










Cryo-EM captures early ribosome assembly in action

Received: 10 November 2022

Accepted: 8 February 2023

Published online: 17 February 2023

 Check for updates

Bo Qin ^{1,8}, Simon M. Lauer ^{1,8}, Annika Balke ², Carlos H. Vieira-Vieira ^{3,4}, Jörg Bürger ^{1,5}, Thorsten Mielke ⁵, Matthias Selbach ^{3,6}, Patrick Scheerer ², Christian M. T. Spahn ¹ ✉ & Rainer Nikolay ^{1,7} ✉

Ribosome biogenesis is a fundamental multi-step cellular process in all domains of life that involves the production, processing, folding, and modification of ribosomal RNAs (rRNAs) and ribosomal proteins. To obtain insights into the still unexplored early assembly phase of the bacterial 50S subunit, we exploited a minimal in vitro reconstitution system using purified ribosomal components and scalable reaction conditions. Time-limited assembly assays combined with cryo-EM analysis visualizes the structurally complex assembly pathway starting with a particle consisting of ordered density for only ~500 nucleotides of 23S rRNA domain I and three ribosomal proteins. In addition, our structural analysis reveals that early 50S assembly occurs in a domain-wise fashion, while late 50S assembly proceeds incrementally. Furthermore, we find that both ribosomal proteins and folded rRNA helices, occupying surface exposed regions on pre-50S particles, induce, or stabilize rRNA folds within adjacent regions, thereby creating cooperativity.

The assembly of ribosomal subunits, the largest ribonucleoprotein particles in the cell, relies on the intrinsic affinities of their rRNA and protein components that interact in a hierarchical manner^{1,2}, involving protein-induced conformational changes in rRNA³. The resulting cooperativity of assembly ensures the completion of ribosomal subunits⁴. The *E. coli* 50S large ribosomal subunit is composed of 33 proteins (L-proteins), one 23S and one 5S rRNA. The 23S rRNA consists of six architectural domains that, stabilized by the L-proteins, define the shape of the 50S subunit with its three protuberances^{5–7}. Cryo-EM studies of bacterial 50S precursors both purified from cells^{8–13} and obtained by in vitro assembly¹⁴ have elucidated the later phase of the maturation pathway.

The most immature 50S precursor states reported contained an already fully formed core, consisting of 23S rRNA domains I, II, III,

VI and early binding L-proteins^{10,12,14}. In this core particle about 50% of all protein and RNA residues are already maturely positioned¹⁴. Further assembly involves the formation of the LI stalk, the central protuberance (CP), the GTPase associated center (GAC), and the stalk base. 50S assembly terminates with folding of the peptidyl transferase center (PTC) and the surrounding rRNA elements referred to as functional core (FC)^{8–10,12,14–16}. The events after structural formation of the large subunit's core appear to be evolutionary conserved. Hence, similar assembly intermediates have been observed for the eukaryotic 60S subunit^{17–19} and large subunit precursors derived from mitochondria, which shed light on the latest period of assembly, detailing the finely tuned steps leading to completion of the subunit's active site^{20–25}, although the involved biogenesis factors differ substantially¹³.

¹Institut für Medizinische Physik und Biophysik, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt Universität zu Berlin, Berlin, Germany. ²Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Institute of Medical Physics and Biophysics, Group Protein X-ray Crystallography and Signal Transduction, Charitéplatz 1, D-10117 Berlin, Germany. ³Proteome Dynamics, Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Robert-Rössle-Str. 10, 13125 Berlin, Germany. ⁴Faculty of Life Sciences, Humboldt Universität zu Berlin, Berlin, Germany. ⁵Microscopy and Cryo-Electron Microscopy Service Group, Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195 Berlin, Germany. ⁶Charité -Universitätsmedizin Berlin, 10117 Berlin, Germany. ⁷Department of Genome Regulation, Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, 14195 Berlin, Germany. ⁸These authors contributed equally: Bo Qin, Simon M. Lauer.

✉ e-mail: christian.spahn@charite.de; nikolay@molgen.mpg.de

In comparison, early assembly of the large ribosomal subunit remains elusive from a structural standpoint. Pioneering studies visualizing nucleolar derived samples from yeast provided first structural insight into intermediate states before core formation has been completed^{26,27}. However, large subunit precursors exploring the earliest stages of their assembly escaped structural analyses so far. Here, we present a biochemical and cryo-EM based analysis of the earliest period of 50S assembly, utilizing an *in vitro* reconstitution system with purified rRNA and L-protein components. We provide high resolution structures of early pre-50S intermediates explaining their structural maturation via parallel assembly routes, starting with a particle consisting of density corresponding to the first -500 nucleotides of the 23S rRNA and the proteins uL22, uL24, and uL29.

Results

Time course of 50S *in vitro* reconstitution

To focus on early events of 50S assembly, we adopted our previous approach¹⁴ and performed step 1 of the *in vitro* reconstitution assay²⁸ as a time course reaction, analyzed sucrose gradient profiles, translation activity of the assembled particles (after a full step 2 incubation), and determined the protein occupancy, using quantitative mass

spectrometry (q-MS) (“Methods”, Fig. 1a and Supplementary Fig. 1a–j). Before heat incubation, the sample migrates in a sucrose gradient as a 33S precursor (Fig. 1b) with reduced amounts of the L-proteins bL32, uL30, uL14, uL29, bL19, bL28 and uL16 (Supplementary Fig. 1j), and exhibits no translation activity (Fig. 1c), in agreement with previous studies²⁹. As expected, both the portion of faster migrating particles and the translation activity increased with the time of incubation (Fig. 1c and Supplementary Fig. 1b–h).

The sucrose gradient profiles revealed that after 3 min of reaction, the sample exhibited its largest complexity with the majority of precursors in the 33S region and a significant amount of faster migrating particles (Fig. 1b). Nevertheless, particles in the 3 min sample contain all L-proteins (except bL31 and bL36) with occupancies of >70% (Supplementary Fig. 1i). Hence, we selected this sample for a detailed cryo-EM analysis. After extensive sorting and multi-particle refinement, we were able to disentangle 16 different states, with nominal resolutions between 3.0 and 6.6 Å (Supplementary Fig. 2, Supplementary Fig. 3, and Supplementary Table 1). While some of the obtained states are related to the 41S-like and 48S-like particles of the late phase of assembly that we have previously obtained with a full 30 min incubation¹⁴, the time-course approach resulted in nine states

