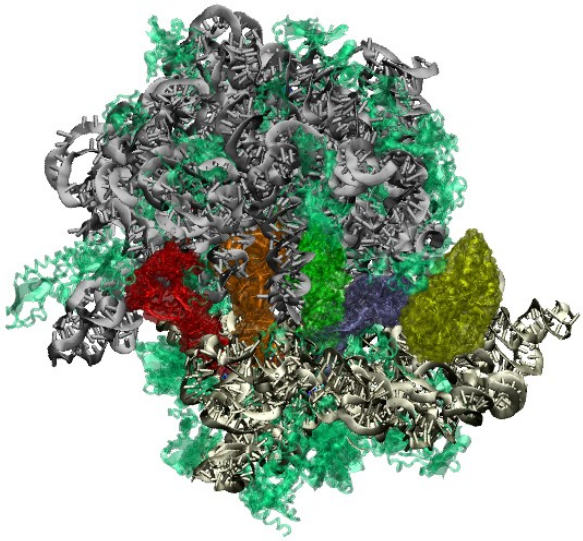
University of Illinois at Urbana-Champaign
Luthey-Schulten Group
NIH Resource for Macromolecular Modeling and Bioinformatics
Computational Biophysics Workshop

Evolution of Translation
The Ribosome



VMD Developer: John Stone

MultiSeq Developers
Elijah Roberts
John Eargle
Dan Wright

Tutorial Authors
Ke Chen
John Eargle
Tyler Earnest
Jonathan Lai
Zan Luthey-Schulten

April 2015

A current version of this tutorial is available at
`http://www.scs.illinois.edu/~schulten/tutorials/ribosome`

Contents

Introduction **3**
Requirements 4

1 The Ribosomal SSU and associated structures: [30 minutes] **4**

2 The Ribosome large subunit (LSU) and associated structures: [30 minutes] **9**
2.1 The peptidyl-transferase center 10

3 Ribosome Origins: [30 minutes] **11**
3.1 Hypothesis on the evolution of the ribosome 11

4 Ribosomal signatures: [60 minutes] **12**
4.1 Definition and classification of the ribosomal signatures 13
4.2 Contribution of ribosomal signatures to phylogenetic separation . 17
4.3 Functional roles of signatures in ribosomal assembly 20

5 Kinetic Model of Ribosome assembly: [30 minutes] **22**

Acknowledgements **26**

Introduction

The ribosome is a large structure found in all living cells that serves as the main translation machinery of the cell. Messenger RNA (mRNA), transcribed from the organism's genome, binds with the ribosome to commence translation to protein. As explained in the previous tutorials [1, 2, 3], many other cellular components, including tRNA, the aminoacyl-tRNA synthetases, and the elongation factors participate in the translation process; however, the ribosome is the central machinery that assembles a protein from a transcribed gene. Solving the structure of the ribosome was awarded the Nobel Prize in Chemistry in 2009 [4]. The bacterial ribosome (70S) consists of a small (SSU or 30S) and large (LSU or 50S) subunit, which bind together around a messenger RNA. Each subunit is made up of rRNA and proteins: the 30S subunit consists of the 16S rRNA subunit and 21 proteins, while the 50S subunit consists of the 23S rRNA subunit, the 5S rRNA subunit, and 34 proteins. The 'S' in this case refers to the Svedberg unit, a measurement of sedimentation during centrifugation, so these numbers do not necessary add up as they would if they referred to mass.

The primary function of the ribosome is to translate a sequence encoded on mRNA into a protein. Proteins are built as linear chains of amino acids by adding one amino acid at a time to a growing chain. The amino acid is characterized by two functional groups, an amino group and a carboxyl group. If the amino group of one amino acid is brought in close proximity to the carboxyl group of a second amino acid, a peptidyl transferase (aminoacyltransferase) reaction can occur, resulting in the loss of a molecule of water (one oxygen atom and one hydrogen atom from the carboxyl group and one hydrogen atom from the amino group) and the formation of a peptide bond between the two amino acids. As all amino acids contain one amino group and one carboxyl group, they can be joined together to form a large chain, or polypeptide. As this reaction occurs in a pocket of the LSU that is primarily rRNA, the ribosome can be thought of as an RNA enzyme, or ribozyme. Specifically, the reaction takes place inside the peptidyl transferase center, which will be discussed in more detail later in the tutorial.

In addition to the peptidyl transferase function, the ribosome also plays a role in maintaining the accuracy of translation by allowing the codon-anticodon interaction between the bound mRNA and the tRNA carrying the next putative amino acid to be joined to the nascent protein chain. The ordering of amino acids in a protein is directly translated from the sequence of nucleotides of the mRNA. The translation between nucleotide *codons*, triplets of nucleotides, and individual amino acids is known as the *genetic code*. This translation is mediated by charged transfer RNA (tRNA) molecules. Each tRNA is charged at the acceptor stem with a specific amino acid corresponding to its anticodon located in the anti-codon stem by an *AARS*. For more details how the *AARS* set the genetic code by charging the tRNAs, please see the tutorials on the amino-acyl tRNA synthetases [1, 2]. When a charged tRNA arrives at the A-site of the ribosome, the anticodon loop of the tRNA is oriented to interact with the next codon of the bound mRNA. If the codon is complementary to

the anticodon, it is released by its carrier molecule, the elongation factor Tu (EF-Tu). If the codon is not complementary to the anticodon, there is a high probability that the tRNA:EF-Tu complex will dissociate and another tRNA will bind. For more details on the behavior of the EF-Tu and tRNA, please see the tutorial on EF-Tu [3].

The ribosome contains three tRNA binding sites, labeled the A-site (aminoacyl), P-site (peptidyl transferase reaction), and the E-site (exit). After the charged tRNA is released into the A-site by the EF-Tu, the existing nascent protein is transfered from the tRNA in the P-site to the the amino group of the bound amino acid on the tRNA in the A-site, extending the chain by one residue. This reaction is catalyzed by the peptidyl tranferase activity of the ribosome. Elongation factor G faciliates ribosome translocation, causing the A-site tRNA to move to the P-site [5]. The newly vacated A-site will be freed to accept the next tRNA. Because all the tRNAs are base-paired with codons in the mRNA, the movement of the tRNAs also moves the mRNA through the ribosome, exposing the next codon to be matched to the next aminoacylated tRNA. This repeats until a *stop codon* is encountered on the mRNA, which is not complementary to any tRNA, but rather binds the release factors, which trigger the release of the protein and the ultimate dissociation of the ribosomal large and small subunits.

Topics addressed in this tutorial are: 1-2) structural aspects of the LSU and SSU [60 minutes]; 3-4) signatures of ribosome evolution that are used to classify organisms in the Phylogenetic Tree of Life [90 minutes]; and 5) kinetic modeling of ribosome assembly [30 minutes]. Intermediates in the assembly of the SSU are analyzed through MD simulations. This tutorial will rely on the paper **Molecular Signatures of ribosomal evolution** by Roberts *et al* [6], which we have provided for you with this tutorial. This tutorial should take approximately three hours to complete.

Requirements

MultiSeq must be correctly installed and configured before you can begin using it to analyze the ribosome. There are a few prerequisites that must be met before this section can be started:

- VMD 1.9.2 or later must be installed. The latest version of VMD can be obtained from http://www.ks.uiuc.edu/Research/vmd/.
- This tutorial requires approximately 250 MB of free space on your local hard disk.

1 The Ribosomal SSU and associated structures: [30 minutes]

1 Before we open a state file, we need to open the Tk Console by clicking on Extensions → Tk Console. Now navigate to the directory TU-

TORIAL_DIR/1.ribosome_structure. Now in the VMD main window, click on File → Load State. From the 1.ribosome_structure, load the state file `ribosome.vmd`. This will load the *Escherichia coli* 50S and 30S subunits containing both the rRNA and the ribosomal proteins, as well as the bound tRNAs in the A-, P-, and E-sites, and the bound EF-Tu. All of these structures are initially hidden with the exception of the EF-Tu and its bound aminoacyl- tRNA.

2 We will first examine the overall structure of the ribosome and highlight some of the particular features discussed in the introduction. In VMD, zoom in on the yellow highlighted region of the elongation factor. This region is known as the amino acid binding pocket, where the amino acid bound to the tRNA sits as the complex migrates to the ribosome. This part of the tRNA is known as the acceptor stem, and the final three nucleotides, those that sit close to the amino acid binding pocket, are always the same: CCA. The tip of the acceptor stem is called the CCA tail (Figure 1).

