Cryo-EM reconstructions of LiCl core particles show strong similarities between disassembly and assembly of the ribosomal 50S subunit

Daniel S. D. Larsson, Sandesh Kanchugal P. & Maria Selmer*

Department of Cell and Molecular Biology, Uppsala University, BMC, Box 596, SE-751 24 Uppsala, Sweden

* To whom correspondence should be addressed. Tel: +46 18 471 4177; Email: maria.selmer@icm.uu.se

Running title: Cryo-EM structures of 50S LiCl core particles

Keywords: RNA structure, RNA folding, RNA-protein interactions, ribosome assembly, cryo-EM

ABSTRACT

Bacterial ribosome assembly is a complex process. Purified 50S ribosomes subjected to high-salt wash removing a subset of ribosomal proteins (r-proteins), were early shown competent for *in vitro* assembly into functional 50S subunits. We here used cryo-EM to determine the structure of such LiCl core particles derived from *E. coli* 50S subunits. A wide range of complexes with large variation in extent of ordered 23S rRNA and occupancy of r-proteins could be identified, and resolved to between 2.8 Å and 9 Å resolution. Many of these particles showed high similarity to *in vivo* and *in vitro* assembly intermediates, supporting the inherent stability or metastability of these states. Similar to states in early ribosome assembly, the main class showed ordered density for 23S rRNA domains 0, I, II, III, VI and the 5'-half of domain IV. In addition, smaller core particles were discovered, which may mimic earlier stages of 50S biogenesis. Our data support a multi-pathway disassembly process based on independent folding blocks, similar but reverse to the assembly process. The study provides examples of dependencies between complex tertiary RNA structure and RNA-protein interactions, exemplifies the need for proteins to stabilize native RNA structure also after folding, and supports the relevance of salt-washed core particles as a mimics of early ribosome assembly intermediates.

INTRODUCTION

The ribosome is a large macromolecular complex performing translation of mRNA into protein in all domains of life. Its intricate assembly is in all organisms a tightly regulated process that has been fine-tuned by evolution. In *E. coli*, fifty-four ribosomal proteins (r-proteins) and three ribosomal RNAs (rRNAs) fold and assemble with the help of a multitude of auxiliary RNA modification enzymes and ribosome assembly factors (reviewed in (Shajani et al. 2011; Rodgers and Woodson 2021)). Despite the complexity of the process, the assembly time for a mature ribosome *in vivo* in exponentially dividing *E. coli* has been estimated to only 2 minutes (Chen et al. 2012).

The structure and assembly of the 50S subunit has been extensively studied *in vitro*, and methods were early developed to reconstitute active subunits from components (Nierhaus and Dohme 1974). During *in vitro* reconstitution, the order of r-proteins associating to the rRNA was found to approximately agree with the order *in vivo* (Herold and Nierhaus 1987; Chen and Williamson 2013). Another *in vitro* approach was to promote the dissociation of r-proteins through high-salt (using *e.g.* LiCl) washing of ribosomal particles. Proteins were shown to dissociate from the core at defined salt concentrations (Nierhaus and Dohme 1974; Nierhaus 1990). We now know that the observed order of leaving r-proteins nearly agrees with the reverse of the *in vivo* assembly order as followed by quantitative mass spectrometry (Chen and Williamson 2013) as well as the *in vitro* reconstitution process (Spillmann et al. 1977). A subset of the of the ribosomal proteins that dissociate during salt wash have also been demonstrated to exchange *in vivo*, presumably as a mechanism by which damaged proteins can be continuously replaced (Pulk et al. 2010).

Our structural understanding of folding and assembly of ribosomes has greatly benefitted from progress in cryogenic electron microscopy (cryo-EM), where classification schemes in single-particle reconstruction (SPR) methods have proven essential to study these heterogeneous particle ensembles. Strategies to enrich *in vivo* ribosome assembly intermediates for these studies include treatment of bacteria with drugs that inhibit ribosome

assembly (Dagley and Sykes 1959; Siibak et al. 2011), knock-down of an r-protein or an assembly factor (Davis et al. 2016; Wang et al. 2020; Seffouh et al. 2019; Zhong et al. 2021) and pull-downs with assembly factors as bait (Zhou et al. 2019; Nikolay et al. 2021).

In one such study, a bL17-depletion strain was used to generate a multitude of ribosomal 50S sub-particles, which could be classified into 4 major classes and a dozen sub-classes (Davis et al. 2016). Based on a detailed comparison of the presence of rRNA helices and ribosomal proteins in the different classes, several cooperative folding blocks were identified. In agreement with previous data showing that ribosome assembly can take different paths upon perturbation of r-protein levels (Siibak et al. 2011), a model of three major routes of LSU folding upon bL17 perturbation was proposed, supporting the robustness of ribosome assembly. Along the different pathways, the central protuberance (CP), the particle base and the L7/L12 stalk formed in different order.

Structural studies of assembly intermediates from LSU *in vitro* reconstitution (Nikolay et al. 2018) showed five distinct precursors that shared many features with the enriched *in vivo* assembly intermediates in the bL17-depleted strain. The later intermediates displayed an intricate variety of folding states of the protuberances and the peptidyl transfer center (PTC).

The salt washed LiCl core particles have been demonstrated to be assembly-competent, allowing reconstitution into active 50S subunits (Nierhaus and Dohme 1974; Spillmann et al. 1977). They have also been shown to work as an *in vitro* substrate for RlmF, an early rRNA modifying enzyme that lacks activity on naked RNA (Sergiev et al. 2008), supporting at least local similarity to states occurring during ribosome assembly *in vivo*. There is to date no high-resolution structure of these particles, but early negative-stain electron microscopy of LiCl washed ribosome preparations showed high degree of heterogeneity where approximately 40% of the ribosomes remained as compact particles lacking the three characteristic protuberances of the LSU (Boublik et al. 1978).

We here set out to further characterize the structure of LiCl core particles using single-particle cryo-EM, with the goal to obtain high-resolution maps and characterize the structure and composition of these particles. We identified 6 major types of 50S sub-particles with a

wide range of sizes–as well as a multitude of variants of each of these. The method allowed observation of core particles smaller than so far identified 50S assembly intermediates and the opportunity to examine effects of removing particular r-proteins. Comparison of these particles enabled us to identify stability dependencies between rRNA secondary structure elements and r-proteins, to reconstruct possible disassembly and assembly pathways and to demonstrate strong similarities between intermediates formed during disassembly and assembly.

MATERIAL AND METHODS

Preparation of LiCl core particles

Ribosomes were purified from strain JW5107 (*ΔrlmF*/*ΔybiN*) of the Keio collection (Baba et al. 2006). For the preparation of tight-coupled 70S ribosomes, the protocol of (Rodnina and Wintermeyer 1995) was used with modifications. Frozen *E. coli* cell pellets (24 g, wet weight) were resuspended in 100 mL of opening buffer containing DNase I (Sigma-Aldrich, USA, 5 µg/ml) and cOmplete EDTA-free protease inhibitor cocktail (Roche, Germany) and the cells were opened using a CF Cell Disrupter (Constant Systems Ltd, UK). The lysate was cleared by centrifugation (Sorvall SS-34 at 16000 RPM for 45 minutes) and filtration (0.45 µm pore size). Crude ribosomes were obtained by pelleting (Beckmann Type 45Ti at 40,000 RPM for 18 hours) 25 ml fractions of lysate through 27 ml cushions of sucrose buffer (with 10.5 mM $Mg(OAc)_2$). The pellets were dissolved in wash buffer (same as the sucrose buffer but with, 7 mM β-mercaptoethanol and no sucrose) and sedimented through sucrose cushions a second time (25 ml of ribosomes and 1.5 ml of sucrose buffer in a Beckmann SW32Ti at 24,800 RPM for 16 hours). The crude ribosomes were separated on pre-formed linear 15–30% sucrose gradients (Beckmann SW32Ti at 19,300 RPM for 17 hours) in TC buffer (buffer C with 20 mM Tris-HCl and 5.25 mM $Mg(OAc)_2$) and the 70S peak was pooled and pelleted (Beckmann Type 45Ti at 45,000 RPM for 18 hours). To obtain individual subunits, 70S

pellets were resuspended in SU buffer (TC buffer but with 3 mM $Mg(OAc)_2$) and subunits separated on pre-formed linear 15–30% sucrose gradients in SU buffer (Beckmann SW32Ti at 19,300 RPM for 16 hours). The 50S subunit fractions were pooled and pelleted (Beckmann Type 45Ti at 45,000 RPM for 18 hours). The pellets were resuspended in LiCl-wash buffer (10 mM Hepes-KOH pH 7.5, 10 mM $Mg(OAc)_2$, 3.5 M LiCl) and incubated for 5 hours on ice. The 50S LiCl core particles were pelleted (Beckmann Type 90Ti at 44,000 RPM for 4 hours), resuspended in SU buffer and separated on pre-formed linear 15–30% sucrose gradients in SU buffer (Beckmann SW32Ti at 19,900 RPM for 16 hours). The 50S LiCl core particle fractions were pooled, pelleted (Beckmann Type 90Ti at 45,000 RPM for 17 hours) and resuspended in storage buffer (20 mM Tris-HCl pH 7.5, 50 mM NH_4Cl, 5 mM $Mg(OAc)_2$, 3 mM β-mercaptoethanol). The concentration of 50S LiCl core particles was measured by 260 nm absorbance (assuming 1 A_{260} unit equals 38 nM) and aliquots were flash frozen and stored at -80 °C.

In vitro methylation assay

Assays were performed in duplicates. Each reaction contained 30 pmol 50S-LiCl particles in reaction buffer (50 mM HEPES-KOH pH 7.5, 20 mM NH_4Cl, 100 mM KCl, 2.5 MgOAc, 8 mM β-mercaptoethanol). Addition of 400 pmol unlabeled S-Adenosyl-L-methionine (Sigma-Aldrich, USA) doped with 16 pmol S-[methyl-^{3}H]-adenosyl-L-methionine (4 Ci/mmol; PerkinElmer, USA) was followed by the addition of 40 pmol RlmF, or buffer (negative control) to a final volume of 30 µl. All reactions were incubated at 37° C for up to 30 min, quenched in 2 ml ice cold 10% trichloroacetic acid and incubated on ice for 10 min. The precipitated reactions were applied to BA85 nitrocellulose filters (Whatman, UK) under vacuum and washed five times with 7 ml of cold 10% trichloroacetic acid. The washed filters were placed in vials containing 5 ml of Filter Safe scintillation cocktail (Schleicher & Schuell, Germany), shaken for 30 min and counted in an LC6500 scintillation counter (Beckman, USA).

Cryo-EM sample preparation

A 300-µl reaction containing 5.2 nmol of C-terminally His-tagged RlmF and 1.2 nmol of 50S LiCl core particles in imaging buffer (50 mM Hepes-KOH pH 7.5, 100 mM NH_4Cl, 10 mM $Mg(OAc)_2$, 0.5 mM EDTA-Tris, 6 mM β-mercaptoethanol, 0.5 mM sinefungin (Sigma-Aldrich, USA), 0.5 U/ml of RiboLock RNase inhibitor (ThermoFisher Scientific, USA)) was incubated with 200 uL of Ni-Sepharose for 30 minutes at 25 °C under gentle agitation in a spin-filter and, subsequently, centrifuged at 0.5 RFC for 3 minutes. The Ni-Sepharose beads were washed twice by incubation for 10 minutes in 500 µl of imaging buffer with 0.2 mM sinefungin, 0.1 U/ml of RiboLock RNase inhibitor and 30 mM imidazole, and the ribosome-RlmF complex eluted in 200 uL of imaging buffer with 0.2 mM sinefungin, 0.1 U/ml of RiboLock RNase inhibitor and 300 mM imidazole. The eluate was sedimented in 50-uL aliquots through 50 µl 50% sucrose cushions in imaging buffer (Sorvall RC M150GX for 90 minutes at 80,000 RPM at 4 °C). The pellets were resuspended in 10 uL of imaging buffer and stored on ice before being flash frozen on cryo-EM grids. The protein content was verified by SDS-PAGE and LC-MS/MS analysis of the lane cut from the SDS-PAGE gel. Sample preparation is summarized in Supplementary Figure S1.