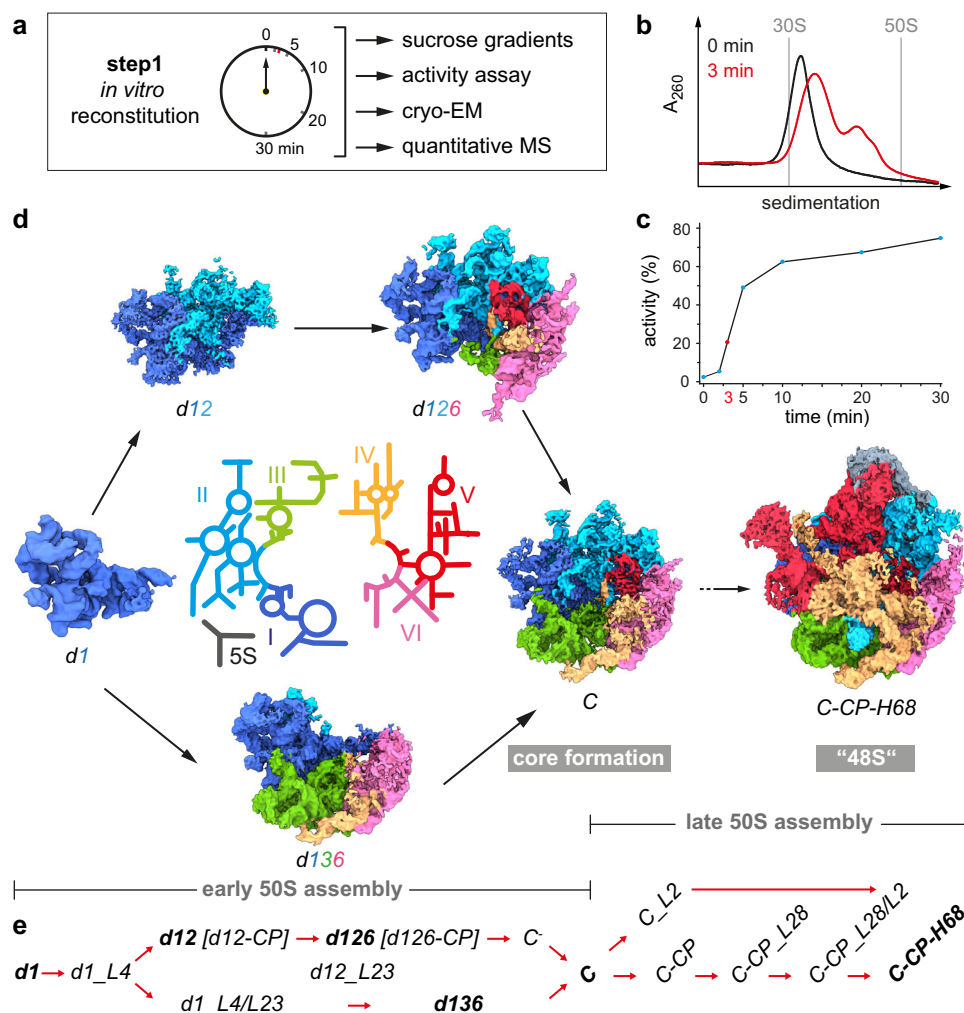


Fig. 1 | Time course of 50S step 1 *in vitro* reconstitution with structural and biochemical analyses. **a** Experimental setup. **b** Sucrose gradient profiles of *in vitro* reconstitution samples incubated for 0 min (black) or 3 min (red), respectively. **c** *In vitro* translation assay indicating the relative activity of subunits after the indicated time of incubation under step 1 conditions plus subsequent incubation under step 2 conditions for 90 min. Data are presented as mean values based on $n = 2$ technical

replicates. **d** In the center, a color-coded 2D map of the 23S rRNA with the individual domains labeled from I–VI (5S, 5S rRNA). Cryo-EM maps of selected states derived from the step 1 reaction after 3 min of incubation appear in the same color-code. **e** Nomenclature for states occurring during early and late 50S assembly. States in bold are shown in **(d)**. *d*, 23S rRNA domain; *C*, 50S core; CP, central protuberance; H68, helix 68. Source data are provided as a Source Data file.

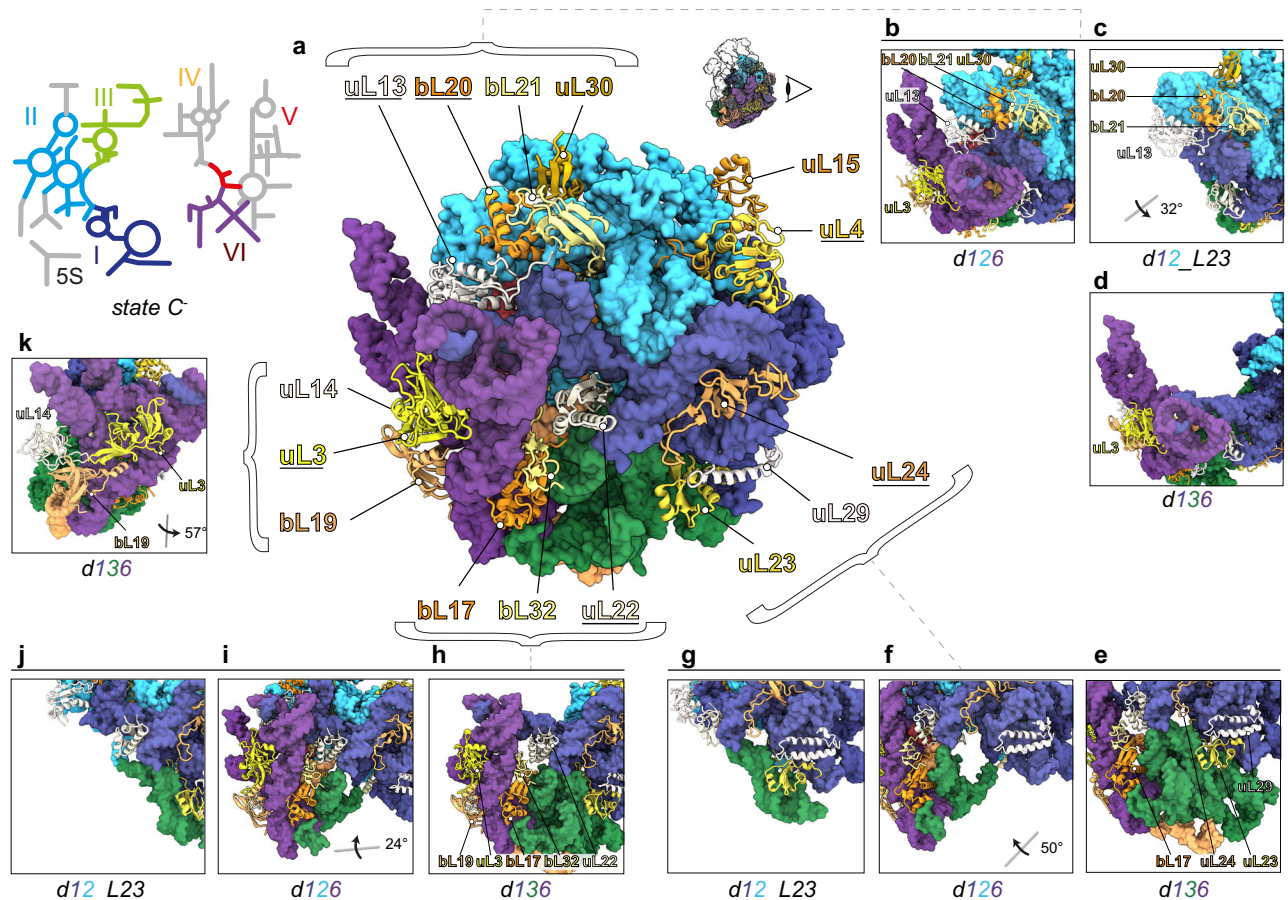


Fig. 2 | Clustering of L-proteins. **a** 2D rRNA map and atomic model of state C in back view. L-proteins appear as cartoons (early essential proteins and uL3 are underlined), rRNA as surface model (lowpass-filtered to 5 Å resolution). Presence and absence of bL21, bL20, and uL30 in states *d126* (**b**), *d12* (**c**), and *d136* (**d**),

respectively. Presence and absence of uL23 in states *d136* (**e**), *d126* (**f**), and *d12* (**g**), respectively. Presence and absence of bL17, bL32, and bL19 in states *d136* (**h**), *d126* (**i**), and *d12* (**j**), respectively. Positions of uL14, uL3 and bL19 in state *d136* (**k**). Viewing angles in (**c**, **f**, **i**, and **k**) are shown relative to the full model in **a**).

identified herein that were clearly less complete than the core particle. Whenever in the following we describe the presence, or appearance of an L-protein, an rRNA element, or an rRNA domain, it refers to a detectable, unambiguous density of this component in the corresponding cryo-EM map. Appearance of such densities can result from a component just binding at that moment or being part of the complex already, and structural rearrangements led to a more stable conformation capable of being captured by cryo-EM based technologies. This caveat is of particular importance in cases where densities for proteins and RNA elements appear simultaneously.

Structure-based pathway for early 50S assembly

The earliest state we identified contained strong density for ~500 bp of domain I and the L-proteins uL22, uL24, and uL29 (referred to as state *d1*, Fig. 1d and Supplementary Fig. 4a). Hence, 50S assembly in vitro starts with the stable formation of the 5' end containing domain I, despite the presence of the entire 23S rRNA molecule. Interestingly, uL24, whose binding site is in the center of *d1*, is found among the first binders (Supplementary Fig. 1i), which explains its role as an assembly initiator²⁹. To better understand the molecular anatomy of the subunit, we arranged selected cryo-EM maps in a logical order and noticed that early assembly can progress at least along two putative routes (Fig. 1d).

