of the 80S ribosome from *Saccharomyces cerevisiae*--tRNA-ribosome and subunit-subunit interactions. *Cell* 107, 373-386.
- Spahn, C. M., Jan, E., Mulder, A., Grassucci, R. A., Sarnow, P. & Frank, J. (2004). Cryo-EM visualization of a viral internal ribosome entry site bound to human ribosomes: the IRES functions as an RNA-based translation factor. *Cell* 118, 465-475.
- Stanley, R. E., Blaha, G., Grodzicki, R. L., Strickler, M. D. & Steitz, T. A. (2010). The structures of the anti-tuberculosis antibiotics viomycin and capreomycin bound to the 70S ribosome. *Nat Struct Mol Biol* 17, 289-293.
- Szymanski, M. & Barciszewski, J. (2002). Beyond the proteome: non-coding regulatory RNAs. *Genome Biol* 3, reviews0005.
- Szymanski, M., Erdmann, V. A. & Barciszewski, J. (2007) Noncoding RNAs database (ncRNAdb). *Nucl Acids Res* 35, D162-D164.
- Temin, H. M. & Mizutani, S. (1970). RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* 226, 1211-1213.
- Thiel, H.-J., Collett, M. S., Gould, E. A., Heinz, F. X., Meyers, G., Purcell, R. H., Rice, C. M. & Houghton, M. (2005). Flaviviridae. In *Virus Taxonomy: VIIIth Report of the International Committee on Taxonomy of Viruses*, pp. 981-998. Edited by C. M. Fauquet, M. A. Mayo, J. Maniloff, U. Desselberger & L. A. Ball. San Diego, CA: Elsevier Academic Press.
- Tu, G. F., Reid, G. E., Zhang, J. G., Moritz, R. L. & Simpson, R. J. (1995). C-terminal extension of truncated recombinant proteins in *Escherichia coli* with a 10Sa RNA decapeptide. *J Biol Chem* 270, 9322-9326.
- Ushida, C., Himeno, H., Watanabe, T. & Muto, A. (1994). tRNA-like structures in 10Sa RNAs of *Mycoplasma capricolum* and *Bacillus subtilis*. *Nucleic Acids Res* 22, 3392-3396.

- Valle, M., Gillet, R., Kaur, S., Henne, A., Ramakrishnan, V. & Frank, J. (2003). Visualizing tmRNA entry into a stalled ribosome. *Science* 300, 127-130.
- Wang, C., Le, S. Y., Ali, N. & Siddiqui, A. (1995). An RNA pseudoknot is an essential structural element of the internal ribosome entry site located within the hepatitis C virus 5' noncoding region. *RNA* 1, 526-537.
- Wang, C., Sarnow, P. & Siddiqui, A. (1994). A conserved helical element is essential for internal initiation of translation of hepatitis C virus RNA. *J Virol* 68, 7301-7307.
- Wassarman, K. M. & Storz, G. (2000). 6S RNA regulates *E. coli* RNA polymerase activity. *Cell* 101, 613-623.
- Wilson, J. E., Powell, M. J., Hoover, S. E. & Sarnow, P. (2000). Naturally occurring dicistronic cricket paralysis virus RNA is regulated by two internal ribosome entry sites. *Mol Cell Biol* 20, 4990-4999.
- Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Jr., Morgan-Warren, R. J., Carter, A. P., Vornrhein, C., Hartsch, T. & Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. *Nature* 407, 327-339.
- Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H. & Noller, H. F. (2001). Crystal structure of the ribosome at 5.5 Å resolution. *Science* 292, 883-896.
- Zhang, A., Altuvia, S. & Storz, G. (1997). The novel oxyS RNA regulates expression of the sigma s subunit of Escherichia coli RNA polymerase. *Nucleic Acids Symp Ser*, 27-28.
- Zhang, A., Altuvia, S., Tiwari, A., Argaman, L., Hengge-Aronis, R. & Storz, G. (1998). The OxyS regulatory RNA represses rpoS translation and binds the Hfq (HF-I) protein. *EMBO J* 17, 6061-6068.
- Zhao, Q., Han, Q., Kissinger, C. R., Hermann, T. & Thompson, P. A. (2008). Structure of hepatitis C virus IRES subdomain IIa. *Acta Crystallogr D Biol Crystallogr* 64, 436-443.
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31, 3406-3415.
- Zwieb, C., Wower, I. & Wower, J. (1999). Comparative sequence analysis of tmRNA. *Nucleic Acids Res* 27, 2063-2071.

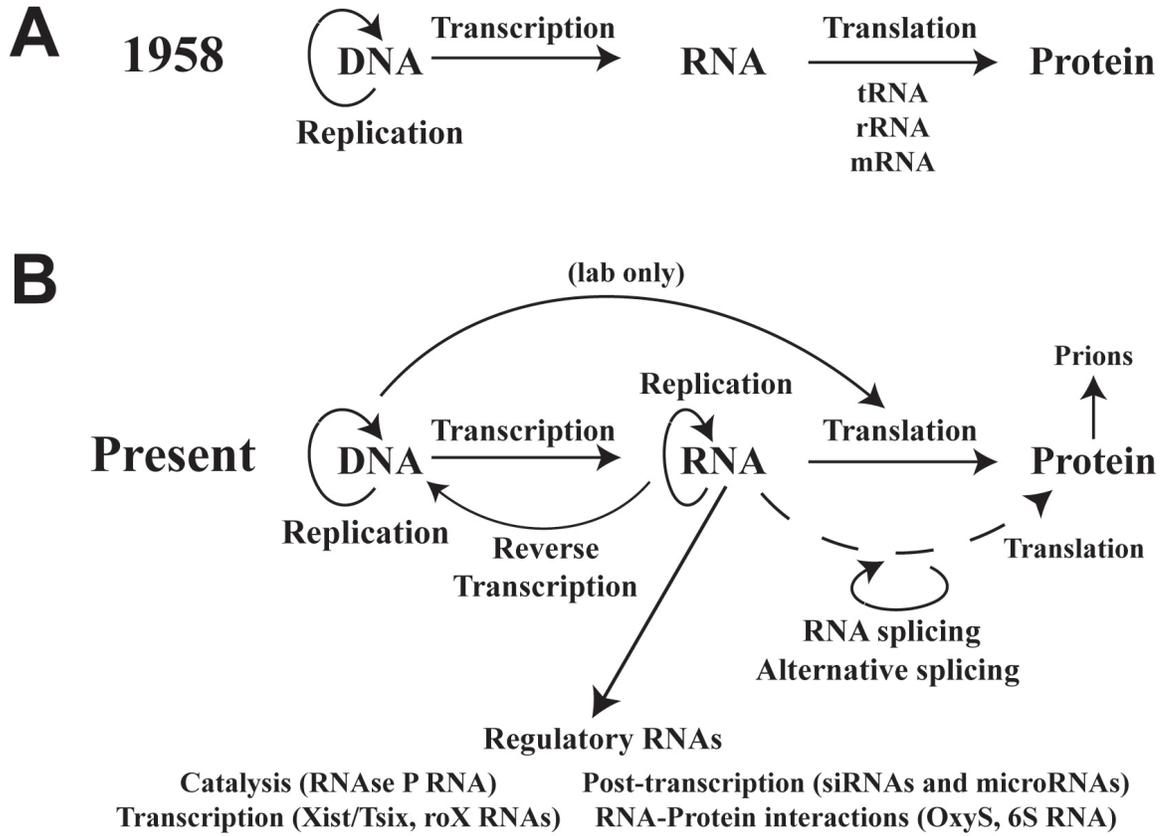


Figure 1.1: Central Dogma of Molecular Biology, then and now. **A)** The original one-gene, one-protein deterministic view of the flow of genetic information proposed by Crick. **B)** A modified scheme of the Central Dogma reflecting recent advances in our understanding of gene expression in both prokaryotes and eukaryotes.

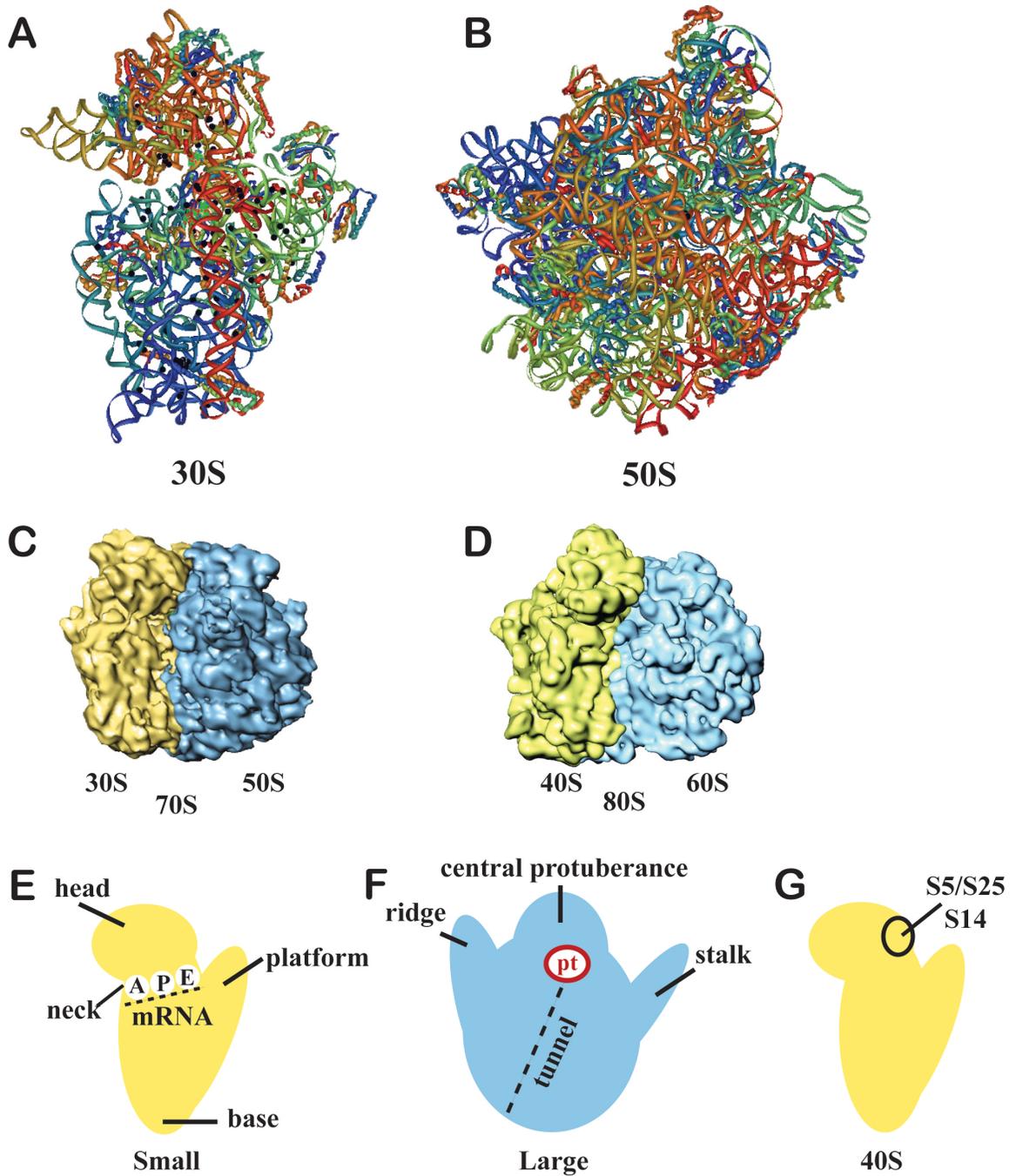


Figure 1.2: Structures of the ribosome and its subunits. **A)** 30S subunit from *Thermus thermophilus* bacterium (Carter *et al.*, 2000). **B)** 50S subunit from *Haloarcula marismortui* archeon (Nissen *et al.*, 2000). **C)** Cryo-EM reconstruction of 70S ribosome from *T. thermophilus* (Adapted from Kaur *et al.*, 2006. Copyright (2006) National

Academy of Sciences, U.S.A.). **D)** Cryo-EM reconstruction of 80S ribosome from *Trypanosoma brucei* (Adapted from Gao *et al.*, 2005. Copyright (2005) National Academy of Sciences, U.S.A.). 30S (yellow) and 50S (blue) subunits are indicated. 40S (yellow) and 60S (blue) subunits are indicated. **E)** General schematic of small ribosomal subunits showing locations of the head, neck, platform and base regions, along with the approximate locations of the A-, P- and E- tRNA binding sites and mRNA binding sites (dotted line). **F)** General schematic of large ribosomal subunits showing the ridge, stalk and central protuberance regions. Indicated are the approximate locations of the peptidyl transferase center (red “pt”) and nascent peptide exit tunnel. **G)** Approximate locations of ribosomal proteins S5, S25 and S14 on the eukaryotic 40S ribosomal subunit (Spahn *et al.* 2004).

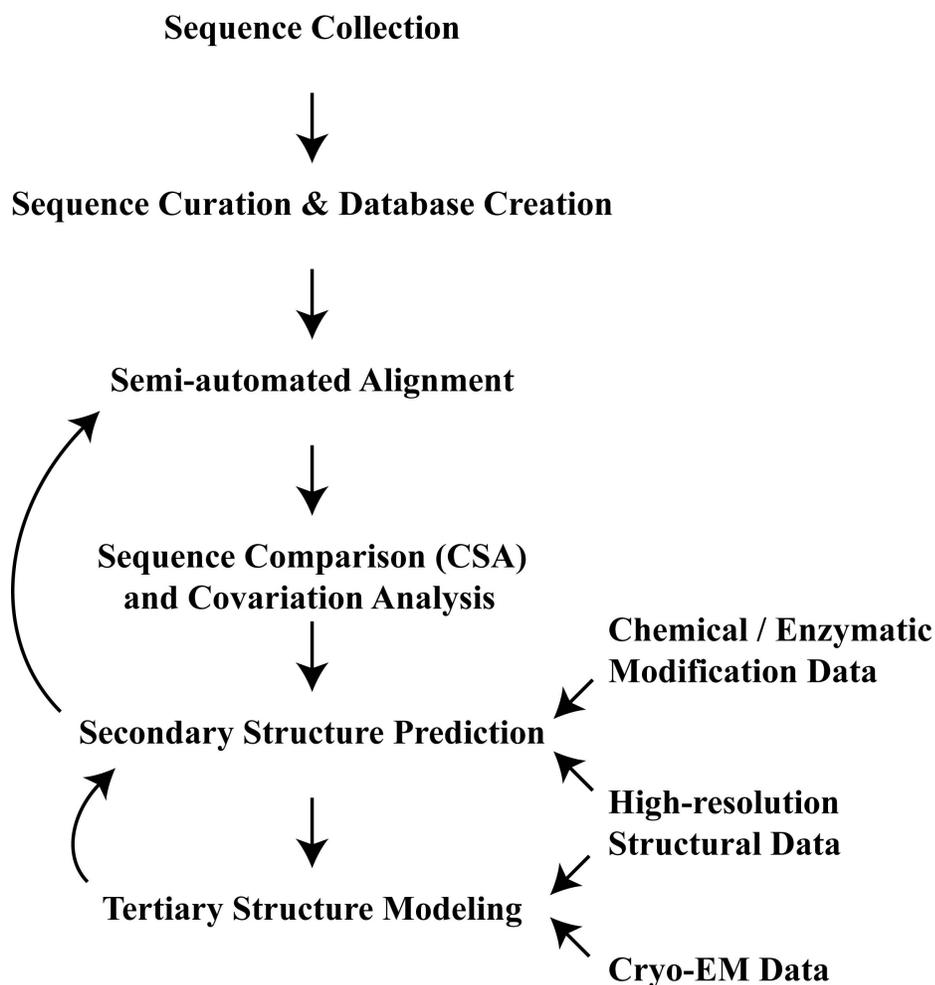


Figure 1.3: Flow of information from the RNA sequence to its tertiary structure model. Sequences were collected, curated and organized into a database (see Chapters 2-5 for details). Results from Comparative Sequence Analysis (CSA) and available structural modification data were used to define the minimal secondary structure of investigated RNAs for use in the structural modeling process (See Chapters 3, 4, and 5). After improving the secondary structure, tertiary structure modeling began. The knowledge base of RNA structure has rapidly expanded due to recent advances in crystallization and NMR techniques. Since nature reuses structural motifs, structural data obtained in previous studies were incorporated into the models in this project, creating

biologically relevant representations. Once the minimal structural models were created, protein interaction sites were cross-referenced in three dimensions to better constrain the models. The models were further refined by incorporation of additional information about the binding of these RNAs to the ribosome.

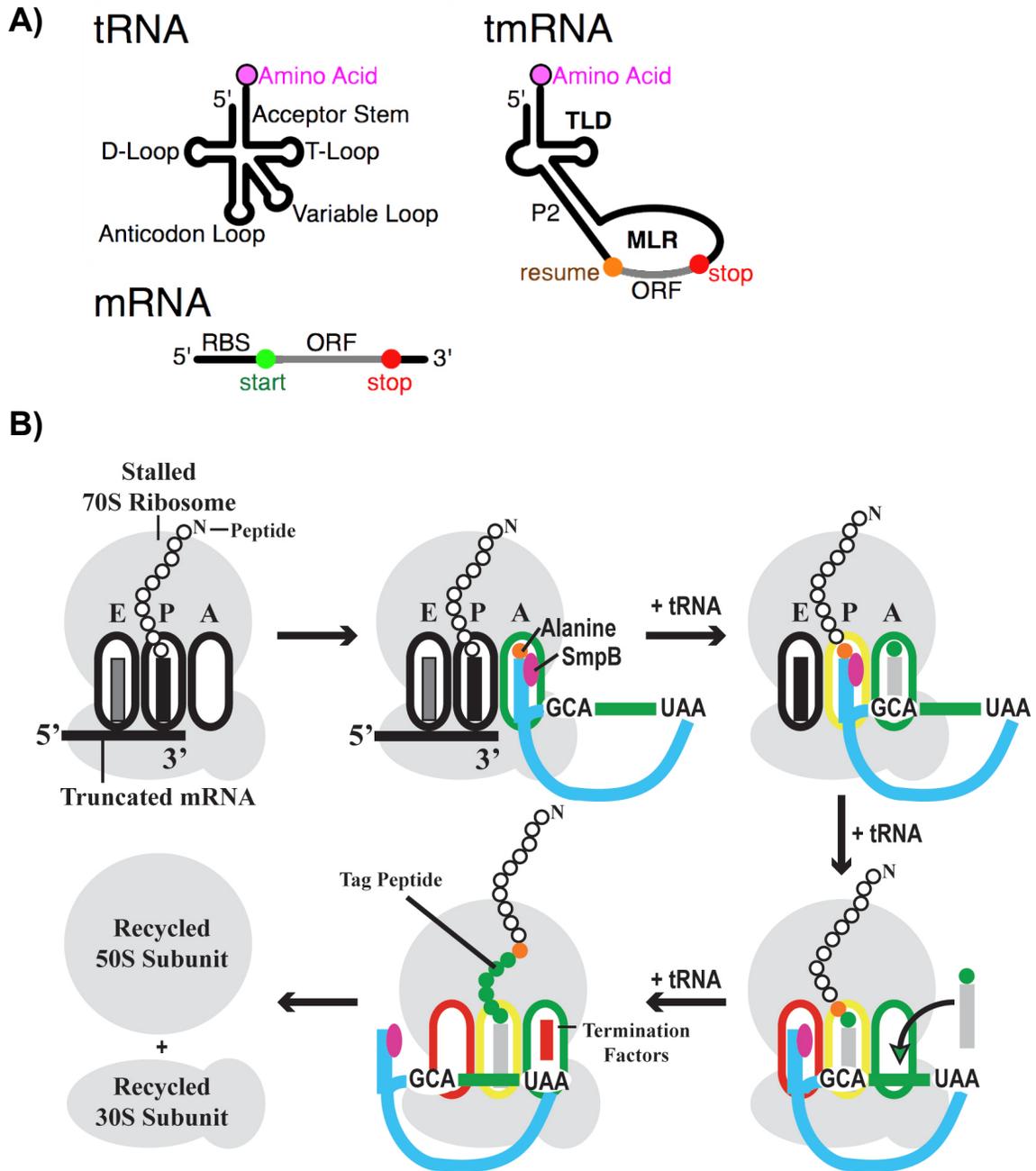


Figure 1.4: tmRNA features and *trans*-translation mechanism. **A)** Comparison of structural features of tmRNA, tRNA and mRNA. Adapted from the Wikipedia article at http://en.wikipedia.org/wiki/Transfer-messenger_RNA. **B)** *Trans*-translation, starts on 70S ribosome stalled at the 3' end of truncated mRNAs. Next, the Alanine-charged (orange) and SmpB-bound (pink) TLD of tmRNA (blue) binds to the empty A site of the

ribosome. After the growing peptide chain encoded by the broken mRNA is transferred to the alanine of the tmRNA, the TLD is shifted to the P site to allow the aminoacylated tRNA to bind to the resume codon in the MLR. Upon translocation, the TLD-SmpB complex moves to the E site. Translation proceeds until the stop codon of the tmRNA is reached, the ribosomal subunits dissociate, and the tagged peptide is released.

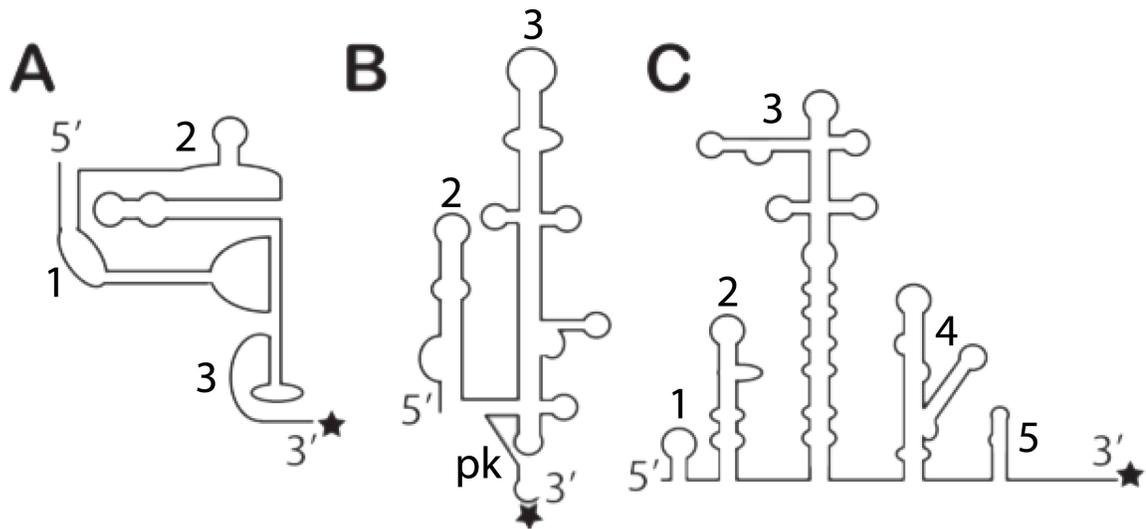


Figure 1.5: Secondary Structures of IRES RNAs. **A)** Secondary structure schematic of Intergenic region (IGR) IRES of CrPV-like viruses in the family *Dicistroviridae* showing domains 1-3. **B)** Secondary structure schematic of HCV IRES RNA showing helices 2, 3 and the pseudoknot. **C)** Secondary structure schematic of FMDV-type IRES RNA showing Domains 1-5. The 5' and 3' ends of RNAs are labeled, with start codons indicated with stars.

CHAPTER 2: THE tmRDB AND SRPDB RESOURCES

Adapted from Andersen, E. S., Rosenblad, M. A., Larsen, N., Westergaard, J. C., Burks, J., Wower, I. K., Wower, J., Gorodkin, J., Samuelsson, T., & Zwieb, C. (2006) *Nucleic Acids Res* 34, D163-168.

ABSTRACT

Maintained at the University of Texas Health Science Center at Tyler, Texas, the tmRNA database (tmRDB) is accessible at the URL <http://rnp.uthscsa.edu/rnp/tmRDB/tmRDB.html> with mirror sites located at Auburn University, Auburn, Alabama (<http://www.ag.auburn.edu/mirror/tmRDB/>) and the Royal Veterinary and Agricultural University, Denmark (<http://tmrdb.kvl.dk/>). The signal recognition particle database (SRPDB) at <http://rnp.uthscsa.edu/rnp/SRPDB/SRPDB.html> is mirrored at <http://srpdb.kvl.dk/> and the University of Goteborg (<http://bio.lundberg.gu.se/dbs/SRPDB/SRPDB.html>). The databases assist in investigations of the tmRNP (a ribonucleoprotein complex which liberates stalled bacterial ribosomes) and the SRP (a particle which recognizes signal sequences and directs secretory proteins to cell membranes). The curated tmRNA and SRP RNA alignments consider base pairs supported by comparative sequence analysis. Also shown are alignments of the tmRNA-associated proteins SmpB, ribosomal protein S1, alanyl-tRNA synthetase and Elongation Factor Tu, as well as the SRP proteins SRP9, SRP14, SRP19, SRP21, SRP54 (Ffh), SRP68, SRP72, cpSRP43, Flhf, SRP receptor (alpha) and SRP receptor (beta). All alignments can be easily examined using a new exploratory browser. The databases provide links to high-resolution structures and serve as depositories for structures obtained by molecular modeling.

INTRODUCTION

Ribosomes extend their repertoire of functions by binding to additional ribonucleoprotein particles (RNPs) that can determine the fate of the protein as it emerges from the large ribosomal subunit. Two such complexes are the transfer-messenger RNP (tmRNP) and the signal recognition particle (SRP). The tmRNP, composed of the tmRNA, small protein B (SmpB) and ribosomal protein S1, rescues bacterial ribosomes stalled on faulty mRNAs. The potentially damaging polypeptides are tagged with a short peptide, released from the ribosome and destroyed by intracellular proteases (reviewed in Karzai *et al.*, 2000). Similarly, the SRP binds to emerging signal sequences and directs secretory protein to cellular membranes (recently reviewed in Halic & Beckmann, 2005). The investigations of tmRNP and SRP combined with the knowledge gained from the high-resolution structures of the ribosome (Ban *et al.*, 2000; Ramakrishnan, 2002; Yusupov *et al.*, 2001) have contributed significantly to our understanding of protein translation and translocation, but many questions remain to be answered. To assist in these ongoing studies, the updated tmRDB and SRPDB resources offer detailed descriptions of the biological roles of tmRNP and SRP, ordered lists of the components and links to high-resolution structures. Secondary structures are supported by comparative sequence analysis. A new alignment browser allows the user to easily explore the alignments.

RESULTS AND DISCUSSION

tmRNA Genes

The tmRDB contains a total of 555 tmRNA sequences in the range of 250 to 434 nucleotides. Because of the continuous rapid emergence of new sequences this dataset is not complete but nevertheless representative. (The tmRNA Website (Gueneau de Novoa & Williams, 2004) should be consulted for the most recent update of new tmRNA sequences.) All bacterial groups, including the Alphaproteobacteria (55 sequences) previously thought to lack tmRNA were found to contain tmRNA genes. Consistent with the evolutionary relationship between bacteria and organelles, tmRNAs were found in most of the chloroplast and mitochondrial genomes. However, tmRNA genes were lacking in the chloroplasts of higher plants. Interestingly, tmRNAs could be identified in the genomes of certain bacteriophages. Most tmRNAs were composed of one continuous molecule, but most Alphaproteobacteria, some Cyanobacteria and some Betaproteobacteria encoded their tmRNA in two parts (Table 2.1), suggesting that this adaptation arose independently (Sharkady & Williams, 2004). The lone one-piece potential alphaproteobacteria *Magnetococcus* MC-1 tmRNA may belong in a separate subgroup of one-piece alphaproteobacteria but there is currently not enough sequence information to justify the creation of a separate subgroup. No tmRNAs were identified in the Archaea or the nuclear genomes of the Eukarya.

Features of tmRNA

The tmRNA sequences were aligned using comparative sequence analysis as described previously for SRP RNA (Larsen & Zwieb, 1991). An outline of the secondary

structure of *Escherichia coli* tmRNA is depicted in Figure 2.1 (left portion). Shown are the tRNA-like domain (TLD), the messenger RNA-like region (MLR), and the pseudoknot (pk) domain (PKD). Modification to the *E. coli* reference structure include the reduction or disappearance of pseudoknots, the appearance of new helices (e.g. in the pk2 of Betaproteobacteria), and complete structural replacements (replacement of pk4 with two tandem pseudoknots in Cyanobacteria). The phylogenetic distribution of the features is summarized in Table 2.1.

tmRNA-encoded Tag-peptides

The 539 tmRNA-encoded tag peptides were characterized by a cluster of hydrophobic amino acids at the C-terminus and a variable length of 8 to 35 amino acids. The resume codon coded predominately for Alanine (474) while the remainder of resume codons coded for Glycine (53), Aspartic acid (5), Valine (1), Leucine (1), Isoleucine (1), Arginine (1), Glutamic Acid (1), Serine (1), or Threonine (1). Whereas tmRNAs usually contain one or two in-frame stop codons in the MLR, the tmRNAs from *Chloroflexus aurantiacus*, *Bacillus megaterium* and *Pediococcus pentosaceus* contain three in-frame stop codons. With the exception of the tag-peptides of *Escherichia coli* and *Bacillus subtilis*, the predicted tag peptide sequences have remained experimentally unconfirmed.

tmRNA-associated Proteins

SmpB

This protein is an essential *trans*-translational co-factor (Karzai *et al.*, 1999) and present in all bacteria. The protein forms quaternary complexes with aminoacylated

tmRNA, EF-Tu and GTP. SmpB mutants which lack the C-terminal tail of the protein bind to ribosomes, but are unable to tag truncated proteins.

Ribosomal Protein S1

Ribosomal protein S1 contains up to six related domains and binds and cross-links to the MLR and pk2 to pk4. The NMR structure of a single Protein S1 RNA-binding domain of *E. coli* has been determined (Bycroft *et al.*, 1997), but little is known about the relative arrangement of the S1 domains at the different functional stages. The alignment suggested four groups of sequences which differed in the number of domains. Overall, domains four, five and six were less conserved and absent in some organisms. The sequences of *Candidatus tremblaya princeps* and *Clostridium acetobutylicum* ATCC 824 stood out as they did not fit well with either of the groups.

Alanyl-tRNA Synthetase

Aminoacylation of tmRNA constitutes a prerequisite step in *trans*-translation, since uncharged tmRNA mutants do not bind to 70S ribosomes *in vivo*. Studies carried out *in vitro* demonstrated that the aminoacyl moiety can be changed without affecting the ability of the tmRNA to participate in protein tagging. The majority of the tmRNAs are expected to be charged with alanine because they possess in their acceptor stem a G-U basepair as the critical determinant for aminoacylation with Alanyl-tRNA Synthetase.

EF-Tu

Elongation factor Tu, found in Bacteria and Eukaryota, forms a ternary complex with GTP and Ala-tmRNA *in vitro* as in regular protein synthesis. Although the association rate constant of Ala-tmRNA for EF-Tu-GTP is lower than that of Ala-tRNA, chemical and enzymatic footprinting indicate that the architecture of this complex closely resembles canonical ternary complexes. EF-Tu primarily interacts with the acceptor arm of the tRNA-like domain of tmRNA.

Phylogeny of tmRNP

A description of the phylogenetic distribution of the secondary structural features of tmRNA based on an alignment of 274 sequences was provided recently (Burks *et al.*, 2005). From the analysis of a total of 555 sequences the following insights into tmRNA phylogeny were obtained with examples provided in Table 2.2: (1) Most tmRNAs consist of a single polynucleotide with a TLD, a relatively unstructured MLR, and a variable number of pseudoknots. Variations in the PKD suggest a preservation of RNA folding without the need for sequence conservation. (2) In Alphaproteobacteria, some Betaproteobacteria, and some Cyanobacteria, the tmRNAs are composed of two chains. These two-piece tmRNAs contain fewer pseudoknots than the typical one-piece tmRNAs. (3) Plastid tmRNAs, unlike their one-piece Cyanobacterium progenitors, have one-piece with a reduced number of pseudoknots. (4) Most mitochondria may be devoid of *trans*-translation because they lack SmpB and contain only very short two-piece tmRNAs which appear to have lost the MLR. A secondary structure of the one-piece *E. coli* tmRNA is seen in Chapter 2 (Figure 2.1).

SRP RNA Genes

A total of 393 SRP RNAs were identified using the procedures described in Materials and Methods. The sequences were arranged into 30 phylogenetic groups including the photosynthetic plastids of red algal origin (except the substantially smaller plastid of the haptophyte *Emiliana huxleyi*) and the chloroplasts of some green algae. 33 organisms had more than one variant. Many novel SRP RNA sequences were found to add to our knowledge of the phylogenetic distribution of the secondary structure features (Table 2.2).

SRP RNA Features

A overview of the SRP RNA secondary structure elements was presented in a recent nomenclature proposal (Zwieb *et al.*, 2005) similar to what is shown in Figure 2.1 (right portion). Several new sequences, e.g. from *Eremothecium gossypii*, *Kluyveromyces waltii* and *K. lactis*, provided additional support for the proposed helices. In the Onygenales group, within Pezizomycotina (*Histoplasma*) and four other species, we found a new helix located toward the 5' end of helix 6. The phylogenetic distribution of the helices feature is shown in Table 2.2. A representative SRP RNA secondary structure diagrams is shown in Figure 2.1B.

Most bacteria, including certain chloroplasts, contained a small SRP RNA of 60 to 115 nucleotides consisting solely of helix 8. The conserved apical tetraloop of this helix typically had the consensus sequence GNRA, with a rare G to U mutations in the first position, but occasionally an URRC (Regalia *et al.*, 2002). In some gram-positive

bacteria (Bacillales and Clostridia groups) and the deeply-branching gram-negative bacteria *Thermotoga maritima*, the SRP RNA was of the archaeal type, but lacked helix 6. The archaeal SRP RNA had a small domain similar to the Alu domain of eukaryotes with a non-consensus UGUNR motif (sometimes UAUNR or CNNNR). In certain Chrenarcheota (*Aeropyrum pernix*) this part seemed to be extended, perhaps forming a helix. The apical loop of the highly conserved helix 8 consisted of four nucleotides in most organisms. Plants and certain fungi, however, possessed six nucleotides in this loop. Recently, we found that *Trichomonas*, *Phytophthora*, and *Entamoeba* have a pentaloop with the consensus sequence G[AT][AT]AA.