For the LC-MS/MS analysis, the sample was digested by trypsin and the peptides were separated in reverse-phase on a C18-column with a 90 minute gradient and electrosprayed to a Q Exactive Orbitrap mass spectrometer (Thermo Finnigan). Tandem mass spectrometry was performed by High-Energy Collision Dissociation fragmentation. The data was analysed using Proteome Discoverer 1.4 (Thermo Fisher Scientific) using the Sequest algorithm against a database of the *E. coli* K-12 BW25113 proteome (the parental strain to the Keio collection). The search criteria for protein identification were set to at least two matching peptides of 95% confidence level per protein.

Cryo-EM grid freezing and data collection

The cryo-EM specimens were prepared on QuantiFoil R 2/2 grids with 2-nm continuous carbon (QuantiFoil Micro Tools GmbH, Germany). The grids were glow-discharged for 40 seconds at 20 mA and 0.38 mBar using an EasiGlow (Ted Pella Inc, USA). Plunge-freezing

was done using a Vitrobot mark IV (ThermoFisher Scientific, USA) using 3 µl of sample at 160 nM ribosome concentration after 30 seconds of incubation on the grid at 100% humidity and 10 °C and blotting for 3.5 seconds. Data was collected at two different occasions on the same Titan Krios microscope from two different grids from the same batch. The microscope was equipped with a Falcon-III direct electron detector used in integrating mode. The nominal magnification was 75,000x (corresponding to a pixel-size of 1.09 Å in the specimen plane) and the acceleration tension was 300 kV. At the first occasion, movies were collected with 66.2 e^-/pixel/second during 0.773 seconds for a total dose of 41.0 $e^-/Å^2$ and the sample stage set to either 0° (3,052 movies) or 15° (2,196 movies) tilt. At the second session, movies were collected with 68.4 e^-/pixel/second during 0.770 seconds for a total dose of 42.2 $e^-/Å^2$ and the sample stage set to 30° tilt (6,120 movies). Data collection parameters are summarized in Supplementary Table S1.

Cryo-EM data processing

The 3D single-particle reconstruction was done using RELION version 3.0 (Zivanov et al. 2018) following the conventional RELION workflow (Scheres 2016). The 0°, 15° and 30° tilt micrographs were initially processed separately and later combined. Movies were motion corrected in 5x5 patches using the RELION implementation of the MotionCor2 algorithm (Zheng et al. 2017). Per-micrograph defocus parameters were estimated using gCTF version 1.18 (Zhang 2016). Particle picking was done using the Laplacian-of-Gaussian algorithm in RELION. Local CTF parameters were estimated for the tilted micrographs using gCTF. An *ab initio* 3-D initial model was generated from primary data to prevent model bias. Multiple rounds of 2-D and 3-D classifications were performed to prune the data down to a set of 384,374 particles. Automatic 3D refinement, per-particle refinement of CTF parameters and particle polishing were performed several times until no further improvement in resolution could be achieved. RELION version 3.1 (Zivanov et al. 2020) was then used to estimate higher order aberrations and anisotropic magnification, followed by further CTF-refinement and particle polishing, to produce the final 2.84-Å consensus reconstruction. The resolution

was estimated by calculating the resolution shell in Fourier space where the correlation between two independently refined half-sets of particles is 0.143 inside of a soft-edged mask. The mask was generated by masking the final reconstruction by a 155-Å spherical mask to remove artifacts at the rim of the reconstruction followed by low-pass filtering to 15 Å, thresholding that map at a sufficiently low value to encompass diffuse density associated with the particle map but excluding solvent noise (0.012), expanding the mask first by 5 pixels, and finally adding a 5 pixel soft edge. Estimation of local resolution and local low-pass filtering was performed using the RELION algorithm using a B-factor determined by fitting to an approximately linear regime at high resolution of the Guinier plot (-32.2 $Å^2$). The efficiency of the orientation distribution (E_{od}) parameter was calculated using cryoEF (Naydenova and Russo 2017). To identify distinct subsets of particle, 3-D classification without alignment was performed based on the consensus reconstruction. The particles in the consensus reconstruction were classified into six classes (chosen based on initial testing with more classes) and each class was refined separately according to the above procedure. Subsequently, classes were further 3-D classified in the same manner in tiers, each time requesting four classes. Data processing parameters are summarized in Supplementary Tables S1 and S2.

Model refinement in cryo-EM maps

A high-resolution crystal structure of the *E. coli* ribosome (PDB-ID: 4YBB) was rigid-body fitted into the consensus map using UCSF Chimera version 1.14 (Pettersen et al. 2004). The pixel size was refined to 1.07 Å by monitoring the map-model fit while varying the pixel size. Large disordered regions were pruned from the model and refinement was performed using real_space_refine in Phenix version 1.19 (Liebschner et al. 2019) with reference model restraints to PDB-ID: 7K00 (Watson et al. 2020) and interactive re-building using Coot version 0.9 (Emsley et al. 2010). In the final deposited model, all nucleotides not supported by the density were removed. No water oxygens or ions were modeled. Figures were prepared using PyMOL version 2.2.2 (Schrödinger Inc, USA), UCSF Chimera 1.14

(Pettersen et al. 2004) and ChimeraX version 1.1.1 (Goddard et al. 2018). Model refinement parameters are summarized in Supplementary Table S1.

Occupancy of structural elements and hierarchical clustering

The atomic model of a mature ribosome (PDB-ID: 4YBB), as well as a model refined against the consensus reconstruction were rigid body fitted to each class reconstruction using UCSF Chimera and the map values were calculated at the position of each heavy atom (*i.e.* non-hydrogen) in the LSU using the phenix.map_value_at_point program. For these calculations, unsharpened maps produced by Relion 3D refinement were used, which are on an absolute gray scale. Maps were low-pass filtered to 5 Å to avoid spurious results due to local misalignments between the maps and the models and the pixel size was adjusted to the calibrated pixel size.

The occupancy was calculated for each respective structural element (chain or SSE in the case of the 23S rRNA) as the fraction of atoms with a map value above a threshold of 0.05. The threshold was confirmed by manual inspection of the maps. The values were normalized according to (Davis et al. 2016), *i.e.* by dividing by the value by the maximum observed value for a particular feature across all classes or to the median values across all features and maps, whatever was larger.

Manual inspection to gauge occupancies of r-protein was performed by identifying the map threshold where the strongest recognizable feature could be seen. Density for 23S rRNA SSE was classified as being near-native, slightly distorted, highly distorted and disordered, which was translated into numerical values 0.1, 0.085, 0.045 and 0, respectively, for clustering and plotting purposes.

The SSE boundaries for the 23S rRNA were taken from (Petrov et al. 2013), specifically the segmentation in the *E. coli* secondary structure map downloaded from the RibosomeGallery on September 22, 2020 (http://apollo.chemistry.gatech.edu/RibosomeGallery/bacteria/E%20coli/LSU/E_coli_LSU_Helices_2.png), which differs slightly from the original publication. SSEs not named in the map

were given the number of the SSE immediately preceding with an additional letter or the next number if available, *cf.* Supplementary Table S3.

Hierarchical clustering using the complete linkage method and the Euclidean metric was calculated using the function scipy.cluster.hierarchy.linkage in the SciPy library (https://www.scipy.org/), both for classes (columns) as well as features (rows). R-proteins which could not be identified in any of the classes were excluded from the analysis (*i.e.* uL1, uL6, bL9, uL10, uL11, bL12, uL15, uL16, bL25, bL27, bL28, bL33, bL35 and bL36).

RESULTS

Sample preparation

LiCl core particles were prepared from active 50S subunits by incubation in a 3.5 M LiCl buffer and purified on a sucrose gradient to limit heterogeneity. Core particles were enriched for RlmF substrate by pull-down with His-tagged A1618 rRNA methyltransferase RlmF and sedimented through a sucrose cushion (Supplementary Figure S1). As previously shown (Sergiev et al. 2008) the LiCl core particles were verified to be methylated *in vitro* by RlmF using tritium labelled S-adenosyl methionine, showing saturation after 10 minutes at ~20% modification (Supplementary Figure S2). RlmF was by SDS-PAGE judged to be present at approximately the same stoichiometry as uL2 (Supplementary Figure S1) and LC-MS/MS analysis showed similar or higher signal as for the stably bound r-proteins. However, no distinct density for the protein could be found in the consensus reconstruction nor in any of the classes. A diffuse density was observed in the region close to the modification site (H49a), but partial signal subtraction and focused classification did not yield any conclusive results (data not shown).

High-resolution cryo-EM reconstruction of the 50S LiCl core particle

Cryo-EM imaging, single-particle reconstruction and a multitiered classification scheme was used to determine the structure of the LiCl core particle (Figure 1). The micrographs showed signs of heterogeneity (Supplementary Figure S3A,D), as reported previously (Boublik et al. 1978), this could also be confirmed by negative-stain TEM (data not shown). Several of the 2-D class averages have one disordered side (Supplementary Figure S3G). Therefore, the data processing followed an agnostic approach where, to avoid selection bias, no external references were used (Henderson 2013). Hence, particle-picking was done without templates and an *ab initio* model was generated from primary data for initial 3-D reconstructions.

LiCl core particles show strong preferred orientation, predominately binding to the carbon support film in one major direction (Boublik et al. 1978). Consequently, conventional data-collection yielded a reconstruction limited to ~14 Å resolution. To improve the angular sampling, two additional data sets were collected with 15° and 30° tilt of the sample stage.

LiCl washing of the LSU produces a range of sub-particles

A consensus reconstruction (Figure 2) from approximately 384,000 particle images produced a map at 2.84-Å resolution, as estimated by the gold-standard $FSC_{0.14}$ criterion (Supplementary Figures S4P). The reconstructed density is dome-like with the ordered part at the 50S solvent side, centered on the expanded nascent-chain exit tunnel (Figure 3A-C), similar to structures of early *in vivo* and *in vitro* 50S assembly intermediates (Davis et al. 2016; Nikolay et al. 2018). None of the three characteristic protuberances of the LSU are ordered and the 5S rRNA molecule is absent.

Classification could identify a wide range of particle classes with varying degree of ordered density (Figure 1, Supplementary Figure S5). The first round resulted in six major classes, numbered 1–6 according to size. Upon further classification, additional distinct complexes could be identified (Figure 1, Supplementary Figure S5). Major classes 1–3 split in two, mainly distinguished by a distinct, but low-resolution, loop-shaped protuberance at a position close to the location of the L7/L12 stalk in the mature 50S particle (indicated with a

star in Supplementary Figure S5). Two major classes, 4 and 6, did not yield any sub-classes, but class 5, the most populated major class, spawned an extensive tree of sub-classes. After four levels of classification, the largest sub-class collected approximately 1/3 of the initial particles and closely resembles the consensus reconstruction.

Most classes could be resolved to better than 4.5 Å $FSC_{0.143}$ resolution, but the classes with the least and most ordered density could only be resolved to approximately 9 Å resolution (Supplementary Figure S6 & S7). Model refinement (Figure 3A–C, Supplementary Figure S4N–O) and some of the analysis and figures are based on the highest resolution map, the consensus reconstruction. The consensus map is very similar to class 5-5 and all conclusions are also valid for this particle. To estimate the effect of the preferred orientation and the tilted data collection, the distribution of Euler angles was plotted (Supplementary Figure S8). All but one class had E_{od} values in the range 0.6–0.8 (Supplementary Table S2), which can be expected for ribosome data sets (Naydenova and Russo 2017).

