STRUCTURAL STUDIES OF RIBONUCLEOPROTEIN

COMPLEXES USING MOLECULAR MODELING

A Dissertation
Presented to
The Academic Faculty

by

Batsal Devkota

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in Applied Biology in the
School of Biology

Georgia Institute of Technology
April 2008

STRUCTURAL STUDIES OF RIBONUCLEOPROTEIN

COMPLEXES USING MOLECULAR MODELING

Approved by:

Dr. Stephen C. Harvey, Advisor
School of Biology
Georgia Institute of Technology

Dr. Roger M. Wartell
School of Biology
Georgia Institute of Technology

Dr. Nael A. McCarty
School of Biology
Georgia Institute of Technology

Dr. Nicholas V. Hud
School of Chemistry and Biochemistry
Georgia Institute of Technology

Date Approved: November 30, 2007

Dedicated to my parents.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

cryo-EM	Cryo-electron microscopy
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
mRNA	Messenger RNA
tRNA	Transfer RNA
rRNA	Ribosomal RNA
GTP	Guanosine triphosphate
GDP	Guanosine diphosphate
EF-Tu	Elongation Factor-Tu
ASL	Anticodon Stem Loop
SH	Small subunit helix
LH	Large subunit helix
PTC	Peptidyl transferase center
L1BD	L1 binding domain
L11BD	L11 binding domain
RMSD	Root Mean Square Deviation
YAMMP	Yet Another Molecular Mechanics Program
YUP	YAMMP Under Python
NMFF	Normal Mode Flexible Fitting
PaV	Pariacoto virus

SUMMARY

The current work reports on structural studies of ribonucleoprotein complexes, *Escherichia coli* and *Thermomyces lanuginosus* ribosomes, and Pariacoto virus (PaV) using molecular modeling. Molecular modeling is the integration and representation of the structural data of molecules as models. Integrating high-resolution crystal structures available for the *E. coli* ribosome and the cryo-EM density maps for the PRE- and POST-accommodation states of the translational cycle, I generated two all-atom models for the ribosome in two functional states of the cycle. A program for flexible fitting of the crystal structures into low-resolution maps, YUP.scx, was used to generate the models. Based on these models, we hypothesize that the kinking of the tRNA plays a major role in cognate tRNA selection during accommodation.

Secondly, we proposed all-atom models for the eukaryotic ribosomal RNA. This is part of a collaboration between Joachim Frank, Andrej Sali, and our lab to generate an all-atom model for the eukaryotic ribosome based on a cryo-EM density map of *T. lanuginosus* available at 8.9Å resolution. Homology modeling and *ab initio* RNA modeling were used to generate the rRNA components.

Finally, we propose a first-order model for a T=3, icosahedral, RNA virus called Pariacoto virus. We used the structure available from x-ray crystallography as the starting model and modeled all the unresolved RNA and protein residues. Only 35% of the total RNA genome and 88% of the protein were resolved in the crystal structure. The generated models for the virus helped us determine the location of the missing N-terminal protein tails. The models were used to propose a new viral assembly pathway for small RNA viruses. We propose that the basic N-terminal tails make contact with the RNA

genome and neutralize the negative charges in RNA and subsequently collapse the

RNA/protein complex into a mature virus. This process is reminiscent of DNA

condensation by positively charged ions.

CHAPTER 1

INTRODUCTION

"...It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material."

- Watson and Crick, 1953

Elucidating structure-function relationships is key to understanding the biology of biomolecules. The above quote from a landmark paper on the structure of the DNA showed how solving the structure of the DNA the authors were able to immediately see how DNA would replicate prior to cell division, the phenomenon that had eluded scientists over many decades. In this work, molecular modeling of various ribonucleoprotein complexes using computational tools has been presented. The primary goal in each case is to elucidate how the structure of each molecule relates to the function by using the models obtained for each complex.

Molecular modeling

Molecular modeling is the collection and integration of data to create a model, a three-dimensional representation of a molecule. The model, which may be either a physical or a computational representation, helps us to visualize the structure of the molecule. A model should satisfy all the structural data available for the molecule that it represents. The greater the amount of data, the better the model, and the more reliable the interpretation of it will be. As more data become available, they can be integrated into the existing model and the model can be further refined. The model should assist in the development of a testable hypothesis that aids in the design of further experiments.

In 1951, Linus Pauling proposed the first model for a biological molecule by postulating the structure of α-helices using his knowledge of chemical bonds and the

chemical structure of proteins (1). Within a few years, Watson and Crick were able to develop a model for the structure of DNA, integrating x-ray fiber diffraction data and the known ratio of purine and pyrimidine bases in the DNA (2). Their elucidation of the structure of DNA immediately paved the way to further understanding other properties of DNA. As mentioned in the first paragraph, the last statement in their famous paper on the structure of DNA clearly shows how a functional mechanism could be proposed by knowing the structure. Ever since then, the importance of structures in understanding the function of molecules has been realized.
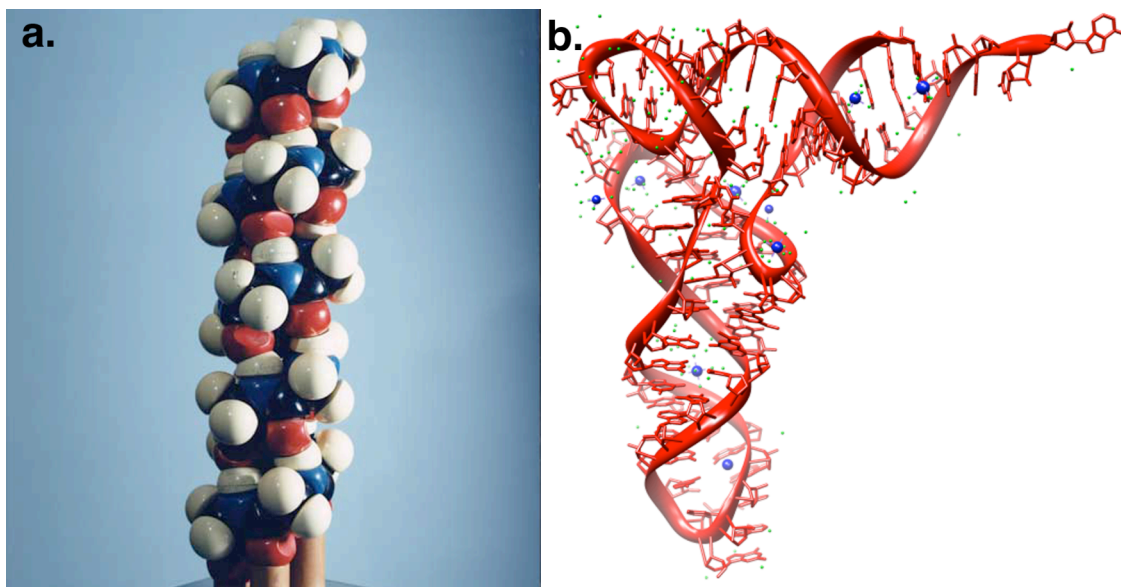
Figure 1.1: Physical and computational models. **(a)** The physical model of α-helix as proposed by Linus Pauling (1). **(b)** Computational model for a tRNA derived from x-ray crystallography (PDB ID: 1EHZ), rendered using Chimera (3). The blue dots represent the Mg^{2+} ions.

The art of molecular modeling started with the use of physical models to solve structures. Pauling used ball and stick models for his α-helices (Figure 1.1a.), while Watson and Crick used metal plates and rods to solve the DNA structure. The use of physical models for molecular models continued until late 1980s, when Walleczek et al. postulated the spatial arrangement of the proteins in the large subunit of ribosome, using

wires to represent the ribosomal RNA and different sized balls to represent the ribosomal proteins (4). Integrating cross-linking data between RNA-RNA and RNA-protein, Brimachome et al. folded the secondary structure of the 16S rRNA into a physical model (5). They were also able to add the proteins from the small subunit to the model using neutron diffraction data, thus by postulating a complete low resolution model for the small subunit of the ribosome (5).

Even though the use of physical models helped advance the field of molecular modeling, the availability of vast amounts of structural data these days requires one to use computational tools to generate models. It is hard to generate an ensemble of models using physical modeling techniques. Also, refining the model would be impossible without rebuilding it from the very beginning. Thus, almost all modeling is now carried out using computational tools. In fact, the term molecular modeling has become synonymous with computational molecular modeling. Advances in computing power and modeling tools have helped the field of molecular modeling greatly. Nowadays, one can easily build, manipulate, analyze and visualize models that consist of tens of thousands of atoms and hundreds of chains. This is an outstanding feat.

The advantage of computer-based modeling is that many models may be generated and analyzed. Refinement is also extremely easy. Another beauty of computer-based molecular modeling is that many forms of data can be used to propose the model. Low-resolution data such as Florescence Resonance Energy transfer (FRET) data, secondary structure information, and data obtained from cross-linking and footprinting experiments may all be used as constraints during computational modeling. Molecular mechanics program such as YAMMP (6) can be used for low-resolution modeling. High-resolution data, such as that obtained from the x-ray crystallography, may be used in conjunction with cryo-EM low-resolution data to propose functionally relevant models. Flexible fitting protocols such as NMFF (7, 8), Situs (9, 10), RsRef (11), and YUP.scx (submitted) can be used for such purposes.

X-ray crystallography

Many experimental techniques exist to solve the structure of a molecule. X-ray crystallography is the most popular and the best method of structure determination for biological macromolecules. The molecule of interest is crystallized and x-ray diffraction data are collected from the crystals. Based on these diffraction data the model of the molecule can be postulated. Because of the large amount of data that is used to solve the structure, the position of each atom can be determined with high confidence. The model obtained from the x-ray crystallography is thus correctly called "the structure" of the molecule. For example the crystal structure of EF-Tu at 2.7Å resolution was solved using more than 80,000 data points (12).

The first crystal structure of a biological macromolecule, myoglobin, was solved by John Kendrew in the late 1950s (13). Upon learning the structure of myoglobin and knowing its function as an oxygen transporter, it was easy to see that conformational changes needed to occur in the molecule to aid in oxygen binding. The structure of myoglobin showed the binding site for the oxygen (haem group) located at the center of the globular protein. It became obvious that the molecule has to be dynamic and possess multiple conformations so as to allow the oxygen to enter the biding sites. There have been many advances in x-ray crystallography made since Kendrew solved the structure of myoglobin. Currently, there are more than 47,000 structures in the Protein Data Bank as compared to just a handful in 1970s (14). With the amount of focus that has been put on getting crystal structures and other type of models, it is clear that structure determination is important to understanding the function of biological molecules.

Another technique for determining high-resolution structures is Nuclear Magnetic Resonance (NMR). The resolution is not comparable to x-ray crystallography, and size is the limiting factor. Generally, molecules greater than 50kDa cannot be easily solved using NMR, even though recent advancements have generated models as high as 100kDa (15). Even though size is the limiting factor, the fact that one does not need to crystallize

the molecule makes it more suitable for studying small molecules that do not easily crystallize and one can study molecules in solution.

Cryo-electron microscopy (cryo-EM)

The development of cryo-EM over the last few decades has added another dimension to the study of biological macromolecules (16, 17). It provides low-resolution data for large macromolecules. The molecules of interest do not need to be crystallized and cryo-EM can capture molecules in their native states, which are valuable advantages over x-ray crystallography. Aqueous solutions containing macromolecular samples are plunged into liquid ethane (-160° C) thus by flash-freezing the solution. This causes the water in the solution to form vitreous ice (glass-like) and in the process traps the molecules in different orientations (18, 19). The use of vitreous ice allows 2-D electron micrographs of the molecules to be captured without any noise from the diffraction of electrons by the ice. Thousands of micrographs of molecules in different orientation are collected. Single particle reconstruction of these micrographs gives a 3-D model. At present, the resolution that can be achieved is only about 6-9Å, thus it is inferior to the resolution of x-ray structures. Only well defined structural features of the molecules, such as α-helices and β-sheets for proteins and RNA helices can be recognized at this resolution. Individual amino acids and nucleotides cannot be resolved. Nevertheless, the increase in resolution from about 45Å for the ribosome that was published in 1991 (20) to 8.9Å for the eukaryotic ribosome that we used in the current study is quite an achievement for the cryo-EM field. This can be attributed to advances in sorting methods, developments in the techniques of electron microscopy, improved image analysis techniques and better protocols for preparing purified samples (21). Within the next few years, we can expect to see cryo-EM density maps with more than 5Å resolution.

Structural studies of ribosome

The ribosome is a large ribonucleoprotein complex, essential for cell viability because of its role in protein synthesis. It decodes the genetic information in a messenger RNA (mRNA) and with the aid of transfer RNAs (tRNA) and other protein cofactors translates it into protein (22-24). The ribosome is composed of approximately 2/3 RNA and 1/3 protein. The prokaryotic ribosome measures approximately 220Å along its longest side. The total mass of the prokaryotic ribosome (70S) is ~2.5 x 10^6 Da. It is composed of two asymmetric subunits (25), large subunit (LSU) and small subunit (SSU), named based on their size (Figure 1.2).

The ribosome was first observed in 1955 by Romanian cell biologist George Palade using electron microscopy (EM), as small dense granules attached to the endoplasmic reticulum in cells (26). Later using EM and other biochemical techniques, it was shown that the ribosome is composed of RNA and protein components and is the site of protein synthesis. As the magnification of EM increased and other methods of elucidating structure emerged, the morphology of the ribosome became more detailed (27, 28). The current knowledge that we have of the ribosome is based on a tremendous amount of work on elucidating the ribosome structure over the last few decades from many labs.
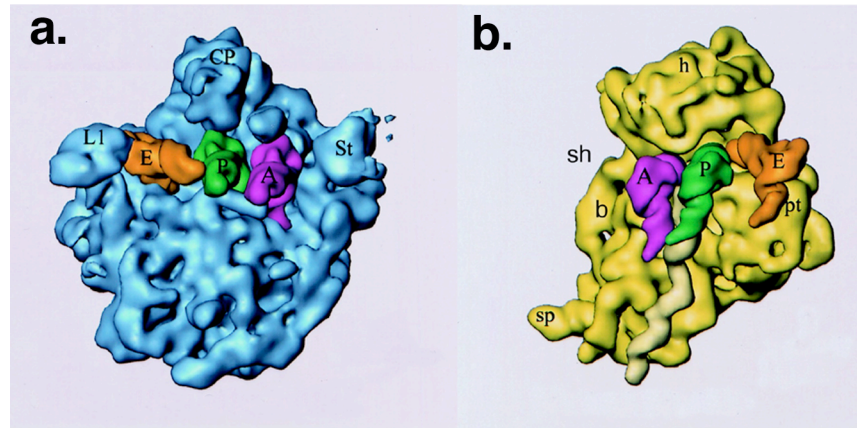
Figure 1.2: Structure of ribosome subunits as determined by cryo-electron microscopy (cryo-EM). (a) Intersubunit view of the large subunit (LSU). Purple, green and orange are the tRNAs at A, P and E-site respectively. L1 – L1 binding domain; CP- central protuberance; St – L7/L12 stalk. (b) Intersubunit view of the small subunit (SSU). tRNAs as in the LSU. h – head; sh – shoulder; b – body; sp- spur; pt –platform.

The major structural features of the SSU are the head, body, platform, and the spur. In between the head and the body, there is a channel called the mRNA channel, through which the mRNA traverses during translation. The initial interaction between the tRNA anticodon loop and mRNA takes place at the decoding center that lies in the SSU. The decoding center lies on top of the penultimate stem, which is a long helix in the SSU that makes contact with the LSU at multiple locations. The LSU has a central protuberance (CP), made up of 5S rRNA and the L1 and L11 binding domains (stalk) as arms (Figure 2). The LSU contains a peptidyl transfer center (residues 2448-2454), where the peptide bond forms between the P-site polypeptide and the A-site amino-acid. [All the numbering for residues are based on *E.coli* ribosome, unless specified otherwise]. A tunnel extends from the PTC to the cytoplasm, from which the nascent protein exits the ribosome. The formation of a complex of SSU (30S) and LSU (50S) subunits, makes a cleft between the two subunits, along which the tRNA moves during the translational cycle. Approximately 12 intersubunit bridges hold the two subunits together (29). The functional sites of the ribosome, and the intersubunit bridges are made almost exclusively of RNA. Thus, the ribosome is also considered a ribozyme (30).

Since Yonath began crystallizing the ribosome more than 25 years ago (31), the development of crystallographic techniques has produced many high resolution crystal structures of ribosomal subunits and its cofactors. As of now, we have a crystal structure of a 30S at 3Å (32, 33) from *T. thermophilus*, a 50S at 2.4Å (34) from *H. marismortui*, a 70S at 5.5Å (35) from *T. thermophilus,* and a 70S from *E. coli* at 3.5Å (29). The crystal structures of different cofactors like tRNA (36), ternary complex (tRNA[Cys].EF-Tu.GDPNP) (12, 37), EF-G (38), Ribosome Recycling Factor (RRF) (39) are also available. Recently, crystal structures of the ribosome complexed with tRNA and mRNA have also been solved (40), along with the ribosome complexed with RRF (41). All these structures have contributed significantly to the elucidation of structure/function relationships of ribosome and its components.

In 1991, Joachim Frank published the first three dimensional (3D) reconstruction of ribosomes based on Cryo-EM (20). The resolution was only about 45Å. Antibiotics like kirromycin, furasidic acid and a non-hydrozylable GTP analog, GDPNP, have been used to stall ribosomes at certain stages and to reconstruct models at different states of translation. Because of this capability, we have cryo-EM density maps of ribosomes (both prokaryotic and eukaryotic) captured at various stages of translational cycle (Table 1.1), which allows us to study the dynamic nature of the ribosome.

Translational cycle

The availability of crystal structures of the ribosomal subunits and cofactors, along with the cryo-EM data at intermediate states of the translational cycle have enabled us to understand the mechanism of translational cycle in great detail (Figure 1.3). The translation of the mRNA for accurate protein synthesis starts with the initial recognition of mRNA by the small subunit. The mRNA wraps around the mRNA channel with its Shine-Delgarno (SD) sequence binding to the anti-SD sequence present at the 3'end of

Table 1.1: Low-resolution cryo-EM density maps available for various states of translational cycle for *E. coli* ribosome.

	Structure	A-site	P-site	E-site	Resolution
Initiation complex					
Allen et al., 2005	70S-fMet-tRNA^{fMet}-IF1,IF2		fMet-tRNA^{fMet}		13.8Å
Gabashvili et al., 2000	70S-fMet-tRNA^{fMet}		fMet-tRNA^{fMet}		11.5Å
Accommodation /Peptidyl transfer					
Valle et al., 2003	Phe-tRNA^{Phe}-Ef-Tu-GDP-kir complex	Phe-tRNA^{Phe}	fMet-tRNA^{fMet}	deacylated tRNA	9.0Å
Valle et al., 2003	fMet-Phe-tRNA^{fMet}	fMet-Phe-tRNA^{Phe}	deacylated tRNA	deacylated tRNA	9.0Å
Pre-translocation / Translocation					
Agrawal et al., 1999	70S-(tRNA)2-EF-G-GMPP(CH2)P complex	tRNA	tRNA		17.5Å – 18.4Å
Agrawal et al., 1999	70S-(tRNA)2-EF-G-GDP-fusidic acid complex		tRNA	tRNA	17.5Å – 18.4Å
Termination					
Rawat et al., 2003	Release complex (wt) RF2		fMTI-tRNA^{Ile}		12.8Å
Rawat et al., 2003	Release complex (mutant) RF2 (GAQ)		fMTI-tRNA^{Ile}		10.9Å
Klaholz et al., 2003	Release complex - RF2	UAA stop codon	tetrapeptidyl tRNA		14.0Å
Valle et al., 2003	70S.tRNA.EF-Tu.tmRNA.SmpB	tmRNA	tRNA		~13-15Å

the 16S rRNA. The initiator tRNA (fMet-tRNA^{fMet}) binds to the P-site to form the pre-initiation complex.

Initiation factors (IF1, IF2 and IF3) help recruit the LSU to the SSU to form the functional 70S ribosome. As soon as the 70S complex forms, the ternary complex (aatRNA-EF-Tu-GTP) brings the amnio-acylated tRNA (aatRNA) to the A-site. If the ternary complex has the correct tRNA that matches the mRNA codon, GTP hydrolysis occurs, which causes the EF-Tu to lose its affinity for the tRNA and the ribosome. Thus, it falls off the ribosome. In the mean time, the tRNA is accommodated into the peptidyl transfer center where the amino-acid it is carrying forms a peptide bond with the terminal amino-acid of the growing polypeptide attached to the P-site tRNA. In this process, the peptide is transferred from P-site to A-site tRNA, and an aminoacid is added to the polypeptide.

Peptide bond formation is followed by translocation where the A-site tRNA translocates to P-site and the deacylated P-site tRNA moves to the E-site. The E-site tRNA then exits the ribosome. The translocation of tRNAs is aided by another elongation factor, EF-G, and is an energy dependent process. The GTP is hydrolyzed, which again changes the conformation of EF-G, causing it to fall off the ribosome. With the void at the A-site, another ternary complex approaches the ribosome and the translation cycle continues.

The cycle stops when the stop codon reaches the A-site. At this stage, the ribosome release factors help cleave the polypeptide off the P-site tRNA and separate the ribosome into its individual components. The subunits are recycled for another cycle of translational cycle.
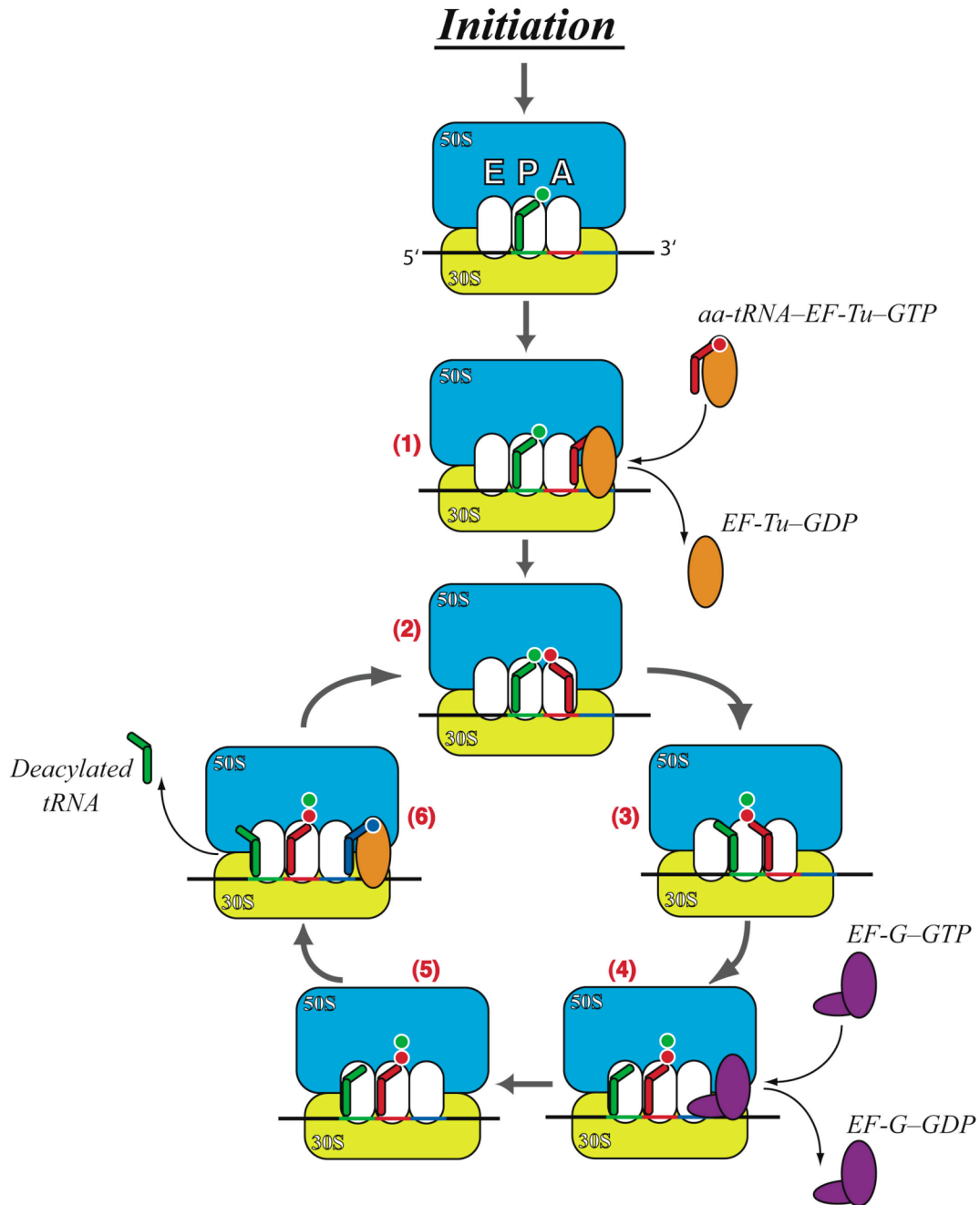
Figure 1.3: Schematic representation of the translational cycle. The initial selection of tRNA takes place at step (1). The second selection step for tRNA occurs while the tRNA (here in red) swings inside the ribosome from step (1) to step (2). Peptidyl transfer occurs during (2). Steps (2) to (6) represent the elongation cycle. The peptide grows by one amino-acid per cycle. When the stop codon reaches the A-site at step (5), the termination factors separate the ribosome into individual components (not shown in the figure). This figure was adapted from Mears, J. PhD Thesis, University of Alabama, 2006.

Fidelity of translation

The most crucial step in translation is the selection of the correct tRNA that corresponds to the mRNA codon. This process is called selection. During selection, the anticodon loop of the tRNA binds with the mRNA codon. If the anticodon matches the codon, then the tRNA is selected. If not, it is rejected. The $\Delta\Delta G$ between binding the cognate tRNA (tRNA with perfect match with mRNA codon) and binding the near-cognate (tRNA with similar match to mRNA codon) is about 3kcal/mol. From Boltzmann's relation, the error rate of incorrect tRNA binding should be about 1 in 100. However, the error rate is about 1 in 10000. Hopfield and Ninio in 1970's in two independent studies proposed that the selection of tRNA occurs twice, once during initial selection and once during accommodation, separated by an irreversible reaction. This makes the two selection processes mutually exclusive, which explains the increase in fidelity (42, 43). After almost four decades, thanks to the results of a series of elegant kinetic experiments, we know the tRNA selection mechanism (44) and the reaction rates of intermediate steps.

Initial selection occurs when amino-acylated tRNA is brought to the ribosome by the elongation factor, EF-Tu. This complex is called a ternary complex. During initial selection, all the non-cognate tRNAs (tRNA anticodon loop that differs substantially to the mRNA codon) get rejected. The forward rate constant of both cognate and near-cognate tRNAs is 20/sec. The interesting observation is that, even though the forward rate constant is the same, the fall off or backward rate constant for cognate and near-cognate is $0.2s^{-1}$ and $20s^{-1}$ respectively. The differential rate for the backward reaction for the cognate and near-cognate shows that an induced fit mechanism occurs in the ribosome that allows the ribosome/tRNA complex to form a more stable complex for cognate tRNA as compared to the near-cognate (44). The crystal structure of the decoding site with cognate tRNA bound has identified the role of universally conserved

nucleotides A1492, A1493 and G530 from the 16S rRNA in stabilizing the helix formed between tRNA and mRNA (45). From cryo-EM data, it has been further observed that the cognate tRNA undergoes kinking between the anti-codon stem and D-stem (46) enabling the tRNA anticodon loop to interact strongly with the mRNA codon. This interaction between the tRNAand mRNA explains the higher fall-off rate for near-cognate tRNA which has weaker interaction with mRNA.

