Dissecting the *in vivo* assembly of the 30S ribosomal subunit reveals the role of RimM and general features of the assembly process

Qiang Guo¹, Simon Goto², Yuling Chen¹, Boya Feng¹, Yanji Xu¹, Akira Muto², Hyouta Himeno², Haiteng Deng¹, Jianlin Lei¹,* and Ning Gao¹,*

¹Ministry of Education Key Laboratory of Protein Sciences, Center for Structural Biology, School of Life Sciences, Tsinghua University, Beijing 100084, China and ²Department of Biochemistry and Molecular Biology, Faculty of Agriculture and Life Science, RNA Research Center, Hirosaki University, Hirosaki 036-8561, Japan

Received August 20, 2012; Revised November 2, 2012; Accepted November 3, 2012

**ABSTRACT**

Ribosome biogenesis is a tightly regulated, multi-stepped process. The assembly of ribosomal subunits is a central step of the complex biogenesis process, involving nearly 30 protein factors *in vivo* in bacteria. Although the assembly process has been extensively studied *in vitro* for over 40 years, very limited information is known for the *in vivo* process and specific roles of assembly factors. Such an example is ribosome maturation factor M (RimM), a factor involved in the late-stage assembly of the 30S subunit. Here, we combined quantitative mass spectrometry and cryo-electron microscopy to characterize the *in vivo* 30S assembly intermediates isolated from mutant *Escherichia coli* strains with genes for assembly factors deleted. Our compositional and structural data show that the assembly of the 3'-domain of the 30S subunit is severely delayed in these intermediates, featured with highly underrepresented 3'-domain proteins and large conformational difference compared with the mature 30S subunit. Further analysis indicates that RimM functions not only to promote the assembly of a few 3'-domain proteins but also to stabilize the rRNA tertiary structure. More importantly, this study reveals intriguing similarities and dissimilarities between the *in vitro* and the *in vivo* assembly pathways, suggesting that they are in general similar but with subtle differences.

**INTRODUCTION**

Ribosome biogenesis is a tightly regulated multi-stepped process, assisted by a wide variety of protein factors, such as transcription factors, endoribonucleases, rRNA helicases and chaperones, rRNA and ribosomal protein modification enzymes and assembly factors (1). As to the 30S subunit, early *in vitro* reconstitution experiments (2–6) have demonstrated that active 30S subunits could be formed from purified ribosomal proteins and 16S rRNA in the absence of other cellular components. The *in vitro* assembly occurs very slowly and requires non-physiological conditions, such as high Mg²⁺ concentration, high ion strength and heat shock. In contrast, the assembly of the 30S subunit *in vivo* starts with rRNA primary transcripts (7) and occurs co-transcriptionally (8) in a much more efficient way, underscoring the essential contribution of assembly factors. In recent years, application of new techniques, such as pulse-chase monitored by quantitative mass spectrometry (PC/QMS) (9), time-resolved X-ray footprinting (10) and time-resolved electron microscopy (11), has brought our understanding of the *in vitro* assembly process to a new level, providing a large amount of valuable kinetic and structural information. Together with earlier work [reviewed in (12)], these data have established that the *in vitro* 30S subunit assembly starts from multiple sites on the 16S rRNA (10), following parallel pathways (9–11) and the free energy of the assembly can be represented by a complex landscape (9). More importantly, kinetic data revealed that for several subsets of 3'-domain proteins, the thermodynamic interdependence does not align well with measured kinetic cooperativity (11,13), and at these locations, the *in vitro* assembly often encounters kinetic traps,
suggesting that assembly factors might be involved in subverting kinetic traps in the assembly landscape (9,11,13).

Over the past two decades, accumulating experimental data, mainly through genetic approaches, has implicated a number of factors, including RbfA, RsgA, KsgA, Era, ribosome maturation factor M (RimM), RimP, RimJ [reviewed in (12)] and YqeH (14,15), in the maturation of the 30S subunit in bacteria. However, the specific molecular roles of most of these factors remain unclear. Among these factors, RimM was first identified as a factor required for a fast growth in rich medium (16). The gene-encoding RimM (yfjA) in Escherichia coli is co-localized to the trmD operon (17) with genes for ribosomal proteins S16 and L19, and a tRNA methyltransferase (TrmD), a hint that RimM might be directly involved in ribosome-related function. Indeed, deletion of RimM confers a slow growth phenotype (18), with accumulation of 16S rRNA precursors and free 30S subunits (19) as well as reduced level of polysomes (20). RimM associates with free 30S subunit in vivo (18,20) and also binds to S19 in vitro (20,21). Moreover, suppressor mutations to the rimM mutant were found on S13 (18) and suppressor mutations to a rimM-Y106AY107A mutant were found on S19, helices 31 and 33b of the 16S rRNA (20).