The LSU is the most stable at the solvent-side

The r-protein content and degree of folded rRNA of all non-redundant classes (*i.e.* with unique sets of particle images) were analyzed using automated occupancy estimations (Supplementary Figure S9, S10) and plotted as heatmaps (Supplementary Figure S11A, B). The corresponding analysis was also done by manual inspection, looking for presence of characteristic secondary and tertiary structure features (Supplementary Table S4, Supplementary Figure S11C, D). The automatic and the manual annotation of r-proteins mostly agreed, but automatic annotation sometimes failed due to protein flexibility, structural differences relative to the mature particle and limited resolution (*cf.* Supplementary Figures S11E vs. S11F; and S11G vs. S11H).

In the majority of the particles, more than half of the 23S rRNA is folded, including domains 0, I, III, VI, most of domain II and the 5’ half of domain IV (*i.e.* H62–H67, from now on referred to as sub-domain $IV_{5'}$) (Figure 2G). The folded regions are found at the solvent-side and particle base of the LSU and to a high degree correlate with the presence of r-

proteins. Several rRNA helices appear as folded but flexible low-resolution features (Figure 3). The three major protuberances and the subunit interface are unstructured in the majority of the particles, meaning that functionally important regions such as the PTC and binding sites for tRNAs and GTPases are disordered (Figure 2G). In addition, many single nucleotides that in the native structure form tertiary contacts with these regions are also not resolved (Figure 2G).

The LiCl core particle classes are analogous to *in vivo* and *in vitro* assembly intermediates

Major class 1 is the so far smallest identified stable core of the LSU, smaller than the smallest particles observed under bL17-depletion (Davis et al. 2016) and *in vitro* reconstitution (Nikolay et al. 2018). The low resolution for this class is most likely caused by structural flexibility and possibly also by heterogeneous presence of r-proteins. Classes 2 through 6 show strong similarities to *in vivo* assembly intermediates accumulated upon bL17 depletion (Supplementary Figure S12), which is quite remarkable since all particles include the early binder bL17 (Davis et al. 2016). They are also highly similar to intermediates from *in vitro* reconstitution (Nikolay et al. 2018). In line with structures from *in vitro* reconstitution, we observe that the particle base is stable in absence of folded protuberances–similar to the C/E pathway in the bL17 depletion model.

The co-variance of occupancy was analyzed by hierarchical clustering to group the particles classes and to identify cooperativity between binding of r-proteins and folding of rRNA (Supplementary Figure S11A–D). The classes grouped into smaller particles (classes 1, 2, 3 and 5-1), medium-sized particles (classes 4 and 5, except 5-1 and 5-8) and larger particles (classes 5-8 and 6) (Supplementary Figure S11I). These are similar to early, intermediate and late 50S assembly intermediates, respectively. In the small particles, most of domains 0, I, III, VI and a core of domain II is folded (H27–64, but excluding H43–44, the distal part of H38 and H42 and in some cases H33–35a). In the medium-sized particles,

more of domain II, as well as sub-domain $IV_{5'}$ and parts of domain V are folded. The large particles are mature-like, but parts of sub-domain $IV_{3'}$ and parts of domain V are unfolded.

Clustering of structural features (rows in Supplementary Figure S11A–D) was rather sensitive to the chosen parameters (*e.g.* reference model, threshold, low-pass filtering, occupancy normalization, exclusion/inclusion of r-proteins, etc.), but the general trends are reproducible and approximately agree with previously identified folding blocks (Davis et al. 2016). An example of a distinction between disassembly and assembly is that the distal part of domain III (*e.g.* H57) in the LiCl washed ribosomes seems more stable and predominantly groups deep in a cluster with the block-1 elements, while during folding (Davis et al. 2016) they are classified in folding block 3 (Supplementary Figure S11A–C).

The misfolded 3' strand of helix H73

Non-native density is present at the base of the L7/L12 stalk in several classes (1-1, 2-1, 3-1, 4 and 5-2), visible as a low-resolution protruding loop at low map thresholds (Figure 4 and Supplementary Figure S5, star). One arm of the loop consists of H97 (Figure 4B). The other arm (which is apparent as a smaller protrusion in 5-2, Figure 4A, white arrow) emerges outside the particle between H1, H94 and H97. The shape suggests it is an rRNA stem-loop, but all nearby helices can be accounted for in the density or rejected based on the geometry. Just inside, H73 in the similar-sized "non-star" particles, is positioned as in the mature ribosome structure, but is absent in all of the "star" classes (Supplementary Figures S13, compare 2-1 vs. 2-2, 4 vs. 5-1, and 5-2 vs. 5-3). Helix H73 is at a 4-way junction, but its 3' strand (nucleotides 2611–2625) can according to prediction form either a smaller or a larger stem loop, compatible in length with the observed non-native densities (Supplementary Figure S14). Therefore, the non-native density in the "star" classes most likely consists of mis-folded H73.

R-proteins leave the 50S according to *in vivo* assembly groups

Across all classes, a number of proteins are always present at high occupancy (uL3, uL4, bL17, bL20, bL21, uL22, uL23, uL29 and bL34; Supplementary Table S4). In addition, uL13 and uL24 are present in most classes. All of these except bL32 have firm support in LC-MS/MS data (Supplementary Table S4). These proteins are all early binders *in vivo*, as identified by pulse-labeling coupled with MS analysis (Chen and Williamson 2013) (Supplementary Figure S15). Furthermore, four proteins that are only found in larger particles (uL2, uL14, bL19 and bL32) bind in the subsequent step *in vivo*.

None of the classes showed clear density for uL1, uL6, bL9, uL10, uL11, bL12, uL15, uL16, bL25, bL27, bL28, bL31, bL33, bL35 or bL36 – although diffuse density due to flexible rRNA prevented the unambiguous assessment of some proteins. Trace levels of uL1, uL5, bL9, uL11, uL15, uL30 and bL31 were detected by LC-MS/MS (Supplementary Table S4), suggesting that some of these r-proteins remain bound to disordered rRNA.

The higher solubility of the CP-associated r-proteins (such as uL5, uL18 and bL25) destabilizes the CP in the LiCl washed ribosomes. The CP and the 5S rRNA are only discernable in the two largest particle classes, 5-8 and 6 (Supplementary Figure S5). Out of the proteins that *in vitro* were linked to 5S rRNA integration (Supplementary Figure S15) (Herold and Nierhaus 1987), uL5 and uL18 can be confirmed at low occupancy in class 6, and uL18 also in class 5-8.

For the three largest r-proteins in the consensus reconstruction (uL2, uL3 and uL4), the long loops and tails that in the native particle interact with the 23S rRNA are invisible (Figure 2I, Supplementary Figure S4L). In some cases, the interacting rRNA helices are unstructured and in other cases they are folded but not in their native position.

Comparison of particle classes show dependencies between different structural elements

The wide range of small 50S sub-particles after LiCl wash allows analysis of the structural consequences of removal of certain r-proteins as well as unfolding of helices and domains of rRNA. Pair-wise comparison between the particle classes thus reveals folding and stability

dependencies that are mostly additive, as depicted schematically in Figure 5. The left branch in the diagram represents particles where the mis-folded H73 appears (classes 5-2, 4, 3-1, 2-1, 1-1). The right branch consists of one path with stronger density for uL2, H34–35 and sub-domain $IV_{5'}$ (5-5, 5-3, 3-2) and one path with stronger density for domain V (5-4, 5-1, 2-2).

Sub-domain III_{tail} can stabilize sub-domain $IV_{5'}$ in absence of uL2

Class 1-1, the smallest of all the reconstructed particles, only consists of domain I (except H25), domain III, domain VI, H26a and H61 of domain 0 and the 5' half of domain IV with particularly strong density for H63 (Supplementary Figure S16A, S17A). This is remarkable, since folding of sub-domain $IV_{5'}$ in other classes coincides with presence of uL2 (see below). Class 1-2 has additional density for the core of domain II, a more mature-like domain 0 and helix H25 of domain I, but less density for domain IV and the distal half of domain III (helices H54–59, from here on referred to as sub-domain III_{tail} (Lanier et al. 2016)).

Based on these structures, we hypothesize that in lieu of uL2, sub-domain III_{tail} can stabilize sub-domain $IV_{5'}$ through contacts between H57 and H58 in domain III and H63 (Supplementary Figure S16B). The correlation between the folding of sub-domain $IV_{5'}$ and sub-domain III_{tail} was also observed in the bL17-depletion model (*i.e.* transitions B to C2 and D1 to E1) (Davis et al. 2016). The observation of a folded or mostly folded sub-domain III_{tail} in all of the particles, independently of both domain IV and uL2, suggests that the former stabilizes the latter rather than vice versa. It has previously been shown that domain III, presumably a late evolutionary addition (Hury et al. 2006), can fold independently of r-proteins or the rest of the 23S rRNA (Athavale et al. 2012). These results suggest that sub-domain III_{tail} promotes folding of domain IV rRNA and thus facilitates the early stages of ribosome biogenesis.

Particles in class 2 show higher stability and homogeneity, and are resolved to significantly higher resolution than class 1, probably because additional r-proteins stabilize the RNA structure. Unlike the smaller particles, they show no density for sub-domain $IV_{5'}$ (Figure 6B and Supplementary Figures S17A). In class 2-1, compared to the native particle,

the whole flank opposite to the CP is shifted 20 Å away from the subunit interface (Supplementary Figures S16C). In class 2-2, this lobe, consisting of sub-domain III_{tail} and helices in domains 0 and VI, is instead in a near-native position correlated with stronger density for the neighboring regions in domain 0 (H25a, H72 and H73) and domain VI (H95 and H97).

The exit tunnel is expanded in the smaller particles, in particular class 2-2 (Figure 3D-F), similar to particle B in Davis *et al.* that also has a partly disordered domain III (Supplementary Figure S12). Classes 1, 2 and 3-1 are possible substrates for RlmF where the A1618 modification site is accessible. These constitute 17% of the total number of particles, in reasonable agreement with 20% of the particles being labeled after *in vitro* methylation.

Class 3-1 is rather similar to 2-1 but H63 is more pronounced and there is more diffuse density in the H33–H35a region (Supplementary Figure S17A). Correlated to this, domain III_{tail} is also closer to its native position, stabilized by H63 (Supplementary Figure S16A) and possibly by the higher occupancy of uL3 and bL17.

In class 3-2, domain II is in a more native position compared to 3-1. It is the smallest particle with unambiguous density for H34–H35 (Figure 6B), correlated with improved positioning of nearby domain III (H52, H56 and H60) and H96 and H101 in domain VI (Supplementary Figure S16).

R-protein uL3 is important for the stability of domains 0 and VI

Class 5-1 shares attributes with both of the smaller particles 2-2 and 3-2, but helix H63 is close-to-natively folded (Supplementary Figure 18A). Similarity to the larger particles in class 5, it has diffuse density extending from H73 spanning the middle of the particle and connecting to H22 (see below).

Higher occupancy of uL3 in 5-1 than in 3-2 seems to cause stronger density for domains 0 and VI (except H96). The same trend can be observed in 2-2 vs 2-1. In the mature particle, the internalized loop 128-153 of r-protein uL3 directly interacts with several of these helices

(H61 and H73 of domain 0 and H94, H96 and H100 of domain VI), suggesting an important role of uL3 in preventing misfolding of H73 and in maintaining its native structure (see above).

R-protein uL2 stabilizes sub-domain $IV_{5'}$

Class 4 has many similarities to class 5-1, but shows the misfolded loop density (see above), which seems to dislocate a lobe at the solvent-side of the particle close to the base of the stalk, including helices H1 and H41 and r-proteins uL13, bL20 and bL21. It also has significantly stronger density for sub-domain $IV_{5'}$ and more mature-like H33–H35a (Supplementary Figure S17A). These differences in domains II and IV seem correlated with the increased occupancy of uL2 (Figure 5 & 6C). The same trend is seen in classes 5-2, 5-3, 5-5 and 5-6 (Supplementary Figure S17B). R-protein uL2 is the largest r-protein in the LSU and makes contact to helices in domains II, III, IV and V. It groups together with H62 and H64–H67 in the clustering analysis (Supplementary Figure S11A,B), as also observed in the bL17 depletion model (Davis et al. 2016).