Initial selection is followed by irreversible GTP hydrolysis. This hydrolysis causes the EF-Tu conformation to lose its affinity for the ribosome and tRNA. EF-Tu falls off the ribosome and the tRNA is accommodated to the PTC. During this accommodation process, the second step of tRNA selection, proofreading, takes place. Proofreading is also a consequence of the same induced fit mechanism since the forward reaction rate for the cognate tRNA is higher than for the near-cognate and backward reaction rate is slower for cognate tRNA and higher for near-cognate (44).
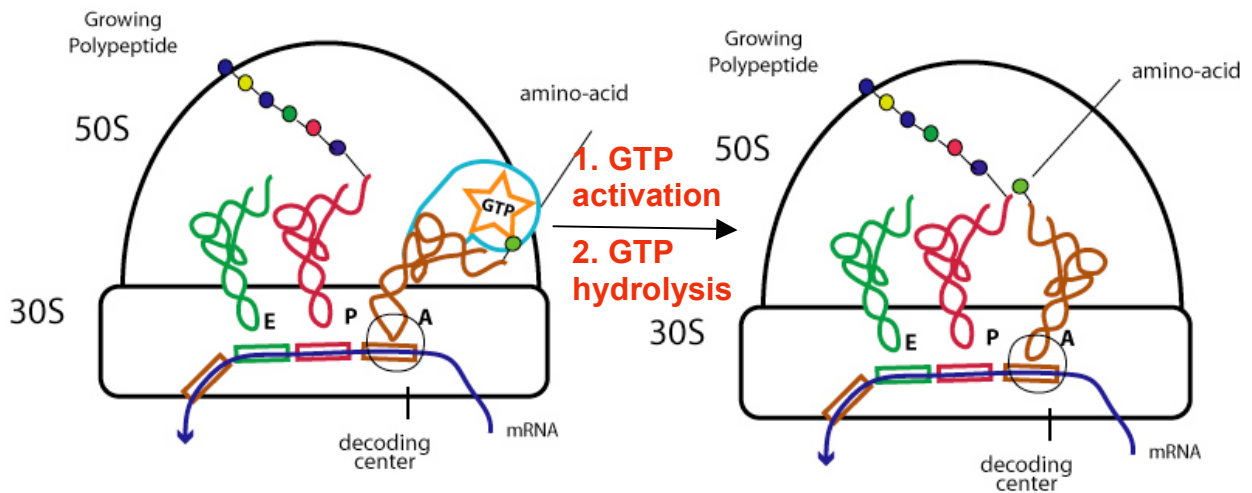
Figure 1.4: Schematic representation of accommodation. This is when the near-cognate tRNA that passes the initial selection get rejected. The brown tRNA initially at the A/T state (tRNA with anticodon bound at the 30S A-site and acceptor stem end bound with EF-Tu) swings its acceptor stem inwards to the peptidyl transfer center where the amino-acid it is carrying is added to the growing polypeptide.

13

Many hypotheses have been put forward to explain the structural basis of this proofreading. The most common hypothesis is that once the cognate tRNA/mRNA complex forms, a signal is transmitted from the decoding center to the functional sites of the 50S subunit (44). This signal transduction induces changes within the ribosome, which allows the cognate tRNA to pass through while rejecting the near-cognate tRNA. Many studies have been done to determine the signaling pathway. Many sites have been suggested to play a role in signal transduction including the tRNA, small subunit, GTPase activating center and L11 binding domain of LSU (46-50). However, even after many experiments and with the availability of the crystal structures the signaling mechanism has yet to be characterized.

In the present work, I will describe our approach to elucidating the structural basis of proofreading with the aid of two ribosomal models derived using cryo-EM densities and the crystal structures for the *E. coli* ribosome.

Structural studies of RNA viruses

Viruses are obligate parasites. All viruses use either DNA or RNA as their genetic material. The genome is encapsidated within a protein coat called the capsid, which may be either enveloped or non-enveloped by membrane. Based on their genetic composition, they can be broadly classified as either RNA or DNA viruses. RNA viruses and DNA viruses differ in their assembly pathways. In RNA viruses, the genome and capsid protein assemble together to form the virion. While, in most cases of the DNA viruses, the DNA genome is pushed inside a preformed capsid by a protein motor in an energy-dependent pathway.

Viruses are interesting particles, because they lie in the boundary between living and non-living. They need host cells to replicate. They enter the host cell and use the host cell's replicating machinery to replicate themselves. In the process they make thousand copies of themselves and destroy the host cell.

14

Studies of viruses are important because they are the causative agents of deadly diseases such as rabies, AIDS, SARS, influenza, etc. Understanding the structure of a virus could lead us to find a cure for the disease it causes, as we could relate the structure of the virus to its function. Knowing the structure-function relationship would help us identify the key structural features, which could be targeted with drugs and be used to alter the function of the virus. Viruses can also be used as model study systems for understanding fundamental biological processes exhibited by viruses such as replication and packaging of genome, etc., because they are simple systems. A detailed study of viruses can help us understand more complex systems.

Structural studies of RNA viruses started as long ago as the1930s when the Tobacco Mosaic Virus (TMV) was first crystallized (51). Theoretical studies helped elucidate the structure of viruses in the early 1950s (52, 53). Harrison et al. published the landmark paper on the structure of Tomato Bushy Stunt Virus (TBSV) at 2.8Å resolution in 1978 (54). Since then, a series of crystal structures of viruses have aided in the structural studies of viruses (55, 56).

Cryo-EM has been used to elucidate the structures of icosahedral viruses since the mid-1980s (16, 17). The advantage of icosahedral symmetry is that fewer particles are required for reconstruction. Thus, icosahedral viruses have been available in subnanometer resolution since the last decade (57, 58). Also, the whole three-dimensional structure can be obtained with cryo-EM while only an asymmetric unit may be obtained in the case of icosahedral viruses with x-ray crystallography. Fitting of x-ray crystal structures into the cryo-EM density maps has helped elucidate the complete structure of the viruses and in some cases understand the dynamic properties of viruses (59, 60).

In our case, we used the x-ray crystal structure available at 2.8Å resolution for Pariacoto virus (PaV) and the cryo-EM maps at 20Å resolution as structural constraints to model the residues missing in the crystal structure. Only 35% of the RNA and 88% of the protein can be determined from the crystal structure. Molecular mechanics with special

modeling techniques was used to generate all-atom models for the virus. The modeling was done initially with coarse-grained modeling, and the models were subsequently converted to all-atom models. These are the first all-atom models for a T=3 virus. With the aid of the models, we determined the location of the missing protein tails. We also have been able to postulate a new assembly pathway for RNA viruses.

Overview of the current studies

In the current study, I present our work on three important ribonucleoprotein complexes (RNPs). The first work on the *E. coli* ribosome was based on high resolution structural data. The crystal structure of the ribosome is at 3.5Å resolution and the cryo-EM is at 8.9Å. Combining these two data sets using YUP.scx helped us generate two all-atom ribosome models in PRE and POST accommodation states with a high level of confidence. Based on these two models, we hypothesized that the differential kinking of cognate versus near-cognate tRNA is the structural basis of proofreading. The second work on modeling the eukaryotic ribosome from *T. lanuginosus* is based on homology models obtained by using the *E. coli* crystal structures as templates, in addition to *ab initio* models for the rRNA of the expansion segments which were generated using MCSYM (61). From the complete model of the rRNA of the eukaryotic ribosome, we were able to determine the location of the expansion segments, the interaction of the expansion segments with each other, and their possible role in intersubunit contacts. The complete model, with the addition of the ribosomal proteins from our collaborators, may give insight into the role of the expansion segments in the increased fidelity of translation in eukaryotes. In the third study, for modeling PaV, no structural data were available for the missing protein tails and 65% of the genome that was within the dodecahedral cage was not resolved by x-ray crystallography. Even the secondary structure of the RNA was not available. However, using the molecular mechanics package, YAMMP, with low-resolution structural data as constraints, we were able to postulate all-atom models for

PaV. Based on the models, we proposed a new assembly pathway for RNA viruses. These structural studies, ranging from high to low resolution, show the beauty of molecular modeling and how molecular modeling can take data at any resolution and help illuminate the structure/function relationships of macromolecules in greater detail.

CHAPTER 2

THE STRUCTURE OF THE E. COLI RIBOSOME BEFORE AND AFTER ACCOMMODATION

Abstract

We present structures for the 70S *E. coli* ribosome, mRNA, and tRNAs for the states immediately preceding (PRE) and following (POST) accommodation. These all-atom models were derived by fitting the crystal structure of the ribosome into cryo-electron microscopy density maps for the two states. The inter-subunit cavity is a small space that provides an energetic barrier to motion of the A-site tRNA during accommodation. Surprisingly, the conformation of that cavity is nearly identical in the PRE and POST structures, even though the A-site tRNA shows dramatic conformational changes during accommodation, with the acceptor stem moving more than 50Å. There is no evidence of induced conformational changes in the ribosome that could serve as a signal from the decoding site to the large subunit, indicating the fidelity of the codon-anticodon interaction. We hypothesize that conformational differences induced in cognate *vs.* near-cognate tRNAs provide the structural basis of proofreading, and we propose experimental and computational tests for this hypothesis.

Introduction

The ribosome catalyzes the translation of genetic information from messenger RNA (mRNA) into the corresponding protein sequence with the help of transfer RNA (tRNA) and other cofactors. When different tRNAs compete for binding to a given mRNA codon, the free energy difference between a perfect codon-anticodon match *vs.* a near match is insufficient to account for the accuracy of translation *in vivo*. In 1974, working independently, John Hopfield and Jacques Ninio both proposed a resolution to

18

this dilemma, postulating that translational fidelity must be based on kinetic effects, rather than on thermodynamics alone (1, 2). They showed that the observed level of translational accuracy could be explained if the codon-anticodon match is tested twice, with the two tests being separated by an irreversible step. This mechanism – kinetic proofreading – is well established as the basis of translational fidelity, but the exact connections between the structural, thermodynamic and kinetic aspects of proofreading are not yet fully understood.

The ribosome is made up of two subunits, traditionally identified by their sedimentation coefficients. The small subunit (30S in *E. coli*) is responsible for monitoring the accuracy of the base pairing match between the mRNA codon and the tRNA anticodon. The large subunit (50S in *E. coli*) contains the peptidyl transferase center (PTC). This is the site where the polypeptide chain is transferred from the P-site tRNA and covalently linked to the amino acid on the A-site tRNA, thus growing one residue longer.

Marina Rodnina and her colleagues have examined the kinetic basis of translational fidelity (3-6). Figure 2.1 is a summary of their kinetic scheme, covering the steps from initial binding of the aminoacylated tRNA to the formation of a new peptide bond. The incoming aminoacylated tRNA is carried into the ribosome by elongation factor Tu (EF-Tu). As it binds to mRNA, the tRNA anticodon stem loop (ASL) makes contact with the A-site of the 30S subunit, while the acceptor stem is still bound to EF-Tu. Noncognate tRNAs fail to bind to the message, but both cognate and near-cognate tRNAs can bind in this configuration, which is often called the A/T state. (Since "A/T" refers only to the tRNA conformation, we prefer to call this the first reading state, emphasizing its place in the Hopfield/Ninio kinetic proofreading scheme.) Here the first test of the codon-anticodon match takes place, with two possible outcomes: the tRNA/EF-Tu complex either falls off the mRNA (k_{-2}) or activates the GTPase activity of EF-Tu (k_3). After the irreversible step of GTP hydrolysis, EF-Tu undergoes a

19
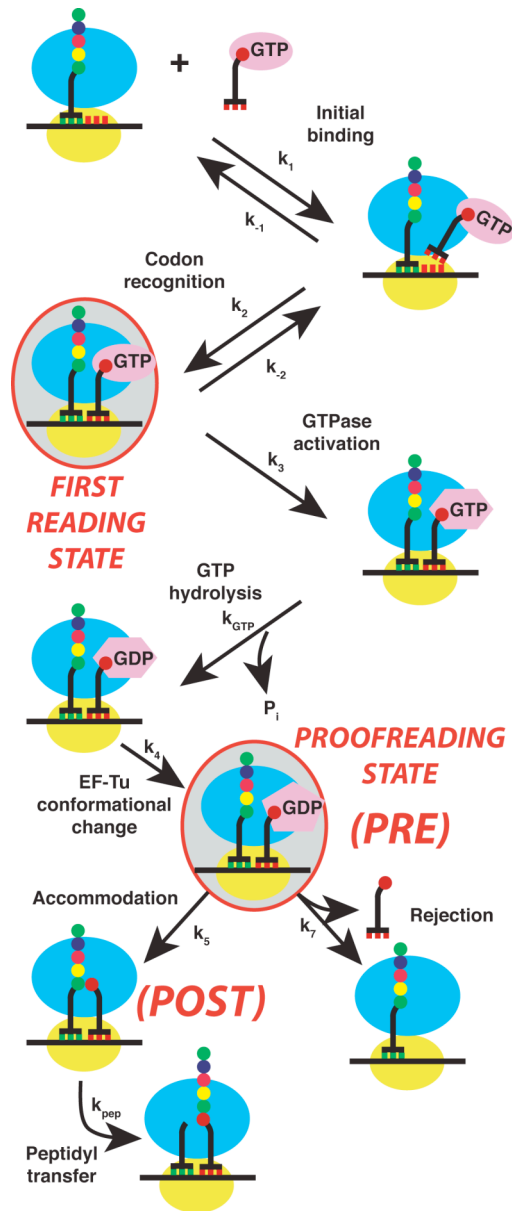
Figure 2.1: Kinetic scheme of Rodnina *et al.*
The ribosome is represented in yellow (small subunit) and blue (large subunit). The P- and A-site codons are shown as green and red triplets, respectively. The P-site is occupied by a tRNA with the growing peptide chain, while the incoming tRNA, charged with the next amino acid (red) is initially bound to EF-Tu/GTP (pink). The first test of the codon/anticodon match (the first reading) occurs right after codon recognition, when the incoming ternary complex either activates GTPase (k_3) or falls off of the message (k_{-2}). The second test (proofreading) comes immediately after the EF-Tu conformational change, when the tRNA either moves into the PTC (accommodation, k_5) or falls off the message and leaves the ribosome (rejection, k_7). Models generated in the present work represent the pre- and post-accommodation states (PRE and POST, respectively). The release of EF-Tu/GDP (k_6) is not shown.

conformational change, releasing the aminoacylated tRNA 3' terminus. The resulting proofreading state is where the second test of the codon-anticodon match occurs: the aminoacylated tRNA either falls off the ribosome (rejection, k_7) or swings through a distance of about 70Å to reach the PTC (accommodation, k_5), leading immediately to the irreversible formation of the peptide bond.

Two results of the kinetic studies are particularly important. First, after codon recognition, the next forward step (GTPase activation) has a larger rate constant (k_3) for cognate tRNA than for near-cognate tRNA, while the backward step (the dissociation of the aminoacyl tRNA/EF-Tu complex) has a smaller rate constant (k_{-2}) for cognate tRNA than for near-cognate tRNA; these differences must require an induced fit mechanism (5, 6). Second, the proofreading step shows similar differences in kinetic behavior, with a higher accommodation rate (k_5) for cognate tRNA than for near-cognate tRNA, and a lower rejection rate (k_7) for cognate tRNA than for near-cognate tRNA. Rodnina *et al.* concluded that induced fit must therefore also be important in proofreading. Independent measurements of the rate constants have confirmed that induced fit almost certainly plays a role in initial selection (7). These authors stated that their data "support" the proposed induced fit mechanism for proofreading as well, although they did not report measurements of the rate constants for accommodation and tRNA rejection.

The kinetic data have important implications for the energetics of the transition pathways for both the first reading and proofreading of the codon/anticodon match (Figure 2.2). The pathway branches at both of these points, and in both cases the energy barrier for the path leading toward peptide bond formation must be lower for the cognate tRNA than for near-cognate, while the reverse must be true for the alternative path, which leads to loss of the A-site tRNA.

What is the structural basis of the induced fit mechanisms in the two steps of proofreading? The first crystal structure of a 30S subunit revealed that the universally conserved nucleotides A1492 and A1493 change conformation to form hydrogen bonds

with a cognate codon-anticodon duplex (8). This was confirmed in later crystallographic studies (9), which dubbed this the "closed conformation", and by our cryo-EM study on the structure of aa-tRNA/EF-Tu/ADP (10). This conformational change in the small subunit is caused by the binding of tRNA to mRNA, and it stabilizes the codon/anticodon complex, so it is clearly important in the first reading of how accurately the anticodon matches the codon (3-5, 10).
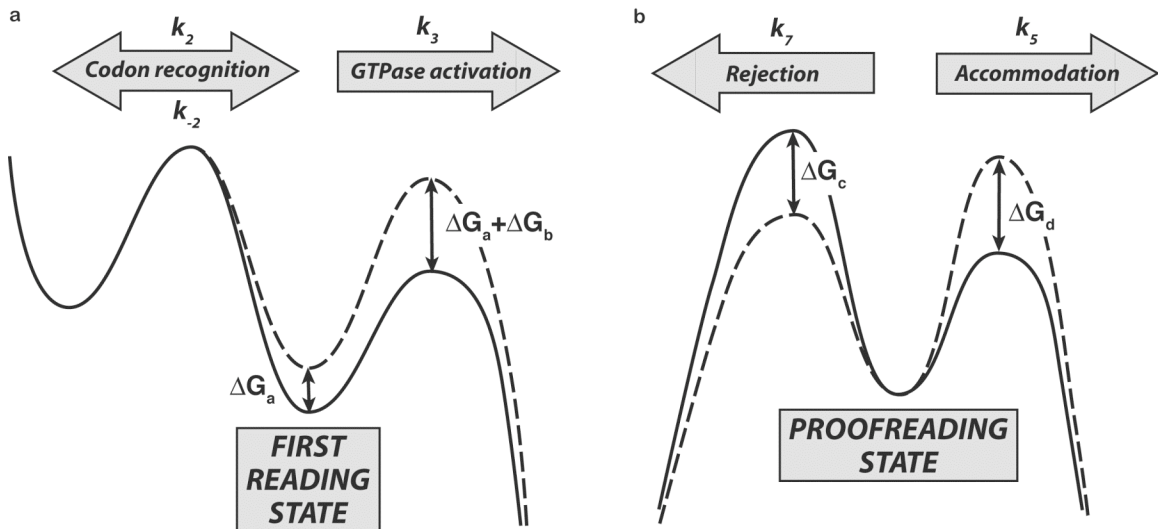
Figure 2.2: Energetics for the pathways associated with tests of the match between anticodon and codon
(a) First reading. Differences in codon-anticodon base pairing and structural changes induced at the ribosomal decoding site stabilize the cognate tRNA (solid line) more than near-cognate tRNAs (dashed line). Part of this stabilization (ΔG_a) occurs at the reading state, but kinetic data show that there must be an additional contribution *(ΔG_b)* as the tRNA passes over the energy barrier leading to GTPase activation, which is irreversible. In this panel, the transition state during codon recognition is taken as the reference state for energy differences. (b) Proofreading. Structural differences facilitate accommodation for cognate tRNA (solid line) relative to near-cognate tRNAs (dashed line), with an energy difference of ΔG_d. The energy barrier for rejection is higher for cognate tRNA than for near-cognate tRNA (ΔG_c). Both rejection and accommodation are irreversible. There is an energetic difference between the cognate and near-cognate cases in the proofreading state, but the ground state differences cannot cause kinetic differences, so the curves have been shifted to coincide at that state, making it the reference state for comparing energy differences.

The present investigation is aimed at a careful examination of the ribosome structure immediately before and after accommodation, with the aim of looking for hints

about the structural basis of the kinetic differences in the behavior of cognate and near-cognate tRNAs during the second test of the codon/anticodon match, *i.e.*, proofreading. Although we cannot catch the ribosome at the transition state of accommodation, we might expect there to be structural differences between the PRE and POST states, particularly in light of the very large motion of the tRNA during accommodation. Any tRNA-induced conformational changes should be related to the structural basis of translational fidelity. Then the task would be to determine how these conformational changes in the ribosome might be exploited at the transition state to lower the energy barrier for cognate tRNA relative to near-cognate tRNAs.

We have built two models, placing the recently determined crystal structures of the *E. coli* 70S ribosome (11, 12) into the cryo-EM density maps for the states immediately prior to and immediately following accommodation (10), then using a flexible fitting algorithm to optimize the agreement between the models and the density maps. Throughout the modeling, we have paid careful attention to maintaining correct stereochemistry. This has produced the first rigorous all-atom models for the ribosome, mRNA and the A-, P- and E-site tRNAs immediately before and immediately after accommodation. The models offer the opportunity to search for significant conformational changes in the ribosome that are induced by the motion of the cognate tRNA during accommodation and that might serve as signals related to the fidelity of the codon-anticodon match.

Results

There are two ways to look for differences between the pre- and post-accommodation states. First, we can superpose the two structures and measure the root mean square deviation (RMSD) between the positions of a given residue in the PRE and POST states. A plot of RMSD *vs.* residue number will then identify those regions whose positions differ most in the two states, *i.e.*, those regions that have undergone the largest

motions. RMSDs help identify global conformational changes, telling us, for example, whether a pair of neighboring residues i and j move a little or a lot during accommodation. But RMSDs do not tell us whether those two residues have moved in the same direction, maintaining the same interaction in the PRE and POST states, or whether they have moved in different directions, making or breaking a critical contact. To evaluate relative motions and search for changes in important contacts, we examine the changes in inter-residue distances, reported in the $N \times N$ *distance difference matrix*, where N is the number of residues. The distance difference between the PRE and POST states for residues i and j is

$$\Delta d_{ij} = d_{ij}^{post} - d_{ij}^{pre}$$

In the search for significant conformational changes, there are two ways to rank distance differences. First, they can be ranked by absolute magnitude. This ranking gives results that are dominated by changes in the relative positions that are far apart in three-dimensional space. For example, the large motions of proteins L1 and L11 and the corresponding RNA domains (L1BD and L11BD) means that if residue i is part of one of these regions, Δd_{ij} will be very large for all residues j that are not part of the same region. Such differences may or may not be functionally significant. A second way to rank distance differences is by the absolute percentage difference. These will be sensitive to local changes and should identify any pairs of residues that are in contact in one conformation but are not in contact in the other.

Deviations from the crystal structure are greatest for peripheral regions

Before examining the PRE and POST models to identify differences between them, we have compared each of them to the crystal structure. Tables A.1 and A.2 summarize two aspects of the protein motions. First, by superposing the crystal structure and the PRE state for a given ribosomal subunit, the global conformational changes can

be examined in terms of the displacement of each protein. (A similar analysis is reported for superposition of the crystal structure and the POST state.) Second, superposition of the crystal structure and the PRE (or POST) structure for a given protein reveals whether or not it has undergone any major conformational changes.

The global conformational changes cause substantial changes in the positions of some of the proteins, particular L11 (see next paragraph). On the other hand, the intramolecular protein conformational changes are all small, with one exception: protein L9 is extended away from the surface of the ribosome in the crystal structure, but it has a more compact conformation on the ribosomal surface in both the PRE and POST states.
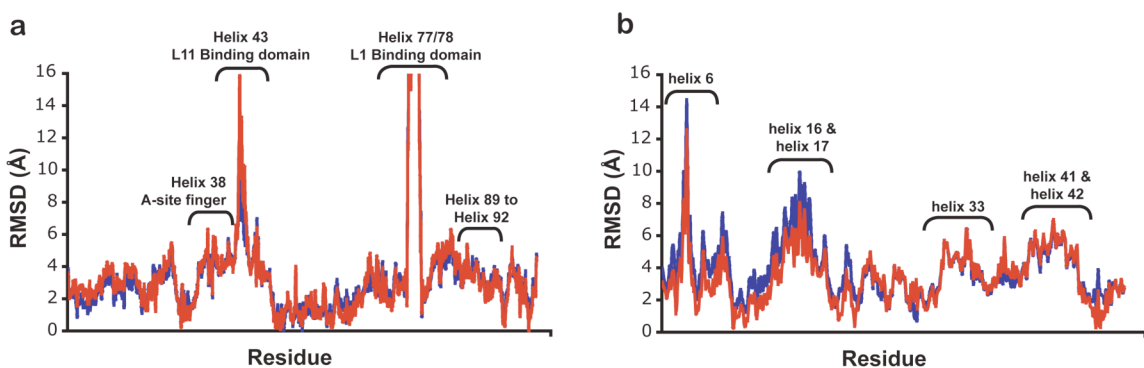
Figure 2.3: Residue-by-residue root mean square differences between the crystal structure and the structures of the PRE (blue) and POST (red) states. (a) 23S RNA. (b) 16S RNA.

Since the 16S and 23S RNAs extend throughout the small and large subunits, respectively, the global conformational changes are accompanied by substantial intramolecular motions for these RNAs. These are analyzed in Figure 2.3 and their magnitudes are mapped onto the three-dimensional structure in Figure 2.4. Overall, the largest deviations occur at the peripheral regions of the ribosome. In the large subunit, the L1 binding domain (L1BD) and the L11 binding domain (L11BD) show the most deviation. Both of these conformational changes are part of the initial binding of the ternary complex to the interface region: the elbow of the tRNA makes contact with the L11BD, while the L1BD moves inwards and makes contact with the E-site tRNA (Figure

2.4a). The latter result is consistent with the observation that the L1 stalk moves towards the E-site upon binding of tRNAs, possibly stabilizing the tRNA/ribosome complex (13, 14).
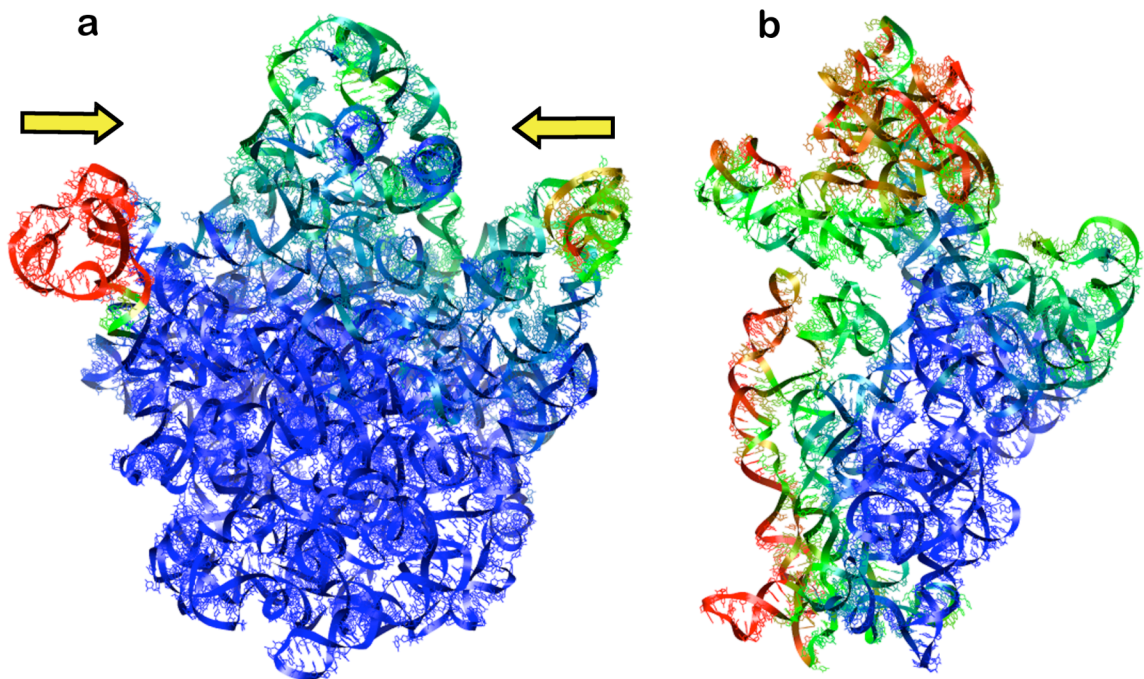
Figure 2.4 Motions in the ribosomal subunits, crystal structure *vs.* the PRE state. (a) Interface view of the large subunit, showing directions of motion for the L1BD (left) and L11BD (right). Color code: RMSD < 3.3Å (blue); 3.3-6.7Å (green), 6.7-10Å (orange) and >10Å (red). (b) Interface view of small subunit; color code: RMSD < 2Å (blue), 2-4Å (green), 4-6Å (orange) and >6Å (red).