Instead of numbering the individual states successively, we designed a more specific nomenclature (Fig. 1e). The early states were designated according to the 23S rRNA domains whose cryo-EM densities they exhibit (e.g., *d1*, *d12*, *d126*, etc.). In order to distinguish states

very similar in composition, additional structural elements were added to their names (e.g., *d1_L4/L23*, densities for domain I plus uL4 and uL23 etc.). Once densities for domains I, II, III, VI, and their interacting proteins are present, the large subunit's core has completely formed (state C). To differentiate states occurring during late assembly, emerging structural features such as the central protuberance, or helix 68 were added to their names (*C-CP*, or *C-CP-H68*).

While state *d1* can mature to *d12* and *d126*, an alternative route involves state *d136*. Both states *d126* and *d136* can convert into state C (*d1236*), a process we refer to as core formation, that enhances base pairing between the 5' and 3' ends of 23S rRNA. Consequently, core formation in vivo can only be accomplished after transcription of the entire 23S rRNA molecule. Subsequently, state C transitions via intermediate steps into state *C-CP-H68* (48S) (Fig. 1d, e and Supplementary Figs. 4 and 5). Interestingly, some of the 23S rRNA domains indeed appear to be folding units that are intrinsically rigid, but flexible relative to one another (Supplementary Movie 1). Domains I and VI exhibit full cryo-EM density within their domain boundaries as soon as they become detectable as parts of discrete states (Fig. 1d; *d1*, *d126*, and *d136*). Defined regions of domains II and III show density in states *d1_L4* and *d1_L4/L23*, respectively, before subsequent concerted appearance of their remaining densities (Supplementary Fig. 4). In the case of domain II only density for the central region appears and extended parts such as A-site finger and GTPase associated center (GAC) form in late assembly. Taken together, early 50S assembly occurs in a domain-wise fashion, while late 50S assembly proceeds incrementally.

Detailed molecular morphogenesis - Contribution of L-proteins

L-proteins concentrating on the back site of the subunit are organized in clusters, containing at least one early assembly protein, providing contact points for later binding proteins and connecting adjacent rRNA domains (Fig. 2a). To exemplify in more detail how certain L-proteins contribute to 50S assembly, we compared states *d1*, *d1_L4* and *d1_L4/L23* (Fig. 3a–e and Supplementary Movie 2). State *d1* is lacking density for uL4 (Fig. 3a), while *d1_L4* exhibits densities for uL4 along with domain II helices H27–31 (Fig. 3b). Similarly, state *d1* is lacking density for uL23 (Fig. 3d), while *d1_L4/L23* exhibits densities for uL23 along with domain III helices H49–53, including the interjacent bulges (Fig. 3e). Thus, uL4 and uL23 appear to contribute to the formation of local seeds to promote further assembly of domains II and III, respectively (Fig. 2b–j). However, while uL4 is present in all downstream assembly intermediates, presence, or absence of uL23 determines, whether assembly proceeds along route 1-3-6 or 1-2-6 (Fig. 2e–g).

Another L-protein that affects the formation of domain III is bL17, whose presence mediates stable folding of H47–49 (Fig. 2h–j). Based on these observations, we conclude that some L-proteins, occupying surface exposed regions on a pre-50S particle, are capable of folding or, stabilizing rRNA elements of adjacent domains. While it is well-known that L-proteins are frequently located at rRNA domain interfaces^{6,7}, and that r-proteins have the capability to remodel rRNA in a cooperative fashion³ this dedicated function of uL4, uL23, bL17, and other L-proteins in ribosome assembly was not apparent (Supplementary Movie 3).

Detailed molecular morphogenesis—Successive manifestation of L-proteins

Globular domains of L-proteins, like in uL24, contribute to the formation of the subunit's crust. Some L-proteins such as uL4, uL3, bL32, and uL2, contain a globular domain (yellow) combined with extended tentacles or loops (red) that interestingly project towards the center of the subunit (Fig. 4a, b). Cryo-EM density for the globular domain of uL4 appears early after formation of domain I, while most of the loop forms with domain II, and the apical tip even later with domain V (Fig. 4c). The globular domain of uL3 critically contributes to formation of domain VI (Fig. 2h, i, k), and the extended loop forms later with domains II, IV and V (Fig. 4d). In addition, bL32 occupies a critical position with its globular domain being part of domain VI (Fig. 2a, h, i) and the extension interacting with domains I, II and V, thereby connecting four rRNA domains (Fig. 4e).

Interestingly, the globular part of uL2 is present on domain III only after core formation in state *C_L2* (Supplementary Fig. 4l) and binding can occur later in state *C-CP_L28/L2* (Supplementary Fig. 4o). Nevertheless, the extended loop starts forming earliest in state *C-CP_L28/L2* and interacts mainly with and stabilizes helices H65, H66 and H69 of domain IV. Only in a mature 50S subunit, density for the whole loop appears that in addition undergoes contacts with domain V (Fig. 4f). Taken together, the flexible tentacles of some L-proteins form later than their globular domains and interact predominantly with nucleotides of the late forming rRNA domains IV and V, possibly by mechanisms involving dynamic sampling, as shown for proteins of the 30S subunit^{30,31}.

Hierarchical arrangement of rRNA

Another principle we observe is the stabilization or folding of rRNA helices, induced by adjacent rRNA elements that are nearly maturely folded. Assembly appears to nucleate at the 5'-end with the formation of domain I to yield state *d1*. To rule out biased image classification by initial alignment, we re-analyzed our cryo-EM data and aligned all particle images to domain VI before proceeding with 3D classification (Supplementary Fig. 6). However, this did not result in discovery of states that would support any nucleation point other than domain I.

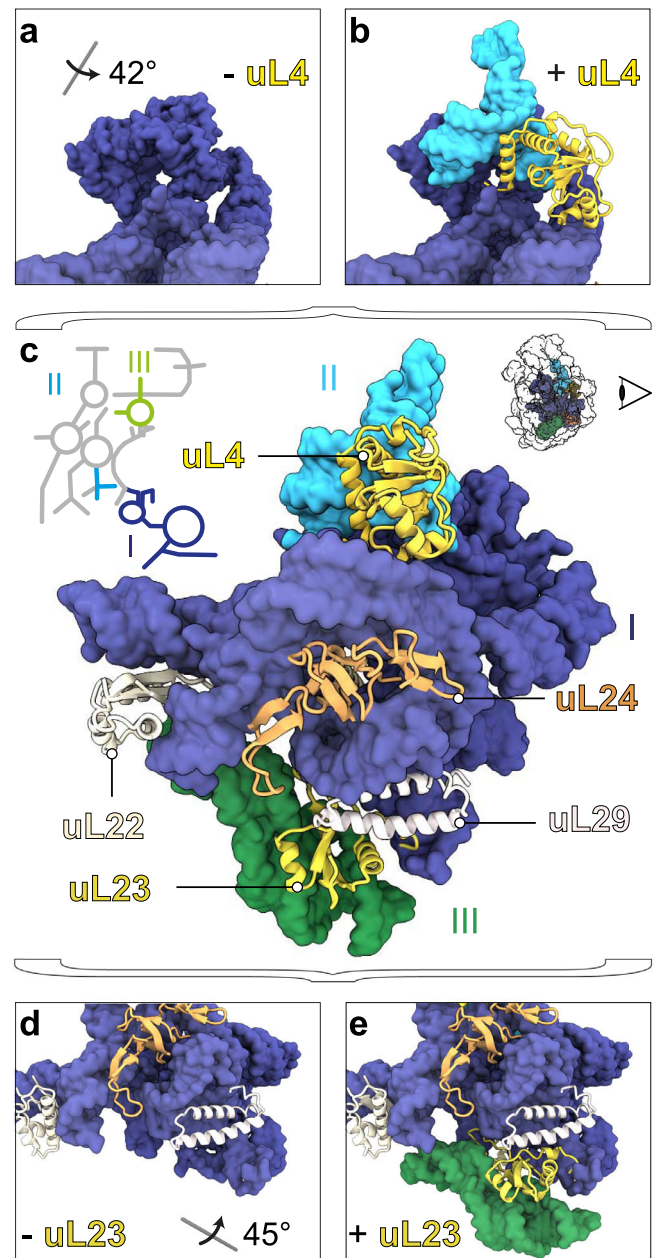


Fig. 3 | Proteins with seeding function. PDB models of early states with rRNA in surface representation and L-proteins as cartoons. Top segments of states *d1* (a) and *d1_L4* (b), full model of state *d1_L4/L23* in back view (c), and bottom segments of states *d1* (d) and *d1_L4/L23* (e) color-coded according to the 23S rRNA 2D map. All viewing angles relative to (c).