Class 5-2 is highly similar to class 4, but the extra density at the stalk-base appears shorter (Figure 4; see above). Class 5-2 also has slightly stronger density for uL24 on the exposed solvent side of domain I, but its presence does not seem to affect the rRNA structure.

Helices H79 and H88 anchor the core of domain V

Class 5-3 shares similarities with both 5-1 and 5-2. As in class 5-1, helices H72 and H73 are folded (Supplementary Figure S13) and diffuse density is connecting H73 and H22 across the particle, likely representing a highly flexible state of H74. The kissing-loop contact between H88 and H22 is present, but similarly to the nearby H29 and H31, H88 is shifted relative to its native position in absence of the stabilizing proteins uL15, bL33 and/or bL35. Also during *in vitro* reconstitution, association of the late-binding r-proteins bL33 and bL35 was shown to coincide with large structural rearrangements of the CP (Nikolay et al. 2018),

thus supporting the correlation between absence of these proteins and rRNA flexibility. Class 5-3 has some density for r-protein uL19 and low levels for neighboring uL14 (Supplementary Figure S16), similar to the first particle along the C-folding pathway in (Davis et al. 2016) (Supplementary Figure S12).

Class 5-4 is quite similar to class 5-3, but with slightly weaker sub-domain $IV_{5'}$ and uL2 (Supplementary Figure S17B) and stronger density for domain V. Helix H88 is more well-defined and there is distinct density for a folded H79 in the native position next to H10. Plenty of diffuse density is connecting these elements and extending in the direction towards H73, and the proximal end of H89 is visible.

Class 5-5, compared to 5-4, shows less distinct density for domain V, although H79 seems ordered. Instead, the density for uL2 and sub-domain $IV_{5'}$ is stronger (Supplementary Figure S17B). The distal part of helix H67 is well-defined, but the 3'-half of domain IV is disordered. This class resembles the second particle in the C-path of bL17-depleted LSU sub-particles (Supplementary Figure S12).

Stability of sub-domain $IV_{3'}$ correlates with stability of domain V in the largest particles

In class 5-6, all the elements that are visible in either of the other smaller particles in class 5 are ordered (Figure 5, Supplementary Figure S17).

Class 5-7 in addition shows strong density for helices H74 and H75 in domain V connecting all the way from H73 to H88 and H79 (Supplementary Figure S17B). Although helix H79 is packed against H10, they are shifted from their native conformation in absence of bL28. Helix H76 also has strong density, which extends into the L1 protuberance formed by the folded but flexible helices H77–78 (Supplementary Figure S5). This is a clear parallel to observations in the bL17-depletion model, where folding of domain V along the C/E paths also starts with H79 to expand towards the two protuberances (Davis et al. 2016). The higher occupancy of uL13 in class 5-7 seems to stabilize helices H1 and H41 while, as in other classes, the relatively strong density for uL14 and uL19 does not appear to influence the rRNA.

In the map of class 5-8 it is, despite limited resolution and anisotropy artifacts, clear that the 3'-half of domain IV is folded, but the distal end of H69 is flexible (Figure 6), agreeing with H69 and H71 remaining unfolded in the largest assembly intermediate in the bL17 depletion model (Davis et al. 2016). Compared to class 5-7, domain V shows more native structure, probably related to the presence of sub-domain $IV_{3'}$ (Figure 6). Helix H68 packs outside H75 and H88, keeping them in place and H69 and H71 pack outside of the folded H93. There is also weak density for the rest of the 3' end of domain V, including H90–92a and the proximal end of H89. Native positioning of H89, however, requires uL16 (Nikolay et al. 2018). Finally, the H82–H87 branch of domain V, that forms part of the CP, shows partial order, and very weak density can be discerned for additional large CP elements (distal end of H38 and 5S rRNA).

The CP is severely destabilized in the LiCl core particles but folds early in ribosome assembly

Class 6, is the only particle where most of the CP rRNA, including the 5S rRNA, is folded. In other aspects, it is similar to class 5-8. The CP is highly depleted of r-proteins, but contains low levels of proteins uL5, uL18 and uL30. The entire CP is shifted away from the subunit-interface side relative to in the mature particle, likely explained by the absence of r-proteins bL33 and bL35, which during *in vitro* reconstitution were shown to assist in adaptation of the native CP conformation (Nikolay et al. 2018).

Folding of helices H49a and H35 demonstrates analogies between ribosome disassembly and assembly

As demonstrated above, the observed classes of disassembled LiCl core particles display high similarity to assembly intermediates. Thus, step-by-step comparison of the smaller to larger particles allows the folding of parts of the LSU to be modelled as an intricate sequential addition of structural layers onto a pre-existing core. As an example, we analyzed the folding of helices around H49a (modification site of RlmF in domain III) and H35 (domain

II) (Figure 7, Supplementary Figure S19). These helices together form tertiary interactions between domains 0, II, III, IV and V in a region stabilized by the positively charged extensions of r-proteins uL2, uL22 and bL34.

Folding and tertiary interactions of H49a

Helix H49a is a stem-loop in domain III that bridges to domain II, close to the nascent-chain exit tunnel (Figure 3D,F). In the mature particle, the stem of H49a packs against helices H49b and H50 and the loop forms interactions with H50, H47 and H35 (Figure 7G). Although most of domain III is folded across all classes, a range of different conformations and tertiary contacts of H49a are observed (Supplementary Figure S19), allowing reconstruction of the sequential assembly of interactions in this region (Figure 7B-F), corresponding to one path of Figure 5. In class 2-2 (Figure 7B), H48 and H49b are folded but H49a is completely disordered. In other early classes (2-1 and 3-1), H49a is folded but points away from H50 at different angles. In class 5-1 (Figure 7C) and 3-2, H49a packs against H49b (in Figure 7G: contacts 1, A-minor contact of A1610/H49a, and 2, A1609/H49a stack with G1332/H50). Although the loop of H49a lacks most of the native tertiary contacts, there is weak density for A1616 connecting to H50 (contact 3). Class 4 is similar but also shows density for the 5'-end of H47 (contact 4, C1615/H49a base-pair with G1271/H47) and for H35. In class 5-3 (Figure 7D) and 5-2, H49a is in a close-to-native conformation. Contact 5 is formed through the stacking between A1618 of H49a and A749 of H35 observed in class 5-4, and the triple-A stack is completed by A1272 of H47 as observed in class 5-5 (Figure 7E). The stack is expected to be further stabilized by the N6 methylation of central A1618 (Sergiev et al. 2016). In classes 5-6 to 5-8 (Figure 7F), there is only a slight distortion of the H49a loop at the flipped-out base of A1614 compared to the native structure where it forms an interaction with the extended loop of uL22, the tip of which (residues 85-95) shows rather weak density in all classes.

Folding and tertiary interactions of H35

Helix H35 is a stem-loop in domain II, which packs outside H49a, where it forms contacts to domains 0 (H73), IV (H64 and H65) and V (H93) (Figure 7G & 3F). H35 is disordered in most small classes such as 2-2 (Figure 7B) and only vaguely discernable in the other small particles. The stability of H35 shows correlation to presence of r-protein uL2 (Figure 7C-E). uL2 interacts with H35a, which packs outside of H35 (Figure 7C) that in turn contacts both H35 (contact 6, A751/H35 stack with A789/H35a) and H65 (Figure 7D). An A-minor contact interlocks H64 and H35 (contact 7, A1681/H64), observed in the more mature-like classes 5-3 and 5-5 (Figure 7D, E), but absent in *e.g.* class 5-1. The further interactions of H65 with the loop of H35 (contact 8, A752/H35 stack with U1781/H65) and of its loop to H73 and H93 (contacts 9, m^5U747/H35 stack with U1613/H73, and 10, A752/H35 base-pair with U2609/H93) and to uL22 are only present in the largest particles such as class 5-8 (Figure 7F), completing the native structure.

DISCUSSION

Core particles of the LSU, where the most loosely attached ribosomal proteins and the 5S RNA have been removed using high-salt wash protocols were first used in studies of ribosome assembly more than half a century ago (Meselson et al. 1964; Homann and Nierhaus 1971). We here provide structural understanding of what these particles are and investigate their similarities and differences to structures occurring during *in vivo* and *in vitro* ribosome assembly.

The r-protein occupancy of our LiCl core particles largely agrees with previous characterizations (Homann and Nierhaus 1971; Spillmann et al. 1977; Nierhaus 1990; Boublik et al. 1978). The spontaneously exchangeable r-proteins *in vitro* and *in vivo* (Pulk et al. 2010) are the most loosely associated ones, a small subset of the ones that can be dissociated with salt. There is no indication that their dissociation and re-association would cause major structural changes, which seems evolutionary important to maintain the active pool of ribosomes.

The wide range of LSU sub-particles that we observe allows analysis of the consequences of removal of specific r-proteins. For some r-proteins, *e.g.* uL3, a primary binder, and uL2, a later binder (Chen and Williamson 2013) there is a strong link between their presence and observation of the native rRNA structure. The central protuberance shows high sensitivity to deproteination and unfolds as proteins uL5, uL16, uL18, bL25, uL30 and uL31 leave. In contrast, it can fold early during ribosome biogenesis as demonstrated in the bL17-depletion model (start of the D path, *cf.* Figure 3C) (Davis et al. 2016) and during LSU *in vitro* reconstitution (Nikolay et al. 2018). The structural stability of domain III seems less dependent on r-proteins. In general, the r-proteins that bind first during assembly are also the ones that stay bound during high-salt wash. The exceptions are uL5, uL15 and uL18 in the CP region, which are early binders that easily dissociate during salt wash (Supplementary Figure S15).

With the exception of the smallest particles, the reconstructed cores show strong similarity to intermediates isolated from bL17-deficient cells or from *in vitro* reconstitution (Supplementary Figure S12) and their protein content agrees with *in vivo* assembly (Chen and Williamson 2013). These similarities between states obtained during disassembly and different variants of assembly strongly support the innate propensity of the rRNA and the r-proteins to adopt distinct, meta-stable complexes. Evolutionary, this inherent property is likely important for the robustness of ribosome assembly upon changing or challenging conditions, such as disabling mutations of an r-protein (Davis et al. 2016). Still, local minima in the energy landscape can lead to trapping of misfolded structures. One such example is the strong non-native secondary structures observed to form by the 3’ strand of helix H73 (Figure 4). This manifests the need for ribosomal proteins not only to facilitate formation of the active RNA structure during co-transcriptional *in vivo* assembly, but also to maintain the native structure in a partly folded structural context. During *in vivo* assembly, RNA helicases allow remodelling of such misfolded structures (Rodgers and Woodson 2021) and during *in vitro* assembly the optimized conditions of temperature and magnesium (Nierhaus and Dohme 1974; Nierhaus 1991) may prevent their formation.

To what extent does ribosome disassembly *in vitro* reverse the process of ribosome assembly *in vivo* or *in vitro*? The disassembly process is sterically constrained; a certain RNA helix or r-protein tail is held in place by outer assembly layers of the ribosome. This may fundamentally be more similar to *in vitro* assembly where the structure has to form starting from the core and continuing outwards. During the co-transcriptional *in vivo* process there is an additional directionality, where at a given time only part of the rRNA can be available for folding and tertiary interactions (reviewed in (Rodgers and Woodson 2021)). Still, the r-proteins that are present in at high occupancy in many classes of core particles are all binding early during *in vivo* assembly (Chen and Williamson 2013), suggesting that directionality is not a main factor in r-protein affinity. We observe that globular r-protein domains are ordered while the positively charged loops and tails are disordered, suggesting that these dissociate from the rRNA before their connected globular domains, in reverse of the assembly procedure. The native packing of the RNA structure sometimes requires such r-protein tails or loops, as an example H49a does not adopt its fully native conformation in absence of the uL22 loop (Figure 7F), the deletion of which has been shown to lead to accumulation of immature LSU particles (Lawrence et al. 2016).