In the 30S subunit, the spur (SH6) has the largest difference between the crystal structure and the current structures (Figs. 2.3b and 2.4b). There is substantial motion where the beak (SH33) meets the shoulder (SH33). Other regions in the head show smaller motions (SH41-42). The proteins in the both subunits move with the corresponding RNA helices, and there are no significant changes in RNA-protein contacts. (As in earlier papers, we use the abbreviation "LH" to designate helices in the large subunit, and "SH" for those in the small subunit.)

Comparison of pre- and post-accommodation ribosome structures

When examined quantitatively, the differences between the PRE and POST states are much smaller than the differences between PRE and the crystal structure, or between POST and the crystal structure. As seen in Tables A.1 and A.2, protein displacements are all very modest during accommodation, except for L11 (whose motion we will discuss presently), and the conformations of all proteins are essentially identical in the PRE and POST states.

The conformational changes within the 16S, 23S and 5S RNAs are correspondingly small. Figure 2.5 provides a residue-by-residue analysis of RMSDs within the 16S and 23S RNAs during accommodation. In the large subunit, the PRE *vs.* POST average RMSDs are about 1Å, and the differences never exceed 8Å. None of the differences over 2Å are near the regions of the interface contacted by the A-site tRNA during accommodation. By comparison, average RMSDs are about 4Å when comparing the crystal structure with either PRE or POST, with peaks of more than 15Å. The largest motions in the large subunit during accommodation are those of L11, the L11BD and the L1BD. tRNA in the A/T conformation makes contact with the L11BD in the pre-accommodation state, possibly straining it, while in the post-accommodation state the tRNA has moved away, allowing the L11BD to relax. Thus, although the L11BD does have a function in initial selection, it presumably plays no role in accommodation once the tRNA leaves the A/T state and begins to swing toward the peptidyl transferase center. The L1BD is far away from the A-site tRNA both before and after accommodation, so it cannot play a direct role in proofreading either.

Similarly, the PRE *vs.* POST RMSD differences in the small subunit are all much smaller than differences between either of these structures and the crystal structure. All RMSD differences for PRE *vs.* POST are less than 4Å (Figure 2.5b), although both states have regions that move by more than 10Å when they are compared to the crystal

structure (Figure 2.3b). Here, too, differences over 2Å are confined to regions distant from those parts of the decoding center and interface that are contacted by the A-site tRNA during accommodation. The highest RMSD value is less than 4Å, and this occurs in SH16 and SH17 (the shoulder), far from the decoding site and far from tRNA contact points on the large subunit.
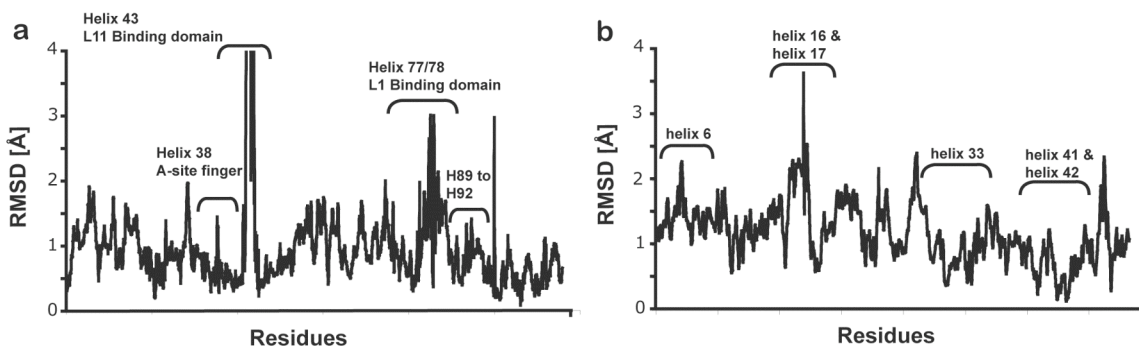
Figure 2.5: Residue-by-residue root mean square differences between the PRE and POST states. (a) 23S RNA. (b) 16S RNA.

Overall, the conformation of the ribosome changes very little between PRE and POST, as the tRNA moves from the A/T to A/A conformation.

The most important question about PRE *vs.* POST is whether or not there is any evidence for the transmission of a signal between the subunits, because examination of the codon/anticodon interaction takes place on the small subunit, while the principal barriers to tRNA motion during accommodation come from collisions with the large subunit. In particular, we might expect there to be noticeable changes in the structures of the inter-subunit bridges as a consequence of the very large motion of tRNA from the A/T to the A/A state. To look for such changes, we have measured the distances between all pairs of residues ij, where residue i is on the small subunit and residue j is on the large subunit; the distance d_{ij} was measured in both the PRE and POST states, and then we constructed a matrix of distance differences, $\Delta d_{ij} = d_{ij}(\text{POST}) - d_{ij}(\text{PRE})$. This matrix was sorted in two ways, first by the magnitude of Δd_{ij}, and second by the magnitude of the percentage difference, $100\Delta d_{ij}/|\langle d_{ij}\rangle|$, where $|\langle d_{ij}\rangle|$ is the average of $d_{ij}(\text{PRE})$ and

d_{ij}(POST). In these calculations, we considered only those pairs where i and j are within 30Å of one another, to search for significant reorganization of any intersubunit bridge. Out of the four largest changes (Table A.5), three are remote from the decoding center and the PTC. The largest percentage change (37%) occurs between residues 1409 in SH44 and 1914 in LH69, but the magnitude (3.4Å) is suspiciously small, considering the resolution of the data. Inspection of the electron density maps and the models in this region does not show convincing evidence of a conformational change. Visual inspection shows that this is not part of a significant structural reorganization.
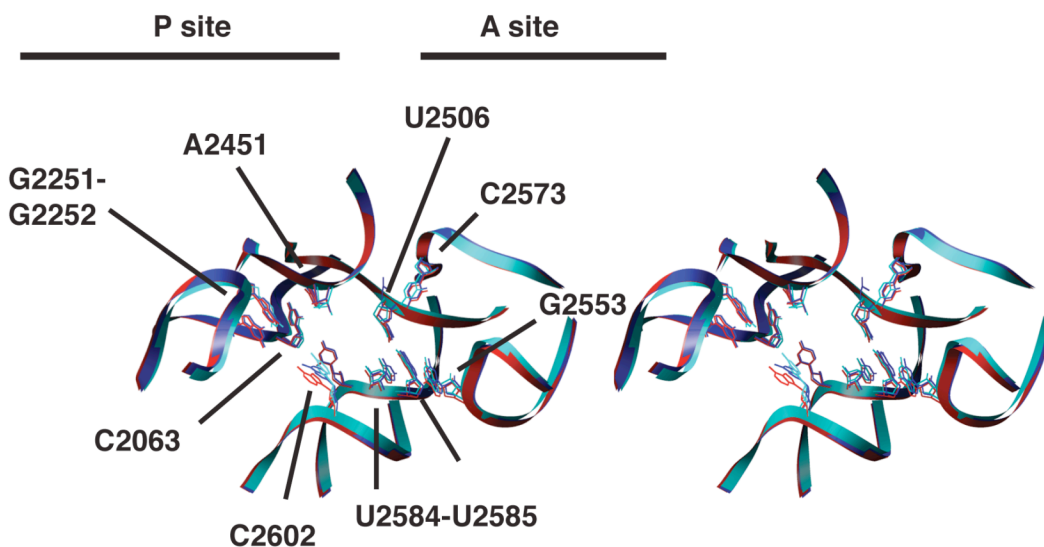
Figure 2.6: Stereo view of the structure of the peptidyl transferase center for the crystal structure (cyan), the PRE state (red) and the POST state (blue). Note all three are essentially identical.

A similar analysis has been used to look for significant conformational changes within either the large subunit (Table A.6) or the small subunit (Table A.7). In the small subunit, the maximum distance differences are less than 5Å, and these are all associated with pairs of residues that are quite far apart and/or quite far from the decoding site. As expected, the most significant differences in the large subunit are associated with the motion of the L11BD discussed above. The distance difference analysis does pick up a 3.7Å change between residues 2056 and 2505, both of which are near the PTC. Once

29

again, an examination of the electron density maps does not show any evidence of a significant conformational change, and a visual inspection of the PTC in the two models does not show evidence of transmission of this small difference to other regions (Figure 2.6).

Rigidity of regions of the ribosome contacted by tRNA during accommodation

The ribosomal regions surrounding the tRNA accommodation pathway are essentially identical in the PRE and POST states (Figure 2.7). The only region near the central cavity between the subunits that shows any significant motion is the L11BD, and, as argued above, this is due to release of the tRNA before accommodation, so this motion cannot play any structural role in proofreading.
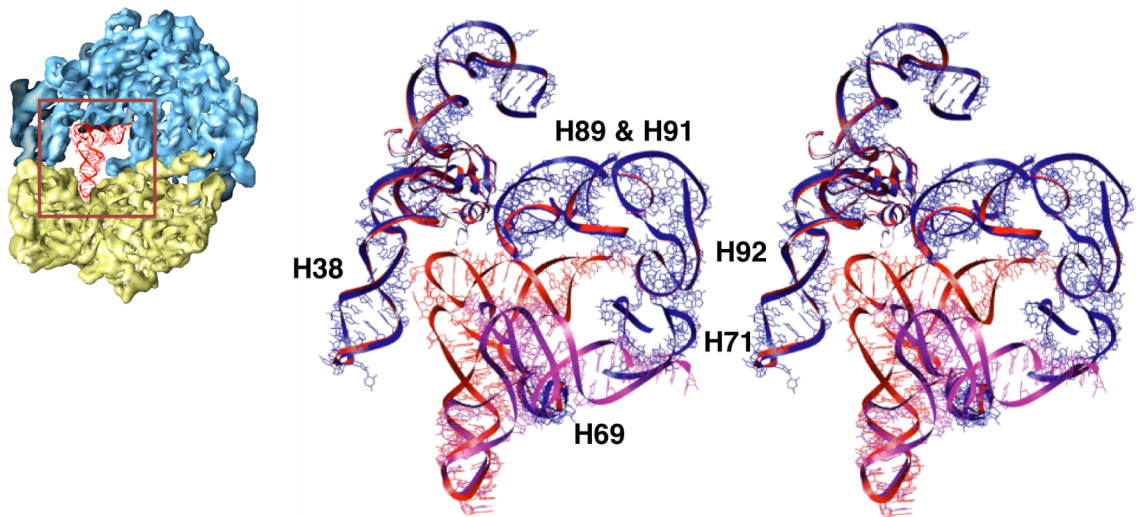
Figure 2.7: Comparison of PRE and POST states for tRNAs and large subunit residues contacted by the A-site tRNA during accommodation
A large motion takes tRNA from the A/T (magenta) to the A/A (red) state. PRE and POST states of the 23S RNA are shown in blue and red, respectively. The conformation of the 23S RNA is essentially fixed, defining a cavity that impedes diffusion of tRNA from the A/T state to the A/A state. The orientation of the view is shown in the cutaway thumbnail figure at the upper left, in which the 30S and 50S subunits are yellow and blue, respectively.

Rigidity of the peptidyl transferase center, and flexibility of tRNA 3'-terminus:

Residues around the peptidyl transferase center have nearly identical conformations in the crystal structure, the PRE state, and the POST state (Figure 2.6). The overlap is nearly perfect, suggesting that the active site of the ribosome is fairly rigid throughout and does not change between the states without tRNA in the PTC (crystal and PRE), and the state where tRNA occupies the PTC (POST). This is consistent with the observation that the catalytic activity of isolated 50S subunit is quite similar to that of the intact 70S ribosome when puromycin is used as the A-site substrate and full-size fMet-tRNA$^{\text{fMet}}$ is used as the P-site substrate (15), a result that led the authors to conclude that no essential conformational changes at the PTC are required for peptide bond formation, nor is any signal transmitted from the decoding site to the PTC. The absence of conformational changes at the PTC is not surprising, considering the discovery that the 50S subunit catalyzes peptide bond formation entirely by bringing the two substrates together, without an active chemical contribution (16, 17).

Finally, we note that the structure of the end of the acceptor stem and the 3'-terminus of A-site tRNA in the POST state differs considerably from that of tRNA in the crystal structure, even though other regions of the tRNA are essentially identical to the crystal structure. In contrast, the crystal structure of tRNA$^{\text{Phe}}$ fits very well into the cryo-EM density of the P- and E-site tRNAs, without significant deformation.

Discussion

The most important result of this work is that the inter-subunit space does not undergo any significant conformational changes during accommodation. This is rather surprising, considering the very large displacement of tRNA from the A/T to the A/A state. How can there be any difference between the accommodation rate constants (k_5) for the cognate and near-cognate tRNAs, in light of the apparent ribosomal rigidity during

this step of the translational cycle? There are no apparent differences in the ribosome structure between the PRE and POST states that could be exploited by the cognate tRNA to distinguish it from near-cognate tRNAs. This is in contrast with the first reading of the codon-anticodon interaction, where the decoding site responds to the structure of the mRNA/tRNA complex, changing conformation during codon recognition (8-10). We are forced to consider the possibility that the ribosome itself does not send a signal between the decoding site and the large subunit that is used to differentiate between cognate and near-cognate codon/anticodon pairing during proofreading.

If we are to preserve the concept of induced fit implied by the kinetic studies, we must revise our thinking a bit. Koshland originally proposed that binding of the substrate induces a conformational change in the enzyme, and that the resulting conformation positions the enzyme's catalytic groups properly (18). The ribosome's conformation does not change appreciably during accommodation, so we must expand the definition of "enzyme" to include the molecule whose conformation does change dramatically, the tRNA. We originally proposed that tRNA unkinking drives accommodation (10), and we now hypothesize that cognate and near-cognate tRNAs have different degrees of kinking in the A/T state. We propose that kinking is differentially stabilized by the combination of the differences in the codon/anticodon interactions and the conformational changes subsequently induced at the decoding site. The result is that cognate tRNA is more kinked than near-cognate tRNA. More conformational energy is stored in the former than the latter, and the difference contributes to ΔG_d. If this hypothesis is correct, the stored conformational energy difference is used to help tRNA over the transition state energy barrier as it diffuses from the A/T conformation to the A/A conformation.

The tRNA must squeeze through a narrow space to reach the peptidyl transferase center (Figure 2.7). The acceptor/T-stem has to pass between the A-site finger (LH38) and helix 92, which terminates in the A-loop. The acceptor/T-stem is about 65Å long, measured from residues ψ55 and C56 in the elbow to A76 at the 3'-terminus, but the

space between the ASF and LH92 is less than 60Å long. Table A.3 shows the interphosphate distances between LH38 and LH92, several of which are less than 60Å. The shortest interatomic distance between LH92 and the ASF is 57Å, from O1P of residue 2558 to O2' of residue 897. While working its way through this space, the acceptor/T-stem is also restricted above (in the view of Figure 2.7) by LH89 and LH91, and below by LH69, which serves as a sort of "rail" inside the elbow, guiding the tRNA during accommodation. Many of the nucleotides in these regions of the 23S RNA are highly conserved.
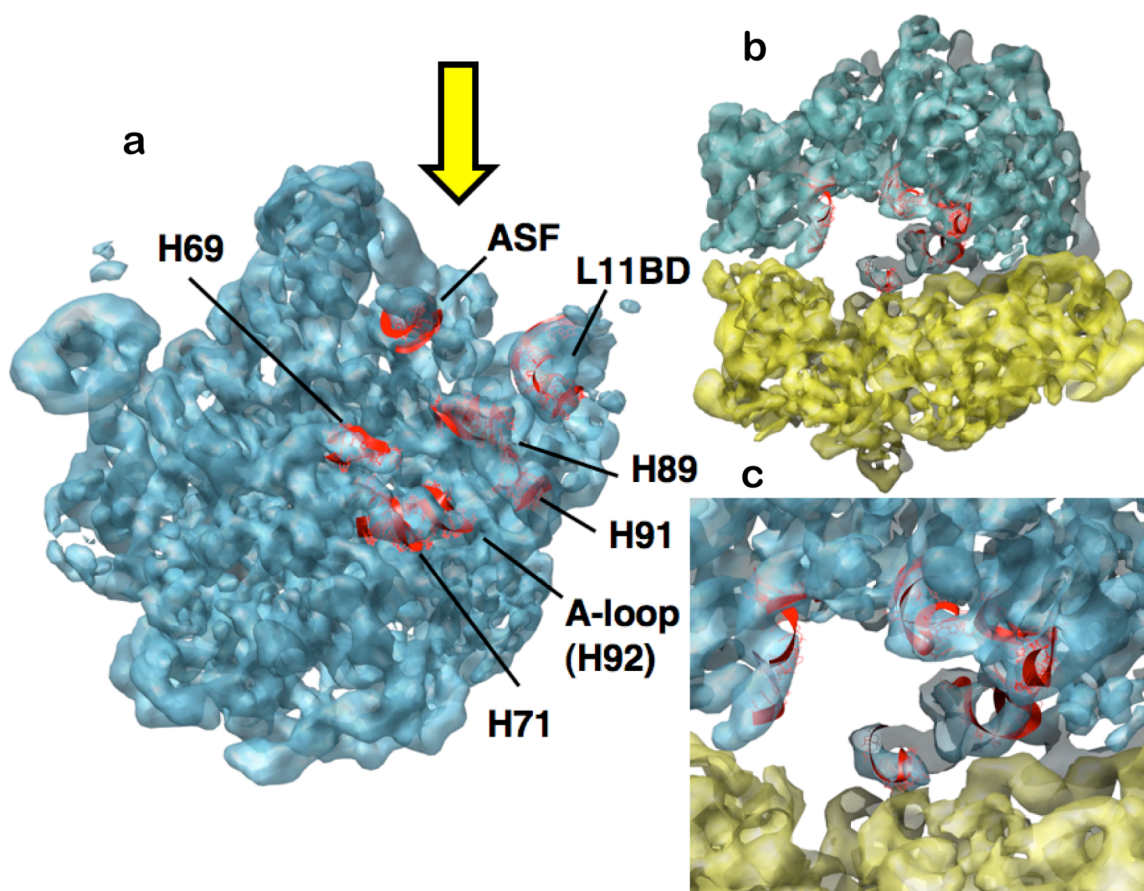
Figure 2.8: tRNA contact map on the large subunit, inferred from our simulation on free tRNA, fit post hoc onto the A-codon and into the ribosome. 50S subunit in blue, 30S subunit in yellow. (a) Interface view. (b and c) View downward into the central cavity from the direction of the arrow in (a).

The Los Alamos group reported a very interesting series of targeted molecular dynamics (MD) simulations (19), identifying ribosomal regions that the tRNA probably

33

contacts during accommodation. They found that the A-loop at the end of LH92 constitutes a substantial barrier to the motion of accommodation, and that the tRNA also interacts with LH38 (the A-site finger), LH69, LH71, LH89, LH91 and the L16 protein at different stages during accommodation. As a complement to those simulations, we have carried out MD simulations examining the unkinking of a model for the A/T tRNA (Caulfield et al., unpublished). Our simulations were done on tRNA bound to mRNA in a ribosome-free model, to ask what the preferred motions of the tRNA would be if it were not hindered by the ribosome. By superposing the codon/anticodon complex from each structure in the simulation onto the codon/anticodon position at the A-site in the ribosome, we were able to identify the regions on the ribosome that would probably hinder tRNA swinging during accommodation. The putative contact regions identified this way are shown in Figure 8; they are very similar to those observed by Sanbonmatsu *et al.* In particular, our MD simulations support their conclusion that the A-site finger, LH71 and residues 2558-2561 from the A-loop are particularly important in hindering accommodation. Furthermore, the average RMSD between the PRE *vs.* POST structures is less than 2.0Å in these regions (Figure 2.5 and Figure 2.7), which further emphasizes their probable role in defining the barrier to tRNA motion during accommodation.

Clearly, some deformation of the tRNA is required to permit accommodation. This view is supported by the molecular dynamics simulations, in which the motion of the acceptor stem was impeded by helix 92, leading to the conclusion that "flexibility of the acceptor stem … is essential for tRNA selection" (19). Although those authors described the barrier created by helix 92 as "surprising", we now suggest that it is an essential component of tRNA selection: the energy required for tRNA deformation during accommodation constitutes the activation energy barrier, and we hypothesize that cognate tRNA has more deformation energy stored in its kink than does near-cognate tRNA. This difference (ΔG_d, Figure 2.2b) would preferentially facilitate accommodation for cognate over near-cognate tRNA.

It has been estimated that cognate tRNA is bound to the small subunit about 100 times (5) to 300 times (20) more tightly than near-cognate tRNA. This difference (ΔG_c in Figure 2.2b) explains why the rejection rate is higher for near-cognate tRNA. By itself, however, it cannot simultaneously account for the slower rejection rate and the faster accommodation rate of cognate tRNA. Tighter binding stabilizes the ground state of the reaction (the proofreading state), so that both the forward and backward barriers in Figure 2.2b would be higher if this were the only difference between the cognate and near-cognate cases. A separate induced conformational change is required, and tRNA kinking fills this need (ΔG_d).

How much energy is stored in the kinked tRNA conformation? The kink deformation arises predominantly from motions that twist the stacked D-stem / anticodon stem, with a total twist of about 22° spread over the five central base pair steps (10). For small twist deformations of RNA double helices, the energy grows quadratically with twist angle, with a twisting force constant k_θ of about 0.05 kcal/(mol•degree^2) (21). The kink energy is $(k_\theta/2)(\Delta\theta)^2$ for each of the five base pair steps, with $\Delta\theta$ ~4.4°. This gives a total energy of about 2.4 kcal/mol for the twist deformation of the cognate tRNA. (The costs of the deformation are paid for by the energy released by binding the EF-Tu•GTP•tRNA ternary complex to the ribosome and by base pairing on formation of the codon/anticodon complex.) If our hypothesis is correct and near-cognate tRNA has less kinking than cognate tRNA, the energy difference for kinking cognate *vs.* near-cognate tRNA should represent a substantial fraction of ~2.4 kcal/mol. The difference in relative rates of accommodation of cognate *vs.* near-cognate tRNAs is on the order of 7× to 100×, which corresponds to an activation free energy difference of 1.1–2.7 kcal/mol for ΔG_d (more properly called $\Delta\Delta G_5^{\neq}$) (22). This represents about 45–100% of the kinking energy for cognate tRNA. Considering the simplicity of the elastic model, this is a remarkable degree of agreement.

This hypothesis is consistent with arguments made many years ago that tRNA flexibility probably has functional consequences (23, 24) and that tRNA is more than a passive partner in translation (25, 26). Since our discovery that tRNA is kinked in the A/T conformation during the PRE state (10), other experiments have suggested a role for tRNA flexibility in translation. D-stem mutations have long been known to affect fidelity (26, 27). Recent studies on one of these, G24A in the D-stem of tRNA^{Trp} (the Hirsh mutation) showed that it changes the rate constants for both GTPase activation (k_3) and accommodation (k_5), leading to the argument that these are manifestations of changes in tRNA flexibility (28). This adds plausibility to the argument that differential kinking in cognate *vs.* near-cognate tRNAs in the PRE state is the structural basis of proofreading.

To complete the kinetic and structural explanation of translational fidelity, this kind of analysis also suggests how induced fit can account for more favorable reaction rates for the cognate tRNA during the first reading (k_3 and k_{-2}, Figure 2.1). The well-known induced conformational changes at the decoding site, involving A1492, A1493, G530 and protein S12, are all complete by the time the structure reaches the reading state, which accounts for ΔG_a (Figure 2.2a). What then accounts for ΔG_b? The conformational changes in the ribosome that comprise the induced fit response must have two separate components, and ΔG_b arises from the second. GTPase activation is more rapid for cognate than near-cognate tRNAs, and these kinetic differences are reflected by the difference in the height of the activation energy barrier. The conformational changes leading to GTPase activation must be different for the cognate and near-cognate cases (ΔG_b), and this sensitivity to the substrate is the second part of the induced fit response in the first reading. A viewpoint similar to this has been advanced by Rodnina and coworkers (3), but the structural basis of these energy differences during GTPase activation remains to be determined.

The static models that we have generated for the pre- and post-accommodation states should provide a framework for additional work towards understanding

proofreading. The landmark MD simulations by the Los Alamos group (19) suggested that passage of the 3'-acceptor terminus of the aa-tRNA complex through a gate in the A-loop constitutes an important barrier to accommodation; although those authors did not say so explicitly, it is possible that this event represents the transition state. Those simulations were based on a homology model, because no crystal structure was available for the 70S ribosome. We believe that those simulations are of sufficient significance that they merit repeating with improved structural models and alternative biasing methods, with an eye to either replicating or challenging the earlier results. The models presented here are based on an *E. coli* crystal structure that was not available when Sanbonmatsu *et al.* began their work. Additionally, since the timescale for accommodation is far beyond the reach of free MD, those authors had to use a biasing algorithm to force the tRNA to move from the A/T to the A/A state. They used the targeted MD method (29). It would be worthwhile to determine how sensitive inferences about accommodation are to biasing, by comparing Sanbonmatsu's results with those based on other biasing methods, such as steered MD (30, 31), Maxwell's demon MD (32) or the Paci and Karplus algorithm (33).

Our model for the structural basis of fidelity during proofreading can be tested both experimentally and computationally. The model predicts that near-cognate tRNAs will have a smaller degree of kinking than cognate tRNA in the PRE state. This should be observable if a ternary complex containing EF-Tu, GDP and a near-cognate tRNA can be captured by x-ray crystallography or cryo-EM in the kirromycin-stalled state. Alternatively, the model could be tested by comparative computer simulations on accommodation of cognate and near-cognate tRNAs, following the protocols outlined in the previous paragraph. If such tests confirm the predicted differences in tRNA kinking, the structural and kinetic basis of translational fidelity will stand fully revealed.

Experimental Procedures

We built and refined the models in three stages. First, we docked crystal structures for the ribosome, tRNAs and mRNAs into the cryo-EM density maps. Second, we used an all-atom real-space flexible fitting refinement algorithm to make the conformational changes required to optimize the fit of the models in the maps. Third, following a few final manual adjustments, modest energy minimization was done to guarantee that the bond lengths and angles fall into standard ranges, and to eliminate any residual van der Waals overlaps, giving a stereochemically correct final model.

Ribosome subunit models

The *E. coli* crystal structures for the 30S subunit (2I2P.pdb) and the 50S subunit (2I2V.pdb), separated from the complete 70S crystal structure (11) served as starting structures for developing models of the ribosomal subunits in the PRE and POST states. The large subunit crystal structure is missing a portion of the A-site finger (ASF), residues 879-897, and a portion of the L1BD (residues 2111-2132 and 2158-2178). The missing portion of the ASF was modeled using MCSYM (34) and the crystal structure of the L1BD (35) was used as a template for the homology model of the missing piece of the rRNA in that region. Homology modeling of the missing pieces was done using the protocol described previously (36).

tRNA and mRNA models

The ribosome crystal structure lacks tRNAs and mRNA. The crystal structure of tRNA[Phe] (1EHZ.pdb) was used for the initial rigid fit into the density of all tRNA sites except for the A/T state, for which we used our previous model for tRNA in the A/T state (10). mRNA from the crystal structure of the complete ribosome from *T. thermophilus* (14) was used as the starting model for mRNA.

