Assembling the Bacterial Large Ribosomal Subunit using Maximum Likelihood Ancestral Sequence Resurrections

by

Joshua G. Stern

Doctoral Dissertation Proposal
School of Biology
Georgia Institute of Technology

Doctoral Committee:

Associate Professor Eric A. Gaucher, Chair
Assistant Professor Brian K. Hammer
Associate Professor Soojin Yi
Assistant Professor Christine M. Dunham (Emory University)
Professor Terry W. Snell

To explore the intersection of phylogenetics and macromolecular engineering, I will build ancient ribosomal particles. I will assemble large subunit-derived particles using individually purified ancestral rRNA and rProtein molecules. Assembly mapping may reveal previously unknown ribosomal functions with implications for polymer, medical, and origin-of-life science. **Characterizing the assembly of ancient ribosome-derived particles is novel, useful, and feasible.**

First, I introduce ancestral resurrection and ribosome engineering. Next, I discuss the motivation and significance of the present work. Finally, I describe my experimental strategy.

1. Introduction

Each cell employs thousands of "protein factories" to translate the DNA/RNA blueprint into protein products that accomplish biochemical tasks. These factories, ribosomes, are macromolecular complexes of diameter ~ 15 nm. In addition to the ribosome's biochemical necessity, ribosomal genes catalog 3.8 billion years of evolutionary history and inform paleogenetic study of extinct nodes in the tree of life (Battistuzzi & Hedges, 2009). Because ribosomal genes encode machinery that has been essential to life since life began, their evolutionary divergence bears witness to the origin of species. **I hypothesize that ancient ribosomal gene products assemble into particles with electrophoretic profiles similar to those of *E. coli* ribosomes.**

In Bacteria and Archaea, each ribosome consists of a large (50*S*) and small (30*S*) subunit **(Fig. 1)**. The 30S helps enforce the genetic code, ensuring that the anticodon of each tRNA complements the codon in the mRNA blueprint. The tRNAs deliver amino acids (protein building blocks) to the ribosome. The 50S catalyzes covalent peptide bond formation between the carboxy- end of the nascent protein and the amino group of the delivered amino acid. In Bacteria, the 30S consists of one ribosomal RNA and 22 ribosomal proteins; the 50S consists of two rRNA molecules and 33 rProteins. Electrostatic interactions between the 58 ribosomal molecules form two complexes, the 50S and 30S, which are stable independently and as a 70S complex. In concert with several elongation factors, the two subunits "ratchet" along the messenger RNA, condensing amino acid building blocks into protein product (Frank & Agrawal, 2000). **I focus on engineering the rProteins and rRNA of the large subunit.**

The presence of ribosomal genes in all living species makes them convenient phylogenetic markers; newly discovered taxa group with previously-known taxa based on rRNA and rProtein sequence similarity. Given a hypothesis, or model, of character replacement rates, bifurcating tree topology, and branch length, phylogenetic algorithms evaluate the likelihood of the data (the sequence alignment). Typical algorithms maximize likelihood using the Monte Carlo method to iteratively propose refinements to an initial tree. Felsenstein (1981) explained how to calculate the likelihood of phylogenetic data given a model.

The resulting phylogenetic trees are the basis for ancestral sequence reconstruction, which infers the most likely nucleic acid or protein sequences at ancestral (internal) tree nodes. Ancestral sequence reconstruction considers **both** the **tree** *and* the **alignment** as "**data**," and alternative ancestral states as alternative models. Using the paradigm of Felsenstein (1981), ancestral reconstruction calculates the probability of alternative models (ancestral character states) given the data (tree and sequence alignment), and the prior probability of the ancestral

state (Yang et al., 1995). For an internal tree node, for each site in the multiple sequence alignment, the ancestral reconstruction is a probability distribution over all possible character states (4 for nucleic acids and 20 for proteins). At an internal node, at each position of the sequence alignment, choosing the ancestral character with the highest probability yields the most probable ancestral sequence.

Ancestral sequences are a cornerstone of evolutionary synthetic biology. Aside from their therapeutic and industrial utility (Gaucher, 2011; Zhou et al., 2012), ancestral phenotypes reveal characteristics of the early Earth's environment. For ancient nodes close to the bacterial cenancestor, Gaucher and coworkers found that ancestors of the translational protein Elongation Factor Thermo-Unstable (EF-Tu) are ***stable*** at thermophilic temperature (Gaucher et al., 2008). This result is consistent with geochemical data indicating that the distant ancestors of extant Bacteria lived in a hot environment. Ancient EFs may also be useful for nuclease-free ribosome-display in-vitro evolution of thermostable enzymes (Zhou et al., 2012). Ancestral enzymes are often promiscuous, and their "hot swappable" character (Cole & Gaucher, 2011) may improve cell-free systems that add to the growing use of synthetic biology.

Using resurrected rProteins and rRNA from an ancient node in the tree of life (Table 1, Figure 2), I will investigate the *in vitro* assembly of the Bacterial ribosome's large subunit.

My approach takes inspiration from the seminal work of Nomura (e.g., Mizushima & Nomura, 1970) and Nierhaus (e.g., Herold & Nierhaus, 1987) whose respective labs demonstrated *in vitro* assembly of the *E. coli* 30S and 50S subunits. Incubating purified small subunit proteins with 16S rRNA leads to intact and active 30S. The same is true for the 50S, although successful assembly of this particle *in vitro* usually requires an intermediate increase in temperature and Mg2++ concentration. The traditional *in vitro* 50S assembly reaction takes 90 minutes, which is much longer than the two minutes[1] required for assembly *in vivo*[2].