Thus, domain I indeed seems to appear first, while the remaining rRNA elements dock onto an already existing RNA surface.

The precursors *d12* (Fig. 5a, b) and *d136* (Fig. 5c, d) exhibit intrinsic flexibility along certain hinge regions located in domain interfaces (red dotted lines), while the individual rRNA domains remain relatively rigid. Nevertheless, the relative domain movements from the most open (Fig. 5a, c) to the close conformation (Fig. 5b, d) allow for the formation of inter-domain contacts between H11/ H37 (Fig. 5b) and H21/ H51 and H11/ H37 (Fig. 5d). In some cases, formed helices remain flexible and are stabilized by tertiary contacts during subsequent assembly. For instance, both H21 and H22 exhibit solid density in state *C*, but both helices reconfigure their shape and conformation and form tertiary contacts once reaching their mature positions (Fig. 5e–g). Further rRNA-assisted rRNA stabilization or

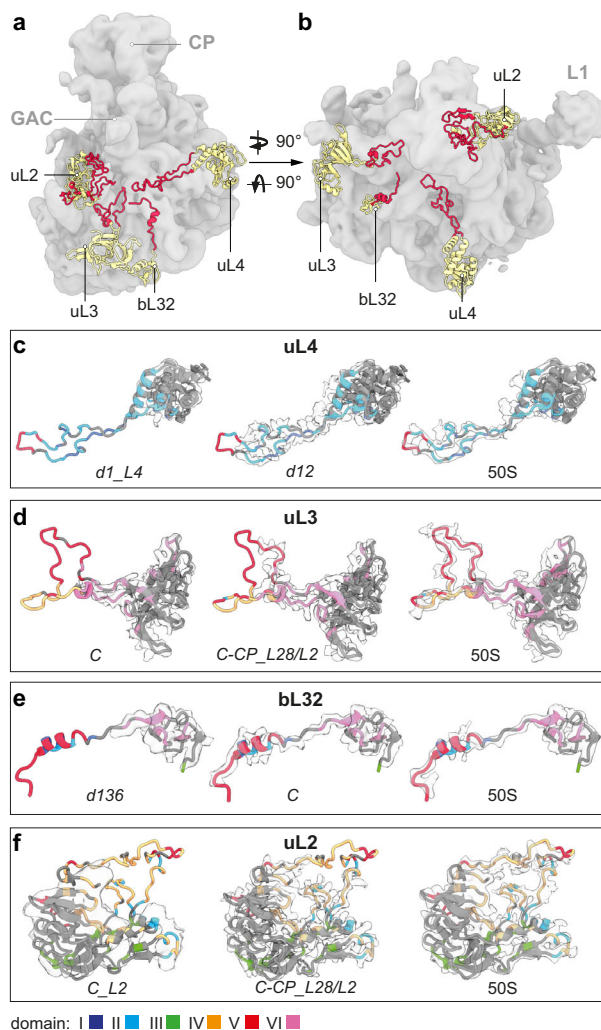


Fig. 4 | L-proteins with extensions. **a** Side view, **b** top view of a lowpass-filtered *C-CP-H68* cryo-EM map (gray) with models of selected L-proteins exhibiting globular domains (yellow) and extensions (red). CP, central protuberance; GAC, GTPase associated center; L1, L1 stalk. **c–f** Cryo-EM maps (light gray) of selected L-proteins, derived from the indicated states, and corresponding atomic models (EMD-22614). Protein models are colored according to their rRNA domain contacts; gray color indicates self-interaction, or surface exposure (“Methods”).

folding becomes apparent, when comparing assembly along the routes 1-2-6 and 1-3-6. While route 1-2-6 initially yields a core particle lacking density for domain IV helices H61-63 (state *C*), maturation along route 1-3-6 culminates in a particle with these parts of domain IV stably folded (state *C* and better defined in state *C-CP_L2/L28*) (Figs. 1d and 5h–j). Hence, initial appearance of domain IV benefits from the presence of the maturely folded domain III. Specifically, rRNA contacts with the already formed domain III helices H56-58 seem to induce formation of domain IV H61 and H63 in state *d136*, and H61-63 in state *C* (Supplementary Fig. 7a–c), thereby initiating the incremental formation of domain IV.

In addition, we noticed that the 5S RNP (ribonucleoprotein particle), consisting of 5S rRNA, uL5, uL18, and bL25, can dock on a pre-50S particle as soon as the central region of domain II has formed (Supplementary Fig. 8a, b). This can occur both in early (*d12-CP*, *d126-CP*) and late assembly (*C-CP*) and goes along with formation of domain II helices H38 (A-site finger) and H42 (GAC), as well as domain V helices H80-H88 (Supplementary Fig. 4d, f, m and Supplementary Fig. 8c), representing additional examples for rRNA induced rRNA stabilization or folding. However, the CP remains mobile and upon rotational

movement of the 5S rRNA towards the subunit interface, H84 and H38 stabilize each other (Fig. 5k, l).

Late 50S assembly

Late 50S assembly, apart from core-stabilizing incorporation of uL2 (Fig. 5m and Supplementary Fig. 7a, d–g), involves stable binding of bL9 and bL28 in state *C-CP_L28*, which stabilizes the base of the L1 stalk (Supplementary Fig. 4o and Supplementary Fig. 7h–i). Next, H68 can form upon integration of bL33 and bL35 (Supplementary Fig. 4q and Supplementary Fig. 7j–k), while bL35 rigidifies the CP (Fig. 5n). In addition, the fixation of H68 seems to involve contacts with uL1. Furthermore, the L1-stalk associated H75 and CP associated H88 provide RNA contacts and stabilize H68. Accordingly, late step 1 assembly terminates with the “48S” particle *C-CP-H68*, sharing most structural features with a mature 50S subunit, but lacking a properly arranged FC. It is remarkable that the evolutionary most ancient part assembles last, both in vivo¹² and in vitro¹⁴. The extensive lack of L-proteins within the FC rationalizes why numerous assembly factors (such as EngA, DbpA, RlmE, and ObgE^{13,32–35}) are required to finalize it in vivo and why high thermal energy is required to activate the particle in vitro¹⁴.

Discussion

Here we show, using the 50S in vitro reconstitution assay in combination with cryo-EM and multi-particle refinement that already after 3 min of reaction as much as 16 structurally distinct precursors have formed. Hence, the structures we report provide unprecedented insights into the detailed progression of 50S in vitro assembly, addressing long-standing questions and revealing fundamental principles of large subunit assembly.

One important question related to in vitro ribosome assembly assays is, how subunit assembly can take place, despite the presence of full 16S or 23S rRNA? In principle, structure formation could initiate at multiple, possibly competing sites. While we cannot rule out this scenario (based on our qMS data), we did not obtain structural evidence for any nucleation point other than domain I (Supplementary Fig. 6). Instead, our data indicate that the assembly of the 50S subunit does not proceed arbitrarily but has its morphogenetic origin in domain I and overall follows a 5′ to 3′ direction.