Density maps and initial fits

The 70S density maps for the pre-accommodation (PRE) and post-accommodation (POST) states (10) were separated into 30S, 50S and tRNAs using SPIDER (37). The maps are available from the electron microscopy database (EMD) as EMD-1055 and EMD-1056 for the PRE and POST states respectively. The crystal structures for the 30S, 50S and tRNAs and mRNA were fitted separately into the corresponding density maps using the rigid body fitting module in Chimera (38). The orientation of each tRNA anticodon with its corresponding mRNA codon was compared with the tRNAs in the *T. thermophilus* crystal structure (14) using the superposition module from the Insight-II software package (Molecular Simulation Inc., San Diego, CA). Manual adjustments of some of the bases were needed to get the correct alignment. We also made some small manual adjustments at the 3' CCA end of the tRNAs in both models, to improve the fits.
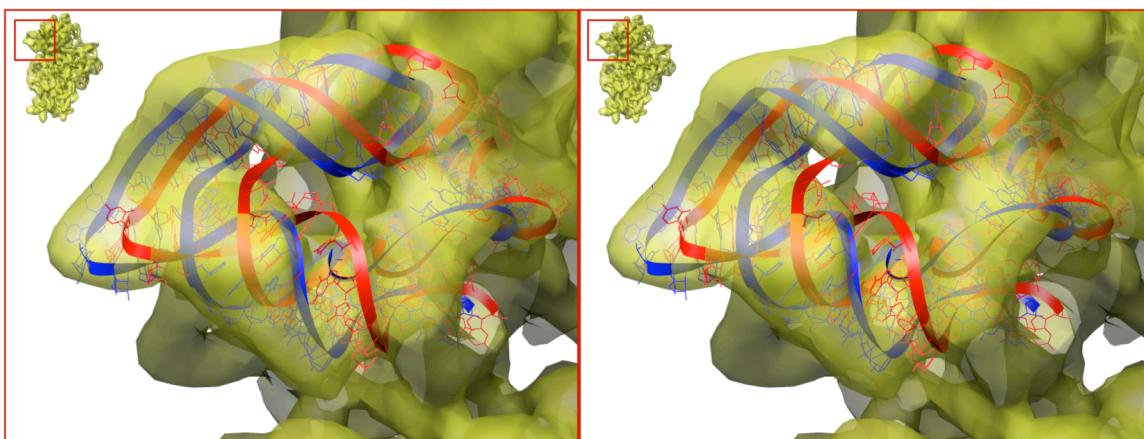
Figure 2.9: Stereo view of the YUP fit for the beak region of the 16S RNA (blue) into the cryo-EM map for the PRE state. The position after globally fitting the crystal structure to the map, but before YUP optimization, is shown in red.

Optimization of the models by flexible fitting

After the initial placement, the fit into the density was optimized with a flexible fitting algorithm using the Emmental sub-package of YUP, our in-house molecular mechanics package (http://rumour.biology.gatech.edu/YammpWeb) (39). The Emmental

module represents the structure by a Gaussian Network Model (GNM). The energy function for the flexible fitting contains terms for scoring the quality of the fit of the model to the density map, plus restraint energies for the GNM and volume exclusion terms, as described below. The optimization protocol uses simulated annealing based on low-temperature molecular dynamics (MD). The model is initially heated up to 10K, then held at 10K for equilibration, then cooled while reducing r_o (see below). The number of steps in each stage is proportional to the number of atoms in the starting model.

Energy terms used during flexible fitting

Density scoring term: E_{SCX}: When we consider the optimal positions of atoms in the model, the peaks in the density map should correspond to wells in the scoring function. In the SCX energy function, cubic voxels of SPIDER density maps are converted to equal-sized spherical wells, with radius r_i. The outer radius of the well (r_o) is the distance up to which any atom feels the force from the density well and is a dynamic parameter during simulated annealing. Large values of r_o tend to smooth out the density map and are used in the initial stages of refinement, while smaller values reproduce the detailed fluctuations in the map. The SCX score is:

$$E_{SCX} = \sum_{p}^{Nwells} \sum_{q}^{Natoms} \begin{cases} -E_0 \omega_q & r_{pq} \le r_i \\ \dfrac{E_0}{(r_o - r_i)^3}(r_o - r_{pq})^2(3r_i - 2r_{pq} - r_o)\omega_q & r_i < r_{pq} < r_o \\ 0 & r_{pq} \ge r_o \end{cases}$$

where ω_q is the density of each voxel normalized to a unit scale, r_{pq} is the distance between atom p and well q, and E_0 is the magnitude of the minimum energy. The latter is a user-defined parameter to adjust the balance between E_{SCX} and the other terms. We used a ratio of 3.5 for r_o/r_i and a value of $E_0 = 0.6$ kcal/mol. The rate of convergence is sensitive to these parameters, but the final models are not.

Gaussian Network restraints: E_{Hooke}: In contrast with other Gaussian network models, which use a coarse-grain approximation, the YUP GNM connects all pairs of non-hydrogen atoms in the structure that lie within a specified cutoff distance D_c (4Å). Each pair is assigned a unique equilibrium bond length b_0 that is identical to the distance when the network is formed. These are enforced with harmonic restraints with a uniform force constant (k_H) of 6 kcal/(mol•Å^2), giving an energy

$$E_{Hooke} = \sum_{r_{ij} < D_c} k_H \left(r_{ij} - b_0 \right)^2$$

where r_{ij} is the distance between atoms i and j.

Soft sphere term for volume exclusion: A non-bond term is required to prevent interpenetration of atoms that are not linked by GNM bonds. We use a semi-harmonic repulsive term, often called a soft-sphere repulsion:

$$E_{ij} = \begin{cases} k_{ij} \left(r_{ij} - r_{ijo} \right)^2 & r_{ij} \leq r_{ijo} \\ 0 & r_{ij} > r_{ijo} \end{cases}$$

where E_{ij} is the soft-sphere interaction energy between atoms i and j; k_{ij} is the soft-sphere force constant for the atom pair ij; r_{ij} is the distance between atom i and j; and r_{ijo} is the minimum distance allowed between the two atoms. k_{ij} was set to 100 kcal/(mol•Å^2), and r_{ijo} = 1.5Å, since the GNM contains all non-hydrogen atoms.

Final refinement of the model

The YUP GNM approach includes all non-hydrogen atoms, so a short cutoff is required to keep the network computationally manageable. As a consequence, the flexible optimization produces some steric clashes. These include a handful of ring penetrations, which we corrected manually. The remaining steric clashes were solved by subjecting the models to gentle energy minimization with NAMD (40), using the AMBER force field

(41). This guarantees that bond lengths and angles lie within the standard ranges, and that there are no unacceptable van der Waals contacts.

Quality of the model

The fit of the various pieces of the model into the density maps is excellent, as judged visually (Figure 2.9). We quantitatively assessed the quality of the models by calculating the cross-correlation coefficients (CCC) comparing the density maps calculated from the models with the experimental density maps. CCC values before and after YUP optimization are reported in Table A.4, which shows considerable improvement in cross-correlation coefficients during optimization. Figure 2.9 shows an example of YUP optimization of the beak region of the small subunit (helix 33).

The stereochemistry of the models was verified using the ADIT validation server of the Protein Data Bank (PDB) (http://pdb.rutgers.edu/validate/). None of the models show unacceptably close van der Waals contacts between the heavy atoms. Fewer than 1% of the bond lengths lie beyond ±0.16Å (6σ) of the lengths specified in the standard dictionary, and over 99% of the bond angles are within 2.6° (6σ) of the standard dictionary angles.

Availability of the models

PDB files are available at http://rumour.biology.gatech.edu/Publications.

References

1.   Ninio, J. 1975. Kinetic amplification of enzyme discrimination. Biochimie 57:587-595.

2.   Hopfield, J. J. 1974. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. Proc Natl Acad Sci USA 71:4135-4139.

3.   Rodnina, M. V., K. B. Gromadski, U. Kothe, and H. J. Wieden. 2005. Recognition and selection of tRNA in translation. FEBS Lett 579:938-942.

4.   Gromadski, K. B., and M. V. Rodnina. 2004. Kinetic determinants of high-fidelity tRNA discrimination on the ribosome. Mol Cell 13:191-200.

5.   Pape, T., W. Wintermeyer, and M. Rodnina. 1999. Induced fit in initial selection and proofreading of aminoacyl-tRNA on the ribosome. EMBO J 18:3800-3807.

6.   Rodnina, M. V., M. Beringer, and W. Wintermeyer. 2007. How ribosomes make peptide bonds. Trends Biochem Sci 32:20-26.

7.   Blanchard, S. C., R. L. Gonzalez, H. D. Kim, S. Chu, and J. D. Puglisi. 2004. tRNA selection and kinetic proofreading in translation. Nat Struct Mol Biol 11:1008-1014.

8.   Carter, A. P., W. M. Clemons, D. E. Brodersen, R. J. Morgan-Warren, B. T. Wimberly, and V. Ramakrishnan. 2000. Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. Nature 407:340-348.

9.   Ogle, J. M., D. E. Brodersen, W. M. Clemons, Jr., M. J. Tarry, A. P. Carter, and V. Ramakrishnan. 2001. Recognition of cognate transfer RNA by the 30S ribosomal subunit. Science 292:897-902.

10.   Valle, M., A. Zavialov, W. Li, S. M. Stagg, J. Sengupta, R. C. Nielsen, P. Nissen, S. C. Harvey, M. Ehrenberg, and J. Frank. 2003. Incorporation of aminoacyl-tRNA into the ribosome as seen by cryo-electron microscopy. Nat Struct Biol 10:899-906.

11.   Berk, V., W. Zhang, R. D. Pai, and J. H. Cate. 2006. Structural basis for mRNA and tRNA positioning on the ribosome. Proc Natl Acad Sci USA 103:15830-15834.

12.   Schuwirth, B. S., M. A. Borovinskaya, C. W. Hau, W. Zhang, A. Vila-Sanjurjo, J. M. Holton, and J. H. Cate. 2005. Structures of the bacterial ribosome at 3.5 A resolution. Science 310:827-834.

13. Valle, M., A. Zavialov, J. Sengupta, U. Rawat, M. Ehrenberg, and J. Frank. 2003. Locking and unlocking of ribosomal motions. Cell 114:123-134.

14. Selmer, M., C. M. Dunham, F. V. t. Murphy, A. Weixlbaumer, S. Petry, A. C. Kelley, J. R. Weir, and V. Ramakrishnan. 2006. Structure of the 70S ribosome complexed with mRNA and tRNA. Science 313:1935-1942.

15. Wohlgemuth, I., M. Beringer, and M. Rodnina. 2006. Rapid peptide bond formation on isolated 50S ribosomal subunits. EMBO Reps 7:699-703.

16. Sievers, A., M. Beringer, M. V. Rodnina, and R. Wolfenden. 2004. The ribosome as an entropy trap. Proc Natl Acad Sci U S A 101:7897-7901.

17. Youngman, E. M., J. L. Brunelle, A. B. Kochaniak, and R. Green. 2004. The active site of the ribosome is composed of two layers of conserved nucleotides with distinct roles in peptide bond formation and peptide release. Cell 117:589-599.

18. Koshland, D. E. 1958. Application of a theory of enzyme specificity to protein synthesis. Proc Natl Acad Sci USA 44:98-104.

19. Sanbonmatsu, K. Y., S. Joseph, and C. S. Tung. 2005. Simulating movement of tRNA into the ribosome during decoding. Proc Natl Acad Sci USA 102:15854-15859.

20. Ogle, J. M., F. V. Murphy IV, M. J. Tarry, and V. Ramakrishnan. 2002. Selection of tRNA by the ribosome requires a transition from an open to a closed form. Cell 111:721-732.

21. Réblová, K., F. Lankas, F. Rázga, M. V. Krasovska, J. Koca, and J. Sponer. 2006. Structure, dynamics and elasticity of free 16S rRNA helix 44 studied by molecular dynamics simulations. Biopolymers 5:504-520.

22. Gromadski, K. B., T. Daviter, and M. Rodnina. 2006. A Uniform Response to Mismatches in Codon-Anticodon Complexes Ensures Ribosomal Fidelity. Mol Cell 21:369-377.

23. Harvey, S. C., M. Prabhakaran, B. Mao, and J. A. McCammon. 1984. Phenylalanine transfer RNA: molecular dynamics simulation. Science 223:1189-1191.

24. Robertus, J. D., J. E. Ladner, J. T. Finch, D. Rhodes, R. S. Brown, and B. F. C. Clark. 1974. Structure of yeast phenylalanine tRNA at 3 angstrom resolution. Nature 250:546-551.

25. Woese, C. R. 2001. Translation: In Retrospect and Prospect. RNA 7:1055-1067.

26. Yarus, M. 1995. tRNA on the Ribosome: a Waggle Theory. In tRNA Structure, Biosynthesis, & Functions. D. Söll, and U. RajBhandary, editors. American Society of Microbiology Press, Washington, D.C. 443-469.

27. Hirsh, D., and L. Gold. 1971. Tryptophan transfer RNA as the UGA suppressor. J Mol Biol 58:459-468.

28. Cochella, L., and R. Green. 2005. An active role for tRNA in decoding beyond codon:anticodon pairing. Science 308:1178-1180.

29. Schlitter, J., M. Engels, and P. Kruger. 1994. Targeted molecular dynamics: a new approach for searching pathways of conformational transitions. J Mol Graph 12:84-89.

30. Isralewitz, B., M. Gao, and K. Schulten. 2001. Steered molecular dynamics and mechanical functions of proteins. Curr Opin Struct Biol 11:224-230.

31. Isralewitz, B., S. Izrailev, and K. Schulten. 1997. Binding pathway of retinal to bacterio-opsin: a prediction by molecular dynamics simulations. Biophys J 73:2972-2979.

32. Harvey, S. C., and H. A. Gabb. 1993. Conformational transitions using molecular dynamics with minimum biasing. Biopolymers 33:1167-1172.

33. Paci, E., and M. Karplus. 1999. Forced unfolding of fibronectin type 3 modules: an analysis by biased molecular dynamics simulations. J Mol Biol 288:441-459.

34. Major, F., M. Turcotte, D. Gautheret, G. Lapalme, E. Fillion, and R. Cedergren. 1991. The combination of symbolic and numerical computation for three-dimensional modeling of RNA. Science 253:1255-1260.

35. Nevskaya, N., S. Tishchenko, S. Volchkov, V. Kljashtorny, E. Nikonova, O. Nikonov, A. Nikulin, C. Kohrer, W. Piendl, R. Zimmermann, P. Stockley, M. Garber, and S. Nikonov. 2006. New insights into the interaction of ribosomal protein L1 with RNA. J Mol Biol 355:747-759.

36. Mears, J. A., M. R. Sharma, R. R. Gutell, A. S. McCook, P. E. Richardson, T. R. Caulfield, R. K. Agrawal, and S. C. Harvey. 2006. A structural model for the large subunit of the mammalian mitochondrial ribosome. J Mol Biol 358:193-212.

37. Frank, J., M. Radermacher, P. Penczek, J. Zhu, Y. Li, M. Ladjadj, and A. Leith. 1996. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. J Struct Biol 116:190-199.

38. Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. Journal of computational chemistry 25:1605-1612.

39. Tan, R. K. Z., A. S. Petrov, and S. C. Harvey. 2006. YUP: A Molecular Simulation Program for Coarse-Grained and Multiscaled Models. J. Chem. Theory Comput. 2:529-540.

40. Nelson, M., W. Humphrey, A. Gursoy, A. Dalke, L. Kalé, R. D. Skeel, and K. Schulten. 1996. NAMD - A parallel object-oriented molecular dynamics program. J. Supercomput. Appl. 10:251-268.

41. Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A second generation force field for the simulation of proteins, nucleic acids and organic molecules. J Am Chem Soc 117:5179-5197.

CHAPTER 3

ALL-ATOM MODEL OF THE EUKARYOTIC RIBOSOMAL RNA

Abstract

The ribosome is a large macromolecular complex that decodes the genetic information in the messenger RNA (mRNA) and translates it into corresponding protein with the help of transfer RNA (tRNA) and other substrates with high accuracy. The presence of ribosomal expansion segments (ES) and additional proteins in eukaryotic ribosomes as compared to its prokaryotic counterpart have been attributed to the higher fidelity of translation in eukaryotes. The roles of these ES and additional proteins have not been fully understood because of the lack of structural data for eukaryotic ribosomes. In the present collaborative work, we postulate atomic models for the eukaryotic ribosome of yeast, *Saccharomyces cerevisiae,* based on the cryo-electron microscopy (cryo-EM) density of *Thermomyces lanuginosus* at 8.9Å resolution. This work describes the modeling of the RNA molecules of the ribosome. The initial model for the RNA molecules is built using high resolution *E. coli* ribosome structures as templates for RNA homology modeling of homologous regions and additional *ab initio* modeling of ES using MCSYM and InsightII. The final model is generated by flexibly fitting and optimizing the initial model into the density using the YUP.scx module from our in-house molecular mechanics tool, YUP.

Introduction

The ribosome is a ribonucleoprotein machine responsible for the translation of genetic message contained in the mRNA to synthesize protein. In eukaryotes, it is composed of a small 40S subunit and a large 60S subunit. The small subunit (SSU) consists of 18S ribosomal RNA (rRNA) and approximately 30 proteins and is responsible

for the decoding of the mRNA. The large subunit consists of three rRNA molecules, 5S, 5.8S and 28S plus about 40 proteins. The large subunit consists of the peptidyl transferase center, where the peptidyl transfer occurs between the growing peptide and the next amino-acid (1). The SSU and LSU come together during the beginning of the translational cycle to form an 80S complex.

Compared to its prokaryotic counterpart, the eukaryotic ribosome consists of additional segments of rRNA and extra proteins mostly at the periphery (Figure 3.1). Because of these additional proteins and RNA components, the eukaryotic ribosome is larger (> 3MDa) compared to prokaryotic ribosome. However, the function of the ribosome is conserved and so is the core region of the subunits of the ribosome where the decoding of the mRNA codon and peptidyl transfer takes place (2). The 40S small subunit consists of the same morphological features, head, body, platform, beak and spur, as that of prokaryotes. The expansion segments (ES3, ES6, ES12) of the small subunit are connected to the platform. Other domains are structurally very similar. The core of the 60S large subunit is also similar to the prokaryotic LSU. The expansion segments are found in all domains, and they form bridges to the 40S subunit at two additional sites (2). Compared to the prokaryotic ribosome, the eukaryotic ribosome also has a higher translational fidelity rate. The presence of these expansion segments and proteins is thought to play some role in this increased rate.

Eukaryotic ribosome has not received the same amount of attention as the prokaryotic ribosome. The main reason is that the crystal structures of eukaryotic ribosome are not yet available. Recently, the availability of low-resolution cryo-EM density maps for eukaryotic ribosome has shed some insight into the structure of the ribosome. The ability of the cryo-EM to capture ribosomes in different functional states has increased our understanding of the functional implications of major structural features (3-5). Spahn et al. generated a model for the eukaryotic ribosome using the prokaryotic ribosome by docking the crystal structures into the cryo-EM reconstructed

47

for the yeast 80S ribosome. The model revealed the conserved and dissimilar regions as compared to the prokaryotes (6). The model generated consists of *E. coli* sequence and is missing the RNA expansion segments and the extra eukaryotic ribosomal proteins.

The current biggest challenge in the studies of the eukaryotic ribosome is the generation of a model for the ribosome in atomic-detail. In this chapter, I describe our part in the effort of proposing all-atom ribosome structure for a eukaryote, *S. cerevisiae*, using low-resolution data from cryo-EM, and secondary structure data along with its homology to the prokaryotic ribosome as structural constraints. We worked on modeling the rRNA. Andrej Sali's group in UCSF is modeling the protein components. Joachim Frank's group provided the cryo-EM maps and also segmented the large 80S complex into individual subunits and further into RNA and protein components. Our contribution is the ribosomal RNA model comprising of 99% of the eukaryotic RNA.

Method

Cryo-EM maps with isolated RNA components

The cryo-EM model was generated at 8.9Å (FSC 0.5 criterion) resolution for *T. lanuginosus* (7). Using the different diffraction intensity of RNA and protein, the RNA and protein components were separated using the method described earlier (6). The separation technique is not absolutely perfect and there may be some density that could be a part of the RNA that is in the derived protein only density or vice versa. Nevertheless, these different maps are invaluable as starting constraints for modeling the rRNA component. We were careful using the map as a basis rather than concrete data.

Alignment of sequences

Alignment of *E. coli* sequences with *S. cerevisiae* sequences from the RNA Comparative Website (8) was used rather than using typical sequence alignment programs such as clustalw that uses only sequence information. This is because the

Figure 3.1: Location of the expansion segments in secondary structure map and in the corresponding cryo-EM maps. (a) Secondary structure of the small subunit with expansion segments colored. (b) The colored segments on the cryo-EM map. (c) and (d) is similar for the large subunit. Figures courtesy of Derek Taylor.

alignment from the website is based on the evolutionary relationship between the sequences using covariance data which is more appropriate in our case.

Expansion segments

Based on the secondary structure map and the alignment, and the previously published data, expansion segments for both the large and small subunits were identified. Many of these expansion segments share the same density, especially in the large subunit (Figure 3.1d), so careful modeling of the RNA was essential so as to not overlap the expansion segments. In some cases, the position of some of the helices had to be moved after the initial fit to fit another expansion segment in the same region.

Secondary structure for the undefined regions

Secondary structures were identified based on comparative genomics. Some expansion segments were missing secondary structure information because they could not be determined using the phylogenetic studies. For these cases, MFOLD was used to predict the secondary structure (9). For cases like ES6 in small subunit, which is a large expansion segments (more than 200 nucleotides), secondary structure has been proposed by using other methods (10). In this case, the secondary structure obtained from experiments was used as a basis for getting the 3-D model. Some loops within the ES of large subunits (eg. ES28, ES39, etc) have not been predicted and MFOLD also could not predict any structures. We modeled them as loops.

Changing the *E. coli* homologous nucleotides to Yeast.

Using the alignment between *E. coli* and yeast obtained from the RNA Comparative Website, conserved RNA helices were identified (8). The identification of homologous regions allowed simple changes to be made in case of base-pair or individual nucleotide changes in loops. Deletion and insertion in the loops were also manually adjusted using the Biopolymer module of the Insight-II software package (Molecular

Simulations, Inc., San Diego, CA). Changes in loops (e.g., from tetra to penta, etc.) were adjusted by searching for a loop with the desired number of bases in crystal structures and replacing the existing loop with the new loop. In most cases some nucleotides needed to be substituted to fit the yeast sequence for the loop. Variable regions such as helix 33 (the beak region of SSU) that changed a lot needed extensive manual intervention.

Secondary to tertiary structure

MCSYM was used to generate tertiary structures for ES based on the secondary structure constraints (11). The MCSYM scripts consist of structural constraints such as pairing partners and the type of interaction between them, connection information between consecutive residues, etc. The program searches through a structural database and gathers all the pairs that satisfy the constraints. Using the residues gathered, it generates a three dimensional model. Appendix B shows an example of a simple MCSYM script.

The structures generated by MCSYM are not energetically favorable. This is because the models generated satisfy only the constraints provided. They do not use any molecular mechanics to energetically classify them. Thus the models obtained from MCSYM need to be carefully analyzed and minimized using molecular mechanics program. In our case, the models were energy minimized using Insight-II discover module. In some cases, the modeled loops were replaced by the crystallographic loops from the *E. coli* ribosome.

Fitting the expansion segments into the density map

Chimera was used to upload the density map. Rough manual fit was done to initially place the structures closer to their appropriate regions. Fit models into maps module of Chimera was used to locally fit the rigid model into the density map (12). This removes as much human biases as possible. If the expansion segments consisted of

51

multiple helical regions, each helix was independently fitted into the density. The helices were separated at the loop regions while fitting and were reconnected after the fit. We assume that helical regions have defined regions in the density maps, and thus can be fitted with more confidence.

Optimizing fit and structure refinement

The cryo-EM density of *T. lanuginosus* is incorporated as a structural restraint for refinement of our model using our YUP.scx module from our in-house molecular mechanics program YUP (13). Energy terms and optimization of YUP.scx protocol is explained in chapter 2.

<div align="center">

Results and Discussions

</div>

Using various molecular mechanics and modeling tools, we have successfully developed all-atom models for eukaryotic ribosomal RNAs. The cryo-EM density available for *T. lanuginosus* was used as constraint to model the ribosomal RNA based on the sequence for *S. cerevisiae*. *T. lanuginosus* and *S. cerevisiae* share more than 95% sequence homology, thus it is appropriate to model *S. cerevisiae* rRNA using *T. lanuginosus* density maps. The model is consistent with an earlier observation that the functional core region is conserved in the eukaryote as compared to bacteria. The overall observation of the expansion segments locations show that they are found mostly at the peripheral regions, away from the active site.

The biggest contribution of this model would be in specifying protein/RNA boundaries and the RNA/RNA boundaries between the expansion segments. Based on the model, the cryo-EM map can be separated into RNA and protein components with better confidence. The model can be further refined using the new data generated. The separated map would provide constraints for further model refinement.

Each of the expansion segment models was individually energy minimized using Insight-II Discover module. After fitting into the density and YUP.scx optimization of the complete model, stereochemical inconsistencies were corrected using conjugate gradient energy minimization with NAMD (14).
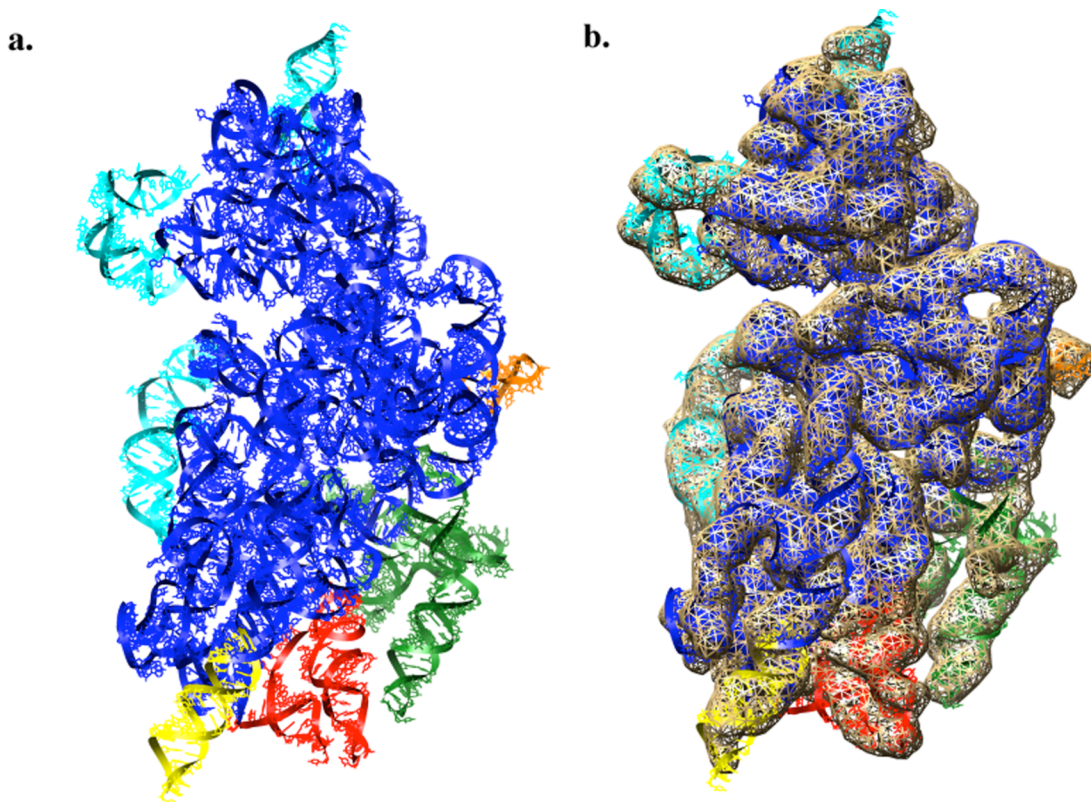
Figure 3.2: (a) Eukaryotic rRNA model for the small subunit. The red (ES3), green (ES6), orange (ES7), and yellow (ES12) represent the expansion segments. The cyan regions are the variable regions. (b) The model inside the cryo-EM density map, showing the fit.

