Ribosome assembly factors modulate the RNA folding landscape to prevent misfolding

Haina Huang[1] & Katrin Karbstein[1,2*]

[1]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, Jupiter, Florida 33458, United States of America

[2]HHMI Faculty Scholar, Chevy Chase, Maryland 20815, United States of America

*Correspondence: kkarbst@scripps.edu

KEYWORDS: quality control; ribosome assembly; RNA folding

Abstract

Ribosomes are among the largest folded RNAs, whose function depends on their structure. Nonetheless, *in vitro* studies indicate a propensity of rRNAs to misfold. We use a combination of DMS-MaPseq, structural analyses, biochemical experiments, and yeast genetics to dissect the final RNA folding steps of the small ribosomal subunit head. Our data demonstrate how a succession of assembly factors prevent the early folding of an RNA loop, thereby allowing an adjacent helical junction to fold first. Deletion of the assembly factors to allow the loop to fold first traps a misfolded junction, demonstrating how assembly factors modulate the folding pathway to enable challenging motifs to fold independent of their interaction partners and guide folding to the native structure. Analysis of ribosome intermediate structures indicates that this role for assembly factors is ubiquitous and suggests that helical junctions present significant obstacles to folding.

RNAs are prone to misfolding because alternative secondary structures are often relatively stable, thus kinetically trapping misfolded intermediates[1-4]. Moreover, RNA structure is built in a very modular manner, such that native and misfolded elements can co-exist in the same structure. As a result, tertiary structure elements can stabilize both native as well as non-native structures[5], thereby furthering RNA's propensity to misfold.

Ribosomes are among the largest folded RNAs. To promote their highly compacted structure, and ultimately their function in translation, they display many long-range secondary or tertiary structure elements. Moreover, the small ribosomal subunit in particular undergoes many conformational changes during the translation cycle, thus endowing it with the capacity to adopt multiple structures. As a result, ribosomal RNAs (rRNAs) are prone to misfold *in vitro*[6-9].

In cells, ribosomes are assembled with the help of a large machinery. Many structures have now illuminated its changing architecture, and are the basis for suggestions that this machinery blocks premature interactions between subdomains[10-19]. Nonetheless, what the consequences would be of such premature interactions between domains, or even smaller structural elements, has not been shown. Moreover, because removing one element of the machinery generally leads to complete disassembly, it has been impossible to test the roles of any element in RNA folding.

Here we use *Saccharomyces cerevisiae* to dissect the folding path of two structural elements of the small subunit head, the 3-way junction between helices 34, 35 and 38 (j34), and the loop of helix 31 (l31), which forms part of the active site. The head is the most versatile structural element of the small subunit and can adopt multiple orientations in relationship to the body, the platform and the large subunit. Perhaps not coincidentally, the head structure is prone to misfolding *in vitro*[6-8] and is the slowest to assemble *in vivo*[11-16,18,20]. In this work, we use a combination of DMS-MaPseq, genetics, structural and biochemical analyses to show that in cells j34 folds before l31. Moreover, we demonstrate a role for a succession of factors throughout the assembly process in preventing the premature folding of l31. When these factors are ablated and l31 is allowed to fold prior

to j34, it stabilizes and thus kinetically traps a misfolded version of j34, blocking further assembly. These data show how the formation of premature contacts within ribosomes can be detrimental for folding and demonstrate a novel role for assembly factors in orchestrating the folding order, thereby perturbing the folding landscape to promote the native structure.

Results

In wild type yeast j34 folds before l31.

Structural analyses have identified two late-folding RNA structural elements in the small subunit head: the three-way junction around helices 34/35/38 (j34) and the loop at the tip of helix 31 (l31)[12,15]. The folding of these RNA elements (as well as the nearby binding of universally conserved ribosomal proteins) is interrogated via the ordered dissociation of the 40S assembly factors (AFs) Ltv1, Enp1 and Rio2 **(Figure 1A)**, in a quality control mechanism that uses the formation of 80S-like complexes to test the ability of the nascent subunit to properly identify the start codon[21]. Interestingly, even though these two structural elements interact with each other in mature 40S **(Figure 1B**), structural evidence[12,15], and our biochemical and genetic data[21] indicate that these elements form successively. Both j34 and l31 are unfolded in pre-40S purified *via* tagged Ltv1, which ensures the occupancy of Ltv1. In contrast, in a later pre-40S intermediate purified *via* tagged Nob1 (and essentially lacking Ltv1), j34 but not l31 is fully resolved, showing j34 can be folded without l31. These data, validated in biochemical analyses[21], also suggest that Ltv1 release triggers j34 folding. Finally, a recent structure of 80S-like ribosomes demonstrates that these lack not just Ltv1, Enp1, and Rio2, but also contain fully folded j34 and l31[18]. Together, these data suggest that j34 folds before l31, and indicate that l31 folding occurs upon either Enp1 or Rio2 release **(Figure 1A)**. To test this model, and better pinpoint when l31 folds, we utilized DMS-MaPseq[22] to dissect rRNA structural differences in different pre40S intermediates (**Figure S1A)**. In brief, pre-40S intermediates were purified via Rio2-TAP to ensure the full occupancy of Rio2, and exposed to DMS, which methylates adenosines (A) or cytidines (C) that are not protected by base pairing, tertiary structure or protein binding. Modified A and C residues are misread during reverse transcription, which is assayed by deep sequencing. Therefore,

by analyzing the frequency of mutations of each A and C, we can quantitatively monitor the structural changes over the entire 18S rRNA. To demonstrate the effectiveness of this method, we first compared the structural changes between pre-40S and mature 40S monitored by DMS-MaPseq to those observed by cryo-EM **(Figure S1B-D, Figure 1C)**. This analysis shows that regions in pre-40S that are more accessible to DMS are located in j34, l31, helix 40 (h40), h28 and the top of h44, as expected from cryo-EM structures[12,15,23] **(Figure S1B)**. Those that are less accessible to DMS in pre-40S overlap the binding sites of cytoplasmic AFs **(Figure S1C)**. Thus, in our hands DMS-MaPseq of purified pre-40S intermediates can faithfully isolate the structure of these intermediates.

To dissect if l31 is folded upon release of Enp1 or Rio2, we took advantage of a previously identified mutation, S15_YRR, which accumulates pre-40S after Enp1 release, while still retaining Rio2 (**Figure 1A**[21]**)**. Thus, if l31 folds after Rio2 release, we would expect it to be mostly unfolded as in pre-40S, while l31 should be largely folded if l31 folds upon Enp1 release. Indeed, the DMS-MaPseq results show that the DMS accessibility of l31 in S15_YRR pre40S is similar to that in WT pre40S **(Figure 1C, S1D&E)** strongly suggesting that l31 remains unfolded after Enp1 release, and thus folds upon Rio2 release. This finding is consistent with the location of Rio2, which is positioned to directly interact with the residues in l31[12,14,15] **(Figure S2A)**. In contrast, two out of three significant changes in the j34 region of S15_YRR compared to WT pre-40S (C1274, A1296) resemble the changes in mature 40S, indicating the folding of j34 in S15_YRR pre-40S, as expected based on our previous data[21]. A1427 also trends towards maturation, although the change is only statistically significant in S15_YRR and not mature 40S. Moreover, upon release of Ltv1 and folding of j34, the tip of h40 also becomes ordered as it now interacts with h37[12,15,23] **(Figure S1F**). Ordering of h40 is reflected in the increased protection of A1400 and A1401 (**Figure S1G)**. Notably, in the S15_YRR mutant this protection has been largely attained (**Figure S1G**). We note that C1284 in j34 is significantly more DMS-accessible in S15_YRR pre-40S than in either WT pre40S or mature 40S. This increased accessibility could be a reflection of the prepositioning of the j34 region for the interaction with h31. Consistent with that proposal, nearby C1279, which is adjacent to the tertiary interaction, also opens up in going from WT to S15_YRR pre-40S. In summary, previous structural[12,15] and genetic[21] data, as well as DMS-MaPseq

herein suggest that j34 folds after Ltv1 release and prior to l31, and that l31 folds upon Rio2 release, immediately prior to or with the formation of 80S-like ribosomes.

Rio2's disordered loop keeps l31 unfolded.

Above, we have shown that Rio2 release leads to folding of l31, consistent with its location adjacent to l31. Intriguingly, an 18 amino acid disordered loop in Rio2 (R129-N146) is positioned to interact with l31 **(Figure S2A)**. We thus wondered if this protein loop inserts itself into l31 to prevent it from adopting its mature tertiary structure. To answer this question, we deleted this loop (Rio2_Δloop[24]) and used DMS-MaPseq to probe its effects on folding of l31.