Nierhaus omitted different combinations of LSU rProteins and rRNAs from *in-vitro* assembly reactions and analyzed the mixtures using ultracentrifugation followed by SDS-PAGE. Determining the rProtein composition of the "heavy particle" in the sucrose gradient revealed that some rProteins depend on the prior binding of others. *In vitro* formation of the full *E. coli* 50S occurs through a contingent series of binding events. **The Nierhaus experiments demonstrate cooperative assembly of 50S particles from individually purified *E. coli* rProteins and rRNA.**

[1] rRNA transcription in *E. coli* proceeds at 42 nt/sec (Gotta et al., 1991). 3000 nt / 42 nt/sec = 71 seconds, and *in vivo* rProtein binding is co-transcriptional with rRNA. Shajani et al. (2011) note two studies indicating that rProtein binding events may occur in two stages, the longer of which is a refolding event on the timescale of seconds. Thus it seems that *in vivo* assembly occurs in less than two minutes.

[2] The slow and non-physiological character of traditional *in-vitro* assembly probably stems from competition between early and late binding rProteins. *In vivo* rProtein synthesis occurs in concord with rRNA synthesis and is further regulated by the presence of other rProteins. *In vivo* ribosome biogenesis is an intricately regulated process (Nierhaus, 2004).

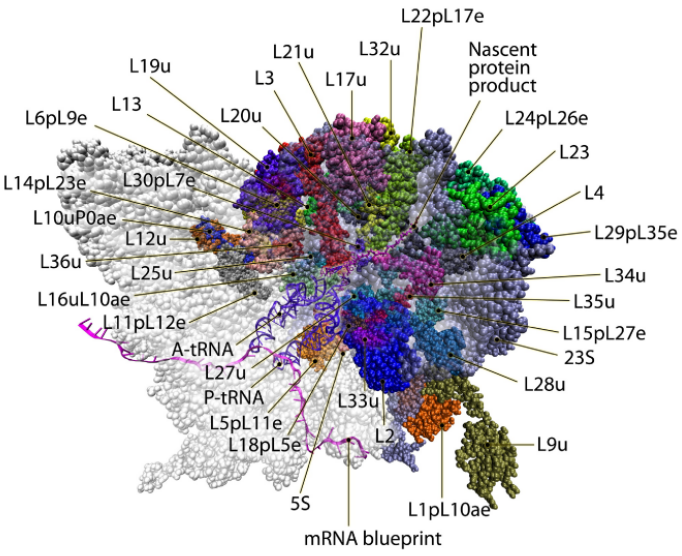

Figure 1. The E. coli 70S ribosome. Portions of the rRNA are hidden to emphasize the rProteins, active site, and nascent peptide. The nascent product is ~ 5.5 nm long. We show the small subunit in white. This image combines renderings of aligned models from several labs (Llano-Sotelo et al., 2010; Agirrezabala et al., 2012; Frauenfeld et al., 2011; Bingel-Erlenmeyer et al., 2008; Zhou et al., 2012; Seidelt et al., 2009; Schmeing et al., 2002) (retrieved from DARC).

Ancestor	Best Hit	% identity
L1pL10ae	L1 [Caldicellulosiruptor owensensis OL]	77.23
L2	L2 [Sulfurihydrogenibium sp. YO3AOP1]	81.09
L3	L3 [Thermotoga maritima MSB8]	81.9
L4	L4 [Thermosipho melanesiensis BI429]	65.61
L5pL11e	L5 [Desulfotomaculum kuznetsovii DSM 6115	77.4
L6pL9e	L6 [Sulfurihydrogenibium sp. YO3AOP1]	74.32
L9u	L9 [Thermobacillus composti KWC4]	60.54
L10uP0ae	L10 [Thermotoga petrophila RKU-1]	58.76
L11pL12e	L11 [Thermotoga neapolitana DSM 4359]	87.94
L12u	L7/L12 [Sulfurihydrogenibium yellowstonense	84.13
L13	L13 [Aquifex aeolicus VF5]	85.31
L14pL23e	L14 [Thermus thermophilus HB8]	86.89
L15pL27e	L15 [Thermotoga maritima MSB8]	77.7
L16uL10ae	L16 [Marinithermus hydrothermalis DSM 148	81.43
L17u	L17 [Nitratireductor aquibiodomus RA22]	69.23
L18pL5e	L18 [Thermotoga maritima MSB8]	90.08
L19u	L19 [Bacillus smithii 7_3_47FAA]	80.36
L20u	L20 [Sulfurihydrogenibium azorense Az-Fu1]	75.42
L21u	L21 [Caldalkalibacillus thermarum TA2.A1]	71.84
L22pL17e	L22 [Thermoanaerobacter tengcongensis MB4	77.48
L23	L23 [Thermotoga neapolitana DSM 4359]	70.65
L24pL26e	L24 [Thermosipho melanesiensis BI429]	75.47
L25u	L25 [Thermotoga sp. RQ2]	52.94
L27u	L27 [Sulfurihydrogenibium azorense Az-Fu1]	84.52
L28u	L28 [Thermotoga neapolitana DSM 4359]	76.47
L29pL35e	L29 [Geobacillus thermodenitrificans NG80-2	72.73
L30pL7e	L30 [Kyrpidia tusciae DSM 2912]	66.13
L31u	L31 [Thermus thermophilus HB8]	68.12
L32u	L32 [Hydrogenivirga sp. 128-5-R1-1]	67.8
L33u	L33 [Dictyoglomus turgidum DSM 6724]	83.67
L34u	L34 [Coxiella burnetii RSA 493]	84.09
L35u	L35 [Syntrophus aciditrophicus SB]	66.67
L36u	L36 [Sulfurihydrogenibium azorense Az-Fu1]	91.89
23S	23S [Thermotoga sp. RQ2]	85.52