Small subunit

ES3 is located at the base of the body of the ribosome between helix 9 and helix 10. The helix stacks on top of helix 9, which is also very variable compared to helix 9 from *E. coli*. [Helix numbering is based on prokaryotic rRNA]. This expansion segment adds another morphological feature in the small subunit, an additional spur or left foot to

the 40S ribosome. The ES3 of *T. lanuginosus* does not interact with ES6 as in other eukaryotic ribosomes (15).
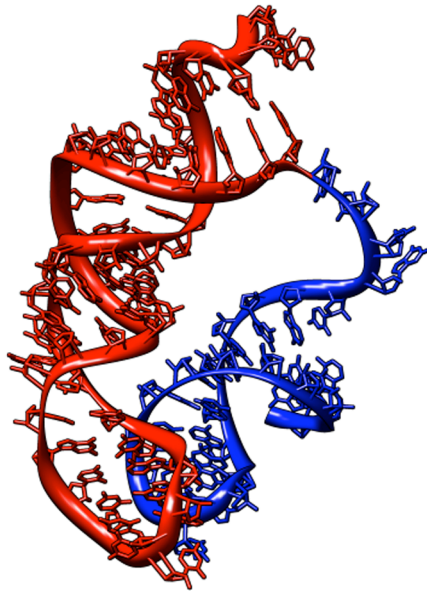
Figure 3.3: Optimized model of the small subunit ES3 (red). Stacks on top of SH9, a variable helix.

ES6 is positioned between helices 21 and 22. The secondary structure map was derived using MFOLD, with modification and cleavage data obtained as a restriction on the number of possible secondary structures derived from the software (10). This is the largest expansion segment in the small subunit and forms the major portion of the surface side of the body of subunit. A small hairpin can be seen sticking out of the RNA-only density. It could not be fitted into the density map without major stereochemical overlap. This is a case where the density previously identified as protein may actually be from the RNA.

Figure 3.4: Optimized model of the small subunit ES6 (red).

ES7 is an extension of only a few basepairs to helix 26. This lies around the mRNA channel, away from other expansion segments.

Located at the base of the penultimate stem, ES12 makes this stem longer in the eukaryotic ribosome when compared to the prokaryotic ribosome. The top of the stem is where the decoding of mRNA takes place, playing a major role in tRNA selection. There may be some relation between high translational fidelity and the longer length of this

penultimate stem. All the expansion segments, except for ES7 (which is just a few base pairs extensions) are clustered at the base of the small subunit (Figure 3.2).
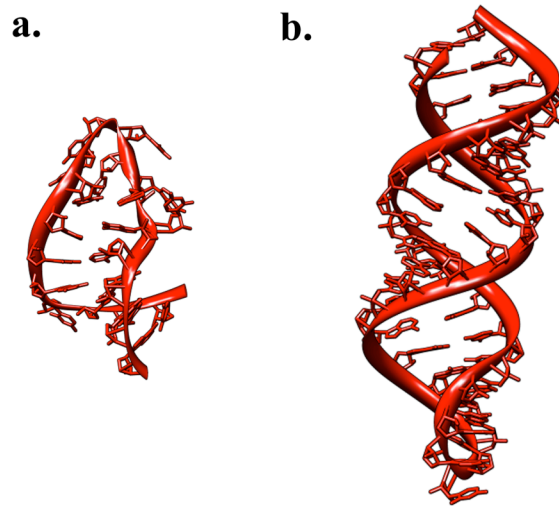
Figure 3.5: Optimized model of the small subunit (a) ES7, and (b) ES12.

Helices h6, h16, h17, h21, h33, h37, h41, h44 show major variability in nucleotide sequences as compared to its prokaryotic counterpart. A surprising observation about these variable regions is that in all the cases these helices in eukaryotic ribosomes are actually shorter than in prokaryotes.

Large subunit

ES3 is located in the 5.8S rRNA, which is homologous to the first 150 bases of the *E. coli* ribosome. It is an extension of a few bases to the already existing loop. The expansion segment lies in the lower part of the subunit in the solvent side.

ES4 is a helix with one strand from the end of the 5.8S rRNA and the other from the start of the 28S rRNA. It lies around where the 23S rRNA ends meet and form the helix in prokaryotes. ES4 replaces the first helix of 23S RNA.
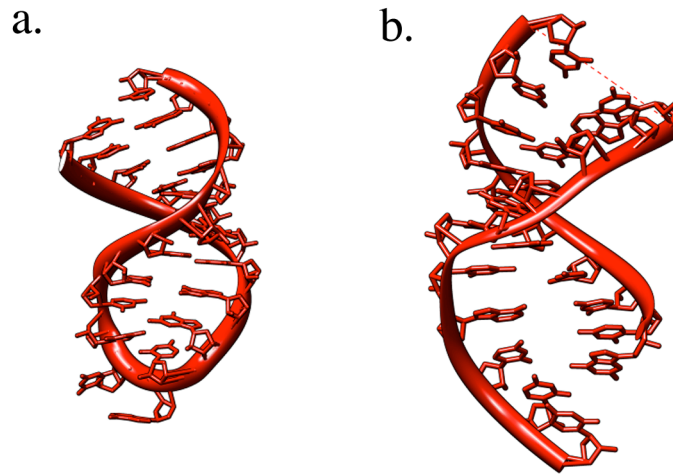
Figure 3.6: Optimized model of the large subunit (a) ES3, and (b) ES4.

Located in Domain I, ES5 is an extension of an existing short loop. It wraps around the lower part of the LSU from the solvent side. The region next to it, helix 18 is variable.
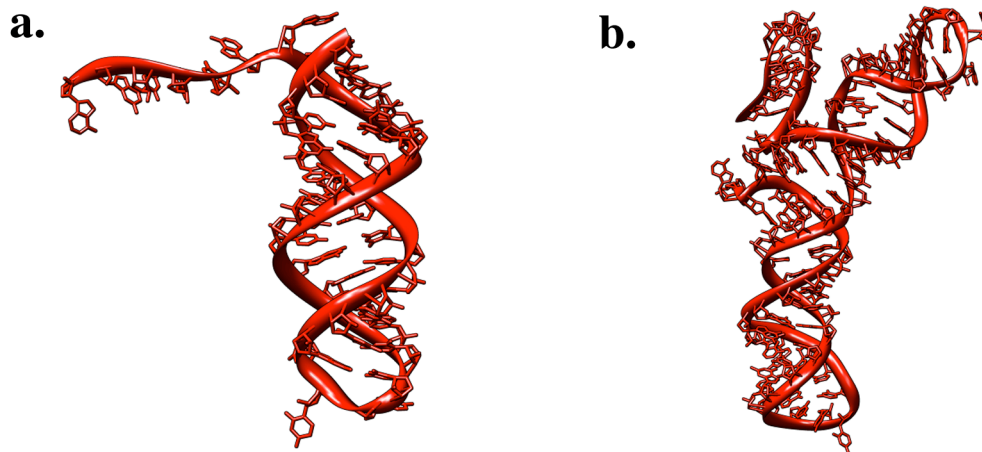
Figure 3.7: Optimized model of the large subunit (a) ES5, and (b) ES9.

ES7 is one of the largest expansion segments in the large subunit with multiple helices. Individual helices were fitted with high confidence and loop regions connected the helices together. This expansion segment interacts with another segment ES39. Both of these expansion segments lie behind the L11 binding domain. This is where the tRNA

makes initial contact with the ribosome. Thus, these expansion segments may play some role in tRNA selection.

Figure 3.8: Optimized model of the large subunit (a) ES7, one of the biggest expansion segments of the large subunit. Fitted into the density map as 5 independent helices and connected together afterwards.

ES12 is an extension of helix 38. It is located at the back of the central protuberance. It may actually play some role in tRNA selection, as it is an extension to the helix that the A-site tRNA contacts with during peptidyl transfer.

ES19 forms a small connection to the pre-existing helix and is located around the lower part of the solvent side. It interacts with ES3.

ES26 is a unique extension because it is an extension from the stem region. Secondary structure shows an extension on both strands of the stem and not an extension to a loop. However, while modeling it turned out that the region from where the

expansion segment extends is actually a sharp turn. Thus, the extension forms the base of the turn.
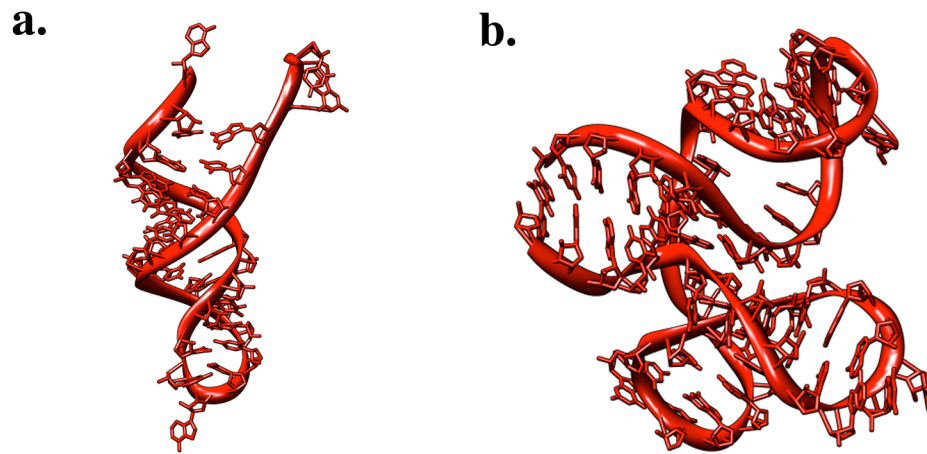
Figure 3.9: Optimized model of the large subunit (a) ES12, and (b) ES26.

ES27 region must be disordered because the density is not clearly defined in the density map. It is also a large segment with multiple long helices. Because of the lack of clear density, the model derived from MC-SYM could not be fitted into the density map.

ES31 has small length helices connected by large loops. The segment is visible from intersubunit side as well and thus could form an intersubunit bridge with the small subunit.

ES41 extends of an already existing helix. It lies on the other side of the large subunit away from ES3, ES4 and ES5.
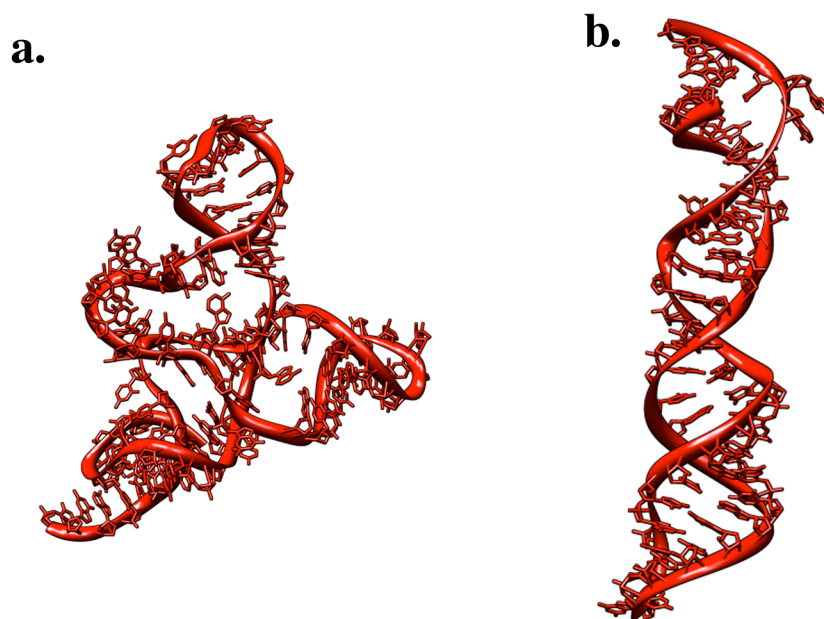
Figure 3.10: Optimized model of the large subunit (a) ES31, and (b) ES41.

Conclusions

This collaborative project is a first attempt of trying to create an all-atom model for a eukaryotic ribosome. As with any model, this model can be further refined with the availability of other structural data. We are in the initial stage of proposing an all-atom model for the eukaryotic ribosome. rRNA models generated would contribute substantially in understanding the eukaryotic ribosome and the role of the expansion segments in protein synthesis.

The major contribution of this model would be in understanding additional protein-RNA interactions in the eukaryotic ribosome. In addition, the model will help us identify the boundaries between expansion segments. The model could be further refined with the additional information about the boundaries. We currently have the model for the ribosomal RNA only. When the coordinates for the proteins become available, detailed analysis of the expansion segments in relation to the additional proteins can be

done. Another contribution of this model would be for solving eukaryotic crystal structure. The model we are proposing could be used as a starting model for refinement when generating an all-atom model from the crystallography data.

The question that needs to be addressed is why the ribosome needs to be bigger in eukaryotes as compared to prokaryotes? What are the roles of these ES is in translational fidelity? Do the expansion segments interact with homologous proteins only? Or do they interact with additional proteins only? The integration of the RNA and protein model, which will generate a complete model for the eukaryotic ribosome will suggest new experiments that will bring us closer to the answers.

References

1. Verschoor, A., S. Srivastava, R. Grassucci, and J. Frank. 1996. Native 3D structure of eukaryotic 80s ribosome: morphological homology with E. coli 70S ribosome. The Journal of cell biology 133:495-505.

2. Doudna, J. A., and V. L. Rath. 2002. Structure and function of the eukaryotic ribosome: the next frontier. Cell 109:153-156.

3. Beckmann, R., C. M. Spahn, J. Frank, and G. Blobel. 2001. The active 80S ribosome-Sec61 complex. Cold Spring Harbor symposia on quantitative biology 66:543-554.

4. Gomez-Lorenzo, M. G., C. M. Spahn, R. K. Agrawal, R. A. Grassucci, P. Penczek, K. Chakraburtty, J. P. Ballesta, J. L. Lavandera, J. F. Garcia-Bustos, and J. Frank. 2000. Three-dimensional cryo-electron microscopy localization of EF2 in the Saccharomyces cerevisiae 80S ribosome at 17.5 A resolution. The EMBO journal 19:2710-2718.

5. Spahn, C. M., M. G. Gomez-Lorenzo, R. A. Grassucci, R. Jorgensen, G. R. Andersen, R. Beckmann, P. A. Penczek, J. P. Ballesta, and J. Frank. 2004. Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation. The EMBO journal 23:1008-1019.

6. Spahn, C. M., R. Beckmann, N. Eswar, P. A. Penczek, A. Sali, G. Blobel, and J. Frank. 2001. Structure of the 80S ribosome from Saccharomyces cerevisiae-- tRNA-ribosome and subunit-subunit interactions. Cell 107:373-386.

7. Sengupta, J., J. Nilsson, R. Gursky, C. M. Spahn, P. Nissen, and J. Frank. 2004. Identification of the versatile scaffold protein RACK1 on the eukaryotic ribosome by cryo-EM. Nature structural & molecular biology 11:957-962.

8. Cannone, J. J., S. Subramanian, M. N. Schnare, J. R. Collett, L. M. D'Souza, Y. Du, B. Feng, N. Lin, L. V. Madabusi, K. M. Muller, N. Pande, Z. Shang, N. Yu,

and R. R. Gutell. 2002. The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. BMC Bioinformatics 3:2.

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

10. Alkemar, G., and O. Nygard. 2006. Probing the secondary structure of expansion segment ES6 in 18S ribosomal RNA. Biochemistry 45:8067-8078.

11. Major, F., M. Turcotte, D. Gautheret, G. Lapalme, E. Fillion, and R. Cedergren. 1991. The combination of symbolic and numerical computation for three-dimensional modeling of RNA. Science 253:1255-1260.

12. Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. Journal of computational chemistry 25:1605-1612.

13. Tan, R. K. Z., A. S. Petrov, and S. C. Harvey. 2006. YUP: A Molecular Simulation Program for Coarse-Grained and Multiscaled Models. J. Chem. Theory Comput. 2:529-540.

14. Nelson, M., W. Humphrey, A. Gursoy, A. Dalke, L. Kalé, R. D. Skeel, and K. Schulten. 1996. NAMD - A parallel object-oriented molecular dynamics program. J. Supercomput. Appl. 10:251-268.

15. Alkemar, G., and O. Nygard. 2003. A possible tertiary rRNA interaction between expansion segments ES3 and ES6 in eukaryotic 40S ribosomal subunits. RNA (New York, N.Y 9:20-24.

CHAPTER 4

STRUCTURAL AND ELECTROSTATIC CHARACTERIZATION OF PARIACOTO VIRUS: IMPLICATIONS FOR VIRAL ASSEMBLY

Abstract

We present the first all-atom model for the structure of a T=3 virus, Pariacoto virus (PaV), which is a non-enveloped, icosahedral RNA virus and a member of the *Nodaviridae* family. The model is an extension of the crystal structure, which reveals about 88% of the protein structure but only about 35% of the RNA structure. Evaluation of alternative models confirms our earlier observation that the polycationic protein tails must penetrate deeply into the core of the virus, where they stabilize the structure by neutralizing a substantial fraction of the RNA charge. This leads us to propose a model for the assembly of small icosahedral RNA viruses: the nonspecific binding of the protein tails to the RNA leads to a collapse of the complex, in a fashion reminiscent of DNA condensation. The globular protein domains are excluded from the condensed phase but are tethered to it, so they accumulate in a shell around the condensed phase, where their concentration is high enough to trigger oligomerization and formation of the mature virus.

Introduction

Pariacoto virus (PaV), a T=3, non-enveloped, icosahedral virus is a member of the Nodaviridae family. It was originally isolated in Peru from the Southern armyworm, *Spodoptera eridania* (1). Its genome consists of two positive-sense ssRNAs (2). RNA1 (3011 nucleotides) codes for protein A, the catalytic subunit for the host RNA replicase, which enables the RNA-dependent RNA replicase to start replicating the viral RNA. RNA2 (1311 nucleotides) codes for capsid precursor protein α. 180 of these α proteins

and the genome assemble together to make up the virus. Ever since it was isolated, PaV has been extensively studied using various techniques (3-6). The relatively small size (20nm in diameter) compared to other RNA viruses, and the ease by which it can be produced in various cell lines (7), make PaV and other members of the Nodaviridae family easy to characterize at the molecular level (8-10).

Structural studies of viruses are very important to understand the protein-protein and protein-RNA interactions as well as to understand assembly pathways in RNA viruses (11-14). In the last few years, many studies have been done on RNA viruses using molecular modeling as a supplementary method when other methods such as x-ray crystallography and cryo-electron microscopy (cryo-EM) do not give sufficient structural information. An all-atom model was derived for a Satellite Tobacco Mosaic Virus (STMV), a T=1 virus, using molecular modeling (15). Subsequently, molecular dynamics was done on the model to study the stability of the protein capsid and the RNA genome (15). Electrostatic interactions between RNA and the protein capsid were studied in Cowpea Chlorotic Mottle Virus (CCMV) by modeling the virus using coarse-grained modeling and representing RNA nucleotides by spheres that were distributed using the Monte Carlo method (16). In addition, electrostatic properties of virus capsids and RNA have also been studied to understand the structural properties and the molecular interactions within the virus (17, 18).

The 3.0Å x-ray crystal structure of PaV reveals an asymmetric unit with three quasi-equivalent protein subunits (A, B and C) and one strand of a 25 base pair RNA duplex (6). Sixty of these units combine to form the icosahedral capsid, with 30 RNA duplexes lying along subunit contacts across the icosahedral 2-fold axes, forming a dodecahedral cage inside the capsid. The A, B, and C subunits (residues 83-321) are folded into an eight-stranded antiparallel β-sandwich, similar to proteins in other nodaviruses. Complementing the x-ray studies, cryo-electron microscopy showed the general overall structure of PaV at 23Å resolution, which matched well with the low

resolution model calculated from the atomic coordinates (6). Cryo-EM also confirmed that the part of the RNA genome that was resolved in the x-ray structure forms the edges of the dodecahedral cage inside the protein capsid.

Although x-ray crystallography and cryo-EM provided a lot of information regarding the PaV structure, they were not able to determine the atomic structure of the complete virus. RNA at the dodecahedral edges accounts for only 35% of the total genome. The remaining 65% of the RNA lies inside the dodecahedral cage and is not resolved in the crystal structure because it lacks icosahedral symmetry. In addition, the 20 vertices at which the RNA duplexes are connected could not be resolved, presumably because different vertices have different structures. Similarly, protein subunit A is missing 6 residues at the N terminal end and 15 at the C-terminal in the crystal structure, while the B and C subunits are missing about 50 residues at the N-terminus and 19 residues at the C-terminus in the crystal structure (6).

In this paper, we report a model for the complete virus and examine the interactions of the basic N-terminus tails with the RNA genome, and their role in the stability of PaV. We used molecular modeling to model the missing 65% of the genome and the unresolved protein residues. We built our models using coarse-grained modeling, representing unresolved nucleotides and amino acids by pseudoatoms and interpolating the pseudoatomic models to all-atom using special algorithms. We generated two all-atom models for the virus that differed in the conformations of the N-terminus protein tails and the extent to which they penetrate into the RNA genome. We tested these against the experimental radial density distributions from cryo-EM, and we evaluated the relative stabilities of the two models by comparing their energies. The result is the first all-atom model for a complete T=3 virus. Further, this effort has led to a new model for the assembly of small, non-enveloped icosahedral RNA viruses.

Methods

RNA modeling

The modeling of the Pariacoto virus genome posed several challenges because of the limited amount of available structural data. To begin with, the secondary structure for the PaV genome is not known. We used a hypothetical secondary structure mapped onto the dodecahedral cage (Figure 4.1). This is the same secondary structure that we proposed earlier (19). Those parts of the RNA genome that do not form the edges of the dodecahedral cage drop inwards towards the center of the capsid as "stalactites". The exact number of these connections is not known, but we used a combination of 3-way junctions and 4-way junctions as structural motifs connecting the RNA on the dodecahedral cage with the RNA in the interior (Figure 4.2). Nothing at all is known about the RNA structure in the interior, so we have to postulate a collection of plausible structures for the stalactites. We used twelve copies of a structure derived from the *E. coli* ribosome domain IV (residues 1764-1988) to represent these. Although the twelve stalactites all have the same initial conformation, these become quite varied during the refinement of the model.

General approach

The volume inside the dodecahedral cage is too small to attach twelve of these stalactites in their initial conformation coming inwards from different vertices without significant interpenetration between them. To solve this problem, we expanded the diameter of the dodecahedral cage by a factor of two, which increased the overall volume of the dodecahedral cage 8-fold. This allowed us to add the stalactites from the vertices. This expanded model was contracted to the actual size in twelve steps, with extensive energy minimization at each step.
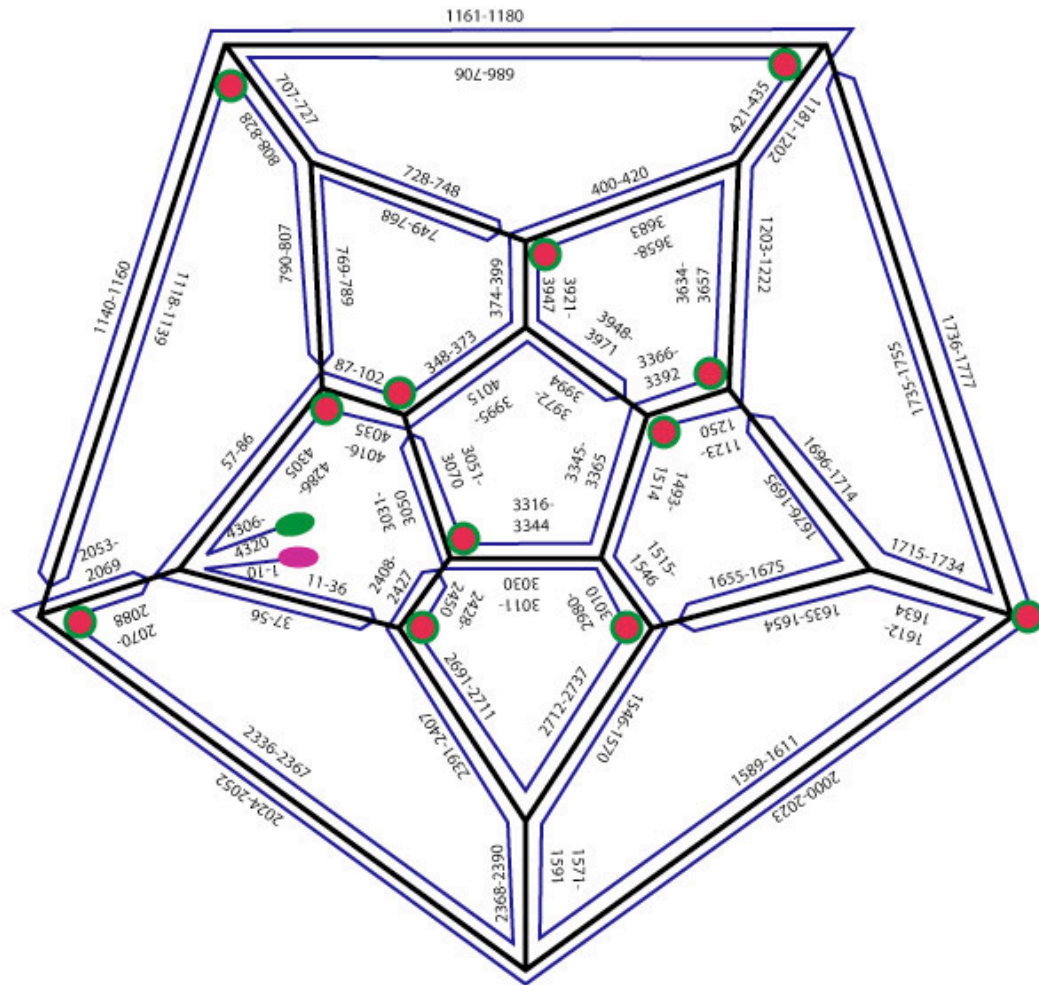
Figure 4.1: Secondary structure map for the Pariacoto virus (PaV) genome, adapted from our earlier model (19). The bipartite genome has been represented as one single strand, since know neither the secondary structure of RNA1 and RNA2, nor the structure of the interactions between them, if any. Pink and green dots represent the 5' and 3' ends, respectively. Red circles with blue borders are the junctions where the stalactites were added to connect with RNA deeper in the interior of the capsid (see text).