In this study, we characterize the immature 30S subunits purified from an E. coli ΔrimM strain biochemically and structurally. Our data indicate that the immature 30S subunits are a collection of assembly intermediates, with the 3′-head domain proteins severely underrepresented, such as S10, S14, S13 and S19. Moreover, protein composition analysis of another category of immature 30S subunits from a ΔrbfAΔrsgA strain shows a different spectrum, with much enhanced levels for these proteins, suggesting that RimM promotes the assembly of these slow binding proteins in vivo. Structural analysis shows that these ΔrimM intermediates also differ largely in rRNA conformation, particularly the rotational position of the 3′-head domain relative to the body domain. An incubation of recombinant RimM with the immature 30S subunits significantly reduces the flexibility of the head domain. More importantly, our data also suggest that the in vivo assembly process occurs along multiple pathways in a certain degree as well, and the rRNA maturation is tightly coupled with ribosomal protein binding. The functional depiction of RimM thus illustrates that there are possible checkpoints along the in vivo assembly pathways where maturation factors come into play to direct the process to more efficient branches.

MATERIALS AND METHODS

Escherichia coli strains

We used E. coli A19 (Hfr, rna-19, gdhA2, his-95, relA1, spoT1, metB1) (22) as the source of the 30S subunit. A19ΔrimM is an A19 derivative in which the rimM gene is replaced by a short peptide gene containing an FRT sequence, constructed as follows. The kanamycin-resistant marker of a rimM disruptant from Keio collection (23), in which the rimM gene has been substituted by an FRT-flanked kanamycin-resistant cassette, was transduced into A19 using phage P1vir to produce an intermediate strain. Then, the kanamycin-resistant cassette was removed from the intermediate strain using an FLP expression plasmid pCP20 (24) to produce the A19ΔrimM strain. A19ΔrbfAΔrsgA is an A19 derivative in which both of the rbfA and rsgA genes are replaced by a short peptide gene containing an FRT sequence, constructed by transducing rbfA::FRT-kan-FRT into A19 and removing kan using pCP20 and then transducing rsgA::FRT-kan-FRT into the resulted strain and removing kan using pCP20. Sources of rbfA::FRT-kan-FRT and rsgA::FRT-kan-FRT are intermediate strains produced during the construction of W3110ΔrbfA (25) and the rsgA-disrupted strain of Keio collection (23), respectively. Both the A19ΔrimM and A19ΔrbfAΔrsgA strains were confirmed with polymerase chain reaction (PCR).

Spot assay and ribosome profile

A19, A19ΔrimM and A19ΔrbfAΔrsgA strains were grown in liquid LB at 37°C to OD 0.8 and diluted to a series of concentrations, 10^0, 10^-1, 10^-2, 10^-3, 10^-4 and 10^-5. Three microliters of each dilution was dropped to a LB plate and incubated at 37°C overnight. The cell extracts from the A19, A19ΔrimM and A19ΔrbfAΔrsgA strains were loaded onto a 10–40% sucrose gradient containing 10 mM Mg(OAc)₂ and centrifuged for 3.5 h at 39000 rpm in a SW41 rotor (Beckman Coulter). The gradients were analyzed with A254 absorbance using a Teledyne ISCO fractionation system.

Immature and mature 30S subunit purification

Escherichia coli cells (A19, A19ΔrimM and A19ΔrbfAΔrsgA strains) grown in LB medium were harvested, lysed and clarified in opening buffer [20 mM Tris–HCl (pH = 7.5), 150 mM NH₄Cl, 10 mM Mg(OAc)₂, and 0.5 mM ethylenediaminetetraacetic acid (EDTA)]. The lysate was loaded onto the top of 5 ml sucrose cushion [200 mM Tris–HCl (pH = 7.5), 150 mM NH₄Cl, 10 mM Mg(OAc)₂, 0.5 mM EDTA and 1.1 M sucrose] and centrifuged for 18 h at 28000 rpm in a 70Ti rotor (Beckman Coulter). The resulting pellets were resuspended in binding buffer and centrifuged through a 10–40% sucrose gradient with 10 mM Mg(OAc)₂ for 7 h at 30000 rpm in a SW32 rotor (Beckman Coulter). Fractions containing the immature 30S and 70S peaks were pooled separately and concentrated with buffer changed to binding buffer for the 30S fractions and to separation buffer [20 mM Tris–HCl (pH = 7.5), 150 mM NH₄Cl and 2 mM Mg(OAc)₂] for the 70S fractions. The 70S fractions were further centrifuged through a 10–40% sucrose gradient with 2 mM Mg(OAc)₂ to get the mature 30S and 50S subunits.