As expected if the Rio2 loop disrupts the l31 structure, l31 is overall less accessible to DMS in pre-40S subunits from Rio2_Δloop cells relative to WT pre-40S, and more similar to l31 from mature 40S **(Figure 2A)**. Of the five residues significantly protected in mature 40S, C1195 and C1197 are already protected in Rio2_Δloop, while C1192 is trending that way, supporting the notion that j31 prematurely foldS in Rio2_Δloop cells. Notably, A1189 and C1190 are becoming overprotected, while A1193 becomes more exposed, indicating some misfolding of this loop. The DMS-MaPseq data also indicate that the h28 and h44 regions start to fold in pre-40S from the Rio2_Δloop cells **(Figure S2B&C)**. This could arise if Rio2, which directly contacts h28 and h44, is mispositioned in this mutant or if folding of l31 affects the folding of h28 and h44 via an allosteric effect.

U1191 within l31 is the target of a cancer-associated acp-modification[25,26], which occurs late in 40S maturation, around the time when 80S-like ribosomes form[27]. l31 is folded in 80S-like ribosomes[18], suggesting that acp-modification occurs after l31 folding, and could thus be an indirect read-out of l31 structure. Thus, we used the previous observation that acp-modification leads to misreading during reverse transcription[25-27] to assess the level of acp-modification in these 40S subunits from the sequencing data of the untreated control cells. Indeed, we observed increased acp-modification of U1191 in Rio2_Δloop pre-40S **(Figure 2B)**, further supporting the notion that l31 is prematurely folded in these intermediates.

Finally, because all studied mutants that affect l31 and j34 impair the formation of 80S-like ribosomes[21], we analyzed the effects from Rio2 loop deletion on the formation of 80S-like ribosomes using a previously described *in vivo* assay[21,28-30]. This assay relies on the accumulation of 80S-like ribosomes, containing pre-18S rRNA (20S rRNA) and 25S rRNA in cells depleted of the ATPase Fap7, as long as 80S-like ribosomes can form [21,28-30]. Indeed, loop deletion shifts the pre-18S rRNA (20S rRNA) from 80S-like ribosomes to 40S ribosomes, under conditions that normally accumulate 80S-like ribosomes (depletion of the ATPase Fap7), indicating a deficiency in 80S-like ribosome formation **(Figure 2C)**. Next, we used Western analysis to test whether the accumulating pre-40S ribosomes retain Ltv1, Enp1 and Rio2. This analysis indicates that release of Ltv1, Enp1 and Rio2 is impaired **(Figure 2D)**, resembling the effects observed from mutation of this loop (Rio2_loop), mutation of Rio2_K105, or the nearby Tsr1 (Tsr1_RK), which also contacts h31[21] **(Figure S2A)**. Importantly, our previous analyses had demonstrated that Ltv1 release at this step requires rearrangements of h31[21]. It appears that premature folding of l31 restricts these rearrangements, impairing Ltv1 release.

Most strikingly, we also observe that the nascent subunits that escape the block to form 80S-like ribosomes retain Rio2, as indicated by the accumulation of Rio2 in the 80S-like fraction **(Figure 2D,** red box**)**. Rio2 is not normally a part of 80S-like ribosomes[18,28]. Moreover, in wild type cells, l31 is only folded once 80S-like ribosomes are formed, i.e. upon Rio2 dissociation[18]. The finding that Rio2 can be retained in 80S-like ribosomes, when its loop is deleted, is consistent with our hypothesis that the loop blocks premature folding of l31, and suggests that the Rio2_Δloop allows the premature folding of l31, such that 80S-like ribosomes retain Rio2. Together, these DMS probing and biochemical analyses provide strong evidence that Rio2's disordered loop keeps l31 unfolded.

Premature folding of l31 leads to misfolding of j34.

We next asked why it is important for Rio2 to keep l31 unfolded until j34 folding is completed. Given that j34 forms a tertiary interaction with l31 in mature 40S **(Figure 1B)**, we hypothesized that premature formation of l31 affects folding of the complex three-way

junction **(Figure 3A)**. To test this hypothesis, we first scrutinized the DMS-MaPseq data of the j34 region from Rio2_Δloop pre-40S **(Figure 3B)**. C1274, A1287 and A1331 (marked with a black asterisk) changed significantly towards mature 40S, although neither reached the level of protection observed in mature 40S. In contrast, A1275, C1284, A1325 (marked with a cyan asterisk) showed changes in Rio2_Δloop pre-40S that are opposite to the changes during maturation, indicating that the changes in j34 are due to misfolding instead of native folding of j34. Furthermore, A1400 and A1401 in h40 do not become protected in Rio2_Δloop intermediates, indicating that the interaction between helices 37 and 40 is not formed, as expected if j34 was misfolded instead of in its native structure **(Figure S1G)**. Lastly, misfolding (instead of natively folding) of j34 in the Rio2_Δloop intermediates is also consistent with impaired Ltv1 release in those cells (**Figure 2D**). Release of Ltv1 leads to folding of j34 [21]. Thus, premature but not native folding of j34 would be expected to promote - not block - Ltv1 release.

To further confirm that j34 is prone to misfolding in Rio2_Δloop cells, we took advantage of previously-described mutations, which destabilize j34 (S20_DE/S20_EYER with Ltv1 bypass[21]) and lead to misfolding of j34 (ΔLtv1[31]), and could therefore be used as molecular probes for j34 folding **(Figure 3C)**. If the premature folding of l31 leads to misfolding of j34 in Rio2_Δloop cells, then we expect these mutations to amplify each other's effects and render the cells synthetically sick. Indeed, quantitative growth measurements demonstrate that the S20_DE/S20_EYER mutations **(Figure 3D)** or Ltv1 deletion **(Figure 3E)** have substantially larger effects in Rio2_Δloop cells than in wild type cells. These data demonstrate the importance of keeping l31 unfolded via the insertion of the Rio2 disordered loop to ensure proper folding of the adjacent j34.

To further test the model that l31 is kept unfolded to protect j34, we used other mutants (Tsr1_RK, Rio2_K105E, Rio2_loop) that impair 40S maturation at the same step as Rio2_Δloop[21] **(Figure 3C)**, and surround l31 **(Figure S2A)**. We combined these mutations with the molecular probes of j34 misfolding described above. As expected, Tsr1_RK, Rio2_K105E, Rio2_loop also exacerbate mutations that destabilize j34 (S20_DE/S20_EYER, **Figure S3A**) or promote j34 misfolding (ΔLtv1, **Figure S3B**),

producing synthetic growth defects. Moreover, these mutants are also rescued by high temperature, which generally alleviates defects from misfolding **(Figure S3C)**. Furthermore, the growth defects of Rio2_K105E and Rio2_loop are exacerbated by low temperature **(Figure S3D)**. Together, our data indicate that Rio2 and Tsr1 cooperate to keep l31 unfolded and that pre-mature folding of l31 leads to misfolding of j34.

l31 is kept unfolded by multiple factors during 40S assembly.

Intriguingly, during 40S subunit maturation l31 is bound by a succession of AFs **(Figure 4A)**. Early during transcription, the snoRNA snR35 binds l31 and specifies the conversion of U1191 into pseudouridine[32]. In a second nucleolar step, the methyltransferase Emg1 (also known as Nep1) modifies the pseudouridine to N1-methylpseudouridine. While this modification is not necessary, Emg1/Nep1 is an essential protein[33]. Deletion of the nucleolar snR35 leads to accumulation of a late cytoplasmic 18S RNA precursor, 20S rRNA[32] **(Figure S4A)**. We therefore wondered if snR35 and Emg1 had a previously unappreciated role in protecting l31 from premature folding early during ribosome assembly. If this is true, then we expect deletion of these factors to block formation of 80S-like ribosomes late in assembly, when their misfolding would be detected in the previously described quality control pathway[21]. To test this prediction we used the *in vivo* assay for formation of 80S-like ribosomes described above[21,28-30]. Indeed, ΔsnR35 impairs the formation of 80S-like ribosomes **(Figure 4B)**. Deletion of Emg1 is not viable, but can be rescued by co-deletion of the nucleolar Nop6[34]. We therefore combined Gal:Emg1 with its suppressor ΔNop6 to test for formation of 80S-like ribosomes. In addition, we used a fractionated cytoplasmic extract **(Figure S4B)**, to ensure that we assay only cytoplastic intermediates and not nucleolar ones, which would be unable to form 80S-like ribosomes as they have no access to 60S subunits. Relative to ΔNop6 cells, depletion of Emg1 in the ΔNop6 background also blocks the formation of 80S-like ribosomes **(Figure 4C).** Furthermore, deletion of snR35, is epistatic to other mutations that lead to premature folding of l31, including Rio2_Δloop, Rio2_loop, as expected if each of these mutations had the same effect: to lead to premature folding of l31 (**Figure S4C**).