Table 1. After inferring ancestral 50S ribosomal sequences, we performed NCBI BLAST searches, using the ancestral predictions as queries. Listed are the top hits for each search as well as the % identity of a pairwise alignment between query and subject.

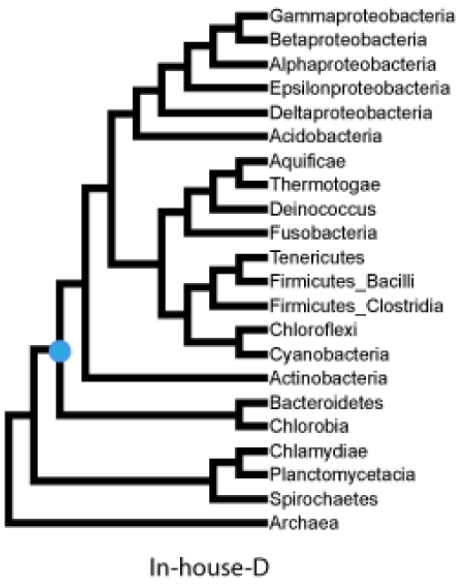

Figure 2. A model of Bacterial evolution, inferred using MrBayes 3.2 and alignments of 17 rProtein families. As far as we know, alignment data for the 50S rProteins fit this model better than any published model. The blue dot indicates our node of interest.

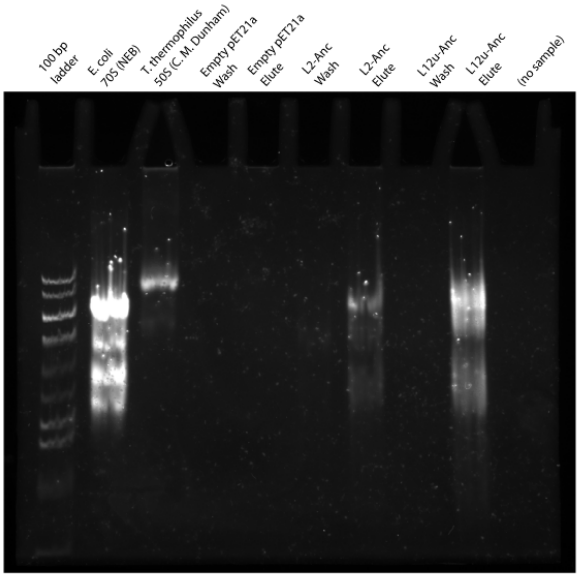

Figure 3. Separation of large and small ribosomal subunits using a composite gel (recipe from C. M. Dunham, personal communication, 2011). Empty plasmid and L2-Anc cultures grew in LB and were induced at OD 0.2 with IPTG to 100 uM. The L12u-Anc culture grew in autoinduction media. Clarified lysates were rocked with 2 mL Ni-NTA 50% slurry (Qiagen) and 10 mM imidazole for 30 mins at 4 °C. Wash volumes were 30-60 mL 20 mM imidazole and 300 mM NaCl. We eluted with 6 mL of 300 mM imidazole, separating into 0.6 mL fractions. Fractions were concentrated on Amicon 100K membranes.

Other workers demonstrated the translational activity of chimeric ribosomes. Friedman (1971) showed that *Bacillus stearothermophilus* ribosomes are more thermostable than *Escherischia coli* ribosomes. Altenberg & Saunders (1971) showed that *E. coli* 30S + *B. stearothermophilus* 50S form active ribosomes that are more thermostable than *E. coli* 50S + *B. stearothermophilus* 30S hybrids. Nomura et al. (1968) reconstituted hybrid 30S using rProteins from *E. coli* and rRNA from *B. stearothermophilus*, and found that these chimeras were active in translation.

Bacterial ribosomal components may be interchangeable. This model fits the published data as well as my initial results. Preliminary experiments suggest that ancestral rProteins assemble in to active *E. coli* ribosomes. Following overexpression of the ancestral L2-6xHis gene, a subpopulation of ribosomes likely contains ancestral L2-6xHis instead of *E. coli* L2. Similarly, ancient L12u-6xHis probably replaces its modern counterpart in some ribosomes, when this gene overexpresses in *E. coli.* I arrive at this interpretation by comparing UV absorbance (not shown) and electrophoretic data **(Figure 3)** of wash fractions with elution fractions. These data are consistent with the interpretation that His-tagged ancestral L12u and L2 bound specifically to E. coli ribosomes, and the His tags bound to an Ni-NTA affinity column, immobilizing chimeric 70S until elution with imidazole.