RimM preparation, tRNA extraction and identification of the 3′ and 5′ ends of the 17S rRNA

Full details are available in the Supplementary Data.

Pelleting assay

Mature or immature 30S subunits (2.5 pmol) were incubated with 30-fold excess of RimM for 15 min at 37°C in binding buffer. The mixture was then layered...
quantitatively examined by Grubbs’ test to remove outliers. Ratios of two or more tryptic peptides from the same protein were used to calculate the means and the standard deviations (Supplementary Table S1).

Cryo sample preparation and cryo-electron microscopy
Cryo-grids for the immature 30S subunits were prepared as previously described (26). The grids were examined in a FEI Tecnai F20 microscope operated at 200 kV, and images were recorded at a nominal magnification of 80,000 x on a Gatan UltraScan 4k CCD camera, under low-dose conditions (~20 e-/Å²). The complex of the immature 30S subunit bound with RimM was formed by an incubation of a 40-fold excess of RimM with the immature 30S subunits at 37°C for 15 min. The grids of the 30S complex were examined in an FEI Titan Krios cryo-TEM operated at 300 kV, and images were collected at a nominal magnification of 59,000 x on an FEI Eagle 4k x 4k CCD camera, under low-dose condition. Data collection was done with AutoEMation software package (27).

Image processing
All the micrographs were decimated by a factor of 2. Particle picking was performed using the SPIDER package (28) with a method based on a locally normalized cross-correlation function (29). The resulting particles (125 x 125 in window size, 2.76 and 3.0 Å in effective pixel size, for the 30S and the 30S complex samples, respectively) were manually verified using a method based on correspondence analysis (30). To ensure the performance of the 2D and 3D analysis, particles were further subjected to another round of manual screen, which finally rendered 164,368 and 94,535 particles for the 30S and 30S complex, respectively. The parameters of the contrast transfer function (CTF) were estimated using SPIDER at the micrograph level. Particles were then CTF corrected using the phase-flipping method (31).

2D image classification was performed using a maximum-likelihood approach (32) with the XMIPP software package (33). Particles from both samples were classified into 100 groups in 100 iterations, and the performance of the classification was monitored by log-likelihood function. To facilitate further comparison, class average images were subjected to a multi-reference alignment to 83 2D projections generated from a cryo-EM map of the mature 30S subunit (26), at an angular interval of 15° (Supplementary Figure S2).

3D classification was performed using a 3D maximum-likelihood approach with XMIPP (34). The initial model was generated by low-pass filtering (60 Å) of a cryo-EM map of the mature 30S subunit (26). Particles from the both samples were classified into five groups in 50 iterations, at an angular sampling of 10°. Refinements of the class structures were performed with SPIDER, following the standard reference projection matching procedures (31), with a gradual decrease of the angular step from 15° to 1°. Amplitude correction to the density maps was performed as previously described (26,35). The final resolutions of the refined density maps were estimated with a soft Gaussian mask approach (36,37) using 0.5 cutoff.
criterion of the Fourier Shell Correlation (Supplementary Table S2).

Atomic model and temperature map building
The head and body domains of a 30S subunit crystal structure [PDB ID: 3OFA, (38)] were docked into the cryo-EM maps as rigid bodies first using Chimera (39), followed by a flexible fitting method based on molecular dynamics simulation (40) in vacuo for 1 000 000 steps with a 0.5-kcal mol\(^{-1}\) scaling factor using NAMD (41). To avoid overfitting, ribosomal protein S2 and S3 were removed from the initial model before flexible fitting due to their low occupancies. For class No. 5 of the immature 30S subunits (Supplementary Table S2), all proteins in the head domain were removed and only the rRNA structure was refined. For better comparison, after fitting, S2 and S3 proteins were added back to the fitted structure using their contacting rRNA helices as reference. The 10 models were aligned using the 30S body domain as reference and 10 temperature maps were constructed in PyMOL (42) by calculating the deviation of the 16S rRNA in the fitted models from the mature 30S structure. The scripts used for root-mean-square deviation (RMSD) calculation and temperature map visualization were downloaded from http://pldserver1.biochem.queensu.ca/~rlc/work/pymol/. Chimera and PyMOL were used for graphic visualization.