These data demonstrate that protection of l31 during the nucleolar stages of assembly is important for the cytoplasmic maturation of 40S ribosomes, presumably to ensure folding of j34. To test this more directly, we tested the genetic interactions between ΔsnR35 and mutants that destabilize j34. As expected, ΔsnR35 is sensitive to all of the j34 destabilizing mutants **(Figure 4D-F)**, an effect that is specific to snR35 deletion and not observed upon deletion of the nearby snR57 **(Figure S4D**). Together, these data demonstrate that snR35, Emg1, Tsr1 and Rio2 function successively to prevent premature folding of loop 31 thereby enabling proper folding of j34.

Discussion

Ordered folding of j34 and l31 is enabled by AFs and promotes native structure

The data herein provide strong evidence for roles of the snoRNA snR35, and the ribosome assembly factors (AFs) Emg1, Rio2 and Tsr1 in keeping the loop of helix 31 (l31) unfolded throughout 40S assembly. This allows for folding of the junction between helices 34, 35 and 38 (j34) prior to l31 (**Figure 1A, 5A).** Thus, these assembly factors reorder the folding landscape: instead of l31 folding prior to j34 (as is the case without these factors), they delay l31 folding to enable prior folding of j34. Furthermore, the data show that when l31 is allowed to fold before j34, by deletion of snR35, or Emg1, or the loop in Rio2 that maintains the unfolded state of l31, j34 is trapped in a misfolded form (**Figure 5B**). Misfolded j34 is identified in subsequent quality control steps that block further maturation[21]. The interplay between j34 and l31 is explained by their direct interaction[23] (**Figure 1B**), as well as the previous observation that tertiary interactions in RNA structure can stabilize not just natively folded but also misfolded intermediates[5]. Thus, by binding to and stabilizing it, l31 traps misfolded j34, thus blocking its refolding.

J34 is prone to misfolding

Misfolding of j34 appears to reflect an intrinsic difficulty of this junction to fold into its native structure, as indicated by *in vitro* folding studies of bacterial 16S rRNA after addition of ribosomal proteins (r-proteins)[6-8]. Both l31 and j34 are highly conserved and present in largely unchanged form in 16S and 18S rRNAs. Similarly, their bound proteins (uS13/Rps18, uS19/Rps15 for l31 and uS3/Rps3, uS10/Rps20, uS14/Rps29 for j34) are

also universally conserved. Reconstitution of bacterial small ribosomal subunits occurs in three stages[9]. An initial assembly step (stage I), where primary and secondary proteins bind to form a misfolded intermediate, termed RI, a heat-dependent refolding step to form RI* (stage II) and a final step (stage III), in which uS10, uS14, uS3, uS2, uS13, and bS21 bind. Probing the differential accessibility of rRNA residues to DMS and kethoxal in each stage has shown that j34 and l31 adopt structure early during assembly[8]. However, the data also suggest that these are misfolded structures. l31 refolds during the heat-dependent folding step, which then allows j34 to refold upon binding of its interacting proteins uS3, uS10 and uS14 in stage III. Notably, l31 also rearranges in stage III. Moreover, probing backbone accessibility during folding leads to similar conclusions and demonstrates a slow second phase of protection for the nucleotides in l31 and j34, with the latter folding even more slowly than the former[6,35]. Thus, the requirement for unfolding of l31 to refold j34 holds in bacteria and reflects a propensity of the junction to misfold.

RimM may chaperone j34 by binding l31 during bacterial assembly

Previous work has also provided evidence for a role of RimM in chaperoning the correct folding of j34, analogous to the roles of Rio2, Emg1 and snR35. When grown in the cold, bacteria lacking RimM specifically accumulate assembly intermediates with misfolded j34[36]. Furthermore, these assembly intermediates are depleted of the j34-interacting uS10/uS14 (the homologs of Rps20 and Rps29)[36],[37]. Additional genetic, structural and biochemical data strongly suggest that RimM indeed chaperones j34 in bacteria by preventing the premature folding of l31 (instead of binding directly to j34). First, RimM binds directly to uS19/Rps15, and docking of RimM onto pre-30S based on this interaction places it in nearly the same position as Rio2, over l31[37,38]. This position is confirmed at very high thresholding of a low-resolution EM-map[37]. Moreover, mutation or depletion of RimM can be rescued by point mutations in uS13/Rps18 and uS19/Rps15 that map adjacent to l31[39]. Finally, rRNA mutations that destabilize h31, or in l31 also suppress the absence or mutation of RimM[38]. Together, these data indicate strongly that RimM, like Rio2 (and snR35, and Emg1) blocks the premature folding of l31, thereby chaperoning j34 by allowing its refolding instead of trapping it in a misfolded state stabilized by tertiary interactions with l31.

The folding of many three-way junctions is choreographed by AFs

In addition to j34, there are 7 other three-way junctions, including j23 (between h22, 23 and 23a) and j26 (between h25, 26 and 26a), as well as j7, a four-way junction between helices 6a, 7, 11 and 12 **Figure S5A**). Analysis of the interactions that these junctions form in mature 40S subunits (**Figure S5B-C)**, combined with their order of formation during maturation (**Figure S5E)** suggests other instances of AF-mediated reordering of the rRNA folding landscape:

In mature 40S, j23 is stabilized by j26 (**Figure S5B)**. Notably, analysis of the structures of pre-40S assembly intermediates[11-13,15-17,19,20,40] indicates that j23 is formed in the earliest 40S structures that are currently available, while j26 is formed only in the subsequent structures (**Figure S5E**). Analysis of *in vitro* bacterial assembly suggests that j23 and j26 fold in stage I[7,8] (although there is evidence for j26 misfolding), and kinetic analysis suggests that j26 acquires structure before j23[6,35]. This difference in the order of folding of j23 and j26 is consistent with AF-mediated delay in folding of j26 *in vivo.* With the exception of Rps22 and Rps23, many ribosomal proteins bound to j26 are later-binding, while j23 binds the early-binding Rps13, Rps14 and Rps1. Moreover, Rps23 is not yet positioned in its final location in early 40S assembly intermediates, instead held away by interactions with Bms1 and Nop14[11,13,20]. These interactions, and the resulting lack of stabilization by Rps23 might contribute to the delay in folding of j26.

Intriguingly, two lines of evidence suggest that *in vivo* h26 folding is delayed by (an unknown) *bacterial* assembly factor as well: First, *in vivo* h26 is accessible to hydroxylradicals in two distinct assembly intermediates, indicating that this is a feature of late bacterial assembly intermediates[36]. Moreover, uS2/Rps0, which binds h26 is a late-binding protein *in vivo*[41]. Thus, analysis of previous studies of 16S and 18S RNA folding *in vitro* and *in vivo* suggests that the order of j23 and j26 folding is reversed by AFs *in vivo*, as described herein for j34 and l31.

Similarly, in mature subunits j7 is stabilized by h44 **(Figure S5C).** Notably, h44 is misoriented in the earliest 40S precursors held via interactions with Utp12 and Utp13, thus enabling the folding of j7 without h44 (which could stabilize misfolded intermediates).

Folding of j43 (between h28, 29, and 43, **Figure S5D**) allows for the placement of h43 between h41, h42, and h30. Thus, to correct j43 misfolding and the resulting misplacement of h43 at a later time, these helices would have to open back up. Evidence for this can be found in the heat-dependent RI→RI* transition in bacterial 16S rRNA[8]. Footprinting data indicate a few early protections in h41, h42 and j43, and then demonstrate dozens of changes within h41, h42, h30, as well as j43 during the heat step. Notably, these include enhancements in the tip of h41, that strongly support unfolding of the structure. Lastly, h41 and h42 demonstrate a small number of additional changes in stage 3, perhaps due to re-closing[7,8]. Together, these data support misfolding of j43 in *in vitro*. These data also provide evidence for early folding of the adjacent j34[7,8], which might help trap misfolded j43.

In contrast, analysis of the (eukaryotic) *in vivo* assembly intermediates indicates that j43 folds before j39 and j34, which stabilize j43 in mature 40S subunits (**Figure S5D, E**). How could folding of j39 and j34 be delayed in eukaryotic cells? The structures[10,11,13,16,19,20] indicate that formation of j39 is disrupted by the AFs Imp4 and Utp1. Moreover, j39 and j34 are stabilized by late-binding ribosomal proteins[23,42](Asc1, Rps17, Rps3, Rps20, and Rps29; note that j39 also binds Rps16, which is early-binding). The binding by Rps16, and eukaryote-specific Rps17, which occurs before Rps3, Rps20 and Rps29 might account for the folding of j39 before j34. Thus, random folding of j43, j39 and j34 *in vitro* appears to lead to misfolding, while AFs may help delay folding of j39 and j34, and thus allow for proper folding of j43.

