

Supplementary Materials for

In-cell architecture of an actively transcribing-translating expressome*

Francis J. O'Reilly^{1,†}, Liang Xue^{2,3,†}, Andrea Graziadei^{1,†}, Ludwig Sinn¹, Swantje Lenz¹, Dmitry Tegunov⁴, Cedric Blötz⁵, Neil Singh⁵, Wim J. H. Hagen², Patrick Cramer⁴, Jörg Stülke⁵, Julia Mahamid^{2,*}, Juri Rappsilber^{1,6,*}

Correspondence to: julia.mahamid@embl.de, juri.rappsilber@tu-berlin.de

This PDF file includes:

Materials and Methods
Supplementary Text
Figs. S1 to S28
Tables S4 to S7
Captions for Movies S1 to S2
References (31-91)

Other Supplementary Materials for this manuscript include the following:

Movies S1-S2
External Data Tables S1-S3

* This manuscript has been accepted for publication in Science. This version has not undergone final editing. Please refer to the complete version of record at <http://www.sciencemag.org/>. The manuscript may not be reproduced or used in any manner that does not fall within the fair use provisions of the Copyright Act without the prior, written permission of AAAS

Materials and Methods

Mycoplasma pneumoniae cultivation for CLMS

M. pneumoniae strain M129 (ATCC 29342) in the 32nd broth passage was used. Bacterial cells were grown at 37°C in 150 cm² tissue culture flasks containing 80 ml of modified Hayflick medium (31): 14.7 g/L Difco™ PPLO (Becton Dickinson, USA.), 20% (v/v) Gibco™ horse serum (New Zealand origin, Life Technologies, Carlsbad, USA), 100 mM HEPES-Na (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; pH 7.4), 1% (w/w) glucose as a carbon source, 0.002% (w/w) phenol red and 1000 U/mL penicillin G.

Crosslinking of the cells

Cells were harvested from media by centrifugation at 16,000 g for 5 min, followed by two washes with 500 µl 1x Phosphate-buffered saline (PBS) buffer. Crosslinker was dissolved freshly in water-free Dimethyl sulfoxide (DMSO) at 50 mM. The crosslinker, either disuccinimidyl suberate (DSS, Thermo Scientific) or disuccinimidyl sulfoxide (DSSO, Thermo Scientific) was quickly added to cells resuspended to a final concentration of 1 mg/ml cells in 1x PBS solution at a final concentration of 2 mM (5% DMSO in final volume). The cells were incubated at 25°C for 30 min under gentle agitation, followed by quenching with 50 mM Tris-HCl (pH 7.5) for 15 min. Cells were pelleted and snap-frozen with liquid N₂ before storing at -80°C.

Cell lysis and proteolysis for LC-MS Analysis

For proteome extraction, the cell pellet was lysed at 20 mg/ml in denaturing solution (6 M Urea, 2 M Thiourea, 50 mM ammonium bicarbonate (ABC, Sigma), 1 mM dithiothreitol (DTT, Sigma)), aided by sonication using a 3.2 mm Qsonica microtip probe (output 1.5, duty cycle 0.2; 20 min per sample; on ice).

Protein concentration of the cell lysate was assessed by Bicinchoninic acid assay (BCA, Pierce) according to the manufacturer's instructions and was diluted to 1 mg/ml with the denaturing solution. The lysate was reduced for 30 min by adding 10 mM DTT, followed by addition of 25 mM iodoacetamide (Sigma) and incubation in the dark for 20 min. The reaction was quenched by adding 10 mM DTT. Proteolysis was carried out using LysC (Pierce) at 1:100 (mass ratio, protease:protein) for 4 h at 37 °C, followed by 1:5 dilution with 50 mM ABC and then trypsin digestion (1:50 (m/m), Pierce) for 16-18 h at 37°C. The reaction was stopped by adding 10% trifluoroacetic acid (TFA, Honeywell) to pH 2-3. Samples were cleaned-up via Stage-tipping (3M) and stored at -20°C until further use, i.e. peptide fractionation.

Peptide fractionation by Strong Cation Exchange Chromatography (SCX)

Peptide digests were eluted from Stage-tips, dried using a vacuum concentrator and resuspended in SCX running buffer A (30% acetonitrile (ACN, VWR GmbH), 10 mM KH₂PO₄ pH 3.0 (Sigma)). Separation of peptides was accomplished using a non-linear gradient with running buffer B (30% ACN, 1 M KCl, 10 mM KH₂PO₄, pH 3.0), as described (32). For the fractionation, a Shimadzu HPLC system (LC-20AD, SIL-20A HT, SPD-M20A, CBM-20A) connected to a PolyLC Polysulfoethyl A column (100 x 2.1 mm, 3 µm, 300 Å) was used. Fractions of 200 µl each were collected over the elution window (ca. 18 column volumes).