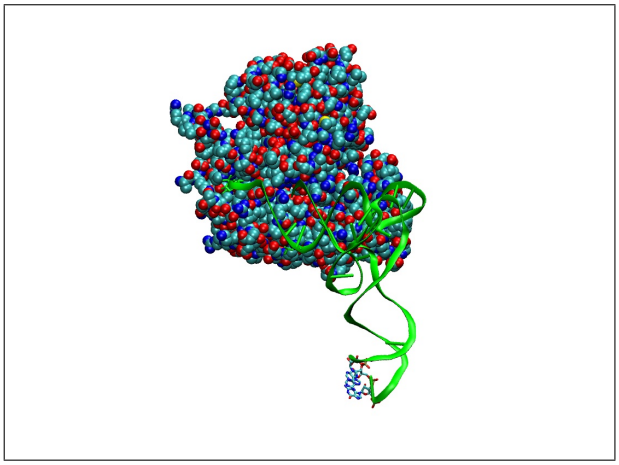


Figure 1: The tRNA bound to the elongation factor

3 Now move to the other side of the tRNA, where three nucleotides have been highlighted in licorice representation. These three nucleotides are the anticodon of the tRNA. Use the Query function of VMD to query the resname of each of these nucleotides (in the order of resid 36, 35, then 34). What is the anticodon of this tRNA? Given that the 'alphabet' of RNA is A, C, U, and G, where A base pairs with U and C base pairs with G, predict the codon to which this tRNA is bound. (The 'resname' of each nucleotide may appear as 'Ar, Cr, Ur, or Gr', the 'r' standing for RNA)

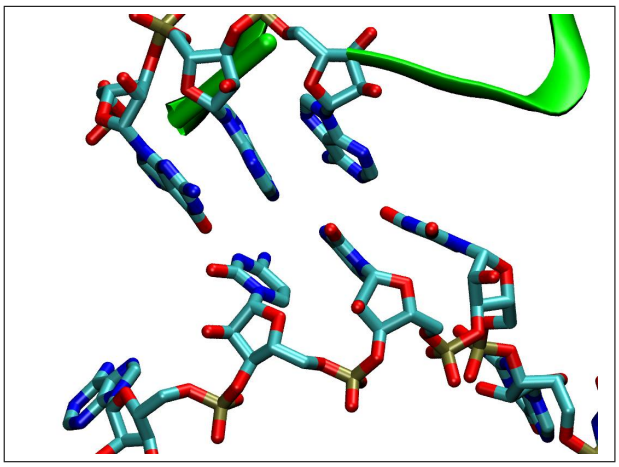


Figure 2: The codon/anticodon

4 Make sure the structure `3FIH.pdb` is displayed in the Selected Molecule drop down box. Now display `chain X` by double-clicking on it to display the mRNA. Note the three nucleotides closest to the anticodon of the tRNA (Figure 2). These three nucleotides represent the codon of the mRNA, which codes for the amino acid currently bound to the tRNA. Use the Query feature of VMD to determine the codon to which the tRNA is base paired. The codon is read from 5' to 3' on the mRNA, which corresponds to the ordering resid 19 20 21. Using the genetic code table shown in Figure 3, determine the amino acid that is about to be added to the growing protein.

5 Now display chain V in the structure by double clicking on the title in the Representations window. This displays the tRNA currently bound in the P-site, or peptidyl transferase site, for which the amino acid has already been bound to the growing protein chain. Create a new representation and type the following as the selected atoms `chain V and resid 34 35 36` to display the anticodon of this tRNA. Color this representation by Name. Change the Drawing Method of this representation to be Licorice. As you did before, determine the type of this tRNA by querying the mRNA nucleotides in the order `resid 16 17 18`.

6 Display chain W now, to display the final tRNA bound to this ribosome. It is clear this tRNA is not as associated with a corresponding codon on the mRNA.

7 Hide chain X. Now, in the VMD representations window, choose `2HGR_SSU_Thermophilus.pdb` in the Selected Molecule drop down box. This is a different structure of the ribosomal 16S, one which contains

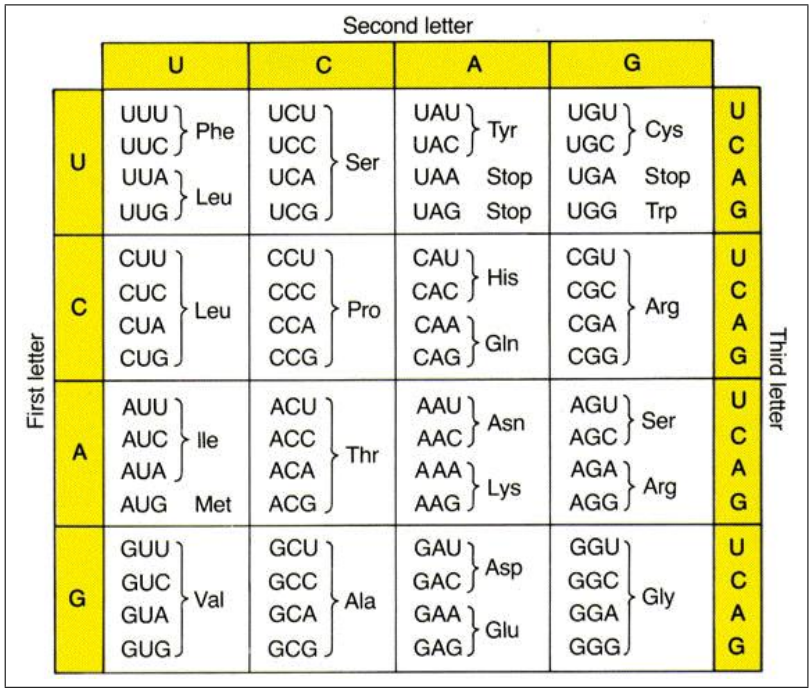


Figure 3: The genetic code

a longer mRNA molecule. We will use this structure to explore how the mRNA is bound by the ribosome. Display the representation chain A and resid 1535 to 1541 and the representation chain 1.

8 You can see that this mRNA is in mostly the same position as the original mRNA. However, this sequence is longer and contains a particular part of the mRNA known as the Shine-Dalgarno sequence. This portion of the mRNA is what is recognized by the ribosome when the mRNA binds and translation begins. The consensus sequence of the Shine-Dalgarno sequence is AGGAGG, though this differs slightly between organisms. The orange residues base pairing with the Shine-Dalgarno sequence are part of the 16S subunit. This complementary sequence is known, appropriately enough, as the anti-Shine-Dalgarno sequence. When the Shine-Dalgarno sequence and anti-Shine-Dalgarno sequence bind, the initiator tRNA:(N-formylmethionine) is recruited and translation begins.

N-formylmethionine. Translation of a protein always begins with the *start codon* AUG. In the genetic code from Figure 3, AUG translated into methionine. However, this methionine derivative (fMet) with a formyl group attached is used instead for the first residue of a protein.

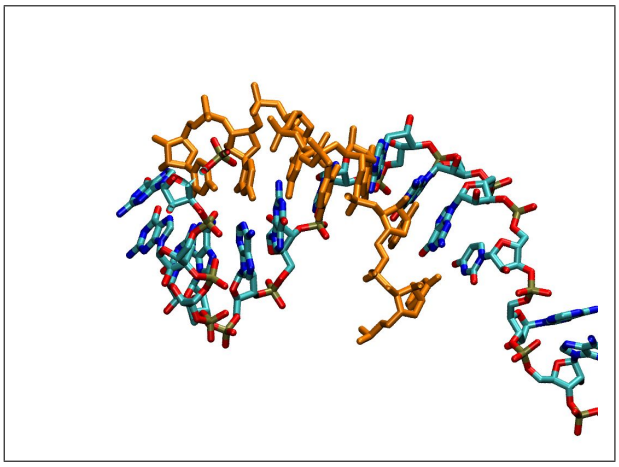


Figure 4: The Shine-Dalgarno sequence on the mRNA and the anti-Shine-Dalgarno sequence on the 16S rRNA

9 Now we can hide the structure `2HGR_SSU_Thermophilus.pdb`. Do this by double clicking on the D next to the structure title in the VMD main window.

10 Now we will display the rRNA portion of the small subunit of the ribosome. In the Selected Molecule drop down box, select `3FIH.pdb`. Display chain

A. The small subunit of the bacterial ribosome contains 1540 nucleotides. Also display the representation chain B C D E F G H I J K L M N O P Q R S T U in VMD, to display the small subunit proteins. Together, this rRNA and 21 proteins comprise the 30S subunit of the bacterial ribosome. Also display chain X to see how the mRNA fits into the small subunit.