The eukaryal SRP RNA was highly variable, particularly with respect to the Alu domain. Secondary structure models were recently presented for the *Saccharomyces* SRP RNAs (Rosenblad *et al.*, 2004; Van Nues & Brown, 2004). These models showed that helices 3 and 4 were missing, and additional helices 9 to 12 had been acquired. We showed recently that the SRP RNA secondary structures of the non-Ascomycota fungi *Phakopsora* and *Rhizopus*, differed from the Ascomycota and were similar to the metazoan SRP RNAs. In Diplomonads and Microsporidia, the small (Alu) domain seemed to have disappeared to leave an SRP RNA composed only of the large (S) domain.

SRP Proteins

SRP9, SRP14 and SRP21

A total of 24 SRP9 protein sequences were identified: 16 sequences from the Metazoa, one each from *Dictyostelium discoideum* and *Entamoeba histolytica*, three from plants, and three from the Alveolata group. SRP14 (a total of 33 sequences) was found in all of eukaryotes examined, including the Fungi. Both SRP9 and SRP14 were absent in all Bacteria and Archaea and some eukaryal groups. SRP21 sequences were identified in 12 fungal genomes. Evidence has been provided that the metazoan SRP9 is homologous to the fungal SRP21 (Rosenblad *et al.*, 2004). This finding was consistent with the indication that a gradual change from SRP9 to SRP21 had occurred in evolution with Pezizomycotina and *Schizosaccharomyces pombe* representing intermediate. However, further studies are required to clarify the functional role of SRP21 in fungi.

SRP19

Protein SRP19 was found in all the examined Eukarya and Archaea. The presence of SRP19 correlated strongly with the appearance of SRP RNA helix 6, thus confirming the important role of SRP19 in the assembly of the large (S) domain (Walter *et al.*, 1983).

SRP54

SRP54, also referred to in Bacteria as Ffh (fifty-four homologue), contains a signal sequence binding pocket (Keenan *et al.*, 1998) and thus is likely to be an essential component of every SRP. The SRPDB lists 115 sequences from all phylogenetic groups.

We identified homologs to the chloroplast Ffh, cpSRP54, in *Arabidopsis*, *Pisum*, *Chlamydomonas* and *Cyanidioschyzon merolae*.

SRP68 and SRP72

31 SRP68 and 34 SRP72 sequences from the Fungi, Metazoa, Mycetozoa, Plants, Alveolata, and Euglenozoa groups were found. Recognizable homologues of these proteins were absent in the Bacteria and Archaea. Both proteins are known to form a heterodimer within the large domain of the mammalian SRP, but relatively little is known about its structure. The SRP72 alignment revealed a new lysine-rich domain, originally identified as Pfam B 7529, which will be added to Pfam (Bateman *et al.*, 2004). A corresponding peptide of 63 amino acids located near the C-terminus of human SRP72 with the consensus PDPXRWLPXXER was shown to bind to SRP RNA with high affinity (Iakhiaeva *et al.*, 2005).

cpSRP43

cpSRP43 is a unique nuclear encoded protein only found in chloroplasts. The protein binds to polypeptides imported to the chloroplast and destined for the thylakoid membrane. cpSRP43 contains four ankyrin repeats at the N-terminus and two chromodomains at the C terminus. It forms a complex with cpSRP54 *via* its chromodomains (Schuenemann *et al.*, 1998).

SRP-associated Proteins

SRP Receptor (alpha) (FtsY)

The SRP receptor is a single polypeptide (FtsY) in the Bacteria and Archaea. In Eukaryotes, the SRP receptor is composed of two subunits, alpha and beta. The alpha subunit is related to FtsY and to SRP54 (Ffh) due to their GTPase domain similarity. Unique to SRP Receptor-alpha (FtsY) are an N-terminal A-region which is thought to be responsible for interacting with the membrane or the beta-subunit of the SRP receptor (reviewed in Halic & Beckmann, 2005).

SRP Receptor (beta)

SRP Receptor (beta) was found in all Eukaryotes including the Fungi. The protein contains a transmembrane anchor and binds to the alpha-subunit of the receptor. Like SRP54 (Ffh), and the alpha subunit (FtsY), the beta subunit also contains a GTP domain.

FlhF

This protein was characterized first as a flagellar gene from *Bacillus subtilis* belonging to the same family of GTP-binding proteins as Ffh and FtsY (Carpenter *et al.*, 1992) suggesting a role on SRP function. However, FlhF was shown recently to be dispensable for protein secretion (Zanen *et al.*, 2004).

Phylogeny of SRP

An extensive inventory of SRP RNA and protein components has allowed us to arrive at a comprehensive view of SRP phylogeny (Table 2.2). Essential elements include

(1) the development of an altered Alu domain in the Ascomycota lacking helices 3 and 4, accompanied by the appearance of protein SRP21. (2) The emergence of the more complex *Saccharomyces* SRP RNAs containing additional helix insertion into helix 5. (3) The retention of a metazoan-type SRP in the Basidiomycota. (4) The appearance of eukaryotic SRPs that lack the typical SRP proteins or the small (Alu) domain. (5) The presence of a much reduced SRP in bacteria and chloroplasts composed of a small RNA and only one protein (Ffh). (6) The conservation of the composition and secondary structure of the Archaeal SRP.

OUTLOOK

Exploring RNA and protein alignments has become increasingly difficult with the growing number of sequences. We have now implemented a browser which allows to display alignments like a map at various zoom levels. The user of the databases can now see more clearly the species- and group-specific differences and focus on features of interest. This tool encourages exploration and is expected to further improve the quality of the alignments.

ACCESS

The data are freely accessible for research purposes at the internet addresses <http://rnp.uthscsa.edu/rnp/tmRDB/tmRDB.html> and <http://rnp.uthscsa.edu/rnp/SRPDB/SRPDB.html> or at the corresponding mirror sites provided in the Abstract. This article should be cited in research projects which use of the tmRDB and SRPDB resources.

MATERIALS AND METHODS

Comparative Sequence Analysis of RNA

New tmRNA sequences as identified at the tmRNA Website (Gueneau de Novoa & Williams, 2004) were merged with the previous tmRNA alignment (Zwieb *et al.*, 2003). New SRP RNAs were identified using SRPscan (Regalia *et al.*, 2002) as well as BLAST (McGinnis & Madden, 2004), RNABOB (Eddy, unpublished), Infernal (Eddy, 2002) and MFOLD (Zuker, 2003). The sequences were placed in phylogenetic order guided by the NCBI Taxonomy (Benson *et al.*, 2000; Wheeler *et al.*, 2000), SARSE (Andersen *et al.*, 2007) and BioEdit (Hall, 1999). Sequences classified as "Unclassified" in NCBI Taxonomy were submitted to BLAST at the tmRNA Website to determine the closest relative based on sequence similarity. Unclassified sequences were aligned to the closest relative based on the results of the BLAST search. Sequences were aligned automatically with CLUSTAL (Higgins *et al.*, 1996) or manually by observing the previously described rules (Larsen & Zwieb, 1991). RNAdbtools (Gorodkin *et al.*, 2001) was applied to confirm compensatory base changes, and check base pairing consistencies, and possible RNA helix extensions.

Protein Alignments

Protein sequences were identified in GenBank (Benson *et al.*, 2000) using BLAST (McGinnis & Madden, 2004) with a subset of representative sequences from the previous versions of tmRDB (Zwieb *et al.*, 2003) and SRPDB (Rosenblad *et al.*, 2003) as queries. The output was examined manually to generate a set of unique sequences for

each protein family. Sequences were aligned using Jalview (Clamp *et al.*, 2004) and CLUSTAL (Higgins *et al.*, 1996) or MUSCLE (Edgar, 2004).

Alignment Browser

The alignments can be viewed, zoomed and scrolled in a WWW-browser under development for genomes by Danish Genome Institute (also directly accessible at <http://www.genomics.dk:8000/RNA>). It currently features only basic navigation, with color-dot, grey-dot and character display, and zoom to any level. Painting of features will be added.

Acknowledgements

We thank Jorgen Kjems for RNA expertise and support, and Florian Müller for the ERNA-3D modeling program. This work was supported by NIH grants GM-58267 to J.W. and GM-49034 to C.Z. J.G. is supported by The Danish Research Council for Technology and Production Sciences and the Danish Center for Scientific Computing. Funding to pay the Open Access publication charges for this article was provided by NIH grant GM-49034 to C.Z.

REFERENCES

- Andersen, E. S., Lind-Thomsen, A., Knudsen, B., Kristensen, S. E., Havgaard, J. H., Torarinsson, E., Larsen, N., Zwieb, C., Sestoft, P., Kjems, J., & Gorodkin, J. (2007). Semiautomated improvement of RNA alignments *RNA* 13, 1850-1859.
- Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* 289, 905-920.
- Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., Studholme, D. J., Yeats, C. & Eddy, S. R. (2004). The Pfam protein families database. *Nucleic Acids Res* 32, D138-141.
- Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Rapp, B. A. & Wheeler, D. L. (2000). GenBank. *Nucleic Acids Res* 28, 15-18.
- Burks, J., Zwieb, C., Muller, F., Wower, I. & Wower, J. (2005). Comparative 3-D modeling of tmRNA. *BMC Mol Biol* 6, 14.
- Bycroft, M., Hubbard, T. J., Proctor, M., Freund, S. M. & Murzin, A. G. (1997). The solution structure of the S1 RNA binding domain: a member of an ancient nucleic acid-binding fold. *Cell* 88, 235-242.
- Carpenter, P. B., Hanlon, D. W. & Ordal, G. W. (1992). flhF, a Bacillus subtilis flagellar gene that encodes a putative GTP-binding protein. *Mol Microbiol* 6, 2705-2713.
- Clamp, M., Cuff, J., Searle, S. M. & Barton, G. J. (2004). The Jalview Java alignment editor. *Bioinformatics* 20, 426-427.
- Eddy, S. R. (2002). A memory-efficient dynamic programming algorithm for optimal alignment of a sequence to an RNA secondary structure. *BMC Bioinformatics* 3, 18.
- Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5, 113.
- Gorodkin, J., Zwieb, C. & Knudsen, B. (2001). Semi-automated update and cleanup of structural RNA alignment databases. *Bioinformatics* 17, 642-645.
- Gueneau de Novoa, P. & Williams, K. P. (2004). The tmRNA website: reductive evolution of tmRNA in plastids and other endosymbionts. *Nucleic Acids Res* 32 Database issue, D104-108.
- Halic, M. & Beckmann, R. (2005). The signal recognition particle and its interactions during protein targeting. *Curr Opin Struct Biol* 15, 116-125.

- Hall, T. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 41, 95-98.
- Higgins, D. G., Thompson, J. D. & Gibson, T. J. (1996). Using CLUSTAL for multiple sequence alignments. *Methods Enzymol* 266, 383-402.
- Iakhiaeva, E., Yin, J. & Zwieb, C. (2005). Identification of an RNA-binding Domain in Human SRP72. *J Mol Biol* 345, 659-666.
- Karzai, A. W., Roche, E. D. & Sauer, R. T. (2000). The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue. *Nat Struct Biol* 7, 449-455.
- Karzai, A. W., Susskind, M. M. & Sauer, R. T. (1999). SmpB, a unique RNA-binding protein essential for the peptide-tagging activity of SsrA (tmRNA). *Embo J* 18, 3793-3799.
- Keenan, R. J., Freymann, D. M., Walter, P. & Stroud, R. M. (1998). Crystal structure of the signal sequence binding subunit of the signal recognition particle. *Cell* 94, 181-191.
- Larsen, N. & Zwieb, C. (1991). SRP-RNA sequence alignment and secondary structure. *Nucleic Acids Res* 19, 209-215.
- McGinnis, S. & Madden, T. L. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res* 32, W20-25.
- Ramakrishnan, V. (2002). Ribosome structure and the mechanism of translation. *Cell* 108, 557-572.
- Regalia, M., Rosenblad, M. A. & Samuelsson, T. (2002). Prediction of signal recognition particle RNA genes. *Nucleic Acids Res* 30, 3368-3377.
- Rosenblad, M. A., Gorodkin, J., Knudsen, B., Zwieb, C. & Samuelsson, T. (2003). SRPDB: Signal Recognition Particle Database. *Nucleic Acids Res* 31, 363-364.
- Rosenblad, M. A., Zwieb, C. & Samuelsson, T. (2004). Identification and comparative analysis of components from the signal recognition particle in protozoa and fungi. *BMC Genomics* 5, 5.
- Schuenemann, D., Gupta, S., Persello-Cartieaux, F., Klimyuk, V. I., Jones, J. D., Nussaume, L. & Hoffman, N. E. (1998). A novel signal recognition particle targets light-harvesting proteins to the thylakoid membranes. *Proc Natl Acad Sci USA* 95, 10312-10316.
- Sharkady, S. M. & Williams, K. P. (2004). A third lineage with two-piece tmRNA. *Nucleic Acids Res* 32, 4531-4538.

- Van Nues, R. W. & Brown, J. D. (2004). Saccharomyces SRP RNA secondary structures: a conserved S-domain and extended Alu-domain. *RNA* 10, 75-89.
- Walter, P. & Blobel, G. (1983) Disassembly and reconstitution of signal recognition particle. *Cell* 34, 525-533.
- Wheeler, D. L., Chappey, C., Lash, A. E., Leipe, D. D., Madden, T. L., Schuler, G. D., Tatusova, T. A. & Rapp, B. A. (2000). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 28, 10-14.
- Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H. & Noller, H. F. (2001). Crystal structure of the ribosome at 5.5 Å resolution. *Science* 292, 883-896.
- Zanen, G., Antelmann, H., Westers, H., Hecker, M., van Dijl, J. M. & Quax, W. J. (2004). FlhF, the third signal recognition particle-GTPase of Bacillus subtilis, is dispensable for protein secretion. *J Bacteriol* 186, 5956-5960.
- Zwieb, C., van Nues, R. W., Rosenblad, M. A., Brown, J. D. & Samuelsson T. (2005) A nomenclature for all signal recognition particle RNAs. *RNA* 11, 7-13.
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31, 3406-3415.
- Zwieb, C., Gorodkin, J., Knudsen, B., Burks, J. & Wower, J. (2003). tmRDB (tmRNA database). *Nucleic Acids Res* 31, 446-447.

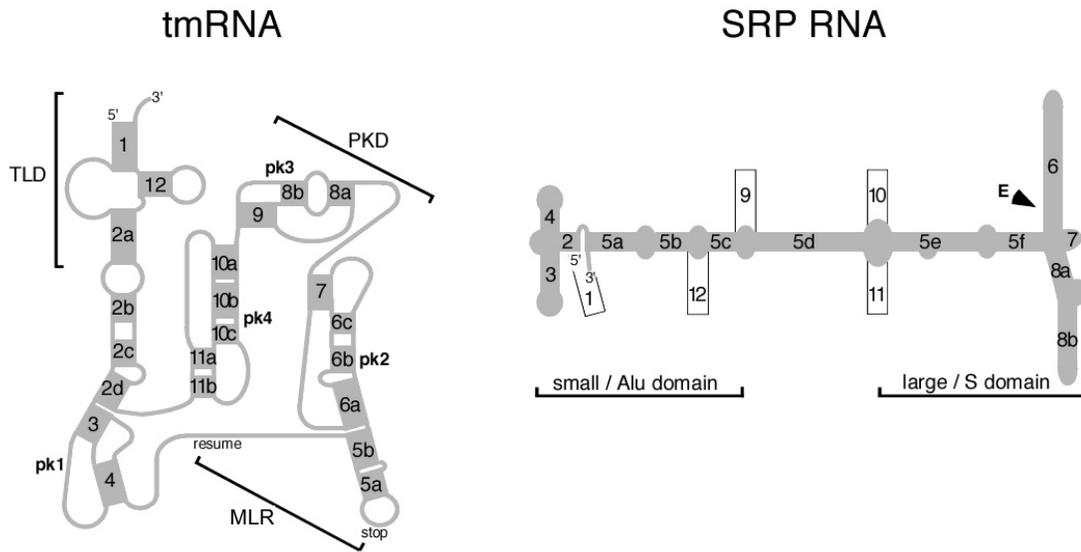


Figure 2.1. Schematic representation of the secondary structures of *Escherichia coli* tmRNA (left) and SRP RNA (right). The tRNA-like (TLD) and pseudoknot (PKD) domains and mRNA-like region (MLR) are indicated. Helices and their sections are numbered from 1 to 12 and letters a to d. The four pseudoknots are labeled pk1 to pk4. The tag peptide-encoding region is located between the resume and stop codon as indicated. In the SRP RNA schematic on the right, the features of the mammalian SRP RNA are shown in gray. Helices are numbered from 1 to 12 with helical sections labeled with letters a to f. The approximate boundaries of the small (Alu) and the large (S) domain are shown. The recently discovered extra helix (E) in the SRP RNAs of the Euglenozoa (see Table 2.2) is indicated by the arrowhead.

2D	Group Species	1	2	3	4	M	5	6	7	8	9	=	10	11	12	SB	S1	RS	Tu	
A	Bacteriophages <i>Bacillus subtilis</i> phage G	x	x	x	x	x	x	x	x	x	x	-	-	-	x	-	-	-	-	
	Bacteriophages CP1639	x	x	x	x	x	x	x	x	x	x	-	x	X	x	x	-	-	-	
	Aquificae <i>Aquifex aeolicus</i>	x	x	x	x	x	x	x	x	x	x	-	x	X	x	x	x	x	x	
	Deinococcus-Thermus <i>Thermus thermophilus</i>	x	x	x	x	x	x	x	x	x	x	-	x	x	x	x	x	x	x	
	Thermodesulfobacteria	x	x	x	x	x	x	x	x	x	x	-	x	x	x	x	x	x	x	
	<i>Thermodesulfobacterium commune</i>																			
	Thermatogae <i>Thermatoga maritima</i>	x	x	x	x	x	x	x	x	x	x	-	x	x	x	x	x	x	x	
	Planctomyces <i>Rhodopirellula baltica</i>	x	x	x	x	x	x	!	!	x	x	-	x	x	x	x	x	x	x	
	Clamydiae/Verrucomicrobia <i>Chlamydia trachomatis</i>	x	x	x	x	x	x	x	x	x	x	-	x	x	x	x	x	x	x	
	Chloroflexi <i>Chloroflexus aurantiacus</i>	x	x	x	x	x	x	x	x	x	x	-	x	x	x	x	x	x	x	
Bacteroides/Chlorobi <i>Bacteroides fragilis</i>	x	x	x	x	x	x	x	x	x	x	-	x	x	x	x	x	x	x		
Bacteroides/Chlorobi <i>Salinibacter ruber</i>	x	x	x	x	x	x	x	?	x	x	-	x	x	x	x	x	x	x		
b	Cyanobacteria <i>Synechystis PCC6803</i>	x	x	x	x	x	x	x	x	x	-	pp	pp	x	x	x	x	x		
	Cyanobacteria <i>Cyanobium gracilis</i>	x	x	x	x	x	x	?	-	-	x	-	-	x	x	x	x	x		
c	Organelles/Chloroplasts <i>Guillardia theta</i>	x	x	x	x	x	x	x	-	-	-	-	-	x	x	x	x	x		
	Organelles/Chloroplasts <i>Thalassiosira pseudonana</i>	x	x	x	x	x	-	-	-	-	-	-	-	x	x	x	x	x		
	Organelles/Mitochondria <i>Reclinomonas americana</i>	x	x	-	-	-	-	-	-	-	-	!	-	-	x	-	x	x		
	Organelles/Mitochondria <i>Jakoba libera</i>	x	x	-	-	-	-	-	-	-	-	-	-	x	-	x	x	x		
	Fibrobacteres/Acidobacteria <i>Fibrobacter succinogenes</i>	x	x	x	x	x	x	x	x	x	-	x	x	x	x	x	x	x		
	Spirochaetes <i>Treponema pallidum</i>	x	x	x	x	x	x	x	x	x	-	x	x	x	x	x	x	x		
	Nitrospirae <i>Leptospirillum species</i>	x	x	x	x	?	?	x	x	x	-	x	x	x	x	x	x	x		
d	Alphaproteobacteria <i>Caulobacter crescentus</i>	x	x	x	x	x	x	x	x	x	x	?	x	x	x	x	x	x		
	Alphaproteobacteria <i>Magnetococcus MC-1</i>	x	x	x	x	x	x	-	x	x	-	-	x	x	x	x	x	x		
	Betaproteobacteria <i>Dechloromonas aromatica</i>	x	x	?	?	x	x	-	-	-	-	x	?	?	x	x	x	x		
	Betaproteobacteria <i>Tremblaya princeps</i>	x	x	x	x	x	-	x	x	x	-	x	x	x	x	x	x	x		
	Betaproteobacteria <i>Neisseria gonorrhoeae</i>	x	x	x	x	x	e	-	-	-	-	?	?	!	x	x	x	x		
	Gammaproteobacteria <i>Francisella tularensis</i>	x	x	x	x	x	e	x	x	x	-	x	x	x	x	x	x	x		
	e	Gammaproteobacteria <i>Escherichia coli</i>	x	x	x	x	x	x	x	x	x	-	x	x	x	x	x	x	x	
		Deltaproteobacteria <i>Geobacter metallireducens</i>	x	x	x	x	x	x	x	x	x	-	x	x	x	x	x	x	x	
		Epsilonproteobacteria <i>Campylobacter jejuni</i>	x	x	x	x	x	x	x	x	x	-	x	x	x	x	x	x	x	
		Fusobacteria <i>Fusobacterium nucleatum</i>	x	x	x	x	x	x	x	x	x	-	x	x	x	x	x	x	x	
	Dictyoglomi <i>Dictyoglomus thermophilum</i>	x	x	x	x	x	x	x	x	?	-	?	x	x	x	x	x	x		
	Actinomycetes <i>Mycobacterium avium</i>	x	x	x	x	x	x	x	x	x	-	x	x	x	x	x	x	x		
	Firmicutes/Bacilli <i>B.subtilis</i>	x	x	x	x	x	x	x	x	x	-	x	x	x	x	x	x	x		
	Firmicutes/Clostridia <i>Clostridium botulinum</i>	x	x	x	x	x	x	x	x	x	-	x	x	x	x	x	x	x		

Table 2.1. tmRNA Features and Representatives. The names of representative species are given for each phylogenetic group in the tmRDB. The column labeled ‘2D’ marks five tmRNA secondary structure examples *a–e* which are shown in more detail in Supplementary Data 1 available at http://nar.oxfordjournals.org/content/34/suppl_1/D163/suppl/DC1. The tmRNA features (helices numbered from 1 to 12) are

shown in the center part of the table. '=' indicates the interruption in the two-part tmRNAs. SB, Protein SmpB; S1, ribosomal protein S1 and its homologues; RS, alanyl-tRNA synthetase; Tu, Elongation Factor Tu. The table cells are annotated as '-', absent; '?', maybe absent or was not found; '!', expected to be present, and 'x', present. 'e' denotes an extra helix; 'pp' is for a tandem pseudoknot.

2D	Group Species	1	2	3	4	5	6	7	8	9	10	11	12	E	T	9	21	14	19	54	68	72	cp54	cp43	
	Plastids <i>Cyanidioschyzon merolae</i>	-	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-	x	X	
	Plastids <i>Arabidopsis thaliana</i>	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x	X	
a	Bacteria <i>Escherichia coli</i>	-	-	-	-	-	-	x	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	
b	Bacteria <i>B.subtilis</i>	x	x	x	x	x	-	x	-	-	-	-	-	-	-	-	-	-	x	-	-	-	-	-	
	Archaea <i>Aeropyrum pernix</i>	x	x	X	X	x	x	-	x	-	-	-	-	-	-	-	-	-	x	x	-	-	-	-	
c	Archaea <i>Methanococcus jannaschii</i>	x	x	x	x	x	x	-	x	-	-	-	-	-	-	-	-	-	x	x	-	-	-	-	
d	Ascomycota <i>Saccharomyces cerevisiae</i>	-	x	-	-	x	x	X	x	x	X	x	x	-	-	-	x	x	x	x	x	x	x	-	-
e	Ascomycota <i>Eremothecium gossypii</i>	-	x	-	-	x	x	X	x	x	-	x	-	-	-	-	x	x	x	x	x	x	x	-	-
f	Ascomycota <i>Coccidioides immitis</i>	-	x	-	-	x	x	x	x	-	x	-	-	-	x	-	-	x	x	x	x	x	x	-	-
g	Ascomycota <i>Schizosaccharomyces pombe</i>	-	x	-	-	x	x	x	x	-	-	-	-	-	-	-	x	x	x	x	x	x	x	-	-
	Basidiomycota <i>Phakospora pachyrhizi</i>	-	x	x	x	x	x	x	x	-	-	-	-	-	-	x	-	x	x	x	x	x	x	-	-
	Microsporidia <i>Encephalitozoon cuniculi</i>	-	?	?	?	x	x	x	x	-	-	-	-	-	-	?	-	?	x	x	?	?	-	-	
h	Metazoa <i>Homo sapiens</i>	-	x	x	x	x	x	x	x	-	-	-	-	-	-	x	-	x	x	x	x	x	x	-	-
	Mycetozoa <i>Dictyostelium discoideum</i>	-	x	x	x	x	x	x	x	-	-	-	-	-	-	x	-	x	x	x	x	x	x	-	-
	Entamoebidae <i>Entamoeba histolytica</i>	-	x	x	x	x	x	x	x	-	-	-	-	-	-	x	-	x	x	x	x	x	x	-	-
	Viridiplantae <i>Arabidopsis thaliana</i>	-	x	x	x	x	x	x	x	-	-	-	-	-	-	x	-	x	x	x	x	x	x	-	-
	Rhodophyta <i>C.merolae</i>	-	?	?	?	!	!	!	!	-	-	-	-	-	-	?	-	?	x	x	x	x	x	-	-
	Heterokonta <i>Phytophthora sojae</i>	-	x	x	x	x	x	x	x	-	-	-	-	-	-	?	-	?	x	x	?	?	-	-	
	Ciliophora <i>Tetrahymena thermophila</i>	-	x	x	s	x	x	x	x	-	-	-	-	-	-	x	-	x	x	x	x	x	x	-	-
i	Apicomplexa <i>Plasmodium falciparum</i>	-	x	X	X	x	x	x	x	-	-	-	-	-	-	x	-	x	x	x	x	x	x	-	-
j	Apicomplexa <i>Theileria annulata</i>	-	x	x	s	x	x	x	x	-	-	-	-	-	-	x	-	x	x	x	x	x	x	-	-
	Euglenozoa <i>Trypanosoma brucei</i>	-	x	x	x	x	x	x	x	-	-	-	-	-	-	x	-	-	x	x	x	x	x	-	-
	Parabasala <i>Trichomonas vaginalis</i>	-	x	x	x	x	x	x	x	-	-	-	-	-	-	!	-	x	x	x	?	?	-	-	
	Diplomonadida <i>Giardia lamblia</i>	-	?	?	?	x	x	x	x	-	-	-	-	-	-	?	-	?	x	x	x	?	-	-	

Table 2.2. SRP RNA features and SRP components ordered by phylogeny. The name of a representative species is given for each group. The column labeled '2D' indicates the secondary structures *a* to *n* shown in Figure 2.1. The RNA features (helices 1 to 12 and

the “extra” helix E) are shown in the left part of the table; proteins SRP9 to SRP72, as well as the chloroplast proteins cp54 and cp43 are indicated on the right portion. The table cells are annotated as '-': absent; '?': maybe absent or is not found; '!': expected to be present; 'x': present; 'X': feature is pronounced and may contain several helical sections; 's': this helix is comparatively small.

CHAPTER 3: COMPARATIVE 3-D MODELING OF tmRNA

Adapted from Burks, J., Zwieb, C., Muller, F., Wower, I. & Wower, J. (2005).

Comparative 3-D modeling of tmRNA. *BMC Mol Biol* 6, 14.

ABSTRACT

Trans-translation releases stalled ribosomes from truncated mRNAs and tags defective proteins for proteolytic degradation using transfer-messenger RNA (tmRNA). This small stable RNA represents a hybrid of tRNA- and mRNA-like regions connected by a variable number of pseudoknots. Comparative sequence analysis of tmRNAs found in bacteria, plastids, and mitochondria provides considerable insights into their secondary structures. Progress toward understanding the molecular mechanism of template switching, which constitutes an essential step in *trans*-translation, is hampered by our limited knowledge about the three-dimensional folding of tmRNA. To facilitate experimental testing of the molecular intricacies of *trans*-translation, which often require appropriately modified tmRNA derivatives, we developed a procedure for building three-dimensional models of tmRNA. Using comparative sequence analysis, phylogenetically-supported 2-D structures were obtained to serve as input for the program ERNA-3D. Motifs containing loops and turns were extracted from known structures of other RNAs and used to improve the tmRNA models. Biologically feasible 3-D models for the entire tmRNA molecule were obtained. The models were characterized by a functionally significant close proximity between the tRNA-like domain and the resume codon. Potential conformational changes which might lead to a more open structure of tmRNA upon binding to the ribosome are discussed. The method, described in detail for the tmRNAs of *Escherichia coli*, is applicable for every tmRNA. Improved, biologically significant molecular models were obtained. These models will guide experimental designs and provide a better understanding of *trans*-translation. The procedure described here for tmRNA is easily adapted for modeling members of other RNA families.

INTRODUCTION

Transfer-messenger RNA (tmRNA), also known as 10Sa RNA or ssrA RNA, is a hybrid of a tRNA-like domain (TLD) and a mRNA-like region (MLR) connected by a variable number of pseudoknots (Zwieb *et al.*, 1999b). TmRNA is a stable and essential component of *trans*-translation, a quality-control process that rescues ribosomes stalled on mRNAs lacking stop codons. During *trans*-translation, ribosomes switch from a defective mRNA (lacking its translation-termination signal) to the MLR of tmRNA. Because the stop codon is provided by the tmRNA, the ribosomes can dissociate and recycle (Withey & Friedman, 1999). As an additional advantage, the tandem translation of the two templates generates a tagged polypeptide which is degraded by housekeeping proteases (Keiler *et al.*, 1996; Tu *et al.*, 1995).

For tagging, tmRNA has to be aminoacylated by aminoacyl-tRNA synthetases (Himeno *et al.*, 1997). Assisted by protein SmpB, the charged tmRNA is delivered to stalled ribosomes as a ternary complex with EF-Tu and GTP. Binding of tmRNA to ribosomes is facilitated by ribosomal protein S1, which interacts with the MLR and pseudoknots but not with the TLD (Barends *et al.*, 2000; Karzai *et al.*, 2000; Karzai & Sauer, 2001; Wower *et al.*, 2000). Recently, cryo-EM revealed the shape of the tmRNA associated with SmpB and EF-Tu in its ribosome-bound form (Valle *et al.*, 2003). Despite this significant progress, high-resolution structures as obtained by NMR and X-Ray crystallography are unavailable and expected to be difficult to obtain in the foreseeable future due to the relatively large size and flexibility of the tmRNA.

In this chapter we used a stepwise procedure for arriving at high-resolution models for the entire tmRNA molecule. First, 2-D structures were obtained by

covariation analysis of a large number of tmRNA sequences. The basepairing information was submitted to the ERNA-3D modeling program (Mueller *et al.*, 1995) to build the helical sections. Structural motifs of the loops and turns were identified in SCOR (Klosterman *et al.*, 2002), high-resolution data were extracted from known structures, and incorporated into the models. Overall, significantly improved 3-D models were obtained which will help to understand the role of tmRNA in *trans*-translation. The described approach could be adopted to obtain high-resolution models of the members of other RNA families.

RESULTS

Identification of tmRNA sequences

The tmRNA sequences were identified previously and subjected to comparative sequence analysis (CSA) as described (Larsen & Zwieb, 1991; Zwieb *et al.*, 1999b). New tmRNA sequences were obtained from the tmRNA website (Williams, 2002), through keyword searches of the literature and GenBank (Benson *et al.*, 2005), or BLAST (Altschul *et al.*, 1990; Altschul *et al.*, 1997), and various genome sequencing projects. The new sequences were subjected to a process that was performed iteratively as described in Materials and Methods to confirm tmRNA identity, remove sequence duplications, and create a meaningful alignment.

New potential tmRNA sequences were maintained as a preliminary alignment in BioEdit (Hall, 1999), separate EMBL-formatted sequence files, and a HTML-formatted phylogenetic list. The sequences were ordered phylogenetically derived from the information in the Ribosomal Database Project (RDP) (Cole *et al.*, 2003). If the organism

name was not listed in the RDP, the sequence was placed next to its closest relative using the NCBI Taxonomy resource (Wheeler *et al.*, 2003).

Selection of tmRNA Sequences

The new sequences were confirmed individually as tmRNAs by comparison with the closest relative using the pairwise alignment feature of BioEdit (Hall, 1999). If there was a lack of obvious similarity, the sequence was inspected for evidence of biological features such as the ability to form a TLD and an open reading frame. Furthermore, the possibility of a two-part tmRNA was considered. A sequence suspected to be a new tmRNA was investigated further by CSA (Larsen & Zwieb, 1991) as described in Materials and Methods.

If the potential new sequence was encoded in two regions as seen in alpha-Proteobacteria and some Cyanobacteria (Keiler *et al.*, 2000), it was compared to the gene sequence of the two-part tmRNA from a closest relative. The sequence was arranged for effective comparison with the one-piece tmRNAs in the alignment. The location of the 3' end of the sequence was found upon comparison with the related sequence. The 5' domain of the tRNA-like part was identified using pairwise alignment procedures to generate a single sequence with a short intervening segment. Each of the 20 new two-part tmRNAs (14 sequences from alpha-Proteobacteria and six from Cyanobacteria) was subjected to this rearrangement.

Comparative Sequence Analysis

Sequences were ordered phylogenetically using the RDP (Cole *et al.*, 2003) as a guide or by pairwise alignment with the closest relative. Identical regions were aligned first and invariant positions were used as signposts. Subsequently, the more similar regions were aligned. Regions of biological significance, such as the resume and stop codons, were then considered. Finally, common secondary structure features were used to align regions that lacked primary structure similarity or biological features. Supported Watson-Crick basepairs and G-U interactions were indicated in the alignment by uppercase letters. Gaps were introduced to account for differences in sequence length and to avoid the alignment of dissimilar regions.