Based on their intrinsic affinities, uL22, uL24, and uL29 interact with their RNA binding sites in domain I and nucleate 50S assembly. Consequently, binding sites for uL4 and bL23 are formed. Next, binding of uL4 and bL23 initiates formation of domains II and III, respectively, and further proteins join, exemplifying the inherent cooperativity of ribosome assembly. Hence, we visualize Nomura’s paradigm, that all the information for ribosome assembly is contained in the structure (i.e., the chemical affinities) of the participating components³⁶. As a practical consequence, atomic models derived from the assembly intermediates will be useful for the design of single molecule approaches and pharmacological agents interfering with critical transitions along the 50S assembly pathway.

After formation of domain I, assembly continues uL4 dependent with *d1_L4* and branches either to route 1-2-6 or 1-3-6 (Fig. 6). The presence of domain II apparently facilitates docking of the CP-forming 5S RNP and as a consequence, route 1-2-6 involves the possibility of CP formation either on *d12*, or *d126* resulting in *d12-CP*, or *d126-CP*, respectively. However, completion of the CP can also occur after the formation of state *C* during incremental late assembly (Supplementary Fig. 5). Route 1-3-6 yields state *C* due to presence of domain III, which is seeded by uL23, and in turn facilitates the incorporation of domain VI independently of domain II (Fig. 6 and Supplementary Fig. 5). Furthermore, additional routes are conceivable. Precursor *d12_L23* could derive from both *d12* and *d1_L4/L23* and constitute an intermediate that matures to state *C* by completion of domain III and simultaneous or subsequent formation of domain VI. Proteins that contribute to the formation of these domains are bL17, bL32, and uL3 (Fig. 6).

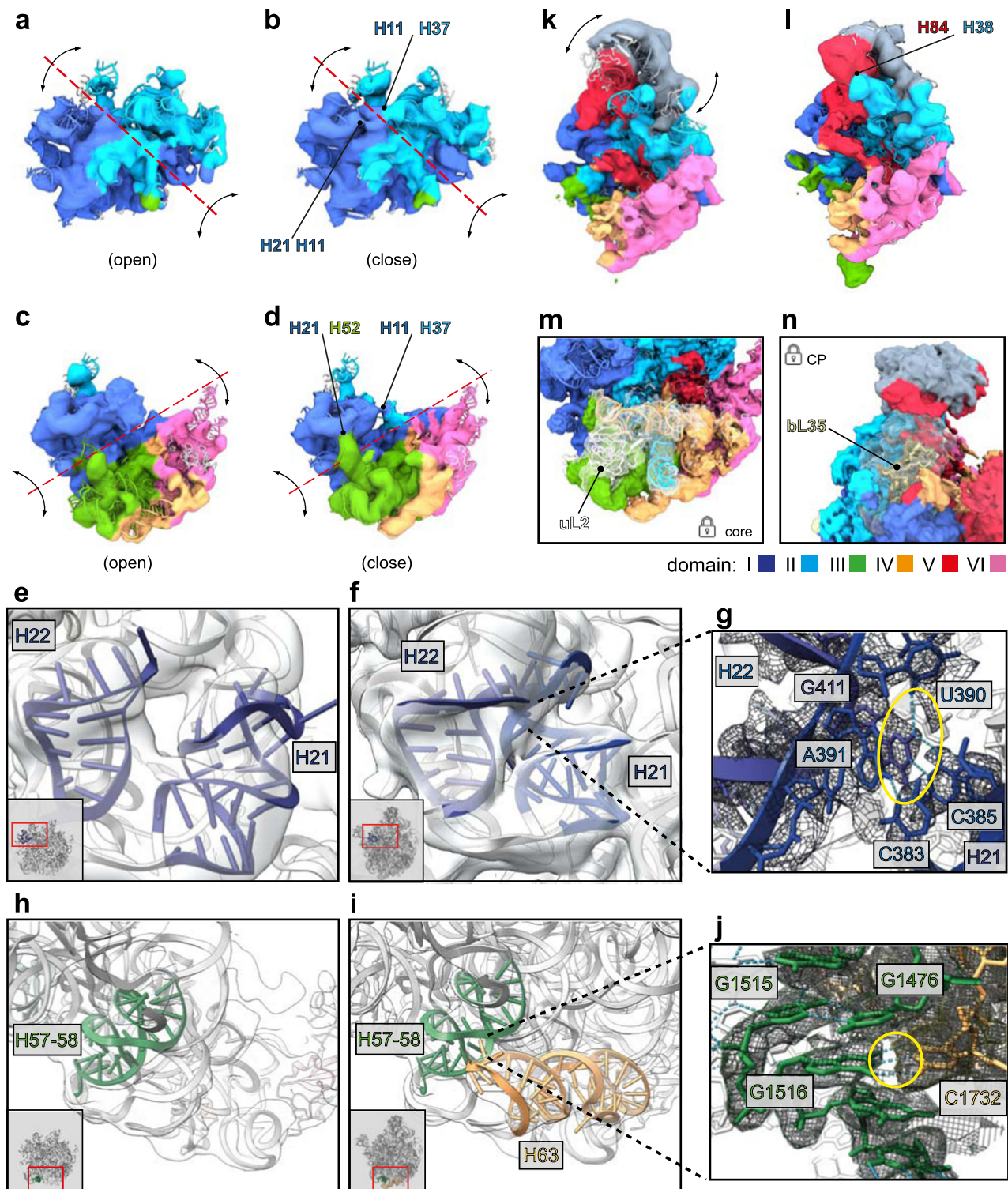


Fig. 5 | Tertiary contacts between rRNA helices. 3D variability maps of states *d12* (a, b), *d136* (c, d), *d126-CP* (k, l). First (left) and last (right) reconstructed frame of each state is shown. Maps were filtered to 8 Å and color-coded according to the six architectural domains of 23S rRNA. Red dotted lines and black arrows indicate axis and directions of observed global movements, respectively. Stabilizing contacts in closed conformations are indicated by black arrows. Interacting regions are labeled. Cryo-EM maps (transparent gray) and PDB models corresponding to rRNA helices H21 and H22 (blue), H57-58 (green), or H63 (orange), respectively. Local

snapshots of state *C* (e, h) and state *C-CP_{L2/L28}* (f, i) contain thumbnails with red rectangles indicating the region of interest. g, j close-ups of regions of interest in state *C-CP_{L2/L28}*. Yellow circles highlight H-bond interactions (light blue dotted lines) between C385-G411 (g) and G1516-C1732 (j). k, l State *d126-CP* exhibiting flexibility in the CP occupying a more upright (k) or more inclined position towards the interface (l). Binding sites of L-proteins uL2 (state *C_{L2}*) (m) and bL35 (state *C-CP-H68*) (n).

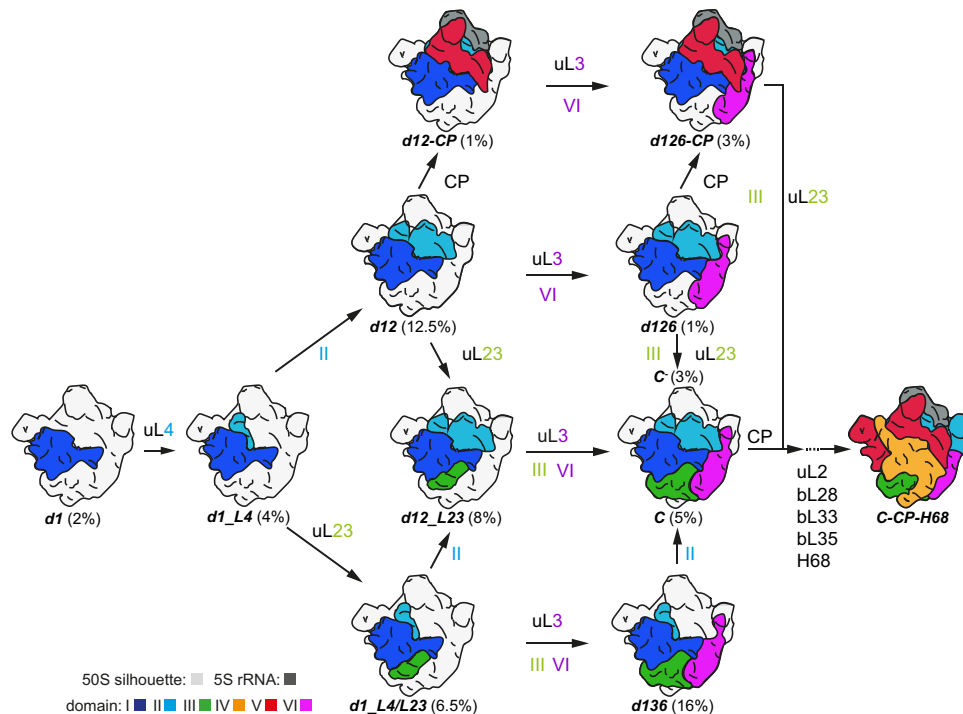


Fig. 6 | Model of early 50S assembly. Models of individual states color-coded as indicated. Numbers in % represent their relative abundance in the sorting scheme. Gray silhouette illustrates density of the mature 50S subunit. The labels uL4, uL23,