11 The ribosomal proteins maintain the stability of the structure, particularly with regards to ribosomal assembly, and several play an important role in the function of the ribosome. Protein S4 helps to maintain translational accuracy. mRNA has secondary structure, but must be a linear chain in order to pass through the ribosome during translation. Protein S4 may assist the mRNA in denaturing its secondary structure. In the VMD representation window, create a new representation. Set the Selected Atoms to be chain G and the Drawing Method to be VDW. Examine the location of S4 with respect to the mRNA.

2 The Ribosome large subunit (LSU) and associated structures: [30 minutes]

1 As discussed in the introduction, the nascent protein is bound to the tRNA present in the P-site, or peptidyl transferase site, of the ribosome. This tRNA is currently displayed in orange in the VMD window. Compare the structure of this tRNA with the structure of the tRNA in the A-site. We have already discussed where the amino acid binding pocket exists on the elongation factor Tu. Based on the structure of the tRNA bound in the A-site, find the location where the amino acid should sit on the tRNA in the P-site (the amino acids do not exist in this structure).

2 In the VMD representations window, choose the file `3FIH.pdb` from the Selected Molecule drop down box. Create a new representation in VMD using selected atoms chain V and resid 74 75 76. Use the drawing method Licorice and coloring method Name. Now the three final residues on the tRNA will be highlighted. This is the CCA tail for the tRNA bound in the P-site. Although the crystal structure from which these coordinates are derived did not include the amino acids bound to the tRNAs, this is where it should be bound.

3 Now hide every representation currently displayed, except for any representation containing chain V. Display the molecule `3FIK.pdb` in the Selected Molecule drop down box and double-click on the representation all. Now the 23S rRNA subunit of the 50S ribosomal subunit is displayed. Notice how the tRNA in the P-site of the ribosome reaches up into the center of the 23S subunit. As we discussed in the introduction, the nascent protein should be currently bound to the tRNA in the P-site, or peptidyl transferase site. Rotate the display. Can you find the 'channel' through the large subunit where the nascent protein should exit?

4 Once you think you have found the nascent protein exit channel, set molecules `3FIK.pdb` as Top in the main VMD window by double clicking on the T column next to the structure name. Choose Tk Console in the VMD Extensions menu. Navigate using the cd command to the 1.ribosome_structure directory and type `source nascent_chain.tcl`. This will draw a sample nascent chain in the exit channel, allowing you to better visualize how the chain will exit the ribosome as it is being synthesized.

5 Ribosomal protein L11 changes conformation to allow the elongation factor to bind, and thus plays an important role in translation. In the VMD representation window, create a new representation. Set the Selected Atoms to be chain I and the Drawing Method to be VDW. Change the molecule in the VMD representations menu Selected Molecule drop down box to be `3FIH.pdb`. Display representations chain A and chain Z to display both the small subunit and the EF-Tu. Examine the location of L11 with respect to the EF-Tu.

2.1 The peptidyl-transferase center

There is evidence that a duplication of a more fundamental RNA structure resulted in the formation of the peptidyl transferase center, where the aminoacyltransferase reaction to extend the nascent protein actually takes place [7, 8]. The PTC in its present form comprises two parts with very nearly identical secondary and tertiary structures. These two parts are the binding sites for the CCA-3' termini of the tRNA for the P- and A-sites. A plausible scenario for the evolution of the ribosome would be that one of these two CCA binding sites duplicated, resulting in the ability to bind two proto-tRNAs inclose proximity allowing a transpeptidation reaction to occur. This complex would likely have been able to synthesize random oligopeptide sequences.

1 Hide chain A. Change the molecule in the VMD representations menu Selected Molecule drop down box to be `3FIK.pdb` Now hide the representation of the entire 23S rRNA (by double-clicking on the all representation) and display the hidden representation with Selected Atoms as nucleic and chain B and (resid 2058 to 2092) .... You can hide the structure 3FIH.pdb in the main VMD window to This will display the portion of the ribosome called the peptidyl transferase center, or PTC. Change the representation of the PTC to VDW, and examine the site where the PTC surrounds the nascent chain and aminoacylated tRNA.

2 Change the representation of the PTC back to NewCartoon. Can you see the symmetry suggested in the text above?

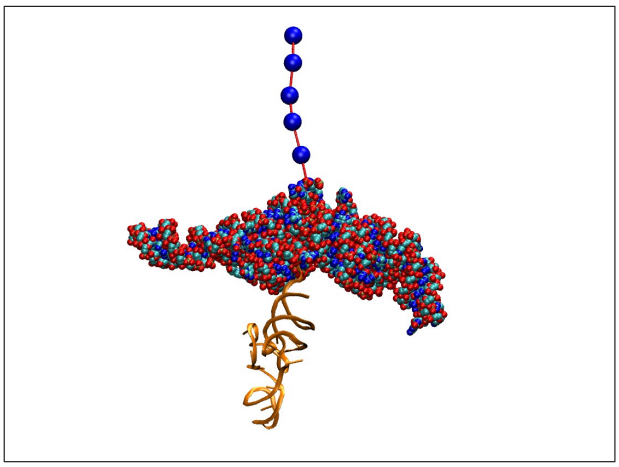


Figure 5: The PTC and nascent chain

3 Ribosome Origins: [30 minutes]

3.1 Hypothesis on the evolution of the ribosome

The ribosome is largely comprised of RNA with proteins decorating the periphery of the structure. All organisms have a similar core structure to their ribosomes, which strongly suggests that the ribosome existed prior to the differentiation of the three domains of life. Analysis of the rRNA tertiary structure has been used to gain insight into how the ribosome evolved before the divergence of the primary organismal lineages [9].

In performing this analysis on the 23S subunit, Bokov and Steinberg observed that a particular interaction motif occurred between domain V and its RNA contacts. The observed motif, an A-minor interaction, involves the interaction of an unpaired adenosine-rich stack with a double helix. In nearly every case of an A-minor interaction with domain V observed, the double helix part of the motif was on domain V. Since an RNA double helix is stable and unpaired stacks of adenosine are not, it was suggested that domain V was an early segment of the 23S and the segments associating with it through A-minor interactions were more recent additions. This suggestion is supported by the fact that domain V contains the pepidyl-transferase center (PTC) which performs the primary function of the ribosome, the addition of amino acids to the nascent polypeptide chain.

Bokov and Steinberg used the A-minor motif to attempt to determine the evolutionary order of assembly of the various elements of the ribosome. To define structurally independent elements, Bokov and Steinberg required that each element have 3' and 5' ends in close proximity (suggesting an insertion that would not otherwise perturb the structure of the remaining portion of the ribosome), that both sides of any double helix be contained in an element, and that

if the element formed an A-minor interaction with the rest of the ribosome that the entire adenosine stack portion of the A-minor be contained in the element and not the double helix portion. Using this method, 19 elements were identified, the deletion of which did not disturb the integrity of the remaining rRNA. Excluding these "layer 1" elements an additional 11 elements were identified which only supported the integrity of the level 1 elements, and could be considered as "layer 2". This process was performed a total of 12 times revealing 59 elements. This method revealed a hierarchy of dependencies, with layer 1 being most likely the latest additions to the 23S and each subsequent layer consisting of progressively more ancient additions to the 23S. The 12 layers constituted 93% of the 23S, with the remaining 7% consisting of a portion surrounding and including the PTC.

1 If the Tk Console is no longer open, open it again by clicking on Extensions → Tk Console. Now navigate to the directory TUTORIAL_DIR/2.ribosome_evolution. Now in the VMD main window, click on File → Load State. From the 2.ribosome_evolution, load the state file `superimposed.vmd`.

2 Close all files in VMD. Load the VMD state file `superimposed.vmd` from the directory 2.ribosomal_evolution on the CD. This is the structure of the large subunit (50S) from *Thermus thermophius*, a 'heat-loving' bacterium originally isolated from a thermal vent. The silver structures consist of the rRNA, while the orange structures are the 50S proteins. Using the VMD representations menu, hide the proteins in this structure by double clicking on the protein representation.

3 In the VMD representations window, select `1S72_LSU_Marismortui.pdb` from the Selected Molecule drop down box. Display the hidden 23S rRNA by double clicking on the nucleic representation. This is the structure of the 50S subunit from *Haloarcula marismortui*, an archaeon found in the Dead Sea. Note the remarkable similarity of the two structures.

4 Based on these two structures, would you predict that the majority of the evolution of the ribosome occurred before or after the divergence of the bacterial and archaeal domains of life?