The existence of secondary structure was determined using covariation analysis as described (Larsen & Zwieb, 1991) (see also Materials and Methods). The alignment was examined to identify compensatory base changes (CBCs) and other covariations. The numbers of CBCs and mismatches between the alignment columns were counted. CBCs provided positive evidence for the existence of a basepair; mismatches provided negative evidence. If the number of compensatory base changes was two times or greater than the number of mismatches, the basepair was considered supported. If a basepair was invariant, no evidence for or against its existence could be gained from CSA. A basepair was considered specific to a particular phylogenetic group if it was proven only in that group.

Quality Control of Sequence Information

To check for the proper assignment of basepairs, the alignment was sent through an automated pipeline of programs from RNAdbTools (Gorodkin *et al.*, 2001). The output was inspected visually and corrections were made manually using the BioEdit program (Hall, 1999). The revised alignment was resubmitted to RNAdbTools, and the review process was repeated until a satisfactory alignment was produced.

TmRNA Alignment

The final alignment contained a total of 274 tmRNA sequences in 16 bacterial phylogenetic groups. A complete phylogenetic list is available at the tmRDB . There was a substantial increase in the number of two-part tmRNAs for a total of 27 two-part tmRNAs in alpha-Proteobacteria (20 tmRNAs), one mitochondrion (one tmRNA), and Cyanobacteria (six tmRNAs). The nine organelle sequences included one from a cyanelle, six from chloroplasts, one from a plastid, and one from the *Reclinomonas americana* mitochondrion. The typical tmRNA was about 350 nucleotides long. The *R. americana* mitochondrion tmRNA was considerably smaller (189 nucleotides), did not contain an ORF and thus may be non-functional. Excluding this exception and any partial tmRNAs, the tmRNA of *Synechococcus* species PCC7009 was the shortest (250 nucleotides), and the longest was from *Chlamydomophila psittaci* (425 nucleotides).

Secondary Structure of tmRNA

The tmRNA secondary structure features were extracted from the alignment and are listed in phylogenetic order in Table 3.1. The representative secondary structure of *Escherichia coli* tmRNA is shown in Figure 3.1.

TLD (Helices 1, 2a and 12)

Although a prominent feature of each tmRNA, the TLD was relatively weakly supported by CSA due to a high degree of sequence conservation. However, the structure of this region is well established by experimental evidence (Hou & Schimmel, 1988; Komine *et al.*, 1994; Ushida *et al.*, 1994).

Helix 1 contained seven basepairs and was usually continuous with the exception of the *Anabaena* species tmRNA, which contained an insertion in the 3'-portion of helix 1. The first pair (1G-C359 in *E. coli* tmRNA) was conserved with one exception in *Alcaligenes eutrophus* where there was a 1U-C345 mismatch possibly due to a sequencing error. The second (2G-C358, *E. coli* numbering, Figure 3.1) and third pair (3G-U357) of helix 1 were invariant and therefore neither supported nor disproven by CSA. The identities of the bases involved in the fourth (4G-C356) and fifth pair varied. The closing pair of helix 1 (7G-C353) was conserved with the exception of a 7U-A388 pair the *Trichodesmium erythraeum* tmRNA. The single-stranded region between helices 1 and 2a ranged from ten in *Dehalococcoides ethenogenes* to 13 nucleotides in one *Clostridium acetobutylicum* sequence.

Helix 2a was equivalent to the anticodon stem of tRNA and contained eight supported basepairs as well as a short variable internal loop in the 5' half of the helix that

occurred in a few sequences. The first position in the helix was a conserved cytosine (C21 in *E. coli*) which formed a weakly-supported basepair with the conserved G333. The partial tmRNA from the chloroplast of *Pavlova lutheri* contained a uracil in the first position, but no information regarding the 3' portion of helix 2a was available.

The T-loop and helix 12 were highly conserved, although many sequences lacked information about helix 12 due to primer annealing during PCR amplification. Helix 12 contained four strongly supported basepairs and a fifth conserved G-C pair (340G-C348 in *E. coli*; Figure 3.1). The *Dehalococcoides ethenogenes* tmRNA had the potential to form a sixth basepair in helix 12. Helix 12 was almost always continuous, except for the tmRNA of *Carboxydothemus hydrogenoformans* which possessed four basepairs and a mismatched U333 and C347. A 331-GG-332 preceded U333 in *C. hydrogenoformans* and followed the conserved 328-GAC-330. Therefore, U333 was unlikely to pair. In the T-loop, the U341 and U342 (*E. coli* tmRNA) seen in most sequences were replaced by two guanines in the tmRNA from the *R. americana* mitochondrion (G79 and G80 in chain B) (Keiler *et al.*, 2000). In the tmRNA from *Caulobacter crescentus*, the nucleotide corresponding to U342 in *E. coli* tmRNA was changed to G62 in chain B.

Helical Sections 2b, 2c and 2d

Overall, sections 2b, 2c, and 2d were well supported. Sections 2a and 2b were separated by a variable loop ranging from one to seven nucleotides in the 5' portion and from one to nine nucleotides in the 3' portion. Sections 2b and 2c had the potential to form a continuously stacked helix (e.g. in *Chlamydophila psittaci* tmRNA). Usually, a bulge of two to six nucleotides separated helical sections 2c and 2d (residues 309-311 in

E. coli tmRNA, Figure 3.1). An asymmetrical loop was present in some sequences (for example, residues 40-41 in chain A, and 27-31 in chain B of *Caulobacter crescentus* tmRNA, see additional file 2: Ccrescentus2D.pdf). Helix 2d was the most conserved of the three helical sections. A 6G-U308 basepair (*E. coli* numbering) in helix 2d was only weakly supported, conserved in most phylogenetic groups, but altered in the Thermatogales, Cyanobacteria, alpha-Proteobacteria, and Gram-positive bacteria. An U-A basepair was possible between these positions (U6 in chain A and 88 in chain B) in the *R. americana* mitochondrion tmRNA, as was a 46A-U334 pair in the *Synechocystis* species PCC6803 tmRNA.

Pseudoknot 1 (Helices 3 and 4)

Pseudoknot 1 (pk1) was well supported. Of the three connecting regions, the two 5'-regions were very short (no or only one residue) while the third was relatively long (one to 11 residues). All pseudoknots in tmRNA followed the same general design (Zwieb *et al.*, 1999a). Most sequences contained helices 3 and 4, with the exception of the tmRNA from *Oenococcus oeni* and the partial sequence from the chloroplast of *Pavlova lutheri*, both of which lacked helix 4 and thus did not form a pseudoknot. Helix 3 usually contained five basepairs. However, a sixth pair was possible in some bacteria. Helix 4 could be split into helical sections 4a and 4b by a bulge seen in 46 sequences (position 57 in *B. anthracis* tmRNA) or an internal loop seen in 52 tmRNA sequences. The adenine-rich terminal loop between the downstream halves of helices 3 and 4 ranged in length from two to 13 nucleotides.

The mRNA-Like Region (MLR)

The MLR consisted of an open reading frame (ORF) preceding helix 5 and varied from 48 (*Heliobacillus mobilis*) to 126 nucleotides (*Odontella sinensis* chloroplast). The resume codon usually coded for alanine, but for glycine in 30 sequences (e.g. *Bacillus anthracis*), aspartic acid in three sequences (e.g. *Staphylococcus epidermidis*), arginine in two uncultured species (FS1 and LEM2), serine in the uncultured species RCA1, and glutamic acid in *Mycoplasma pulmonis*.

Helix 5 was the least supported helix. One to three stop codons were located within the helix 5 loop. A single UAA stop codon was present in 157 sequences. UAG (17 sequences) or UGA (10 sequences) were used less frequently. In 85 sequences there were two in-frame stop codons, where UAA was always the first codon, followed by another UAA (73 sequences), UAG (10 sequences) or UGA (2 sequences). Two sequences (*Bacillus megaterium* and *Chloroflexus aurantiacu*) were found to contain three tandem in-frame stop codons.

Pseudoknot 2 (Helices 6 and 7)

Pseudoknot 2 was well supported and similar in overall design to pk1. Helical sections 6b and 6c showed a potential to form a continuous helix in *Thermotoga maritima*. In beta-Proteobacteria, 6b was replaced by a short hairpin 6d (Zwieb *et al.*, 1999b). Helix 6d was observed also in three tmRNAs of the gamma-Proteobacteria *Acidithiobacillus ferrooxidans* and *Francisella tularensis*.

Pseudoknot 3 (Helices 8 and 9)

Pseudoknot 3 was well supported but missing in Cyanobacteria and the organelles (Table 3.1). Helical sections 8a and 8b were likely to be continuously stacked because a single helix was present in some species such as *Aquifex aeolicus*. The unusual purine-rich internal loop between helical sections 8a and 8b was present in most gamma-Proteobacteria suggesting a special function.

Pseudoknot 4 (Helices 10 and 11)

This feature was well supported and was similar in design to the other tmRNA pseudoknots. Helical sections 10a and 10b had the potential to stack because a single helix was present in *Prevotella intermedia*. In some Cyanobacteria sequences, however, pk4 was replaced with two smaller tandem pseudoknots.

Secondary Structure Prediction of the MLR

Because CSA was unable to determine secondary structure for a large portion of the MLR, energy calculations were carried out aimed to predict structure for the single-stranded portion of the open reading frame. The region corresponding to residues 79-107 of *E. coli* tmRNA (Figure 3.1) was extracted from the alignment. A representative alignment of 197 sequences was submitted to Mfold (Zuker, 2003). Each sequence had the potential to form at least one helix, designated “m” (see Figure 3.1). Two or more adjacent helices were predicted for 17 sequences. The number of basepairs varied from two in *Chloroflexus aurantiacus* to ten in *Mycoplasma pulmonis*.

Secondary Structures of *E. coli* tmRNA

Secondary structures were determined for all sequences in the alignment but only the sequence of *E. coli* tmRNA was extracted, diagrammed, and processed for 3-D modeling. The 363-nucleotide tmRNA of the gamma-Proteobacterium *Escherichia coli* represented the typical tmRNA containing the TLD, the MLR, and four pseudoknots (pk1 to pk4) encompassing the pseudoknot domain (PKD). The 90-GCA-92 resume triplet coded for alanine. Two in-frame UAA stop codons (positions 120-125) were located within the terminal loop of helix 5 (Figure 3.1). Three basepaired regions (shown boxed) were only weakly supported by CSA. Helix m (residues 87-98) was predicted only by energy calculations. A slightly different helix involving residues 88-100 was suggested by footprinting of *E. coli* tmRNA (Felden *et al.*, 1997). The evidence for the 112U-A133 basepair was weak, but was included due to the possibility of extending helix 5 (Materials and Methods). Helical section 5a (residues 108-111 and 134-137) was enlarged by the weakly supported 108G-C137, 110U-A135 and 111U-G134. The 109C-G136 pair was disproven. In helix 10ab, the basepair between 256G-C275 was only weakly supported. Helix 10ab (residues 248-256 and 274-283) could be extended by the boxed 257U-G274 pair.

Tertiary Structure Modeling and Visualization of tmRNA

ERNA-3D, a program developed to model RNA in three dimensions (Mueller *et al.*, 1995), was used on an SGI workstation as described in Materials and Methods. *E. coli* tmRNA was selected because this tmRNA is the subject of extensive research. *B. anthracis* tmRNA was chosen as an example of a tmRNA from a Gram-positive

bacterium, and *C. crescentus* tmRNA was selected because it represents a two-part tmRNA.

In order to create the initial models, the sequence and basepairing information were entered into an ERNA-3D input file to automatically generate A-form RNA for the helices sections and specify the single-stranded regions using ERNA-3D's algorithm (Mueller *et al.*, 1995). Since ERNA-3D avoided an XYZ coordinate system as reference for the user, the manipulation of the model from the viewer's perspective was simple and intuitive. The coordinates of each model were saved in PDB format (Sussman *et al.*, 1999) for compatibility with other molecular modeling programs. Motifs (listed in Table 3.2) were selected to model the loops and turns of a particular tmRNA. ERNA-3D selection files were generated to define clusters and place the motif in 3-D without disturbance to the rest of the model. The 3-D cursor box was used to manipulate a cluster in three-dimensional space, similar to the manipulation of a section of a physical model.

Numerous high-resolution structures determined by NMR or X-ray crystallography represented a rich source of detailed information for defining biologically meaningful motifs. The SCOR database (Klosterman *et al.*, 2002) provided a way to find suitable templates. In rare cases when a SCOR search for a motif did not result in a acceptable match (e.g. motif 9, Table 3.2), the nucleotides were positioned manually in ERNA-3D. Otherwise, the coordinates were obtained from the Protein Data Bank PDB (Berman *et al.*, 2000), extracted using the program Swiss-PDBViewer (Guex & Peitsch, 1997), and imported into ERNA-3D. The source motif and the region to be modeled were selected as separate clusters and aligned in three dimensions using common features (usually a shared basepair). Once superimposed, the coordinates of the residues in the

source motif were copied onto the corresponding residues in the model. The template was then deleted, leaving a biologically meaningful structure. The backbone connections between the motif and the rest of the model were inspected visually and, if needed, manual adjustments were made to correct bond lengths and angles.

As an example of the motif modeling process, the purine-rich loop in *E. coli* pk3 (positions 204-206 and 223-225) was constructed using a similar loop in the 30S ribosomal subunit. First, the purine-rich loop was defined as motif 11a (Table 3.2), and used to search the SCOR database. Positions 780-782 and 800-802 in the structure of the *Thermus thermophilus* 30S ribosomal subunit (Wimberly *et al.*, 2000) were found to conform to the motif. The 30S ribosomal subunit coordinates (1J5E.pdb in this case) were downloaded from the PDB and displayed using Swiss-PDBViewer. The coordinates of the loop and the closing basepairs were extracted and inspected to confirm that the structure was compatible. The clustered regions were aligned with the ends of helical sections 8a and 8b at the basepairs 203U-G226 and 207A-U222 of the *E. coli* model and 779C-G803 and 783C-G799 of the template. Template positions 780-AAA-782 and 800-GUA-802 were then copied onto 204-GGA-206 and 223-GAA-225 of the model. The template was deleted and the bond lengths and angles involving the atoms of the phosphates of residues U203, G222, A206, and U222 were adjusted.

In some instances, the tmRNA sequence alignment was reinvestigated using ideas derived from the 3-D model. For example, the alignments of pk1 in *Bacillus anthracis* tmRNA and relatives was changed from a two nucleotide bulge (56-AU-57) between helical sections 4a and 4b to a more feasible and equally well supported one-nucleotide bulge (not shown). The alignment of helix 10 in pk3 in *B. anthracis* tmRNA and relatives

was altered from a 237C-A269 mismatch and an asymmetrical loop (C239 and 266-GU-267) to a single looped-out C269. The alignment of pk3 of *Caulobacter crescentus* and relatives was changed from four basepairs and a weakly supported fifth pair in helix 8 (between 174G-C196 of chain A) to a four-basepair structure (not shown).

Finally, information about spatial neighborhoods as obtained from cross-linking experiments was introduced. Interactions between the D- and T-loops were incorporated from a previous model (Zwieb *et al.*, 2001) that was based on cross-links observed in the *E. coli* TLD (motif 2 in Table 3.2). A cross-link observed between positions the stop codon loop and C154/C155 in pk2 of *E. coli* tmRNA (Wower *et al.*, unpublished) was considered along with the previously-discovered covariation (Kelley *et al.*, 2001) between C44 and C66 (*E. coli* numbering, Figure 3.1). Finally, the models were inspected for correct bond angles and distances.

3-D Model of *E. coli* tmRNA

The model shown as a ribbon diagram in Figure 3.3 consists of a compacted MLR and PKD with the TLD extending from the body of the molecule due to the near-coaxial stacking of the helix 2 sections. The coordinates for the TLD were taken from a previous model (Zwieb *et al.*, 2001) which is based on two cross-linked sites, one formed between nucleotides U9/U10 near the 5' end and nucleotides C346/U347 in the T loop, the other involving residues at positions 25–28 and 326–329 within helix 2a (motif 2 in Table 3.2). Important features of the TLD include the non-Watson-Crick base pairs formed by 19-GA-20 and 333-GA-334 which have been confirmed by site-directed mutagenesis (Hanawa-Suetsugu *et al.*, 2001).

A very efficient UV-induced cross-link observed between the stop codon loop of helix 5 and pk2 of *E. coli* tmRNA (Wower *et al.*, unpublished) introduced a considerable constraint of helices 5, 6, and 7, and, as has been shown recently, is consistent with the cryo-EM structure of ribosome-bound tmRNA of the initial stage of *trans*-translation (Valle *et al.*, 2003). Also considered was the previously-discovered covariation (Kelley *et al.*, 2001) between C44 and C66 (*E. coli* numbering, Figure 3.1) which was thought to determine the orientation of helix 2 in relation to helix 3 and thus the approximate angle by which the TLD protrudes. The 44/66 covariation is strongly supported (26 covariations *versus* four mismatches) in an alignment of 143 representative sequences (not shown). Since this is a non-Watson-Crick covariation, it is difficult to propose a precise structure in this region. More extensive studies will be required to better understand the nature and structural significance (if any) of the 44/66 covariation.

The distance between the 3' end and A231 in pk3 was 180 Å, and 70 Å between the outside edges of pk1 and pk4. Helix 5 and pk2 were positioned in a parallel fashion due to a cross-link observed between the stop-codon loop and pk2. The nucleotides in the bulge between helical sections 6a and 6b (motif 9, Table 3.2) were adjusted manually to allow for a close fit of helix 5 and pk2. The four pseudoknots were arranged in a loop with the resume codon positioned near the internal loop between helices 2a and 2b (motif 3a, Table 3.2).

Considering that the pseudoknots are likely to constitute relatively independent structural units, conformational changes might occur around the connecting single strands, as well as in the MLR and the weakly-supported helix m. TmRNA may become less flexible when bound to proteins such as SmpB and ribosomal protein S1. EF-Tu,

however, likely binds to the coaxially-stacked helices 1 and 12 (Nissen *et al.*, 1995), and therefore appears to have little effect on the conformation of the TLD. Protein SmpB was found to bind near helix 2a (Barends *et al.* 2001; Wower *et al.*, 2002), has two RNA binding sites (Dong *et al.*, 2002), and thus could make additional contacts with other regions. Protein S1 is the largest ribosomal protein, has been shown to be close to numerous sites, and to be required for the binding of tmRNA to the ribosome (Wower *et al.*, 2001). Since S1 is a flexible, beadlike protein (Subramanian *et al.*, 1983) it may not restrict the conformational potential of the tmRNA molecule. Instead, the protein may instill some constraint to the large central loop formed by the PKD and the MLR. Because S1 is known to melt helices in mRNAs (Bear *et al.*, 1976), it is also possible that it unwinds helix m and exposes the resume codon and the preceding nucleotides U85 and A86 for efficient *trans*-translation (Ivanov *et al.*, 2002; Williams *et al.* 1999).

The tmRNA model shows the resume codon in close proximity to the internal loop formed between helical sections 2a and 2b. This arrangement would allow the ribosome to "jump" a relatively short distance from the end of the broken mRNA onto the ORF of tmRNA. In a recent cryo-EM study of the initial stage of *trans*-translation (Valle *et al.*, 2003), the tRNA-like region, SmpB, EF-Tu, and part of pk4 were located in the A site of the ribosome. We suggest that this more open arrangement is made possible due to the flexibility of tmRNA, the melting of helix m, and/or a change in conformation induced by the binding of tmRNA to the ribosome (Figure 3.4). The opening of the central loop seems to be accompanied by a rotation of the TLD around the helix 2 axis (compare Figures 3.4A and 3.4B) and thus might properly align the resume codon with the 3'-end of broken mRNA in the ribosomal decoding center. At the later stages of the

transit of tmRNA across the ribosome even more dramatic conformational changes were shown to disrupt helix 5 and the pseudoknotted regions (Wower et al, 2005). These downstream alterations are likely mediated not by protein S1 but by the intrinsic helicase activity of the ribosome (Takyar *et al.*, 2005) and are required to maintain the ribosomal subunits in close proximity to the unfolded tmRNA in order to monitor *trans*-translation.

DISCUSSION

We have compared a growing number of tmRNA sequences from all groups of bacteria to produce an alignment from which the secondary structure of any tmRNA could be easily extracted. Most basepairings were supported by phylogenetic evidence, whereas only a few helical sections required energy calculations. Uncertainties in assignment of basepairs, such as the pseudoknot region of chloroplasts and one-piece cyanobacterial tmRNAs, may be eliminated in the future when more sequences will become available.

The common layout of the secondary structures indicated a similar function in all bacteria. The number and size of the pseudoknots varied, supporting the idea that the pseudoknots may only enhance the essential functions carried by the TLD and the MLR (Wower *et al.*, 2004). Differences in the secondary structure features were usually not random but occurred between groups of related organisms. For example, helix 6d was present only in the beta- and three close relatives of the gamma-Proteobacteria. Whether these group-specific features are responsible for differences in the *trans*-translation mechanism remains to be determined. However, strategies that exploit these differences,

for example for developing new antibiotics targeted at a specific group of bacteria, can now be envisioned.

In principle, tertiary structure models of any tmRNA in the alignment could be built using the described procedures. Here, we have shown how to generate an updated model of *E. coli* tmRNA (Zwieb *et al.*, 1999a). The TLD mimicked the L-shape of canonical tRNA (Sussman *et al.*, 1978) and may be necessary for proper association of tmRNA with the EF-Tu, SmpB, and subsequent binding to the ribosomal A site. The lack of a D-stem was suggested to confer flexibility (Stagg *et al.*, 2001), but SmpB may be responsible for stabilizing this region (Barends *et al.*, 2001; Gutmann *et al.*, 2003).

Differences in the shapes of the tmRNA models (e.g. the angle between helix 2 and the main body of the molecule) may be an indicator of a significant level of flexibility. Conformational changes might occur in regions for which no helical structure could be predicted either by CSA or energy calculations. Such regions might include the MLR, the single strands connecting the pseudoknots, and the weakly-supported helices m and 5 (Figures 3.3 & 3.4). Based on a long-range cross-link, large scale structural changes likely involve the stop codon loop and pk2 since these regions do not basepair but are in close proximity (Wower *et al.*, unpublished).

TmRNA may become less flexible when bound to proteins such as SmpB and ribosomal protein S1. EF-Tu, however, which likely binds to the coaxially-stacked helices 1 and 12 (Nissen *et al.*, 1995), appears to have little effect on the RNA conformation. Protein SmpB was found to bind near helix 2a (Barends *et al.*, 2001; Wower *et al.*, 2002), has two RNA binding sites (Dong *et al.*, 2002), and thus could make additional contacts with other regions. Protein S1 is the largest ribosomal protein, was

shown to be close to numerous tmRNA sites, and is required for tmRNA to bind to the ribosome (Wower *et al.*, 2000). S1 is a flexible, beadlike protein (Subramanian, 1983) and thus may not restrict the conformational potential of the tmRNA molecule. Instead, the protein may instil some constraint to the large loop formed by the PKD and the MLR. Because the protein is known to melt helices in mRNAs (Bear *et al.*, 1976), it is possible that it unwinds helix m and exposes the resume codon and the preceding nucleotides U85 and A86 for efficient *trans*-translation (Ivanov *et al.*, 2002; Williams *et al.*, 1999).

The tmRNA models show the resume codon in close proximity to the loop formed between helical sections 2a and 2b. This arrangement would allow the ribosome to “jump” a relatively short distance from the end of the broken mRNA onto the ORF of tmRNA. In a recent cryo-EM study of the initial stage of *trans*-translation (Valle *et al.*, 2003), the tRNA-like region, SmpB, EF-Tu, and part of pk4 were located in the A site of the ribosome. Overall, the tmRNA was arranged similar to a previous model (Zwieb *et al.*, 1999a) with the pseudoknots around the “beak” of the 30S subunit. We suggest that this more open arrangement is made possible due to the flexibility tmRNA, the melting of helices m and 5, and/or a change in conformation induced by the binding of tmRNA to the ribosome (Figure 3.4). At the later stages of the transit of tmRNA across the ribosome even more extensive conformational changes might involve the disruption of the pseudoknotted regions (Wower *et al.*, 2005). These alterations are likely required to maintain the ribosomal subunits in close proximity to the tmRNA in order to monitor *trans*-translation.

ORIGINAL CONCLUSIONS

This study significantly advances our understanding of *trans*-translation by providing biologically feasible 3-D models for the entire tmRNA molecule. Although the modeling of one tmRNA was described here, 3-D models of every tmRNA can be extracted from the alignment. The models are characterized by a functionally meaningful close proximity between the TLD and the resume codon. Conformational changes induced by binding of tmRNA to SmpB, ribosomal protein S1, and the ribosome suggest a transformation of a free, compact tmRNA to a more open, ribosome-bound structure. The comparative modeling approach described here for tmRNA is easily adapted for other RNA classes.

FUTURE DIRECTIONS AND EVOLUTION OF THE *E. COLI* tmRNA MODEL

Two crystal structures of the TLD in complex with SmpB were published by Gutmann *et al.* (2003) and Bessho *et al.* (2007). To investigate the potential for our modeling procedure to be ported to studying the structure of RNA-protein complexes, we decided to take advantage of information for the entire TLD in the Bessho group crystal structure of the TLD-SmpB complex (PDB ID 2CZJ). The coordinates of residues 1-13, 14-24, 31-35, and 46-72 in the *T. thermophilus* TLD were copied onto residues 1-13, 15-25, 328-332 and 333-359 in the *E. coli* model. An inserted G14 was manually positioned between the coordinates of residues 13 and 15 facing away from SmpB. The structural coordinates of the TLD in the 2CZJ coordinates of the TLD-SmpB complex were then deleted, leaving the SmpB coordinates behind in a biologically relevant conformation and configuration bound to the TLD of *E. coli* tmRNA. During this process, the loop between

sections 2a and 2b was updated using the coordinates of the internal loop with an interrupted stack formed by residues 2436-2349 and 2469-2473 in PDB structure 1J5A, selected due to the correct size (five nucleotides in 5' half, four in 3' half) and twist of helix 2 closer to that seen in cryo-EM studies. The coordinates of residues 2469-2473 of 1J5A were copied onto the coordinates of *E. coli* tmRNA model residues 29-33, and 2436-2439 (1J5A) were used to model residues 321-324 of *E. coli* tmRNA. Source coordinates were deleted, leaving the model shown in Figure 3.4E.

While this model may not answer the question of the location of the resume codon in the tmRNA on the ribosome, the shape of the TLD-SmpB complex suggests that the area of the cryo-EM difference map originally assigned to a subsection of helix 2 by Valle and colleagues may actually be the location of the SmpB molecule, with the rest of the bulk of the TLD difference map encompassing the TLD itself and EF-Tu.

The model is constantly improving with each new piece of the *trans*-translation puzzle that comes to light. The model can be and is improved by advancements in X-ray crystallography, NMR and cryo-EM techniques.

METHODS

Comparative Sequence Analysis

The tmRNA sequences were arranged in phylogenetic order using information available in the RDP (Cole *et al.*, 2003). When the phylogenetic order could not be determined, the sequence was placed next to the closest relative as determined by the ClustalW plug-in of BioEdit (Hall, 1999; Thompson *et al.*, 1994). The sequences were made available at the tmRDB (Zwieb *et al.*, 2003).

Aligning was done manually using BioEdit (Hall, 1999) with details described previously (Larsen & Zwieb, 1991). Briefly, closely related sequences were aligned first. Then, invariant positions were used as guides to align the dissimilar regions. Next, common secondary structure elements were identified by observing covariations and find support for basepairs, tertiary interactions, or other structural features. Compensatory base changes (CBCs) were observed if a change in one residue of a Watson-Crick or G-U pair was compensated by a second change to conserve pairing. Two residues were mismatched if they did not form a Watson-Crick or G-U pair. CBCs and mismatches were counted to determine positive and negative evidence in order to prove or disprove the existence of a particular pair. A basepair was considered proven if there was at least twice as much positive than negative evidence. Invariant pairs provided neither positive nor negative evidence. If a basepair was proven in one phylogenetic group and disproven in another, the basepair was considered to be specific to that group.

The alignment and suggested CBCs were checked using RNAdbTools (Gorodkin *et al.*, 2001) to eliminate incorrectly-paired nucleotides, suggest extensions of helices, and determine the phylogenetic support for each basepair. Weakly supported basepairs adjacent to supported basepairs were considered an extension of the helix and usually included in the secondary structures (Figure 3.1).

3-D Model Building

The secondary structure information was used as input for ERNA-3D (Mueller *et al.*, 1995) installed on an SGI workstation running IRIX 6.5. ERNA-3D generated A-form RNA for each helix and calculated the conformations of single-stranded regions.

The models were examined using CrystalEyes stereovision goggles and an StereoGraphics infrared emitter. Structural motifs were identified using SCOR (Klosterman *et al.*, 2002) and the coordinates were obtained from the Protein Data Bank (PDB) (Berman *et al.*, 2000), extracted using Swiss-PDBViewer (Guex & Peitsch, 1997) and superimposed onto the model. Data obtained from site-directed mutagenesis, cross-linking experiments, or the literature were incorporated, and bond lengths and angles were adjusted manually to produce biologically feasible models. The final models were saved in PDB format and viewed in iMol to create the ribbon diagrams shown in Figures 3.3 and 3.4.

REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* 215, 403-410.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25, 3389-3402.
- Barends, S., Karzai, A. W., Sauer, R. T., Wower, J. & Kraal, B. (2001). Simultaneous and functional binding of SmpB and EF-Tu-TP to the alanyl acceptor arm of tmRNA. *J Mol Biol* 314, 9-21.
- Barends, S., Wower, J. & Kraal, B. (2000). Kinetic parameters for tmRNA binding to alanyl-tRNA synthetase and elongation factor Tu from Escherichia coli. *Biochemistry* 39, 2652-2658.
- Bear, D. G., Ng, R., Van Derveer, D., Johnson, N. P., Thomas, G., Schleich, T. & Noller, H. F. (1976). Alteration of polynucleotide secondary structure by ribosomal protein S1. *Proc Natl Acad Sci U S A* 73, 1824-1828.
- Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Wheeler, D. L. (2005). GenBank. *Nucleic Acids Res* 33 Database Issue, D34-38.
- Bessho, Y., Shibata, R., Sekine, S., Murayama, K., Higashijima, K., Hori-Takemoto, C., Shirouzu, M., Kuramitsu, S., & Yokoyama, S. (2007). Structural basis for functional mimicry of long-variable-arm tRNA by transfer-messenger RNA. *Proc Natl Acad Sci U S A* 104, 8293-8298.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Res* 28, 235-242.
- Cole, J. R., Chai, B., Marsh, T. L., Farris, R. J., Wang, Q., Kulam, S. A., Chandra, S., McGarrell, D. M., Schmidt, T. M., Garrity, G. M. & Tiedje, J. M. (2003). The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic Acids Res* 31, 442-443.
- Dong, G., Nowakowski, J. & Hoffman, D. W. (2002). Structure of small protein B: the protein component of the tmRNA-SmpB system for ribosome rescue. *Embo J* 21, 1845-1854.
- Felden, B., Himeno, H., Muto, A., McCutcheon, J. P., Atkins, J. F. & Gesteland, R. F. (1997). Probing the structure of the Escherichia coli 10Sa RNA (tmRNA). *Rna* 3, 89-103.

- Gorodkin, J., Zwieb, C. & Knudsen, B. (2001). Semi-automated update and cleanup of structural RNA alignment databases. *Bioinformatics* 17, 642-645.
- Guex, N. & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18, 2714-2723.
- Gutmann, S., Haebel, P. W., Metzinger, L., Sutter, M., Felden, B. & Ban, N. (2003). Crystal structure of the transfer-RNA domain of transfer-messenger RNA in complex with SmpB. *Nature* 424, 699-703.
- Hall, T. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 41, 95-98.
- Hanawa-Suetsugu, K., Bordeau, V., Himeno, H., Muto, A. & Felden, B. (2001) Importance of the conserved nucleotides around the tRNA-like structure of Escherichia coli transfer-messenger RNA for protein tagging. *Nucleic Acids Res* 29, 4663–4673.
- Himeno, H., Sato, M., Tadaki, T., Fukushima, M., Ushida, C. & Muto, A. (1997). In vitro trans translation mediated by alanine-charged 10Sa RNA. *J Mol Biol* 268, 803-808.
- Hou, Y. M. & Schimmel, P. (1988). A simple structural feature is a major determinant of the identity of a transfer RNA. *Nature* 333, 140-145.
- Ivanov, P. V., Zvereva, M. I., Shpanchenko, O. V., Dontsova, O. A., Bogdanov, A. A., Aglyamova, G. V., Lim, V. I., Teraoka, Y. & Nierhaus, K. H. (2002). How does tmRNA move through the ribosome? *FEBS Lett* 514, 55-59.
- Karzai, A. W., Roche, E. D. & Sauer, R. T. (2000). The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue. *Nat Struct Biol* 7, 449-455.
- Karzai, A. W. & Sauer, R. T. (2001). Protein factors associated with the SsrA.SmpB tagging and ribosome rescue complex. *Proc Natl Acad Sci U S A* 98, 3040-3044.
- Keiler, K. C., Shapiro, L. & Williams, K. P. (2000). tmRNAs that encode proteolysis-inducing tags are found in all known bacterial genomes: A two-piece tmRNA functions in Caulobacter. *Proc Natl Acad Sci U S A* 97, 7778-7783.
- Keiler, K. C., Waller, P. R. & Sauer, R. T. (1996). Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* 271, 990-993.
- Kelley, S. T., Harris, J. K. & Pace, N. R. (2001). Evaluation and refinement of tmRNA structure using gene sequences from natural microbial communities. *Rna* 7, 1310-1316.