These initial results do not compel the interpretation that ancient rRNA cooperates with ancient rProteins to assemble a ribosomal or subribosomal 50S-type particle *in vitro*. However, they are consistent with the model that ribosome assembly is a) *intrinsic* to the isolated ribosomal machinery (Nierhaus, 2004), and b) ***conserved*** **throughout Bacteria.**

How **conserved is Bacterial 50S assembly cooperation? This thesis connects sequence evolution with phenotype evolution. I propose to build ancient 50S-like particles, using rProteins and rRNA unknown to modern biology (Table 1). The *in vitro* assembly experiments that I propose ask: *how conserved is Bacterial 50S assembly cooperation?***

2. Motivation

Assembling ancestral ribosomal machinery will enrich the origin of life, macromolecular assembly, and *in vitro* translation fields.

If the RNA world hypothesis is true, and emergence of the peptidyltransferase center of the ribosomal RNA marked the beginning of the genetic code and life as we know it (Fox & Naik, 2004), then it seems plausible that the road from the RNA world to the RNA+protein[3] world is a **series of rRNA and rProtein duplication events[4]. The experiments proposed here explore the RNA world hypothesis**.

I am interested in the specificity with which ancestral rRNA – rProtein assembly cooperations occur. Are the binding events of ancestral assembly the same, different, or more flexible than those of modern assembly? **Has the dependency network (Nierhaus, 2004) evolved from**

[3] See Dworkin et al. (2003) for a discussion of alternative origin-of-life scenarios, and a rationale for why the RNA world to RNA+protein world model seems most likely.

[4] For example, Agmon et al. (2005) observed that the peptidyltransferase center situates within a structurally symmetrical region of the 23S rRNA. These authors found no evidence for sequence homology between the two sides of the RNA. However, rapid divergence can obscure homology.

one specific state to another, from dense to sparse, or from sparse to dense (Fig. 4)? Any of these scenarios would be interesting to the ribosome and origin-of-life community. Ancestral enzymes often have more promiscuous activity than their modern-day counterparts (Cole & Gaucher, 2011). If this is the case for ribosomal proteins, then one rProtein may be able to stand in for several others. Such a scenario would be consistent with rProtein families sharing common ancestors – that is, **several or all of the rProteins may be ancient paralogs**[5,6].

I will assay rProtein binding dependency for the "second-to-last" common ancestor of Bacteria[7]. These tests may not show a change in rProtein/ rRNA assembly cooperation. Such a result would not necessarily refute the paralog hypothesis, if the vast majority of Bacterial rProtein specialization occurred along the branch leading from the cenancestor of life to the cenancestor of Bacteria. On the other hand, if rProteins underwent significant subfunctionalization during Bacterial divergence, we may find evidence of ancestral assembly promiscuity in these binding assays.

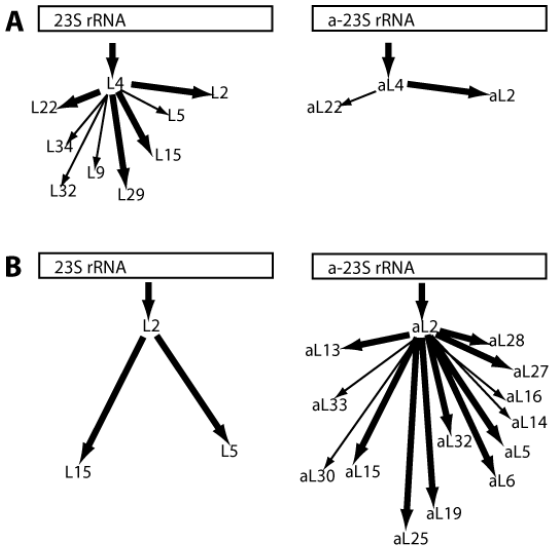

Figure 4. A. Ancestor with sparse dependency network suggesting that ribosomes evolved promiscuous assembly cooperation over time. **B.** Ancestor with dense dependency network suggesting that ribosomes evolved specific assembly cooperation over time. (Dependency maps in left column are for modern E. coli, from Nierhaus lab. Maps in right column are speculative.)

My thesis centers on ribosome assembly *in vitro*. However, collaborative work in the Gaucher lab could have the potential to identify commercially relevant properties of ribosome-derived particles in the area of translational activity (e.g., Cole & Gaucher, 2012). A deeper understanding of natural solutions to macromolecular assembly and function may inform human approaches to real-world challenges. Initial analysis of our ancestral rProteins and rRNA indicate that they will be thermostable upon purification. Individual rProteins and rRNA purified from ancestral and *E. coli* sequences would create a large space of potential chimeric ribosomes that could result from assembly reactions taking components from different sources. Each chimera could stimulate a different set of productive questions useful for increasing the yield and longevity of *in vitro* translation systems.

The components of the ancestral ribosome in this thesis will probably be purified by over expression in *E. coli*, because this approach is presently more economical than *in vitro* translation such as the PURE system[8]. However, an eventual consequence of this project is

[5] For examples of ancient paralogs, see Iwabe et al. (1989) or Gogarten et al. (1992).

[6] I am not the first to have this idea (Jue et al., 1980; Fox, 2010), but to my knowledge I am the first to propose an experiment that might support it.

[7] My trees describe the divergence of 115 Bacteria and 94 Archaea. Assuming that the root lies between these two domains of life, the "second-to-last" common ancestor of Bacteria is the earliest Bacterial node whose sequences can be reconstructed reliably. This constraint exists because some of the Bacterial rProteins are found only in Bacteria, and have no Archaeal outgroup.