Initial RNA genome model

The crystal structure of PaV (1F8V.pdb) is available from the RCSB Protein Data Bank (20). The dodecahedral RNA cage was generated by applying the BIOMT TRANSFORMATION matrix given in the file, using the oligomer generator tool in the Viper database (21). The vertex structures were defined by the secondary structure (Figure 4.1). Each vertex had either three or four extensions of RNA coming out of it

(Figure 4.2). Small hairpin loops were added at twelve vertices, as stubs to which the
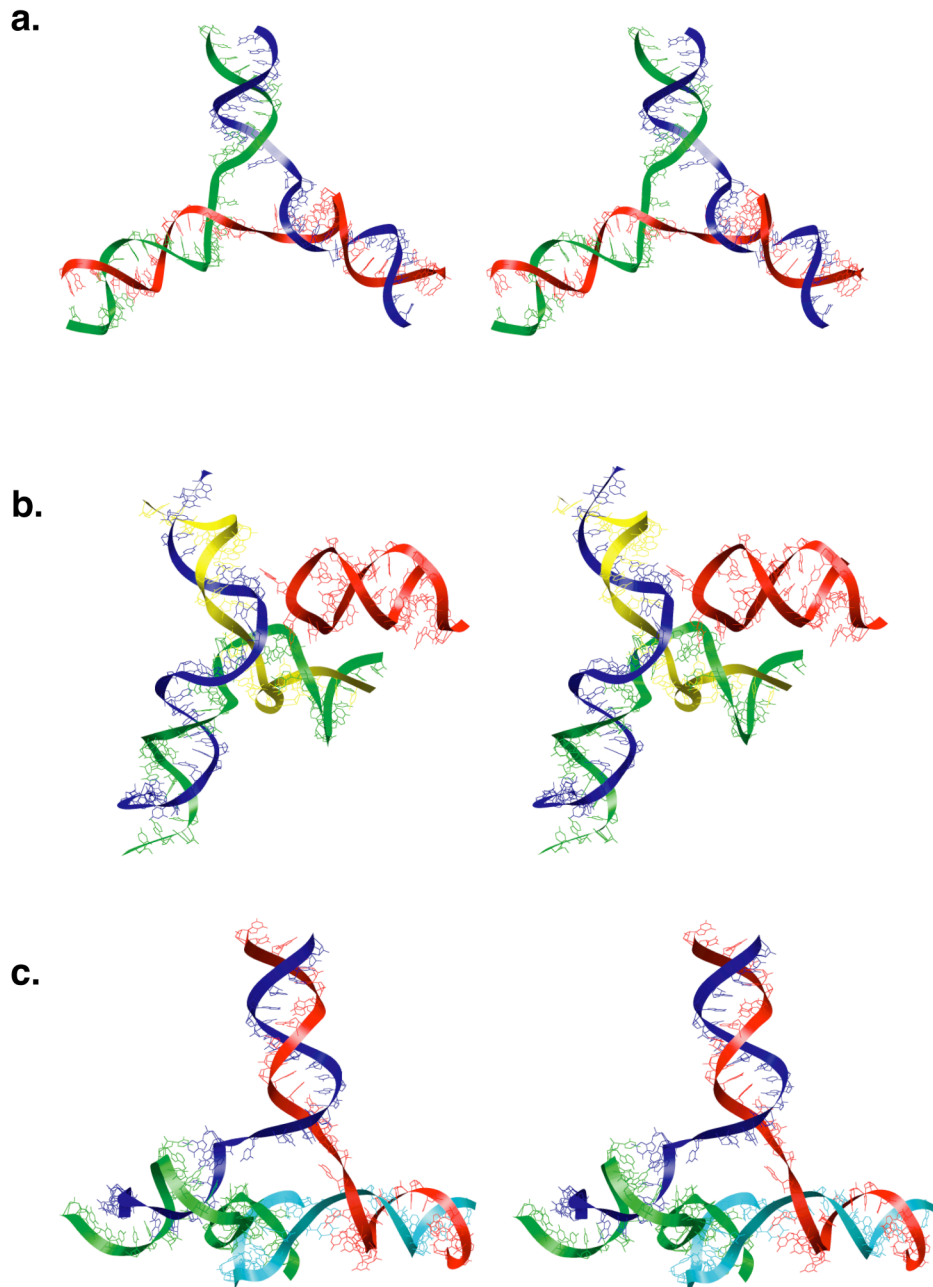
Figure 4.2: Stereo images of model junctions. **a.** A typical three-way junction. RNA duplexes line on three adjacent edges of dodecahedral cage, and there is no stalactite at the vertex. **b.** Another type of three-way junction, connecting duplexes on two edges with a stalactite. The stalactite is attached to the green and yellow helix. There is a stem-loop on the third edge, coming from a neighboring vertex (red). **c.** A four-way junction, connecting duplexes on three edges with a helix (blue and green) that is the attachment point for a stalactite.

stalactites were subsequently added. We cut the RNA duplex on each edge in half, fixing each half to the appropriate vertex. This initial model was generated on a Silicon Graphics workstation using the Builder module of INSIGHT II graphics software. This initial model (Figure 4.3a) contained all 4322 RNA residues.
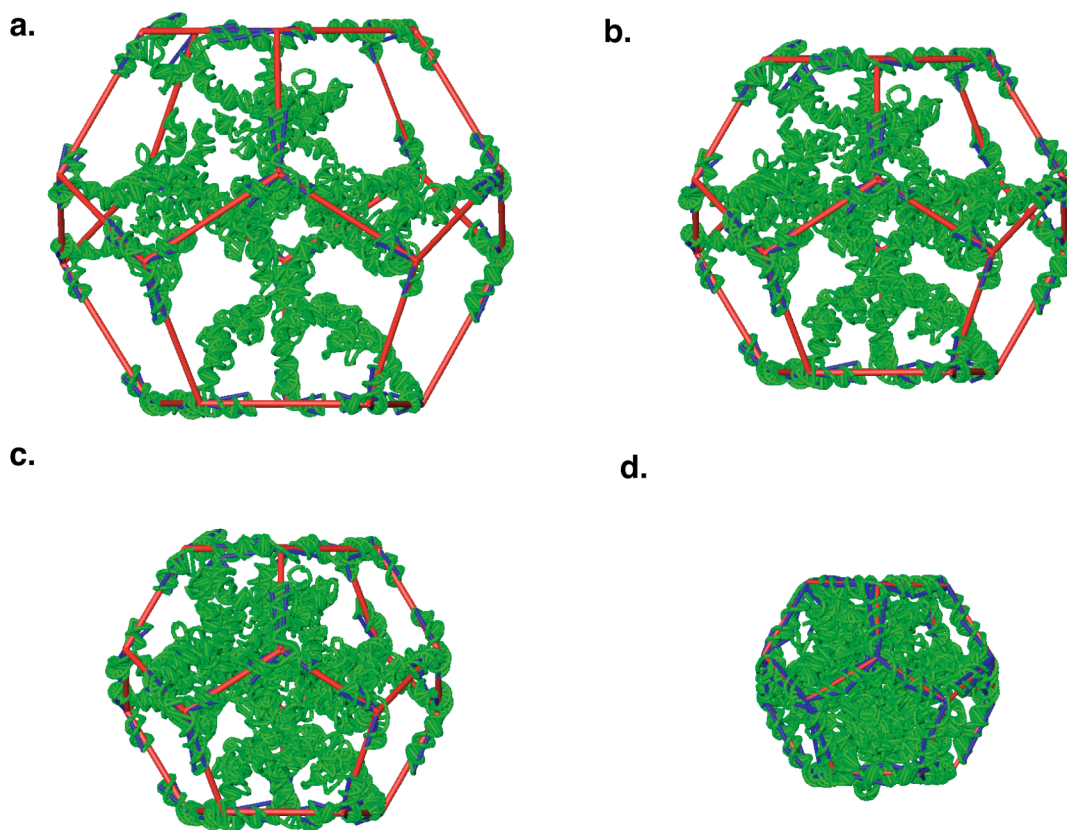
Figure 4.3: Minimization protocol for the viral RNA. **a.** The initial model with the diameter of the dodecahedral cage doubled (red lines). RNA duplexes are cut at the middle and rigidly attached to their corresponding vertex atoms. Pseudobonds from each vertex atom to the edge of the RNA duplex are represented as blue lines. These bonds restrain the crystallographic regions during minimization. The stalactites can be seen inside the dodecahedral cage. The volume of the cage is eight times the volume of the actual cage. **b.** The model after four rounds of minimization, at about six times the actual volume. **c.** The model after eight rounds of minimization, at about three times the actual volume. **d.** The final model, after twelve rounds of minimization.

Coarse-grained modeling and minimization

 The initial model is quite large and the experimental data available for modeling are quite limited, so coarse-grained modeling is appropriate for refining the model. We converted the all-atom initial model to coarse-grain representation, with each nucleotide represented by a pseudoatom at the phosphate position. A more complete description of this "all-P" model is available elsewhere (22), along with a full description of the corresponding force field. Twenty pseudoatoms were also added at the vertices of the dodecahedral cage, to form a framework that could be easily expanded and contracted; we call these "vertex pseudoatoms".

 The edges of the dodecahedral cage were decreased to the original length in multiple steps, decreasing the ideal bond length (b_0) of the expanded framework in 5Å steps and minimizing until convergence after each step (Figure 4.3). The minimization was done using our in-house molecular mechanics package, YAMMP (22). The harmonic energy terms used in the minimization are tabulated in Table 4.1. Since all the terms used in the potential energy function of all-P models are harmonic, full minimization of the model should lead to zero energy, if all restraints can be satisfied.

 During minimization, the stalactite RNAs were free to move and adjust their conformations, to avoid steric overlap. They had softer force constants in the energy terms than did the RNA domains on the dodecahedral cage (Table 4.1). The crystallographic regions were restrained by using strong force constants in the energy terms, and by the addition of pseudobonds connecting each vertex pseudoatom to the ends of the RNA duplexes on each edge (Figure 4.2). These regions did not deviate significantly from the crystal structure during the contraction/minimization cycles.

Table 4.1: Energy terms used in RNA genome modeling

Energy terms	Types	Equation	Force constant
Bond	Crystallographic	$E_b = k_b(b - b_0)^2$	k_b=20kcal/mol
	Stalactites	where b_0 is the distance in the initial model derived from crystal structure.	k_b = 2 kcal/mol
Angle	Crystallographic	$E_\theta = k_\theta(\theta - \theta_0)^2$	k_θ=20kcal/mol
	Stalactites	where θ_0 is the angle in the initial model	k_θ = 2kcal/mol
Improper torsion	Crystallographic	$E_i = k_i(i - i_0)^2$	k_i =20 kcal/mol
	Stalactites	where i_0 is the improper torsion between four atoms in the initial model	k_i = 2kcal/mol
Non-bond exclusion		$E_{nbn} = k_{ij}(d - d_{ij})^2$, if $d < d_0$, $d_0 = 10\text{Å}$	k_{ij} = 2kcal/mol
NOE term		$E_{noe} = \begin{cases} k_{hi}(r - r_{hi})^2, & \text{if } r > r_{hi} \\ 0, & \text{if } r \leq r_{hi} \end{cases}$ where r is the distance from the center to atom i. r_{hi} was changed during each step of minimization.	k_{hi} = 2kcal/mol
Stud	Stud atom was kept at the center to keep the RNA within a certain radius.	$E_{st} = k_{st}[(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2]$ where (x,y,z) is the current position of atom i and (x_0,y_0,z_0) is the desired position.	k_{st} = 40kcal/mol

Conversion of the refined coarse-grained RNA model to an all-atom model

Generating an all-atom model from phosphate positions is a challenging problem. The bond and angle restraints in the all-P models are based on observed distributions of P-P distances and P-P-P angles in the Nucleic Acid Database (22). With only these restraints, there is no way to guarantee that groups of four or more successive P atoms in any all-P model will have a conformation that corresponds to any real RNA structure. As a consequence, all-atom models can be generated fairly easily in double-helical regions, but all-atom models for other regions (loops, bulges, single-strands) are necessarily more speculative. This is not inappropriate, considering the modesty of our overall goal: generate a plausible RNA model, in terms of connectability along the backbone and the

absence of serious steric problems. A more rigorous structural effort would not be justified, because we don't know the actual secondary structure of the PaV RNAs, and there are no high- or intermediate-resolution data on the RNA structure, except within the dodecahedral cage.
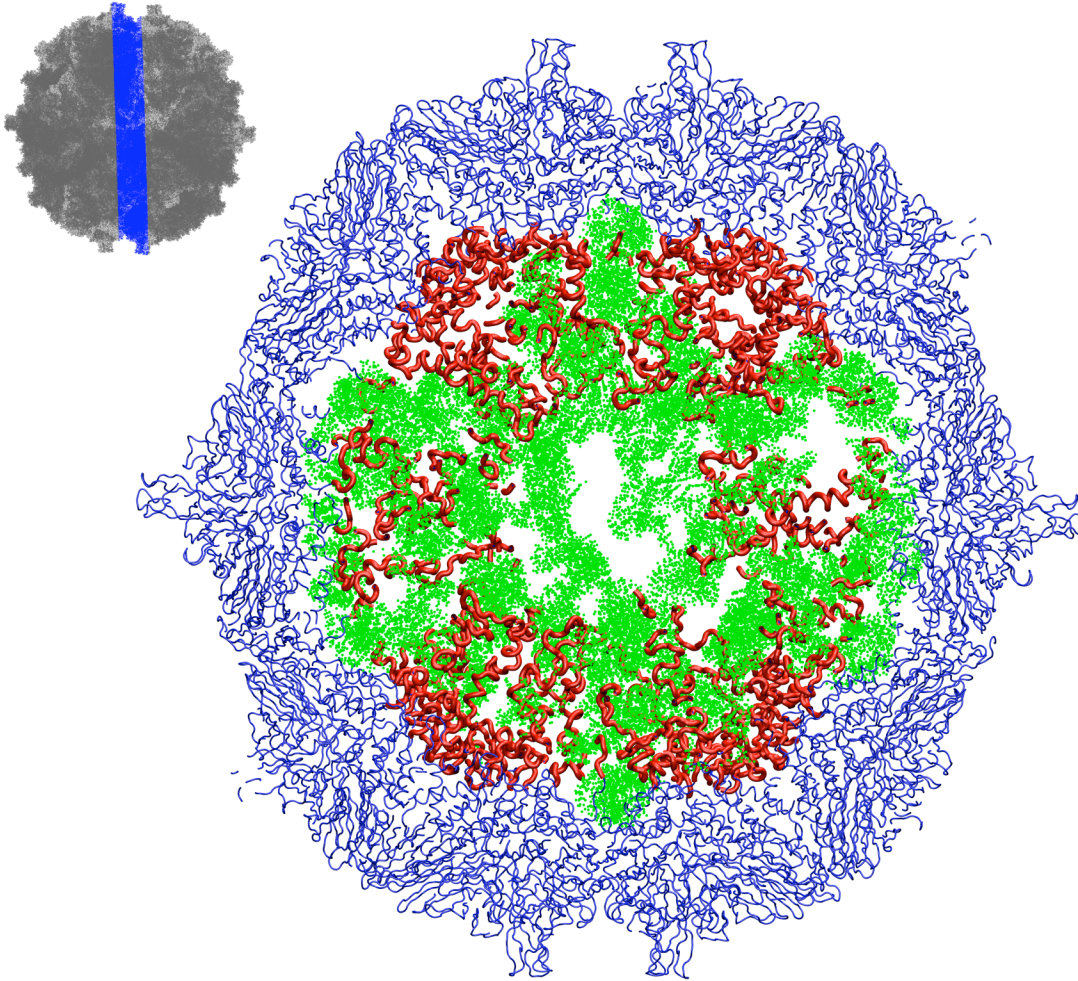
Figure 4.4: A 20Å slice through the center of Model_8. Protein residues seen in the crystal structure are colored blue, while noncrystallographic residues are red. The RNA is green. The protein tails reach very close to the center of the structure.

Briefly, the procedure used here builds all-atom models using a database of nucleotide conformations derived from all RNA-containing structures in the PDB as of April, 2006. In base-paired regions, four phosphate positions (0 and +1 on each strand) serve as anchor points, and a pair of nucleotides from the database must be fit to the

structure, one on each strand. In non-base-paired regions, the four anchor phosphates are those -1, 0, +1 and +2 relative to the nucleotide being placed. The compatibility of all examples in the database with a particular position is assessed by requiring that the base be identical to the one being modeled, and that the root mean square deviations of the four phosphate positions in the example be within 1.5Å of the anchor phosphates in the all-P model. Only examples that pass this compatibility test are kept within the search space of each nucleotide.

The modeling problem then becomes one of exploring the search space of the whole molecule to determine which combination of examples gives the most plausible structure, where plausibility is defined as the lowest energy (van der Waals plus electrostatics, using the AMBER 8 force field). This optimizes base pairing and stacking, while minimizing steric clashes. Searching is done in a piecewise fashion, focusing on individual regions, to optimize performance. The most plausible structure is then refined by optimization of the ribose conformations, followed by energy minimization and a short annealing of the entire model, using molecular dynamics.

Protein modeling

For modeling the missing protein residues, we followed a similar methodology as in the case of RNA modeling, expanding the capsid, adding missing amino acids, and then shrinking the capsid back to its original size in multiple steps, with minimization at each step. Coarse-grain modeling was the initial step in modeling the missing residues of the capsid proteins. After refinement of the coarse-grain model was complete, it was converted to an all-atom model, followed by final refinement.

Coarse-grained protein modeling

First, the capsid was expanded three times in length by simply multiplying the coordinates of the capsid atoms by 3. The crystallographic residues facing towards the RNA were converted into a model where two consecutive residues are represented by a pseudoatom (2C-model). The rest of the crystallographic residues were represented by twelve pseudoatoms each, defining the face, edge and the vertices of the equilateral triangle of each asymmetric unit. The missing N-terminal residues were generated in extended linear form pointing towards the RNA genome at the center. C-terminal residues were generated as a random coil. Residues for both the N- and C-terminal tails were represented by one pseudoatom per residue (Figure C.2).

The starting capsid model was scaled back down to the original size in a series of steps, testing different scaling factors and Van der Waals (vdw) diameters for the pseudoatoms of the protein tails. We examined scaling ratios between 0.95 to 0.99, finding that different scaling ratios did not significantly affect the configurations of the protein tails (data not shown). However, changing the vdw diameters from 8 to 12 Å significantly affected the penetration of the protein tails into the RNA genome (Figure 4.5b). The resulting structures, designated model_8 and model_12, have dramatically different conformations for the protein tails. In model_8, the tails penetrate deeply into the RNA core, while they lie on the outside of the RNA core in model_12.

All-atom protein modeling

Model_8 and model_12 were converted into all-atom representation using PULCHRA (22). This program converts Cα models to all-atom models using a rotamer library prepared from the statistics of Cα distances in the PDB. The complete all-atom models, including all residues of the RNA genome and the capsid proteins, were energy minimized with NAMD, using the CHARMM forcefield.

74

Electrostatic calculations:

Calculations of the electrostatic potential were performed using the Adaptive Poisson-Boltzmann Solver (APBS) (23). CHARMM27 forcefield radii and charges were assigned to the minimized all-atom structures of Model_8 and Model_12 using the PDB2PQR (24) routine, yielding a charge of $+46e$ for each of the 60 capsomers and $-4320e$ for the RNA genome, where e is the charge on the proton. This resulted in the a net charge of $-1560e$ for the complete virus. The nonlinear version of the Poisson-Boltzmann equation was solved numerically on a 225×225×225 grid with an initial grid spacing of 2.0 Å, followed by focusing with the grid spacing reduced to 1.5 Å. The dielectric constants of the interior and exterior of the macromolecules were set to 10 and 78.5, respectively. The ionic strength was set to 100mM, using only monovalent ions. The resulting potentials were mapped onto the solvent accessible surface area of the models generated at the coarse-grained level and visualized using Chimera (25).

Model Evaluation

The coarse-grained pseudoatomic model of the genome was checked for the presence of possible knots using the "knot" program (26). Our RNA model does not contain any knots. The all-atom genome model reconstructed from the pseudoatomic model was also checked for interpenetration of rings and correct stereochemistry using PROCHECK, provided in the RCSB PDB website (http://www.pdb.org). There are no ring penetrations or other stereochemical problems. The RNA and protein distributions inside the complete all-atom models of the virus were compared with the native virus by generating density maps and corresponding radial density distribution functions (Figure 4.5) from the final all-atom models, using SPIDER (27).

Results and Discussions

The 65% of the genome that was not resolved in the crystal structure was generated and packaged within the dodecahedral cage. Even though all twelve stalactites had the same starting structures, they have significantly different conformations in the final model (Figure C.1). The protein tails missing in the crystal structure were also generated, and their final conformations also vary significantly from one another in the final model.

The generation of two models for PaV that differ in the distribution of the N-terminus protein tails offers an opportunity to study their role in stabilizing the virus. The different positions of the tails in the two models are reflected in different density distributions (Figure 4.5). In model_12 most of the tails are packed in a shell around 100Å from the center, which is between the genome and capsid. For model_8, many protein tails were able to penetrate deep inside the genome, and they contribute significantly to the density peak at a radius of about 50Å (Figure 4.5a). Peaks around this radius have been found in PaV (Figure 4.5b) and in other nodaviruses (19). Thus, structurally model_8 is structurally more consistent with native viruses than model_12. This is also consistent with density maps in Flock House virus (FHV), which is closely related to PaV. The radial density distribution for wild type FHV has a peak at $R\sim32\text{Å}$, but that peak is missing in mutant FHV in which 30 amino acids have been deleted from the amino terminus (19).

Single point energy calculation of the two models showed that model_8 is also energetically more favorable than model_12. The electrostatic interaction energy between the RNA and the capsid of PaV is much lower for model_8 (-3910 kcal/mol) than for model_12 (-523 kcal/mol). This agrees with the observations drawn from the structural data (Figure 4.5): the protein tails that penetrate deep into the core of the virus stabilize PaV by neutralizing a large fraction of the charge of the RNA genome.
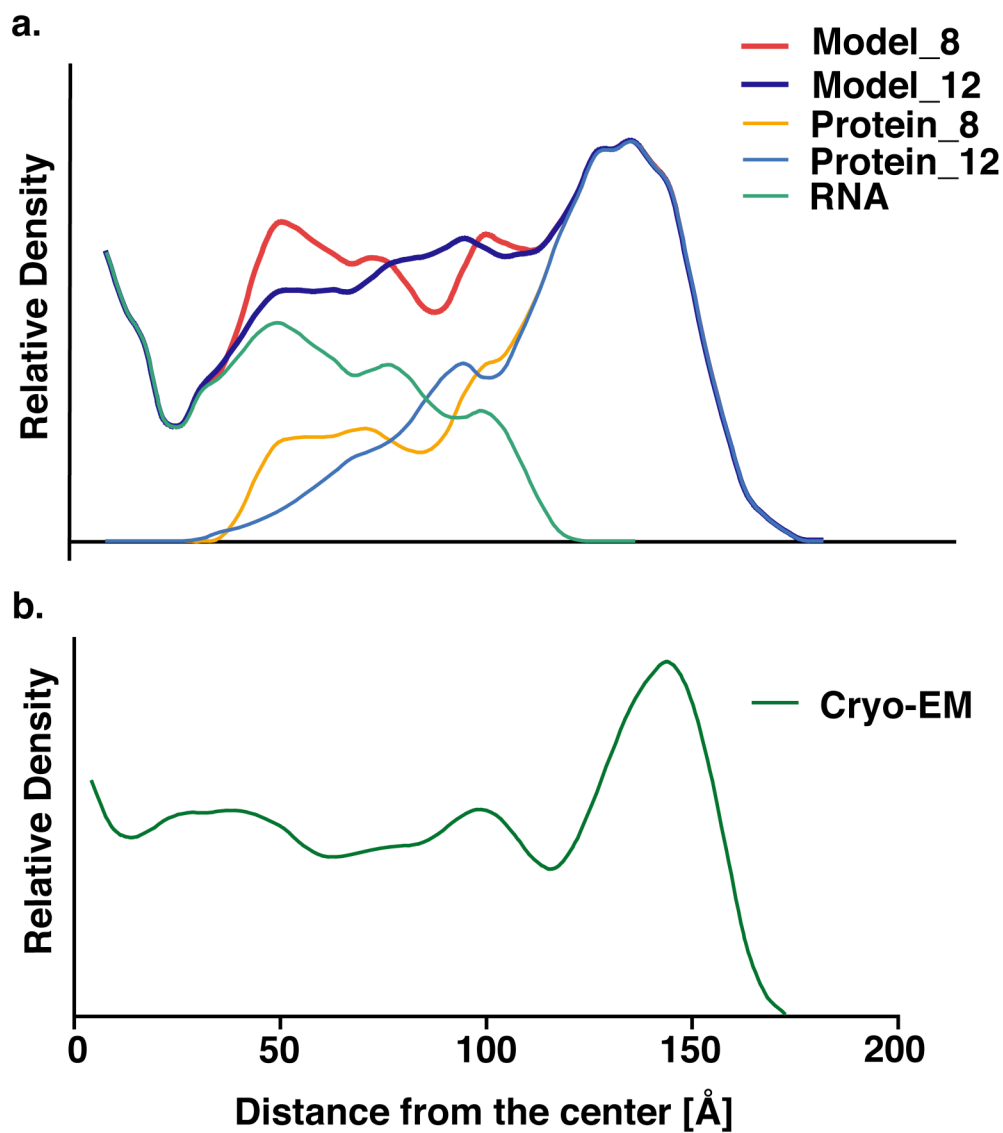
Figure 4.5: Comparison of model radial density distributions with the experimental distribution. **a.** Density distributions have been separated into RNA and protein components for model_8 and model_12. The peak at around 50Å for model_8 is due to the major contribution of the protein tails that penetrate deeply into the RNA core. For model_12, most of the protein tails are packed in a shell at a radius of ~100Å. **b.** Experimental cryo-EM density distribution.
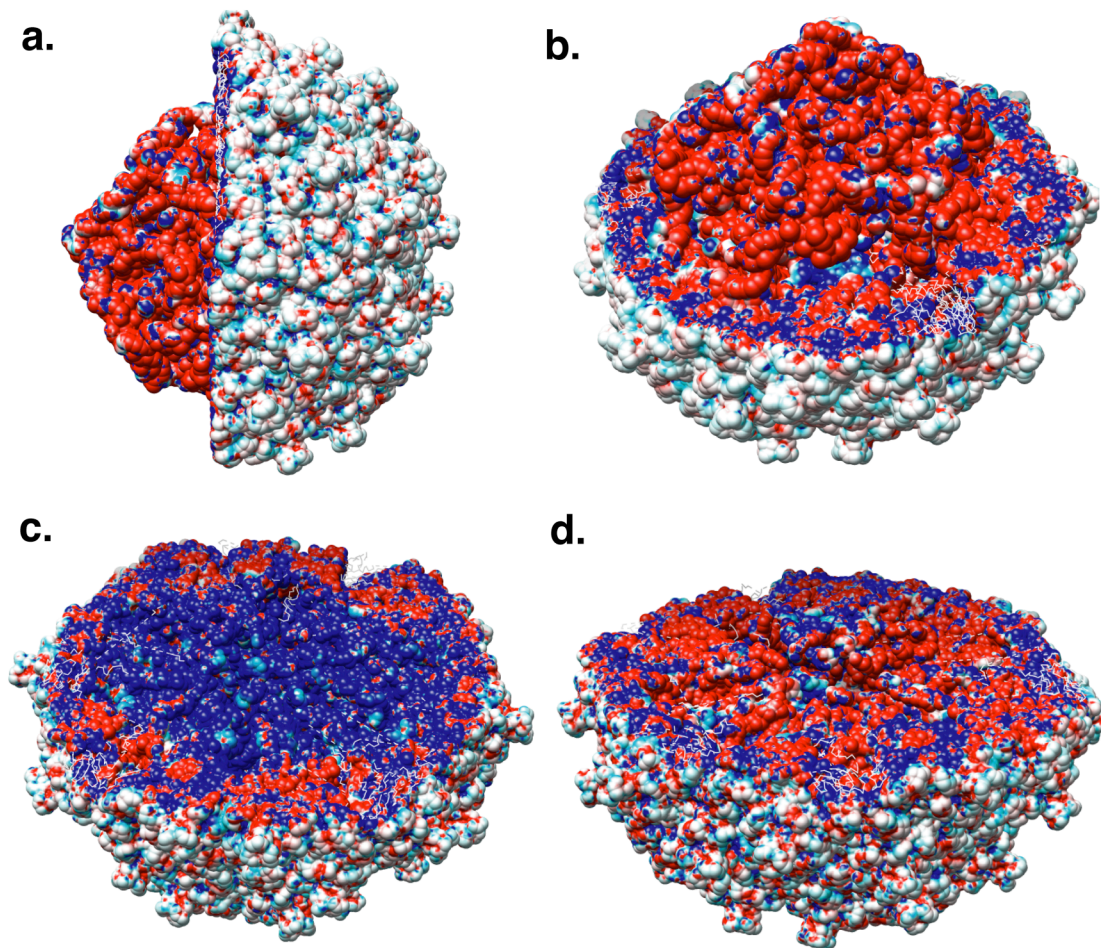
Figure 4.6: Electrostatic potential mapped onto the solvent-accessible surface area of PaV. The potential of the entire virus is mapped onto the surface of the RNA and one hemisphere of the capsid shell: **a**. side view; **b.** top view. **c.** Potential of the empty capsid mapped onto the surface of one hemisphere of the capsid. **d.** Potential of the entire virus mapped onto the surface of an empty hemisphere of capsid proteins. The color code of the electrostatic potential ranges from -5 kT/e (red) to 5 kT/e (blue).