RESULTS
Construction of a series of E. coli A19 strains
RNase I is the major non-specific endoribonuclease localized in periplasm and often found to be in 30S subunit fractions in cell extracts (43). To avoid undesired degradation of the rRNA precursors in the immature 30S subunits during the sample preparation, we chose the RNase I defective A19 strain (22) as the source of the 30S subunits. In this genetic background, we further constructed strains with the rimM gene deleted and with both rsgA and rbfA genes deleted. The two resulting strains grow poorly on LB medium and show an accumulation of free 30S subunits (Figure 1). Interestingly, both the cell growth test (Figure 1A) and the ribosome profile analysis (Figure 1B) show that the deletion of rimM is more deleterious. As a result, there is an intermediate peak between the 30S and 50S peaks, probably representing immature 50S precursors caused by globally decreased protein production in the ΔrimM strain (Figure 1B).

Compositional characterization of the immature 30S subunits from the A19 ΔrimM and ΔrbfAΔrsgA strains
RNA gel analysis shows that the rRNAs in the 30S fractions from the ΔrimM strain and the ΔrbfAΔrsgA strain are 16S rRNA precursors (Figure 2A), indicating that these free 30S subunits are indeed immature 30S particles. Identification of the two sets of 16S rRNA precursors by a previously established 5′-3′-rapid amplification of complementary DNA ends (RACE) technique (44) reveals that a majority of these precursors are unprocessed at both the 5′- and 3′-ends (Supplementary Figure S1). The protein gel analysis shows that some ribosomal proteins, e.g. S2 and S3, are underrepresented in the ΔrimM sample (Figure 2B). The compositional heterogeneity suggests that the immature 30S subunits from the ΔrimM strain are a collection of in vivo assembly intermediates that are different in protein composition.

To determine the protein levels, similar to a previously established quantification method (45), we used a QMS technique based on TMT labeling (46). The QMS data reveal that the levels of S21, S10, S14, S13, S19, S3, S2 and S5 are dramatically reduced in the ΔrimM sample (<50% of those in the mature 30S subunits (Figure 2C and Supplementary Table S1). Most of them are secondary and tertiary binding proteins from the 3′-head domain of the 30S subunit, except that S21 and S5 are tertiary binder from the central domain and the 5′-domain, respectively. S21 is known to easily dissociate in solution (47) and is therefore not included for further analysis. Thus, these data clearly demonstrate that the deletion of RimM causes a severe delay in the assembly of the 3′-domain of the 30S subunit in vivo (Figure 2C and Supplementary Figure S3). Among these 3′-domain proteins, S7 has the highest occupancy (81%) in the ΔrimM sample, which is in accordance with the in vitro assembly map that S7 is a primary binder and directs the binding of all the rest 3′-domain proteins (48).

In contrast, the protein composition of the immature 30S subunits from the ΔrbfAΔrsgA sample shows intriguing difference and similarity (Supplementary Figure S3). The most underrepresented protein in the ΔrbfAΔrsgA sample is still S21 (30%), followed by S7, S2, S10, S11 and S19 (49–63%) (Figure 2C and Supplementary Table S1), clearly showing a different pattern. Although many 3′-domain proteins, such as S10, S13, S14, S19 and S3 are also underrepresented, their levels are significantly higher than those in the ΔrimM sample (Figure 2D and Supplementary Table S1). In fact, the ΔrbfAΔrsgA sample has a higher level for almost all the proteins, compared with the ΔrimM sample (Figure 2C and Supplementary Figure S3). For example, S10, S13 and S14 have an over 2-fold increase and S21, S19 and S3 have a moderate increase.
from 1.5- to 2-folds. Interestingly, two primary proteins, S7 and S4, display significantly lower levels in the \(\text{C}1\text{rbfA} / \text{C}1\text{rsgA}\) sample than in the \(\text{C}1\text{rimM}\) sample (Figure 2C and D). Taking together, an evident pattern is that the immature 30S subunits from the \(\text{C}1\text{rbfA} / \text{C}1\text{rsgA}\) strain have significantly higher occupancies for all the secondary and tertiary binding proteins in the 3\(\text{0}\)-domain (Figure 2E), suggesting that their 3\(\text{0}\)-head domains are indeed further matured with more proteins incorporated.

This immediately suggests that a role of RimM in vivo is to promote the binding of 3\(\text{0}\)-domain proteins, since the immature 30S subunits from the \(\Delta\text{rbfA}\Delta\text{rsgA}\) strain likely resemble a stage downstream the RimM action. In agreement with this conclusion, the \(\text{in vitro}\) kinetic data show that RimM accelerates the binding of some head domain proteins, S19, S10 and S3 (49).

Structural characterization of the immature 30S subunits from the \(\Delta\text{rimM}\) strain

To explore the structural heterogeneity of the immature 30S subunits, we applied the cryo-EM single-particle method to our sample. First, a reference-free 2D image classification technique based on maximum-likelihood optimization (32) was employed to estimate the level of structural variation in the cryo-EM particles. The 2D analysis reveals that a large number of the class average images show smeared densities on the head domain of the 30S subunit. In contrast, densities in these average images corresponding to the body domain are nicely resolved, and the features of the body domain could be easily identified (Supplementary Figure S2). This suggests that the immature 30S subunits are truly composed of multiple assembly intermediates, with a highly flexible head domain and a rather rigid body domain.