The striking structures of early pre-90S assembly intermediates indicate a role for the AFs in separating the subdomains from each other[11,13,20]. It was speculated that this could ensure that functionality such as ligand binding can be adopted only in later states[43]. Moreover, it was implied that this could also ensure folding of the subdomains[10,11]. How

and why this would help, however, was not discussed. Together, the data herein, as well as the above reanalysis of previous studies indicates a general role for AFs in re-ordering the folding landscape to prevent the premature adoption of tertiary contacts. In this manner, they avoid kinetic traps by modulating the RNA folding landscape. The need for such separate and ordered assembly is presumably exacerbated, when multiple contacts between subdomains can form.

A role for snoRNAs in modulating RNA folding.

The data herein show that one role of snR35 (in addition to the pseudouridylation of U1191) is to keep l31 unfolded. Interestingly, j4, j7, j33, j34, j39, and j43 are all bound by snoRNAs, while j23 and j26 are not[44]. Thus, six out of eight of these junctions could be chaperoned by snoRNAs. Of the remaining two, j23 is folding early. Globally, most junctions include snoRNA binding sites, and moreover, 19/33 snoRNAs bind at helical junctions indicating that chaperoning their folding could be a previously underappreciated general role for snoRNAs. Interestingly, because the misfolding of j34 caused by the absence of the nucleolar snR35 or Emg1 is not detected by quality control mechanisms until the latest cytoplasmic stages, this also illustrates how cytoplasmic assembly defects can be caused by nuclear events. This is important, because it provides a note of caution when interpreting the timing of assembly defects.

Materials and Methods

Plasmids and yeast strains.

Yeast strains **(Table S1)** were generated by PCR-mediated homologous recombination[45], and confirmed by PCR, serial dilution and western blotting if antibodies were available. snoRNA deletion strains were also confirmed via Northern blotting. Plasmids **(Table S2)** were constructed using standard cloning techniques and confirmed via sequencing.

Growth curve measurements.

Cells were grown in YPD or glucose minimal media (if the supplemented plasmid is not essential) overnight, and then diluted into fresh YPD for 3-6 hours before inoculating into 96-well plates (Thermo Scientific) at a starting OD600 between 0.04 to 0.1. A Synergy.2 plate reader (BioTek) was used to record OD600 for 48 hours, while shaking at 30 °C,

unless otherwise specified (37°C or 26.5°C). Doubling times were calculated using data points within the mid-log phase. Data were averaged from at least 6 biological replicates of 3 different colonies and 2 independent measurements. Statistical analyses for each measurement are detailed in the respective figure legend.

DMS MaPseq sample preparation.

Pre-40S intermediates were purified using Rio2-TAP via IgG and calmodulin beads as previously described[46]. Mature 40S were purified as described[47]. The purified ribosomes were treated at 30°C for 5min with 1%DMS (Sigma-Aldrich) in the presence of 80mM HEPES pH7.4, 50mM NaCl, 5mM $Mg(OAc)_2$, 0.2uM RNaseP RNA. DMS reactions were stopped by addition of 0.4-vol buffer (1M β-ME, 1.5M NaOAc, pH 5.2) and purified using phenol chloroform precipitation.

DMS MaPseq RNA library preparation.

Generation of the RNA sequencing library was adapted from a previous protocol[48]. RNAs were fragmented with 20mM $MgCl_2$ at 94°C for 10min, before separation on a denaturing 15% gel. RNA fragments between 50-80nt were cut out and eluted overnight in RNA gel extraction buffer (0.3M NaOAc, pH5.5, 1mM EDTA, 0.25% v/v SDS). RNA fragments were dephosphorylated using T4 PNK (NEB), ligated to the adaptor using T4 RNA ligase II truncated (NEB) and gel purified. Reverse transcription was carried out with TGIRT III (InGex) in 50mM Tris HCl, pH 8.3, 75mM KCl, 3mM MgCl2, 1mM dNTPs, 5mM DTT, 10U SUPERase·In (Invitrogen), for 1.5hrs at 60°C. RNAs were hydrolyzed by addition of 250mM NaOH and the cDNAs were circularized using CircLigase II (Lucigen). The circularized cDNAs were sent to Genomics Core at Scripps Florida for deep sequencing using either single-end or pair-end Illumina NextSeq 500 platform.

DMS MaPseq data processing.

Processing of the DMS MaPseq data followed the pipeline from GitHub code (https://github.com/borisz264/mod_seq). Briefly, raw sequencing reads were first trimmed of their adaptor sequences via Skewer[49] (Adaptor sequences: AGATCGGAAGAGCACACGTCTGAACTCCAGTCA for single-end sequencing reads and R1: AGATCGGAAGAGCACACGTCTGAACTCCAGTCA, R2: AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT for pair-end sequencing reads.) Shapemapper 2.0[50] was used to trim the first and last 5nt of each read as well as low-

quality nucleotides, align reads to *S. cerevisiae* 20S rRNA (Saccharomyces Genome Database), and count mutations and read coverage at each rRNA position. Raw sequencing data are deposited in **xxx**.

DMS MaPseq data analysis:

The first 5 and last 5nt of 20S rRNA as well as nucleotides 1766-1785 were excluded because of their low read depth. Native 18S rRNA modifications sites were analyzed separately. We first separately averaged the mutational rates of A, C, G and U in each sample (DMS-treated and untreated, and each mutant). This analysis confirmed that DMS only increases the mutational rate of A and C, as expected because only DMS-modifications at A and C lead to mutations during reverse transcription. All subsequent analyses were thus done only with A and C. To calculate the DMS accessibility for each nucleotide, we divided that nucleotide's mutational rate by the average mutational rate of all untreated A (or C) nucleotides, before subtracting the value for the untreated sample from the DMS-treated sample.

$$DMS\ accessibility_i = \frac{m_{i,+DMS}}{\text{average (A or C)}_{-DMS}} - \frac{m_{i,-DMS}}{\text{average (A or C)}_{-DMS}}$$

Where m_i is the mutational rate of nucleotide i, either with (+DMS) or without (-DMS) DMS. At least three biological replicates were collected for each sample. Raw mutational rates are summarized in **Sheet S1**.

In vivo subunit joining assay.

Cells were grown to mid-log phase in galactose before inoculating into YPD for at least 16 hours to deplete Fap7. Formation of 80S-like ribosomes was assayed by sucrose gradient fractionation and Northern blotting as previously described[21,28-30], and the fraction of 20S rRNA in 80S-like ribosomes quantified. Dissociation of assembly factors was assayed by Western blotting, and the fraction of each AF in pre-40S ribosomes quantified. Cellular fractionation for ΔNop6,Gal:Emg1 cells was carried out exactly as previously described[29]. At least two biological replicates were assayed for each mutant.

Quantification and statistical analysis.

Western blots were imaged using the ChemiDoc MP Imaging System from Biorad after applying luminescence substrates (Invitrogen). Northern blots were visualized by

Typhoon™ FLA 7000(GE Healthcare). Image Lab (Biorad) and Quantity One (Biorad) were used to analyze Western blots and Northern blots, respectively.

Statistical analyses were performed using GraphPad Prism version 8.3 (GraphPad Software, San Diego, California). Statistical tests were used as indicated in the respective figure legends.

Oligos used in this study are summarized in **Table S3**.

Details of cell types and plasmids used in each experiment are listed in **Table S4**.

Supplementary Materials comprising 5 Supplemental Figures, 3 Supplemental Tables and 1 supplemental sheet are available online.

Acknowledgements

We thank Dr. Karunadharma from the Scripps Florida Genomics Core for library preparation and sequencing, Z. Zhao for technical support with the running and debugging of Github codes and members of the Karbstein laboratory for discussion and comments on the manuscript.

This work was supported by NIH grants R01-GM117093, R01-GM086451, R35-GM136323, and HHMI Faculty Scholar grant 55108536 to K.K., and a Farris Foundation fellowship to H.H..

The authors declare no competing financial interests.