[8] *In vitro* transcription of rRNA, absent the S30 extract, is unlikely to be useful for assembling active ribosomes, because of the guaranteed absence of methylation.

improved-yield *in vitro* translation; future ribosome assembly projects may be entirely *in vitro*.

3. Experimental strategy

To conduct a survey of ribosomal sequences in Bacteria and Archaea, I used BLAST and keyword searches of the NCBI and the PDB, as well as a literature review[9]. **Fig. 1** and **Table 1** summarize this census; every family name corresponds to a set of rProtein or rRNA sequences, and every sequence corresponds to one taxon of the 209 considered by this study. I chose taxa with completely sequenced genomes, that had been used in previous phylogenetic studies[10], and that seemed diverse.

I aligned the 23S rRNA using structure as a guide. I aligned the *H. marismortui* (3G71), *E. coli* (3ORB), and *T. thermophilus* (2X9S) 23S rRNA models using UCSF Chimera. I expanded this seed alignment using the ARB, GenBank, and Sanger databases. I used CLUSTAL to generate initial alignments of the rProteins and rRNA. I inspected and adjusted each alignment by hand, using MacClade.

Evolutionary models that consider rate correlations of base-paired sites in RNA estimate the likelihood of the data more accurately than those models that ignore this parameter (Jow et al., 2002). I used RNAVIEW and VARNA to extract the 23S basepair information in dot-bracket notation from *E. coli* and *H. marismortui* PDB files. I used PHASE to infer the consensus secondary structure from the aligned dot-bracket strings.

This project reconstructs sequences using published phylogenies as well as phylogenies resulting from my own analysis. I performed a RAxML analysis (CIPRES cluster) for a concatenated rProtein alignment assuming that each site evolved independently and that the site-specific rates were Gamma-distributed (WAG model). I also performed a MrBayes 3.2 run for 7000000 generations (WAG+G) (Biocluster). This run allowed the gamma distribution to differ for each rProtein family. This extra parameter contributed to a much longer runtime for MrBayes over RAxML. However, the extra effort yielded a phylogeny with significantly better likelihood than that generated by RAxML, as determined by the Approximately Unbiased test (Shimodaira & Hasegawa, 2001). My rProtein alignment data fit the MrBayes and RAxML models significantly better than published models, as well as the model I generated using the 23S alignment.

Most evolutionary models do not consider insertions and deletions. To reconstruct ancestral character states when gaps are considered as a potential ancestral state, I used the GASP algorithm (Edwards & Shields, 2004) as well as the GLOOME webserver (Cohen et al., 2010).

Based on the GASP or GLOOME output I generated a set of alignments that include only those sites inferred as present in the node of interest. I ran these alignments in PAML (or PHASE, for the rRNA), and used the best-fitting topology as determined earlier. These runs generated the ancestral sequences that I am using to build an ancient ribosome or sub-ribosomal particle.

---

[9] (Nierhaus, 2004; Arnold & Reilly, 1999; Sykes et al., 2010; Tatusov, 1997; Lecompte, 2002; Yutin et al., 2012)

[10] (Battistuzzi & Hedges, 2009; Brochier-Armanet et al., 2011; Wu & Eisen, 2008).

