
Assembly of ribosomes in eukaryotes

JOHN WOOLFORD

Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA

What we knew before RNA

Since their discovery more than 65 years ago, ribosomes have been a prototype to understand the biology of ribonucleo-protein particles (RNPs) and catalytic RNAs. However, their complexity makes figuring out the pathway of their assembly quite challenging. We now know that more than 350 protein and RNA molecules participate in yeast ribosome assembly, and many more in metazoa. Assembly intermediates are heterogeneous and dynamic, and resemble mature ribosomes too closely to isolate them by traditional means. Those of us who face these challenges want to understand how the 7000–14,000 nucleotides of pre-rRNA are folded, covalently modified, processed to remove spacer sequences, and bound to 79–80 different ribosomal proteins (r-proteins) to create this dynamic nanomachine in eukaryotes.

Nomura and Nierhaus's heroic experiments in the 60's and 70's to reconstitute bacterial ribosomes *in vitro* laid the foundation, showing that the information necessary for assembly is contained in the component r-proteins and mature rRNAs, and that assembly is hierarchical and cooperative. However, reconstitution of active subunits was inefficient and had to be carried out under non-physiological conditions. We imagined that additional activities occurring *in vivo* would enable the accurate and efficient assembly necessary for cells to grow and divide. Groundbreaking work done in the '70s by Penman, Perry, Darnell, and especially Warner and Planta, provided an outline of the ribosome assembly pathway in eukaryotes. They discovered the pre-rRNA processing intermediates, and showed that they are methylated and packaged into RNPs containing r-proteins as well as unidentified non r-proteins that they forecast would assist assembly (assembly factors, AFs). Their initial results also led them to predict that subunits underwent their last steps of maturation in the cytoplasm. Early experiments of J. Steitz, Gerbi, and Tollervey showed that small nucleolar RNAs (snoRNAs) participate in ribosome assembly.

Tools that have fueled us

The birth of the journal *RNA* in 1995 coincided roughly with the re-emergence of ribosome biogenesis research from a relatively quiet sabbatical to a period of remarkable progress. Entry of ribosome assembly research into mid-log phase was propelled by classical and molecular genetic screens in yeast, to discover and initially characterize AFs. Development of genomic and proteomic tools enabled an important breakthrough, the first isolation of ribosome assembly intermediates from cells. Epitope-tagged AFs were used to affinity-purify pre-ribosomes, and biologist-friendly mass spectrometry facilitated identification of the co-purifying AFs. The inventory of proteins identified as AFs quickly soared to 200+, a complexity that even Nomura found daunting. These included many potential RNA binding proteins as well as NTPases. Determining which proteins and pre-rRNAs co-purified with each other, and assaying the effects of depleting each factor, told us approximately when each factor enters and exits from pre-ribosomes and for which pre-rRNA processing step each AF is first required. More recently, nucleolar proteomics and more facile functional genomics in metazoa have identified 286 human proteins necessary for ribosome production, including 74 not found in yeast.

Determination of the atomic structure of bacterial and eukaryotic ribosomes then transformed our field, providing a platform for much more sophisticated experimental design and interpretation. However, these structures do not reveal the pathway of assembly. We still need to learn exactly how each AF and r-protein interact with each other and with pre-rRNA as assembly proceeds. The development of methods combining protein-RNA crosslinking with high throughput RNA sequencing has enabled us to gain important information about how these proteins contact RNA in pre-ribosomes. Advances in RNA structure probing, together with high throughput sequencing, are starting to reveal how pre-rRNA structure is remodeled during assembly. Most promising is the development of improved cryo-EM technology that can be used to generate near-atomic models of the structure of assembly intermediates.

Corresponding author: jw17@andrew.cmu.edu

Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.050633.115>. Freely available online through the *RNA* Open Access option.

© 2015 Woolford This article, published in *RNA*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

What have we learned in the last 20 years?

Subunit biogenesis begins with transcription of 35S-47S pre-rRNA by RNA polymerase I, and the pre-5S rRNA by RNA polymerase III. Assembly starts immediately. Local folding and modification of pre-rRNA, processing of spacers at the 5' end of the RNA, and protein binding occur co-transcriptionally. Among the earliest events are methylation of riboses and conversion of uridines to pseudo-uridines, to stabilize pre-rRNA folding. These reactions are mediated by methyltransferases and pseudo-uridine synthases associated with snoRNAs that base-pair with 100–200 different target sequences in pre-rRNA. It remains largely a mystery how so many snoRNPs find their targets in what appears to be a very short interval of the assembly pathway. Presumably, relatively open RNA structures must be established or maintained for these modifications to occur before nascent particles are compacted. Not only modifications, but also simply the base-pairing of the snoRNAs with pre-rRNA could enable proper pre-rRNA folding. Interestingly, these RNA modifications occur primarily in the active sites of each subunit, the rRNA sequences comprising the decoding and peptidyltransferase centers (DC and PTC). However, no single modification is essential, and not every potential target is modified in every ribosome. Whether such structural heterogeneity creates functionally heterogeneous ribosomes, i.e. ribosomes specialized for translation of a subset of mRNAs, has become a provocative issue.