Author contributions

H. Huang performed all experiments. H. Huang and K. Karbstein designed the experiments and wrote the paper.

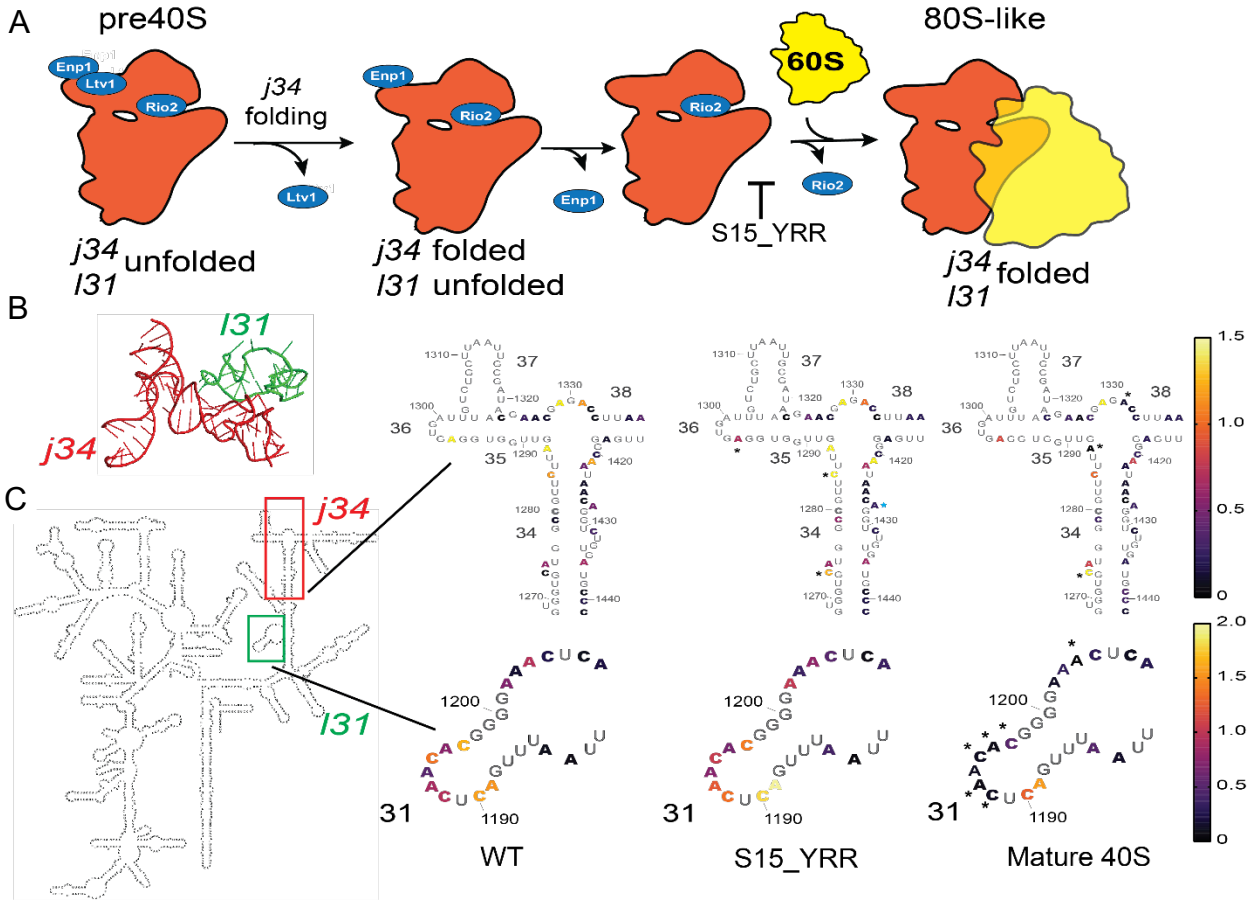

Figure 1. In WT yeast j34 folds before l31

A. Schematic release order of Ltv1, Enp1 and Rio2 prior to formation of 80S-like ribosomes. The folding status of j34 and l31 is indicated under each intermediate.

B. L31 (green) and j34 (red) interact in mature 40S ribosomes (extracted from PDB ID 3J77).

C. (Left) The locations of j34 and l31 on the secondary structure of *Saccharomyces cerevisiae* 18S (from RiboVision) are highlighted with red and green boxes, respectively. (Right) DMS accessibility of nucleotides within the j34 and l31 regions of 18S rRNA from WT pre40S, S15_YRR pre40S and mature 40S. The color of each nucleotide corresponds to its DMS accessibility as indicated in the scale bar. Changes that are statistically significant (p_{adj}<0.05) using a two-way ANOVA analysis are marked with a black asterisk. Changes that that do not rise to this statistical significance are marked with a cyan asterisk. n≥3.

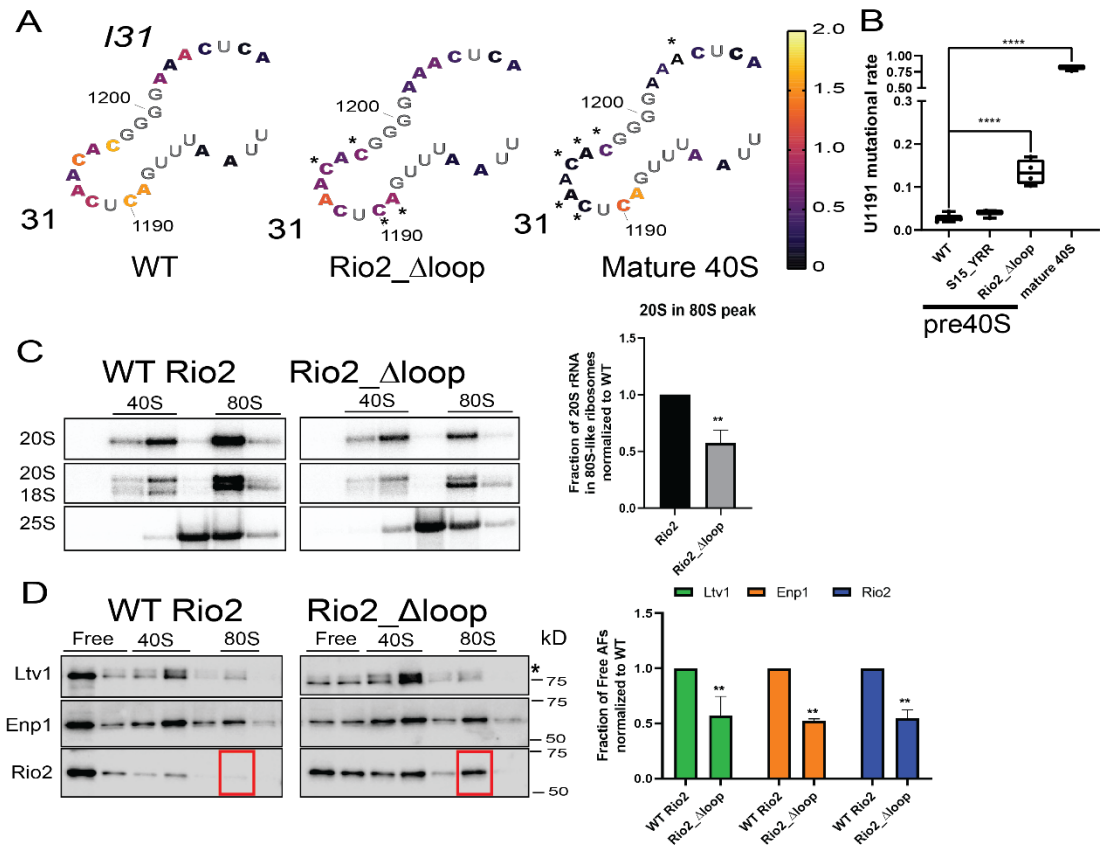

Figure 2. Rio2's disordered loop keeps I31 unfolded.

A. DMS accessibility of nucleotides within I31 of 18S rRNA from WT pre40S, Rio2_Δloop pre40S and mature 40S. The color of each nucleotide corresponds to its DMS accessibility as indicated in the scale bar. Note that the first (WT pre-40S) and third panel (mature 40S) are identical to those in Figure 1B, and shown here for direct comparison. Changes that are statistically significant (p_{adj}<0.05) using a two-way ANOVA analysis are marked with a black asterisk. n≥3.

B. Mutational rate at U1191 from untreated WT pre40S, S15_YRR pre40S, Rio2_Δloop pre40S and mature 40S ribosomes. Significance was tested using a one-way ANOVA test. ****, p<0.0001. n≥3.