- Klosterman, P. S., Tamura, M., Holbrook, S. R. & Brenner, S. E. (2002). SCOR: a Structural Classification of RNA database. *Nucleic Acids Res* 30, 392-394.
- Komine, Y., Kitabatake, M., Yokogawa, T., Nishikawa, K. & Inokuchi, H. (1994). A tRNA-like structure is present in 10Sa RNA, a small stable RNA from *Escherichia coli*. *Proc Natl Acad Sci U S A* 91, 9223-9227.
- Larsen, N. & Zwieb, C. (1991). SRP-RNA sequence alignment and secondary structure. *Nucleic Acids Res* 19, 209-215.
- Mueller, F., Doring, T., Erdemir, T., Greuer, B., Junke, N., Osswald, M., Rinke-Appel, J., Stade, K., Thamm, S. & Brimacombe, R. (1995). Getting closer to an understanding of the three-dimensional structure of ribosomal RNA. *Biochem Cell Biol* 73, 767-773.
- Nissen, P., Kjeldgaard, M., Thirup, S., Polekhina, G., Reshetnikova, L., Clark, B. F. & Nyborg, J. (1995). Crystal structure of the ternary complex of Phe-tRNAPhe, EF-Tu, and a GTP analog. *Science* 270, 1464-1472.
- Stagg, S. M., Frazer-Abel, A. A., Hagerman, P. J. & Harvey, S. C. (2001). Structural studies of the tRNA domain of tmRNA. *J Mol Biol* 309, 727-735.
- Subramanian, A. R. (1983). Structure and functions of ribosomal protein S1. *Prog Nucleic Acid Res Mol Biol* 28, 101-142.
- Sussman, J. L., Abola, E. E., Lin, D., Jiang, J., Manning, N. O. & Prilusky, J. (1999). The protein data bank. Bridging the gap between the sequence and 3D structure world. *Genetica* 106, 149-158.
- Sussman, J. L., Holbrook, S. R., Warrant, R. W., Church, G. M. & Kim, S. H. (1978). Crystal structure of yeast phenylalanine transfer RNA. I. Crystallographic refinement. *J Mol Biol* 123, 607-630.
- Takyar, S., Hickerson, R.P. & Noller, H.F. (2005) mRNA helicase activity of the ribosome. *Cell* 120, 49-58.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22, 4673-4680.
- Tu, G. F., Reid, G. E., Zhang, J. G., Moritz, R. L. & Simpson, R. J. (1995). C-terminal extension of truncated recombinant proteins in *Escherichia coli* with a 10Sa RNA decapeptide. *J Biol Chem* 270, 9322-9326.
- Ushida, C., Himeno, H., Watanabe, T. & Muto, A. (1994). tRNA-like structures in 10Sa RNAs of *Mycoplasma capricolum* and *Bacillus subtilis*. *Nucleic Acids Res* 22, 3392-3396.

- Valle, M., Gillet, R., Kaur, S., Henne, A., Ramakrishnan, V. & Frank, J. (2003). Visualizing tmRNA entry into a stalled ribosome. *Science* 300, 127-130.
- Wheeler, D. L., Church, D. M., Federhen, S., Lash, A. E., Madden, T. L., Pontius, J. U., Schuler, G. D., Schriml, L. M., Sequeira, E., Tatusova, T. A. & Wagner, L. (2003). Database resources of the National Center for Biotechnology. *Nucleic Acids Res* 31, 28-33.
- Williams, K. P. (2002). The tmRNA Website: invasion by an intron. *Nucleic Acids Res* 30, 179-182.
- Williams, K. P., Martindale, K. A. & Bartel, D. P. (1999). Resuming translation on tmRNA: a unique mode of determining a reading frame. *Embo J* 18, 5423-5433.
- Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Jr., Morgan-Warren, R. J., Carter, A. P., Vornrhein, C., Hartsch, T. & Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. *Nature* 407, 327-339.
- Withey, J. & Friedman, D. (1999). Analysis of the role of *trans*-translation in the requirement of tmRNA for lambdaimmP22 growth in Escherichia coli. *J Bacteriol* 181, 2148-2157.
- Wower, I.K., Zwieb, C. & Wower, J. (2005). Transfer-messenger RNA unfolds as it transits the ribosome. *RNA* 11, 668-673.
- Wower, I. K., Zwieb, C. & Wower, J. (2004). Contributions of pseudoknots and protein SmpB to the structure and function of tmRNA in *trans*-translation. *J Biol Chem* 279, 54202-54209.
- Wower, I. K., Zwieb, C. W., Guven, S. A. & Wower, J. (2000). Binding and cross-linking of tmRNA to ribosomal protein S1, on and off the Escherichia coli ribosome. *Embo J* 19, 6612-6621.
- Wower, J., Zwieb, C. W., Hoffman, D. W. & Wower, I. K. (2002). SmpB: a protein that binds to double-stranded segments in tmRNA and tRNA. *Biochemistry* 41, 8826-8836.
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31, 3406-3415.
- Zwieb, C., Gorodkin, J., Knudsen, B., Burks, J. & Wower, J. (2003). tmRDB (tmRNA database). *Nucleic Acids Res* 31, 446-447.
- Zwieb, C., Guven, S. A., Wower, I. K. & Wower, J. (2001). Three-dimensional folding of the tRNA-like domain of Escherichia coli tmRNA. *Biochemistry* 40, 9587-9595.
- Zwieb, C., Mueller, F. & Wower, J. (1999a). Comparative three-dimensional modeling of tmRNA. *Nucl Acids Symp Ser* 41, 200-204.

Zwieb, C., Wower, I. & Wower, J. (1999b). Comparative sequence analysis of tmRNA.
Nucleic Acids Res 27, 2063-2071.

tmRNA (*Escherichia coli*)

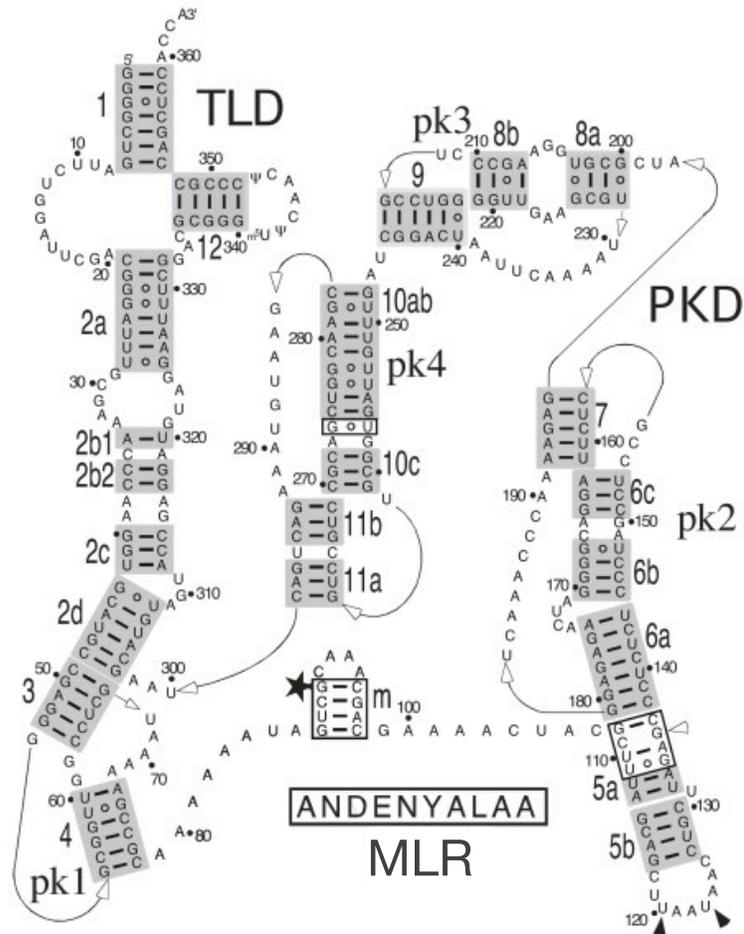


Figure 3.1. Secondary structure of *E. coli* tmRNA. Phylogenetically-supported helices are highlighted in gray and numbered from 1 to 12. The 5' and 3' ends are indicated. Arrows represent connections from 5' to 3'. Residues are numbered in increments of ten. Weakly supported regions and basepairs are shown in boxes. The disproven potential pairing of C109 with G136 is labeled with an open arrowhead. The star labels the first nucleotide of the resume codon. The tag peptide sequence is shown below the mRNA-like region. The stop codons are indicated with solid arrowheads. Three domains are distinguished: the tRNA-like domain (TLD), the mRNA-like region (MLR), and the pseudoknot domain (PKD).

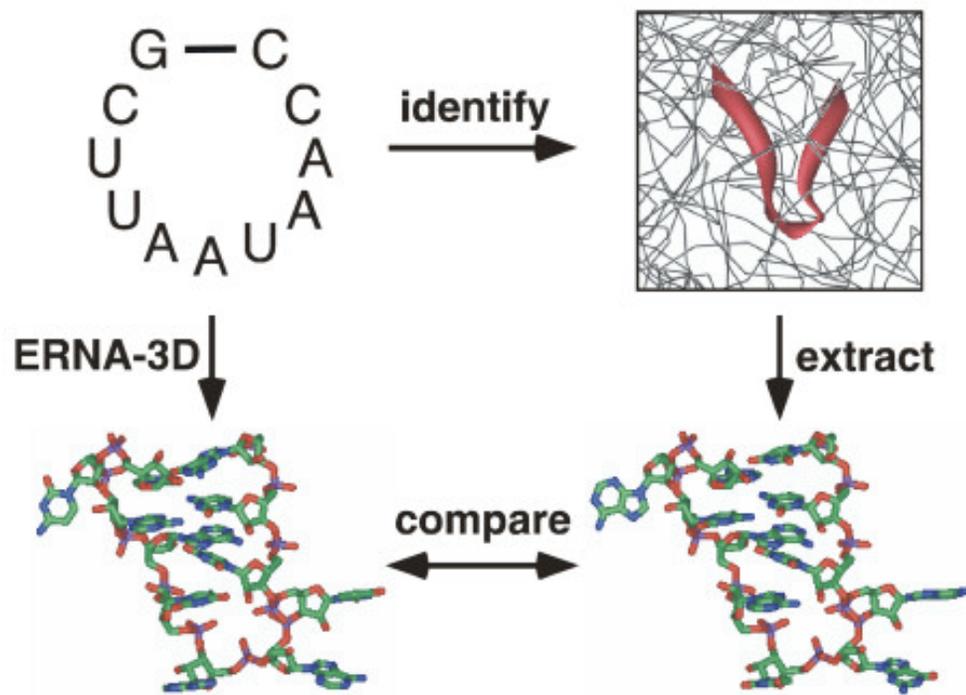


Figure 3.2. Motif modeling procedure. Motifs, for example the nonamer-loop shown in the top-left panel, were identified in the known high-resolution structures (top-right) with the help of SCOR (Klosterman *et al.*, 2002). The PDB coordinates were extracted (bottom-right) and compared with the 3-D model generated by ERNA-3D (bottom-left) to deduce relevant models.

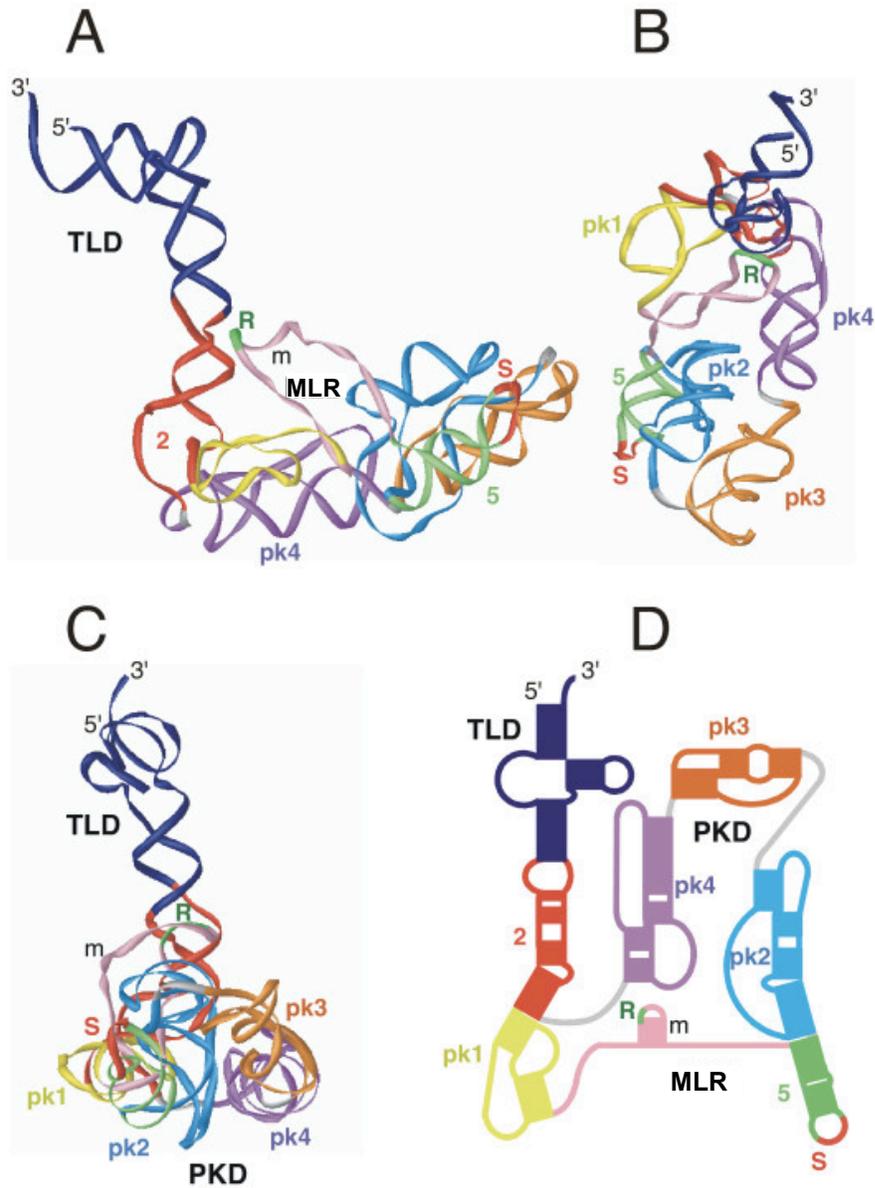


Figure 3.3. 3-D model of *Escherichia coli* tmRNA. The 3-D model of *Escherichia coli* tmRNA is viewed as a ribbon diagram from the side in panel **A**, the top in panel **B**, and in panel **C** turned by approximately 90° around the y-axis in relation to **A**. Panel **D** shows a representation of the corresponding 2-D structure using the identical coloring scheme. Labeled are the 5' and 3' ends, the resume (R) and stop codons (S), and the three regions (TLD, MLR, PKD). The figure was produced with iMol and the PDB coordinates of

additional file Ecoli-closed.pdb are available at
<http://rnp.uthscsa.edu/rnp/tmRDB/tmRDB.html>.

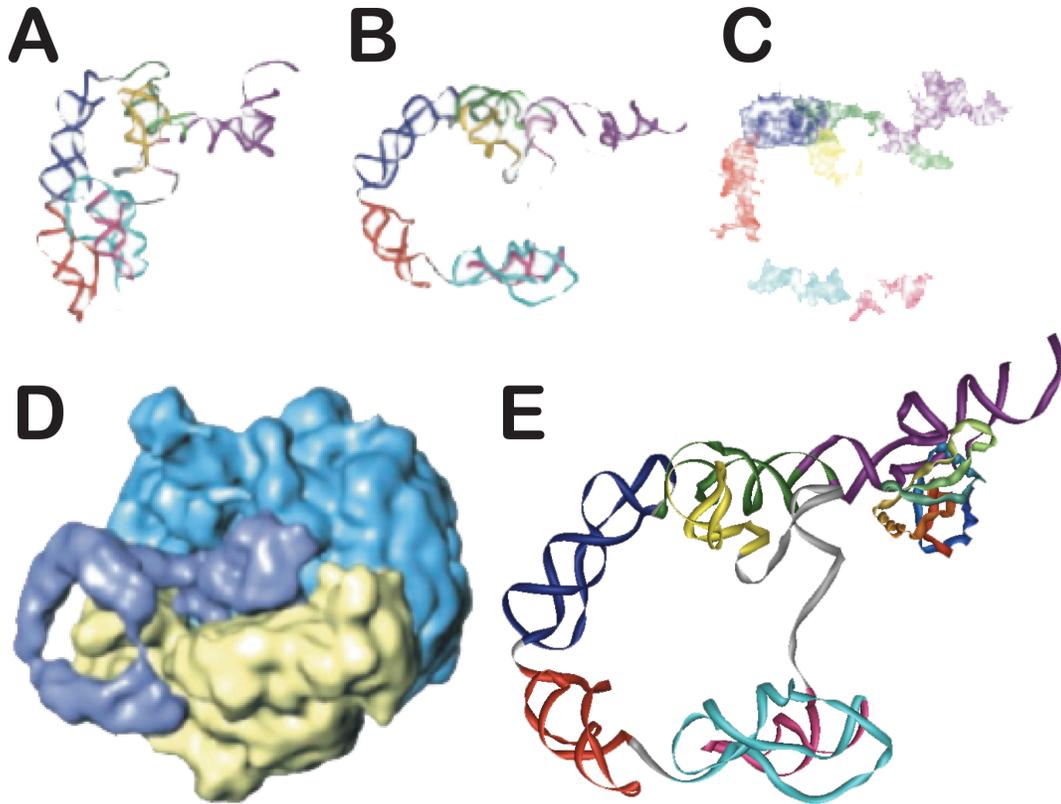


Figure 3.4. Conformational changes in *Escherichia coli* tmRNA. **A)** Closed form of the *E. coli* model as shown in Figure 3.3. **B)** Open conformation adjusted to more closely resemble the ribosome-bound form as determined by cryo-EM (Valle *et al.*, 2003) using additional file Ecoli-open.pdb. **C)** Coordinates extracted from the cryo-EM model (Valle *et al.*, 2003). The TLD is shown in dark purple, helix 2 in green, pk1 in yellow, helix 5 in pink, pk2 in turquoise, pk3 in red, and pk4 in dark blue. **D)** Electron density map of the 50S subunit in light blue, the 30S subunit in yellow, and the bound tmRNA (in the absence of ribosomal protein S1) in dark blue (from Valle *et al.*, 2003). **E)** Recently updated *E. coli* tmRNA incorporating crystal structure of TLD-SmpB from *T. thermophilus* and revised loop 2ab motif as described in text.

Genetic Group	TLD	MLR	Pk1	pk2	pk3	pk4	Other
Thermophilic Oxygen Reducers							
Thermatogales							
Green Non-sulfur & Bacteria relatives							
Flexibacter Cytophaga Bacteroides							
Green Sulfur Bacteria							
Planctomyces & relatives							
Cyanobacteria					1	1,2	3
Plastids					-	-	
Mitochondria		-	-	-	-	-	
Fibrobacter/Acidobacter & relatives							
Spirochetes & relatives							
Proteobacteria, alpha				4		5	3
Proteobacteria, beta				6			3
Proteobacteria, gamma				6			
Proteobacteria, delta							
Proteobacteria, epsilon							
Fusobacteria							
Gram Positive Bacteria							

Table 3.1. Phylogenetic distribution of tmRNA features. The tRNA-like domain (TLD), mRNA-like region (MLR), and the four pseudoknots pk1 to pk4 are shown on the top. Other features peculiar to a phylogenetic group are in the right column. White fields indicate the presence, dark gray fields the absence of a feature. Light gray suggests structural modifications as noted: (1) certain Cyanobacteria lack these pseudoknots. (2) One-chain cyanobacterial tmRNAs contain two smaller tandem pseudoknots named pk4a and 4b. (3) The tmRNAs of some species in this group consist of two basepaired molecules (Keiler *et al.*, 2000; Williams, 2002). (4) The genus *Rickettsia* and its relatives lack pk2. (5) pk4 of the alpha-Proteobacteria has been reduced to a single helix (named helix 11). (6) Some species in this group contain an additional helix (helix 6d).

Motif	SCOR class	tmRNA Res.	Source Res.	Coordinates	Comments
1		1–7, 353–363	1–7, 12–22	1IKD.pdb (chain W)	ACCA end and G3-U357 pair
2		8–28, 325–352	8–28, 325–352	tmx-34.pdb from tmRDB	
3a	<i>internal loop with unpaired stacked bases</i>	29–33, 321–324	1775–1779, 1765–1768	1JJ2.pdb	
3b	<i>stacked duplex with one non-WC pair</i>	C35, A319		ERNA-3D	
4	<i>stacked duplex with two non-WC pairs</i>	38–39, 315–316	2874–2875, 2882–2883	1JJ2.pdb	
5		309–311		ERNA-3D	
6a	<i>pseudoknot</i>	49–78	1–33	1RNK.pdb	pk1
6b	<i>tetraloop</i>	87–98	5–8	1AFX.pdb	the only YRRR tetraloop in SCOR
7	<i>nonallop</i>	118–126	1834–1842	1JJ2.pdb	
8	<i>one unpaired and stacked U</i>	U131	U30	1B36.pdb	
9		171–174		ERNA-3D	
10a	<i>stacked duplex with two non-WC pairs</i>	149–150, 165–166	288–289, 363–364	1JJ2.pdb	
10b	<i>pseudoknot</i>	138–196		6a	pk2
11a	<i>internal loop</i>	204–206, 223–225	780–782, 800–802	1J5E.pdb	
11b	<i>pseudoknot</i>	197–247		6a	pk3
12	<i>stacked duplex with one non-WC pair</i>	G258, A273	A-G6, B-A27	420D.pdb	
13a	<i>stacked duplex with one non-WC pair</i>	C266, U296		ERNA-3D	
13b	<i>pseudoknot</i>	248–299		6a	pk4

Table 3.2. Structural motifs used for the *Escherichia coli* tmRNA model. Shown in columns one to four are the motif numbers in bold, their SCOR classification (Klosterman *et al.*, 2002), the residue positions in the tmRNA model and the source structure. Column five lists the filenames containing the atomic coordinates that were derived from the PDB (Sussman *et al.*, 1999), the tmRDB (Zwieb *et al.*, 2003), or were generated by ERNA-3D (Mueller *et al.*, 1995).

CHAPTER 4: *IN SILICO* ANALYSIS OF IRES RNAs OF FOOT-AND-MOUTH DISEASE VIRUS AND RELATED PICORNAVIRUSES

Burks, J., Zwieb, C., Mueller, F., Wower, I.K., & Wower, J. *In Silico* Analysis of IRES RNAs of Foot-and-Mouth Disease Virus and Related Picornaviruses.

ABSTRACT

Foot-and-Mouth Disease Virus (FMDV) uses an Internal Ribosome Entry Site (IRES), a highly structured segment of its genomic RNA, to hijack the translational apparatus of the infected host. Computational analysis of 161 Type II picornavirus IRES RNA sequences yielded FMDV and closely related picornavirus RNA secondary structures which included only base pairs supported by comparative or experimental evidence. The deduced helical sections provided the foundation for a three-dimensional model of FMDV IRES RNA. The model was further constrained by incorporation of available data derived from chemical and enzymatic modification experiments as well as high-resolution information about IRES RNA-bound proteins. A hypothetical model for IRES-ribosome interaction is proposed.

INTRODUCTION

Foot-and-mouth disease (FMD) is an acute, highly contagious infection of cloven-hoofed animals (Grubman & Baxt, 2004; Mahy, 2005). It is caused by Foot-and-Mouth Disease Virus (FMDV), an *Aphthovirus* of the family *Picornaviridae*. The genomic RNA of this virus contains one open reading frame (ORF) flanked by 5'- and 3'-untranslated regions (5'- and 3'-UTR). Translation of the ORF produces a large polyprotein that is post-translationally cleaved into a number of structural and non-structural proteins (reviewed in Belsham, 2005). To synthesize the polyprotein, FMDV internally initiates translation in a cap-independent process facilitated by the internal ribosome entry site (IRES), a segment of the 5'-UTR that is required for binding the viral genomic RNA to ribosomes and recruiting canonical translation initiation factors (Belsham & Bostock, 1988; Belsham & Brangwyn, 1990; Kuhn *et al.*, 1990). The activities of IRES RNAs are stimulated by several RNA-binding proteins provided by the infected host (Belsham, 2005; Jackson, 2002; Pestova *et al.*, 2001; Pilipenko *et al.*, 2000).

FMDV IRES RNA is classified as a Type II picornavirus IRES RNA secondary structure (Wimmer *et al.*, 1993). Structure-function relationships in the FMDV IRES RNA are interpreted using two different secondary structure models. The first is derived from comparative analysis of 5'-UTR sequences from three FMDV strains, four Encephalomyocarditis virus (EMCV) strains, three Theiler's murine encephalomyelitis virus (TMEV) strains, as well as DMS modifications and RNase V1 and S1 cleavages of strain RRR of EMCV (Pilipenko *et al.*, 1989). The secondary structure model published by Fernandez-Miragall and Martinez-Salas (2003) relies on mfold energy minimization calculations (Mathews *et al.*, 1999; Zuker, 2003) and structural probing data of domain

III of transcribed BVDV IRES RNA. Additional probing of the structure suggested regions in domains II and IV (Fernandez-Miragall & Martinez-Salas, 2007; Fernandez-Miragall *et al.*, 2006). Both models share five domains (denoted I-V in FMDV IRES RNA), but differ with respect to their base pairing patterns.

Recently, cryo-electron microscopy (cryo-EM), X-ray crystallography and NMR spectroscopy provided insights into the three-dimensional arrangements of structures of and in IRES RNAs from Hepatitis C Virus (HCV; *Flaviviridae*), Classical Swine Fever Virus (CSFV; *Flaviviridae*), and the intergenic region (IGR) IRESes of Cricket Paralysis-like Virus (CrPV; *Dicistroviridae*) and *Plautia Stali* Intestine Virus (PSIV; *Dicistroviridae*) (Boehringer *et al.*, 2005; Collier *et al.*, 2002; Kieft *et al.*, 2002; Locker *et al.*, 2007; Lukavsky *et al.*, 2003; Lukavsky *et al.*, 2000; Pfingsten *et al.*, 2006; Rijnbrand *et al.*, 2004; Schuler *et al.*, 2006; Spahn *et al.*, 2001; Spahn *et al.*, 2004; Zhao *et al.*, 2008). Despite highly different sequences, HCV and CrPV IRES RNAs bind to the neck and platform of the 40S subunit (Boehringer *et al.*, 2005; Kieft *et al.*, 2001; Kieft, 2008). However it is unclear where or how the structurally distinct picornavirus IRES RNAs bind to the ribosome.

To facilitate the study of these important viruses and to better understand the structure and functions of Type II picornavirus IRES RNAs, we compared 161 sequences *in silico* to identify covarying base pairs. Comparative sequence analysis (CSA) was proven effective in the construction of reliable secondary structures of ribosomal RNAs, transfer RNA, signal recognition particle RNA and transfer-messenger RNA (Holley, 1968; Larsen & Zwieb, 1991; Woese *et al.*, 1980; Zwieb *et al.*, 1999). Our studies yield a revised model of the secondary structure of FMDV IRES RNA supported by both

covarying base pairs and available biochemical data for reported for functional FMDV IRES RNAs. The updated secondary structure model was used to investigate the possibilities for FMDV IRES RNA in three dimensions in its free form, when bound to IRES-associated proteins and the 40S ribosomal subunit.

RESULTS AND DISCUSSION

Comparative Sequence Analysis

Sequences of FMDV IRES RNAs collected from Rfam and GenBank were aligned using BioEdit (see Methods) (Benson *et al.*, 2009; Griffiths-Jones *et al.*, 2003; Hall, 1999). To derive secondary structure, we observed covariations which maintained base pairs of the Watson-Crick type (A-U, G-C and G-U) despite differences in sequence. Such compensatory base changes (CBCs) supported the existence of base paired regions because, during evolution, random single mutations that disrupt pairing would not have been compensated for by mutations that restored stability unless required. A mismatch provided negative evidence for a base pair, whereas an invariant pair provided neither positive nor negative support for its existence (Larsen & Zwieb, 1991). A base pair was considered supported if there were at least twice as many CBCs as there were mismatches (Larsen & Zwieb, 1991). Invariant canonical pairs may be supported if neighbored by one compensatory pair as an extension of a helical section. The covariation analyses were assisted by RNAdbTools and the Semi-Automated RNA Sequence Editor (SARSE) (Gorodkin *et al.*, 2001; Andersen *et al.*, 2007).

Due to the observed high levels of sequence conservation, analysis of the 129 FMDV IRES RNAs provided an insufficient number of CBCs. Therefore, the alignment

was expanded to include IRES RNA sequences from related aphthoviruses, cardioviruses (including EMCV and Mengovirus) and parechoviruses (including Ljungan viruses) obtained from GenBank and Rfam (Benson *et al.*, 2009; Griffiths-Jones *et al.*, 2003).

The sequences were grouped according to immunological serotype and viral taxonomic classification, followed by realignment using CLUSTALW and manual editing (Thompson *et al.*, 1994). The final alignment contained 161 nonredundant sequences from aphthoviruses (FMDV; 129 sequences from serotypes A, Asia1, O, C, SAT1, SAT2 and SAT3), cardioviruses (EMCV and TMEV; 20 sequences), and parechoviruses (12 sequences).

Secondary Structure of the FMDV IRES RNA

Figure 4.1 shows the derived secondary structure diagram of the FMDV IRES RNA. CSA and corroborating experimental evidence (discussed below) support 30 helical sections arranged in five distinct domains D1 through D5, which correspond to domains I-V in the secondary structure model proposed by Pilipenko *et al.* (1989). The poorly supported helix of the functionally obsolete D1 was not investigated in detail in any of the Type II RNA sequences analyzed (Belsham & Brangwyn, 1990; Jang & Wimmer, 1990). Below, we describe the properties of domains D2 through D5 and highlight structural differences between IRES RNAs of FMDV and related picornaviruses.

Domain 2

This region consists of four helical sections 2a-2d. Sections 2a and 2b are separated by a symmetric loop in FMDV, but an asymmetric loop exists in all other Type II IRES RNAs. Section 2b in FMDV results from base pairing between residues 288-289 and 326-327 as initially proposed (Pilipenko *et al.*, 1989). Present analysis does not support the 287-288 and 324-325 base pairings suggested by energy minimization (Fernandez-Miragall & Martinez-Salas, 2003). Sections 2b and 2c are separated by a short symmetric loop in FMDV, but they are continuously stacked in the cardiovirus and parechovirus RNAs. In EMCV and TMEV, 2a and 2bc are separated by an asymmetric loop of five to seven nucleotides in the 5' portion and two to five nucleotides in the 3' portion. Another asymmetric loop connects helical sections 2c and 2d in all genera. Helix 2 is capped by a pyrimidine-rich loop containing two to five residues which are likely to be part of a binding site for polypyrimidine tract binding protein (PTB) (Luz & Beck, 1991).

Domain 3

This domain is composed of two unequally conserved subdomains (Pilipenko *et al.*, 1989). The variable subdomain (D3V) contains sections 3a-3i, whereas the conserved subdomain (D3C) is composed of section 3j and helices 3.1-3.5 (Pilipenko *et al.*, 1989). Sections 3a-3j are present in all Type II picornavirus IRES RNAs and may exhibit substantial stacking. In the cardiovirus and Ljungan parechovirus RNAs, two additional sections (3k and 3l, not shown) are inserted between 3h and 3i.

Helical sections 3i-3j are well supported by CSA, whereas helix 3.1 is consistent

with earlier chemical probing of EMCV IRES RNA (Pilipenko *et al.*, 1989). CSA provides weak support for helix 3.5 because the distal and proximal base pairs are compensatory and their neighboring invariant pairs 228G-C242 (O1K numbering) and 230G-C240 can be supported by extension. Chemical and enzymatic modification analyses indicated that helix 3.5 existed in analyzed EMCV IRES RNA but may have been absent in FMDV IRES RNA samples in later experiments with *in vitro* transcribed RNA (Fernandez-Miragall & Martinez-Salas, 2003; 2007; Fernandez-Miragall *et al.*, 2006; Pilipenko *et al.*, 1989). The functionally important A/C-rich loop of helix 3.5 was accessible to single-strand RNA-specific RNase S1 in EMCV but not the proposed base pairs of helix 3.5 (Kaminski *et al.*, 1994; Lopez de Quinto & Martinez-Salas, 1997; Pilipenko *et al.*, 1989; Pilipenko *et al.*, 2000) (the A/C-rich loop as well as the CRAAA loop of FMDV IRES was shown to be DMS accessible by Fernandez-Miragall, 2003, 2006 and 2009 in *in vitro* transcribed RNA). It could be possible that this structure may exist in equilibrium. Helix 3.2b is capped by a GNRA tetraloop (where N is any nucleotide and R is a purine) in all Type II IRES RNA (Robertson *et al.*, 1999). Mutational analysis in FMDV strain C-s8c1 demonstrated that this tetraloop is essential for IRES RNA activity (Lopez de Quinto & Martinez-Salas, 1997).

The invariant base pairs of helix 3.3 (Figure 4.1) cannot be proved or disproved by CSA. However, helix 3.3 and its conserved pentaloop (CRAAA) are indirectly supported by data derived from DMS modification as well as cleavage by RNases V1 and S1 of EMCV IRES (Pilipenko *et al.*, 1989). One exception to the pentaloop conservation (CGCAA) occurs in FMDV O Akesu/58 and o1argentina iso5 (GenBank Accessions AF511039 and AY593814). The first base pair of helix 3.4 is an invariant G-C. This

structure is well supported by DMS modification data and RNase V1 and S1 cleavage analyses (Pilipenko *et al.*, 1989).

Domain 4

D4 contains helical sections 4a-4c and helices 4.1 (D4C) and 4.2 (D4V) (Pilipenko *et al.*, 1989). Section 4a is supported by compensatory mutations of the IRES RNA in FMDV strain C-s8c1 as well as DMS modifications and cleavages by RNases V1 and S1 of EMCV strain R (Lopez de Quinto *et al.*, 2001; Lopez de Quinto & Martinez-Salas, 2000; Pilipenko *et al.*, 1989). Sections 4a and 4b are separated by an internal loop of two to four residues on each strand, suggested by mutagenesis to be required for proper FMDV IRES RNA-mediated translation and binding of eIF4G (Lopez de Quinto *et al.*, 2001; Lopez de Quinto & Martinez-Salas, 2000).

Helical section 4.1a is highly conserved and supported by protection from DMS modification and cleavage by RNases V1 and S1 of EMCV (Pilipenko *et al.*, 1989). The last two base pairs of section 4.1a are required for IRES activity as indicated by the analysis of compensatory mutations (Bassili *et al.*, 2004). Section 4.1b is composed of four invariant base pairs and supported indirectly by the observation that disruption of the stack or perturbation of the sequence disrupt binding of initiation factors eIF4G and eIF4B (Bassili *et al.*, 2004). Moreover, section 4.1b is separated from sections 4.1a and 4.1c by two conserved bulges (residues 359-GA-360 and 328-AC-329, respectively). In FMDV strain O/SKR/2002 (GenBank accession AY312589), one uridine is inserted into the GA bulge. Helix 4.1 is capped by a loop of two to 13 nucleotides. Such diloops have been previously observed in ribosomal RNA (Jucker & Pardi, 1995).