uL3 and the numbers II, III, and VI indicate when cryo-EM densities for parts of these proteins or entire 23S rRNA domains, respectively, become detectable. CP central protuberance; H68 helix 68.

Accordingly, early 50S assembly is characterized by hierarchical arrangements of rRNA, resulting in branches, but on the other hand, generates cooperativity and ultimately culminates in convergence along the assembly pathway. Due to the interwoven architecture and the ring-shaped arrangement of domains I, II, III, and VI, domain VI can be incorporated via routes 1-2-6 or 1-3-6. While uL24 promotes the formation of the 5'-domain I, the late assembling uL3 plays a critical role in the structural organization of the 3'-end-containing domain VI. As we do not observe a state consisting only of domains I and VI, their contact areas may be too small to drive stable incorporation of domain VI in the absence of either domain II or III. This suggests that domain I together with domain VI cradle the missing domain in *d126*, or *d136*, to promote the appearance of domain III or II, respectively, to conclude core formation and early assembly (Fig. 6). Thus, our analysis indicates that coordinated structural formation of the two terminal domains of 23S rRNA is a decisive event that drives further assembly.

Structural data regarding earliest stages of large subunit assembly are lacking from both the prokaryotic and the eukaryotic domain. Interestingly, a structural analysis of nucleolar pre-60S assembly captured a state, consisting of 28S rRNA domains I and II (similar to our state *d12*), that was thought to proceed linearly to a 60S subunit²⁶. Surprisingly, a second study found evidence for parallel assembly pathways²⁷. While we find similar states, confirming the evolutionary conservation of the process, our analysis starts with a particle whose assembly has just initiated (*d1*). In addition, we can take advantage of the multitude of states we were able to obtain after thorough sorting and trace how decisive steps are taken towards a mature 50S subunit, involving multiple parallel routes of assembly.

We took advantage of the fact that the *in vitro* reconstitution occurs in the absence of ribosome assembly factors. Hence, precursor states approaching at rate limiting steps of the reaction are not as rapidly processed as is the case *in vivo* and *in vitro* systems utilizing assembly factors. Dong et al.³⁷ used a system to study 50S assembly that involves *in vitro* transcribed ribosomal rRNA (iSAT) and combined it with a cryo-EM analysis. Their setup is supposed to be closer to the

in vivo situation, since in addition to vectorially produced rRNA, it also utilizes S150 extract containing all relevant ribosome assembly factors. Interestingly their analysis yielded roughly the same structural states, suggesting that the states we obtained are of physiological relevance. As our system enriches certain states (such as *d12*, *d12_L23*, and *d136*), we have been able to refine highly resolved maps of these interesting early pre-50S intermediates.

In addition, we noticed that the assembly pathways that we describe structurally are in very good agreement with the Nierhaus assembly map of the 50S subunit³⁸ (Supplementary Fig. 5b).

Taken together, the combination of time-limited *in vitro* 50S assembly and cryo-EM allowed us to obtain unprecedented insights into the early assembly phase of the bacterial large ribosomal subunit. However, to obtain a more detailed understanding of the assembly process, fluorescence-based real time experiments and approaches clarifying the mechanistic role of ribosome assembly factors will be required.

Methods

50S *in vitro* reconstitution

The assay was performed according to³⁹ with the modifications described in¹⁴ as follows. For the step 1 reaction of the 50S reconstitution, seven samples were prepared in individual tubes corresponding to 7 time points. In each of the tubes, 6.25 equivalent units of TP50 (total protein of the 50S subunit derived from *E. coli* CAN20-12E (RNase I II⁻ D⁻ BN⁻)⁴⁰) and 6.25 A₂₆₀ rRNA (purified 23S rRNA + 5S rRNA) were mixed on ice and incubated at 44 °C using a thermo-block (Eppendorf, Thermomixer comfort) in Rec4 buffer (20 mM HEPES/KOH pH7.6, 400 mM NH₄Ac, 4 mM Mg(Ac)₂, 0.2 mM EDTA, 5 mM 2-mercaptoethanol). A first aliquot was taken instantaneously (0 min), and further samples were taken after 2, 3, 5, 10, 20, and 30 min of incubation. All reactions were stopped by placing the individual tubes on ice. For an activity assay (*in vitro* translation assay), 36 pmol of material from the individual time points were adjusted to 20 mM Mg(Ac)₂ concentration and incubated for 90 min at 50 °C.

Analytical sucrose density gradient ultra-centrifugation

18 pmol material from the individual time course were subjected to analytical sucrose density gradient (10–30%) ultra-centrifugation (SW40 rotor, 26,000 rpm (-85,000 g) for 17 h) in Tico buffer (20 mM HEPES/KOH pH7.6 on ice, 30 mM KAc, 6 mM Mg(Ac)₂, 4 mM 2-mercaptoethanol).

In vitro translation assay (poly(U)-dependent poly(Phe) assay)

6 pmol isolated native 50S subunits (50S), reconstituted 50S subunits (50S rec) from each time point of step 1 that had been subjected to a full step 2 incubation, were mixed with 12 pmol 30S subunits and incubated in assay buffer (20 mM HEPES/KOH pH7.6, 150 mM NH₄Ac, 6 mM Mg(Ac)₂, 0.05 mM EDTA, 4 mM 2-mercaptoethanol, 0.05 mM spermine, 2 mM spermidine, 3 mM ATP, 1.5 mM GTP, 5 mM acetyl-phosphate, 83 μM [¹⁴C] Phe (22 dpm/pmol) (Hartmann analytic GmbH), 0.83 mg/ml poly(U), 0.34 mg/ml tRNA^{bulk} and S100 extract) for 60 min at 37 °C. Subsequently, 30 μl BSA (1% stock solution) and 2 ml TCA (10%) were added. Samples were briefly vortexed, incubated for 15 min at 90 °C and 5 min on ice. Precipitated proteins were immobilized on glass filters, washed two times with 5% TCA and once with ether/ethanol (1:1). The filters were incubated in scintillation liquid (Roth, Rotiszint eco plus) and the amount of incorporated ¹⁴C Phe was determined using a scintillation counter (Wallac 1409 liquid scintillation counter). The assay was performed in duplicates and mean values were calculated.

Quantitative mass spectrometry analysis

108 pmol of the individual samples from the step 1 time course reaction were adjusted to total volume 180 μl and loaded on 900 μl sucrose cushion (20% in Tico buffer). Ultra-centrifugation was performed (TLA-110 rotor, 65,000 rpm (-151,000 g) for 2.5 h). After removing the supernatant, the pellet was briefly washed and dissolved in Tico buffer. 36 pmol of each sample were analyzed by quantitative mass spectrometry according to the workflow described in ref. ¹³. For protein comparisons per sample, iBAQ intensities were normalized to uL24, which is an early assembly protein of the large bacterial ribosomal subunit. Fold changes to the reference purified mature 50S was calculated from normalized iBAQ values. The ribosomal stoichiometry of individual L-proteins in each sample was calculated as the percentual fraction of the value of corresponding L-proteins in mature 50S. Therefore, 100% stoichiometry represents a protein that is as abundant as its counterpart in the mature 50S subunit. Experiments were performed in duplicates and data were processed with Heatmap Illustrator 1.0 (CUCKOO Workgroup) and Origin8 (OriginLab Corporation).