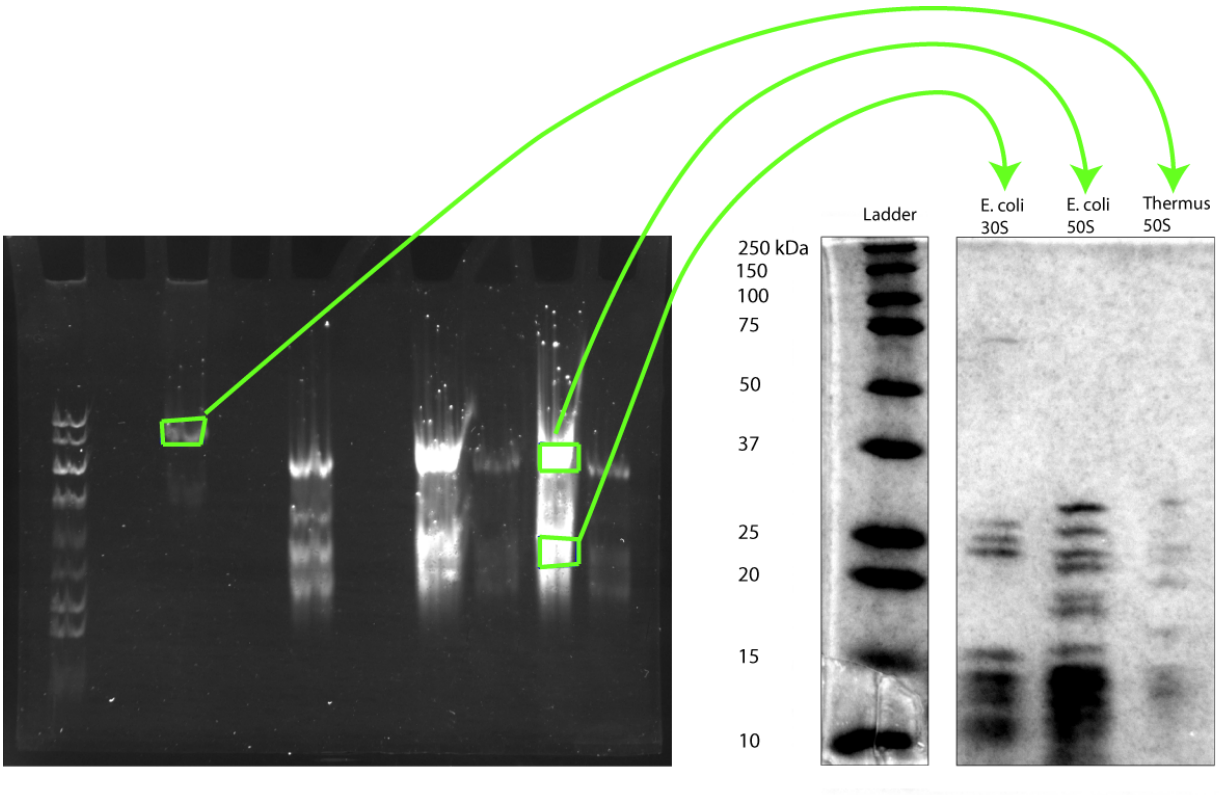

Figure 5. At left, a composite agarose-acrylamide gel showing separation of large and small ribosomal subunits, stained with EtBr. Lanes (L to R): 100 bp ladder, 6 uL 350 nM T. thermophilus 50S, 6 uL 350 nM E. coli 70S, 15 uL 350 nM E. coli 70S, 40 uL 350 nM E. coli 70S. Three bands were cut from the gel as shown, heated in 30 uL Laemmli loading buffer for 10 mins at 95°C, and loaded into the wells of a gel for SDS-PAGE (right). SDS gel was stained with Coomassie.

My genes will have terminal His-tags and restriction sites to enable removal if appropriate. The imidazole side-chains in a histidine tag bind Ni-NTA with an affinity of 14 nM (Knecht et al., 2009). I will use Ni-NTA columns to individually purify the ancestral components under denaturing conditions[11]. Ancestral rRNA purification could follow (Leonov et al., 2003): the ancestral rRNA would contain a binding site for streptavidin. I would load a streptavidin column with clarified lysate from the ancestral rRNA culture. Eluting with 10 mM biotin would yield chimeric *E. coli* ribosomes incorporating ancestral 23S or 5S rRNA. The proteins could be removed by performing an RNeasy MaxiPrep (Qiagen). I may also try adding an MS2 binding site to the 23S (Youngman & Green, 2005), depending on the success of the streptavidin approach. I will probably use plasmid pK4-16 to express the ancient 23S in *E. coli*. This plasmid contains the rrnB operon (Quan et al., 2010; Yale CGSC).

Given any combination of rProteins and rRNA, the Nomura/ Nierhaus dependency experiments, in vitro, detected the subset of rProteins that participated in assembly. By performing hundreds of assembly reactions using different combinations of rProteins with rRNA, the Nomura and Nierhaus labs showed that certain proteins do not assemble in to the ribosome unless other proteins have already been incorporated or are present in the reaction. **The synthesis of these**

[11] Semrad et al., 2004 found that about half of the E. coli rProteins form inclusion bodies upon overexpression. Therefore, my purifications will probably be under denaturing conditions.

experiments yielded "dependency maps" (Nierhaus, 2004). In the traditional LSU setup, a subset of rProteins incubates at 44 °C for 20 mins at a concentration of 1400 nM each, along with 700 nM of 23S, in a volume of 200 uL, with ionic conditions as detailed in Röhl et al, 1982. The mixture is layered on to a sucrose density gradient and spun at 250,000 g for 2.75 hours. The "heavy" particles as indicated by an A_{260} trace are isolated from the gradient, and subjected to overnight acid treatment, which causes the rProteins to precipitate. The precipitate is isolated, redissolved in denaturing conditions, run on SDS-PAGE, and the gel is silver-stained. Thus, all rProteins that bound to the assembly intermediate are present in the gel.

My approach to assembly mapping is similar to the method from Nomura and Nierhaus. However, composite gels may be used to isolate rRNA-rProtein complexes in lieu of ultracentrifugation. Although its usefulness for preparing intact ribosomes is unclear, this method seems suitable for my specific purpose of monitoring rProteins that were bound to rRNA. rRNA-containing bands can be cut from the composite gel and loaded into the well of an SDS-Polyacrylamide gel **(Fig. 5)**. SDS-PAGE serves as the readout for rProtein-23S assembly. It seems feasible to add electrophoretic dimensions, or MALDI, to resolve co-migrating bands. **This thesis will focus on assembly dependencies that seem interesting,** and leave an exhaustive mapping to future studies that use high-throughput approaches (e.g., Sykes et al., 2010).

This project is a contribution toward the longer-term goal of taking life apart and putting it back together again. Integrating *in vitro* translation, transcription, and replication into a self-sustaining system will reinvent life on human terms. Over the next few decades, this innovation will advance renewable energy, material science, and medicine. Our present work in building ancient ribosomal particles explores a fundamental question in molecular evolution and advances a synthetic biology more amenable to rational design.