Helix 4.2 consists of the well supported sections 4.2a and 4.2b in aphthoviruses, cardioviruses and most parechoviruses, with a pyrimidine-rich loop of five to eight residues. The invariant residues C378, U379 and U381 are present in all 4.2b stem-loops. Section 4.2b is replaced by two smaller helices (4.2c and 4.2d, not shown) separated by 5'-GGGUAGAA-3' in Ljungan parechoviruses. These helices are capped by a four to seven residue loop and a tetraloop, respectively (Johansson *et al.*, 2002).

The single strand which connects sections 4.2a and 4c contains five invariant adenine residues in FMDV (strain C-s8c1) and TMEV (strain GDVII) IRES RNAs (GenBank accessions AJ133357 and M20562). A sixth adenine residue is found in all strains of EMCV, except EMCV-30 (GenBank accession AY296731) which contains one additional guanine residue (LaRue *et al.*, 2003; Pevear *et al.*, 1988).

Domain 5 and the 3' region of the IRES RNA

In FMDV and most parechoviruses, helix 5 consists of sections 5a (five base pairs) and 5b (four base pairs) separated by one residue, but seven to eight base pairs are found in cardioviruses and Ljungan parechoviruses. Helix 5 is capped by a loop of up to four residues and connects to a single-stranded region *via* two to nine pyrimidines leading up to the AUG start codon used by the majority of cardioviruses to initiate protein synthesis (Kaminski *et al.*, 1990; Kong & Roos, 1991). Most strains of FMDV initiate at an AUG triplet located further downstream, but many experimental FMDV IRES RNA constructs can express their downstream proteins using the first start codon (Belsham, 1992; 2005; Sangar *et al.*, 1987). No recognizable consensus motif was found in the

sequence linking helix 5 and the AUG start codon. This variable region typically consists of 22-25 residues, but some sequences contain up to 42 residues.

Distribution of Conserved Elements In FMDV IRES RNA Secondary Structure

Overall, helical sections 2b, 2c, 3a, 3b, 3d through 3i, 3j and helix 4.2, as well as the connecting regions between domains proved to be the least conserved regions. Clusters of invariant residues were identified in sections 2a, 3c, 3g, 3.1, 3.2, 3.3, 3.5, 4a-4c, 4.1 and 5, as well as unpaired residues 44, 50, 54-56, 62-63, 67, 93, 112-113, 141-146, 177, 180, 197, 200-201, 269-270, 288-289, 316, 328-329, 338-341, 359-360, 397-401, 404, 423 and 428-430 (Figure 4.1). This uneven distribution of the conserved elements likely results from restrictions imposed not only by the IRES RNA structure, but also specific interactions with associated proteins and the ribosome.

Modeling the Three-Dimensional Structure of FMDV IRES RNA

In earlier studies we demonstrated that RNA secondary structures resulting from CSA could be used as reliable blueprints for building meaningful tertiary structure models (Burks *et al.*, 2005). In this process, base pairing information is entered into the molecular modeling program ERNA-3D to generate a preliminary three-dimensional model containing A-form RNA for the helical sections while the single-stranded segments adopt conformations according to ERNA-3D's built-in algorithm (Mueller *et al.*, 1995). In order to obtain biologically feasible loop structures, the sequence for each loop was used to search the Structure Classification of RNA database (SCOR) for similar structures in the Protein Data Bank (PDB) (Berman *et al.*, 2002; Klosterman *et al.*, 2002).

The coordinates of the loops were copied onto the preliminary model (Figure 4.2) as specified in Table 4.2.

Modeling Constraints

Covariation analysis is not limited to defining secondary structure, but can also be used to identify possible long-range interactions (Larsen & Zwieb, 1991). Attempts to identify canonical RNA-RNA tertiary interactions for constraining the model were unsuccessful. Therefore, because FMDV IRES RNA interacts with host proteins, we used available data for these interactions to constrain the model in three dimensions. FMDV IRES RNA binds to polypyrimidine tract binding protein (PTB), IRES *Trans*-acting Factor 45 (ITAF45), initiation factors eIF4G, eIF4A, eIF3 eIF2 and other proteins, but detailed structural information for complexes between these proteins and RNA is limited to PTB (Lopez de Quinto *et al.*, 2001; Lopez de Quinto & Martinez-Salas, 2000; Luz & Beck, 1990; 1991; Monie *et al.*, 2007; Oberstrass *et al.*, 2005; Pilipenko *et al.*, 2000). PTB binds FMDV IRES as a monomer of four RRM-motif-containing RNA-binding domains, of which the third and fourth are considered to be prominent for FMDV and were shown to bind to helix 2 and a segment containing helix 5 and the 3' polypyrimidine tract of FMDV IRES RNA (Amir-Ahmady *et al.*, 2005; Maris *et al.*, 2005; Monie *et al.*, 2007; Monie *et al.*, 2005; Oh *et al.*, 1998; Perez *et al.*, 1997; Simpson *et al.*, 2004; Song *et al.*, 2005). The NMR structure of PTB domains three and four in complex with two 5'-CUCUCU-3' hexamers (PDB ID 2ADC; Oberstrass *et al.*, 2005) was used as a three-dimensional ruler to constrain U53 and U59 relative to the residues at positions 441-446 (see Table 4.2). These choices are consistent with the requirement of U54 for PTB

binding, the protection of nucleotides 53-UU-54 and 439-447 from CMCT modification (Table 4.1), the observation that purine substitutions in 439-UUUC-442 interfere with PTB binding, and the sequence similarity between 441-UCCUUU-446 in FMDV IRES and 532-CUCUCU-537 of the NMR structure (Kolupaeva *et al.*, 1996; Luz & Beck, 1991; Oberstrass *et al.*, 2005; Pilipenko *et al.*, 2000; Song *et al.*, 2005).

Three-Dimensional FMDV IRES RNA Model

The model is elongated with approximate dimensions of 80Å by 90Å by 300Å (Figure 4.3). Helix 5 and the third and fourth domains of PTB are located in the fork formed by helices 2 and 4.2. The coaxial stack of helix 2 is interrupted by the asymmetric loop between sections 2c and 2d. Sections 3a-3i are seen as an elongated region connected to the cluster of sections 3i, 3j and helices 3.1 and 3.5. Sections 4a-4c and helices 4.1 and 4.2 form three stacks. Helix 5 is nestled between helices 4.1 and 4.2, sections 4a-4c, and PTB (Figure 4.3). This model is compatible with chemical or enzymatic protection data (see Table 4.1) including those that were not utilized for its construction, as well as with low-resolution electron microscopy data available for the free form of FMDV IRES RNA (Beales *et al.*, 2003). Protections from chemical modifications or enzymatic cleavages due to PTB may be caused by direct binding or indirectly through conformational changes. Protections from RNase P1 and T1 cleavages in the loop between helical sections 2c and 2d (Figure 4.3, Table 4.1) may be explained by the interaction of the second PTB RNA-binding domain as this domain has the potential to be close to residues 63-65. The numerous protections in FMDV IRES RNA helix 3 caused by PTB may be due to the binding of either the first or second PTB RNA-

binding domains or the concerted binding of the third and fourth PTB RNA-binding domains and ITAF45 (Pilipenko *et al.*, 2000). Such conformational changes seem likely because enhanced accessibility to CMCT modifications and RNase T1 cleavages after PTB/ITAF45 binding have been demonstrated (Pilipenko *et al.*, 2000). Furthermore PTB, in concert with ITAF45, is thought to play a role as an RNA-folding chaperone and induces conformational changes throughout the IRES RNA (see Table 4.1) (Kolupaeva *et al.*, 1996; Luz & Beck, 1991; Pilipenko *et al.*, 2000; Song *et al.*, 2005).

In contrast to the binding data for PTB and FMDV IRES and while this manuscript was in preparation, Kafasla *et al.* (2009) showed that domains 2 and 3 of PTB bound to EMCV IRES RNA domains K (4.2 in FMDV IRES) and H (2), respectively, using Fe-II-directed hydroxyl radical probing experiments. The comparison of the EMCV IRES RNA secondary structure (not shown) and data of Kafasla *et al.* with the secondary structure of FMDV IRES RNA support the arrangements of the domains in the model of FMDV IRES RNA, suggesting that domain I of EMCV IRES (and presumably helix 3 of FMDV IRES) is located independently from domains H (2), J (4.1) and L (5) which may be located in a separate cluster due to clustered patterns of cleavage caused by close proximity to one PTB domain or another. What remains to be seen is if the order of subunit binding is unique in both FMDV and EMCV IRES RNAs, and if this is due to nucleotide or secondary structure placement in the binding site of the PTB domains. Crystallography could resolve this discrepancy in the future.

Topography of the IRES Ribonucleoprotein Complex

Each picornavirus has different requirements for protein factors (reviewed in Belsham, 2005; Belsham & Jackson, 2000). Figure 4.4 shows the complex network of interactions that occur during FMDV IRES RNA-mediated translation initiation. *In vitro* studies demonstrated that Type II IRES RNA recruitment to the 40S ribosomal subunit requires eIF4G, eIF4A, eIF3 and eIF2 (Pestova *et al.*, 1996a). FMDV additionally requires PTB and ITAF45 (Pilipenko *et al.*, 2000). Initiation factor eIF4B may bind helix 5 and promote the formation of 48S complexes on either FMDV or EMCV IRES RNAs (Lopez de Quinto *et al.*, 2001; Meyer *et al.*, 1995).

Protection and toeprint data are available for eIF4G (Kolupaeva *et al.*, 1998; Pilipenko *et al.*, 2000). This initiation factor is part of the eIF4F complex, along with eIF4A and the cap-binding protein eIF4E (Pause *et al.*, 1994a; Svitkin *et al.*, 2001). Both eIF4G and eIF4A are required for translation of Type II picornavirus IRES RNAs (Pause *et al.*, 1994b; Pestova *et al.*, 1996a; Pestova *et al.*, 1996b; Svitkin *et al.*, 2001). The C-terminal portion of eIF4G contains the RNA, eIF3 and eIF4A binding sites, and plays a central role in organizing the 48S complex (Gross *et al.*, 2003; Kolupaeva *et al.*, 1998; Korneeva *et al.*, 2000; Korneeva *et al.*, 2001; LeFebvre *et al.*, 2006; Morino *et al.*, 2000; Pestova *et al.*, 1996a; Pestova *et al.*, 1996b). *In vitro* mutagenesis studies demonstrated that the binding site for eIF4G is located in helix 4 of the FMDV IRES RNA (Bassili *et al.*, 2004). Disruptions of sections 4a and 4b have detrimental effects on eIF4G binding to FMDV IRES RNA, and the loop between sections 4a and 4b does not tolerate any mutations (Lopez de Quinto *et al.*, 2001; Lopez de Quinto & Martinez-Salas, 2000). The observed biochemical effects caused by the binding of eIF4G (alone or in complex with

eIF4A) are almost exclusively located in helices 4 and 5 (Figure 4.3, Table 4.1) (Kolupaeva *et al.*, 1998; Pilipenko *et al.*, 2000). Residues in the 4ab loop, sections 4b, 4c, the A-rich loop between 4.2a and 4c, and helices 4.1 and 4.2 are protected from chemical modifications or enzymatic cleavage (Figure 4.3), while eIF4G increases the accessibility of G393 in section 4.2b and 411-CC-412 in section 4a to RNase V1 cleavage (Kolupaeva *et al.*, 1998).

In vitro, eIF4A binds to an FMDV IRES RNA fragment composed of helices 4 and 5, and the 3' polypyrimidine tract, but not when the fragment lacks helix 4.1 (Stassinopoulos & Belsham, 2001). It is unclear whether eIF4G or eIF4A are individually responsible for the 413-UG-414 eIF4F and eIF4G:eIF4A toeprints, the 419-GGU-421 eIF4G:eIF4A toeprint, or the protections from RNase T1 cleavage observed at residues G317, G359 and G402 (Kolupaeva *et al.*, 1998; Pilipenko *et al.*, 2000).

The multisubunit mammalian initiation factor eIF3 was shown to bind to eIF4G (Browning *et al.*, 2001; Morris-Desbois *et al.*, 2001; Zhou *et al.*, 2005; Zhou *et al.*, 2008). This interaction is of critical importance for the delivery of the 40S ribosomal subunit to the initiation site of the EMCV IRES RNA (LeFebvre *et al.*, 2006). eIF3 binds to the 40S ribosomal subunit and was reported to make multiple contacts with the FMDV IRES RNA, including but not limited to helix 5 (Fraser *et al.*, 2004; Lopez de Quinto *et al.*, 2001; Pacheco *et al.*, 2008).

FMDV IRES RNA on the 40S Ribosomal Subunit

The FMDV IRES RNA protein complex interacts with the 40S ribosomal subunit during translation initiation. The topography of the initiation complex was approximated

using available protein:ribosome and IRES:ribosome interaction data. Information about the interactions of eIF4G and eIF3 with FMDV IRES RNA was the most useful (Korneeva *et al.*, 2000; Siridechadilok *et al.*, 2005). Although the eIF3 binding site on FMDV IRES RNA is unknown, cryo-EM studies of the eIF4G:eIF3 and eIF3:40S complex place the FMDV-associated eIF4G near the ribosomal E site (Siridechadilok *et al.*, 2005). FMDV IRES:eIF4G interactions were not modeled because the structure of this complex is unknown. However, eIF4G protection of sections 4a-4c and helices 4.1 and 4.2 (Figure 4.3 and Table 4.1) supports the localization of this region near eIF4G (Kolupaeva *et al.*, 1998; Siridechadilok *et al.*, 2005). This placed helices 2, 4.1, 4.2, and 5 and PTB at the E site of the ribosome (Figure 4.5), consistent with the models of eIF4G:eIF3 and eIF3:40S complexes (Siridechadilok *et al.*, 2005).

With helix 4.1 located close to the E site, the AUG start codon was easily positioned near the P site without perturbing the remainder of the model (Figure 4.5). This placement was consistent with data from toeprinting analysis of FMDV IRES RNA bound to rabbit reticulocyte ribosomes which demonstrated protections 15 nucleotides downstream of the first start codon (462-AUG-464) (Pilipenko *et al.*, 2000).

Additional clues about the topography of picornavirus IRES RNA:ribosome complexes came from a cross-link between EMCV IRES RNA and ribosomal protein S25 (rpS25) in *Drosophila* (Nishiyama *et al.*, 2007). This protein cross-links to stem loop IV of PSIV IRES RNA (Nishiyama *et al.*, 2007; Pflingsten *et al.*, 2006; Spahn *et al.*, 2004). The second domain of HCV IRES RNA cross-links to ribosomal protein S5 (rpS5) from HeLa cells (Fukushi *et al.*, 2001). rpS5 is located at the “back” of the head of the 40S subunit and cross-links to rpS25 suggesting close proximity between these proteins

(Schuler *et al.*, 2006; Spahn *et al.*, 2004; Tolan & Traut, 1981; Uchiumi *et al.*, 1981). Involvement of rpS5 and rpS25 in functions of IRES RNAs from picornaviruses, flaviviruses and dicistroviruses suggest that these proteins constitute important elements for proper IRES RNA positioning (Pfungsten & Kieft, 2008). We propose that helix 4.1 is close to these ribosomal proteins.

The IRES RNA:protein complex model is positioned such that helix 4.1 is near the platform (Figure 4.5). Helices 3-3.5 and the 5' end are seen outside of the 40S subunit. The start codon of each model is located in the P site. This particular arrangement is plausible in EMCV IRES RNA because translation starts at this codon and would allow proximity required to form a cross-link between the IRES and rpS25, and in FMDV IRES RNA which initiates translation either at the 462-464 start codon (in experimental constructs) or the preferred start codon 84 nucleotides downstream (Belsham, 1992; Sangar *et al.*, 1987). For the second functioning start codon of FMDV IRES RNA to reach the P site, the ribosome presumably scans the FMDV RNA (Andreev *et al.*, 2007; Belsham, 1992; Belsham & Jackson, 2000). If scanning does in fact happen (assisted by eIF1 and eIF1A, which bind helix 5), the FMDV IRES RNA could be outside the ribosome by the time the second, preferred functional start codon enters the P site (Pacheco *et al.*, 2008; Passmore *et al.*, 2007).

CONCLUSIONS

We generated plausible three-dimensional models of the FMDV IRES RNA based on extensive comparative analysis of Type II picornavirus IRES RNA sequences. The models suggest that FMDV IRES RNA similarities with other IRES RNAs at the tertiary

structure level including a close proximity between helix 4.1 and rpS5 and rpS25. As in HCV IRES RNA, FMDV IRES RNA helix 3 may be found outside the ribosome and located away from helices 2, 4.2 and 5. The models will be useful as guides in future studies of the interactions between Type II Picornavirus IRES RNAs, their associated protein factors and the ribosome.

METHODS

Collection and Cataloguing of Available Type II Picornavirus IRES Sequences

The Rfam alignment for *Aphthovirus* IRESes was combined with sequences collected from GenBank using a keyword search using “FMDV”, “FMDV IRES”, “EMCV”, “EMCV IRES”, “TMEV”, “TMEV IRES”, and results of BLAST searches with representative sequences (Benson *et al.*, 2009; Griffiths-Jones *et al.*, 2003). Duplicates were removed and each sequence was given a unique identifier. Sequences were grouped by genus, virus and, if applicable, serotype.

Comparative Sequence Analysis

The sequences were aligned using regions of identity and similarity. For ambiguous regions, RNA secondary structure features were used to align the sequences (Larsen & Zwieb, 1991). The alignment is available in fasta format at <http://rnp.uthscsa.edu/rnp/IRES/FMDVIRESRNA.zip>.

Three-Dimensional Molecular Modeling

The first model was constructed using the sequence of the FMDV IRES RNA from strain O1K (GenBank accession number X00871 positions 252-716). The modeling process was described previously (Burks *et al.*, 2005; see Chapter 3). Briefly, the secondary structure information was used as input for ERNA-3D (Mueller *et al.*, 1995) installed on an SGI workstation running IRIX 6.5 and equipped with CrystalEyes stereovision goggles and a StereoGraphics infrared emitter. The initial model was modified by consulting SCOR (Klosterman *et al.*, 2002). The coordinates of relevant structures were obtained from the PDB (Berman *et al.*, 2002) (see Tables 4.2 and 4.3) and extracted using Swiss-PdbViewer (Guex & Peitsch, 1997). Experimental data obtained from the literature were considered to validate the model. Finally, bond angles and lengths were corrected to produce biologically feasible conformations. The pdb coordinates of the model are available at <http://rnp.uthscsa.edu/rnp/IRES/FMDVIRESRNA.zip>.

ACKNOWLEDGEMENTS

This research was supported by grants from the Alabama Agricultural Experiment Station Foundation to I.K.W. and J.W. and an Auburn University Biogrant to J.W. Publication costs were supported in part by the Upchurch Fund for Excellence. J.M.B. was supported by the National Science Foundation under Grant No. 0091853 and NSF-EPS 0447675. The molecular graphics image in Figure 4.5 was produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081).

REFERENCES

- Amir-Ahmady, B., Boutz, P. L., Markovtsov, V., Phillips, M. L. & Black, D. L. (2005). Exon repression by polypyrimidine tract binding protein. *RNA* 11, 699–716.
- Andersen, E., Lind-Thomsen, A., Knudsen, B., Kristensen, S., Havgaard, J., Torarinsson, E., Larsen, N., Zwieb, C., Sestoft, P., Kjems, J. & Gorodkin, J. (2007). Semiautomated improvement of RNA alignments. *RNA* 13, 1850-1859.
- Andreev, D. E., Fernandez-Miragall, O., Ramajo, J., Dmitriev, S. E., Terenin, I. M., Martinez-Salas, E. & Shatsky, I. N. (2007). Differential factor requirement to assemble translation initiation complexes at the alternative start codons of foot-and-mouth disease virus RNA. *RNA* 13, 1366-1374.
- Bassili, G., Tzima, E., Song, Y., Saleh, L., Ochs, K. & Niepmann, M. (2004). Sequence and secondary structure requirements in a highly conserved element for foot-and-mouth disease virus internal ribosome entry site activity and eIF4G binding. *J Gen Virol* 85, 2555-2565.
- Beales, L., Holzenburg, A. & Rowlands, D. (2003). Viral internal ribosome entry site structures segregate into two distinct morphologies. *J Virol* 77, 6574-6579.
- Belsham, G. J. (1992). Dual initiation sites of protein synthesis on foot-and-mouth disease virus RNA are selected following internal entry and scanning of ribosomes in vivo. *EMBO J* 11, 1105-1110.
- Belsham, G. J. (2005). Translation and replication of FMDV RNA. *Curr Top Microbiol Immunol* 288, 43-70.
- Belsham, G. J. & Brangwyn, J. K. (1990). A region of the 5' noncoding region of foot-and-mouth disease virus RNA directs efficient internal initiation of protein synthesis within cells: involvement with the role of L protease in translational control. *J Virol* 64, 5389-5395.
- Belsham, G. J. & Jackson, R. (2000). Translation initiation on picornavirus RNA. In *Translational Control of Gene Expression*, pp. 869-900. Edited by N. Sonenberg, J. W. B. Hershey & M. B. Mathews. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. (2009). GenBank. *Nucleic Acids Res* 37, D26-31.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G. Bhat, T.N., Weissig, H., Shindyalov, I.N. & Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Res* 28-235-242. <http://www.pdb.org>.

- Boehringer, D., Thermann, R., Ostareck-Lederer, A., Lewis, J. & Stark, H. (2005). Structure of the hepatitis C virus IRES bound to the human 80S ribosome: remodeling of the HCV IRES. *Structure* 13, 1695-1706.
- Browning, K. S., Gallie, D. R., Hershey, J. W., Hinnebusch, A. G., Maitra, U., Merrick, W. C. & Norbury, C. (2001). Unified nomenclature for the subunits of eukaryotic initiation factor 3. *Trends Biochem Sci* 26, 284.
- Burks, J., Zwieb, C., Müller, F., Wower, I. & Wower, J. (2005). Comparative 3-D modeling of tmRNA. *BMC Mol Biol* 6, 14.
- Collier, A.J., Gallego, J., Klinck, R., Cole, P.T., Harris, S.J., Harrison, G.P., Aboul-Ela, F., Varani, G. & Walker, S.A (2002) A conserved RNA structure within the HCV IRES eIF3-binding site. *Nat Struct Biol* 9, 375-380.
- Fernandez-Miragall, O. & Martinez-Salas, E. (2003). Structural organization of a viral IRES depends on the integrity of the GNRA motif. *RNA* 9, 1333-1344.
- Fernandez-Miragall, O. & Martinez-Salas, E. (2007). In vivo footprint of a picornavirus internal ribosome entry site reveals differences in accessibility to specific RNA structural elements. *J Gen Virol* 88, 3053-3062.
- Fernandez-Miragall, O., Ramos, R., Ramajo, J. & Martinez-Salas, E. (2006). Evidence of reciprocal tertiary interactions between conserved motifs involved in organizing RNA structure essential for internal initiation of translation. *RNA* 12, 223-234.
- Fraser, C. S., Lee, J. Y., Mayeur, G. L., Bushell, M., Doudna, J. A. & Hershey, J. W. (2004). The j-subunit of human translation initiation factor eIF3 is required for the stable binding of eIF3 and its subcomplexes to 40 S ribosomal subunits in vitro. *J Biol Chem* 279, 8946-8956.
- Fukushi, S., Okada, M., Stahl, J., Kageyama, T., Hoshino, F. B. & Katayama, K. (2001). Ribosomal protein S5 interacts with the internal ribosomal entry site of hepatitis C virus. *J Biol Chem* 276, 20824-20826.
- Gorodkin, J., Zwieb, C. & Knudsen, B. (2001). Semi-automated update and cleanup of structural RNA alignment databases. *Bioinformatics* 17, 642-645.
- Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A. & Eddy, S. R. (2003). Rfam: an RNA family database. *Nucleic Acids Res* 31, 439-441.
- Gross, J. D., Moerke, N. J., von der Haar, T., Lugovskoy, A. A., Sachs, A. B., McCarthy, J. E. & Wagner, G. (2003). Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E. *Cell* 115, 739-750.
- Grubman, M. J. & Baxt, B. (2004). Foot-and-mouth disease. *Clin Microbiol Rev* 17, 465-493.

- Guex, N. & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18, 2714-2723.
- Hall, T. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 41, 95-98.
- Holley, R. W. (1968). Experimental approaches to the determination of the nucleotide sequences of large oligonucleotides and small nucleic acids. *Prog Nucleic Acid Res Mol Biol* 8, 37-47.
- Jackson, R. J. (2002). Proteins Involved in the Function of Picornavirus Internal Ribosomal Entry Sites. In *Molecular Biology of Picornaviruses*, pp. 171-183. Edited by B. Semler, L. & E. Wimmer. Washington, D.C. 20036-2904: ASM Press.
- Jang, S. & Wimmer, E. (1990). Cap-independent translation of encephalomyocarditis virus RNA: structural elements of the internal ribosomal entry site and involvement of a cellular 57-kD RNA-binding protein. *Genes Dev* 4, 1560-1572.
- Jang, S. K., Krausslich, H. G., Nicklin, M. J., Duke, G. M., Palmenberg, A. C. & Wimmer, E. (1988). A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J Virol* 62, 2636-2643.
- Johansson, S., Niklasson, B., Maizel, J., Gorbalenya, A. E. & Lindberg, A. M. (2002). Molecular analysis of three Ljungan virus isolates reveals a new, close-to-root lineage of the Picornaviridae with a cluster of two unrelated 2A proteins. *J Virol* 76, 8920-8930.
- Jucker, F. M. & Pardi, A. (1995). Solution structure of the CUUG hairpin loop: a novel RNA tetraloop motif. *Biochemistry* 34, 14416-14427.
- Kafasla, P., Morgner, N., Pöyry, T. A., Curry, S., Robinson, C. V. & Jackson, R. J. (2009). Polypyrimidine tract binding protein stabilizes the encephalomyocarditis virus IRES structure via binding multiple sites in a unique orientation. *Mol Cell* 34, 556-68.
- Kaminski, A., Belsham, G. J. & Jackson, R. J. (1994). Translation of encephalomyocarditis virus RNA: parameters influencing the selection of the internal initiation site. *EMBO J* 13, 1673-1681.
- Kaminski, A., Howell, M. T. & Jackson, R. J. (1990). Initiation of encephalomyocarditis virus RNA translation: the authentic initiation site is not selected by a scanning mechanism. *EMBO J* 9, 3753-3759.
- Kieft, J., Zhou, K., Grech, A., Jubin, R. & Doudna, J. (2002). Crystal structure of an RNA tertiary domain essential to HCV IRES-mediated translation initiation. *Nat Struct Biol* 9, 370-374.

- Kieft, J., Zhou, K., Jubin, R. & Doudna, J. (2001). Mechanism of ribosome recruitment by hepatitis C IRES RNA. *RNA* 7, 194-206.
- Kieft, J. S. (2008). Viral IRES RNA structures and ribosome interactions. *Trends Biochem Sci* 33, 274-283.
- Klosterman, P., Tamura, M., Holbrook, S. & Brenner, S. (2002). SCOR: a Structural Classification of RNA database. *Nucleic Acids Res* 30, 392-394.
- Kolupaeva, V., Hellen, C. & Shatsky, I. (1996). Structural analysis of the interaction of the pyrimidine tract-binding protein with the internal ribosomal entry site of encephalomyocarditis virus and foot-and-mouth disease virus RNAs. *RNA* 2, 1199-1212.
- Kolupaeva, V., Pestova, T., Hellen, C. & Shatsky, I. (1998). Translation eukaryotic initiation factor 4G recognizes a specific structural element within the internal ribosome entry site of encephalomyocarditis virus RNA. *J Biol Chem* 273, 18599-18604.
- Kong, W. P. & Roos, R. P. (1991). Alternative translation initiation site in the DA strain of Theiler's murine encephalomyelitis virus. *J Virol* 65, 3395-3399.
- Korneeva, N. L., Lamphear, B. J., Hennigan, F. L., Merrick, W. C. & Rhoads, R. E. (2001). Characterization of the two eIF4A-binding sites on human eIF4G-1. *J Biol Chem* 276, 2872-2879.
- Korneeva, N. L., Lamphear, B. J., Hennigan, F. L. & Rhoads, R. E. (2000). Mutually cooperative binding of eukaryotic translation initiation factor (eIF) 3 and eIF4A to human eIF4G-1. *J Biol Chem* 275, 41369-41376.
- Kuhn, R., Luz, N. & Beck, E. (1990). Functional analysis of the internal translation initiation site of foot-and-mouth disease virus. *J Virol* 64, 4625-4631.
- Larsen, N. & Zwieb, C. (1991). SRP-RNA sequence alignment and secondary structure. *Nucleic Acids Res* 19, 209-215.
- LaRue, R., Myers, S., Brewer, L., Shaw, D. P., Brown, C., Seal, B. S. & Njenga, M. K. (2003). A wild-type porcine encephalomyocarditis virus containing a short poly(C) tract is pathogenic to mice, pigs, and cynomolgus macaques. *J Virol* 77, 9136-9146.
- LeFebvre, A. K., Korneeva, N. L., Trutschl, M., Cvek, U., Duzan, R. D., Bradley, C. A., Hershey, J. W. & Rhoads, R. E. (2006). Translation initiation factor eIF4G-1 binds to eIF3 through the eIF3e subunit. *J Biol Chem* 281, 22917-22932.
- Locker, N., Easton, L. & Lukavsky, P. (2007). HCV and CSFV IRES domain II mediate eIF2 release during 80S ribosome assembly. *EMBO J* 26, 795-805.

- Lopez de Quinto, S., Lafuente, E. & Martinez-Salas, E. (2001). IRES interaction with translation initiation factors: functional characterization of novel RNA contacts with eIF3, eIF4B, and eIF4GII. *RNA* 7, 1213-1226.
- Lopez de Quinto, S. & Martinez-Salas, E. (1997). Conserved structural motifs located in distal loops of aphthovirus internal ribosome entry site domain 3 are required for internal initiation of translation. *J Virol* 71, 4171-4175.
- Lopez de Quinto, S. & Martinez-Salas, E. (2000). Interaction of the eIF4G initiation factor with the aphthovirus IRES is essential for internal translation initiation in vivo. *RNA* 6, 1380-1392.
- Lukavsky, P., Kim, I., Otto, G. & Puglisi, J. (2003). Structure of HCV IRES domain II determined by NMR. *Nat Struct Biol* 10, 1033-1038.
- Lukavsky, P., Otto, G., Lancaster, A., Sarnow, P. & Puglisi, J. (2000). Structures of two RNA domains essential for hepatitis C virus internal ribosome entry site function. *Nat Struct Biol* 7, 1105-1110.
- Luz, N. & Beck, E. (1990). A cellular 57 kDa protein binds to two regions of the internal translation initiation site of foot-and-mouth disease virus. *FEBS Lett* 269, 311-314.
- Luz, N. & Beck, E. (1991). Interaction of a cellular 57-kilodalton protein with the internal translation initiation site of foot-and-mouth disease virus. *J Virol* 65, 6486-6494.
- Mahy, B.W.J. (2005) Introduction and History of Foot-and-Mouth-Disease Virus. *Curr Top Microbiol Immunol* 288, 1-8.
- Maris, C., Dominguez, C. & Allain, F. H. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J* 272, 2118-2131.
- Mathews, D., Sabina, J., Zuker, M. & Turner, D. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* 288, 911-940.
- Meyer, K., Petersen, A., Niepmann, M. & Beck, E. (1995). Interaction of eukaryotic initiation factor eIF-4B with a picornavirus internal translation initiation site. *J Virol* 69, 2819-2824.
- Monie, T. P., Perrin, A. J., Birtley, J. R., Sweeney, T. R., Karakasiliotis, I., Chaudhry, Y., Roberts, L. O., Matthews, S., Goodfellow, I. G. & Curry, S. (2007) Structural insights into the transcriptional and translational roles of Ebp1. *EMBO J* 26, 3936-3944.

- Monie, T. P., Hernandez, H., Robinson, C. V., Simpson, P., Matthews, S. & Curry, S. (2005). The polypyrimidine tract binding protein is a monomer. *RNA* 11, 1803–1808.
- Morino, S., Imataka, H., Svitkin, Y. V., Pestova, T. V. & Sonenberg, N. (2000). Eukaryotic translation initiation factor 4E (eIF4E) binding site and the middle one-third of eIF4GI constitute the core domain for cap-dependent translation, and the C-terminal one-third functions as a modulatory region. *Mol Cell Biol* 20, 468-477.
- Morris-Desbois, C., Rety, S., Ferro, M., Garin, J. & Jalinot, P. (2001). The human protein HSPC021 interacts with Int-6 and is associated with eukaryotic translation initiation factor 3. *J Biol Chem* 276, 45988-45995.
- Mueller, F., Doring, T., Erdemir, T., Greuer, B., Junke, N., Osswald, M., Rinke-Appel, J., Stade, K., Thamm, S. & Brimacombe, R. (1995). Getting closer to an understanding of the three-dimensional structure of ribosomal RNA. *Biochem Cell Biol* 73, 767-773.
- Nakashima, N. & Uchiumi, T. (2009). Functional analysis of structural motifs in dicistroviruses. *Virus Res* 139, 137-147.
- Nishiyama, T., Yamamoto, H., Uchiumi, T. & Nakashima, N. (2007). Eukaryotic ribosomal protein RPS25 interacts with the conserved loop region in a dicistroviral intergenic internal ribosome entry site. *Nucleic Acids Res* 35, 1514-1521.
- Oberstrass, F. C., Auweter, S. D., Erat, M., Hargous, Y., Henning, A., Wenter, P., Reymond, L., Amir-Ahmady, B., Pitsch, S., Black, D. L. & Allain, F. H. (2005). Structure of PTB bound to RNA: specific binding and implications for splicing regulation. *Science* 309, 2054-2057.
- Oh, Y. L., Hahm, B., Kim, Y. K., Lee, H. K., Lee, J. W., Song, O., Tsukiyama-Kohara, K., Kohara, M., Nomoto, A. & Jang, S. K. (1998). Determination of functional domains in polypyrimidine-tract-binding protein. *Biochem J* 331 (Pt 1), 169-175.
- Pacheco, A., Reigadas, S. & Martinez-Salas, E. (2008). Riboproteomic analysis of polypeptides interacting with the internal ribosome-entry site element of foot-and-mouth disease viral RNA. *Proteomics* 8, 4782-4790.
- Passmore, L. A., Schmeing, T. M., Maag, D., Applefield, D. J., Acker, M. G., Algire, M. A., Lorsch, J. R. & Ramakrishnan, V. (2007). The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. *Mol Cell* 26, 41-50.
- Pause, A., Belsham, G. J., Gingras, A. C., Donze, O., Lin, T. A., Lawrence, J. C., Jr. & Sonenberg, N. (1994a). Insulin-dependent stimulation of protein synthesis by phosphorylation of a regulator of 5'-cap function. *Nature* 371, 762-767.