C. Northern blots of 10-50% sucrose gradients from Fap7-depleted cells expressing plasmid-encoded wild type (WT) Rio2 or Rio2_Δloop proteins. Fap7 was depleted in glucose for over 16h to accumulate 80S-like ribosomes. The positions where 40S and 80S ribosomes sediment are indicated. (Right) Quantifications of the data on the left indicate the efficiency of formation of 80S-like ribosomes. Specifically, 20S rRNA in 80S fractions are first normalized to total 20S. This fraction of 20S in 80S-like ribosomes in Rio2_Δloop cells is further normalized to that in WT Rio2. Data are shown as mean with standard deviation. Significance was tested using an unpaired test. **, p<0.01. n=3.

D. Western blots for Ltv1, Enp1 and Rio2 from the gradients in panel C. Quantifications are shown on the right. AFs in free fractions are first normalized to their total signal before normalizing the Rio2_Δloop sample to WT Rio2. Data are shown as mean with standard deviation. Significance was tested using a two-way ANOVA test. **, p<0.01. n=2.

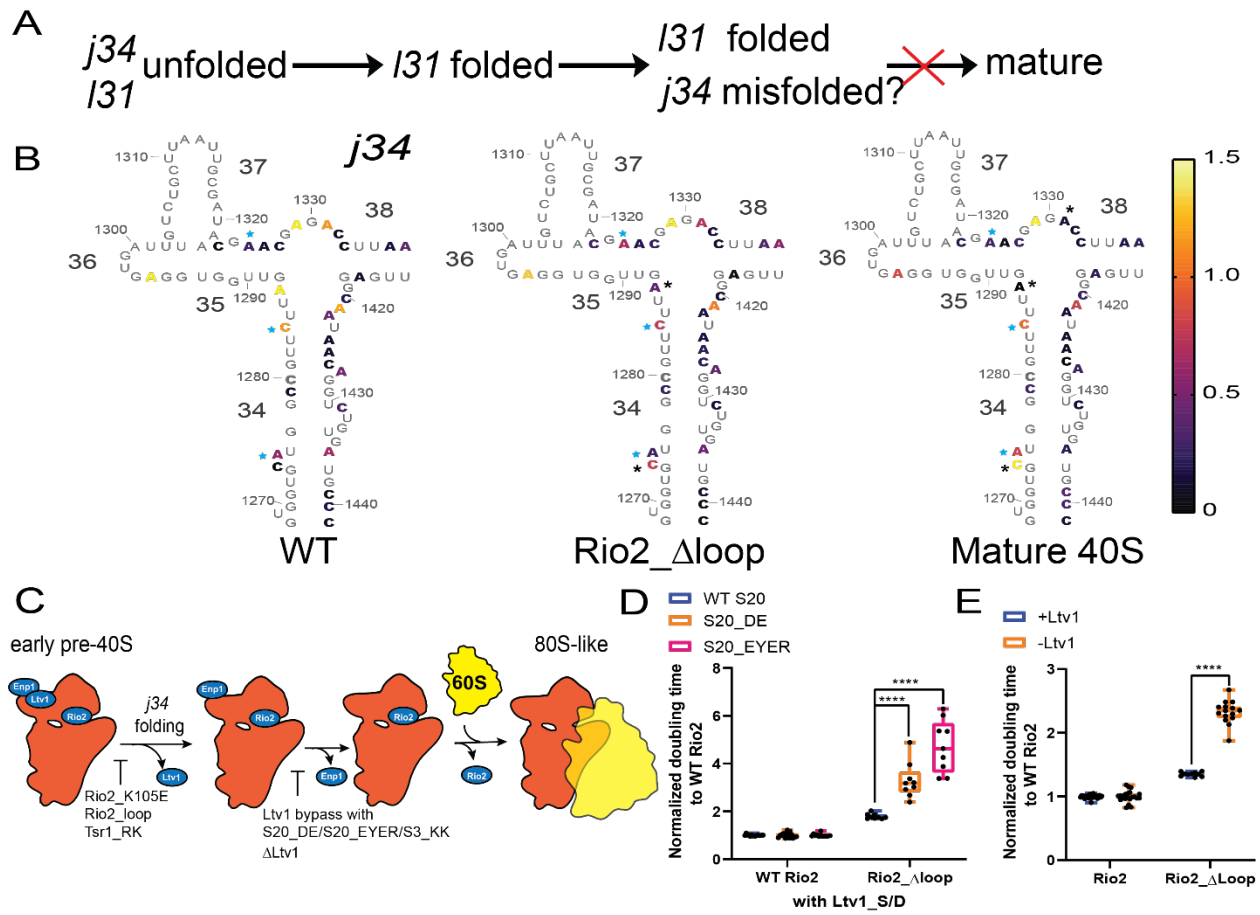

Figure 3. Premature folding of l31 leads to misfolding of j34.

A. Model that l31 pre-mature folding traps misfolded j34 and blocks 40S maturation.
B. The color of each nucleotide corresponds to its DMS accessibility as indicated in the scale bar. Note that the first (for WT pre40S) and third panel (for mature 40S) are identical to those in Figure 1B. Changes that are statistically significant using a two-way ANOVA analysis comparing nucleotides of the entire 18S region to WT pre-40S are marked with black asterisk. *, P<0.05. Changes that are opposite in Rio2_Δloop and mature 40S are marked with a cyan asterisk. n⩾3.
C. Schematic release order of Ltv1, Enp1 and Rio2 prior to formation of 80S-like ribosomes. The step impaired by each mutation is indicated.
D. Normalized doubling times of ΔLtv1,Gal:Rio2,Gal:S20 cells supplied with plasmids encoding Ltv1_S/D and WT Rio2 or Rio2_Δloop and WT S20 or S20_DE or S20_EYER. Significance was tested using a two-way ANOVA test. ****, p<0.0001. n≥9.
E. Normalized doubling times of Gal:Rio2 or ΔLtv1,Gal:Rio2 cells supplied with plasmids encoding WT Rio2 or Rio2_Δloop. Significance was tested using a two-way ANOVA test. ****, p<0.0001. n≥10.

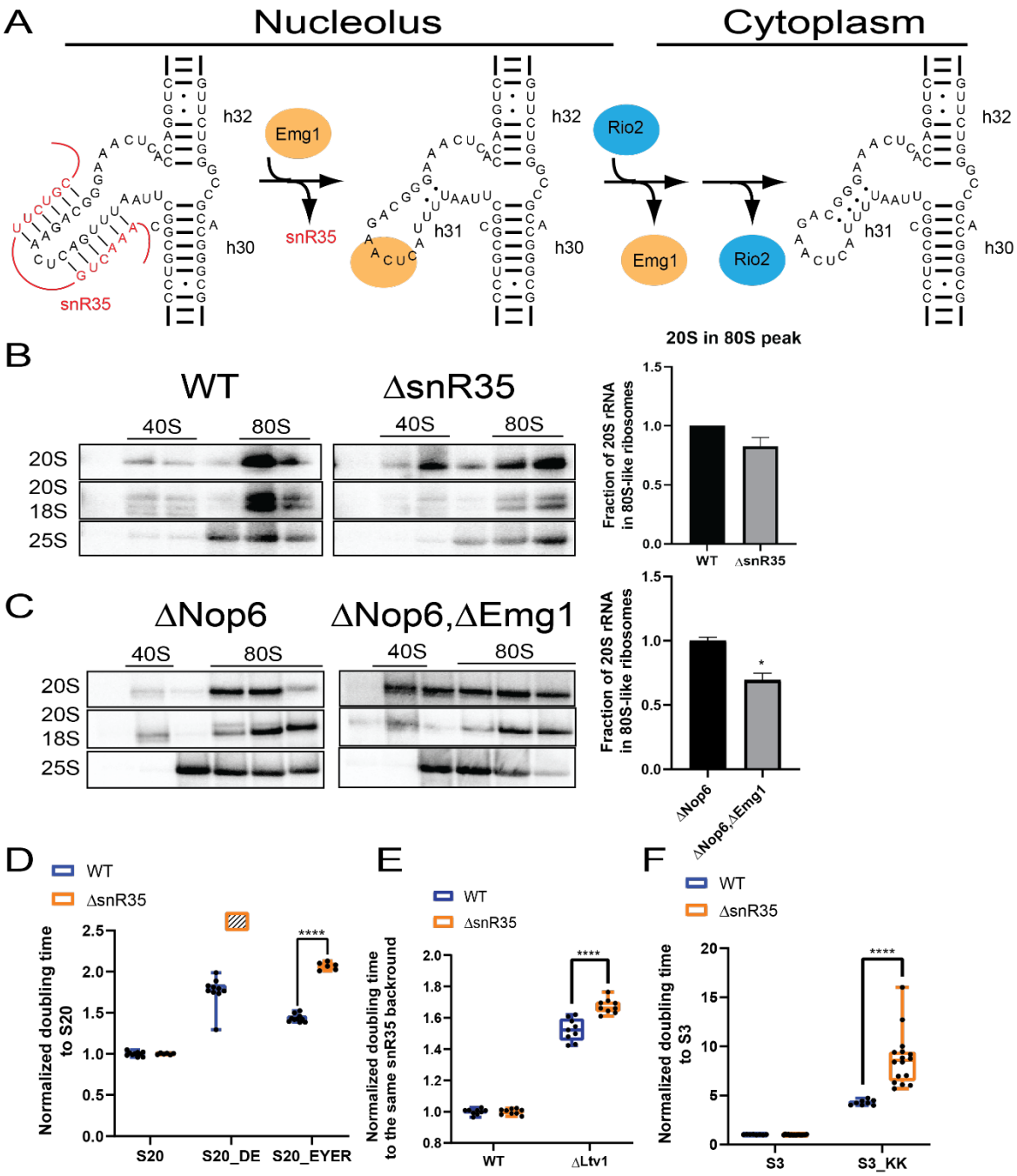
A
Nucleolus
Cytoplasm
h32
h30
h31
Emg1
Rio2
snR35
B
WT
ΔsnR35
40S
80S
20S
18S
25S
20S in 80S peak
Fraction of 20S rRNA in 80S-like ribosomes
C
ΔNop6
ΔNop6,ΔEmg1
D
E
F
Normalized doubling time to S20
S20
S20_DE
S20_EYER
Normalized doubling time to the same snR35 backround
ΔLtv1
Normalized doubling time to S3
S3
S3_KK
****