The reconstructed series of particles subjected to r-protein dissociation corroborates the multi-pathway of assembly (Siibak and Remme 2010; Davis et al. 2016; Davis and Williamson 2017) by demonstrating the semi-independent stability of folding blocks of rRNA and r-proteins. In particular, since parts of domains IV and V can remain structured while the other unfolds, we can from the LiCl core particles derive two distinct pathways where either of these blocks independently remain folded (Figure 5).

We have here shown that removal of loosely associated r-proteins with a high-salt wash allows access to the smallest stable cores of the large subunit. These particles show density for only 4–8 r-proteins, but additional ones may be bound to disordered rRNA regions. Compared to the well-studied early, co-transcriptional 30S assembly process (Rodgers and Woodson 2021), little is known about the corresponding steps of 50S assembly. However, this study provides examples of concepts that have been described for the 30S such as the

sequential reduction of structural variability of RNA by binding of r-proteins, specifically in regions that require long-range interactions to reach their native structure. There is a clear similarly between how r-proteins during assembly recognize and bind to structured regions and further contribute to the maturation of the nearby structure, and how dissociation of r-proteins seems to start with the disordering of their extensions and loops coupled to loosening of rRNA structure to be followed by dissociation of the folded r-protein domains and unfolding of the respective rRNA binding sites. In addition, we observe a multi-pathway disassembly process, in agreement with the complex “assembly landscape” of the 30S subunit that was proposed as analogous to the energy landscape in protein folding (Talkington et al. 2005). These observations support that the core particles in some respects can be used as mimics of transient early assembly intermediates, for example in studies of maturation factors.

In conclusion, the *in vitro*-generated 50S LiCl core particles represent a collection of intrinsically stable or meta-stable complexes, from which we can learn about inherent structural properties of ribonucleoprotein particles, and specifically about states that are likely to occur during 50S ribosome assembly.

DATA AVAILABILITY

Atomic coordinates for the consensus structure have been deposited in the Protein Data Bank under accession code 7ODE. All maps have been deposited in the Electron Microscopy Data Bank with accession codes EMD-12826 for the consensus reconstruction and EMD-12828 through EMD-12841, EMD-12843 and EMD-12844 for the different classes.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

FUNDING

This work was supported by [grants 2016-06264, 2017-03827] from the Swedish Research Council to MS.

ACKNOWLEDGEMENTS

The cryo-EM data was collected with the aid of Julian Conrad and Marta Carroni at the Solna node of the Cryo-EM Swedish National Facility funded by the Knut and Alice Wallenberg, Family Erling Persson and Kempe Foundations, SciLifeLab, Stockholm University and Umeå University.

AUTHOR CONTRIBUTIONS

SK performed protein purification and prepared the sample complex with DL. DL prepared ribosomes and did all cryo-EM work. MS conceived the project and obtained funding. DL and MS wrote the manuscript. All authors approved the manuscript.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest with the contents of this article.

REFERENCES

Athavale SS, Gossett JJ, Hsiao C, Bowman JC, O'Neill E, Hershkovitz E, Preeprem T, Hud N V., Wartell RM, Harvey SC, et al. 2012. Domain III of the T. thermophilus 23S rRNA folds independently to a near-native state. *Rna* **18**: 752–758.

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H. 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: The Keio collection. *Mol Syst Biol* **2**.

Boublik M, Spiess E, Roth HE, Hellmann W, Jenkins F. 1978. Structure of LiCl core particles of 50 S ribosomal subunits from Escherichia coli by electron microscopy. *Cytobiologie* **18**: 309–319.

Chen SS, Sperling E, Silverman JM, Davis JH, Williamson JR. 2012. Measuring the dynamics of E. coli ribosome biogenesis using pulse-labeling and quantitative mass

spectrometry. *Mol Biosyst* **8**: 3325–3334.

Chen SS, Williamson JR. 2013. Characterization of the ribosome biogenesis landscape in E. Coli using quantitative mass spectrometry. *J Mol Biol* **425**: 767–779.

Dagley S, Sykes J. 1959. Effect of drugs upon components of bacterial cytoplasm. *Nature* **183**: 1608–1609.

Davis JH, Tan YZ, Carragher B, Potter CS, Lyumkis D, Williamson JR. 2016. Modular Assembly of the Bacterial Large Ribosomal Subunit. *Cell* **167**: 1610-1622.e15.

Davis JH, Williamson JR. 2017. Structure and dynamics of bacterial ribosome biogenesis. *Philos Trans R Soc B Biol Sci* **372**: 20160181.

Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. *Acta Crystallogr Sect D Biol Crystallogr* **66**: 486–501.

Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, Ferrin TE. 2018. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **27**: 14–25.

Henderson R. 2013. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proc Natl Acad Sci U S A* **110**: 18037–18041.

Herold M, Nierhaus KH. 1987. Incorporation of six additional proteins to complete the assembly map of the 50 S subunit from Escherichia coli ribosomes. *J Biol Chem* **262**: 8826–8833.

Homann HE, Nierhaus KH. 1971. Ribosomal Proteins: Protein Compositions of Biosynthetic Precursors and Artifical Subparticles from Ribosomal Subunits in Escherichia coli K 12. *Eur J Biochem* **20**: 249–257.

Hury J, Nagaswamy U, Larios-Sanz M, Fox GE. 2006. Ribosome origins: The relative age of 23S rRNA Domains. *Orig Life Evol Biosph* **36**: 421–429.

Lanier KA, Athavale SS, Petrov AS, Wartell R, Williams LD. 2016. Imprint of Ancient Evolution on rRNA Folding. *Biochemistry* **55**: 4603–4613.

Lawrence MG, Shamsuzzaman M, Kondopaka M, Pascual C, Zengel JM, Lindahl L. 2016. The extended loops of ribosomal proteins uL4 and uL22 of Escherichia coli contribute to

ribosome assembly and protein translation. *Nucleic Acids Res* **44**: 5798–5810.

Liebschner D, Afonine P V., Baker ML, Bunkoczi G, Chen VB, Croll TI, Hintze B, Hung LW, Jain S, McCoy AJ, et al. 2019. Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr Sect D Struct Biol* **75**: 861–877.

Meselson M, Nomura M, Brenner S, Davern C, Schlessinger D. 1964. Conservation of ribosomes during bacterial growth. *J Mol Biol* **9**: 696–711.

Naydenova K, Russo CJ. 2017. Measuring the effects of particle orientation to improve the efficiency of electron cryomicroscopy. *Nat Commun* **8**: 629.

Nierhaus K. 1990. Reconstitution of ribosomes. In *Ribosomes and Protein Synthesis* (ed. G. Spedding), pp. 161–189, Oxford University Press, Oxford.

Nierhaus KH. 1991. The assembly of prokaryotic ribosomes. *Biochimie* **73**: 739–755.

Nierhaus KH, Dohme F. 1974. Total reconstitution of functionally active 50S ribosomal subunits from Escherichia coli. *Proc Natl Acad Sci U S A* **71**: 4713–4717.

Nikolay R, Hilal T, Qin B, Mielke T, Bürger J, Loerke J, Textoris-Taube K, Nierhaus KH, Spahn CMT. 2018. Structural Visualization of the Formation and Activation of the 50S Ribosomal Subunit during In Vitro Reconstitution. *Mol Cell* **70**: 881-893.e3.

Nikolay R, Hilal T, Schmidt S, Qin B, Schwefel D, Vieira-Vieira CH, Mielke T, Bürger J, Loerke J, Amikura K, et al. 2021. Snapshots of native pre-50S ribosomes reveal a biogenesis factor network and evolutionary specialization. *Mol Cell*. **81**: 1200-1215.e9

Petrov AS, Bernier CR, Hershkovits E, Xue Y, Waterbury CC, Hsiao C, Stepanov VG, Gaucher EA, Grover MA, Harvey SC, et al. 2013. Secondary structure and domain architecture of the 23S and 5S rRNAs. *Nucleic Acids Res* **41**: 7522–7535.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera - A visualization system for exploratory research and analysis. *J Comput Chem* **25**: 1605–1612.

Pulk A, Liiv A, Peil L, Maiväli Ü, Nierhaus K, Remme J. 2010. Ribosome reactivation by replacement of damaged proteins. *Mol Microbiol*. **75**: 801-814

Rodgers ML, Woodson SA. 2021. A roadmap for rRNA folding and assembly during transcription. *Trends Biochem Sci*. **46**: 889-901

Rodnina M V., Wintermeyer W. 1995. GTP consumption of elongation factor Tu during translation of heteropolymeric mRNAs. *Proc Natl Acad Sci U S A* **92**: 1945–1949.

Scheres SHW. 2016. Processing of Structurally Heterogeneous Cryo-EM Data in RELION. In *Methods in Enzymology*, Vol. 579 of, pp. 125–157, Academic Press.

Seffouh A, Jain N, Jahagirdar D, Basu K, Razi A, Ni X, Guarné A, Britton RA, Ortega J. 2019. Structural consequences of the interaction of RbgA with a 50S ribosomal subunit assembly intermediate. *Nucleic Acids Res* **47**: 10414–10425.

Sergiev P V., Golovina AY, Osterman IA, Nesterchuk M V., Sergeeva O V., Chugunova AA, Evfratov SA, Andreianova ES, Pletnev PI, Laptev IG, et al. 2016. N6-Methylated Adenosine in RNA: From Bacteria to Humans. *J Mol Biol* **428**: 2134–2145.

Sergiev P V., Serebryakova M V., Bogdanov AA, Dontsova OA. 2008. The ybiN Gene of Escherichia coli Encodes Adenine-N6 Methyltransferase Specific for Modification of A1618 of 23 S Ribosomal RNA, a Methylated Residue Located Close to the Ribosomal Exit Tunnel. *J Mol Biol* **375**: 291–300.

Shajani Z, Sykes MT, Williamson JR. 2011. Assembly of bacterial ribosomes. *Annu Rev Biochem* **80**: 501–526.

Siibak T, Peil L, Dönhöfer A, Tats A, Remm M, Wilson DN, Tenson T, Remme J. 2011. Antibiotic-induced ribosomal assembly defects result from changes in the synthesis of ribosomal proteins. *Mol Microbiol* **80**: 54–67.

Siibak T, Remme J. 2010. Subribosomal particle analysis reveals the stages of bacterial ribosome assembly at which rRNA nucleotides are modified. *Rna* **16**: 2023–2032.

Spillmann S, Dohme F, Nierhaus KH. 1977. Assembly in Vitro of the 50 S subunit from Escherichia coli ribosomes: Proteins essential for the first heat-dependent conformational change. *J Mol Biol* **115**: 513–523.

Talkington MWT, Siuzdak G, Williamson JR. 2005. An assembly landscape for the 3OS ribosomal subunit. *Nature*. **438**: 628-632

Wang W, Li W, Ge X, Yan K, Mandava CS, Sanyal S, Gao N. 2020. 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 U S A* **117**: 15609–15619.

Watson ZL, Ward FR, Méheust R, Ad O, Schepartz A, Banfield JF, Cate JHD. 2020. Structure of the bacterial ribosome at 2 Å resolution. *Elife* **9**: 1–62.

Zhang K. 2016. Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**: 1–12.

Zheng SQ, Palovcak E, Armache JP, Verba KA, Cheng Y, Agard DA. 2017. MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**: 331–332.

Zhong ED, Bepler T, Berger B, Davis JH. 2021. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat Methods*. **18**: 176-185

Zhou D, Zhu X, Zheng S, Tan D, Dong MQ, Ye K. 2019. Cryo-EM structure of an early precursor of large ribosomal subunit reveals a half-assembled intermediate. *Protein Cell* **10**: 120–130.

Zivanov J, Nakane T, Forsberg BO, Kimanius D, Hagen WJH, Lindahl E, Scheres SHW. 2018. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**: 1–22.

Zivanov J, Nakane T, Scheres SHW. 2020. Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1. *IUCrJ* **7**: 253–267.