Sample preparation for cryo-EM

12 pmol of the 3 min sample purified via sucrose cushion ultra-centrifugation, as described above, were diluted twofold in adaptation buffer (20 mM HEPES/KOH pH7.6, 4 mM Mg(Ac)₂, 0.2 mM EDTA, 5 mM 2-mercaptoethanol) to achieve a final concentration of 288 pmol/ml and spotted on glow-discharged holey carbon grids (RI.2/1.3 copper 400 mesh holey carbon, without additional carbon film on top (carbon-free)), (Quantifoil Micro Tools GmbH) and cryo plunged in liquid ethane after blotting using a Vitrobot device (MK4). For structural analyses of ribosome assembly, we generally use grids without thin carbon to minimize a preferred orientation of the particles.

Cryo-electron microscopy and data processing

Data were collected on a FEI TecnaiG2 Polara equipped with a Gatan K2 Summit detector operated in super-resolution mode at 300 kV at a calibrated pixel size of 0.625 Å. Movies were acquired for 10 s applying a total electron dose of 63 e⁻/Å².

Data processing

Movies were aligned and dose-weighted using Defocus values were estimated using Gctf⁴¹. Templates for particle picking were generated in SPIDER⁴². Therefore, density maps were generated from atomic models of state 1 (PDB: 6GC7) and state 5 (6GBZ), subsequently low pass filtered to 20 Å, followed by projection into 84 equally distributed orientations, respectively. Orientation images were averaged into four projections using Xmipp3 2D classification⁴³ classification for each template, resulting in eight distorted averages (diameter: 140–230 Å) for particle picking. Particles were picked using templates in Gauto-match (developed by K. Zhang). Particle images were extracted and normalized, using Relion 3.0⁴⁴ with a box size of 600 and Fourier cropped to 150 for sorting. If not stated otherwise, Cryosparc v3.1⁴⁵ was used for the identification and refinement of final classes. Initial sorting was achieved using ab-initio classification, followed by two rounds of heterogeneous refinement to recover ribosomal particles (Supplementary Fig. 2a). Remarkably, recovered ribosomal particle classes exhibited density for the architectural domains 1 and 2 of the 23S rRNA only in the first, and density for domain 1 only in the second heterogeneous refinement. Selected particles were then pooled and aligned to a consensus map using Homogenous refinement. Remaining non-ribosomal particles were sorted out using reference-free 2D classification. Initial identification of different 50S precursor classes was achieved using 3D classification without alignment in Relion 3.1. Therein, best classification results were achieved, when particles were locally aligned to domain 1 of the 23S rRNA using Local refinement in prior. The 3D classification resulted in six distinct classes (*dI*, *dI2*, *dI36*, *C*, *dI26-CP*, *C-CP*). All particles were re-aligned and re-assigned using identified classes as templates and heterogeneous refinement (Supplementary Fig. 2b). An additional 3D ice template was generated as non-ribosomal bait template by ab-initio reconstruction of respective classes from the 2D classification. Re-assigned particles were refined and remaining structural heterogeneity was sorted using hierarchical 3D variability clustering. When necessary, additional rounds of multi-class ab-initio reconstruction (class similarity = 0) were performed, and non-structured maps were used as bait templates in one round of heterogeneous refinement to sort out shiny particles. The most mature precursor, *C-CP*-H68, still showed marked fragmented density for uL16, bL35, and H68, indicating remaining structural heterogeneity. However, due to low particle numbers, it was not sorted further. After completion of particle classifications, sub-classes were refined using non-uniform refinement at a pixel size of 1.25 Å.

Model building

Atomic models of previously identified 50S assembly intermediates (PDB: 6GC0, 6GC4, 6GC7, 6GC8) were used for modeling. Initially, models were rigid body docked in ChimeraX⁴⁶, followed by rigid body fitting of individual L-proteins. RNA regions and L-proteins were removed when densities were missing or strongly fragmented. The adjustment of structured elements was performed by iterative model building and real-space refinement into the EM-density using Coot 0.9.6⁴⁷, Phenix 1.19⁴⁸ and ERRASER⁴⁹ considering secondary structure restraints. Models for maps with resolutions above 4 Å were subjected to a final round of geometry minimization in Phenix for restoration of geometry of the RNA. For the analysis of L-protein mediated sampling (Fig. 3), full-length models of respective L-proteins were fitted into EM maps of indicated states. A 50S Cryo-EM map (EMD-222614) was used for comparison⁵⁰. Models were colored according to any rRNA contacts the corresponding 23S architectural domains within 4 Å distance. Self-contacts and surface exposure were colored in gray. Local Cryo-EM maps are shown at 3 Å distance to atomic models using the ChimeraX zone tool. For better visualization, noise and small fragmented regions were excluded by using the surface dust tool.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Cryo-EM density maps and atomic models are stored in EMDB and PDB as follows: *dI* (EMD-16509/ PDB ID 8C9C), *dI_L4* (EMD-16508/ PDB ID 8C9B), *dI_L4/L23* (EMD-16507/ PDB ID 8C9A), *dI2* (EMD-16506/ PDB ID 8C99), *dI2_L23* (EMD-16505/ PDB ID 8C98), *dI36* (EMD-16504/ PDB ID 8C97), *dI26* (EMD-16503/ PDB ID 8C96), *dI2-CP* (EMD-16502/ PDB ID 8C95), *dI26-CP* (EMD-16501/ PDB ID 8C94), *C* (EMD-16500/ PDB ID 8C93), *C* (EMD-16499/ PDB ID 8C92), *C_L2* (EMD-16498/ PDB ID 8C91), *C-CP* (EMD-16497/ PDB ID 8C90), *C-CP_L28* (EMD-16496/ PDB ID 8C8Z), *C-CP_L28/L2* (EMD-16495/ PDB ID 8C8Y), *C-CP-H68*: (EMD-16494/ PDB ID 8C8X). Mass spectrometry proteomics data have been deposited to ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository with the dataset identifier PXD030312. All additional data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Source data are provided with this paper.