Next, a multi-structure refinement method (34) was used to investigate possible metastable structural intermediates at the 3D level. As a result, the particles were grouped into five classes, and as expected, the five cryo-EM maps (at 12–14 Å resolution) display dramatic conformational differences (Figure 3). However, similar to a previous cryo-EM study on the immature 30S subunits from a \(\Delta\text{rsgA}\) strain (50), we did not find
significant densities that could be attributed to the unprocessed ends of the 17S rRNA.

To facilitate the quantitative comparison of the structural data, we built pseudo-atomic models for the five cryo-EM maps using a flexible fitting technique (40). Based on these models, five temperature maps for the 16S rRNA were then constructed according to their deviations from the structure of the mature 30S subunit (Figure 3K–O). Structural difference can be directly identified from these maps. First, the conformational difference is dominated by a relatively rigid rotational movement of the head domain, which changes the interdomain orientation between the head and the body domains. Especially, one class has a nearly 60° rotated head domain (Figure 3E, J and O). This rotated structure, derived from nearly one-third of all the particles (Supplementary Table S2), is in fact very similar to one of the Group II in vitro assembly intermediates discovered in a time-resolved electron microscopy study (11), which was shown to miss nearly all the 3-domain proteins. In addition to the rotation, in two classes (Figure 3K and M), the channel between the head and the body domains is closed up, resulting in a narrow down of the mRNA entrance. Second, four of the five maps show very incomplete, fragmented densities for helix 44 of the 16S rRNA, except for one group (Figure 3A), which is close to the conformation of a mature 30S subunit and also has a less rotated head domain. Along with the conformational difference at helix 44, the decoding center is also sharply different among the five maps (Supplementary Figure S4). Third, in agreement with the QMS data, these structures differ in protein composition, as exemplified by S2 and S3 (Figure 3F–J). In fact, none of the structures has a full occupancy for both factors, and interestingly the occupancy of S2 has no correlation with the occupancy of S3 (Supplementary Table S2). This observation appears to align well with the in vitro assembly data showing that S2 and S3 could bind in independent order and the prior binding of S2 ahead of S3 leads to kinetically trapped intermediates (11).

In summary, as seen in the temperature maps, the head domain of the 16S rRNA is highly mobile in the immature 30S subunits. It is known that the motion between the head domain and the body domain is intrinsic and is believed to be required for the dynamic interaction with translational components (51). However, the head domain rotation observed in our structures is in a much larger scale, suggesting that hypo-level of proteins in the 3-domain increases its flexibility. This structural observation demonstrates that the in vivo intermediates from the ΔrimM strain vary not only in protein composition but also in rRNA conformation.

**Structural characterization of the ΔrimM immature 30S subunits bound with RimM**

Next, we examined the binding preference of RimM to the immature and mature 30S subunits by pelleting assay. While RimM shows almost no binding to the mature
30S subunit, it indeed binds to the immature subunit, with a low affinity (Figure 4). In contrast, both RfbA and RsgA show a considerably higher affinity to the mature 30S subunit containing the 16S rRNA than RimM does, and especially, RsgA displays a strong preference to the mature 30S subunit (25). Therefore, similar to our QMS data, the binding preferences of these factors also suggest that RimM acts, prior to RfbA and RsgA, in the in vivo assembly pathway.

We then sought to explore possible structural changes of the immature 30S subunits upon RimM binding. Using the same 2D and 3D image classification techniques, we found that the addition of RimM to the immature 30S subunits seems to stabilize the 30S head domain. At the 2D level, class averages of particles from the RimM-treated sample still show smeared densities in the head region, but the total fraction of the particles with an unstable head domain is significantly smaller (Supplementary Figure S2). About 18% of particles from the untreated sample display apparent instability in the head domain, whereas the percentage in the treated sample is decreased to 11%.

At the 3D level, similarly, we classified the RimM-treated data into five groups, and these cryo-EM maps (at 15–19 Å resolution) also differ in conformation and protein composition (Figure 5). First, the head domain rotation is apparently in a much less scale, as seen in the temperature maps (Figure 5K–O and Supplementary Figure S4), implying the final accommodation of helix 44 is probably a later event, not related to RimM binding. Third, as expected, the occupancies of S2 and S3 are both very low, but surprisingly, the levels of S2 and S3 seem to be even lower than the untreated sample (Supplementary Table S2). This finding suggests that RimM stabilizes the immature 30S subunits in a conformation that disfavors S2 and S3 binding, implying that S2 and S3 binding might be later events in the assembly pathway.