- Pause, A., Methot, N., Svitkin, Y., Merrick, W. C. & Sonenberg, N. (1994b). Dominant negative mutants of mammalian translation initiation factor eIF-4A define a critical role for eIF-4F in cap-dependent and cap-independent initiation of translation. *EMBO J* 13, 1205-1215.
- Pelletier, J., Kaplan, G., Racaniello, V. R. & Sonenberg, N. (1988). Cap-independent translation of poliovirus mRNA is conferred by sequence elements within the 5' noncoding region. *Mol Cell Biol* 8, 1103-1112.
- Perez, I., McAfee, J. G. & Patton, J. G. (1997). Multiple RRM's contribute to RNA binding specificity and affinity for polypyrimidine tract binding protein. *Biochemistry* 36, 11881-11890.
- Pestova, T., Hellen, C. & Shatsky, I. (1996a). Canonical eukaryotic initiation factors determine initiation of translation by internal ribosomal entry. *Mol Cell Biol* 16, 6859-6869.
- Pestova, T., Kolupaeva, V., Lomakin, I., Pilipenko, E., Shatsky, I., Agol, V. & Hellen, C. (2001). Molecular mechanisms of translation initiation in eukaryotes. *Proc Natl Acad Sci U S A* 98, 7029-7036.
- Pestova, T., Shatsky, I. & Hellen, C. (1996b). Functional dissection of eukaryotic initiation factor 4F: the 4A subunit and the central domain of the 4G subunit are sufficient to mediate internal entry of 43S preinitiation complexes. *Mol Cell Biol* 16, 6870-6878.
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25, 1605-1612.
- Pevear, D. C., Luo, M. & Lipton, H. L. (1988). Three-dimensional model of the capsid proteins of two biologically different Theiler virus strains: clustering of amino acid difference identifies possible locations of immunogenic sites on the virion. *Proc Natl Acad Sci U S A* 85, 4496-4500.
- Pfingsten, J., Costantino, D. & Kieft, J. (2006). Structural basis for ribosome recruitment and manipulation by a viral IRES RNA. *Science* 314, 1450-1454.
- Pfingsten, J. S. & Kieft, J. S. (2008). RNA structure-based ribosome recruitment: lessons from the Dicistroviridae intergenic region IRESes. *RNA* 14, 1255-1263.
- Pilipenko, E., Blinov, V., Chernov, B., Dmitrieva, T. & Agol, V. (1989). Conservation of the secondary structure elements of the 5'-untranslated region of cardio- and aphthovirus RNAs. *Nucleic Acids Res* 17, 5701-5711.
- Pilipenko, E. V., Pestova, T. V., Kolupaeva, V. G., Khitrina, E. V., Poperechnaya, A. N., Agol, V. I. & Hellen, C. U. (2000). A cell cycle-dependent protein serves as a template-specific translation initiation factor. *Genes Dev* 14, 2028-2045.

- Rijnbrand, R., Thiviyanathan, V., Kaluarachchi, K., Lemon, S. & Gorenstein, D. (2004). Mutational and structural analysis of stem-loop IIIc of the hepatitis C virus and GB virus B internal ribosome entry sites. *J Mol Biol* 343, 805-817.
- Robertson, M., Seamons, R. & Belsham, G. (1999). A selection system for functional internal ribosome entry site (IRES) elements: analysis of the requirement for a conserved GNRA tetraloop in the encephalomyocarditis virus IRES. *RNA* 5, 1167-1179.
- Sangar, D. V., Newton, S. E., Rowlands, D. J. & Clarke, B. E. (1987). All foot and mouth disease virus serotypes initiate protein synthesis at two separate AUGs. *Nucleic Acids Res* 15, 3305-3315.
- Schuler, M., Connell, S. R., Lescoute, A., Giesebrecht, J., Dabrowski, M., Schroer, B., Mielke, T., Penczek, P. A., Westhof, E. & Spahn, C. M. (2006). Structure of the ribosome-bound cricket paralysis virus IRES RNA. *Nat Struct Mol Biol* 13, 1092-1096.
- Simpson, P. J., Monie, T. P., Szendroi, A., Davydova, N., Tyzack, J. K., Conte, M. R., Read, C. M., Cary, P. D., Svergun, D. I., Konarev, P. V., Curry, S. & Matthews, S. (2004). Structure and RNA interactions of the N-terminal RRM domains of PTB. *Structure* 12, 1631-1643.
- Siridechadilok, B., Fraser, C., Hall, R., Doudna, J. & Nogales, E. (2005). Structural roles for human translation factor eIF3 in initiation of protein synthesis. *Science* 310, 1513-1515.
- Song, Y., Tzima, E., Ochs, K., Bassili, G., Trusheim, H., Linder, M., Preissner, K. T. & Niepmann, M. (2005). Evidence for an RNA chaperone function of polypyrimidine tract-binding protein in picornavirus translation. *RNA* 11, 1809-1824.
- Spahn, C., Kieft, J., Grassucci, R., Penczek, P., Zhou, K., Doudna, J. & Frank, J. (2001). Hepatitis C virus IRES RNA-induced changes in the conformation of the 40S ribosomal subunit. *Science* 291, 1959-1962.
- Spahn, C. M., Jan, E., Mulder, A., Grassucci, R. A., Sarnow, P. & Frank, J. (2004). Cryo-EM visualization of a viral internal ribosome entry site bound to human ribosomes: the IRES functions as an RNA-based translation factor. *Cell* 118, 465-475.
- Stassinopoulos, I. A. & Belsham, G. J. (2001). A novel protein-RNA binding assay: functional interactions of the foot-and-mouth disease virus internal ribosome entry site with cellular proteins. *RNA* 7, 114-122.
- Svitkin, Y. V., Pause, A., Haghighat, A., Pyronnet, S., Witherell, G., Belsham, G. J. & Sonenberg, N. (2001). The requirement for eukaryotic initiation factor 4A

(eIF4A) in translation is in direct proportion to the degree of mRNA 5' secondary structure. *RNA* 7, 382-394.

- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22, 4673-4680.
- Tolan, D. R. & Traut, R. R. (1981). Protein topography of the 40 S ribosomal subunit from rabbit reticulocytes shown by cross-linking with 2-iminothiolane. *J Biol Chem* 256, 10129-10136.
- Uchiumi, T., Terao, K. & Ogata, K. (1981). Identification of neighboring protein pairs cross-linked with dimethyl 3,3'-dithiobispropionimidate in rat liver 40S ribosomal subunits. *J Biochem* 90, 185-193.
- Wimmer, E., Hellen, C. U. & Cao, X. (1993). Genetics of poliovirus. *Annu Rev Genet* 27, 353-436.
- Woese, C. R., Magrum, L. J., Gupta, R., Siegel, R. B., Stahl, D. A., Kop, J., Crawford, N., Brosius, J., Gutell, R., Hogan, J. J. & Noller, H. F. (1980). Secondary structure model for bacterial 16S ribosomal RNA: phylogenetic, enzymatic and chemical evidence. *Nucleic Acids Res* 8, 2275-2293.
- Zhao, Q., Han, Q., Kissinger, C., Hermann, T. & Thompson, P. (2008). Structure of hepatitis C virus IRES subdomain IIa. *Acta Crystallogr D Biol Crystallogr* 64, 436-443.
- Zhou, C., Arslan, F., Wee, S., Krishnan, S., Ivanov, A. R., Oliva, A., Leatherwood, J. & Wolf, D. A. (2005). PCI proteins eIF3e and eIF3m define distinct translation initiation factor 3 complexes. *BMC Biol* 3, 14.
- Zhou, M., Sandercock, A., Fraser, C., Ridlova, G., Stephens, E., Schenauer, M., Yokoi-Fong, T., Barsky, D., Leary, J., Hershey, J., Doudna, J. & Robinson, C. (2008). Mass spectrometry reveals modularity and a complete subunit interaction map of the eukaryotic translation factor eIF3. *Proc Natl Acad Sci U S A* 105, 18139-18144.
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31, 3406-3415.
- Zwieb, C., Wower, I. & Wower, J. (1999). Comparative sequence analysis of tmRNA. *Nucleic Acids Res* 27, 2063-2071.

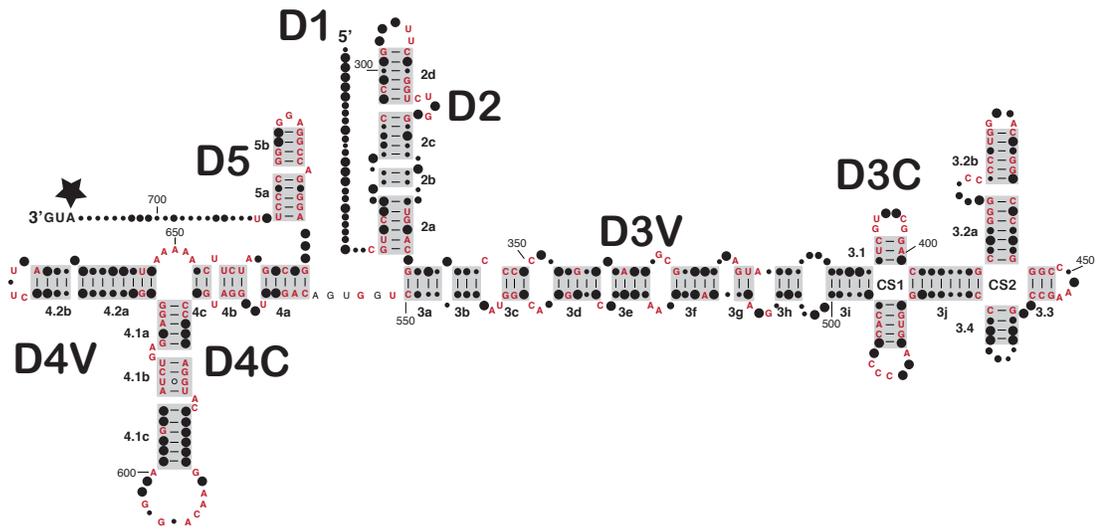


Figure 4.1: Secondary structure diagram of FMDV O1K IRES RNA supported by CSA and biochemical data. Invariant residues are shown in red. Conservation is indicated with large dots representing 90% or greater conservation, medium size dots 75-89% and small dots 74% or less. The 5' and 3' ends are shown, and positions are numbered every ten residues. Domains are numbered D1-D5. Conserved and variable subdomains are denoted C and V, respectively. Helices are numbered in the 5' to 3' direction and helical sections are indicated with letters. Helical insertions are named with a period followed by a number. The first utilized start codon is indicated in bold and with a star.

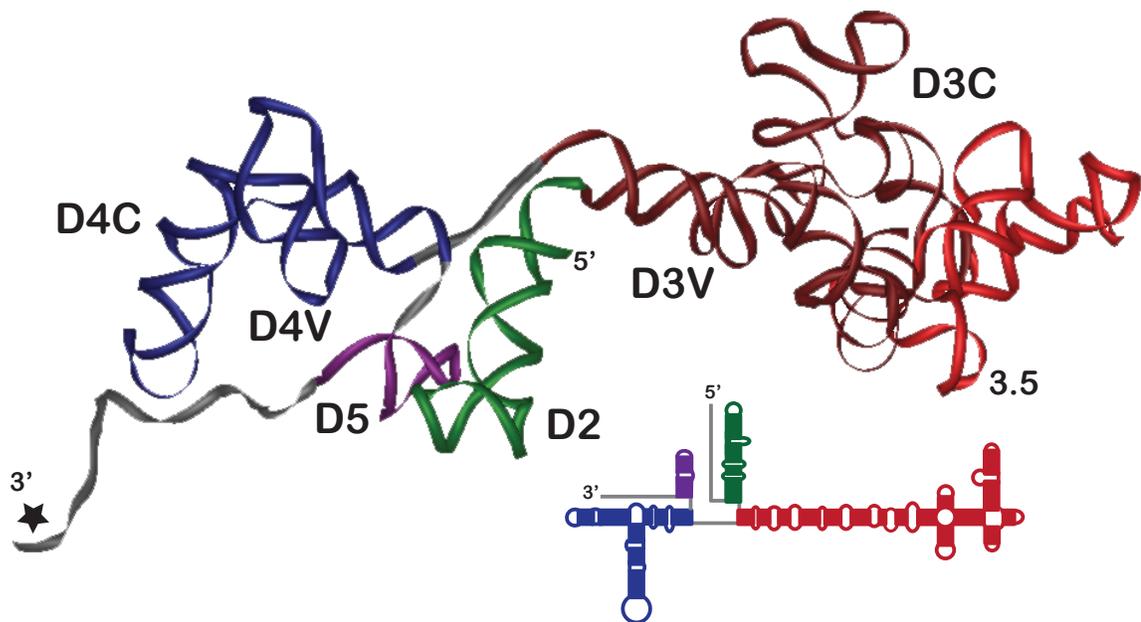


Figure 4.2: Three-dimensional model of FMDV O1K IRES RNA. Domains are colored as in the secondary structure schematic. The 5' and 3' ends and domains are indicated, as are the conserved and variable subdomains in D3 and D4. The location of the formed helix 3.5 is indicated in the bright red structure of D3C, and the structure resulting from the melting of helix 3.5 is shown in dark red. The start codon is marked with a star. The figure was generated using iMol (<http://www.pirx.com/iMol/overview.shtml>).

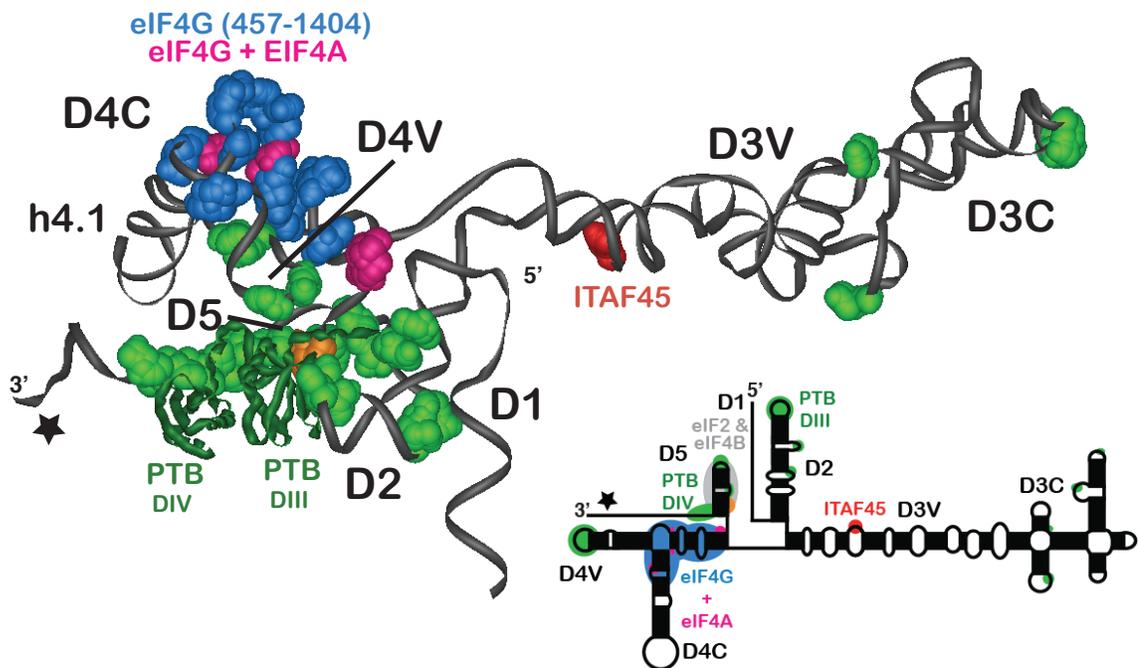


Figure 4.3: Protein protections in FMDV. Residues protected from modification by PTB (green spheres), ITAF45 (red), the middle domain of eIF4G alone (blue) or eIF4G-eIF4A or eIF4F complexes (pink) as listed in Table 4.1. Orange spheres represent protections of residues 419-421 by either PTB, eIF4G:eIF4A or eIF4F complexes. PTB domains three and four are shown. The start codon is indicated in with a star. The approximate location of the binding sites for eIF2 and eIF4B are shown in gray on the secondary structure. The figure was generated using iMol (<http://www.pirx.com/iMol/overview.shtml>).

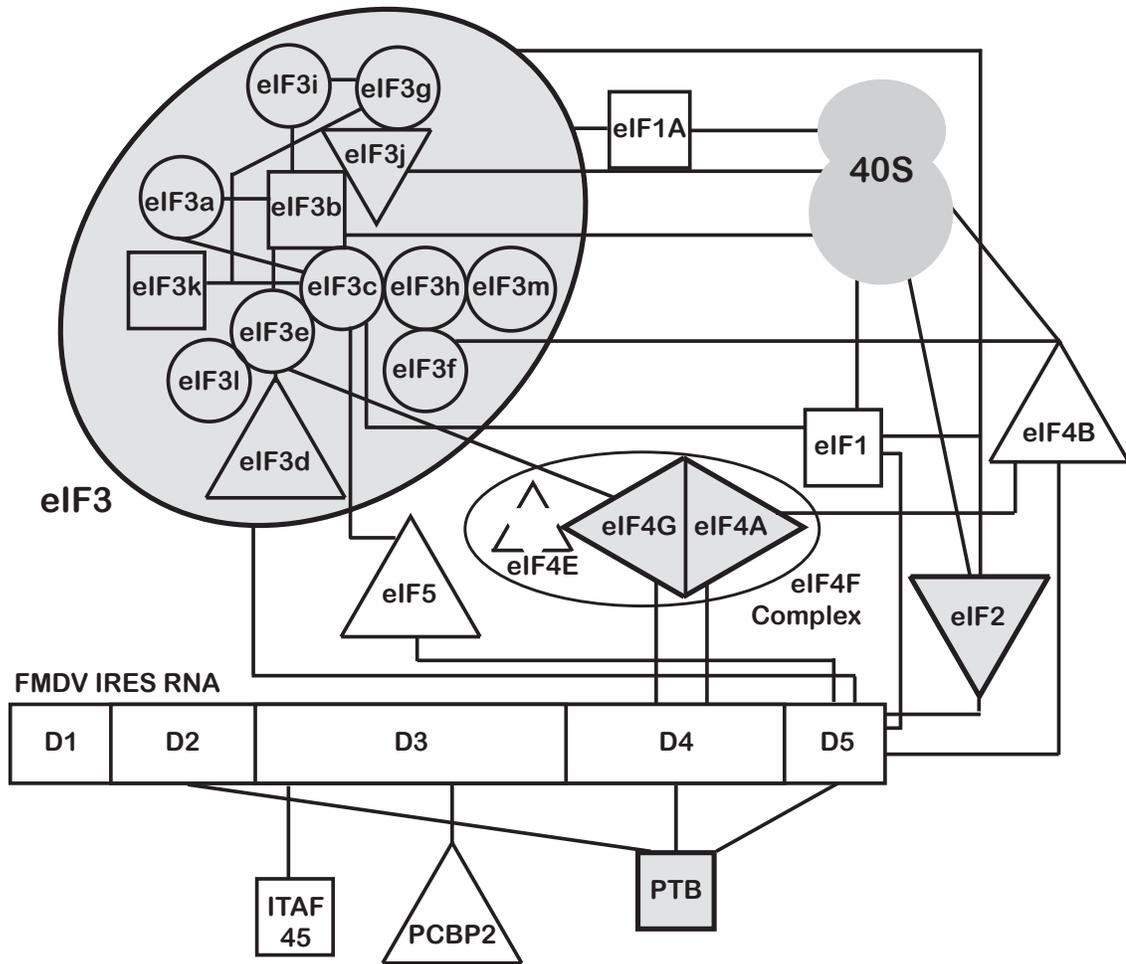


Figure 4.4: Map of interactions between FMDV IRES RNA domains, initiation factors and the 40S ribosomal subunit. The FMDV IRES RNA domains are indicated. The minimal set of proteins required for 48S complex formation on FMDV IRES RNA is shown in gray. The 40S ribosomal subunit is represented by the gray snowman. Proteins for which complete or partial high-resolution structural data are available are shown as squares or triangles, respectively; otherwise, they are indicated as circles. Interactions are shown as lines or close placements.

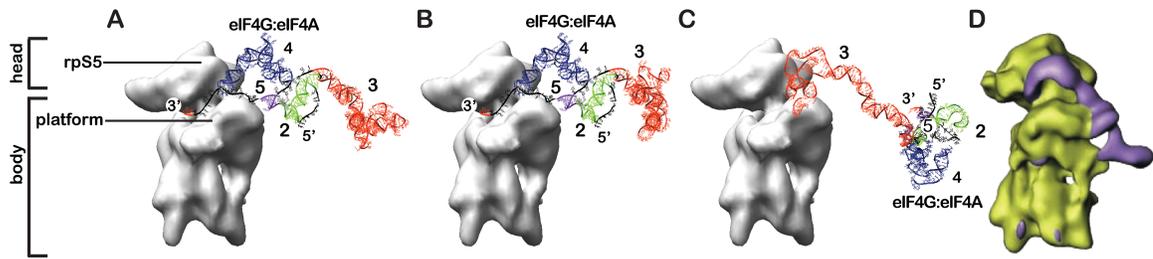


Figure 4.5: Placement of FMDV IRES RNA on the cryo-EM surface representation of the human 40S subunit. **A)** Model with intact helix 3.5. **B)** Model with melted helix 3.5. **C)** Model in B with proposed D3 tRNA-like region near the E site showing bulk of model far away from 40S subunit toward solvent due to length of D3 stack. **D)** Cryo-EM representation of HCV IRES RNA bound to rabbit reticulocyte 40S ribosomal subunit for reference (from Spahn *et al.*, 2001. Reprinted with permission from AAAS.). The coloring of the FMDV IRES RNA models in A-C is as in Figure 4.2. The first start codon (red) is indicated. Figure 4.5 was generated using UCSF Chimera and the electron density map of the naked human 40S subunit (EMDB ID 1092; Spahn *et al.*, 2004) retrieved from the Electron Microscopy Database (Pettersen *et al.*, 2004).

D	Feature	Model Res.	Experimental Res.	E	M	Proteins	R	
2	2d loop	53-UUU-55	305-UUU-307 (O1K)	p	C	PTB	2	
	2cd loop	63-UA-64	405-CU-406 (O1K)	p	P1	PTB	1	
	2cd loop	G65	G317 (O1K)	p	T1	PTB	4	
	2c	70G	413G (O1K)	p	P1	PTB	1	
	2bc loop	72A	415A (O1K)	p	P1	PTB	1	
	3	3cd loop	G99	G351 (O1K)	p	T1	ITAF45	4
		3d	G103	G351 (O1K)	e	T1	PTB	4
		3.1	G149	G492 (O1K)	p	P1	PTB	1
		3.2ab loop	U168	A511 (O1K)	p	P1	PTB	1
		3.2b loop	A179	A522 (O1K)	p	P1	PTB	1
3.5		A232	A575 (O1K)	p	P1	PTB	1	
3.5		G239	G491 (O1K)	e	C	ITAF45	4	
3.5		G239	G491 (O1K)	e	T1	PTB, ITAF45	4	
4		G301, G303	G553, G555 (O1K)	e	T1	PTB, ITAF45	4	
		U302	G554 (O1K)	e	C	ITAF45	4	
	4a	G308	C684 (EMCV R)	p	V1	eIF4G (457-1404)	3	
	4ab loop	311-AA-312	687-AA-688 (EMCV R)	p	D	eIF4G (457-1404)	3	
	4b	A314	A690 (EMCV R)	p	V1	eIF4G (457-1404)	3	
	4c	G317	G569 (O1K)	p	T1	eIF4G+eIF4A	4	
	4c	G317	G569 (O1K)	e	T1	PTB	4	
	4.1ab loop	G359	G611 (O1K)	p	T1	eIF4G+eIF4A	4	
	4.1ab loop	A360	A724 (EMCV R)	p	D	eIF4G (457-1404)	3	
	4.1a	G361	U725 (EMCV R)	p	C	eIF4G (457-1404)	3	
	4.2b loop	U379, U381	U631, U633 (O1K)	p	C	PTB	2	
	4.2ab loop	386-UU-387	758-UU-759 (EMCV R)	p	C	PTB	2	
	4.2a	391-CG-392	763-CG-764 (EMCV R)	p	V1	eIF4G (457-1404)	3	
	4.2a	G393	A765 (EMCV R)	e	V1	eIF4G (457-1404)	3	
	4.2a	U395	G767 (EMCV R)	p	V1	eIF4G (457-1404)	3	
	A-rich loop	A397	A770 (EMCV R)	p	D	eIF4G (457-1404)	3	
	A-rich loop	397-AA-398	648-AA-649 (O1K)	t	TP	eIF4F	4	
	A-rich loop	398-AAAA-401	771-AAAA-774 (EMCV R)	p	D	eIF4G (457-1404)	3	
	4c	G402	G653 (O1K)	p	T1	eIF4G+eIF4A	4	
	4c	C403	C776 (EMCV R)	p	V1	eIF4G (457-1404)	3	
	4bc loop	U404	G777 (EMCV R)	p	V1	eIF4G (457-1404)	3	
	4b	405-UC-406	778-UC-779 (EMCV R)	p	V1	eIF4G (457-1404)	3	
	4a	411-CC-412	783-GC-784 (EMCV R)	e	V1	eIF4G (457-1404)	3	
	4a	413-UG-414	664-UG-665 (O1K)	t	TP	eIF4G+eIF4A	4	
	4a	413-UG-414	664-UG-665 (O1K)	t	TP	eIF4F	4	
	4a	G414	G665 (O1K)	e	T1	PTB, ITAF45	4	
	5	5a	419-GGU-421	670-GGU-672 (O1K)	p	T1	PTB	4
		5a	419-GGU-421	670-GGU-672 (O1K)	p	T1	eIF4G+eIF4A	4
		5a	420-GU-421	671-GU-672 (O1K)	t	TP	eIF4F	4
		5a	421-UG-422	672-UG-673 (O1K)	e	C	ITAF45	4
		5a	G422	G764 (O1K)	p	P1	PTB	1
		5ab loop	A423	A765 (O1K)	p	P1	PTB	1
		5b	C424	C766 (O1K)	p	P1	PTB	1
5b		G427	G678 (O1K)	e	C	ITAF45	4	
5b		429-GG-430	680-GG-681 (O1K)	e	T1	PTB	4	
5b		430-GUCG-433	681-GCCG-664(O1K)	t	TP	PTB	4	
5b		G434	G685 (O1K)	e	C	ITAF45	4	
5b		U439	U690 (O1K)	t	TP	PTB	4	
5b		U439	U690 (O1K)	p	C	PTB	2	
3'		3' Poly(Y)	440-UU-441	691-UU-692(O1K)	p	C	PTB	2
		3' Poly(Y)	440-UU-441	691-UU-692(O1K)	t	TP	PTB	4
	3' Poly(Y)	442-CC-443	811-CC-812 (EMCV R)	p	V1	PTB	2	
	3' Poly(Y)	444-UUU-446	813-UUU-815 (EMCV R)	p	C	PTB	2	
	3' Poly(Y)	444-UUUA-447	695-UUUU-698 (O1K)	p	C	PTB	2	

Table 4.1: Summary of biochemical data used as modeling constraints. Columns 1 and 2 indicate domains and RNA secondary structure features as annotated in Figure 4.1. Residue positions in the model (IRES RNA from FMDV strain C-s8c1) are given in column 3. The fourth column details the corresponding residues in the referenced experiments. All strains are FMDV unless indicated. Protections (p) or enhanced accessibility (e) to chemicals or enzymes are shown in columns 5 and 6. Methods are chemical modification using CMCT (CMCT) or DMS (DMS), enzymatic cleavage reactions using RNases T1 (RNase T1) or RNase V1 (RNase V1), or toeprinting experiments (toeprint). Proteins and polypeptides with residue positions are shown in column 7. The plus sign indicates a protein-protein complex. References are 1, Luz & Beck, 1991; 2, Kolupaeva *et al.*, 1996; 3, Kolupaeva *et al.*, 1998; Pilipenko *et al.*, 2000.

Feature	SCOR class	PDB	Res.	Target Res.	Notes
1	Stacked duplex with two non-WC pairs	1N66	17-18 5-6	34-35 76-77	
2	Stacked duplex with two non-WC pairs	1MUV	A:4-5 B:4-5	39-39 72-73	
3	About 90 degree turn with short stacked bases	1P5M	11-15	62-66	
4a	Pentaloop with two bases in the main stack	1L8V	235-239	51-55	
4b		2ADC	540-541	53-54	1
5		ERNA-3D		84	
6	Stacked duplexes with one non-WC pair	1KP7	5, 26	89, 294	
7	Loops with interrupted stack	1FJG	228-229 133-135	93-94 288-290	
8	Loops with two extruded helical single strands	1MJI	35-37 45-47	98-100 282-284	
9	One stacked unpaired base flanked by non WC pair	1JJ2	1466 1476-1477	106 275-276	
10	About 90 degree turn with short stacked bases	1J5A	2306-2308 2365-2367	111-113 268-270	
11	Loops with interrupted stack	1JJ2	H:2610-2613 H:2545-2546	119-122 261-262	
12	Loops with base triples, no dinucleotide platform	2LDZ	F:24-25 E:6-9	124-125 256-259	
13	Stacked duplexes with four non-WC pairs	1NA2	F:20-23 E:7-10	129-132 249-252	
14		1KH6	4-5 49-50 8-10 21-23 24-26 35-37 39-43 44-48	136-137 243-244 138-140 147-149 150-152 224-226 227-231 239-242	2
15	Hexaloop with six bases in the main stack	1E8O	111-116	141-146	
16	Heptaloop	1NKW	O:2795-2801	232-238	
17		1KH6	8-10 21-23 4-7 49-50 39-43 44-48 24-26 35-37	207-209 216-218 203-206 194-195 158-162 189-193 219-221 155-157	3
18		ERNA-3D		165-170	
19	GNRA tetraloop	1KH6	29-32	177-180	
20	Pentaloop	2NOQ	116-120	197-202	

21	Tetraloop	2NOQ	85-88	211-214	4
22		ERNA-3D		299-304	
23	Loops with an extruded helical single strand	1LNG	175-177 153-154	310-312 408-409	
24	Stacked duplexes with one non-WC pair	1J7T	41, 5	316, 404	
25		ERNA-3D		397-401	
26	Several looped out bases	1J5E	129-130	359-360	
27	Loops with an extruded helical single strand	1JJ2	1287-1288	328-329	
28	Tridecaloop: Kissing hairpin thirteen base loop	1BAU	6-18	336-348	
29	One looped-out G	1J5E	31	388	
30	Hexaloop with one base in 5' stack, one in 3' stack, one in both stacks	1FQZ	12-17	378-383	
31		ERNA-3D		415-417	
32	One looped-out G	1J5E	31	423	
33	Triloops with three looped-out purines	1J5A	1186-1188	428-430	
34		2ADC	533-537	442-446	

¹After feature 4a was modeled, residues 540 and 541 of 2ADC were superimposed onto residues 53 and 54 while retaining the connections between 52/53 and 54/55. ²Because the source HCV feature has three extra residues in the juncture (A5, A6, U38), a slight modification was made to connect residues 137 and 138, and 226 and 227. ³A slight modification to this feature was made so residues 157 and 158 were connected, as HCV has extra residues in the juncture. ⁴The remainder of the adjacent helical section was copied as part of feature 7.

Table 4.2: Features used to model the FMDV IRES RNA. Structural features and their SCOR classifications (Klosterman *et al.*, 2002) are shown on the left. The sources of the coordinates (PDB or ERNA-3D) for a given feature are given in the PDB column. The residues positions involved in the feature are indicated in the Target Res. column. Coordinates for polypyrimidine-tract binding protein (PTB) domains three and four along with their bound RNA hexamers were from 2ADC.pdb.

CHAPTER 5: COMPARATIVE STRUCTURAL STUDIES OF BOVINE VIRAL
DIARRHEA VIRUS IRES RNA

Burks, J., Zwieb, C., Müller, F., Wower, I. K. & Wower, J. Comparative Structural
Studies of Bovine Viral Diarrhea Virus IRES RNA.

ABSTRACT

The internal ribosomal entry site (IRES) RNA of Bovine Viral Diarrhea Virus (BVDV) has been implicated in virus propagation. To gain insight into the structure and potential function of the BVDV IRES RNA we collected and aligned 663 of its sequences. The majority of sequences belonged to either genotype 1 or 2, but a third previously unidentified group was distinctly different. Compensatory Watson-Crick and wobble G-U pairs were investigated to establish phylogenetically supported secondary structures for each of the BVDV IRES RNA sequences. Conservation levels varied between 49 and 80 percent overall. Two highly variable regions corresponded to residues 209-220 and 298-307 in BVDV-1b Osloss. The extensively folded BVDV IRES RNAs were composed of helices 2, 3 and 4. Helix 2 consisted of five helical sections. Helix 3 contained sections 3a to 3j as well as six helical insertions 3.1 to 3.6. Sections 3a and 3b together with helices 3.6 and 4 formed an RNA pseudoknot. Three-dimensional modeling of the BVDV-1b Osloss IRES RNA showed it to be elongated with approximate dimensions of 170 Å by 65 Å by 90 Å. The model could be placed on the 40S ribosomal subunit similar to how the HCV IRES RNA was arranged. The IRES RNA-ribosome complex predicted a proximity between helix 2 of the BVDV IRES and ribosomal proteins S5 and S25.