4 Ribosomal signatures: [60 minutes]

The term ribosomal signatures was coined by Carl Woese and used by him as one form of evidence to define and distinguish the three domains of life: Bacteria, Archaea and Eucarya. The universal phylogenetic tree (UPT, Figure: 7) constructed from the 16S ribosomal RNA (rRNA) shows a so-called canonical pattern, in which all taxa group into three distinct clusters with the eucaryotic and archaeal subbranches closer to each other than to bacteria. This further confirms the signal we could see simply from the signatures. 20 years later, with

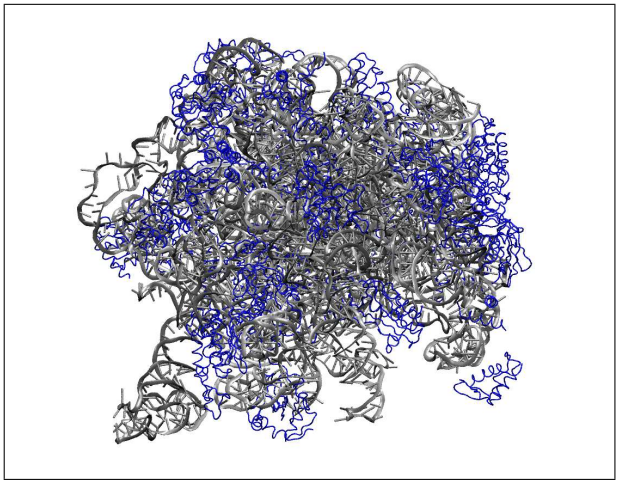


Figure 6: 50S subunit from *H. marismortui*

the help of the rapid growth of the massive genomic and structural data, we are able to extend the signature notion and use the information it conveys to study the evolution of cells and origin of life.

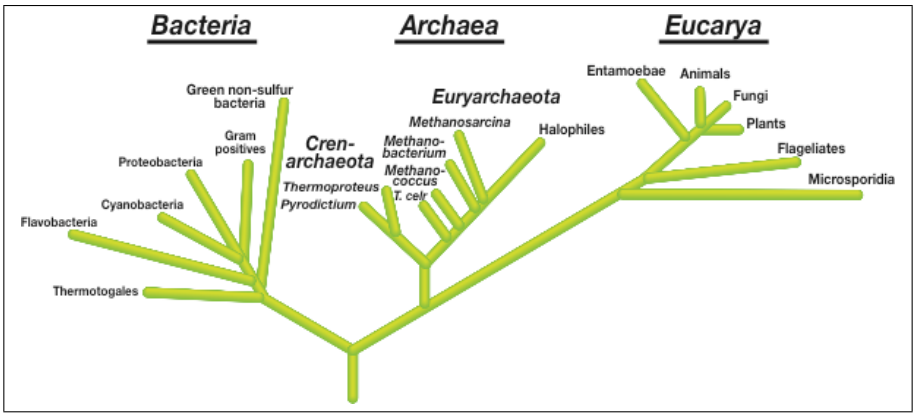


Figure 7: The tree of life

4.1 Definition and classification of the ribosomal signatures

Ribosomal signatures are regions on the ribosome that are constant and unique to a particular domain of life. The two general kinds of signatures are defined based on characteristics of the rRNA. Sequence signatures are positions in

rRNA's primary structure whose compositions remain constant in one domain of life but occur rarely in the other domains. Structural signatures are regions in its secondary and/or tertiary structure that have a unique configuration in a given domain, and they could be further classified into three subtypes: i) insertions or deletions (indels) that are characteristically present in one domain of life but absent in another, ii) regions of the rRNA in which the secondary (and therefore tertiary) structure differs between two domains, and iii) regions that are similar in secondary structure but differ in their tertiary conformation.

1 Delete all files out of the main VMD window before you move on to the next step.

2 We have prepared for you two sets of multiple sequence alignments that were created with MultiSeq. The first set of multiple sequence alignments is for domain V of the 23S rRNA for bacteria and archaea. We will analyze these sequences using MultiSeq and reproduce the results of the paper by Roberts *et* al [6]. Open MultiSeq in VMD by clicking on Extensions → Analysis → MultiSeq

Ribosomal Domains. The 23S rRNA, or large subunit of the bacterial ribosome, is formed from six domains (See Figure 8 for an illustration of the domains of the 23S in a secondary structure representation). Domain V contains the peptidyl transferase center, the enzyme 'active site' of the ribosome, where the aminoacyltransferase reaction takes place to elongate the nascent protein chain.

3 Now we can identify the locations of the sequence signatures on a part of the 23S domain of the ribosome. To do this, we need to load an alignment of several 23S rRNA sequences across the bacteria and archaea, and use this alignment to calculate the sequence signatures. These signatures can then be mapped onto the structure of 23S, so we can understand the types of places sequence signatures are found. Start a new session of MultiSeq by clicking on File → New Session.

4 We are going to load a saved state of MultiSeq. Go to File → Load Session and navigate to the directory 3.ribosomal_signatures. Highlight the file signatures.multiseq and click Open. This will load the MultiSeq session file. This contains an alignment of several archaeal and bacterial 23S sequences for domain V of the 23S. The 23S is made up of six domains, and domain V consists of the peptidyl transferase center and surrounding regions. See Figure 8.

5 Click View → Coloring → Signatures. In the resulting dialog box, choose Archaea and Bacteria. Click OK. (It may take a minute for the signature calculation to complete).

.

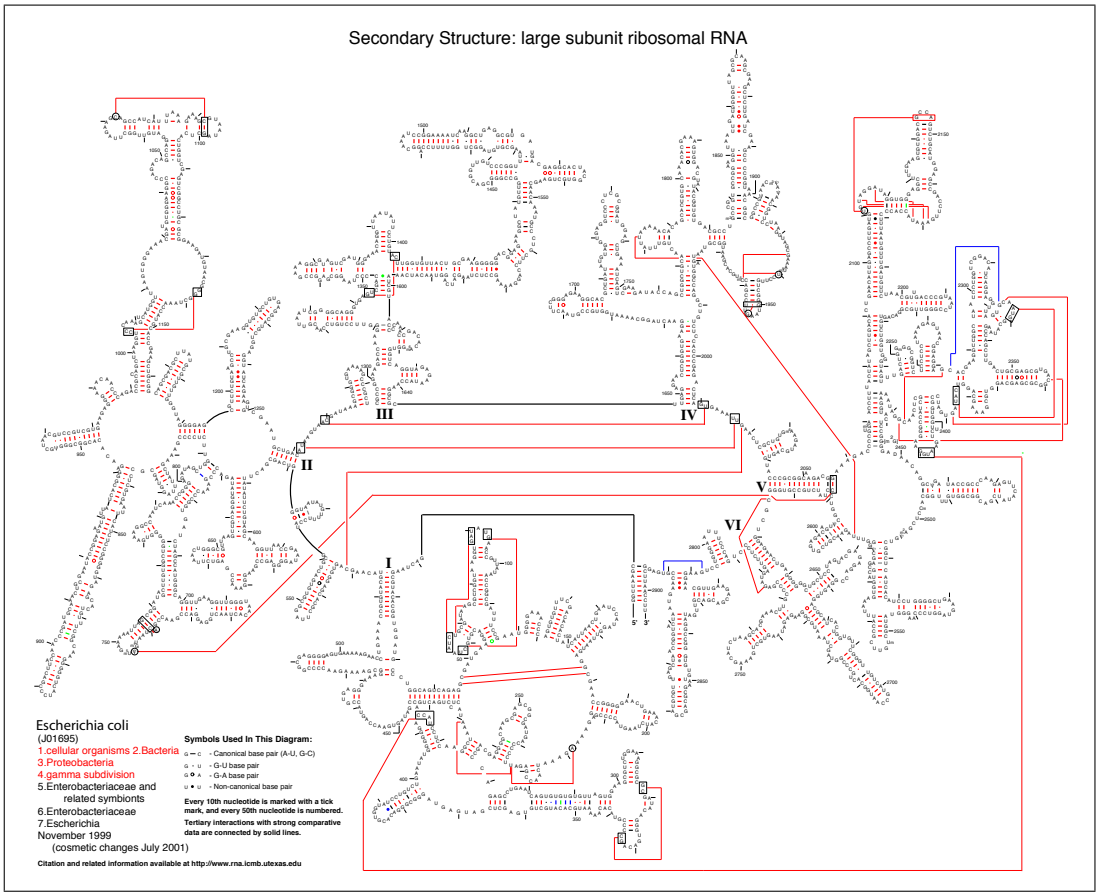


Figure 8: Secondary structures of the 23S, divided into domains

6 Once the calculation completes, scroll to the right in the MultiSeq window until you see the alignments. Note the colored columns. The light blue columns represent nucleotides that are conserved across both bacteria and archaea. The dark blue columns represent nucleotides that are conserved within each domain of life, but are *not* conserved across both domains of life. These are the sequence signatures. We are going to examine one of them on the structure of the ribosome 23S.