References

- Kaczanowska, M. & Ryden-Aulin, M. Ribosome biogenesis and the translation process in *Escherichia coli*. *Microbiol. Mol. Biol. Rev.* **71**, 477–494 (2007).
- Shajani, Z., Sykes, M. T. & Williamson, J. R. Assembly of bacterial ribosomes. *Annu. Rev. Biochem.* **80**, 501–26 (2011).
- Stern, S., Powers, T., Changchien, L. I. M. & Noller, H. F. RNA-protein interactions in 30S ribosomal subunits: folding and function of 16S rRNA. *Science* **244**, 783–90 (1989).
- Rodgers, M. L. & Woodson, S. A. A roadmap for rRNA folding and assembly during transcription. *Trends Biochem. Sci.* **46**, 889–901 (2021).
- Noller, H. F. et al. Secondary structure model for 23S ribosomal RNA. *Nucleic Acids Res.* **9**, 6167–6189 (1981).
- Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**, 905–20 (2000).
- Harms, J. et al. High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell* **107**, 679–88 (2001).
- Li, N. et al. Cryo-EM structures of the late-stage assembly intermediates of the bacterial 50S ribosomal subunit. *Nucleic Acids Res.* **41**, 7073–7083 (2013).
- Jomaa, A. et al. Functional domains of the 50S subunit mature late in the assembly process. *Nucleic Acids Res.* **42**, 3419–35 (2014).
- Ni, X. et al. YphC and YsxC GTPases assist the maturation of the central protuberance, GTPase associated region and functional core of the 50S ribosomal subunit. *Nucleic Acids Res.* **44**, 8442–55 (2016).
- Wang, W. et al. Loss of a single methylation in 23S rRNA delays 50S assembly at multiple late stages and impairs translation initiation and elongation. *Proc. Natl Acad. Sci. USA* **117**, 15609–15619 (2020).
- Davis, J. H. et al. Modular assembly of the bacterial large ribosomal subunit. *Cell* **167**, 1610–1622.e15 (2016).
- Nikolay, R. et al. Snapshots of native pre-50S ribosomes reveal a biogenesis factor network and evolutionary specialization. *Mol. Cell* **81**, 1200–1215.e9 (2021).
- Nikolay, R. et al. Structural visualization of the formation and activation of the 50S ribosomal subunit during in vitro reconstitution. *Mol. Cell* **70**, 881–893.e3 (2018).
- Kater, L. et al. Construction of the central protuberance and L1 stalk during 60S subunit biogenesis. *Mol. Cell* **79**, 615–628.e5 (2020).
- Lancaster, L., Lambert, N. J., Maklan, E. J., Horan, L. H. & Noller, H. F. The sarcin-ricin loop of 23S rRNA is essential for assembly of the functional core of the 50S ribosomal subunit. *RNA* **14**, 1999–2012 (2008).
- Wu, S. et al. Diverse roles of assembly factors revealed by structures of late nuclear pre-60S ribosomes. *Nature* **534**, 133–137 (2016).
- Kargas, V. et al. Mechanism of completion of peptidyltransferase centre assembly in eukaryotes. *Elife* <https://doi.org/10.7554/eLife.44904> (2019).
- Zhou, Y., Musalgaonkar, S., Johnson, A. W. & Taylor, D. W. Tightly-orchestrated rearrangements govern catalytic center assembly of the ribosome. *Nat. Commun.* **10**, 958 (2019).
- Chandrasekaran, V. et al. Visualizing formation of the active site in the mitochondrial ribosome. *Elife* **10**, e68806 (2021).
- Cheng, J., Berninghausen, O. & Beckmann, R. A distinct assembly pathway of the human 39S late pre-mitoribosome. *Nat. Commun.* **12**, 4544 (2021).
- Cipullo, M., Gesé, G. V., Khawaja, A., Hällberg, B. M. & Rorbach, J. Structural basis for late maturation steps of the human mitoribosomal large subunit. *Nat. Commun.* **12**, 3673 (2021).
- Hillen, H. S. et al. Structural basis of GTPase-mediated mitochondrial ribosome biogenesis and recycling. *Nat. Commun.* **12**, 3672 (2021).
- Jaskolowski, M. et al. Structural Insights into the Mechanism of Mitoribosomal Large Subunit Biogenesis. *Mol. Cell* **79**, 629–644.e4 (2020).
- Lenarčič, T. et al. Stepwise maturation of the peptidyl transferase region of human mitoribosomes. *Nat. Commun.* **12**, 3671 (2021).
- Kater, L. et al. Visualizing the assembly pathway of nucleolar pre-60S ribosomes. *Cell* **171**, 1599–1610.e14 (2017).
- Sanghai, Z. A. et al. Modular assembly of the nucleolar pre-60S ribosomal subunit. *Nature* **556**, 126–129 (2018).
- Nierhaus, K. H. & Dohme, F. Total reconstitution of functionally Active 50S Ribosomal Subunits from *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **71**, 4713–4717 (1974).
- Nierhaus, K. H. The assembly of prokaryotic ribosomes. *Biochimie* **73**, 739–755 (1991).
- Duss, O., Stepanyuk, G. A., Puglisi, J. D. & Williamson, J. R. Transient protein-RNA interactions guide nascent ribosomal RNA folding. *Cell* **179**, 1357–1369.e16 (2019).
- Rodgers, M. L. & Woodson, S. A. Transcription increases the cooperativity of ribonucleoprotein assembly. *Cell* **179**, 1370–1381.e12 (2019).
- Zhang, X. et al. Structural insights into the function of a unique tandem GTPase EngA in bacterial ribosome assembly. *Nucleic Acids Res.* **42**, 13430–13439 (2014).
- Nicol, S. M. & Fuller-Pace, F. V. The ‘DEAD box’ protein DbpA interacts specifically with the peptidyltransferase center in 23S rRNA. *Proc. Natl Acad. Sci. USA* **92**, 11681–11685 (1995).
- Arai, T. et al. Single methylation of 23S rRNA triggers late steps of 50S ribosomal subunit assembly. *Proc. Natl Acad. Sci. USA* <https://doi.org/10.1073/pnas.1506749112> (2015).
- Feng, B. et al. Structural and functional insights into the mode of action of a universally conserved Obg GTPase. *PLoS Biol.* **12**, e1001866 (2014).
- Traub, P. & Nomura, M. Structure and function of *E. coli* ribosomes. V. Reconstitution of functionally active 30S ribosomal particles from RNA and proteins. *Proc. Natl Acad. Sci. USA* **59**, 777–784 (1968).
- Dong, X. et al. Near-physiological in vitro reconstitution of 50S subunit reveals parallel assembly pathway. *Biophys. J.* **122**, <https://doi.org/10.1016/j.bpj.2022.11.2620> (2022).
- Herold, M. & Nierhaus, K. H. Incorporation of six additional proteins to complete the assembly map of the 50S subunit from *Escherichia coli* ribosomes. *J. Biol. Chem.* **262**, 8826–8833 (1987).

39. Dohme, F. & Nierhaus, K. H. Total reconstitution and assembly of 50 S subunits from *Escherichia coli* ribosomes in vitro. *J. Mol. Biol.* **107**, 585–599 (1976).
 40. Deutscher, M. P., Marlor, C. W. & Zaniewski, R. Ribonuclease T: new exoribonuclease possibly involved in end-turnover of tRNA. *Proc. Natl Acad. Sci. USA* **81**, 4290–4293 (1984).
 41. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
 42. Frank, J. et al. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–9 (1996).
 43. de la Rosa-Trevin, J. M. et al. Xmipp 3.0: an improved software suite for image processing in electron microscopy. *J. Struct. Biol.* **184**, 321–328 (2013).
 44. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, e42166 (2018).
 45. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. CryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
 46. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
 47. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–32 (2004).
 48. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–21 (2010).
 49. Chou, F. C., Sripakdeevong, P., Dibrov, S. M., Hermann, T. & Das, R. Correcting pervasive errors in RNA crystallography through enumerative structure prediction. *Nat. Methods* **10**, 74–76 (2013).
 50. Watson, Z. L. et al. Structure of the bacterial ribosome at 2 Å resolution. *Elife* **9**, 1–62 (2020).
- data and built atomic models. A.B. and P.S. evaluated and improved atomic models; C.H.V.-V. and M.S. performed qMS and analyzed the data. J.B. and T.M. participated in the grid preparation, operation of the microscopes, data acquisition, and processing; C.M.T.S. and R.N. supervised the study. R.N. drafted the manuscript. All authors contributed to writing of the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-36607-9>.

Correspondence and requests for materials should be addressed to Christian M. T. Spahn or Rainer Nikolay.

Peer review information *Nature Communications* thanks Joseph Davis and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

Acknowledgements

We would like to thank Anett Unbehaun, Timo Fluegel and Matthew Kraushar for intensive scientific discussions, excellent suggestions, and text editing. This work was funded by Bundesministerium für Bildung und Forschung (BMBF 16GW0300 to C.M.T.S.) and the Human Frontier Science Program Organization (HFSP-Ref. RPG0008/2014, to C.M.T.S.). In addition, this work was supported by the Deutsche Forschungsgemeinschaft (DFG) through the cluster of excellence Unifying Systems in Catalysis (UniSysCat) under Germany's Excellence Strategy-EXC 2008-390540038 (to A.B., C.M.T.S., and P.S.).

Author contributions

B.Q. performed all biochemical experiments and prepared sample for cryo-EM and qMS; S.M.L. processed, analyzed, and visualized structural