Therefore, our structural analysis of the cryo-EM images from the RimM-treated sample shows that in addition to the role in ribosomal protein assembly, RimM also has a role in stabilizing the rRNA tertiary structure in the 3'-domain.

**Binding position of RimM on the 30S subunit**

The pelleting assay indicates that the affinity of RimM to the immature 30S subunits is very low (Figure 4). This apparently sets an obstacle for us to analyze contact sites of RimM on the 30S subunit in detail. Fortunately, RimM binds to S19 in vitro (20,21) and could co-crystallize with S19 (PDB ID: 3A1P). Therefore, the binding position of RimM could be deduced using S19 as a reference (Figure 6), given that RimM does not change its contacts in the context of the immature 30S subunit. In fact, the cryo-EM maps of the RimM-treated immature 30S subunits, although prepared with a 40-fold excess of RimM, have limited densities at locations expected to have RimM bound when the maps are displayed at a 3σ level (Figure 5). Densities corresponding to RimM begin to appear in lower threshold (Supplementary Figure S6). Nevertheless, we could compare the average densities statistically, within a 3D binary mask generated from the aligned RimM crystal structure. Consistently, the densities at RimM-bound region in the cryo-EM maps from the RimM-treated sample are significantly higher than those from the untreated sample (Supplementary Figure S6). This analysis, albeit rather preliminary, proves that RimM is present in these cryo-EM maps.

The structure of RimM is composed of two β-barrels containing domains (21). While the C-terminal domain is shown to interact with S19, the N-terminal domain closely resembles a tRNA-binding domain of EF-Tu (21), suggesting the ability of RimM to bind to the 16S rRNA. Consistently, alignment of the structure of the RimM-S19 complex immediately places the N-terminal domain of RimM at the junction of several helices, such as h29, h30 and h42 (Figure 6). Since, prior to RimM binding, the 30S assembly is in a stage with very limited 3'-domain protein incorporated (Figure 2C), the binding of RimM at this multi-helices interface might stabilize the rRNA conformation globally and therefore allows a faster and/or more stable binding of 3'-domain proteins.

**DISCUSSION**

The role of RimM in the assembly of the 30S subunit

It is known that disturbance to protein translation might affect the subunit assembly in an indirect way, due to a shortage in the ribosomal protein production. Consequently, the impaired subunit assembly in E. coli strains with genes for assembly factors deleted stems not only from the defective assembly process itself but also from a reduced supply of ribosomal proteins. Nevertheless, in this study, the composition of the in vivo intermediates from the ΔrimM and ΔrbsAΔrsgA strains clearly displays a non-uniform level of ribosome proteins, with mostly the 3'-domain proteins significantly underrepresented (Figure 2), suggesting that the secondary effect caused by impaired translation in these strains is negligible and does not over-shadow the assembly defects.
The role of RimM uncovered in the present work in fact serves as a perfect illustration to the proposed general function of assembly factors, i.e. to subvert possible kinetic traps caused by mis-folded rRNA or rate-limiting binding of certain proteins, during the \textit{in vivo} assembly process (11,13,49). Previous kinetic data showed that the binding of 30-domain proteins is not obligatory to the prebinding of S7 (13), while in contrast, prebinding of S7 and S19 together dramatically accelerates the binding of the rest 30-domain proteins (13), indicating the presence of a rate-limiting S7-independent assembly pathway for S19 (13,52). Consistently, our data show that in the \textit{C1 rimM} sample, S19 is among the most underrepresented proteins, whereas in the \textit{C1 rbgA/rsgA} sample, the level of S19 is dramatically increased (50), suggesting that RsgA acts at a very late stage when most of the components are already in place. Second, intermediates from the \textit{C1 rimM} strain show underrepresented levels for all secondary and tertiary binding proteins from the 30-domain. Interestingly, the \textit{C1 rimM} intermediates are to some extent similar in protein composition to a previously identified \textit{in vivo} 21S intermediates (44,53), but differ from the \textit{in vitro} RI intermediates (54). The close resemblance of the \textit{C1 rimM} intermediates to the naturally populated \textit{in vivo} 21S intermediates suggests that they represent an early stage during the assembly of the 30-domain, likely the entry stage. Last, in contrast, the intermediates from the \textit{C1 rbgA/rsgA} strain, with only a subset of 30-domain proteins largely underrepresented, do not resemble any of the known intermediates, indicating that they represent a novel set of intermediates roughly in-between.