7 Take a look at the VMD display. In the Representations menu, make sure `2HGQ_LSU_Thermophilus.pdb` is showing. Now change the representation of chain A to be NewCartoon. Finally, click on the Trajectory tab, and with chain A as the selected atoms, click on the Set and Autoscale buttons.

8 In MultiSeq, click on View → Highlight Color → Yellow. Then click on View → Highlight Style → VDW. This will highlight the nucleotides you select in MultiSeq as yellow, space-filling VDW spheres in the VMD display.

9 Next, we will load the nascent chain into this structure to easily visualize where the protein exits. In VMD, click on Extensions → Tk Console. Using the cd command, navigate to the directory TUTORIAL_DIR/3.ribosomal_signatures. Now type `source nascent_chain.tcl`. This will load a simple representation of a protein exiting the ribosome.

10 Now scroll across the MultiSeq window to column 2061. This column should be highlighted dark blue as a sequence signature, and should be adenosine (A) for the bacteria and guanosine (G) for the archaea. At the very bottom of the MultiSeq window, in the row with the title 2HGQ_LSU_Thermophilus, click on column 2061 to highlight it in yellow, as in Figure 9.

11 Now check the VMD main window. Note that the highlighted nucleotide has appeared as a yellow VDW representation right next to the protein sitting in the ribosomal exit channel. This sequence signature interacts directly with the exiting protein! Now load the structure `antibiotics.pdb` from the directory 3.ribosomal_signatures. In VMD, open the representations menu. Make sure `antibiotics.pdb` is showing in the Selected Molecule drop down box, and make the Drawing Method VDW. Color this antibiotic Orange.

12 Now take a look at the VMD display. Note that the antibiotic is binding to the exact sequence signature in the exit channel that we highlighted earlier. This antibiotic is erythromycin, and it binds in the exit channel of bacterial ribosomes, blocking protein synthesis. Since it binds to a sequence signature, this nucleotide differs between bacteria and archaea/eukarya. As a result, this antibiotic will not bind in ribosomes of other domains of life, and only the bacteria are killed by erythromycin.

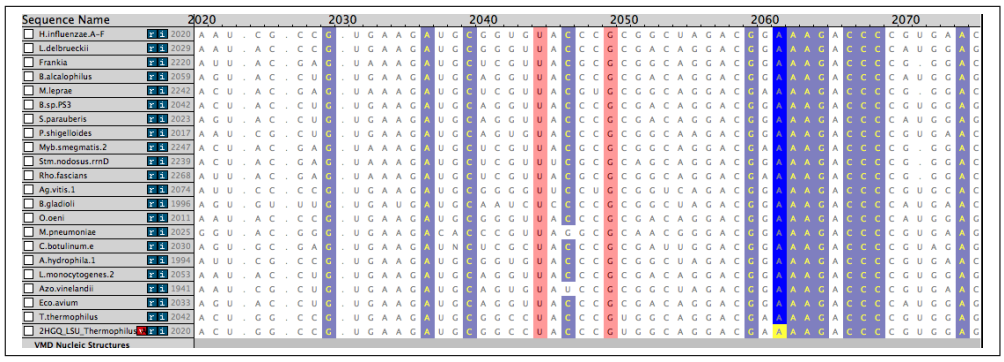


Figure 9: Highlight the sequence signature in MultiSeq

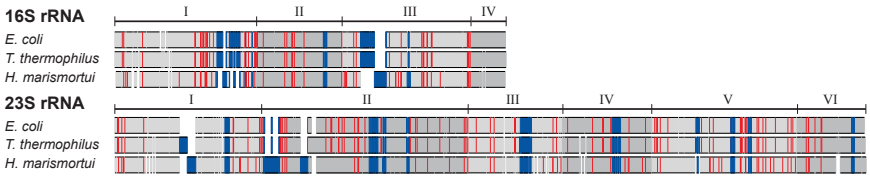


Figure 10: Distribution of the presence of bacterial and archaeal signatures in 90,000 environmental 16S rRNA sequences [6]

4.2 Contribution of ribosomal signatures to phylogenetic separation

Phylogenetic separation between bacteria and archaea can be measured by the distance between roots of the bacteria and archaea subbranches on a canonical tree. As shown in Figure: 10, sequence signatures are distributed throughout both the 16S and 23S rRNA, and they are estimated to constitute about 5 % of the nucleotides in each molecule. In order to quantify the contributions signatures make to the phylogenetic separation between bacteria and archaea, we constructed trees of the 16S rRNA as well as the 23S rRNA with and without those sequence signature positions. The result shows that the $\sim$ 5 % sequence signatures are responsible for 42 % of the distance between bacteria and archaea for the 16S rRNA and 28 % for the 23S rRNA.

1 To begin analysis of the signatures in the 16S rRNA in bacteria and archaea, we will first load a pre-aligned set of sequences into MultiSeq. Please start a new session of MultiSeq by clicking on the File $\rightarrow$ New Session. Click Yes on the prompt to clear the current session and begin the new session. Now click on File $\rightarrow$ Import Data. Import the file `16S.AB.full-aligned.fasta` from the 3.ribosomal_signatures directory.

2 Here we will load a premade phylogenetic tree calculated using RAxML

and examine the contribution of the sequence signatures to the difference between the two groups. In MultiSeq, click on Tools → Phylogenetic Tree. Make sure All Sequences is selected, and check the From File checkbox. Click the Browse button. Navigate to the directory 3.ribosomal_signatures, and highlight `16S.AB.full-aligned.tre`. Click Open. Now click OK to create the tree.

3 To more easily visualize the separation between the domains of life, we will color the background of the tree. Click on View → Background Color → Taxonomy → Domain of Life. Now click View → Leaf Text → Taxonomy → Domain of Life. You can see that the archaea seem to be splitting the bacterial group. We can correct this by rerooting the tree. Click anywhere on the long line connecting the bacterial group to mark it. Now select View → Reroot tree at selected point to reroot the tree.

4 Now we will measure the distance between the bacterial and archaeal groups. Click on the root of the bacterial subtree to highlight it with a yellow point. Holding down the CTRL key (Windows) or the COMMAND key (Mac), click on the root of the archaeal subtree. Now the distance between the two subtrees is printed in the bottom left hand corner of the Tree Viewer window. Note the distance, which should be a fractional number.

5 Now we will generate a new tree, one generated in the absence of the sequence signatures you generated above. Move back to the MultiSeq window. Again click on Tools → Phylogenetic Tree and this time load the file `16S.AB.noseqsigs-aligned.tre`. Reroot this tree as you did before and calculate the distance between the archaeal and bacterial subtrees. Note the distance between them.

6 Calculate the percentage change in distance between the bacterial and archaeal groups with and without the sequence signatures. Does this number agree with the calculated value of 42% given above?

In addition to the sequence alignment, structure-based techniques have been used to identify structural signatures, which are structural features present in one domain of life that are not present in another. Six structural signatures were found for the 16S rRNA and 14 were found for the 23S rRNA. The same tree construction procedure was applied for the structural signatures and the decrease in the separation between the bacterial and archaeal subbranches while structural signatures were cut out was 8 % for the 16S and 16 % for the 23S, significantly less than that for the sequence signatures. However, a phylogenetic approach that explicitly includes the structural modeling of indels might be a more reliable alternative for evaluating the phylogenetic contribution of the structural signatures. Though this could only be done with the four existing structures we have (three Bacteria *Thermus thermophilus*, *Escherichia coli* and

Figure 11: Rerooted archaea and bacteria phylogenetic tree

Deinococcus radiodurans and one Archaea *Haloarcula marismortui*), it gave an estimation of 50 % reduction in distance upon removing the structural signatures — a value comparable to the contributions of the sequence signatures.

4.3 Functional roles of signatures in ribosomal assembly

Ribosomal signatures are not only molecular fossils that enable us to define taxonomy and study the origin of life, but are also important for us to study the dynamics of ribosomal assembly in the modern cells. In two consecutive papers [6, 10], we showed that h16 on the 16S rRNA and the N-terminus of r-protein S4 are the largest bacterial structure signatures that interact with each other using sequence and phylogeny analysis as well as structural modeling. In subsequent studies [11, 12, 13, 14], we combined molecular dynamics simulations, single-molecule Förster resonance energy transfer (smFRET) and biochemistry experiments to explore the dynamic landscape of these signature regions. The results (Figure: 12) show that the rRNA signature adopts multiple metastable states prior to S4 binding, and that interactions between rRNA signature h16 and the intrinsically disordered N-terminal signature of S4 are designed to speed up molecular recognition during initiation of ribosomal assembly. The two recently solved eukaryotic ribosome structures [15, 16] give us new data on these structure signatures. In this section, we will observe and compare these signatures on the *E. coli* and *Saccharomyces cerevisiae* (yeast) ribosomal SSU structures.