FIGURE LEGENDS

Figure 1. Single-particle reconstruction and hierarchical classification of LiCl core particles. The width of the branches in the diagram reflects the number of particle images (italic font) in each class (bold font) in the tiered classification scheme. Classes are sorted from left to right approximately according to the amount of ordered density. Images show the reconstructions in “crown view” with the L1 protuberance to the left and the L7/L12 stalk to the right (*cf.* Supplementary Figure S5 for details). Minor classes that did not produce a stable reconstruction by themselves, and hence were discarded, are not shown.

Figure 2. Consensus reconstruction of the 50S LiCl core particle (**A)** The map (gray) in “crown view” is filtered according to the local resolution and shown at map level 0.1, together with a rigid-body fitted model of the mature 50S particle (PDB ID 4YBB). The rRNA is colored according to (G) and shown with the subunit interface at the front. (**B-D**) Same as (A), rotated in 90° increments around the Y-axis. (**E,F**) The model of the mature 50S and the map shown separately in the same view as in (A). (**G**) Secondary-structure map of the LSU rRNA (modified from (Petrov et al. 2013)). Folded parts in the consensus particle are indicated with filled circles and outlined with a dashed line. Disordered helices, loops or single nucleotides are shown as white dots and labeled in italics. The SSE elements are colored in orange, indigo, blue, purple, yellow or red tones for domains 0–VI, respectively. (**H,I**) Present r-proteins from the rigid-body fitted model shown as space-filling spheres together with map. Disordered loops in proteins uL2 (228–272), uL3 (128–153) and uL4 (48–83) are shown as cartoon. View along the exit tunnel, seen from the solvent-side (**H**) and the subunit interface side (**I**). Additional views in Supplementary Figure S4.

Figure 3. Local resolution and the expanded exit tunnel. (**A**) Consensus reconstruction seen from the interface-side, centered on the nascent-chain exit tunnel, colored according to local resolution, with the refined atomic model colored according to Figure 2G. (**B**) As (A), rotated

90°. (**C**) Central slice through the map (rectangle in (A)) in the same orientation as in (B) showing the expanded exit-tunnel and the absent CP. The rigid-body fitted model of the mature 50S particle (PDB ID 4YBB) is colored as in Figure 2G. The solid line indicates the extent of the ordered density and the dashed line indicates the extent of the mature particle. (**D-F**) Class 2-2 in same views as in (A-C). The model is the same as in (A-B). In this particle, sub-domain $IV_{5'}$ and parts of domains II, III are disordered compared to the consensus structure, in particular nucleotide A1618 in H49a. The dashed line in (F) indicates the extent of the consensus reconstruction in (C). All maps are locally lowpass filtered according to the local resolution and are displayed at level 0.13.

Figure 4. Non-native density close to the base of the L7/L12 stalk. (**A**) In classes 1-1, 2-1, 3-1 and 4, the density of H97 extends and loops back close to H1 (asterisk). In class 5-2, density for the upper (black arrow-head) and lower (white arrow-head) arm are not connected. Analogous classes 1-2, 2-2, 3-2, 5-1 and 5-3, respectively, do not show this extra density. Maps are filtered according to the local resolution and are displayed at level 0.05. Parts of the mature particle structure (PDB ID 4YBB) is shown for reference. (**B**) Same parts of model 4YBB as in (A), in same view (colors as in Figure 2G). (**C**) Overview.

Figure 5. Relationship between observed particle classes based on presence of structural features. Arrows indicate decreasing order upon disassembly. Lines (colored as in Figure 2G) enclose particles with significant density for specific rRNA helices or r-proteins, labels indicate rRNA helices that unfold or r-proteins that dissociate. Particles are shown as in Supplementary Figure S5. Asterisks indicate misfolded H73 structures (Figure 4).

Figure 6. Structural dependencies between presence of uL2 and degree of folding of 23S rRNA domain IV and H33–35a in domain II. Maps (gray) are filtered according to estimated local resolution and displayed at a threshold of 0.08. The reference model of the mature particle (PDB ID 4YBB) is colored with RNA as in Figure 2G and proteins in white (details

and additional views in Supplementary Figures S17 & S18). (**A**) Overview showing the relevant region of the consensus reconstruction. (**B**) Helices H33–35a (blue) and H58 (pink) are absent in class 2-2 but present in class 3-2. (**C**) Folding of helices H62–67 (yellow) correlates with presence of r-protein uL2 (white). (**D**) The 3' half of domain IV (H68–71) folds outside of H74, H75 and H88 in domain V, as seen in larger class-5 particles.

Figure 7. Folding of the H49a-H35 region in different particles suggests a local order of assembly. (**A-F**) Maps showing varying degrees of folded structure in this region. Lines indicate density that differ between maps (dashed line indicate weak/diffuse density). The maps are not sharpened or low-pass filtered and shown at level 0.05. Supplementary Figure S19 shows the same view for all particle classes. (**A**) The overview shows the consensus map and a mature model, PDB ID 4YBB, colored as in Figure 2G. (**B**) In class 2-2, H48 is folded, but not H49a. (**C**) Density for H34, H35 and H35a outside of H49a in class 5-1. (**D-E**) Occupancy for H64, H65 and uL2 is partial in class 5-3, but close to full in 5-5. (**F**) Helices H90, H93 and H74 are folded in class 5-8. (**G**) Schematic view of stabilizing interactions in the H49a-H35 region. Tertiary contacts for H49a and H35 are indicated with solid lines and numbered in approximate assembly order, as deduced from the maps (see text for numbers). Dashed lines indicate contacts to r-proteins. (**H**) Structure of helices and r-proteins in (G) in the mature 50S (PDB ID 4YBB).

FIGURES for Larsson *et al.* 2021

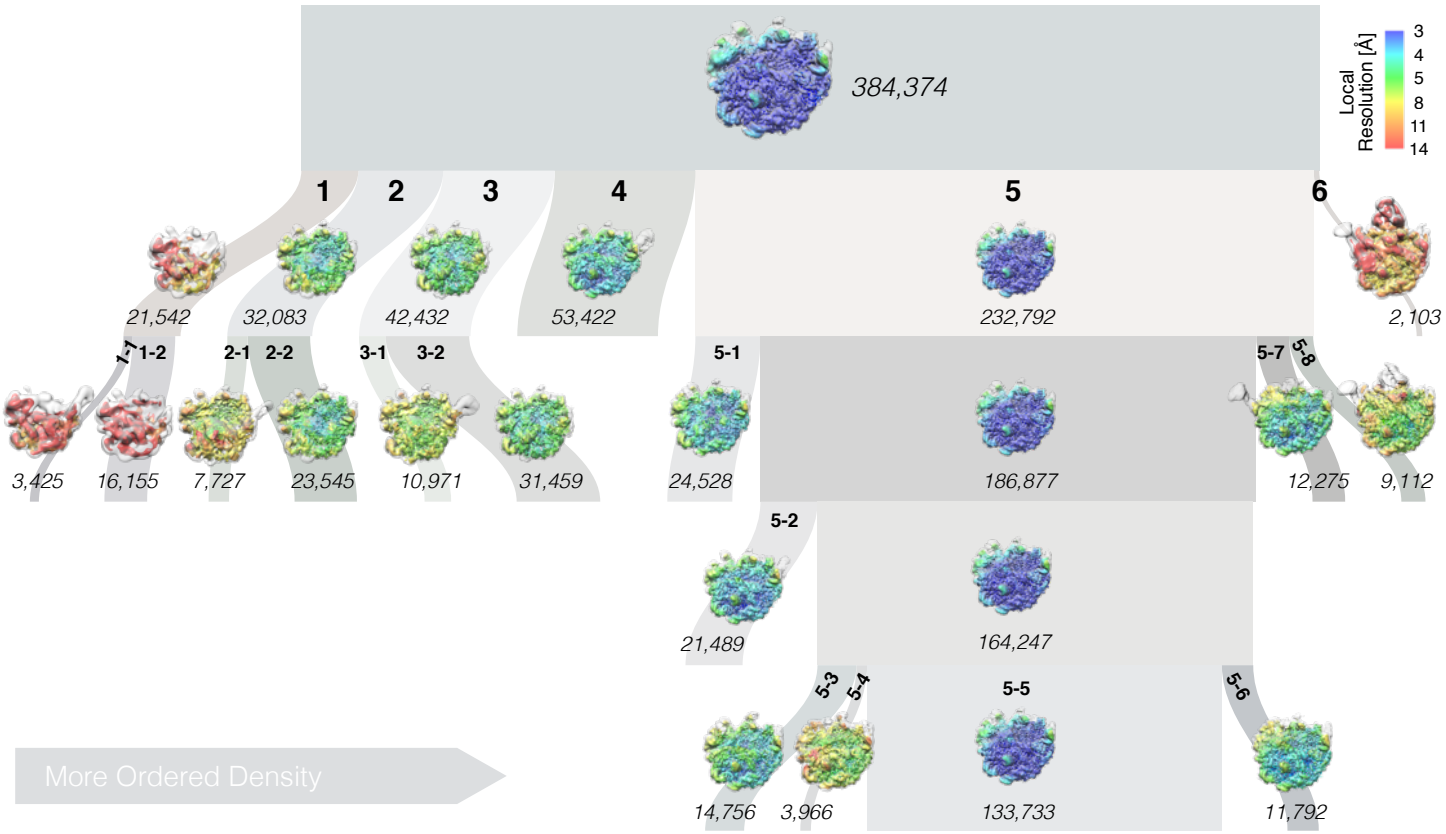

Figure 1

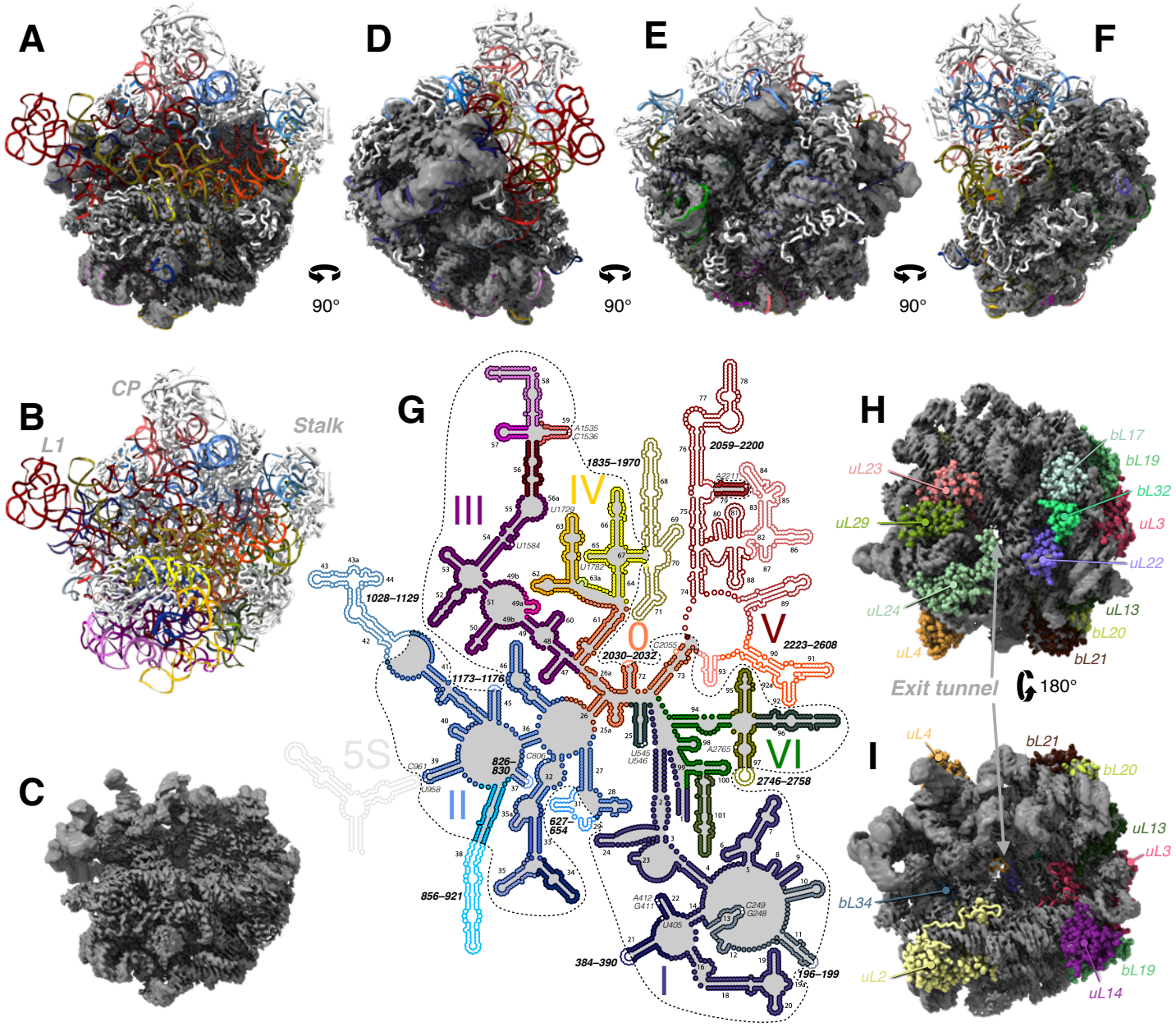
A
D
E
F
90°
90°
90°
B
CP
L1
Stalk
C
G
III
IV
V
VI
II
I
O
5S
1835–1970
2059–2200
2223–2608
2746–2758
1028–1129
1173–1176
2030–2032
826–830
627–654
856–921
384–390
196–199
H
bL17
bL19
bL32
uL23
uL29
uL3
uL22
uL24
uL13
bL20
uL4
bL21
Exit tunnel
180°
I
uL4
bL21
bL20
uL13
uL3
bL34
uL2
bL19
uL14