In yeast, four snoRNAs, U3, U14, snr10, and snr30, participate in steps other than modification, namely early cleavage events at the 5' end of pre-rRNA. U3 snoRNA, the best studied thus far, functions as an RNA chaperone. It base-pairs with at least six sites spread over 1500 nucleotides, and thus might organize this domain as a proper substrate for cleavage. U3 also base-pairs with sequences that can form the conserved central pseudoknot, perhaps to prevent premature formation of this structure. U3 is but one component of a very large complex, the processome, which is necessary for compaction of pre-40S particles, seen as knobs at the 5' ends of nascent transcripts by EM of “Miller spreads” of transcribing rDNA. The processome assembles hierarchically from six protein sub-complexes and at least 40 other proteins, and might facilitate the “sno storm” as well as the subsequent early cleavage and compaction events.

Most r-proteins associate with the earliest pre-ribosomes, but establish stabler interactions with rRNA as assembly proceeds, analogous to the transition from encounter complexes to mature structures described for bacterial subunit reconstitution. Systematic depletion or knockdown of r-proteins provided a global view for how this might happen in vivo. Assembly proceeds hierarchically from one neighborhood to the next within each subunit. The body substructure of the small subunit (SSU) containing the 5' and central domains of 18S rRNA is assembled before the head structure containing the 3' major domain of 18S rRNA. The former

step is important for the earliest cleavage events at the A₀ and A₁ sites, while the latter is necessary for co-transcriptional cleavage at the A₂ site, to release the pre-40S particle from the transcription complex. Assembly of the large subunit (LSU) begins with binding of r-proteins to domains I, II, and VI of LSU rRNA to create a particle able to undergo processing of the ITS1 spacer RNA. These steps are coupled with subsequent construction of the polypeptide exit tunnel and cleavage within the ITS2 spacer. This leads to stabilization of the inter-subunit domain, including formation of the neighborhood surrounding the central protuberance (CP) and the PTC. Thus, assembly proceeds largely 5' to 3' with respect to RNA for both subunits, with the notable exception that domain VI at the 3' end of 25S rRNA assembles early. Joining of the 5' and 3' ends of this long RNA seems to be required to initiate subsequent processing and remodeling steps of the pre-60S particle.

A major challenge remaining is to determine the precise functions of AFs. For example, are the numerous NTPases present in pre-ribosomes remodelers, timers, or proofreaders? An example of remodeling and structural proofreading mediated by NTPases is evident from late nuclear steps of LSU biogenesis. Assembly of the GTPase Nog2 triggers ITS2 cleavage and stimulates significant remodeling of late nuclear pre-60S particles by the AAA ATPase Rea1. This conformational change may trigger Nog2 GTPase activity to release Nog2 from pre-ribosomes, allowing its replacement in the same site by the nuclear export adaptor Nmd3. Thus, remodeling and proofreading may be coupled with licensing for export. Do the nineteen different DEAD-box proteins (DBPs) found in yeast pre-ribosomes drive assembly by remodeling RNA or RNP structure, and if so, exactly what are their substrates? A few of these ATPases have been implicated in release of snoRNAs. One intriguing hypothesis is that DBPs might participate in kinetic proofreading, by distinguishing on-pathway from stalled substrates.

The most important steps in subunit maturation, completion of construction of the active sites and proofreading of their functionality, are saved for the end of the pathway, in the cytoplasm. In an ordered series of reactions, release of the remaining AFs is coupled with association of the last few r-proteins. The late removal of these AFs might serve several purposes: delaying final steps of RNA folding to allow earlier steps to be completed properly, and preventing RNA misfolding into kinetic traps. These last domains to be remodeled also overlap with binding sites for translation factors or with inter-subunit bridges. Thus, the presence of late AFs in cytoplasmic pre-ribosomes could prevent premature association of nascent subunits with functional translational machinery. Yet, some test-driving does occur. Late pre-40S particles engage in a translation-like cycle with 60S subunits, perhaps to test the functionality of translation factor binding sites. Likewise, in late pre-60S particles r-protein L10 is used to detect proper assembly of the P-site and the GTPase activation center. This occurs through a network of

interactions that also triggers release of the last AFs whose presence could block interaction of 60S subunits with 40S subunits. Thus, newly made LSUs are licensed for entry into the pool of active ribosomes.

Other important unanswered questions

We now have a detailed outline of the pathway of ribosome assembly in yeast, including a mostly complete list of the players involved. Tools are in place to investigate the pathway at higher resolution in yeast and also in metazoa. Bacterial subunit reconstitution *in vitro* proceeds through multiple alternative pathways. To what extent does this occur *in vivo* and in eukaryotes? Could this be mediated

by differential pre-rRNA modification, post-translational modification of r-proteins, r-protein isoforms, or simply the absence of certain r-proteins from some active ribosomes, to generate specialized ribosomes? The functions of most r-proteins and AFs have been explored by their depletion or knockdown, which may only reveal their first function or the effects of destabilizing pre-rRNPs. Examination of informative missense alleles should provide deeper insight into function. How are misassembled ribosomes identified and dealt with, and when they escape detection, how does their dysfunction lead to diseases? Of course, the real frontier remains with RNA. The ultimate challenge is to learn how the very long pre-rRNAs are efficiently folded into functional forms.