**Functional interplay of assembly factors**

In addition to the two sets of intermediates described in this study, another set of \textit{in vivo} intermediates, isolated from a \textit{ΔrsgA} strain, was also quantitatively analyzed (50). The comparison of these quantitative data from different genetic background would enable us to identify the temporal relationship of assembly factors.

First, assembly intermediates isolated from a \textit{ΔrsgA} strain only have a very small subset of tertiary binding proteins (S21, S2 and S3) largely underrepresented (50), suggesting that RsgA acts at a very late stage when most of the components are already in place. Second, intermediates from the \textit{ΔrimM} strain show underrepresented levels for all secondary and tertiary binding proteins from the 30-domain. Interestingly, the \textit{ΔrimM} intermediates are to some extent similar in protein composition to a previously identified \textit{in vivo} 21S intermediates (44,53), but differ from the \textit{in vitro} RI intermediates (54). The close resemblance of the \textit{ΔrimM} intermediates to the naturally populated \textit{in vivo} 21S intermediates suggests that they represent an early stage during the assembly of the 30-domain, likely the entry stage. Last, in contrast, the intermediates from the \textit{ΔrbgA/ΔrsgA} strain, with only a subset of 30-domain proteins largely underrepresented, do not resemble any of the known intermediates, indicating that they represent a novel set of intermediates roughly in-between.
Therefore, the protein spectra of the above three sets of intermediates clearly suggest an order for the actions of these factors (Figure 6C), which is consistent with previous genetic and biochemical data (19,25,55). It must be noted that it is difficult to unambiguously timestamp other biogenesis factors in the assembly pathway due to the lack of biochemical data, although genetic data have suggested both functional redundancy and hierarchy for some assembly factors [reviewed in (12)]. Nevertheless, if we view the in vivo assembly as a multi-branched process, the seemingly function redundancy among assembly factors is merely a sign of altered contribution of different, inter-connected assembly pathways. As shown in Figure 6C, the late-stage assembly in vivo starts with an in vivo 21S intermediate (44,53) and proceeds along a highly efficient pathway in the presence of all assembly factors. Assembly factors come in play at different time points to assist certain kinetically disfavored assembly events. The disruption of a factor or a combination of factors would avert the assembly to less efficient branches and cause accumulation of a certain category of kinetically trapped intermediates. Consistent with this view, most of the E. coli genes for assembly factors are not essential. The remaining question is whether these kinetically trapped intermediates from various genetic background with different factors disrupted truly represent genuine snapshots of the assembly process in the normal condition, or different ‘dead-end’ products that are otherwise elusive in the normal condition.

The in vivo assembly of the 30S 3’-domain also follows parallel pathways

Early chemical probing of the 16S rRNA conformation (56), as well as recent kinetic measurement of the protein binding (9,11), showed that the 3’-domain assembly is the latest event during the in vitro assembly of the 30S subunit, coincident with the 5’- to 3’-transcription order. On the other hand, accumulating evidences (9–11,56) suggest that a major general feature of the in vitro assembly of the 30S subunit is that the process proceeds along multiple routes. In this study, we isolated the in vivo assembly intermediates from two genetically modified E. coli strains. The most remarkable feature in the protein spectra of these two sets of intermediates is that they both severely lack 3’-domain proteins, suggesting that the maturation of the 3’-domain is also a rate-limiting process in vivo. These quantitative data also indicate that the intermediates from both strains are very heterogeneous in ribosomal protein composition, which means they do not represent a single populated assembly intermediate state, but rather a collection of multiple-related intermediates with more than one metastable state enriched. These differently prepared intermediates, although with a recognizable temporal relationship, cannot be easily reconciled by a single continuous assembly pathway.
With respect to the occupancies of individual proteins, there are a number of exceptions to the well-accepted Nomura assembly map (5). To name a few, first, an unexpected observation of our QMS data is that two primary proteins (S4 and S7) in the \( \Delta rstA \Delta rsgA \) sample show decreased levels compared with the \( \Delta rimM \) sample. S4 is thermodynamically required for subsequent binding of S16, S12 and S5 to the 3'-domain (5), and more importantly, S4 was shown to have a global stabilization effect on the 3'-domain (57). S7 is the only primary protein in the 3'-head domain and directs the following binding of S9, S13 and S19 (5,48). However, in the \( \Delta rstA \Delta rsgA \) sample, the occupancy of S4 and S7 (Figure 2) is even lower than its follower proteins. Similarly, the level of S7 was also reported to be lower than its follower S9, S13 and S19 in the \( \Delta rsgA \) sample (50). Second, levels of S10 and S14 are lower than their follower tertiary proteins S2 and S3 in the \( \Delta rimM \) sample, suggesting that S3 could bind independent of S10. Third, the level of S2 is lower than S3 in the \( \Delta rstA \Delta rsgA \) sample, although S3 binding is thermodynamically dependent on the prior binding of S2 in the Nomura map.