Figure 2

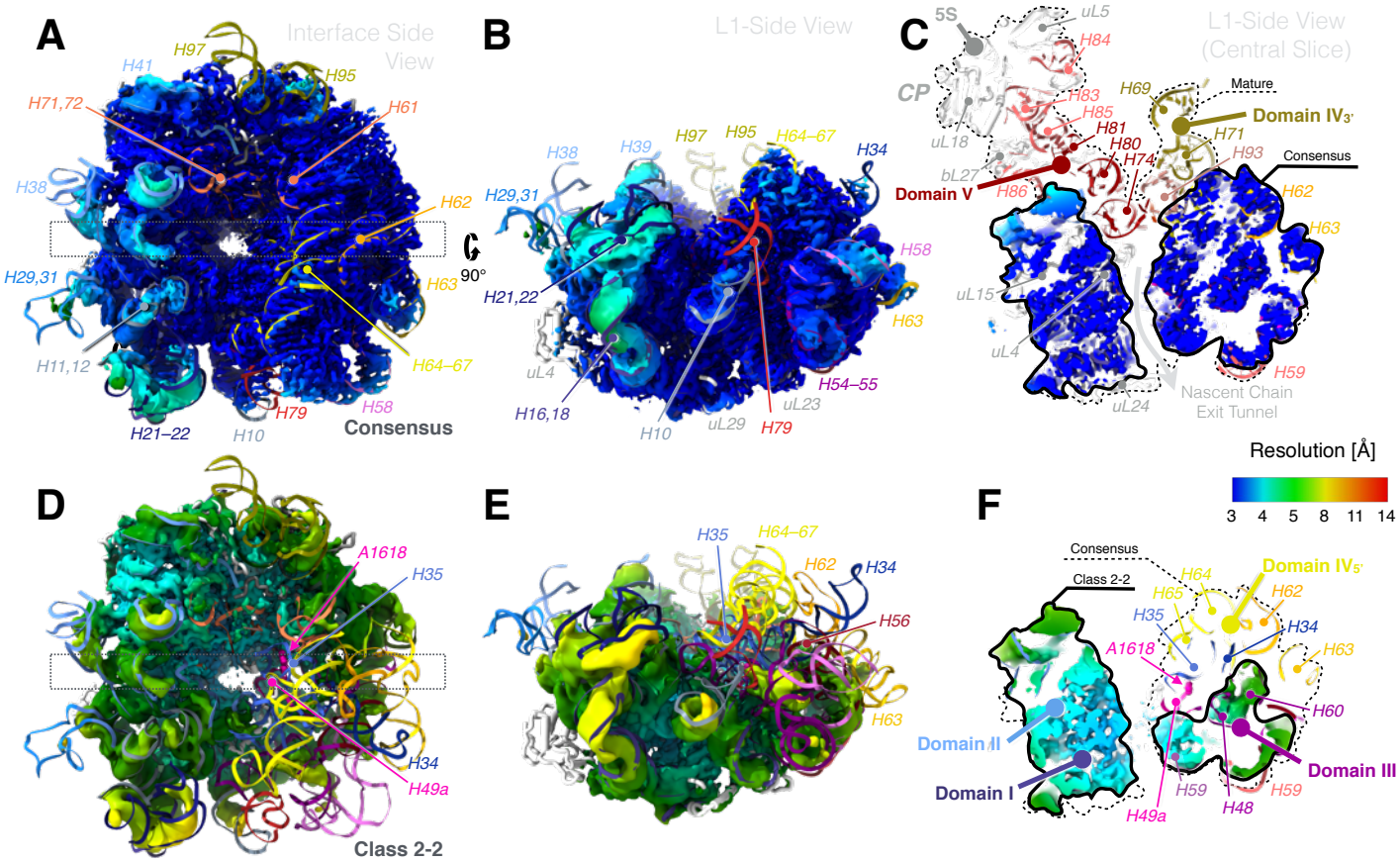
A
Interface Side View
H97
H41
H95
H71,72
H61
H38
H62
H29,31
H63
H11,12
H64–67
H79
H58
H21–22
H10
Consensus
B
L1-Side View
H38
H39
H97
H95
H64–67
H34
H29,31
90°
H58
H21,22
H63
uL4
H54–55
H16,18
uL23
H10
uL29
H79
C
5S
uL5
H84
L1-Side View (Central Slice)
CP
H83
H69
Mature
H85
Domain IV3'
H81
H71
uL18
H80
H74
H93
Consensus
bL27
Domain V
H86
H62
H63
uL15
uL4
H59
Nascent Chain Exit Tunnel
uL24
Resolution [Å]
3 4 5 8 11 14
D
A1618
H35
H34
H49a
Class 2-2
E
H35
H64–67
H62
H34
H56
H63
F
Consensus
Class 2-2
H64
Domain IV5'
H65
H62
H35
H34
A1618
H63
H60
Domain II
Domain III
Domain I
H59
H59
H49a
H48

Figure 3

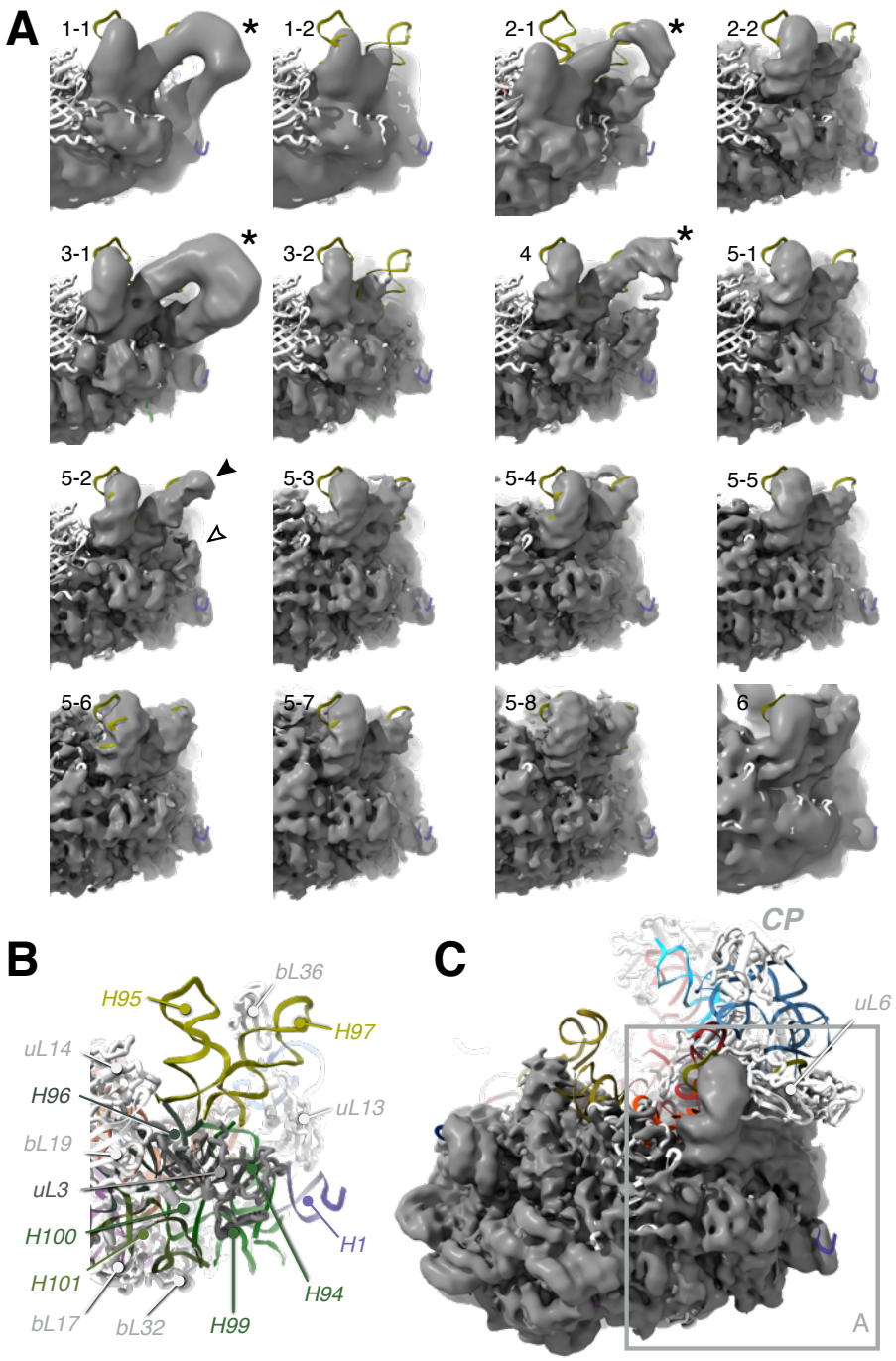
A
1-1
1-2
2-1
2-2
3-1
3-2
4
5-1
5-2
5-3
5-4
5-5
5-6
5-7
5-8
6
B
H95
bL36
H97
uL14
H96
uL13
bL19
uL3
H100
H1
H101
H94
bL17
bL32
H99
C
CP
uL6
A