Most Probable Schedule

Nov 2012: Finalize sequences
Nov 2012 – May 2013: Individually purify rProteins
- Nov 2012 – Dec 2012: Purify Anc-L15pL27e, L15pL27e, Anc-L20u, L20u, Anc-L13, L13
- Jan 2013: Purify L3, L5pL11e, L24pL26e, L4, L22pL17e, Anc-L3, Anc-L5pL11e, Anc-L24pL26e, Anc-L4, Anc-L22pL17e
- Feb 2013: Purify L1pL10ae, L23, and/or others (E. coli & Ancestral)
- Mar 2013: Purify L29pL35e, L17u, and/or others (E. coli & Ancestral)
- Apr 2013: Purify L34u, L33u, and/or others (E. coli & Ancestral)
- May 2013: Purify L19u, L30pL7e, and/or others (E. coli & Ancestral)

December 2012 – April 2013: Purify rRNA
- Dec 2012: Design affinity tag insertions
- Jan 2013: Site-directed mutagenesis on pK4-16 to add restriction sites if necessary
- Mar 2013: Insert ancestral and/or E. coli tagged 23S in to pK4-16
- Apr 2013: Transform, culture, harvest, purify

December 2012 – May 2013: Solubility optimization
January 2013 – January 2014: Conduct assembly experiments
February 2014 – May 2014: Analyze data and draw conclusions
June 2014: Thesis defense

Budget (excluding wages):

```
$4000  ..... Synthetic genes
$1000  ..... Media & reagents
$1000  ..... Nitrile gloves, RNase-free tubes & tips
$1700  ..... BugBuster (Novagen)
$300   ..... RNase inhibitor (NEB)
$550   ..... DNase I (NEB)
$1000  ..... Ni-NTA Agarose (Qiagen)
$1000  ..... Streptavidin Agarose (Pierce)
$1000  ..... Mass spectrometry services
$450   ..... RNeasy MaxiPrep (Qiagen)
$1000  ..... Amicon concentrators (Millipore)

$13000 ..... Total
```

References

Agirrezabala X, et al. Structural characterization of mRNA-tRNA translocation intermediates. P Natl Acad Sci USA (2012) 109:6094-6099.

Agmon F, Bashan A, Zarivach R, Yonath A. Symmetry at the active site of the ribosome: structural and functional implications. Biol Chem (2005) 386:833-844.

Altenburg LC, Saunders GF. Properties of Hybrid Ribosomes Formed from Subunits of Mesophilic and Thermophilic Bacteria. Journal of Molecular Biology (1971) 55:487-&.

Arnold RJ, Reilly JP. Observation of Escherichia coli ribosomal proteins and their posttranslational modifications by mass spectrometry. Anal Biochem (1999) 269:105-112.

Battistuzzi FU, Hedges SB. A major clade of prokaryotes with ancient adaptations to life on land. Mol Biol Evol (2009) 26:335-343.

Bingel-Erlenmeyer R, et al. A peptide deformylase-ribosome complex reveals mechanism of nascent chain processing. Nature (2008) 452:108-U113.

Brochier-Armanet C, Forterre P, Gribaldo S. Phylogeny and evolution of the Archaea: one hundred genomes later. Curr Opin Microbiol (2011) 14:274-281.

Cohen O, Ashkenazy H, Belinky F, Huchon D, Pupko T. Gloome: Gain Loss Mapping Engine. Bioinformatics (2010) 26:2914-2915.

Cole MF, Gaucher EA. Utilizing natural diversity to evolve protein function: applications towards thermostability. Current Opinion in Chemical Biology (2011) 15:399-406.

Cole, M. F. and E. A. Gaucher, 2012. Engineering the translation machinery to accept unnatural amino acids. The Sixth Peptide Engineering Meeting, Emory University Conference Center-Atlanta, Georgia.

Dworkin JP, Lazcano A, Miller SL. The roads to and from the RNA world. J Theor Biol (2003) 222:127-134.

Ederth J, Mandava CS, Dasgupta S, Sanyal S. A single-step method for purification of active His-tagged ribosomes from a genetically engineered Escherichia coli. Nucleic Acids Research (2009) 37.

Edwards RJ, Shields DC. GASP: Gapped ancestral sequence prediction for proteins. BMC Bioinformatics (2004) 5.

Felsenstein J. Evolutionary Trees from DNA-Sequences - a Maximum-Likelihood Approach. Journal of Molecular Evolution (1981) 17:368-376.