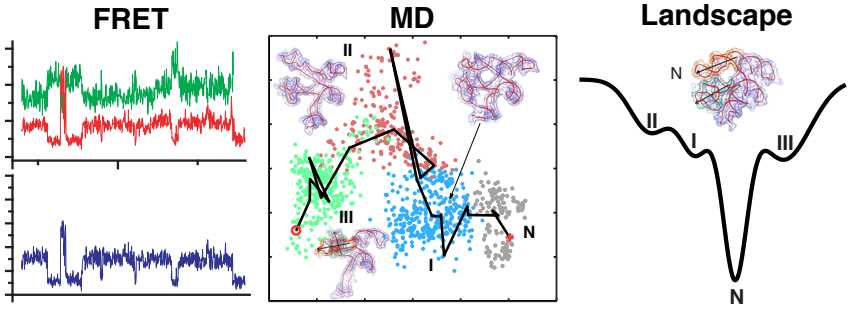


Figure 12: All-atom MD simulations and smFRET experiments identify multiple metastable conformations of the native five-way junction in the 16S *E. coli* without presence of S4. These states are important for the correct binding of S4 at the onset of ribosome biogenesis.

1 Clear all MultiSeq sessions and delete all molecules in the VMD main window before you move on to the next step.

2 Load the *E. coli* ribosomal SSU structure directly from online. Click File → New Molecule... in the VMD main window, then type 2i2p in the Filename

box and click Load button. Load the *S. cerevisiae* SSU structure (PDB code: 3o2z) in the same way. This will take a while as the each file contains over 50,000 atoms. After you are done, you should see the structure in the OpenGL display window and sequences of each chain loaded into the MultiSeq window.

3 Notice that sequences titled "302z_a" to "3o2z_h" under the VMD Nucleic Structure group contain all question marks(?). This is because these chains are unassigned secondary structure and have unnatural residues. Click on any of the question marks, you should see the three letter code (UNK) for its residue on the status bar at the bottom of the MultiSeq window. As they are not relevant here, you may simply delete them all. There are also question marks towards the end of sequences "3o2z_1" and "2i2p_A", which are magnesium ions and ligand osmium (III) hexammine. You may also delete these by clicking Edit → Enable Editing → Full, then highlighting all the question marks you want to delete, and hitting DELETE. Don't forget to change the editing ability to off to avoid any accidental changes to the sequences.

4 In order to compare the two structures, we will perform structure alignment between the two RNA molecules. Mark the 16S rRNA sequence of *E. coli* "2i2p_A" and 18S rRNA sequence of *S. cerevisiae* "3o2z_1". Click Tools → Stamp Structural Alignment. In the resulting window, choose Marked Structures and keep all the default parameters. Click OK.

5 After the alignment is done, you may color it by conservation of the structures. Click View → Coloring → Apply to Marked. Click View → Coloring again and this time choose Qres. You will see that the marked sequences as well as the structures in the OpenGL Display window are colored accordingly, with blue indicating conserved structures and red representing dissimilar structures (see Figure 13).

It is clear from the structural alignment that, the core of SSU is conserved across *Bacteria* and *Eukarya*, especially the groove where mRNA binds and the interface with LSU. Eukaryotic extension elements in the 18S rRNA are mostly located on the periphery of the SSU. h16 (*E. coli* structure 2i2p: chain A and resid 406 to 436; *S. cerevisiae* structure 3o2z: chain 1 and resid 477 to 512), which is located at the entrance of the mRNA on the SSU, is different in the two structures in length, secondary structure, and tertiary contacts.

6 Create representations for the bacterial r-protein S4 (2i2p: chain D) and eukaryotic r-protein S9A (3o2z: chain E). You should be able to see the two proteins occupying the same binding site created by h16, h17, and h18 on the rRNA.

Though the alignment was not done on S4/S9A proteins directly, the two proteins are reasonably aligned as a result of the structurally aligned rRNA. This

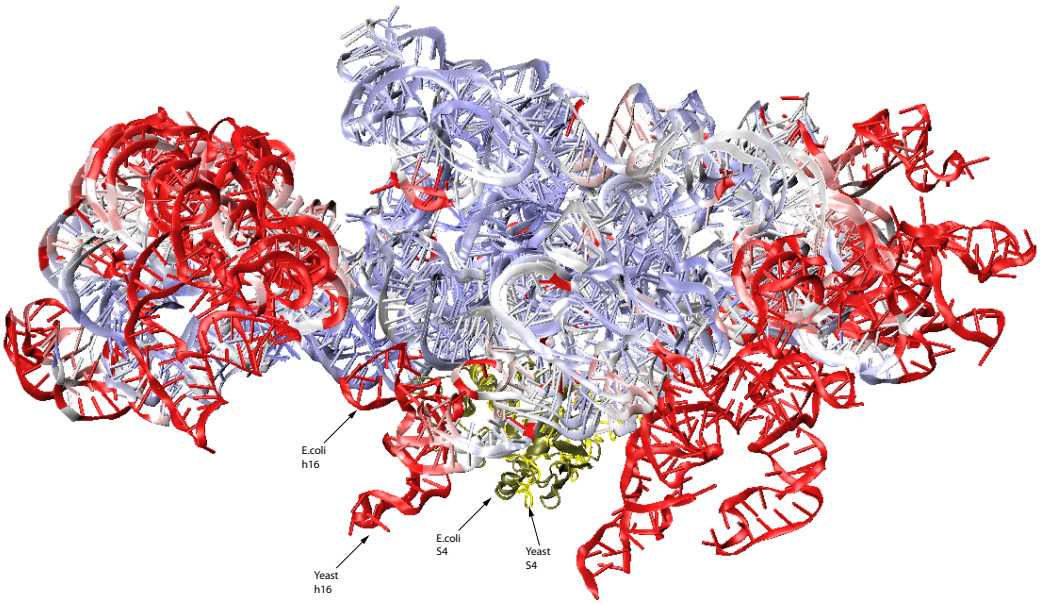


Figure 13: Structurally aligned SSU structure of *E.coli* and *S. cerevisiae*.

shows the conservation of RNA:protein interactions despite of the differences in rRNA and protein sequence or structure. Take a closer look at the proteins; the C-terminus domain is conserved except for a ∼40-residue insertion. Unfortunately, the N-terminus of the yeast S9A is missing in this structure. To see the differences, you might turn to the newest *Tetrahymena thermophila* strucuture (PDB code: 2xzm) and do the same exercise as above. This structure is even more complicated, with a few more ribosomal proteins binding to the 18S rRNA. The naming of r-proteins may be different in different organisms. Please refer to http://www.pdb.org/ for chain information for the structure.

5 Kinetic Model of Ribosome assembly: [30 minutes]

In bacteria, the biogenesis of a ribosome [19] requires a number of critical steps: (1) the transcription of ribosomal RNA and r-protein mRNA from the multiple ribosomal operons; (2) the synthesis of the r-proteins, which is regulated on the translational level based on organization of the r-protein operons in the genome; (3) post-transcriptional processing and modification of both the ribosomal RNA (rRNA) and r-proteins; and (4) the highly coordinated assembly of r-proteins and rRNA towards the mature ribosomal subunits. These events occur in parallel throughout the cell cycle.