Figure 4. l31 is kept unfolded by multiple factors during 40S assembly.

A. The order of assembly factors binding to l31 in the nucleolus and cytoplasm.
B. Northern blots of 10-50% sucrose gradients from Fap7 depleted cells with and without snR35. Fap7 was depleted in glucose for over 16h to accumulate 80S-like ribosomes. Quantifications are shown on the right. Data are shown as mean with standard deviation. n=2.
C. Northern blots of 10-50% sucrose gradients from the cytoplasmic fraction of ΔNop6,Gal:Fap7 and ΔNop6,Gal:Emg1,Gal:Fap7 cells. Fap7 was depleted in glucose for over 16hrs. Quantifications are shown on the right. Data are shown as mean with standard deviation. Significance was tested using an unpaired t-test. *, $p<0.05$. n=2.
D. Normalized doubling times of ΔLtv1,Gal:S20 or ΔLtv1,ΔsnR35,Gal:S20 cells supplied with plasmids encoding Ltv1_S/D and WT S20 or S20_DE or S20_EYER. Cells with Ltv1_S/D and S20_DE in ΔsnR35 do not grow, and are therefore shown as a box filled with diagonal lines. Significance was tested using a two-way ANOVA test. ****, $p<0.0001$. $n\geq6$.
E. Normalized doubling times of BY4741, ΔLtv1, ΔsnR35, ΔLtv1ΔsnR35 cells. Significance was tested using a two-way ANOVA test. ****, $p<0.0001$. n=9.
F. Normalized doubling times of ΔLtv1,Gal:S3 or ΔLtv1,ΔsnR35,Gal:S3 cells supplied with plasmids encoding Ltv1_S/D and WT S3 or S3_KK. Significance was tested using a two-way ANOVA test. ****, $p<0.0001$. $n\geq8$.

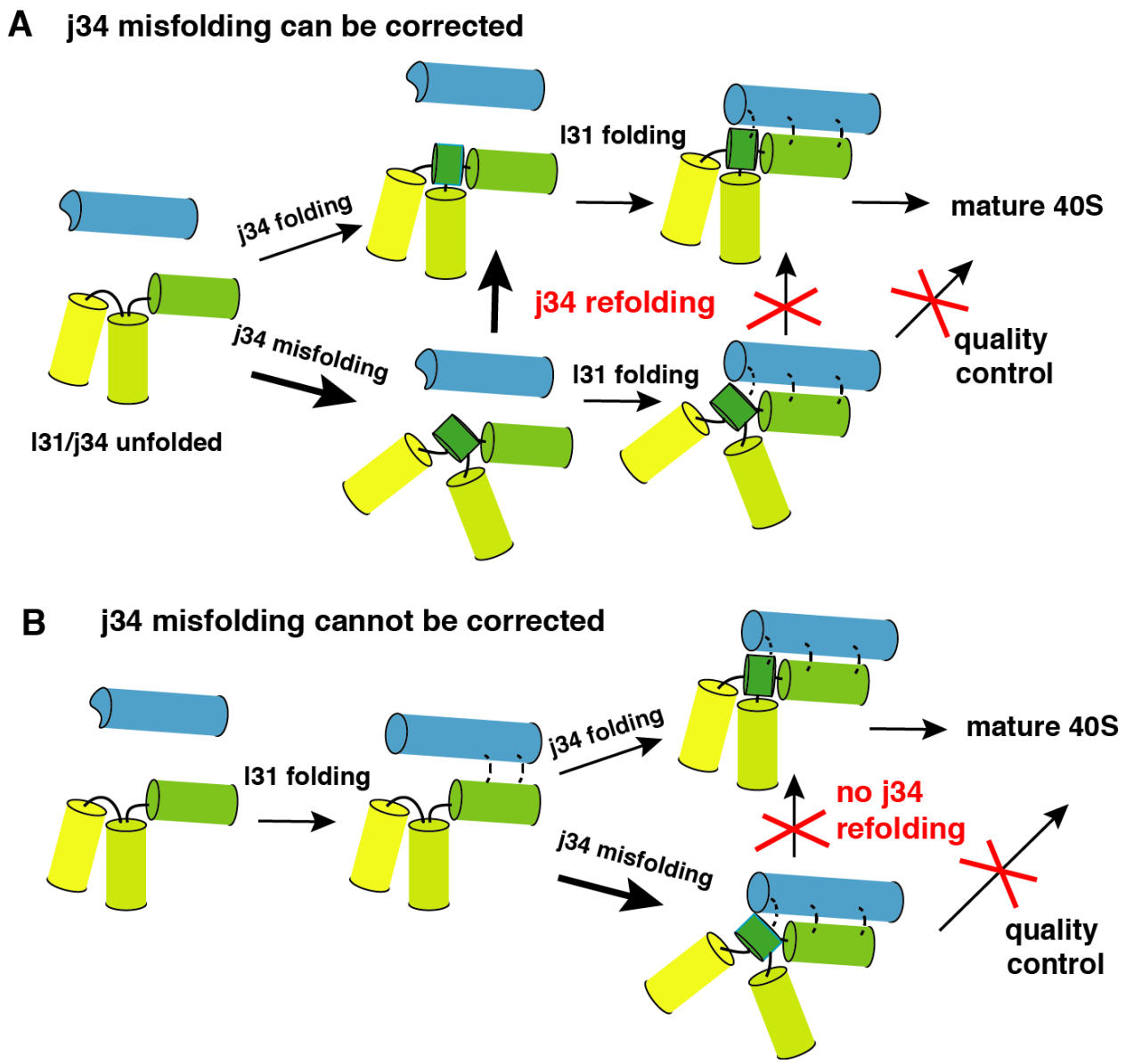

Figure 5. AFs choreograph rRNA folding to prevent misfolding.

A. During its folding j34 (green disk) partitions between folding to its native form (top) or a misfolded intermediate (bottom). This intermediate can refold, until stabilized by l31 (blue), but cannot mature. Thus, folding of j34 before l31 allows for refolding of j34 from its misfolded intermediate to the native junction.
B. When l31 folds before j34 (in the absence of snR35, Emg1, or the Rio2 loop), the misfolded state of j34 is stabilized by l31, thereby trapping it and blocking refolding and further maturation.

References.