INTRODUCTION

Bovine viral diarrhea virus (BVDV) manifests itself as a disease with a wide spectrum of symptoms and causes major economic loss for bovine producers worldwide (Baker, 1995). BVDV is a positive single-stranded, uncapped RNA virus which belongs to the Pestivirus genus and the Flaviviridae family (Brock *et al.*, 1992; Collett *et al.*, 1988a; Collett *et al.*, 1988b; Lindenbach *et al.*, 2007; Renard *et al.*, 1987; Thiel *et al.*, 2005). Of the three major subgenotypes 1a, 1b and 2, 1b predominates in North America (Flores *et al.*, 2002; Fulton *et al.*, 2003; Fulton *et al.*, 2005; Fulton *et al.*, 1997; Fulton *et al.*, 2009; Pellerin *et al.*, 1994; Ridpath and Bolin, 1998; Ridpath *et al.*, 1994; Ridpath *et al.*, 2000).

The approximately 12.5 kb BVDV RNAs contain at their 5' end an untranslated region (5' UTR) with an internal ribosomal entry site (IRES) (Brown *et al.*, 1992; Collett *et al.*, 1988b). This site initiates translation of a single polyprotein in concert with eukaryotic initiation factor eIF2, eIF3, the ribosome, initiator transfer RNAs, and GTP (Chon *et al.*, 1998; Myers *et al.*, 2001; Pestova and Hellen, 1999; Poole *et al.*, 1995; Purchio *et al.*, 1984). Although the role of the IRES in replication of BVDV is still poorly understood, the importance of this region for the propagation of related viruses, such as Hepatitis C Virus (HCV) and chimeric HCV-Poliioviruses, has been clearly demonstrated (Friebe *et al.*, 2001). Disrupting the IRES RNA structure or preventing the binding of translation factors to the IRES RNAs of HCV and Foot and mouth disease virus (FMDV) abolished viral replication and disease (Buratti *et al.*, 1997; reviewed by Dasgupta *et al.*, 2004; Pawlotsky *et al.*, 2007). Similar strategies for interfering with BVDV propagation require a detailed knowledge of the structure of the BVDV IRES RNA.

BVDV strains NADL (Collett *et al.*, 1988b), SD-1 (Deng and Brock, 1992) (subgenotypes 1a), Osloss (De Moerlooze *et al.*, 1993) (subgenotype 1b), and BVDV-2 isolate 890 (Ridpath and Bolin, 1995) possess IRES RNAs composed of about 310 nucleotides. Each is preceded by an AUG start codon located approximately 75 nucleotides from the 5' end of the viral RNA. Sequence comparisons and calculations of RNA secondary structure using minimal energy parameters suggested two main domains, II and III (Brown *et al.*, 1992; Deng and Brock, 1993). Site-directed mutagenesis of certain sites abolished the translation of reporter protein indicating that they were important for the propagation of BVDV (Chon *et al.*, 1998; Poole *et al.*, 1995). The presence of an RNA pseudoknot was supported by compensatory mutations which restored translation (Moes and Wirth, 2007).

No high-resolution structural data are currently available for BVDV IRES RNA. A partial structure of domain II from the related pestivirus Classical Swine Fever Virus (CSFV) IRES was obtained in solution (Locker *et al.*, 2007). Structures of IRES RNA domain II of the more distantly related hepacivirus Hepatitis C Virus (HCV) have been solved by NMR (Lukavsky *et al.*, 2003) and X-Ray crystallography (Dibrov *et al.*, 2007; Zhao *et al.*, 2008), as were the structures of several stem loops of HCV domain III (Collier *et al.*, 2002; Klinck *et al.*, 2000; Lukavsky *et al.*, 2000; Rijnbrand *et al.*, 2004). Also, high-resolution data are available for the junction of four HCV IRES RNA helices (Kieft *et al.*, 2002). Cryo-EM structures of the HCV IRES RNA are known both on and off the rabbit reticulocyte 40S ribosomal subunit (Spahn *et al.*, 2001) or bound to the human 80S ribosome (Boehringer *et al.*, 2005). Conformational changes in domain II of

the HCV IRES RNA upon the binding of the antisense inhibitor Isis-11 have been revealed recently (Paulsen *et al.*, 2010).

In the presented work we assembled BVDV IRES RNA sequences from GenBank and Rfam (Benson *et al.*, 2009; Gardner *et al.*, 2009; Griffiths-Jones *et al.*, 2003) and used stringent comparative sequence analysis to determine RNA secondary structure. Using phylogenetically supported base pairings, a three-dimensional model of the BVDV IRES RNA was constructed. Biophysical data from the related IRES RNAs were considered if no data was available for the predicted BVDV structure. The results provide a solid foundation for the identification of targets and therapeutic agents which may interfere with BVDV survival.

RESULTS AND DISCUSSION

Identification of BVDV IRES RNA Sequences

BVDV IRES RNA sequences were extracted from the Pestivirus IRES RNA alignment (IRES_Pesti, ID RF00209) deposited in the Rfam database (Gardner *et al.*, 2009; Griffiths-Jones *et al.*, 2003). GenBank was searched with a representative subset of the obtained seed sequences, and keywords such as "BVDV" and "Bovine viral diarrhea" were used to find additional IRES RNA candidates. All sequences were merged into a single preliminary alignment file for examination as described in Material and Methods. False positive and redundant sequences were removed and the longest sequence of an otherwise identical source sequence was retained. The sequences were arranged according to genotype and subgenotype by inspecting the GenBank records and their associated literature sources. Highly similar sequences for which a subgenotype

assignment was unavailable were placed in close proximity to sequences of known type. If that was not possible the closest available match was made. Each entry was given a unique identifier composed of "BVDV" followed by the genotype or subgenotype assignment (if known) and the GenBank accession number.

Upon completion of the analysis, 663 unique BVDV IRES RNA sequences remained in the collection. The majority were unclassified subtypes with 559 sequences belonging to genotype 1 and 59 sequences being of genotype 2. The remaining 45 sequences were distributed among the 15 different subgenotypes of genotype 1. Many sequences lacked residues near the 5'-end of the viral RNA. The full lengths of the 16 unique BVDV genomes available in GenBank suggested that the size of the IRES RNAs varied between 313 (Accession EF101530) and 320 (AF145967) residues. The BVDV-1b strain Osloss IRES RNA (M96687) contained 314 nucleotides at positions 72 to 386 of its genome and served as a reference.

Comparative Sequence Analysis (CSA) of BVDV IRES RNAs

To improve the quality of the alignment, structural features containing phylogenetically supported base pairs were identified in the sequence similarity-based alignment of BVDV IRES RNAs. Adjustments to the alignment were made by identifying covarying compensatory base changes (CBCs) with the help of the SARSE alignment editor and its associated tools described in Materials and Methods. Nucleotide differences between the sequences were examined for changes which maintained Watson-Crick (A-U, G-C) and wobble G-U pairs. CBCs supported the existence of a base pair, while mismatches provided evidence against its existence (Larsen and Zwieb, 1991).

Invariant residues could not be used to prove or disprove a pair as it was impossible to observe a CBC or mismatch. A pair was considered supported if there was at least twice as much positive evidence for its existence as there was negative evidence. Unless clearly disproven, potential Watson-Crick and G-U pairs were included if they were positioned next to a supported pair. The adjusted alignment was annotated with a pairing mask to encode all identified base pairs. It is available in FASTA format in Supplementary Materials and is also accessible at <http://rnp.uthscsa.edu/IRES/BVDVIRESRNA.zip>.

Inspecting the identity thresholds of the alignment with Jalview (Material and Methods) distinguished three major groups of sequences. The largest group contained 575 genotype 1 sequences (GenBank Accessions M31182 to AF268278 in the alignment). The second group (74 sequences, AB003619 to U65055) was composed of a mixture of sequences which were annotated as belonging to either genotype 1 or 2. A third group of 10 BVDV-2 sequences (AB019150 to AB019174) suggested that some of the reported serotypical or genotypical assignments may have to be reexamined in the future.

The level of sequence conservation across the phylogenetically supported alignment varied between 49 and 80 percent. Two variable regions corresponded to residues 209-220 and 298-307 in the BVDV-1b Osloss IRES RNA (Figure 5.1). The 3' portion of the alignment possessed seven invariant residues corresponding to A266, G278, A292, A318, C323, G325, and A328 of BVDV-1b Osloss. Analysis of the terminal regions of the 28 full-length sequences suggested a conserved 5' region (residues 73-126 in BVDV-1b Osloss) and a nearly invariant 3' region immediately preceding the AUG start codon (residues 365-386). Conserved alignment positions presumably

represented sites which were directly or indirectly responsible for virus propagation by maintaining the structural integrity of the BVDV RNAs or by mediating the binding of IRES RNA-associated proteins and translation factors.

BVDV IRES RNA Secondary Structure

The alignment pairing mask made it possible to extract the secondary structure of each sequence. Because of its disease-causing economical impact, we used the Osloss strain for reference. The BVDV IRES RNAs were extensively folded and composed of three helices named helix 2, 3 and 4 (Figure 5.1). Two helical sections in the 5' UTR, annotated as 1a and 1b, were not displayed in the secondary structure diagram as they were located approximately ten residues upstream of the IRES and had been shown to be dispensable for viral replication (Chon *et al.*, 1998). Helices 3 and 4 participated in the formation of a previously proposed pseudoknot (Moes and Wirth, 2007). Published experimental data were considered as described below, including results from site-directed mutagenesis experiments of the BVDV (Moes and Wirth, 2007) and closely related CSFV IRES RNA (Fletcher and Jackson, 2002; Kolupaeva *et al.*, 2000b; Pestova *et al.*, 1998; Rijnbrand *et al.*, 1997) as well as the high-resolution structures of homologous IRES RNA regions (Collier *et al.*, 2002; Dibrov *et al.*, 2007; Kieft *et al.*, 2002; Lukavsky *et al.*, 2003; Lukavsky *et al.*, 2000; Zhao *et al.*, 2008).

Helix 2

The residues which participated in the formation of helix 2 were represented in only 34 of the aligned BVDV IRES RNA sequences and appeared to be conserved. Due to these limitations, we were unable to extract a sufficient number of CBCs to reliably predict secondary structure and thus refrained from indicating the conservation levels of helix 2 (Figure 5.1). Chemical or enzymatic modification and site-directed mutagenesis data were unavailable for helix 2 of BVDV or the closely related CSFV. Thus, the base pairings were deduced from the homologous structures of the HCV and CSFV IRES RNAs (Figure 5.2). Despite these constraints, predictions could be made with regards to the secondary structure and composition of the related BVDV IRES RNA region.

Helix 2 consisted of five helical sections, annotated as sections 2a to 2e and separated by internal loops of up to six residues (Locker *et al.*, 2007; Lukavsky *et al.*, 2003). Section 2a was relatively variable and contained six to eight base pairs. Its 5' portion was exclusively composed of pyrimidine and its 3' portion of purine residues. Sections 2a and 2b were connected at their 5' side without interruption or by one unpaired residue which was typically an adenine. The 3'-portion of internal loop was characterized by four to six nucleotides which were predominantly adenine residues. This loop contained one more residue than the homologous structural element of the CSFV IRES RNA (Locker *et al.*, 2007).

The four base pairs of section 2b were invariant and included a non-canonical A-G pair which was also present in the high-resolution structure of the homologous region of the CSFV IRES RNA (Locker *et al.*, 2007). In many sequences, 2b connected to section 2c with one adenine residue (A87; Figure 5.1). The invariant helical section 2c

contained four canonical and two non-Watson-Crick A-G pairs. It connected to 2d with a guanine (G116). Section 2d was typically composed of four base pairs with purines in the 5' and pyrimidines in its 3' portion. An additional A-G pair was observed in a BVDV type 2 sequence (GenBank Accession FJ493479). In BVDV.2___.AF145967 an uracil residue was inserted between the nucleotides at positions U114 and U115 (Figure 5.1). Section 2d made its 5' connection to helical section 2e via a conserved adenine (A98) which appeared to be characteristic of the BVDV IRES RNAs because it was absent in the CSFV and HCV IRES RNA sequences (Figure 5.2). Helical section 2e was composed of three invariant Watson-Crick G-C pairs. The sequence of the terminal loop of helix 2 was also highly conserved even among BVDV, CSFV and HCV. The residues corresponding to positions 104 and 108 in BVDV were conserved as purines and pyrimidines, respectively. There was no evidence for an A-U pair (A107 with U102, Figure 5.1) which had been implicated previously (Lukavsky *et al.*, 2003). The 31 sequences with information about the region between helices 2 and 3 suggested direct connections, although in one sequence (AF145967) an adenine was inserted between BVDV-1b Osloss positions 138 and 139 (Figure 5.1).

Helix 3

Helix 3 encompassed the largest portion of the BVDV IRES RNA and displayed an intricate arrangement of 10 helical sections (labeled a to j) and six helical insertions, annotated as 3.1 to 3.6 according to naming conventions used in previous nomenclature proposals (Burke *et al.*, 1987; Zwieb *et al.*, 2005).

Section 3a consisted of five to seven base pairs and was well supported by CBCs. This section was connected to 3b directly or separated by one residue in the 5' portion and up to four nucleotides in the 3' portion thus forming an asymmetric loop in some of the secondary structures. Covariation analysis supported the five or six base pairs in section 3b. The insertion of a uracil into section 3b was observed in the BVDV-1 isolate PT810 (GenBank Accession Z79766), a cytosine was inserted in the sequences with the accession GQ985459 and an adenine in EU051825. One cytosine residue was lacking in AF104019, AF356505, AY671985 and AY671986. The base pairings in helical sections 3a and 3b were supported by site-directed mutagenesis designed to compensate for function in the BVDV IRES RNAs (Moes and Wirth, 2007) and CSFV (Fletcher and Jackson, 2002) (See Table 5.1). In agreement with the proposed pairing of U348 in the BVDV IRES RNAs, the equivalent U335 in the CSFV IRES RNA was cleaved by the double-strand specific RNase V1 (Kolupaeva *et al.*, 2000b; Table 5.2). Similarly, the single-stranded nature of the residues between sections 3a and 3b conformed with the finding that the residue corresponding to BVDV A351 (CSFV G337) was accessible to RNase T1 (Kolupaeva *et al.*, 2000b; Table 5.2). Sections 3b and 3c connected typically without interruption or were separated by one nucleotide in seven of the BVDV IRES RNA sequences.

Section 3c typically contained two supported base pairs (U152 with A320 and G153 with C319, Figure 5.1), but only one base pair was observed in five of the sequences (Accessions U63479, AF220247, AY671980, AY944297, and AY273777). Sections 3c and 3d were separated by two adenines with two more residues added in two of the sequences (Accessions L20923 and L20926). Section 3d was composed of either

five or six supported pairs, but one or two mismatches disrupted this section in GQ985459, AY323878, AY671980, U65032, U65033, AY671985, AY273154 and AY273777. Typically, sections 3d and 3e connected via an invariant uracil, yet an additional residue was observed in four sequences (AY671980, AY763085, U65032, and U65033). Section 3e consisted of three to five base pairs which included an invariant G-C Watson-Crick pair at positions 163 and 263 (Figure 5.1). 3e connected without the insertion of a residue to section 3f in the 5' half, but contained one unpaired residue in the 3' half which tended to be an adenine.

Three of the four base pairs of section 3f were supported by CBCs with the fourth being an invariant G-C pair corresponding to G167 and C257 (Figure 5.1). This section was separated from helix 3.1 by two adenine residues in the majority the BVDV IRES RNA sequences. A cytosine was added in AB042705, AB042711, and AB042664, a uracil residue in AY159536, a guanine in L20918, and an adenine in AB042663. One adenine was substituted with a guanine in AY763045 and DQ973172. The AM749823 sequence contained only one adenine between 3f and 3.1.

Helical insertion 3.1 consisted of five covarying base pairs and an invariant 175C-G186 pair. This helix was capped by a highly conserved AGUA tetraloop. A GUC loop was observed instead in the sequence with the accession AM749823. Two unpaired residues, typically purines, separated helix 3.1 from section 3g.

The high level of conservation observed for 3g (Figure 5.1) provided an insufficient number of CBCs, but this section was included as being base paired because of the known coordinates of the homologous feature of the HCV IRES RNA (Kieft *et al.*, 2002). It joined section 3h via an asymmetric loop of one or two residues in the 5'

portion, and three to six residues in the 3' portion. The two central base pairs of section 3h were supported by CBCs and flanked by two highly conserved base pairs. Sections 3h and 3i were joined by an asymmetric loop containing an invariant uracil in the 5' portion and four to five residues in the 3' portion.

CSA supported the four base pair section 3i. It was separated from 3j by two conserved symmetrically arranged cytosine residues. Interestingly, 44 type 1 sequences connected instead via a CA/CC internal loop. The secondary structure of the sequence with the accession EU034170 displayed a similar CGC/UAC loop, but no such loop was observed in any of the type 2 sequences. Section 3j consisted of four or five phylogenetically supported base pairs and was capped by a loop of three to 10 residues. This variable loop corresponded to one of the two least conserved regions of the BVDV IRES RNA alignment (discussed above).

Helix 3.2 typically formed a direct connection with its preceding section 3g, but one residue was inserted in a small number of type 1 sequences (AF417989, AJ304390, EU034170, L20921, L20927, Z73248, and AF417988). Two residues were inserted in one sequence (EU034172). Being highly conserved, helix 3.2 was supported by a small number of CBCs. This helix had been observed also by NMR and crystallographic analyses of the homologous region in HCV IRES RNA (Kieft *et al.*, 2002; Rijnbrand *et al.*, 2004). A conserved adenine joined helix 3.2 and section 3f, and an additional residue was inserted in a four type 1 sequences (Accessions AF417989, L20918, L32880, and U65030). Section 3e and helix 3.3 were usually connected with two or three nucleotides (usually an UA), although in one case (EU034170) five nucleotides participated in the single strand.

Helix 3.3 consisted of four to nine CBC-supported base pairs and a highly conserved guanidine-rich loop with three invariant guanines at its center (Figure 5.1). Helices 3.3 and 3.4 were connected with three or four residues. Helix 3.4 was composed of a stem of four to seven Watson-Crick base pairs and a loop with three to seven residues. A direct connection between the 5' end of section 3c and 3' end of helix 3.5 was observed in most of the predicted secondary structures with exceptions seen in AY671980, AY944297 and AB019154 where one nucleotide (GQ985459) or two nucleotides (GQ985459) linked both helices. Helix 3.5 was highly conserved, contained two supported base pairs with one additional pair supported as an extension of the helix, and complied with the high-resolution structure of the homologous HCV IRES RNA regions (Lukavsky *et al.*, 2000). The helix was capped by a GAUA tetraloop. Helices 3.5 and 3.6 were joined directly, but one residue was inserted occasionally (Accessions AY273158, EU034174, EU034175, and AB019154).

Relying on covariations between C334 and G342, helix 3.6 contained two base pairs. Due to a large number of mismatches the potential pairings between U335 and either A340 or G341 were disproved and not included (Figure 5.1). Also disproved were the potential pairings between 341-GG-342 and 365-CC-366 (Moes & Wirth, 2007) due to extensive mismatches between these positions. Helix 3.6 joined section 3b directly with one exception in BVDV.2a___.AM749823 which contained a surprisingly large insertion of five residues. Helix 3.6 joined helix 4 via one or two nucleotides at the 3' end of helix 3 and up to one nucleotide at the 3' end of helix 4.

Helix 4

The 448 sequences which contained information about helix 4 showed that it was composed of four or five base pairs with the pair corresponding to 367U-A340 being well supported by several CBCs. Residues C337, A338, U367, C368 and C371 were invariant and did not provide support or disprove for pairing. The 339G-C368 pair displayed two compensatory changes and five mismatches and thus was not clearly disproved. According to the rule explained in Materials and Methods, the 339G-C368 pair was included because it was adjacent to the well-supported 367U-A340 pair.

The existence of helix 4 had been established experimentally by compensatory mutations in the IRES RNAs of BVDV (Moes and Wirth, 2007) and CSFV (Fletcher and Jackson, 2002; Kolupaeva *et al.*, 2000b; Pestova *et al.*, 1998; Rijnbrand *et al.*, 1997) (Table 5.1). Furthermore, the 336G-C371, 337C-G370 and 338A-U369 pairs of helix 4 were included because, unlike 341-GG-342 and 365-CC-366, they were not clearly disproved by our covariation analysis. The 3' portion of helix 4 connected to section 3a via a single-stranded region composed of four to 11 nucleotides that contained several adenine residues. Between helix 4 and the start codon were 12 essentially invariant residues which suggested that this region performed an important function. Most likely, upon binding of the IRES binds to the 40S ribosomal subunit, it engaged in the proper positioning of the start codon.

BVDV IRES RNA Pseudoknot

A pseudoknot forms when one or more nucleotides in a hairpin loop base pair with nucleotides outside of the loop. Pseudoknots have been known features of many

medium-size and large RNA molecules which are involved in translation and other cellular processes (Brierley *et al.*, 2008). Within the phylogenetically supported secondary structures of the BVDV IRES RNAs a pseudoknot engaged sections 3a and 3b together with helices 3.6 and 4 in much the same as had been proposed previously (Moes and Wirth, 2007). Using site-directed compensatory mutations as well as chemical and enzymatic probing, a similar pseudoknot was shown to exist also in the CSFV IRES RNA (Fletcher and Jackson, 2002; Kolupaeva *et al.*, 2000b; Pestova *et al.*, 1998; Rijnbrand *et al.*, 1997) and the more distantly related HCV IRES RNA (Berry *et al.*, 2010; Kieft *et al.*, 2001; Wang *et al.*, 1995).

A Three-Dimensional Model of the BVDV IRES RNA

The three-dimensional structure of a BVDV IRES RNA had not been determined experimentally. We used the phylogenetically supported secondary structure of the Osloss strain to generate a three-dimensional model and attempt to gain insights into the molecule's function and possible mechanism of BVDV IRES-mediated translation initiation. For structural comparisons, a three-dimensional model of the HCV-1b IRES RNA (Accession GU451224) was constructed using the available secondary structure (Spahn *et al.*, 2001).

Initial models were generated by entering the base pair information into ERNA-3D (Mueller *et al.*, 1995) to generate A-form RNA for the helical sections and calculate preliminary conformations of single-strands using the built-in algorithm (see Material and Methods). The helices were arranged in three-dimensions to form continuous chains. If applicable, high-resolution data from NMR or X-ray crystallography were incorporated

to generate models with biologically meaningful components (see Tables 5.3 and 5.4). The three-dimensional space was explored using stereo vision with knowledge of the available biochemical data, including those obtained from site-directed mutagenesis and enzymatic and chemical modification experiments (Fletcher and Jackson, 2002; Kolupaeva *et al.*, 2000b; Moes and Wirth, 2007; Pestova *et al.*, 1998; Rijnbrand *et al.*, 1997; Tables 5.1 and 5.2; Tables 5.3 and 5.4). The placements of sections in the BVDV and HCV IRES RNA models were further assisted by fitting to the cryo-EM surface map (PDB ID 2AGN; Boehringer *et al.*, 2005). The atomic pdb-formated coordinates of the models were viewed using iMol (<http://www.pirx.com/iMol>) as displayed in Figures 5.3A (BVDV-1b Osloss) and 5.3B (HCV-1b).

The BVDV IRES RNA model had an overall elongated shape with the dimensions of 170 Å by 65 Å by 90 Å, similar to the dimensions of the corresponding HCV model (170 Å by 85 Å by 90 Å). The BVDV pseudoknot permitted considerable movement due to the single-strands formed by U335, G341, 351-AU-352, and 359-366 (Figure 5.1). Recent molecular dynamics calculations suggested that such conformational changes occurred in the equivalent regions of the HCV IRES RNA (Lavender *et al.*, 2010). The four-way junction (3b, 3c, 3.5 and 3.6) of the BVDV IRES RNA was be stabilized by a 318A-U328 Watson-Crick pair which was supported by compensatory mutations in the HCV IRES RNA (Easton *et al.*, 2009). Two CBCs and two mismatches neither supported not disproved this interaction in BVDV, but it was included because of its feasibility in three dimensions and favorable constrain of the pseudoknot position.

Sections 3d, 3e, and helices 3.3 and 3.4 of BVDV formed a four-way junction at approximately the same location as a three-way junction in the HCV IRES RNA (Figure

5.3). Helix 3.3 of the CSFV and BVDV IRES RNAs appeared to correspond to helix 3.3 of the HCV IRES RNA and helix 3.3 of CSFV was likely involved in ribosomal binding (Jubin *et al.*, 2000; Kolupaeva *et al.*, 2000b). This helix was placed as suggested by cryo-EM (Boehringer *et al.*, 2005). In order to accommodate helix 2, the pseudoknot and the four-way junction, sections 3d and 3e were not coaxially stacked in contrast to recent suggestions (Ouellet *et al.*, 2010). Helix 3.4 was present only in the pestivirus IRES RNAs and absent in HCV (Brown *et al.*, 1992) but apparently did not interfere with the binding of eIF3 (Sizova *et al.*, 1998) or the 40S subunit (Kolupaeva *et al.*, 2000b).

BVDV IRES RNA on the 40S Ribosomal Subunit

Our three-dimensional BVDV IRES RNA model was based on stringent phylogenetic comparisons and several other constraints obtained from a wide variety of sources. We explored the model in relation to the surface of the 40S ribosomal subunit as described in Materials and Methods. Independently, we modeled the HCV IRES RNA and placed it onto the 40S ribosomal subunit in ways that were consistent with cryo-EM data for the initial IRES-40S binding stage (Spahn *et al.*, 2001). Because of the similar shapes of the BVDV and HCV IRES RNA models, both could be oriented as had been observed by cryo-EM of complexes with eIF3 (Siridechadilok *et al.*, 2005), 40S subunits (Spahn *et al.*, 2001) or 80S ribosomes (Boehringer *et al.*, 2005) (Figure 5.4).

Helix 2 was located at the “back” of the head of the 40S subunit as had been observed for the HCV IRES RNA (Spahn *et al.*, 2004; Tolan and Traut, 1981; Uchiumi *et al.*, 1981). The C83 residue in helix 2 of HCV IRES RNA was recently shown to cross-link to ribosomal proteins S14 and S16, and helix 3.3 nucleotides A275 and G263 were

shown to cross-link to ribosomal proteins S3a, S14 and S16 in the same study (Babaylova *et al.*, 2009). Due to similar positioning and structural constraints, these proteins would be expected to be close to BVDV helices 2 and 3.3, respectively. Proximity between HCV helix 2 and ribosomal proteins S5 and S25 (Landry *et al.*, 2009) and a cross-link to S5 (Fukushi *et al.*, 2001) would be expected to be features of the BVDV IRES RNA-ribosome complex. Helix 3 extended toward the solvent side despite the presence of the BVDV-specific helix 3.4. The juncture of helices 3.3, 3.4 and sections 3d and 3e was positioned near the body of the 40S subunit. Helix 3.3 was located near a contact region of eIF3 with 40S subunit (Siridechadilok *et al.*, 2005).

CONCLUSIONS

We identified significant differences between the BVDV and HCV IRES RNA secondary structures with respect to pseudoknot organization, the length of helix 2, and the presence of BVDV helix 3.4. Despite these distinctions, the models converged in three dimensions and adopted similar arrangements when placed onto the 40S ribosomal subunit. Therapeutic agents designed to interfere with IRES function would be expected to affect a wide range of viruses. However, with the detailed information provided here it may now be possible to consider interventions specifically directed at the IRES RNA of BVDV.

METHODS

Comparative Analysis of BVDV IRES RNA Sequences

BVDV IRES RNA sequences were extracted from the FASTA-formatted Rfam Pestivirus IRES alignment (version 9.1) located at <http://rfam.sanger.ac.uk/family/RF00209> (Gardner *et al.*, 2009). Additional sequences were identified using keywords and the Entrez search engine at <http://www.ncbi.nlm.nih.gov/sites/gquery> (Benson *et al.*, 2009). A NetBlast search was initiated with a representative subset of the data to obtain similar sequences (Altschul *et al.*, 2009; McGinnis and Madden, 2004). The data were examined using the BioEdit sequence alignment editor (Hall, 1999). Sequences were grouped by genotype and sequence similarity and preliminarily aligned using CLUSTAL (Higgins *et al.*, 1996). Duplicate entries were removed until each sequence in the alignment was unique. To prove or disprove Watson-Crick G-C, A-U, and G-U wobble base pairs, the alignment of the unique BVDV IRES RNA sequences was examined and adjusted with the SARSE editor (Andersen *et al.*, 2007) and externally linked programs of the RNAdbTools suite (Gorodkin *et al.*, 2001). Properties of the alignment, such as the extent of the conserved and variable regions were inspected using JalView (Waterhouse *et al.*, 2009). The alignment was made available for download at <http://rnp.uthscsa.edu/IRES/BVDVIRESRNA.zip>.

Molecular Modeling of IRES RNAs

The sequence and base pair information were entered into the ERNA-3D program (Mueller *et al.*, 1995) installed on a Windows PC workstation equipped with CrystalEyes

stereovision goggles and StereoGraphics infrared emitter. ERNA-3D created preliminary coordinates for A-form RNA in helical sections and calculated the initial conformations of single-stranded regions. RNA loops and other structural elements with similarity to known high-resolutions structures were identified in the PDB (Berman *et al.*, 2002) and in SCOR at <http://scor.berkeley.edu/> (Klosterman *et al.*, 2002) (Tables 5.3 and 5.4) followed by incorporating the coordinates into the models. Manual adjustments were made in ERNA-3D with consideration of the published data described in Results and Discussion. Tables listing the structural features are provided as Tables 5.3 and 5.4. The final atomic coordinates of the BVDV (bvdv1b_osloss_ires_model.pdb) and HCV (hcv1b_ires_model.pdb) IRES RNA models are accessible at <http://rnp.uthscsa.edu/IRES/BVDVIRESRNA.zip>. The IRES RNA models were placed onto the cryo-EM surface structure of the human 40S ribosomal subunit (Spahn *et al.*, 2004; EMDB ID 1092) obtained at http://emsearch.rutgers.edu/atlas/1092_summary.html (Heymann *et al.*, 2005) using the discussed constraints.

ACKNOWLEDGEMENTS

This research was supported by grants from the Alabama Agricultural Experiment Station Foundation and the Alabama Cattlemen's Association to I.K.W. and J.W. and an Auburn University Biogrant to J.W. Publication costs were supported in part by the Upchurch Fund for Excellence. J.M.B. was supported by the National Science Foundation under Grant No. 0091853 and NSF-EPS 0447675. The molecular graphics image in Figure 5.4 was produced using the UCSF Chimera package from the Resource

for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081).

REFERENCES

- Altschul, S. F., Gertz, E. M., Agarwala, R., Schaffer, A. A. & Yu, Y. K. (2009). PSI-BLAST pseudocounts and the minimum description length principle. *Nucleic Acids Res* 37, 815-824.
- Andersen, E. S., Lind-Thomsen, A., Knudsen, B., Kristensen, S. E., Havgaard, J. H., Torarinsson, E., Larsen, N., Zwieb, C., Sestoft, P., Kjems, J. & Gorodkin, J. (2007). Semiautomated improvement of RNA alignments. *RNA* 13, 1850-1859.
- Babaylova, E., Graifer, D., Malygin, A., Stahl, J., Shatsky, I. & Karpova, G. (2009). Positioning of subdomain III_d and apical loop of domain II of the hepatitis C IRES on the human 40S ribosome. *Nucleic Acids Res* 37, 1141-1151.
- Baker, J. C. (1995). The clinical manifestations of bovine viral diarrhea infection. *Vet Clin North Am Food Anim Pract* 11, 425-445.
- Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Sayers, E. W. (2009). GenBank. *Nucleic Acids Res* 38, D46-51.
- Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., Feng, Z., Gilliland, G. L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J. D. & Zardecki, C. (2002). The Protein Data Bank. *Acta Crystallogr D Biol Crystallogr* 58, 899-907.
- Berry, K. E., Waghray, S. & Doudna, J. A. (2010). The HCV IRES pseudoknot positions the initiation codon on the 40S ribosomal subunit. *RNA* 16, 1559-1569.
- Boehringer, D., Thermann, R., Ostareck-Lederer, A., Lewis, J. D. & Stark, H. (2005). Structure of the hepatitis C virus IRES bound to the human 80S ribosome: remodeling of the HCV IRES. *Structure* 13, 1695-1706.
- Brierley, I., Gilbert, R. J. C. & Pennell, S. (2008). RNA pseudoknots and the regulation of protein synthesis. *Biochemical Society Transactions* 36, 684-689.
- Brock, K. V., Deng, R. & Riblet, S. M. (1992). Nucleotide sequencing of 5' and 3' termini of bovine viral diarrhea virus by RNA ligation and PCR. *J Virol Methods* 38, 39-46.
- Brown, E. A., Zhang, H., Ping, L. H. & Lemon, S. M. (1992). Secondary structure of the 5' nontranslated regions of hepatitis C virus and pestivirus genomic RNAs. *Nucleic Acids Res* 20, 5041-5045.
- Buratti, E., Gerotto, M., Pontisso, P., Alberti, A., Tisminetzky, S. G. & Baralle, F. E. (1997). In vivo translational efficiency of different hepatitis C virus 5'-UTRs. *FEBS Lett* 411, 275-280.