Ribosomal assembly involves the cooperation of many of molecular compo-

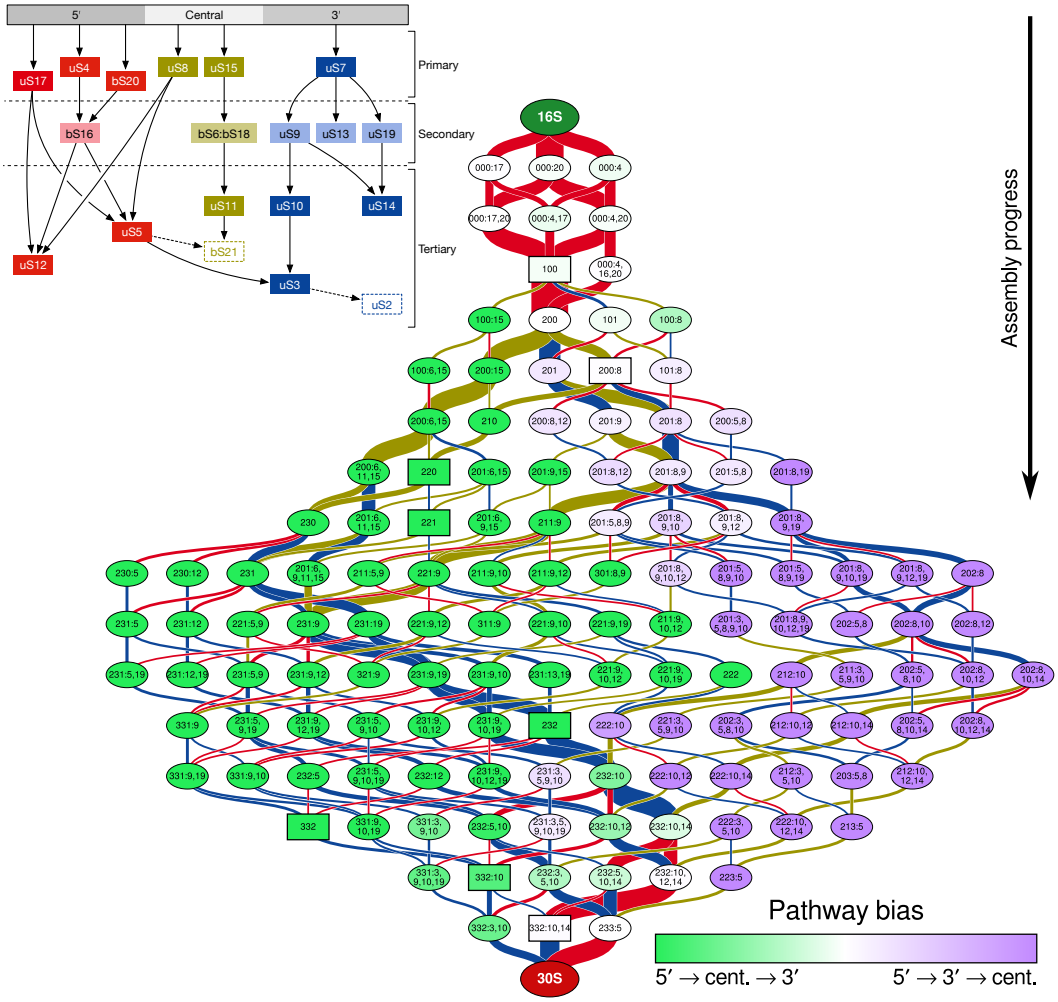


Figure 14: *In vitro* kinetic model for 30S assembly at 15 °C [17]. Each node represents an assembly intermediate, labeled according to which proteins are bound. A three digit number describes the set of r-proteins bound to each domain (5′-, central-, and 3′- respectively). For example, state 201 means that all of the primary and secondary proteins in the 5′ and all of the primary proteins in the 3′ are bound. All remaining r-proteins are listed after the three digit number. The edges connecting the intermediates represent the r-protein binding reactions. The width represents the total amount of intermediate converted by that reaction, and the color indicates the binding domain of that protein (5′-red, central-yellow, and 5′-blue.) The color of each node indicates its bias toward its use of the two assembly pathways. Green indicates that clustering of protein binding order trajectories have indicated that this species is more likely to take part in the 5′ → central → 3′ pathway. Predicted assembly intermediates from pulse/chase qMS and cryoEM [18] are represented using rectangles.

nents. The 30S small subunit (SSU), tasked with the initial binding of mRNA and its decoding, is composed of the 16S rRNA and 21 r-proteins, where as the 50S large subunit (LSU), tasked with the assembly of protein through peptide bond formation, is composed of the 5S and 23S rRNA and 33 ribosomal proteins. These 54 proteins diffuse through the cell to find the rRNA and intermediates through parallel pathways. In addition, approximately 20 assembly cofactors are engaged to facilitate the process at various assembly stages. The rich complexity of 30S assembly process attracted Nomura, *et. al* [20], who first observed how the stability of the binding of certain r-protein depend on the prior binding of other r-protein, constructing a hierarchical dependency map of the assembly process at temperatures optimal for the growth of *Escherichia coli* (37 °C) from equilibrium reconstitution experiments. The Nomura map is shown in (Figure 14). Within each domain, the primary binding r-proteins can bind directly to the rRNA while the secondary and tertiary r-proteins depend on the primary r-proteins. Progress in biophysical approaches has increased our understanding of *in vitro* ribosomal self-assembly through the protein assisted dynamics of RNA folding [21, 22, 14], and the kinetic cooperativity of protein binding [23, 18, 24, 25, 26]. All of the studies suggest that assembly of the *E. coli* 30S subunit proceeds through multiple parallel pathways, starting with the proteins associated with the 5′ domain of the 16S rRNA binding first, followed by the central domain proteins, and finally the 3′ domain proteins.

Using the Nomura map of thermodynamic binding dependencies and protein incorporation kinetic data, we have constructed comprehensive *in vitro* kinetic models (Figure: 14) that capture the topology of the protein RNA interaction network and reproduce the protein binding kinetics of assembly, starting from the bare 16S rRNA or from pre-prepared assembly intermediates [23, 18]. This model reproduces the binding kinetics for all of the r-proteins and is consistent with an assembly mechanism inferred from cryo-electron microscopy (cryoEM) of 30S assembly intermediates [18, 24]. A key prediction from this model is the presence of two distinct assembly pathways that bifurcate from state 200 (Figure: 14).

In this section, we will analyze the results from three different molecular dynamics simulations to probe the structural changes near the state 200 bifurcation point. In each MD simulation, a folded 16S rRNA with different r-proteins bound was allowed to unfold. The three simulations are: state201 (containing proteins **S4**, **S17**, **S20**, **S16**, and **S7**); state201:8,19 (containing proteins S4, S17, S20, S16, S7, **S8** and **S19**); and state201:8,9,19 (containing proteins S4, S17, S20, S16, S8, S7, S19 and **S9**).

1 Open a new VMD session before proceeding to the next step.

2 Load the `state201.psf` and `state201.dcd` from the 4.ribosome_assembly directory. You can easily do this by clicking on the File → New Molecule menu tab on the VMD Main Window. A new window entitled "Molecule File Browser" should appear. Please click on the Browse button (located to the right of the Filename textbox) and navigate to the 4.ribo-

some_assembly directory. Double click on the `state201.psf` file and click the button "Load". Repeat for `state201.dcd`.

3 In the VMD Main Window, a new entry, "state2007.psf", should appear. This entry should contain 37,613 atoms and 34 frames (representing about 80 ns from a MD simulation). Change the representation of the molecule by clicking on the Graphics → Representation menu tab on the VMD Main Window. A new window entitled "Graphical Representations" should appear. In the middle of the "Graphical Representations", a textbox with the labels "Style", "Color", and "Selection" should appear. Initially, in this textbox, there should be an entry labeled "Lines", "Names", and "all". If your entry looks different from this, do not worry as we will be changing the representation. Please click on the above entry. The text should now be highlighted. Under the "Draw Style" menu tab, please change the "Coloring Method" and "Drawing Method" to the following: "SegName" and "NewCartoon" respectively. Click the button "Apply" if the representation on your VMD OpenGL Display Window did not change. Run the simulation forwards and backwards several times.

4 As you run the simulation back and forth, you should notice that part of the system unfolds. We will now quantify the degree to which the system unfolds. To do this, please open the Tcl/Tk console. You can do this by clicking on the Extensions → Tk console menu tab on the VMD Main Window.

5 Now we are going to define atom-selections for different parts of the system. Figure 15a shows the different atom-selections. In the Tk console, please type the following:

```
set atmsel1 [atomselect top "noh and resid 1060 to 1197"]
set atmsel2 [atomselect top "noh and resid 935 to 950 \
            1231 to 1247 to 1290 to 1380"]
set nf [molinfo top get numframes]
set wp [open "dist.state201.dat" "w"]
for {set i 0} {$i <$nf} {incr i} {
    $atmsel1 frame $i
    $atmsel2 frame $i
    set cAtmsel1 [measure center $atmsel1]
    set cAtmsel2 [measure center $atmsel2]
    puts $wp [veclength [vecsub $cAtmsel1 $cAtmsel2]]
}
close $wp
```

This will print out the center of mass separation for different parts of the system to a text file named ‘‘`dist.state201.dat`’’.

6 Repeat for the simulations: `state201:8,19` and `state201:8,9,19`. Make sure that you change the filename ‘‘`dist.state201.dat`’’ lest you want to overwrite your own files.

7 Plot the center of mass separation for `state201`, `state201:8,19`, and `state201:8,9,19`. The plot should be similar to the one shown in Figure 15b. Your results should show that S9 prevents the system from opening up.