Figure 4.6 depicts the electrostatic potential mapped onto the solvent accessible surface area of PaV. The external surface of PaV is almost neutral (Figure 4.6a), whereas the interior of the virus bears both positive charges (the protein tails) and negative charges (RNA). The lower panels of Figure 4.6 show the potential calculated for the virus without (Figure 4.6c) and with (Figure 4.6d) RNA, mapped onto the surface of the empty capsid. The positively charged tails (blue in Figure 4.6c) are fully neutralized and even reveal some negative potential on their surface due to the close proximity of RNA. The

latter observation is probably due to the fact that the total charge of RNA is almost factor of two greater than that of the capsid.

CONCLUSIONS

There are three pieces of evidence that the polycationic protein tails penetrate deeply into the interior of nodavirus capsids. First, mutant FHV that lack 30 N-terminal amino acids lack the 32Å peak seen in cryo-EM radial density distribution profiles for wild-type FHV (19). Second, our model 8 reproduces the experimental radial density distribution much better than model 12, and tails in the former penetrate much deeper into the capsid than those in the latter model. Finally, electrostatic calculations show that deep penetration of the tails has a stabilizing effect, because of more efficient neutralization of the RNA charge.

This observation has important implications for viral assembly.
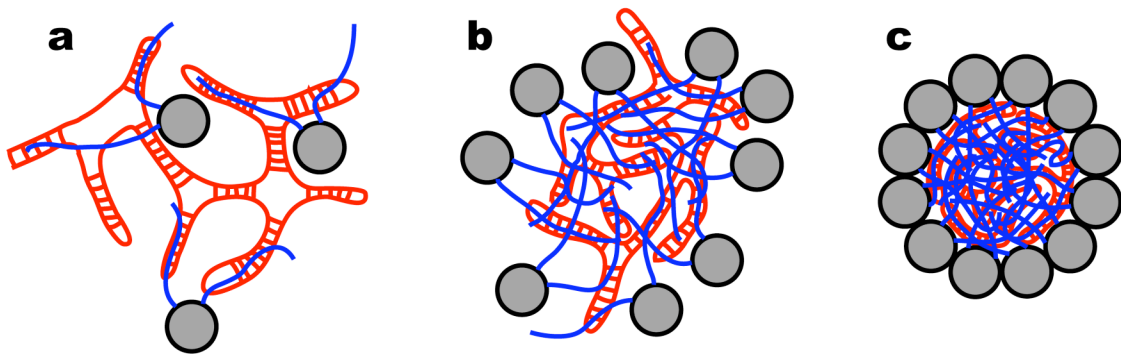
Figure 4.7: Model for assembly of small icosahedral RNA viruses. **a.** The polycationic N- and C-terminal protein tails bind nonspecifically to the RNA genome. **b.** When enough proteins are bound and the RNA charge is sufficiently neutralized, the complex collapses, in a process much like the condensation of DNA by polyvalent cations. The globular domains of the capsid proteins are tethered to the condensed RNA but are squeezed out and form a shell around it. **c.** The local concentration of the globular domains is high enough to promote oligomerization, leading to the formation of the mature capsid.

The assembly of small icosahedral RNA viruses like PaV and FHV is quite different from bacteriophage. Interactions between phage capsid proteins are strong

enough that capsids assembly spontaneously. The DNA genome is then forced into the pre-formed capsid by an ATP-dependent motor; there is little or no attraction between the DNA and the capsid proteins, in order to promote ejection of the genome upon infection of the host bacterium. In contrast, protein-protein interactions are weak in nodaviruses (capsids do not assemble spontaneously), and RNA-protein interactions are strongly attractive.

We propose a simple mechanism for the assembly of nodaviruses. Positively charged protein tails bind to the RNA (Figure 4.7a), with RNA replication, protein synthesis and RNA-protein binding occurring very closely in time and space (28, 29). When a sufficient quantity of the RNA charge is neutralized, the resulting complex collapses in a process reminiscent of DNA condensation (Figure 4.7b). We believe that most of these interactions are nonspecific, although in the mature virus there is evidence of a specific interaction between RNA2 and the N-terminal tail (30). In addition, the crystal structure (6) shows ordered interactions between the RNA and 36 N-terminal residues of subunit A, and between the RNA and eight residues of the C-terminus of subunit A, although the identity of the RNA in those interactions cannot be determined. We hypothesize that the globular domains of the capsid proteins are squeezed to the outside of the collapsed state, as shown in figure 4.7b. This provides a sufficiently high local concentration that the relatively weak protein-protein affinity is overcome, leading to oligomerization and the formation of the mature capsid (Figure 4.7c).

One remarkable observation suggests that this mechanism might apply to many single-stranded viruses. Belyi and Muthukumar examined 16 wild-type and 3 mutant viruses (both DNA and RNA viruses) with genomes ranging from about 1 kb to 12 kb (31). They found that the ratio of the genome size to the net charge on the terminal protein tails is 1.61 ± 0.03, an unexpectedly uniform ratio. Such a narrow range might be explained by our model, because the initial collapse would require sufficient charge neutralization to overcome RNA-RNA repulsions, but not so much as to lock the

condensed state into a fixed configuration that could preclude the structural flexibility necessary for fitting the condensed mass into the final capsid structure.

This model provides a simple mechanistic basis for explaining how the relatively weakly associating proteins can force RNA into a small compact volume: the very strong electrostatic interactions between the polyanionic RNA and the polycationic protein tails provide a sufficiently favorable change in enthalpy to overcome the unfavorable entropic penalty associated with the dramatic reduction in RNA conformational space. It seems highly unlikely that a compact RNA structure would form first, followed by the formation of the protein capsid around it, as suggested earlier (15). The former is opposed by very strong forces, while the latter is driven by only weak ones.

In summary, we present the first all-atom model of a complete T=3 virus. Although there are insufficient experimental data to allow the development of a completely rigorous model, our model is consistent with all the available data, and it is sterically plausible. Most important, it leads to a simple mechanistic explanation of the assembly of small icosahedral RNA viruses. It will be exciting to test this model both experimentally and computationally.

References

1.   Zeddam, J. L., J. L. Rodriguez, M. Ravallec, and A. Lagnaoui. 1999. A noda-like virus isolated from the sweetpotato pest spodoptera eridania (Cramer) (Lep.; noctuidae). J Invertebr Pathol 74:267-274.

2.   Krishna, N. K., and A. Schneemann. 1999. Formation of an RNA heterodimer upon heating of nodavirus particles. Journal of virology 73:1699-1703.

3.   Johnson, K. N., J. L. Zeddam, and L. A. Ball. 2000. Characterization and construction of functional cDNA clones of Pariacoto virus, the first Alphanodavirus isolated outside Australasia. Journal of virology 74:5123-5132.

4.   Johnson, K. N., L. Tang, J. E. Johnson, and L. A. Ball. 2004. Heterologous RNA encapsidated in Pariacoto virus-like particles forms a dodecahedral cage similar to genomic RNA in wild-type virions. Journal of virology 78:11371-11378.

5.   Johnson, K. N., and L. A. Ball. 2003. Virions of Pariacoto virus contain a minor protein translated from the second AUG codon of the capsid protein open reading frame. The Journal of general virology 84:2847-2852.

6.   Tang, L., K. N. Johnson, L. A. Ball, T. Lin, M. Yeager, and J. E. Johnson. 2001. The structure of pariacoto virus reveals a dodecahedral cage of duplex RNA. Nature structural biology 8:77-83.

7.   Schneemann, A., W. Zhong, T. M. Gallagher, and R. R. Rueckert. 1992. Maturation cleavage required for infectivity of a nodavirus. Journal of virology 66:6728-6734.

8.   Fisher, A. J., B. R. McKinney, J. P. Wery, and J. E. Johnson. 1992. Crystallization and preliminary data analysis of Flock House virus. Acta crystallographica 48 (Pt 4):515-520.

9.   Mori, K., T. Nakai, K. Muroga, M. Arimoto, K. Mushiake, and I. Furusawa. 1992. Properties of a new virus belonging to nodaviridae found in larval striped jack (Pseudocaranx dentex) with nervous necrosis. Virology 187:368-371.

10.   Reinganum, C., J. B. Bashiruddin, and G. F. Cross. 1985. Boolarra virus: a member of the Nodaviridae isolated from Oncopera intricoides (Lepidoptera: Hepialidae). Intervirology 24:10-17.

11.   Klug, A. 1999. The tobacco mosaic virus particle: structure and assembly. Philosophical transactions of the Royal Society of London 354:531-535.

12.   Namba, K., R. Pattanayek, and G. Stubbs. 1989. Visualization of protein-nucleic acid interactions in a virus. Refined structure of intact tobacco mosaic virus at 2.9 A resolution by X-ray fiber diffraction. Journal of molecular biology 208:307-325.

13.   Namba, K., and G. Stubbs. 1986. Structure of tobacco mosaic virus at 3.6 A resolution: implications for assembly. Science (New York, N.Y 231:1401-1406.

14. Reddy, V. S., H. A. Giesing, R. T. Morton, A. Kumar, C. B. Post, C. L. Brooks, 3rd, and J. E. Johnson. 1998. Energetics of quasiequivalence: computational analysis of protein-protein interactions in icosahedral viruses. Biophysical journal 74:546-558.

15. Freddolino, P. L., A. S. Arkhipov, S. B. Larson, A. McPherson, and K. Schulten. 2006. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. Structure 14:437-449.

16. Zhang, D., R. Konecny, N. A. Baker, and J. A. McCammon. 2004. Electrostatic interaction between RNA and protein capsid in cowpea chlorotic mottle virus simulated by a coarse-grain RNA model and a Monte Carlo approach. Biopolymers 75:325-337.

17. Konecny, R., J. Trylska, F. Tama, D. Zhang, N. A. Baker, C. L. Brooks, 3rd, and J. A. McCammon. 2006. Electrostatic properties of cowpea chlorotic mottle virus and cucumber mosaic virus capsids. Biopolymers 82:106-120.

18. Chin, K., K. A. Sharp, B. Honig, and A. M. Pyle. 1999. Calculating the electrostatic properties of RNA provides new insights into molecular interactions and function. Nature structural biology 6:1055-1061.

19. Tihova, M., K. A. Dryden, T. V. Le, S. C. Harvey, J. E. Johnson, M. Yeager, and A. Schneemann. 2004. Nodavirus coat protein imposes dodecahedral RNA structure independent of nucleotide sequence and length. Journal of virology 78:2897-2905.

20. Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. 2000. The Protein Data Bank. Nucleic acids research 28:235-242.

21. Shepherd, C. M., I. A. Borelli, G. Lander, P. Natarajan, V. Siddavanahalli, C. Bajaj, J. E. Johnson, C. L. Brooks, 3rd, and V. S. Reddy. 2006. VIPERdb: a relational database for structural virology. Nucleic acids research 34:D386-389.

22. Malhotra, A., R. K. Tan, and S. C. Harvey. 1994. Modeling large RNAs and ribonucleoprotein particles using molecular mechanics techniques. Biophysical journal 66:1777-1795.

23. Baker, N. A., D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. Proceedings of the National Academy of Sciences of the United States of America 98:10037-10041.

24. Dolinsky, T. J., J. E. Nielsen, J. A. McCammon, and N. A. Baker. 2004. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic acids research 32:W665-667.

25. Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. Journal of computational chemistry 25:1605-1612.

26. VanLoock, M. S., B. A. Harris, and S. C. Harvey. 1998. To knot or not to knot? Examination of 16S ribosomal RNA models. J Biomol Struct Dyn 16:709-713.

27. Frank, J., M. Radermacher, P. Penczek, J. Zhu, Y. Li, M. Ladjadj, and A. Leith. 1996. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. Journal of structural biology 116:190-199.

28. Venter, P. A., N. K. Krishna, and A. Schneemann. 2005. Capsid protein synthesis from replicating RNA directs specific packaging of the genome of a multipartite, positive-strand RNA virus. Journal of virology 79:6239-6248.

29. Venter, P. A., and A. Schneemann. 2007. Assembly of two independent populations of flock house virus particles with distinct RNA packaging characteristics in the same cell. Journal of virology 81:613-619.

30. Marshall, D., and A. Schneemann. 2001. Specific packaging of nodaviral RNA2 requires the N-terminus of the capsid protein. Virology 285:165-175.

31. Belyi, V. A., and M. Muthukumar. 2006. Electrostatic origin of the genome packing in viruses. Proceedings of the National Academy of Sciences of the United States of America 103:17174-17178.

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

Molecular models are any form of representations of the structures of molecules. Models can be physical objects, simple figures, computational representations, etc. Models incorporate structural data available for the molecules, and get better as more data are integrated into building it. Models are built because visualizing the structure is more helpful and intuitive than raw data in the form of points or density maps. A molecular model may be used for various purposes ranging from understanding structure-function relationships of the molecules, refining experimentally determined structures, and proposing new hypotheses that can be experimentally tested and verified.. Computer-based models are essential in cases when structural determination of molecules may be too expensive or impossible using experimental technologies such as x-ray crystallography or NMR.

I used molecular modeling to study ribonucleoprotein complexes that could not be studied using other experimental techniques. In each of the systems, I incorporated all the structural and experimental data available to propose the most accurate models possible. Using the derived models, I was able to postulate new hypothesis on translational fidelity and viral assembly pathway.

Structural basis of translational fidelity

The most important contribution of my work is the generation of two models for the *E. coli* ribosome in different states of the translational cycle. The PRE and POST-accommodation states of the ribosome have been captured by cryo-EM (see Chapter 1). The cryo-EM maps are at 8.9Å resolution, which is the highest resolution for the *E. coli* ribosome currently available (46). In conjunction with the density map, I used the high-

resolution crystal structure of the *E. coli* ribosome available at 3.5Å (40) to propose these two models. I used Yup.scx module of our in-house molecular mechanics package, YUP, to flexibly fit the crystal structure into the low resolution density maps (62). Currently, it is not possible to generate crystal structures for these two states, as it is very hard, if not impossible, to capture the ribosomes in these two states and get crystals from them.

On the basis of these models, I proposed a new hypothesis on how the ribosome selects cognate tRNA with high fidelity. It has been shown that the high fidelity of the translation is due to selection of the tRNA that occurs in two steps, during initial selection and then during proofreading (42, 43). 1% of the near-cognate tRNAs pass the initial selection step. Of the near-cognate tRNAs pass through the initial selection only 1% of the tRNAs make it to the PTC. The accommodation rates for the cognate tRNA and the near-cognate tRNA are $7s^{-1}$ and $0.1s^{-1}$ respectively (44, 63). When tRNA enters the ribosome and makes initial contact with the mRNA, it induces conformational changes within the ribosome (64, 65).

Earlier hypothesis have suggested that the conformational changes induced during initial selection induces a signal that gets propagated from the decoding center in the SSU to the PTC in the LSU (44, 49, 50, 63). This signal induces changes in the ribosome that allows the cognate tRNA to get to the PTC and undergo peptidyl transfer and in the mean time reject near-cognate tRNAs from getting to the PTC. The mechanism or the residues involved in this signal transformation are not known, even with the generation of all the high-resolution structures. However, from the models that I generated for PRE (with tRNA away from the PTC) and POST (with tRNA at the PTC), I noticed that there are no significant structural changes in the tRNA corridor within the ribosome from PRE to POST. I also noticed that the path for tRNA is very narrow, which is unlike what is generally believed to be that after EF-Tu falls off, tRNA has enough space to get to PTC. The narrow space actually provides energetic barrier that the tRNA has to overcome before getting to the PTC. Based on these observations, I propose that the ribosome

provides steric hindrances for the passage of tRNA from PRE to POST state, be it a cognate or near-cognate. This reduces the need for signal transformation from the decoding center to PTC. So, if there is no signal transferred from DC to PTC, how is there a differential rate between cognate and near-cognate tRNA during accommodation? The answer lies in the kinking of tRNA that occurs during initial selection of tRNA. The kinking of tRNA is not its intrinsic property, thus the energy required for kinking has to be compensated by the favorable interactions that the tRNA makes with the ribosome. I propose that there is more kinking, thus more conformational energy stored, in cognate tRNA because of more interactions of it with the ribosome as compared to near-cognate. Thus, when EF-Tu falls off the ribosome leaving the tRNA, cognate tRNA unkinks and has enough energy to squeezes through the narrow corridor provided by the ribosome and get to the PTC, while near-cognate tRNA falls off. This explains the differential accommodation rate for the cognate and near-cognate tRNA.

MD simulations are being performed by Thomas Caulfied on the models with some biasing to move the A/T state tRNA to the A/A state within the ribosome, to test the hypothesis presented. The transition states from the simulations will show the tRNA pathway and the different residues that are hindered. This would show that tRNA does not have a free path to the PTC, rather it needs some energy to go over the barrier created by ribosome. The availability of these models that are based on cryo-EM, would give a more realistic pathway to tRNA as compared to earlier studies because the earlier models were homology models for the ribosome (66). With the aid of MD we can get the transition state of tRNA as it moves from PRE accommodation to the POST accommodation state. Similarly, cryo-EM map of the ribosome with near-cognate tRNA and EF-Tu could actually prove the differential kinking between cognate and near-cognate tRNA during initial selection. Currently, cryo-EM data is only available for the PRE-state with cognate tRNA and EF-Tu. If somehow near-cognate tRNA could also be

captured, we would be able to compare the differential kinking in tRNAs, which would test our hypothesis.

The structure of eukaryotic ribosome

The eukaryotic ribosome, because of its complexity and unavailability of the crystal structures, has not received as much attention as its prokaryotic counterpart. It consists of more rRNA and proteins in both subunits. Part of the rRNA of the subunits form a core that is homologous to the prokaryotic rRNA. The additional eukaryotic rRNAs are found mostly in the periphery and are called expansion segments (ES). The eukaryotic ribosome has a higher fidelity rate (10-100x) in translation compared to the prokaryotic ribosome, despite the fact that the core where decoding and peptidyl transfer take place is almost identical to the prokaryotic version. The only apparent differences are the ES that are located far away from the active sites. The obvious question is what role do these ES play in translational fidelity, if any, in eukaryotes? Or are they involved in some other functions? To answer these questions, we decided to start by modeling the structure of the ES using the cryo-EM data available for the fungus, *T. lanuginosus*. Understanding the structure of the rRNA and proteins and their interactions will help explain the mechanism of translation and the role of these ES in eukaryotes.

Based on the available secondary structure information and using cryo-EM density maps as constraints, I generated models for 95% of the eukaryotic ribosomal RNA molecules. I used homology modeling and software for predicting 3-D from the secondary structure, MC-SYM, to propose the first generation all-atom model of the eukaryotic rRNAs. The models that I generated are the first-order model for the structure of eukaryotic rRNA molecules. The rRNA model with ribosomal proteins will help us define the RNA/protein boundaries in the cryo-EM map.

In addition to exploring the role of the ES, this model could also be used as a starting model when the eukaryotic ribosome is crystallized and diffraction data obtained

from these crystals. It could be used for overcoming the phase problems while solving the structure. Just like any other model, as more data become available – the quality of the model will improve. Understanding the eukaryotic ribosome in greater detail has therapeutic implications as well. One could use the structural difference between prokaryotic and eukaryotic ribosome to design drugs. It may also shed some light on the most glaring question in the field of ribosome: why does the ribosome need to be so large?

Structural and Electrostatic Characterizations of Pariacoto virus

Finally, I generated an all-atom model for PaV. This project started with a series of questions such as what is the structure of the RNA genome inside the dodecahedral cage? Where are the missing N- and C-terminal protein tails located? How do the electrostatic interactions between the RNA and the proteins stabilize the virus? Crystal structures for the asymmetric unit of the virus at 3.0Å resolution and the cryo-EM map at 20Å resolution do not give any structural information about the core of the virus. It is because the inside of the virus is not symmetric. Icosahedral averaging during structure determination eliminates the structural data needed to solve the structure. Using coarse-grained modeling, I was able to propose a plausible first-generation model for PaV. The coarse-grained model was interpolated to an all-atom model such that electrostatic characterization of the virus could be done.

To test the effect of the missing N-terminus tails, we generated two models that differed in VDW contact distance between the RNA and protein. One had a contact distance of 8Å (model_8) and the other had a contact distance of 12Å (model_12). More protein tails penetrated the RNA genome in model_8 as compared to model_12. Model_8 also had lower electrostatic energy, implying that the protein tails need to penetrate inside so as to stabilize this highly charged RNA. The fact that the N-term tails are highly basic also suggests that they play a crucial role in stabilizing the large negative charge on RNA

genome. So if the protein penetrates inside, and RNA and protein package together in RNA virus, then the RNA and protein should have extensive interactions even during pre-assembly. Reminiscent to DNA condensation by positively charged ions, the positively charged residues in the N-terminus tail could condense the RNA and protein components together, thus by assembling the virus. When the proteins that are interacting with RNA assemble to a virus, the RNA interacting with the A-subunit makes up the edges of the dodecahedral cage while the ones interacting with the B and C-subunits should be in the core. The globular portion of the capsomers lie at the outside forming the capsid.

The model that we generated consists of few deficiencies. During modeling electrostatic effects were ignored. This is inaccurate because the electrostatic effects contribute significantly in the packaging. Also, we generated the genome first and then added protein around it. In RNA viruses, RNA and protein package together. In future, this model could be refined using electrostatics during minimization. Also, if proteins and RNA could be expanded and minimized together to the actual size then the model could be further refined.

Summary

Overall, using computer-based approaches for molecular modeling, I generated models for various ribonucleoprotein complexes. In each case, analyses of the models were done within the limits of the experimental data used for modeling. The models for *E. coli* ribosome at two states were based on very high-resolution data. Because of this, I was able to deduce and analyze models at atomic details. Second project on the structure of the eukaryotic ribosome was based on data at lower resolution. No atomic-level interpretation of the data was done and the analyses were limited to locating protein/RNA and RNA/RNA boundaries. Finally, the models obtained for PaV were based on even lower resolution data. The analysis done on these models were very speculative. In

summary, the advantages of computer-based modeling in being able to integrate data in any level of resolution and to generate models have been shown. Careful analysis of the models could lead to new hypotheses, which may lead to the design of innovative experiments, which may someday lead to the understanding of fundamental biological phenomenon such as translational fidelity and viral assembly.

APPENDIX A

ANALYSIS OF PRE- AND POST- ACCOMMODATION MODELS

Table A.1: Root mean square displacements in the large subunit.
When the PRE and POST models of the large subunit are superposed, the RMSD of the atoms in the 5S RNA is 0.89Å. The contribution of intramolecular conformational changes to this motion is 0.24Å, obtained by directly superposing the two models of 5S RNA. Intramolecular motions in the 23S RNA are analyzed in figures 2.3 and 2.5. See text in chapter 2 for discussion of L9.

Chain	RMSD (Å): PRE *vs.* crystal	RMSD (Å): POST *vs.* crystal	RMSD (Å): PRE *vs.* POST (Global)	RMSD (Å): PRE *vs.* POST (Intramolecular)
5S	**3.58**	**3.12**	**0.89**	**0.24**
23S	**4.02**	**3.77**	**1.24**	**0.94**
All Proteins	**4.46**	**3.83**	**2.17**	–
L2	2.69	2.52	1.39	0.75
L3	2.81	2.74	0.90	0.56
L4	2.89	2.25	0.94	0.49
L5	2.97	2.96	0.99	0.56
L6	3.66	3.48	0.66	0.53
L9	*11.46**	*11.58**	*2.91**	*2.48**
L11	11.22	6.16	9.05	0.75
L13	2.68	2.63	0.46	0.30
L14	3.00	2.87	0.93	0.71
L15	3.48	2.72	1.15	0.61
L16	2.58	2.45	0.68	0.49
L17	2.95	2.58	0.77	0.21
L18	4.67	4.00	0.96	0.40
L19	3.71	3.34	1.01	0.57
L20	3.04	3.42	1.35	1.28
L21	3.32	3.25	0.67	0.63
L22	1.74	1.27	0.88	0.49
L23	2.39	2.71	1.54	0.75
L24	2.91	2.27	1.06	0.51
L25	3.90	3.47	0.77	0.58
L27	3.25	2.94	0.82	0.41
L28	3.10	2.02	1.33	0.30
L29	3.74	2.69	1.55	0.37
L30	2.97	2.87	0.60	0.44
L32	3.33	2.79	1.01	0.60
L33	2.12	1.28	0.97	0.54
L34	2.74	2.21	1.01	0.13
L35	2.71	2.74	0.74	0.64

Table A.2: Root Mean Square Displacements in the Small Subunit

RMSDs, determined using the superposition procedure described in the caption to Table A.1. Intramolecular RMSDs within the 16S RNA are analyzed in Figure 2.3b.

Chain	RMSD (Å): PRE *vs.* crystal	RMSD (Å): POST *vs.* crystal	RMSD (Å): PRE *vs.* POST (Global)	RMSD (Å): PRE *vs.* POST (Intramolecular)
16S	**4.42**	**3.65**	**2.69**	**0.54**
All Proteins	**4.32**	**4.07**	**1.49**	–
S2	3.32	3.49	2.27	0.80
S3	3.55	3.36	1.41	0.75
S4	5.02	5.30	1.74	0.70
S5	2.99	3.07	1.64	0.60
S6	5.31	4.21	2.32	0.56
S7	5.98	5.13	1.93	1.67
S8	1.77	1.98	1.72	0.62
S9	4.92	4.25	1.84	0.84
S10	4.08	3.64	1.69	0.95
S11	4.91	4.51	1.61	0.53
S12	3.16	3.19	1.99	1.59
S13	6.97	6.46	1.46	0.78
S14	5.01	4.59	0.96	0.56
S15	2.75	2.40	1.58	0.54
S16	4.00	4.34	1.63	0.62
S17	1.48	1.91	1.17	0.65
S18	3.74	3.41	2.00	0.70
S19	6.14	5.70	0.83	0.44
S20	2.05	1.88	1.69	0.48
S21	4.46	3.74	2.02	0.66

Table A.3: Inter-phosphate Distances Hindering tRNA Movement During Accommodation.

The table reports phosphate-phosphate distances between nucleotides in LH92 (rows) and nucleotides in the A-site finger (LH38) (columns). The tRNA T/acceptor stem is about 65Å long from ψ55/C56 in the elbow to A76 at the 3'-acceptor terminus. Distances shorter than this create steric conflicts and are shown in red. Distances are rounded to the nearest 1Å.

H92 / H38	895	896	897	898	899
2554	82	77	76	71	72
2555	78	74	72	67	68
2556	75	70	68	63	64
2557	69	64	63	58	60
2558	67	62	62	57	59
2559	69	64	64	60	62
2560	74	68	69	65	68
2561	78	73	74	70	73
2562	82	76	77	73	76

Table A.4: Cross-Correlation Coefficients Between Models and Density Maps

	Before YUP Fit	After YUP Fit
Pre-Accommodation:		
30S	0.80	0.87
50S	0.82	0.86
tRNAs[1]	0.76	0.84
Post-Accommodation:		
30S	0.78	0.84
50S	0.78	0.83
tRNAs[2]	0.69	0.80

Notes:

      1: Includes tRNAs in the A/T, P and E sites

      2: Includes tRNAs in the A/A, P and E sites

The cross-correlation coefficient (CCC) between the experimental density map r^{exp} and a density map from a model r^{model} is defined as

$$CCC = \frac{\sum_{ijk} \rho^{exp}(i,j,k)\rho^{model}(i,j,k)}{\left[\sum_{ijk}\left(\rho^{exp}(i,j,k)\right)^2\left(\rho^{model}(i,j,k)\right)^2\right]^{1/2}}$$

where the sums are taken over all voxels *i,j,k* in the maps.

Table A.5: Intersubunit Motions

The top 200 distance differences for all pairs of residues within 30Å of one another have been examined. Sets of distances were clustered together if the residues lie in the same helices. The pair with the largest deviation for each region is shown in the table.