Figure 4

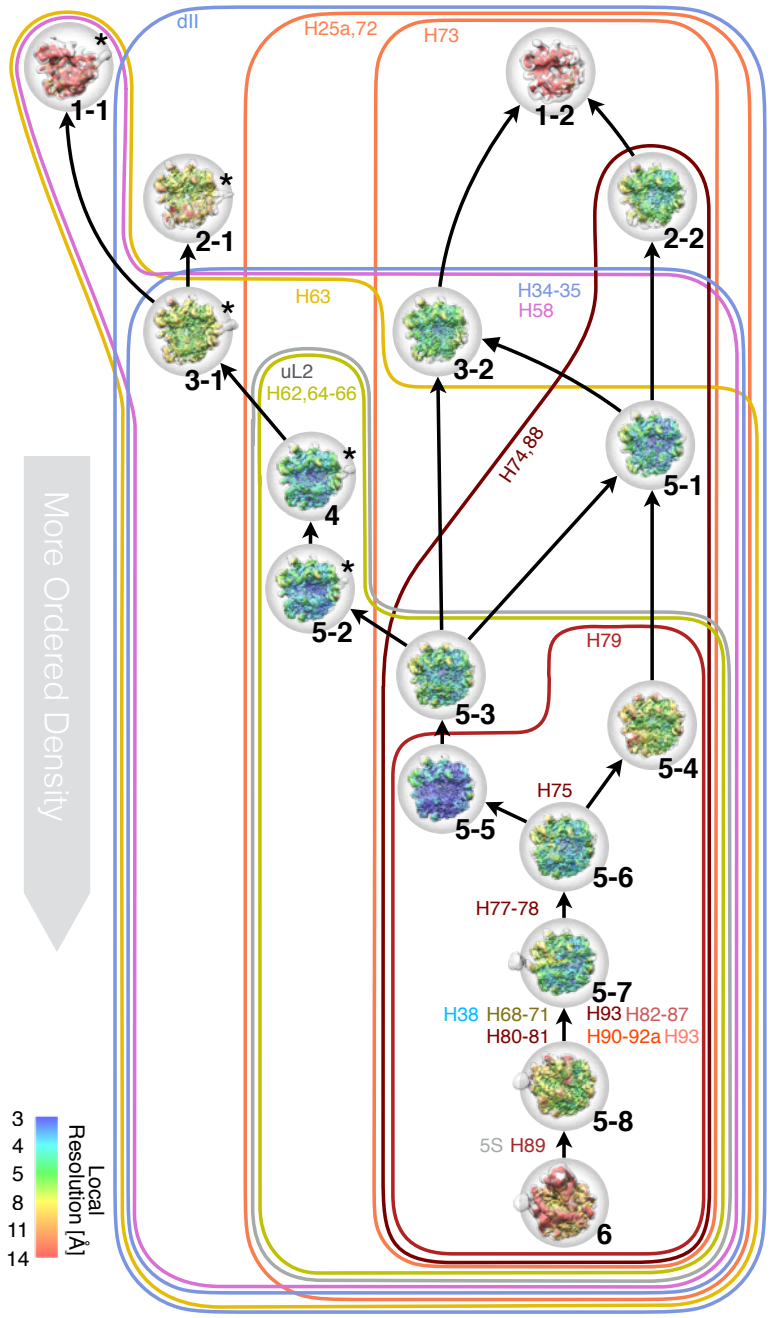
dII
H25a,72
H73
1-1
1-2
2-1
2-2
H63
H34-35
H58
3-1
3-2
uL2
H62,64-66
H74,88
5-1
4
5-2
H79
5-3
5-4
H75
5-5
5-6
H77-78
5-7
H38 H68-71
H80-81
H93 H82-87
H90-92a H93
5-8
5S H89
6
More Ordered Density
3
4
5
8
11
14
Local Resolution [Å]

Figure 5

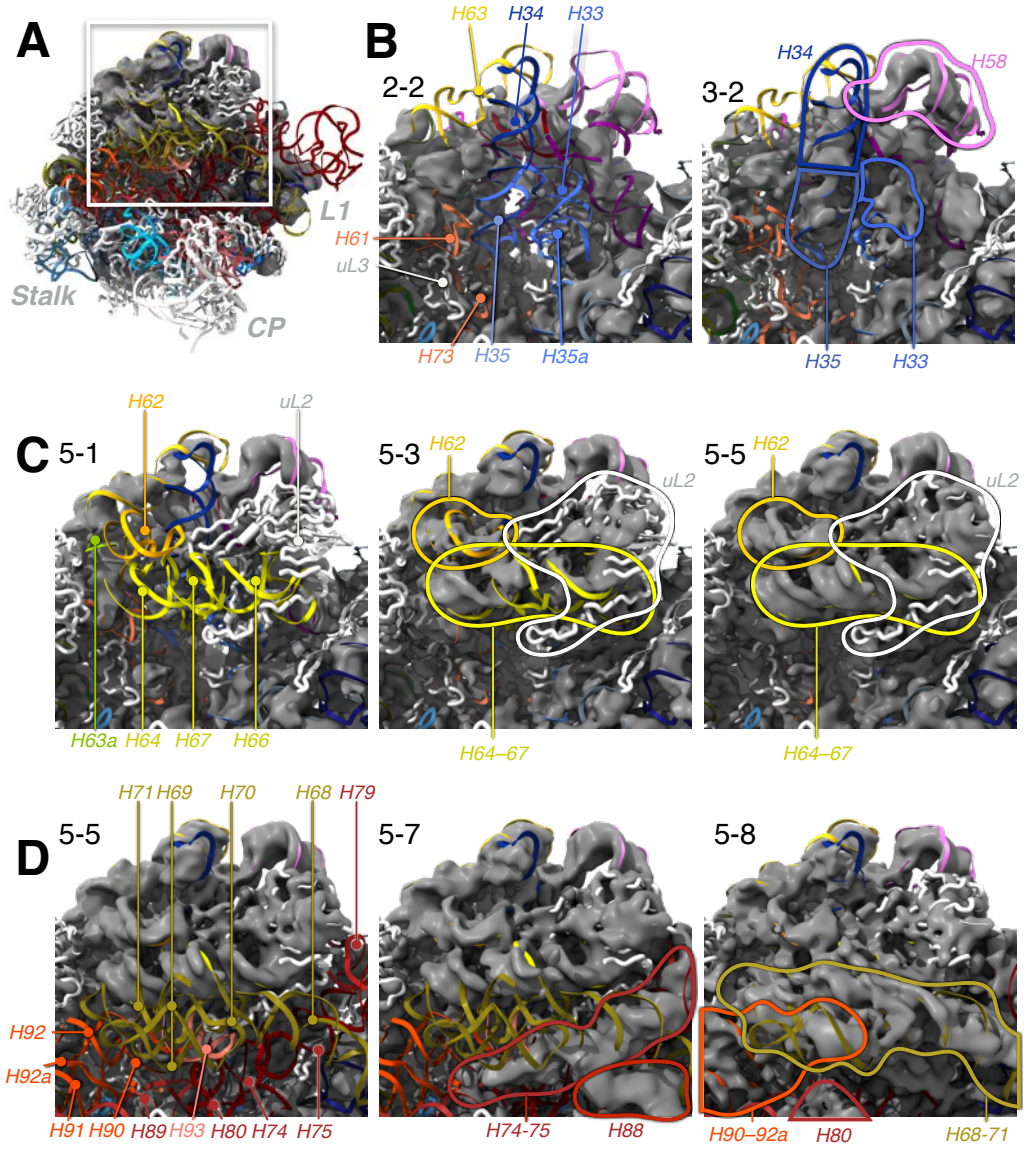
A
L1
Stalk
CP
B
2-2
H63
H34
H33
H61
uL3
H73
H35
H35a
3-2
H34
H58
H35
H33
C
5-1
H62
uL2
H63a
H64
H67
H66
5-3
H62
uL2
H64–67
5-5
H62
uL2
H64–67
D
5-5
H71
H69
H70
H68
H79
H92
H92a
H91
H90
H89
H93
H80
H74
H75
5-7
H74-75
H88
5-8
H90–92a
H80
H68-71

Figure 6

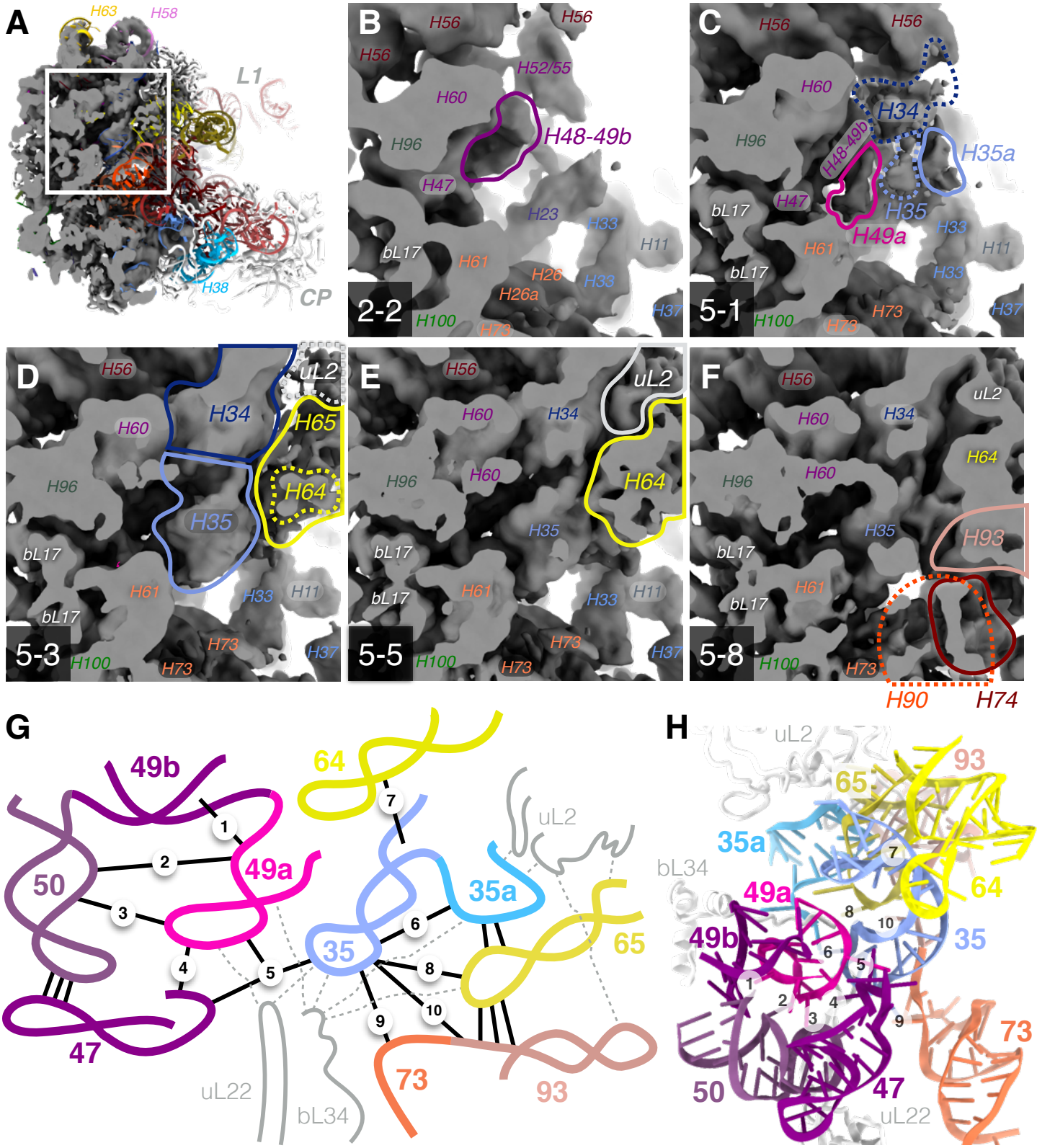
A
H63
H58
L1
H38
CP
B
H56
H56
H56
H52/55
H60
H48-49b
H96
H47
H23
H33
H11
bL17
H61
H26
H33
H26a
H37
2-2
H100
H73
C
H56
H56
H60
H34
H96
H48-49b
H35a
bL17
H47
H35
H49a
H33
H61
H11
bL17
H33
5-1
H100
H73
H73
H37
D
H56
uL2
H34
H65
H60
H96
H64
H35
bL17
H61
H33
H11
bL17
H73
5-3
H100
H73
H37
E
H56
uL2
H60
H34
H96
H60
H64
H35
bL17
H61
H33
H11
bL17
H73
5-5
H100
H73
H37
F
H56
uL2
H60
H34
H64
H96
H60
H35
H93
bL17
H61
bL17
5-8
H100
H73
H90
H74
G
49b
64
1
7
uL2
2
49a
50
3
35a
6
35
65
4
5
8
10
9
47
uL22
bL34
73
93
H
uL2
93
65
35a
7
bL34
49a
64
8
10
49b
35
6
5
1
2
4
3
9
73
50
47
uL22

Figure 7