Felsenstein, Inferring Phylogenies, 2004
Fox GE. Origin and Evolution of the Ribosome. Csh Perspect Biol (2010) 2.
Fox, G.; Naik, A. In The Genetic Code and the Origin of Life; 2004; pp 92-105.
Frank J, Agrawal RK. A ratchet-like inter-subunit reorganization of the ribosome during translocation. Nature (2000) 406:318-322.
Frauenfeld J, et al. Cryo-EM structure of the ribosome-SecYE complex in the membrane environment. Nature Structural & Molecular Biology (2011) 18:614-U127.
Friedman SM. Heat Stabilities of Ribosomal Subunits and Reassociated Ribosomes from Bacillus-Stearothermophilus. Journal of Bacteriology (1971) 108:589-&.
Gaucher E, Govindarajan S, Ganesh O. Palaeotemperature trend for Precambrian life inferred from resurrected proteins. Nature (2008) 451:704-708.
Gaucher EA. Variants of Ancestral Uricases and Uses Thereof (2011) United States of America: Georgia Tech Research Corporation.
Gogarten JP, Taiz L. Evolution of Proton Pumping Atpases - Rooting the Tree of Life. Photosynth Res (1992) 33:137-146.
Herold M, Nierhaus KH. Incorporation of six additional proteins to complete the assembly map of the 50 S subunit from Escherichia coli ribosomes. J Biol Chem (1987) 262:8826-8833.
Iwabe N, Kuma K, Hasegawa M, Osawa S, Miyata T. Evolutionary Relationship of Archaebacteria, Eubacteria, and Eukaryotes Inferred from Phylogenetic Trees of Duplicated Genes. P Natl Acad Sci USA (1989) 86:9355-9359.
Jow H, Hudelot C, Rattray M, Higgs PG. Bayesian phylogenetics using an RNA substitution model applied to early mammalian evolution. Molecular Biology and Evolution (2002) 19:1591-1601.
Jue RA, Woodbury NW, Doolittle RF. Sequence homologies among E. coli ribosomal proteins: evidence for evolutionarily related groupings and internal duplications. J Mol Evol (1980) 15:129-148.
Knecht S, Ricklin D, Eberle AN, Ernst B. Oligohis-tags: mechanisms of binding to Ni2+-NTA surfaces. J Mol Recognit (2009) 22:270-279.
Lecompte O, Ripp R, Thierry JC, Moras D, Poch O. Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. Nucleic Acids Res (2002) 30:5382-5390.
Leonov AA, Sergiev PV, Bodganov AA, Brimacombe R, Dontsova OA. Affinity purification of ribosomes with a lethal G2655C mutation in 23 S rRNA that affects the translocation. Journal of Biological Chemistry (2003) 278:25664-25670.
Liu PF, Avramova LV, Park C. Revisiting absorbance at 230 nm as a protein unfolding probe. Anal Biochem (2009) 389:165-170.
Llano-Sotelo B, et al. Binding and action of CEM-101, a new fluoroketolide antibiotic that inhibits protein synthesis. Antimicrob Agents Chemother (2010) 54:4961-4970.
Mizushim.S, Nomura M. Assembly Mapping of 30s Ribosomal Proteins from E-Coli. Nature (1970) 226:1214-&.
Nierhaus K. Assembly of the prokaryotic ribosome. In: Protein synthesis and ribosome structure--Nierhaus KH, Wilson DN, eds. (2004) Weinheim: Wiley-VCH. 85-105.
Quan S, Skovgaard O, Squires CL. Escherichia coli Ribosomal RNA deletion strain (2010).
Schmeing T, et al. A pre-translocational intermediate in protein synthesis observed in crystals of enzymatically active 50S subunits. Nature Structural Biology (2002) 9:225-230.
Schulze H, Nierhaus KH. Minimal Set of Ribosomal Components for Reconstitution of the Peptidyltransferase Activity. Embo Journal (1982) 1:609-613.
Seidelt B, et al. Structural Insight into Nascent Polypeptide Chain-Mediated Translational Stalling. Science (2009) 326:1412-1415.
Semrad K, Green R, Schroeder R. RNA chaperone activity of large ribosomal subunit proteins from Escherichia coli. Rna-a Publication of the Rna Society (2004) 10:1855-1860.

Shajani Z, Sykes MT, Williamson JR. Assembly of Bacterial Ribosomes. Annu Rev Biochem (2011) 80:501-526.
Shimodaira H, Hasegawa M. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics (2001) 17:1246-1247.
Swaney S, McCroskey M, Shinabarger D, Wang ZG, Turner BA, Parker CN. Characterization of a high-throughput screening assay for inhibitors of Elongation Factor P and ribosomal peptidyl transferase activity. J Biomol Screen (2006) 11:736-742.
Sykes MT, Shajani Z, Sperling E, Beck AH, Williamson JR. Quantitative Proteomic Analysis of Ribosome Assembly and Turnover In Vivo. Journal of Molecular Biology (2010) 403:331-345.
Tatusov RL, Koonin EV, Lipman DJ. A genomic perspective on protein families. Science (1997) 278:631-637.
Traub P, Nomura M. Structure and Function of Escherichia Coli Ribosomes .6. Mechanism of Assembly of 30-S Ribosomes Studied in Vitro. Journal of Molecular Biology (1969) 40:391-&.
Wang HH, et al. Multiplexed in Vivo His-Tagging of Enzyme Pathways for in Vitro Single-Pot Multienzyme Catalysis. Acs Synth Biol (2012) 1:43-52.
Wu M, Eisen JA. A simple, fast, and accurate method of phylogenomic inference. Genome Biol (2008) 9:R151.
Yang ZH, Kumar S, Nei M. A New Method of Inference of Ancestral Nucleotide and Amino-Acid-Sequences. Genetics (1995) 141:1641-1650.
Youngman EM, Green R. Affinity purification of in vivo-assembled ribosomes for in vitro biochemical analysis. Methods (2005) 36:305-312.
Yutin N, Puigbo P, Koonin EV, Wolf YI. Phylogenomics of Prokaryotic Ribosomal Proteins. PLoS One (2012) 7.
Zhou J, Lancaster L, Trakhanov S, Noller HF. Crystal structure of release factor RF3 trapped in the GTP state on a rotated conformation of the ribosome. Rna-a Publication of the Rna Society (2012) 18:230-240.
Zhou Y, Asahara H, Gaucher EA, Chong SR. Reconstitution of translation from Thermus thermophilus reveals a minimal set of components sufficient for protein synthesis at high temperatures and functional conservation of modern and ancient translation components. Nucleic Acids Research (2012) 40:7932-7945.