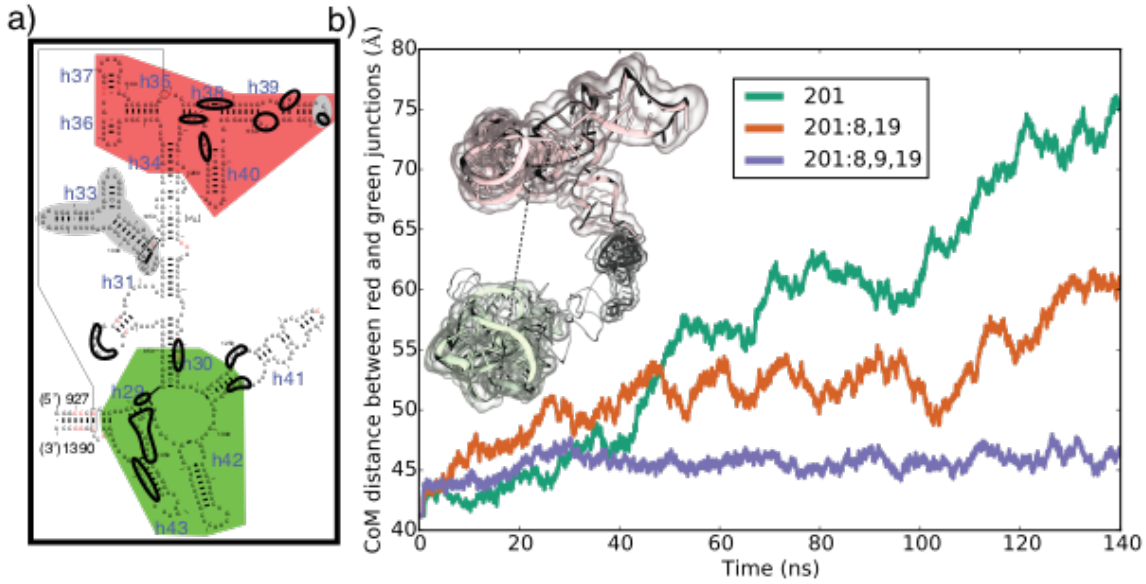


Figure 15: (a) Secondary structure diagram of the 3′ domain with the center of masses defined. Center of masses are computed from the lower four-way junction helices h29, h30, h41-h43 (green) and the upper three-way junction helices h34-h40 (red). These centers are separated by the structural signature—marked in gray circles—h33 and numerous sequence signatures [27]. Nucleic acid bases circled in black make contact to S9 (<5Å) in the folded SSU. (b) Time traces of center of mass distances in the 3′ domain. Traces show how S9 brings together different helices in the 3′ domain.

Acknowledgments

Development of this tutorial was supported by the National Institutes of Health (P41-GM104601-23 - Resource for Macromolecular Modeling and Bioinformatics), National Science Foundation (MCB 12-44570 Evolution of Translation From Molecules to Cells), and the Department of Energy (BER ORNL ABI).

References

[1] L. Li, A. Sethi, and Zan Luthey-Schulten. Evolution of translation: Class-i aminoacyl-trna synthetases. 2012.

[2] P. O'Donoghue, A. Sethi, B. Dhaliwal, A. Sethi, and Zan Luthey-Schulten. Evolution of biomolecular structure: Class-ii aminoacyl-trna synthetases. 2008.

[3] K. Chen, J. Eargle, Z. Ghaemi, J. Lai, and Zan Luthey-Schulten. Evolution of translation ef-tu: trna. 2014.

[4] http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2009/press.html.

[5] Peter B Moore. How should we think about the ribosome? *Annual review of biophysics*, 41:1–19, 2012.

[6] E. Roberts, A. Sethi, Montoya J., Woese C.R., and Luthey-Schulten Z. Molecular signatures of ribosomal evolution. *PNAS*, 105(37):13953–13958, 2008.

[7] Poul Nissen, Jeffrey Hansen, Nenad Ban, Peter B. Moore, and Thomas A. Steitz. The Structural Basis of Ribosome Activity in Peptide Bond Synthesis. *Science*, 289(5481):920–930, 2000.

[8] I. Agmon, A. Bashan, R Zarivach, and A Yonath. Symmetry at the active site of the ribosome: structural and functional implications. *Biol. Chem.*, 386:833–844, 2005.

[9] Konstantin Bokov and Sergey V. Steinberg. A hierarchical model for evolution of 23s ribosomal rna. *Nature*, 457:977–980, 2009.

[10] K. Chen, E. Roberts, and Z. Luthey-Schulten. Horizontal gene transfer of zinc and non-zinc forms of bacterial ribosomal protein S 4. *BMC evolutionary biology*, 9(1):179, 2009.

[11] K. Chen, J. Eargle, K. Sarkar, M. Gruebele, and Z. Luthey-Schulten. Functional Role of Ribosomal Signatures. *Biophysical journal*, 99(12):3930–3940, 2010.

[12] K. Chen, J. Eargle, J. Lai, H. Kim, T. Ha, T. Abeysirigunawardena, M. Mayerle, S. Woodson, and Z. Luthey-Schulten. Assembly of the fiveway junction in the ribosomal small subunit using hybrid MD/Gō simulation. 116:6819–6831, 2012.

[13] Jonathan Lai, Ke Chen, and Zaida Luthey-Schulten. Structural intermediates and folding events in the early assembly of the ribosomal small subunit. 42:1333513345, 2013.

[14] Hajin Kim, Sanjaya Abeysirigunawardena, Ke Chen, Megan Mayerle, Kaushik Ragunathan, Zaida Ann Luthey-Schulten, Taekjip Ha, and Sarah Woodson. Protein-guided RNA dynamics during early ribosome assembly. *Nature*, 506:334–338, Feb. 2014.

[15] A. Ben-Shem, L. Jenner, G. Yusupova, and M. Yusupov. Crystal structure of the eukaryotic ribosome. *Science*, 330(6008):1203, 2010.

[16] J. Rabl, M. Leibundgut, S.F. Ataide, A. Haag, and N. Ban. Crystal Structure of the Eukaryotic 40S Ribosomal Subunit in Complex with Initiation Factor 1. *Science*, 331(6018):730, 2011.

[17] Tyler Earnest, Jonathan Lai, Ke Chen, Mike Hallock, Jamie R. Williamson, and Zan Luthey-Schulten. Towards a whole-cell model of ribosome biogenesis: Kinetic modeling of ssu assembly. 2015 [Submitted].

[18] A.M. Mulder, C. Yoshioka, A.H. Beck, A.E. Bunner, R.A. Milligan, C.S. Potter, B. Carragher, and J.R. Williamson. Visualizing Ribosome Biogenesis: Parallel Assembly Pathways for the 30S Subunit. 330(6004):673–677, October 2010.

[19] Magdalena Kaczanowska and Monica Rydén-Aulin. Ribosome biogenesis and the translation process in *E. coli*. 71(3):477–494, 2007.

[20] W.A. Held, B. Ballou, S. Mizushima, and M. Nomura. Assembly Mapping of 30S Ribosomal Proteins from *Escherichia coli*: Further Studies. 249(10):3103–3111, 1974.

[21] T. Adilakshmi, P. Ramaswamy, and S.A. Woodson. Protein-Independent Folding Pathway of the 16S rRNA 5′ domain. 351(3):508–519, August 2005.

[22] Tadepalli Adilakshmi, Deepti L Bellur, and Sarah a Woodson. Concurrent Nucleation of 16S Folding and Induced Fit in 30S Ribosome Assembly. 455(7217):1268–72, October 2008.

[23] Anne E Bunner, Andrea H Beck, and James R Williamson. Kinetic Cooperativity in *Escherichia coli*30S Ribosomal Subunit Reconstitution Reveals Additional Complexity in the Assembly Landscape. 107(12):5417–22, March 2010.

[24] D. G. Sashital, C. A. Greeman, D. Lyumkis, C. Potter, B. Carragher, and J. R. Williamson. A combined quantitative mass spectrometry and electron microscopy analysis of ribosomal 30s subunit assembly in *E. coli*. page eLife 2014;10.7554/eLife.04491, 2014.

[25] M.W.T. Talkington, G. Siuzdak, and J.R. Williamson. An Assembly Landscape for the 30S Ribosomal Subunit. 438(7068):628–632, 2005.

[26] M.T. Sykes and J.R. Williamson. A Complex Assembly Landscape for the 30S Ribosomal Subunit. 38:197–215, 2009.

[27] E. Roberts, A. Sethi, J. Montoya, C.R. Woese, and Z. Luthey-Schulten. Molecular signatures of ribosomal evolution. 105(37):13953–13958, 2008.