The Nomura map was derived by single protein omission reconstitution experiments with fully processed 16S rRNA under equilibrium conditions and therefore does not necessarily reflect the true order of serials of binding events during assembly (52). In addition to our QMS data, deviations from the Nomura map have already been observed from both in vivo and in vitro studies. Previous genetic data show that S15, a primary protein in the central domain, is dispensable for the 30S assembly in vivo (58). Furthermore, in vitro kinetic data from Williamson group based on PC-QMS (11,13) or fluorescence triple correlation spectroscopy (52) indicate that S9 and S19 could bind independent of S7 (13,52), and S2 could bind independent of S3 (11). All these data suggest that there are hidden assembly pathways that could not be directly inferred from the Nomura map. Thus, the seemingly discrepancy between our QMS data and the Nomura map could be easily reconciled if we view the in vivo assembly process as a highly branched network. In the presence of not fully processed 17S rRNA and the absence of certain assembly factors, the 30S assembly in vivo takes alternative, kinetically inefficient, pathways that are not predicted by the Nomura map. Therefore, the assembly intermediates isolated from these different assembly factor-deficient mutants might represent intermediates kinetically trapped in various parallel branches in the assembly network.

In summary, although the number of possible assembly pathways in vivo is limited by the co-transcriptional nature and the presence of assembly factors in the in vivo condition, our data provide additional strong evidence to the emerging idea that the in vivo assembly also proceeds along parallel pathways in a certain degree (54,58).

The maturation of the 16S rRNA 3'-domain in vivo is highly coupled with protein assembly

Our structural data reveal that the in vivo assembly intermediates differ largely in rRNA conformation. Through the integration of structural (26,50) and QMS data [the present work and (50)], we could draw a conclusion that the 3'-domain of the 16S rRNA maturates in a progressive manner in vivo, paralleling with the ribosomal protein assembly.

First, assembly intermediates from \( \Delta rimM \) cells show dramatic conformational differences in the position of the 3'-domain (Figure 3 and Supplementary Figure S2). In contrast, cryo-EM structures of intermediates from \( \Delta rsgA \) cells (50) also vary in the 3'-domain position, but with a much smaller scale. Second, the long helix 44 of the 3'-minor domain is highly flexible and is almost invisible in the structures of intermediates from \( \Delta rimM \) cells (Figures 3 and 5). However, cryo-EM structures of 30S intermediates from \( \Delta rsgA \) cells (50), which also contain 17S rRNA, display well-resolved densities for helix 44, except for the upper decoding center region. This suggests that helix 44 adapts its mature conformation only at a very late stage. In fact, hydroxyl radical probing data (10,59) already showed that the full accommodation of helix 44 is a late event even when the experiments were performed with the 16S rRNA. Third, further downstream is the cryo-EM structure of the 30S–RsgA complex, which displays an almost identical conformation to the mature 30S subunit (26).

Based on the above structural comparisons, the maturation of the 3'-domain of the 16S rRNA follows the transcription order in a progressive manner, first the 3'-head domain, next the 3'-minor domain (Figure 6C), and more importantly, the conformational maturation is coupled with the gradually increased protein level (Figure 6C). Therefore, our data have revealed another common characteristic shared by the in vivo and the in vitro processes, i.e. a high cooperativity between protein binding and rRNA folding (10,11,60,61).

ACCESSION NUMBERS

The cryo-EM maps have been deposited in the EMDataBank (EMDB codes 5500-5504 and 5506-5510 for the immature 30S subunits with or without the RimM treatment, respectively). The atomic models of the 16S rRNA have been deposited in the Protein Data Bank (PDB codes 3J28, 3J29, 3J2A, 3J2B, 3J2C, 3J2D, 3J2E, 3J2F, 3J2G, 3J2H).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Materials and Methods, Supplementary Tables 1 and 2, Supplementary Figures 1–7 and Supplementary References [26,32,44].

ACKNOWLEDGEMENTS

We thank Nara Institute of Science and Technology, National BioResource Project and National Institute of Genetics for providing the \( \Delta rimM \) strain from the Keio collection and Dr Krisana Asano for construction of an overproducing system of RimM. We also thank the
FUNDING
Ministry of Science and Technology of China [2010CB912401, 2010CB912402 and 2013CB910404]; the National Natural Science Foundation of China [31170677] and the Japan Society for the promotion of Science [23380054]. Funding for open access charge: Ministry of Science and Technology of China.

Conflict of interest statement. None declared.

REFERENCES

of the E. coli 70S ribosome at 11.5 Å resolution. Cell, 100, 537–549.