1. Herschlag, D., Bonilla, S. & Bisaria, N. The Story of RNA Folding, as Told in Epochs. *Cold Spring Harb Perspect Biol* **10**(2018).
2. Woodson, S.A. RNA folding pathways and the self-assembly of ribosomes. *Acc Chem Res* **44**, 1312-9 (2011).
3. Weeks, K.M. Review toward all RNA structures, concisely. *Biopolymers* **103**, 438-48 (2015).
4. Ganser, L.R., Kelly, M.L., Herschlag, D. & Al-Hashimi, H.M. The roles of structural dynamics in the cellular functions of RNAs. *Nat Rev Mol Cell Biol* **20**, 474-489 (2019).
5. Russell, R., Tijerina, P., Chadee, A.B. & Bhaskaran, H. Deletion of the P5abc peripheral element accelerates early and late folding steps of the Tetrahymena group I ribozyme. *Biochemistry* **46**, 4951-61 (2007).
6. Adilakshmi, T., Bellur, D.L. & Woodson, S.A. Concurrent nucleation of 16S folding and induced fit in 30S ribosome assembly. *Nature* **455**, 1268-72 (2008).
7. Holmes, K.L. & Culver, G.M. Analysis of conformational changes in 16 S rRNA during the course of 30 S subunit assembly. *J Mol Biol* **354**, 340-57 (2005).
8. Holmes, K.L. & Culver, G.M. Mapping structural differences between 30S ribosomal subunit assembly intermediates. *Nat Struct Mol Biol* **11**, 179-86 (2004).
9. Held, W.A., Mizushima, S. & Nomura, M. Reconstitution of Escherichia coli 30 S ribosomal subunits from purified molecular components. *J Biol Chem* **248**, 5720-30 (1973).
10. Cheng, J., Kellner, N., Berninghausen, O., Hurt, E. & Beckmann, R. 3.2-Å-resolution structure of the 90S preribosome before A1 pre-rRNA cleavage. *Nature Structural & Molecular Biology* **24**, 954-964 (2017).
11. Barandun, J. et al. The complete structure of the small-subunit processome. *Nat Struct Mol Biol* **24**, 944-953 (2017).
12. Heuer, A. et al. Cryo-EM structure of a late pre-40S ribosomal subunit from Saccharomyces cerevisiae. *Elife* **6**(2017).
13. Sun, Q. et al. Molecular architecture of the 90S small subunit pre-ribosome. *Elife* **6**(2017).
14. Johnson, M.C., Ghalei, H., Doxtader, K.A., Karbstein, K. & Stroupe, M.E. Structural Heterogeneity in Pre-40S Ribosomes. *Structure* **25**, 329-340 (2017).
15. Scaiola, A. et al. Structure of a eukaryotic cytoplasmic pre-40S ribosomal subunit. *EMBO J* (2018).
16. Cheng, J. et al. 90S pre-ribosome transformation into the primordial 40S subunit. *Science* **369**, 1470-1476 (2020).
17. Du, Y. et al. Cryo-EM structure of 90S small ribosomal subunit precursors in transition states. *Science* **369**, 1477-1481 (2020).
18. Rai, J. et al. An opened pre-40S platform enables temporal regulation of Fap7. *bioRxiv*, 617910 (2020).
19. Cheng, J. et al. Thermophile 90S Pre-ribosome Structures Reveal the Reverse Order of Co-transcriptional 18S rRNA Subdomain Integration. *Mol Cell* **75**, 1256-1269 e7 (2019).
20. Chaker-Margot, M., Barandun, J., Hunziker, M. & Klinge, S. Architecture of the yeast small subunit processome. *Science* **355**(2017).

21. Huang, H., Ghalei, H. & Karbstein, K. Quality control of 40S ribosome head assembly ensures scanning competence. *Journal of Cell Biology* (2020).
22. Zubradt, M. et al. DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat Methods* **14**, 75-82 (2017).
23. Ben-Shem, A. et al. The structure of the eukaryotic ribosome at 3.0 A resolution. *Science* **334**, 1524-9 (2011).
24. Ferreira-Cerca, S. et al. ATPase-dependent role of the atypical kinase Rio2 on the evolving pre-40S ribosomal subunit. *Nature Structural & Molecular Biology* **19**, 1316-+ (2012).
25. Babaian, A. et al. Loss of m(1)acp(3)Psi Ribosomal RNA Modification Is a Major Feature of Cancer. *Cell Rep* **31**, 107611 (2020).
26. Meyer, B. et al. Ribosome biogenesis factor Tsr3 is the aminocarboxypropyl transferase responsible for 18S rRNA hypermodification in yeast and humans. *Nucleic Acids Res* **44**, 4304-16 (2016).
27. Hector, R.D. et al. Snapshots of pre-rRNA structural flexibility reveal eukaryotic 40S assembly dynamics at nucleotide resolution. *Nucleic Acids Res* **42**, 12138-54 (2014).
28. Strunk, B.S., Novak, M.N., Young, C.L. & Karbstein, K. A translation-like cycle is a quality control checkpoint for maturing 40S ribosome subunits. *Cell* **150**, 111-21 (2012).
29. Ghalei, H. et al. Hrr25/CK1delta-directed release of Ltv1 from pre-40S ribosomes is necessary for ribosome assembly and cell growth. *J Cell Biol* **208**, 745-59 (2015).
30. Ghalei, H. et al. The ATPase Fap7 Tests the Ability to Carry Out Translocation-like Conformational Changes and Releases Dim1 during 40S Ribosome Maturation. *Mol Cell* **67**, 990-1000 e3 (2017).
31. Collins, J.C. et al. Ribosome biogenesis factor Ltv1 chaperones the assembly of the small subunit head. *Journal of Cell Biology* **217**, 4141-4154 (2018).
32. Liang, X.H., Liu, Q. & Fournier, M.J. Loss of rRNA modifications in the decoding center of the ribosome impairs translation and strongly delays pre-rRNA processing. *RNA* **15**, 1716-28 (2009).
33. Meyer, B. et al. The Bowen-Conradi syndrome protein Nep1 (Emg1) has a dual role in eukaryotic ribosome biogenesis, as an essential assembly factor and in the methylation of Psi1191 in yeast 18S rRNA. *Nucleic Acids Res* **39**, 1526-37 (2011).
34. Buchhaupt, M., Kotter, P. & Entian, K.D. Mutations in the nucleolar proteins Tma23 and Nop6 suppress the malfunction of the Nep1 protein. *FEMS Yeast Res* **7**, 771-81 (2007).
35. Adilakshmi, T., Lease, R.A. & Woodson, S.A. Hydroxyl radical footprinting in vivo: mapping macromolecular structures with synchrotron radiation. *Nucleic Acids Res* **34**, e64 (2006).
36. Clatterbuck Soper, S.F., Dator, R.P., Limbach, P.A. & Woodson, S.A. In vivo X-ray footprinting of pre-30S ribosomes reveals chaperone-dependent remodeling of late assembly intermediates. *Mol Cell* **52**, 506-16 (2013).
37. Guo, Q. et al. Dissecting the in vivo assembly of the 30S ribosomal subunit reveals the role of RimM and general features of the assembly process. *Nucleic Acids Res* **41**, 2609-20 (2013).

38. Lovgren, J.M. et al. The PRC-barrel domain of the ribosome maturation protein RimM mediates binding to ribosomal protein S19 in the 30S ribosomal subunits. *RNA* **10**, 1798-812 (2004).
39. Bylund, G.O., Persson, B.C., Lundberg, L. & Wikström, P. A novel ribosome-associated protein is important for efficient translation in Escherichia coli. *Journal of bacteriology* **179**, 4567-4574 (1997).
40. Hunziker, M. et al. Conformational switches control early maturation of the eukaryotic small ribosomal subunit. *Elife* **8**(2019).
41. Sashital, D.G. et al. A combined quantitative mass spectrometry and electron microscopy analysis of ribosomal 30S subunit assembly in E. coli. *Elife* **3**(2014).
42. Svidritskiy, E., Brilot, A.F., Koh, C.S., Grigorieff, N. & Korostelev, A.A. Structures of yeast 80S ribosome-tRNA complexes in the rotated and nonrotated conformations. *Structure* **22**, 1210-1218 (2014).
43. Kornprobst, M. et al. Architecture of the 90S Pre-ribosome: A Structural View on the Birth of the Eukaryotic Ribosome. *Cell* **166**, 380-393 (2016).
44. Yang, J. et al. Mapping of Complete Set of Ribose and Base Modifications of Yeast rRNA by RP-HPLC and Mung Bean Nuclease Assay. *PLoS One* **11**, e0168873 (2016).
45. Longtine, M.S. et al. Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast* **14**, 953-61 (1998).
46. Strunk, B.S. et al. Ribosome assembly factors prevent premature translation initiation by 40S assembly intermediates. *Science* **333**, 1449-53 (2011).
47. Acker, M.G., Kolitz, S.E., Mitchell, S.F., Nanda, J.S. & Lorsch, J.R. Reconstitution of yeast translation initiation. *Methods Enzymol* **430**, 111-45 (2007).
48. McGlincy, N.J. & Ingolia, N.T. Transcriptome-wide measurement of translation by ribosome profiling. *Methods* **126**, 112-129 (2017).
49. Jiang, H., Lei, R., Ding, S.W. & Zhu, S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* **15**, 182 (2014).
50. Busan, S. & Weeks, K.M. Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2. *RNA* **24**, 143-148 (2018).