30S	50S	PRE distance (Å)	POST distance (Å)	Difference (Å)	Remarks:
1409 Top of the decoding center (SH 44)	1914 (LH69)	10.7	7.3	3.4 (37.4%)	This is the largest percent difference in the intersubunit regions. See text and Fig 2.4.
685 (SH 23)	2145 (LH78-L1BD)	18.1	22.4	4.3 (21.3%)	L1BD is at the periphery so this change is not interesting.
898 (SH27)	720 (LH34)	13.3	16.3	3.0 (20.6%)	LH34 is very far from the PTC so chances of it being involved in signal transmission is low.
1333 (Bottom of SH42)	2309 (LH84)	34.3	28.3	6.0 (19.3%)	SH42 is in the head of the small subunit and LH 84 is the P-site finger. Both regions are remote from the decoding center and PTC.

Table A.6: Motions Within the Large Subunit

We ranked the top 500 distance differences for all pairs of residues within 50Å of one another. Results are reported both as absolute distance differences and also as percentage differences. The pairs were clustered together by helices. These are the clusters that show the maximum differences. The pair with the largest deviation in each cluster is shown in the table.

Residue 1	Residue 2	PRE distance (Å)	POST distance (Å)	Difference (Å)	Remarks:
884 (ASF)	1066 (L11BD)	28.8	36.8	8.0	ASF does not move. The difference is due to the motion of L11BD which moves away from tRNA binding site in POST state. Most of the significant differences are in this cluster.
1067 (L11BD)	2470 (LH89)	30.9	25.4	5.5	This motion is also due to L11BD moving between two states.
2056 (near PTC)	2505 (near PTC)	6.7	10.5	3.7 (43%)	2405 moves relatively more than other residues, but it is quite far from the CCA end of both A and P-site tRNAs. (This is the biggest percentage difference within the 50S residues). Around the PTC the change is very small.

Table A.7: Motions Within the Small Subunit:

We ranked the top 500 distance differences for all pairs of residues within 50Å of one another. These are the regions that show the maximum motion. The pairs were clustered together by helices. These are the clusters that show the maximum differences. The pair with the largest deviation in each cluster is shown in the table.

Residue 1	Residue 2	PRE distance (Å)	POST distance (Å)	Difference (Å)	Remarks:
81	159	50.0	45.3	4.65	Between spur (helix 6) and helix 8; found at the bottom of the 30S.
1001	1031	19.7	15.2	4.53	Bases from the beak region. 1031 lies at the tip of the beak - so it moves a bit.
1177	1378	31.3	26.9	4.41	1177 lies at the cytoplasmic surface and 1378 lies around the neck (quite far).
1378	1505	38.5	42.6	3.91	Some motion between h43 and h45. This is closer to the decoding center. However, the initial distance is large (> 35Å)

AN EXAMPLE OF A MCSYM-SCRIPT

```
// MC-SYM script for modeling the ES12 of Small
// subunit
// lsu_es12.mcc
// Author          : Batsal Devkota
// Created On       : June 27, 2006

// Script MC-Sym for LSU ES12 helix

sequence (r A1 GUCCU UGUUA CUUAA UUGAA CGUGG ACAUU)
// Loop ES12

// Conformations ----------------------------------------

residue
(
//loop ES12
A1 A4    {helix }   1
A5 A6    { type_A  }10
A7 A9    { helix }  1
A10 A18 { type_A } 20
A19 A21 { helix }  1
A22 A23 { type_A } 10
A24 A27 { helix }  1
A28 A30 { type_A } 10
)

// Relations ----------------------------------------

connect
(
A1 A4 { helix }  1
A4 A5   { stack }  10
A5 A6  { stack } 10
A6 A7{ stack } 10
A7 A9  { helix } 1
A9 A10 { stack } 10
A10 A11  { stack } 10
A11 A12 { stack } 10
A12 A13  { adjacent } 30
A13 A14  { adjacent } 30
A14 A15 { adjacent } 30
A15 A16  { adjacent } 30
```

```
A16 A17  { stack } 10
A17 A18 { stack } 10
A18 A19 { stack } 10
A19 A21  {helix } 1
 A21 A22 { stack } 10
 A22 A23 { stack } 10
 A23 A24  { stack } 10
 A24 A27 { helix } 1
 A27 A28 { adjacent} 1
 A28 A29 { adjacent } 10
 A29 A30 {adjacent } 10
)

pair (

 A1 A27  (51)      10
 A2 A26  (51)       10
 A3 A25  (51)     10
 A4 A24  (51)     10
 A5 A23  { XVI }     10
 A6 A22  { XXVIII}     10
 A7 A21  (51)      10
 A8 A20  (51)      10
 A9 A19  (51)       10
 A10 A18  { VIII }   10
 )

// Building -------------------------------------------

lsu_es12 = backtrack
(
 (A1 A27)
 (A1 A2 A26)
 (A27 A28 A29 A30)
 (A2 A3 A25)
 (A3 A4 A24)
 (A4 A5 A23)
 (A5 A6 A22)
 (A6 A7 A21)
 (A7 A8 A20)
 (A8 A9 A19)
 (A9 A10 A18)
 (A10 A11)
 (A18 A17)
 (A11 A12 A13)
 (A17 A16 A15 A14)
)
```

```
// Constraint ------------------------------------------

lsu_es12_cache = cache(
lsu_es12
rmsd (1.0 align base_only no_hydrogen)
)

adjacency (lsu_es12 1.0 4.0)

res_clash
(
lsu_es12
fixed_distance 1.0
all
no_hydrogen
)

// Exploration ------------------------------------------

explore
(
lsu_es12_cache
fileName_pdb(
"/SGI/600/bdevkota/mcsym/yeast/models/lsu_es12 -1.1-
%03d.pdb" )
)
```
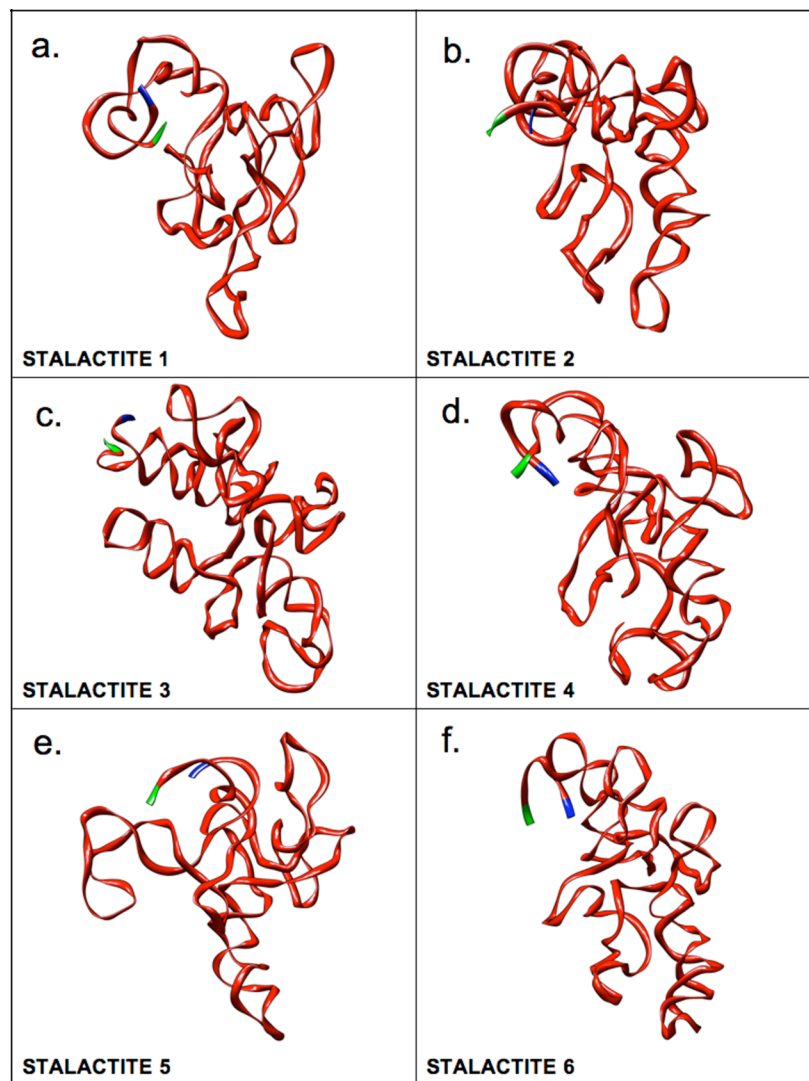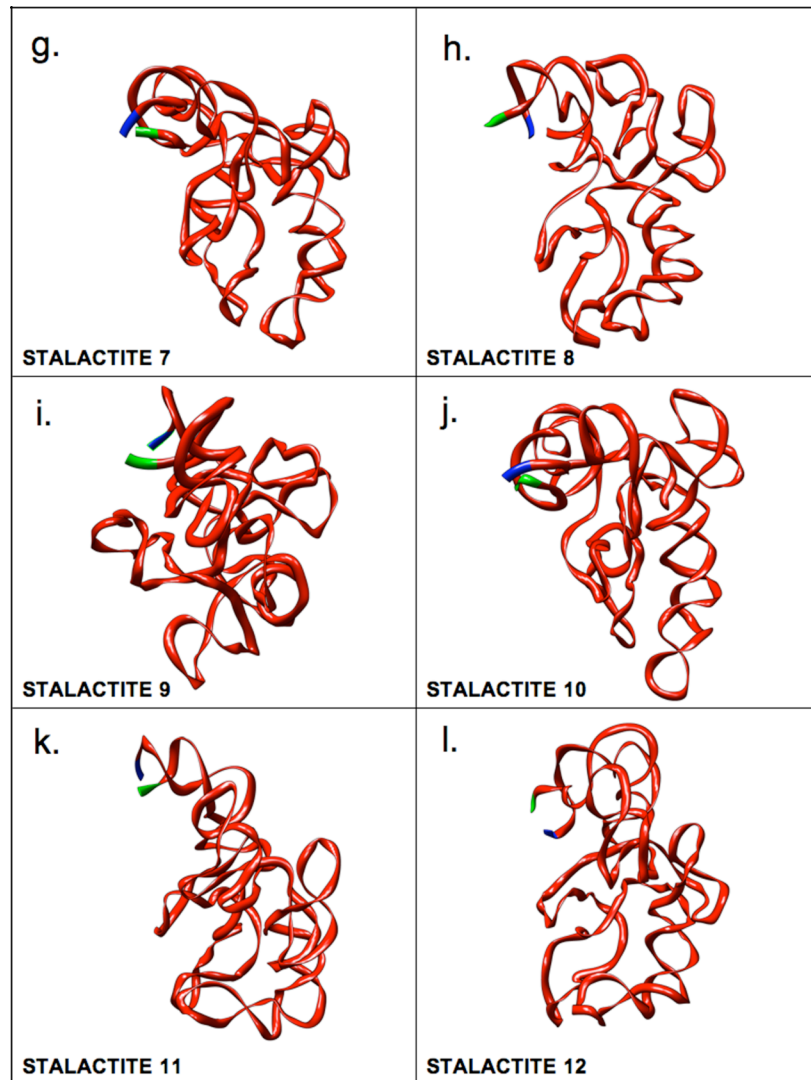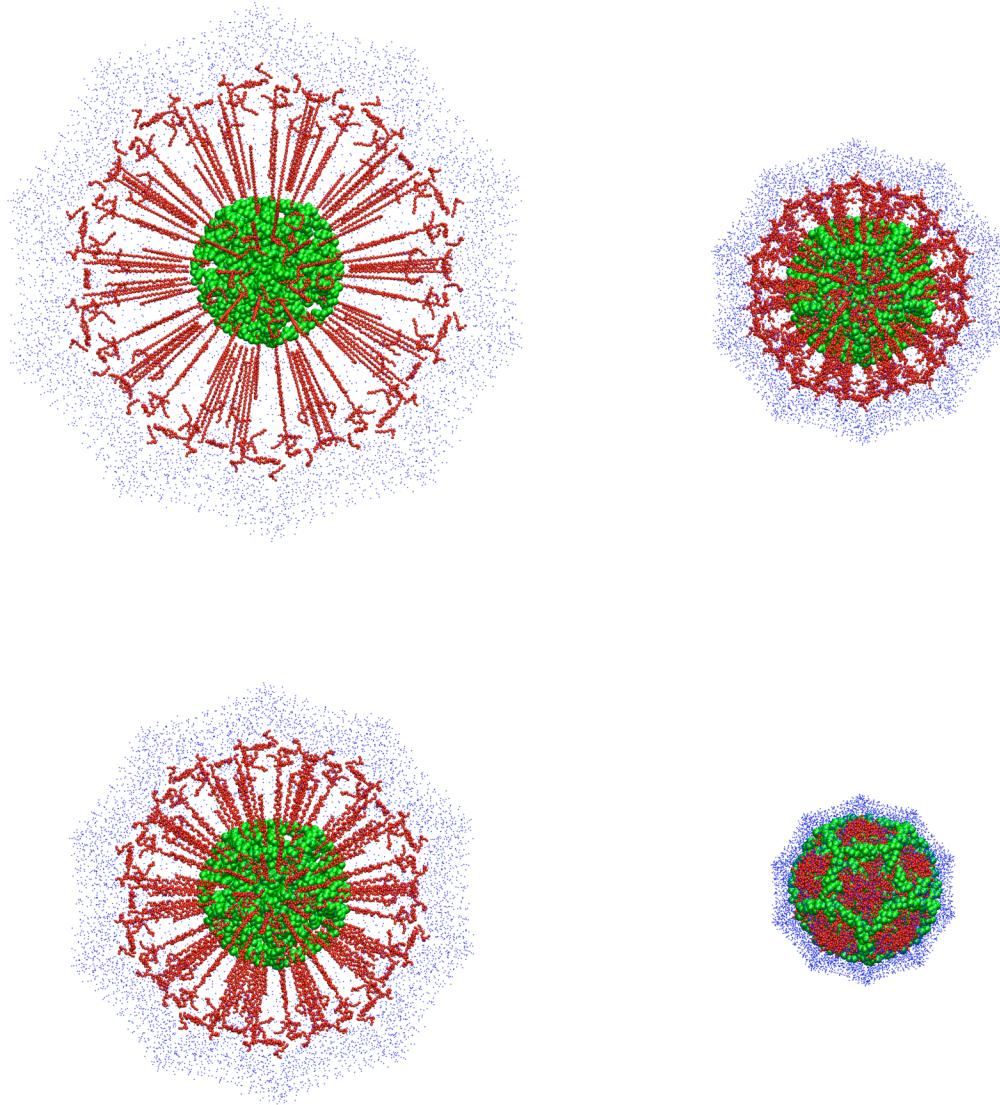
APPENDIX C

SUPPLEMENTARY FIGURES FOR THE STRUCUTRAL STUDIES

OF PARIACOTO VIRUS

Figure C.1: Stalactites have different conformations in the final model. 5' and 3' ends are labeled blue and green, respectively.

Figure C.2: Refinement of the protein tails (red) with RNA genome (green) at the center. Tails are added after expanding the capsid to three times its normal size. Refinement consists of successive rounds of scaling and minimization. The RNA genome (green) and globular regions of the capsid proteins (blue) are held fixed during minimization, while the flexible tails (red) are pulled toward the center of the virus and are allowed to adjust their conformations.

REFERENCES

1. Pauling, L., and R. B. Corey. 1951. Atomic coordinates and structure factors for two helical configurations of polypeptide chains. Proceedings of the National Academy of Sciences of the United States of America 37:235-240.

2. Watson, J. D., and F. H. Crick. 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 171:737-738.

3. Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. Journal of computational chemistry 25:1605-1612.

4. Walleczek, J., D. Schuler, M. Stoffler-Meilicke, R. Brimacombe, and G. Stoffler. 1988. A model for the spatial arrangement of the proteins in the large subunit of the Escherichia coli ribosome. The EMBO journal 7:3571-3576.

5. Brimacombe, R., J. Atmadja, W. Stiege, and D. Schuler. 1988. A detailed model of the three-dimensional structure of Escherichia coli 16 S ribosomal RNA in situ in the 30 S subunit. Journal of molecular biology 199:115-136.

6. Tan, R. K., and S. C. Harvey. 1993. Yammp: Development of a Molecular Mechanics Program Using the Modular Programming Method. J. Comp. Chem. 14:455-470.

7. Tama, F., O. Miyashita, and C. L. Brooks, 3rd. 2004. Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. Journal of structural biology 147:315-326.

8. Tama, F., O. Miyashita, and C. L. Brooks, 3rd. 2004. Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis. Journal of molecular biology 337:985-999.

9. Wriggers, W., R. A. Milligan, and J. A. McCammon. 1999. Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. Journal of structural biology 125:185-195.

10. Wriggers, W., and S. Birmanns. 2001. Using situs for flexible and rigid-body fitting of multiresolution single-molecule data. Journal of structural biology 133:193-202.

11. Korostelev, A., R. Bertram, and M. S. Chapman. 2002. Simulated-annealing real-space refinement as a tool in model building. Acta crystallographica 58:761-767.

12. Nissen, P., M. Kjeldgaard, S. Thirup, G. Polekhina, L. Reshetnikova, B. F. Clark, and J. Nyborg. 1995. Crystal structure of the ternary complex of Phe-tRNAPhe, EF-Tu, and a GTP analog. Science (New York, N.Y 270:1464-1472.

13. Kendrew, J. C., G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, and D. C. Phillips. 1958. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature 181:662-666.

14. Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. 2000. The Protein Data Bank. Nucleic acids research 28:235-242.

15. Hiller, S., G. Wider, T. Etezady-Esfarjani, R. Horst, and K. Wuthrich. 2005. Managing the solvent water polarization to obtain improved NMR spectra of large molecular structures. Journal of biomolecular NMR 32:61-70.

16. Adrian, M., J. Dubochet, J. Lepault, and A. W. McDowall. 1984. Cryo-electron microscopy of viruses. Nature 308:32-36.

17. Vogel, R. H., S. W. Provencher, C. H. von Bonsdorff, M. Adrian, and J. Dubochet. 1986. Envelope structure of Semliki Forest virus reconstructed from cryo-electron micrographs. Nature 320:533-535.

18. Agrawal, R. K., and J. Frank. 1999. Structural studies of the translational apparatus. Curr Opin Struct Biol 9:215-221.

19. Agrawal, R. K., R. K. Lata, and J. Frank. 1999. Conformational variability in Escherichia coli 70S ribosome as revealed by 3D cryo-electron microscopy. Int J Biochem Cell Biol 31:243-254.

20. Frank, J., P. Penczek, R. Grassucci, and S. Srivastava. 1991. Three-dimensional reconstruction of the 70S Escherichia coli ribosome in ice: the distribution of ribosomal RNA. The Journal of cell biology 115:597-605.

21. Frank, J. 2002. Single-particle imaging of macromolecules by cryo-electron microscopy. Annu Rev Biophys Biomol Struct 31:303-319.

22. Crick, F. H. 1958. On protein synthesis. Symposia of the Society for Experimental Biology 12:138-163.

23. Crick, F. H. 1962. The genetic code. Scientific American 207:66-74.

24. Tissieres, A., H. K. Mitchell, and U. M. Tracy. 1974. Protein synthesis in salivary glands of Drosophila melanogaster: Relation to chromosome puffs. Journal of molecular biology 85:389-398.

25. Tissieres, A., and J. D. Watson. 1958. Ribonucleoprotein particles from Escherichia coli. Nature 182:778-780.

26. Palade, G. E. 1955. A small particulate component of the cytoplasm. The Journal of biophysical and biochemical cytology 1:59-68.

27. Lake, J. A. 1981. The ribosome. Scientific American 245:84-97.

28. Lake, J. A. 1982. Ribosomal subunit orientations determined in the monomeric ribosome by single and by double-labeling immune electron microscopy. Journal of molecular biology 161:89-106.

29. Schuwirth, B. S., M. A. Borovinskaya, C. W. Hau, W. Zhang, A. Vila-Sanjurjo, J. M. Holton, and J. H. Cate. 2005. Structures of the bacterial ribosome at 3.5 A resolution. Science (New York, N.Y 310:827-834.

30. Cech, T. R. 2000. Structural biology. The ribosome is a ribozyme. Science 289:878-879.

31. Yonath, A., J. Mussig, and H. G. Wittmann. 1982. Parameters for crystal growth of ribosomal subunits. J Cell Biochem 19:145-155.

32. Schluenzen, F., A. Tocilj, R. Zarivach, J. Harms, M. Gluehmann, D. Janell, A. Bashan, H. Bartels, I. Agmon, F. Franceschi, and A. Yonath. 2000. Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. Cell 102:615-623.

33. Wimberly, B. T., D. E. Brodersen, W. M. Clemons, Jr., R. J. Morgan-Warren, A. P. Carter, C. Vonrhein, T. Hartsch, and V. Ramakrishnan. 2000. Structure of the 30S ribosomal subunit. Nature 407:327-339.

34. Ban, N., P. Nissen, J. Hansen, P. B. Moore, and T. A. Steitz. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. Science 289:905-920.

35. Yusupov, M. M., G. Z. Yusupova, A. Baucom, K. Lieberman, T. N. Earnest, J. H. Cate, and H. F. Noller. 2001. Crystal structure of the ribosome at 5.5 A resolution. Science 292:883-896.

36. Shi, H., and P. B. Moore. 2000. The crystal structure of yeast phenylalanine tRNA at 1.93 A resolution: a classic structure revisited. Rna 6:1091-1105.

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

38. Hansson, S., R. Singh, A. T. Gudkov, A. Liljas, and D. T. Logan. 2005. Crystal structure of a mutant elongation factor G trapped with a GTP analogue. FEBS Lett 579:4492-4497.

39. Wilson, D. N., F. Schluenzen, J. M. Harms, T. Yoshida, T. Ohkubo, R. Albrecht, J. Buerger, Y. Kobayashi, and P. Fucini. 2005. X-ray crystallography study on ribosome recycling: the mechanism of binding and action of RRF on the 50S ribosomal subunit. Embo J 24:251-260.

40. Selmer, M., C. M. Dunham, F. V. t. Murphy, A. Weixlbaumer, S. Petry, A. C. Kelley, J. R. Weir, and V. Ramakrishnan. 2006. Structure of the 70S ribosome complexed with mRNA and tRNA. Science (New York, N.Y 313:1935-1942.

41. Weixlbaumer, A., S. Petry, C. M. Dunham, M. Selmer, A. C. Kelley, and V. Ramakrishnan. 2007. Crystal structure of the ribosome recycling factor bound to the ribosome. Nature structural & molecular biology 14:733-737.

42. Hopfield, J. J. 1974. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. Proceedings of the National Academy of Sciences of the United States of America 71:4135-4139.

43. Ninio, J. 1974. A semi-quantitative treatment of missense and nonsense suppression in the strA and ram ribosomal mutants of Escherichia coli. Evaluation of some molecular parameters of translation in vivo. Journal of molecular biology 84:297-313.

44. Pape, T., W. Wintermeyer, and M. Rodnina. 1999. Induced fit in initial selection and proofreading of aminoacyl-tRNA on the ribosome. The EMBO journal 18:3800-3807.

45. Ogle, J. M., A. P. Carter, and V. Ramakrishnan. 2003. Insights into the decoding mechanism from recent ribosome structures. Trends in biochemical sciences 28:259-266.

46. Valle, M., A. Zavialov, W. Li, S. M. Stagg, J. Sengupta, R. C. Nielsen, P. Nissen, S. C. Harvey, M. Ehrenberg, and J. Frank. 2003. Incorporation of aminoacyl-tRNA into the ribosome as seen by cryo-electron microscopy. Nature structural biology 10:899-906.

47. Blanchard, S. C., R. L. Gonzalez, H. D. Kim, S. Chu, and J. D. Puglisi. 2004. tRNA selection and kinetic proofreading in translation. Nature structural & molecular biology 11:1008-1014.

48. Rodnina, M. V., M. Beringer, and W. Wintermeyer. 2007. How ribosomes make peptide bonds. Trends in biochemical sciences 32:20-26.

49. Rodnina, M. V., K. B. Gromadski, U. Kothe, and H. J. Wieden. 2005. Recognition and selection of tRNA in translation. FEBS Lett 579:938-942.

50. Rodnina, M. V., and W. Wintermeyer. 2001. Fidelity of aminoacyl-tRNA selection on the ribosome: kinetic and structural mechanisms. Annual review of biochemistry 70:415-435.

51. Stanley, W. M. 1935. Isolation of a Crystalline Protein Possessing the Properties of Tobacco-Mosaic Virus. Science (New York, N.Y 81:644-645.

52. Crick, F. H., and J. D. Watson. 1956. Structure of small viruses. Nature 177:473-475.

53. Caspar, D. L., and A. Klug. 1962. Physical principles in the construction of regular viruses. Cold Spring Harbor symposia on quantitative biology 27:1-24.

54. Harrison, S. C., Olson, A. J., Schutt, C. E. , Winkler, F. K., Bricogne, G. . 1978. Tomato bushy stunt virus at 2.9 Å resolution. Nature 276:368-373.

55. Wilson, I. A., J. J. Skehel, and D. C. Wiley. 1981. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution. Nature 289:366-373.

56. Hogle, J. M., M. Chow, and D. J. Filman. 1985. Three-dimensional structure of poliovirus at 2.9 A resolution. Science (New York, N.Y 229:1358-1365.

57. Bottcher, B., S. A. Wynne, and R. A. Crowther. 1997. Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. Nature 386:88-91.

58. Conway, J. F., N. Cheng, A. Zlotnick, P. T. Wingfield, S. J. Stahl, and A. C. Steven. 1997. Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. Nature 386:91-94.

59.	Lata, R., J. F. Conway, N. Cheng, R. L. Duda, R. W. Hendrix, W. R. Wikoff, J. E. Johnson, H. Tsuruta, and A. C. Steven. 2000. Maturation dynamics of a viral capsid: visualization of transitional intermediate states. Cell 100:253-263.

60.	Heymann, J. B., N. Cheng, W. W. Newcomb, B. L. Trus, J. C. Brown, and A. C. Steven. 2003. Dynamics of herpes simplex virus capsid maturation visualized by time-lapse cryo-electron microscopy. Nature structural biology 10:334-341.

61.	Major, F., and R. Griffey. 2001. Computational methods for RNA structure determination. Current opinion in structural biology 11:282-286.

62.	Tan, R. K. Z., A. S. Petrov, and S. C. Harvey. 2006. YUP: A Molecular Simulation Program for Coarse-Grained and Multiscaled Models. J. Chem. Theory Comput. 2:529-540.

63.	Gromadski, K. B., and M. V. Rodnina. 2004. Kinetic determinants of high-fidelity tRNA discrimination on the ribosome. Mol Cell 13:191-200.

64.	Ogle, J. M., D. E. Brodersen, W. M. Clemons, Jr., M. J. Tarry, A. P. Carter, and V. Ramakrishnan. 2001. Recognition of cognate transfer RNA by the 30S ribosomal subunit. Science (New York, N.Y 292:897-902.

65.	Ogle, J. M., F. V. Murphy, M. J. Tarry, and V. Ramakrishnan. 2002. Selection of tRNA by the ribosome requires a transition from an open to a closed form. Cell 111:721-732.

66.	Sanbonmatsu, K. Y., S. Joseph, and C. S. Tung. 2005. Simulating movement of tRNA into the ribosome during decoding. Proceedings of the National Academy of Sciences of the United States of America 102:15854-15859.

VITA

BATSAL DEVKOTA

Batsal Devkota was born in Dandeldhura, Nepal. He attended Guheshwori, St. Xavier's and Budhanilkantha High School in Kathmandu, Nepal for his schooling. He received B.Sc. in Biochemistry and Molecular Biology from the University of New Hampshire, Durham, NH in May 2000. He worked in Variagenics, Inc. in Cambridge, MA and Everest Biotech in Kathmandu, Nepal before coming to Georgia Tech to pursue a doctorate in Applied Biology. When he is not working on his research, Mr. Devkota enjoys playing chess, backgammon, tennis and watching football.