- Burke, J. M., Belfort, M., Cech, T. R., Davies, R. W., Schweyen, R. J., Shub, D. A., Szostak, J. W. & Tabak, H. F. (1987). Structural conventions for group I introns. *Nucleic Acids Res* 15, 7217-7221.
- Chon, S. K., Perez, D. R. & Donis, R. O. (1998). Genetic analysis of the internal ribosome entry segment of bovine viral diarrhoea virus. *Virology* 251, 370-382.
- Collett, M. S., Anderson, D. K. & Retzel, E. (1988a). Comparisons of the pestivirus bovine viral diarrhoea virus with members of the flaviviridae. *J Gen Virol* 69 (Pt 10), 2637-2643.
- Collett, M. S., Larson, R., Gold, C., Strick, D., Anderson, D. K. & Purchio, A. F. (1988b). Molecular cloning and nucleotide sequence of the pestivirus bovine viral diarrhoea virus. *Virology* 165, 191-199.
- Collier, A. J., Gallego, J., Klinck, R., Cole, P. T., Harris, S. J., Harrison, G. P., Aboul-Ela, F., Varani, G. & Walker, S. (2002). A conserved RNA structure within the HCV IRES eIF3-binding site. *Nat Struct Biol* 9, 375-380.
- Dasgupta, A., Das, S., Izumi, R., Venkatesan, A. & Barat, B. (2004). Targeting internal ribosome entry site (IRES)-mediated translation to block hepatitis C and other RNA viruses. *FEMS Microbiol Lett* 234, 189-199.
- De Moerlooze, L., Lecomte, C., Brown-Shimmer, S., Schmetz, D., Guiot, C., Vandenberg, D., Allaer, D., Rossius, M., Chappuis, G., Dina, D. & *et al.* (1993). Nucleotide sequence of the bovine viral diarrhoea virus Osloss strain: comparison with related viruses and identification of specific DNA probes in the 5' untranslated region. *J Gen Virol* 74 (Pt 7), 1433-1438.
- Deng, R. & Brock, K. V. (1992). Molecular cloning and nucleotide sequence of a pestivirus genome, noncytopathic bovine viral diarrhoea virus strain SD-1. *Virology* 191, 867-869.
- Deng, R. & Brock, K. V. (1993). 5' and 3' untranslated regions of pestivirus genome: primary and secondary structure analyses. *Nucleic Acids Res* 21, 1949-1957.
- Dibrov, S. M., Johnston-Cox, H., Weng, Y. H. & Hermann, T. (2007). Functional architecture of HCV IRES domain II stabilized by divalent metal ions in the crystal and in solution. *Angew Chem Int Ed Engl* 46, 226-229.
- Easton, L. E., Locker, N. & Lukavsky, P. J. (2009). Conserved functional domains and a novel tertiary interaction near the pseudoknot drive translational activity of hepatitis C virus and hepatitis C virus-like internal ribosome entry sites. *Nucleic Acids Res* 37, 5537-5549.
- Fletcher, S. P. & Jackson, R. J. (2002). Pestivirus internal ribosome entry site (IRES) structure and function: elements in the 5' untranslated region important for IRES function. *J Virol* 76, 5024-5033.

- Flores, E. F., Ridpath, J. F., Weiblen, R., Vogel, F. S. & Gil, L. H. (2002). Phylogenetic analysis of Brazilian bovine viral diarrhoea virus type 2 (BVDV-2) isolates: evidence for a subgenotype within BVDV-2. *Virus Res* 87, 51-60.
- Friebe, P., Lohmann, V., Krieger, N. & Bartenschlager, R. (2001). Sequences in the 5' nontranslated region of hepatitis C virus required for RNA replication. *J Virol* 75, 12047-12057.
- Fukushi, S., Okada, M., Stahl, J., Kageyama, T., Hoshino, F. B. & Katayama, K. (2001). Ribosomal protein S5 interacts with the internal ribosomal entry site of hepatitis C virus. *J Biol Chem* 276, 20824-20826.
- Fulton, R. W., Ridpath, J. F., Confer, A. W., Saliki, J. T., Burge, L. J. & Payton, M. E. (2003). Bovine viral diarrhoea virus antigenic diversity: impact on disease and vaccination programmes. *Biologicals* 31, 89-95.
- Fulton, R. W., Ridpath, J. F., Ore, S., Confer, A. W., Saliki, J. T., Burge, L. J. & Payton, M. E. (2005). Bovine viral diarrhoea virus (BVDV) subgenotypes in diagnostic laboratory accessions: distribution of BVDV1a, 1b, and 2a subgenotypes. *Vet Microbiol* 111, 35-40.
- Fulton, R. W., Saliki, J. T., Burge, L. J., d'Offay, J. M., Bolin, S. R., Maes, R. K., Baker, J. C. & Frey, M. L. (1997). Neutralizing antibodies to type 1 and 2 bovine viral diarrhoea viruses: detection by inhibition of viral cytopathology and infectivity by immunoperoxidase assay. *Clin Diagn Lab Immunol* 4, 380-383.
- Fulton, R. W., Whitley, E. M., Johnson, B. J., Ridpath, J. F., Kapil, S., Burge, L. J., Cook, B. J. & Confer, A. W. (2009). Prevalence of bovine viral diarrhoea virus (BVDV) in persistently infected cattle and BVDV subtypes in affected cattle in beef herds in south central United States. *Can J Vet Res* 73, 283-291.
- Gardner, P. P., Daub, J., Tate, J. G., Nawrocki, E. P., Kolbe, D. L., Lindgreen, S., Wilkinson, A. C., Finn, R. D., Griffiths-Jones, S., Eddy, S. R. & Bateman, A. (2009). Rfam: updates to the RNA families database. *Nucleic Acids Res* 37, D136-140.
- Gorodkin, J., Zwieb, C. & Knudsen, B. (2001). Semi-automated update and cleanup of structural RNA alignment databases. *Bioinformatics* 17, 642-645.
- Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A. & Eddy, S. R. (2003). Rfam: an RNA family database. *Nucleic Acids Res* 31, 439-441.
- Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 41, 95-98.
- Heymann, J. B., Chagoyen, M. & Belnap, D. M. (2005). Common conventions for interchange and archiving of three-dimensional electron microscopy information in structural biology. *J Struct Biol* 151, 196-207.

- Higgins, D. G., Thompson, J. D. & Gibson, T. J. (1996). Using CLUSTAL for multiple sequence alignments. *Methods Enzymol* 266, 383-402.
- Jubin, R., Vantuno, N. E., Kieft, J. S., Murray, M. G., Doudna, J. A., Lau, J. Y. & Baroudy, B. M. (2000). Hepatitis C virus internal ribosome entry site (IRES) stem loop III_d contains a phylogenetically conserved GGG triplet essential for translation and IRES folding. *J Virol* 74, 10430-10437.
- Kieft, J. S., Zhou, K., Grech, A., Jubin, R. & Doudna, J. A. (2002). Crystal structure of an RNA tertiary domain essential to HCV IRES-mediated translation initiation. *Nat Struct Biol* 9, 370-374.
- Kieft, J. S., Zhou, K., Jubin, R. & Doudna, J. A. (2001). Mechanism of ribosome recruitment by hepatitis C IRES RNA. *RNA* 7, 194-206.
- Klinck, R., Westhof, E., Walker, S., Afshar, M., Collier, A. & Aboul-Ela, F. (2000). A potential RNA drug target in the hepatitis C virus internal ribosomal entry site. *RNA* 6, 1423-1431.
- Klosterman, P. S., Tamura, M., Holbrook, S. R. & Brenner, S. E. (2002). SCOR: a Structural Classification of RNA database. *Nucleic Acids Res* 30, 392-394.
- Kolupaeva, V. G., Pestova, T. V. & Hellen, C. U. (2000). Ribosomal binding to the internal ribosomal entry site of classical swine fever virus. *RNA* 6, 1791-1807.
- Landry, D. M., Hertz, M. I. & Thompson, S. R. (2009). RPS25 is essential for translation initiation by the Dicistroviridae and hepatitis C viral IRESs. *Genes Dev* 23, 2753-2764.
- Larsen, N. & Zwieb, C. (1991). SRP-RNA sequence alignment and secondary structure. *Nucleic Acids Res* 19, 209-215.
- Lavender, C. A., Ding, F., Dokholyan, N. V. & Weeks, K. M. (2010). Robust and generic RNA modeling using inferred constraints: a structure for the hepatitis C virus IRES pseudoknot domain. *Biochemistry* 49, 4931-4933.
- Lindenbach, B. D., Thiel, H.-J. & Rice, C. M. (2007). Flaviviridae: The Viruses and Their Replication. In *Fields Virology*, 5 edn, pp. 1101-1152. Edited by P. H. Knipe, P. M. Howley, D. E. Griffin, R. A. Lamb, M. A. Martin, B. Roizman & S. E. Straus. Philadelphia, PA: Lohmann Williams & Wilkins.
- Locker, N., Easton, L. E. & Lukavsky, P. J. (2007). HCV and CSFV IRES domain II mediate eIF2 release during 80S ribosome assembly. *EMBO J* 26, 795-805.
- Lukavsky, P. J., Kim, I., Otto, G. A. & Puglisi, J. D. (2003). Structure of HCV IRES domain II determined by NMR. *Nat Struct Biol* 10, 1033-1038.

- Lukavsky, P. J., Otto, G. A., Lancaster, A. M., Sarnow, P. & Puglisi, J. D. (2000). Structures of two RNA domains essential for hepatitis C virus internal ribosome entry site function. *Nat Struct Biol* 7, 1105-1110.
- McGinnis, S. & Madden, T. L. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res* 32, W20-25.
- Moes, L. & Wirth, M. (2007). The internal initiation of translation in bovine viral diarrhea virus RNA depends on the presence of an RNA pseudoknot upstream of the initiation codon. *Virology* 361, 124.
- Mueller, F., Doring, T., Erdemir, T., Greuer, B., Junke, N., Osswald, M., Rinke-Appel, J., Stade, K., Thamm, S. & Brimacombe, R. (1995). Getting closer to an understanding of the three-dimensional structure of ribosomal RNA. *Biochem Cell Biol* 73, 767-773.
- Myers, T. M., Kolupaeva, V. G., Mendez, E., Baginski, S. G., Frolov, I., Hellen, C. U. & Rice, C. M. (2001). Efficient translation initiation is required for replication of bovine viral diarrhea virus subgenomic replicons. *J Virol* 75, 4226-4238.
- Ouellet, J., Melcher, S., Iqbal, A., Ding, Y. & Lilley, D. M. J. (2010). Structure of the three-way helical junction of the hepatitis C virus IRES element. *RNA* 16, 1597-1609.
- Paulsen, R. B., Seth, P. P., Swayze, E. E., Griffey, R. H., Skalicky, J. J., Cheatham, T. E., 3rd & Davis, D. R. (2010). Inhibitor-induced structural change in the HCV IRES domain IIa RNA. *Proc Natl Acad Sci U S A* 107, 7263-7268.
- Pawlotsky, J. M., Chevaliez, S. & McHutchison, J. G. (2007). The hepatitis C virus life cycle as a target for new antiviral therapies. *Gastroenterology* 132, 1979-1998.
- Pellerin, C., van den Hurk, J., Lecomte, J. & Tussen, P. (1994). Identification of a new group of bovine viral diarrhea virus strains associated with severe outbreaks and high mortalities. *Virology* 203, 260-268.
- Pestova, T. V. & Hellen, C. U. (1999). Internal initiation of translation of bovine viral diarrhea virus RNA. *Virology* 258, 249-256.
- Pestova, T. V., Shatsky, I. N., Fletcher, S. P., Jackson, R. J. & Hellen, C. U. (1998). A prokaryotic-like mode of cytoplasmic eukaryotic ribosome binding to the initiation codon during internal translation initiation of hepatitis C and classical swine fever virus RNAs. *Genes Dev* 12, 67-83.
- Poole, T. L., Wang, C., Popp, R. A., Potgieter, L. N., Siddiqui, A. & Collett, M. S. (1995). Pestivirus translation initiation occurs by internal ribosome entry. *Virology* 206, 750-754.

- Purchio, A. F., Larson, R., Torborg, L. L. & Collett, M. S. (1984). Cell-free translation of bovine viral diarrhoea virus RNA. *J Virol* 52, 973-975.
- Renard, A., Schmetz, D., Guiot, C., Brown-Shimmer, S., Dagenais, L., Pastoret, P. P., Dina, D. & Martial, J. A. (1987). Molecular cloning of the bovine viral diarrhoea virus genomic RNA. *Ann Rech Vet* 18, 121-125.
- Ridpath, J. F. & Bolin, S. R. (1995). The genomic sequence of a virulent bovine viral diarrhoea virus (BVDV) from the type 2 genotype: detection of a large genomic insertion in a noncytopathic BVDV. *Virology* 212, 39-46.
- Ridpath, J. F. & Bolin, S. R. (1998). Differentiation of types 1a, 1b and 2 bovine viral diarrhoea virus (BVDV) by PCR. *Mol Cell Probes* 12, 101-106.
- Ridpath, J. F., Bolin, S. R. & Dubovi, E. J. (1994). Segregation of bovine viral diarrhoea virus into genotypes. *Virology* 205, 66-74.
- Ridpath, J. F., Neill, J. D., Frey, M. & Landgraf, J. G. (2000). Phylogenetic, antigenic and clinical characterization of type 2 BVDV from North America. *Vet Microbiol* 77, 145-155.
- Rijnbrand, R., Thiviyanathan, V., Kaluarachchi, K., Lemon, S. M. & Gorenstein, D. G. (2004). Mutational and structural analysis of stem-loop III C of the hepatitis C virus and GB virus B internal ribosome entry sites. *J Mol Biol* 343, 805-817.
- Rijnbrand, R., van der Straaten, T., van Rijn, P. A., Spaan, W. J. & Bredenbeek, P. J. (1997). Internal entry of ribosomes is directed by the 5' noncoding region of classical swine fever virus and is dependent on the presence of an RNA pseudoknot upstream of the initiation codon. *J Virol* 71, 451-457.
- Siridechadilok, B., Fraser, C., Hall, R., Doudna, J. & Nogales, E. (2005). Structural roles for human translation factor eIF3 in initiation of protein synthesis. *Science* 310, 1513-1515.
- Sizova, D. V., Kolupaeva, V. G., Pestova, T. V., Shatsky, I. N. & Hellen, C. U. (1998). Specific interaction of eukaryotic translation initiation factor 3 with the 5' nontranslated regions of hepatitis C virus and classical swine fever virus RNAs. *J Virol* 72, 4775-4782.
- Spahn, C. M., Jan, E., Mulder, A., Grassucci, R. A., Sarnow, P. & Frank, J. (2004). Cryo-EM visualization of a viral internal ribosome entry site bound to human ribosomes: the IRES functions as an RNA-based translation factor. *Cell* 118, 465-475.
- Spahn, C. M., Kieft, J. S., Grassucci, R. A., Penczek, P. A., Zhou, K., Doudna, J. A. & Frank, J. (2001). Hepatitis C virus IRES RNA-induced changes in the conformation of the 40s ribosomal subunit. *Science* 291, 1959-1962.

- Thiel, H.-J., Collett, M. S., Gould, E. A., Heinz, F. X., Meyers, G., Purcell, R. H., Rice, C. M. & Houghton, M. (2005). Flaviviridae. In *Virus Taxonomy: VIIIth Report of the International Committee on Taxonomy of Viruses*, pp. 981–998. Edited by C. M. Fauquet, M. A. Mayo, J. Maniloff, U. Desselberger & L. A. Ball. San Diego, CA: Elsevier Academic Press.
- Tolan, D. R. & Traut, R. R. (1981). Protein topography of the 40 S ribosomal subunit from rabbit reticulocytes shown by cross-linking with 2-iminothiolane. *J Biol Chem* 256, 10129-10136.
- Uchiumi, T., Terao, K. & Ogata, K. (1981). Identification of neighboring protein pairs cross-linked with dimethyl 3,3'-dithiobispropionimidate in rat liver 40S ribosomal subunits. *J Biochem* 90, 185-193.
- Wang, C., Le, S. Y., Ali, N. & Siddiqui, A. (1995). An RNA pseudoknot is an essential structural element of the internal ribosome entry site located within the hepatitis C virus 5' noncoding region. *RNA* 1, 526-537.
- Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189-1191.
- Zhao, Q., Han, Q., Kissinger, C. R., Hermann, T. & Thompson, P. A. (2008). Structure of hepatitis C virus IRES subdomain IIa. *Acta Crystallogr D Biol Crystallogr* 64, 436-443.
- Zwieb, C., van Nues, R. W., Rosenblad, M. A., Brown, J. D. & Samuelsson, T. (2005). A nomenclature for all signal recognition particle RNAs. *RNA* 11, 7-13.

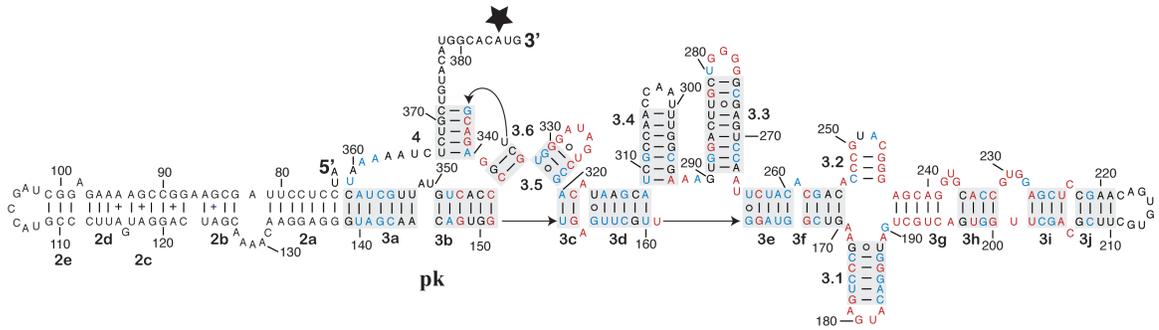


Figure 5.1. The secondary structure of BVDV-1b strain Osloss IRES RNA (GenBank accession M96687). Residues which are invariant or more than 99 percent conserved are shown in red for helices 3 and 4; conservations between 95-99% are indicated in blue. Base pairs supported by covariation analysis are shown on a gray background. The 5' and 3' ends are indicated, as are helical sections and the start codon (star). Sections in helices 3 and 4 are colored gray to reflect support from covariation analysis or mutagenesis, respectively. Sections in helix 2 are not colored due to insufficient information in the alignment.

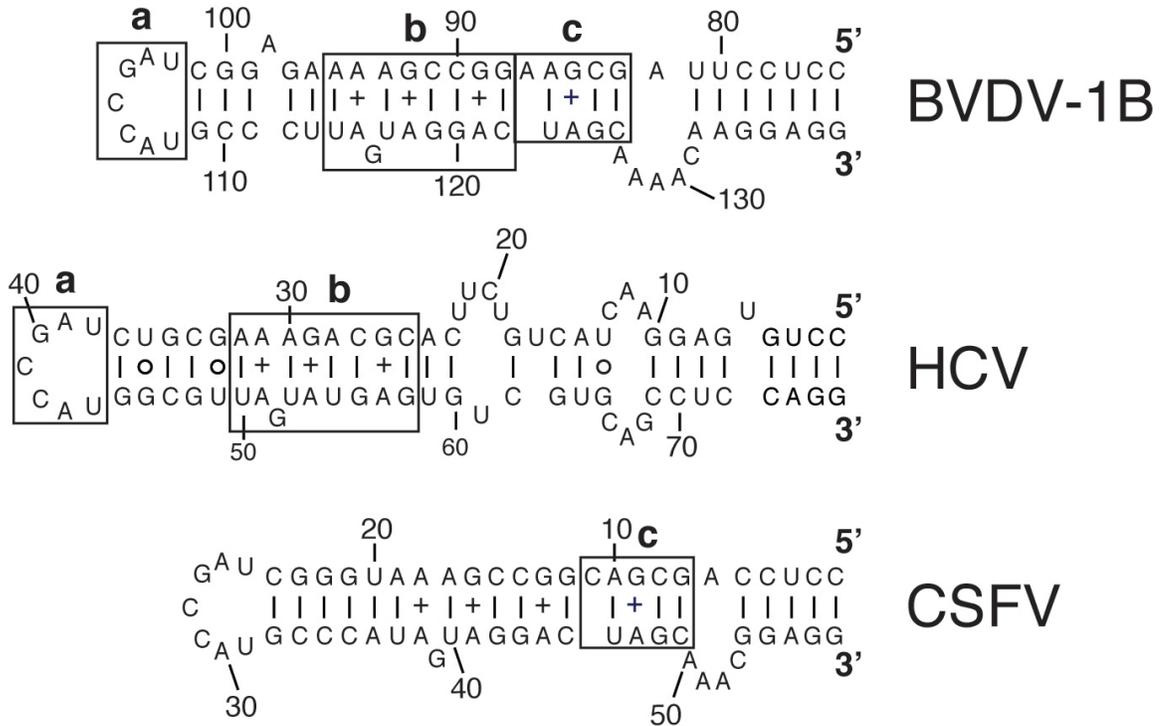


Figure 5.2. Secondary structure of helix 2 of BVDV-1b IRES RNA (top) in comparison with those derived from NMR analysis of HCV (middle; Lukavsky et al., 2003) and CSFV (bottom; Locker et al., 2007) IRES RNAs. Feature a (residues 38-44) and feature b (25-32 and 50-58) are from P15P.pdb. Feature c is from residues 7-11 and 46-49 of 2HUA.pdb. Canonical base pairs are indicated with dashes or circles (G-U pairs) and non-canonical A-G and A-A pairs with plus signs.

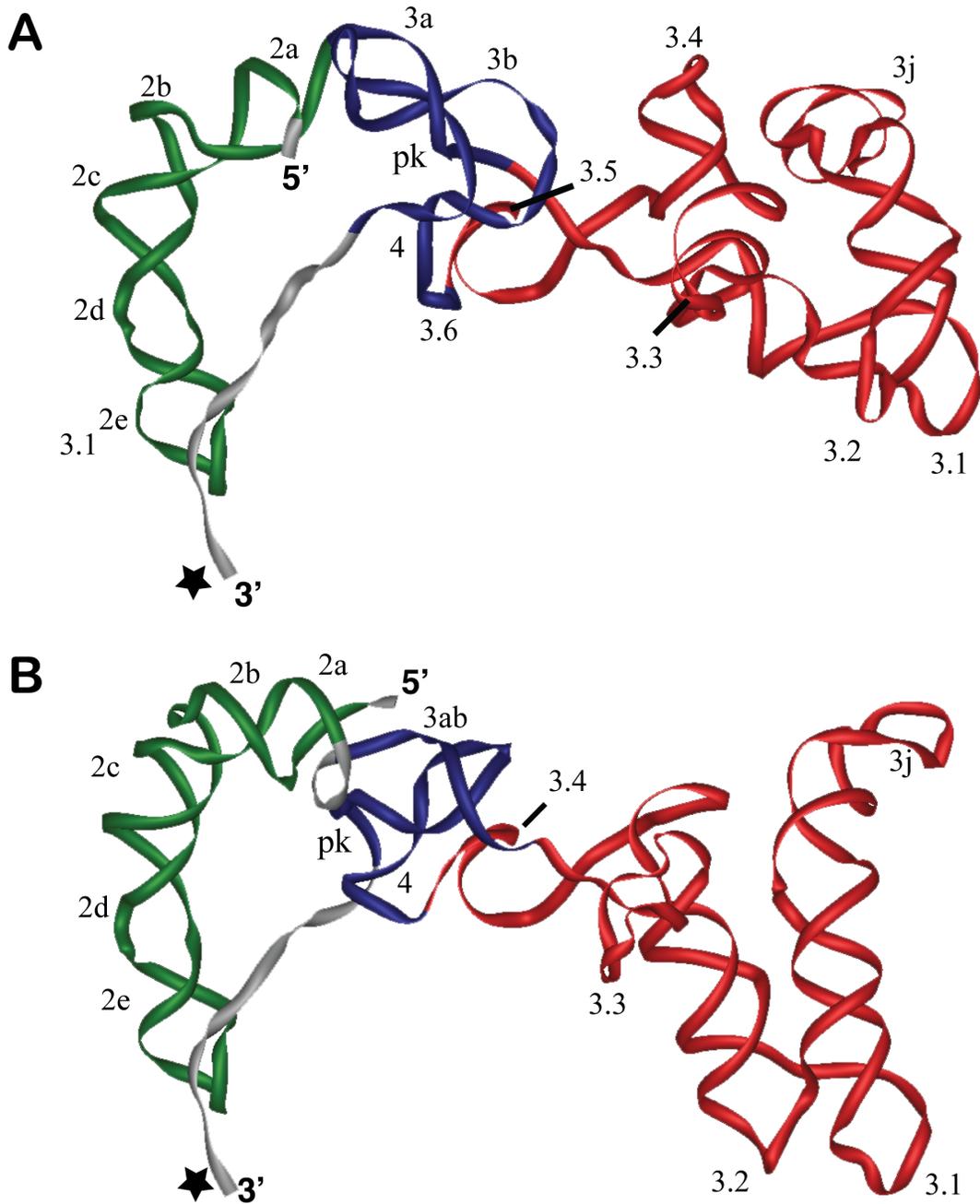


Figure 5.3. Three-dimensional IRES RNA Models. **A)** Model of BVDV subgenotype 1b IRES RNA showing helix 2 (green), the pseudoknot (blue) and helix 3 (red). Features are indicated as in Figure 5.1. **B)** Representation of HCV subgenotype 1b IRES RNA derived from the secondary structure published by Spahn *et al.*, 2001. Regions are colored as in 3A.

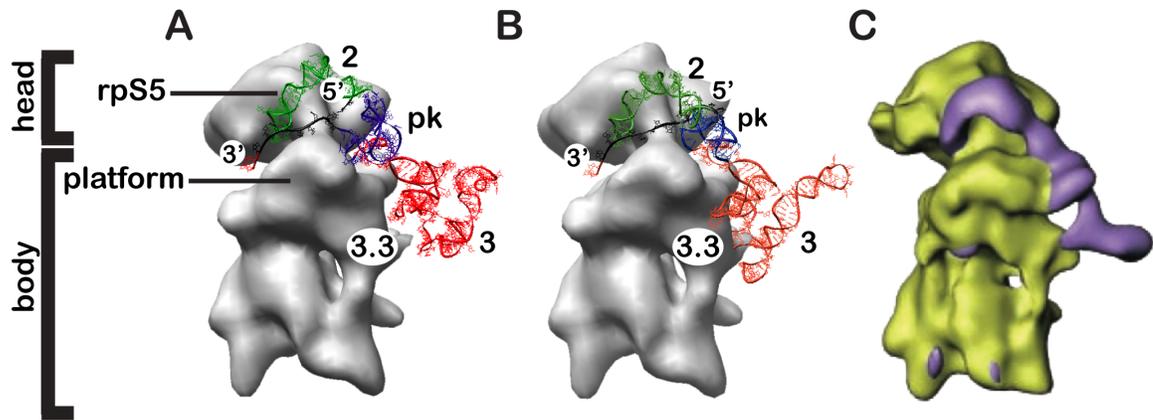


Figure 5.4. Placement of the IRES RNA models on the 40S ribosomal subunit. A) Model of BVDV IRES RNA (Figure 5.3A) on the 40S ribosomal subunit (gray, EMD ID 1092; Spahn *et al.*, 2004). **B)** Representation of HCV IRES RNA (Figure 5.3B) on the 40S ribosomal subunit. **C)** HCV IRES RNA (purple) bound to rabbit reticulocyte 40S ribosomal subunit (yellow) as observed by cryo-electron microscopy (from Spahn *et al.*, 2001. Reprinted with permission from AAAS.). Colors of the BVDV and HCV IRES RNA models are as shown in Figure 5.3. The head, body, platform, approximate location of rpS5, 5' and 3' ends and the start codons (red) are indicated.

BVDV Residues	CSFV Residues	Helices	Effects	Expression	References
139-144	131-133	3a	d	-	Fletcher & Jackson (2002)
353-358	342-344	3a	d	-	
139-144, 353-358	131-133, 342-344	3a	c	+	
322-323	309-310	3.5	d	-	Kolupaeva <i>et al.</i> (2000)
333-334	320-321	3.6	d	-	
344-346	331-333	3b	d	-	
336-341	323-328	4	d	-	Rijnbrand <i>et al.</i> (1997)
366-371	356-361	4	d	-	
336-341, 366-371	323-328, 356-361	4	c	+	
339-340	325-326	4	d	-	Fletcher & Jackson (2002), Pestova <i>et al.</i> (1998)
368-369	357-358	4	d	-	
339-340, 368-369	325-326, 357-358	4	c	+	
340-341	327-328	4	d	(+)	Fletcher & Jackson (2002)
366-367	355-356	4	d	-	
340-341, 366-367	327-328, 355-356	4	c	+	
139-150		3a	d	-	Moes & Wirth (2007)
344-358		3b	d	-	
139-150, 344-358		3a, 3b	c	+	
139-143		3a	d	-	
339-342		3.6, 4	d	-	
365-368		4	d	-	
339-342, 365-368		3.6, 4	c	+	

Table 5.1. Mutations and compensatory changes of the proposed base pairs of the BVDV or CSFV IRES RNA pseudoknots and their effects on IRES-mediated translation relative to wild-type. Positions in BVDV-1b IRES RNA corresponding with the mutated residues in the CSFV IRES RNA are given in the first and second columns. Helical sections are given in the third column. Intents of mutations (disruption or compensation of base pairs) are shown in the fourth column with their effect on translation of a reporter product given in the fifth. Reporter products measured in references are as follows: Chloramphenicol acetyltransferase (Rijnbrand *et al.*, 1997), truncated influenza NS1 protein (Fletcher and Jackson, 2002; Kolupaeva *et al.*, 2000b; Pestova *et al.*, 1998), or Luciferase (Moes and Wirth, 2007). Pluses indicate production of reporter product and minuses loss of expression. Mutation of nt 327-328 (+) in CSFV decreased expression to half of wild-type levels in contrast to the 355-356 mutant (loss of expression). References are indicated in the last column.

BVDV-1b Residues	CSFV Residues	Helices	Methods	Effects
G148	G139	3b	T1	p
A259	A250		V1	e
G200	G188		T1	p
G236	G219		T1	e
G271	U262	3.3	V1	p
277-GGG-279	268-GGG-270		T1	p
295-CG-296	284-AU-285	3.4	V1	p
G339	G326	4	T1	p
U348	U335	3b	V1	e
A351	G337		T1	e
G373	G362		T1	p
379-GG-380	358-GG-359		T1	p
A381	A61		V1	p
386-GG-387	375-GG-376		T1	p
G388	G378		T1	p
U389	U379		V1	p
A396	C385		V1	p
C397	A386		V1	p
399-AAU-401	388-UUU-390		V1	p

Table 5.2. Expected modifications in BVDV IRES RNA based on enzymatic modifications of wild-type CSFV IRES RNA due to binding of the rabbit reticulocyte 40S ribosomal subunit. Locations of corresponding nucleotides in BVDV-1b domains (D) or features are shown in the first and second columns. The corresponding BVDV-1b Osloss IRES RNA residues are indicated in the first column, followed by the actual modified positions in CSFV IRES RNA (Kolupaeva *et al.*, 2000b). Modifications located in helical sections are indicated in the third column for reference. Methods used to enzymatically modify the RNA (RNases T1 or V1) are indicated in the fourth column. In the last column, protections (p) and enhanced accessibility (e) to RNases are indicated.

Feature	Residues	Coordinates	Description
1	138-139	ERNA-3D	Connection between helix 2 and pseudoknot
2	335, 341, 359-367, 350-351,	ERNA-3D	Single-stranded connections in pseudoknot
3	154,318	ERNA-3D	^a
4	161, 265-266, 290-292	ERNA-3D	Single-stranded connections in four-way junction between pseudoknot section 3b, helices 3.4 and 3.5 and section 3c
5	A259	1NBS (A130)	One looped out base
6	171-172, 189-190, 254	ERNA-3D	Single-stranded connections in four-way junction between section 3e, helices 3.1 and 3.3 and section 3f
7	179-182	1F85 (6-9)	Hexaloop
8	196, 236-238	1KOC (38, 7-9)	Several looped out bases
9	201, 228-231	1E7K (44, 29-32)	About 90 degree turn
10	206, 223	1J5A (1219, 1253)	Stacked duplex with non-WC pair
11	211-218	1FFZ (2069-2076)	Octaloop
12	247-250	1IDV (4-7)	Tetraloop
13	276-280	1J5A (1794-1798)	Pentaloop
14	300-304	1JJ2 (1469-1473)	Pentaloop
15	324-329	1F85 (5-10)	Tetraloop (325-328) ^b
16	372-386	ERNA-3D	Single-stranded region including the start codon

^aA318 is involved in base pair 318A-U327, which was constructed in ERNA-3D by positioning U327 of Tetraloop feature 15 in proper Watson-Crick base pairing coordinates with A318, while retaining the phosphate connectivity between A320 and G321. ^bThis hairpin loop is conserved in HCV, GBV, CSFV and BVDV 1 and 2. Since the invariance of the loop means that compensatory changes can not be determined, the whole motif was copied. Rijnbrand *et al.* (2004) classifies this structure as a tetraloop, while SCOR lists the loop as a hexaloop.

Table 5.3: Table of the features used for generating a three-dimensional model of the BVDV IRES RNA. Residue positions involved in the structural feature are indicated in the Residues column. The sources of the coordinates (PDB or ERNA-3D) for a given feature are given in the Coordinates column. The last column provides a description of the feature. Features used to model helix 2 of BVDV IRES RNA are as described in Results and Discussion.

Feature	Residues	Coordinates	Description
1	42-43	ERNA-3D	Two nucleotides preceding helix 2
2	119-124	ERNA-3D	Single-stranded region connecting helix 2 and the pseudoknot.
3	303-304, 313-314, 324	ERNA-3D	Pseudoknot
4	136, 288	ERNA-3D	^a
6	154-155, 228	1KH6 (6-7, 38)	Single-strands connecting four-way junction
7	162-165	1F85 (6-9)	Tetraloop
8	176, 223	1P5P (26, 57)	Stacked duplex with one non-WC pair
9	181-187, 210-218	1KP7 (21-27, 4-12)	Internal loops separated by two base pairs and flanked by two base pairs. Residues 185 and 212 are involved in a non-WC base pair, while 183 and 214-216 are involved form an unpaired, unstacked A flanked by non-WC pairs. The entire feature was used
10	192-205	1FJE (5-18)	Tetradecaloop
11	247-250	1IDV (4-7)	Tetraloop
5	G243	1NBS (A130)	One looped out base
12	253-279	1F84 (2-28)	Helix 3.3, including internal and hairpin loops
13	280-283	ERNA-3D	Single-strand connecting three-way junction.
14	294-299	1F85 (5-10)	Tetraloop (295-298) ^b
15	331-344	ERNA-3D	Single-stranded region including the start codon

^aA288 is involved in base pair 288A-U297, which was constructed in ERNA-3D by positioning U297 of Tetraloop feature 14 in proper Watson-Crick base pairing coordinates with A288, while retaining the phosphate connectivity between U290 and G291. ^bThis hairpin loop is conserved in HCV, GBV, CSFV and BVDV 1 and 2. Since the invariance of the residues in the helix means that compensatory changes can not be determined, the whole motif was copied. Rijnbrand *et al.* (2004) classifies this structure as a tetraloop, while SCOR lists the loop as a hexaloop.

Table 5.4. Table of the features used for generating a three-dimensional model of the HCV IRES RNA. Residue positions involved in the structural feature are indicated in the Residues column. The sources of the coordinates (PDB or ERNA-3D) for a given feature are given in the Coordinates column. The last column provides a description of the feature. The features used to model helix 2 of HCV IRES RNA are as described in Results and Discussion.