Collected fractions of interest from five runs were pooled, desalted using Stage-Tips and stored at -20°C.

Peptide fractionation by Size-Exclusion Chromatography (SEC)

Dried peptide digests or fractions were resuspended in SEC running buffer (30% ACN, 0.1% TFA) and injected onto a Superdex Peptide 3.2/300 column running at 50 µl/min flow rate connected to an Äkta Pure system. Fractions of 100 µl were collected and dried in a vacuum concentrator (33, 34) and stored at -20°C.

Peptide fractionation by hydrophilic Strong Anionic Exchange Chromatography (hSAX)

Peptides from Stage-Tips were dried and subsequently resuspended in hSAX running buffer A (20 mM Tris-HCl, pH 8.0). Sample was injected onto a Dionex IonPac AS24 column (2 x 250 mm) running with 150 µl/min operated at 15°C under Unicorn 7.1 on an Äkta Pure system. Elution of peptide mixtures was achieved by applying an exponential gradient with buffer B (20 mM Tris-HCl, pH 8.0, 1 M NaCl; gradient identical to SCX gradient), while collecting fractions of 150 µl. Obtained fractions of interest were acidified with TFA to pH 2-3, pooled, desalted by Stage-tipping and stored at -20°C.

Two-dimensional peptide fractionation for each CLMS dataset

Four datasets were collected to achieve significant depth of analysis (Fig. S1A). Dataset 1 - Peptides from cells crosslinked with DSS were fractionated by SCX. Then 11 fractions enriched for crosslinked peptides were fractionated individually by SEC as a second dimension. Three fractions enriched for crosslinked peptides from each SEC run were collected, resulting in 33 total fractions for MS acquisition.

Dataset 2 - Proteins were extracted and subjected to native SEC, as described in (35), and all fractions with protein complexes of molecular weight larger than 150 kDa were pooled for digestion and peptide separation as in Dataset 1.

Dataset 3 - Peptides from cells crosslinked with DSS were fractionated by SCX. Then 9 fractions enriched for crosslinked peptides were fractionated individually by hSAX into 8 pools as a second dimension. This resulted in 72 total fractions for MS acquisition.

Dataset 4 - Peptides from cells crosslinked with DSSO were fractionated by SCX to give 9 fractions enriched for crosslinked peptides. Each fraction was fractionated individually by hSAX, resulting in 9x15 total hSAX fractions.

LC-MS/MS acquisition of crosslinked samples

Acquisition of crosslinked peptide spectra was performed on a Fusion Lumos Tribrid Mass Spectrometer (Thermo Fisher Scientific, San Jose, USA) connected to an Ultimate 3000 UHPLC system (Dionex, Thermo Fisher Scientific, Germany). Mobile phase A (0.1% (v/v) formic acid), and mobile phase B (80% (v/v) acetonitrile, 0.1% (v/v) formic acid). The samples were dissolved in 1.6% acetonitrile (Fluka), 0.1% formic acid (Fluka) and separated on an Easy-Spray column (C-18, 50 cm, 75 µm internal diameter, 2 µm particle size, 100 Å pore size) running with 300 nl/min flow rate using optimized gradients for each offline-fraction (ranging from 2% B to 55% B over 62.5, 92.5 or 152.5 min (for datasets 1, 2-3, and 4 respectively), then to 55% in 2.5 min and to 95% in 2.5 min). Fractions were injected twice when more than 3 µg peptide material was available.

The MS data were acquired in data-dependent mode using the top-speed setting with a 3 second cycle time. For every cycle, the full scan mass spectrum was recorded in the Orbitrap at a resolution of 120,000 in the range of 400 to 1,600 m/z. Ions with a precursor charge state between +3 and +6 were isolated and fragmented. The fragmentation regime depended on the crosslinking reagent employed: For the non-cleavable linker DSS, higher-energy collisional dissociation (HCD) energies optimized for mass and charge of a precursor species were applied (36). For the cleavable DSSO crosslinker, stepped fragmentation energies of 18, 24 and 30 normalized collision energy (NCE) were used that were found appropriate (37). The fragmentation spectra were then recorded in the Orbitrap with a resolution of 30,000. Dynamic exclusion was enabled with single repeat count and 60 second exclusion duration.

Identification and statistical validation of crosslinked peptides

Raw data from mass spectrometry was processed using MSconvert (version 3.0.11729) (38) including denoising (top 20 peaks in 100 m/z bins) and conversion to mgf-file format. Masses of precursor peptides were recalibrated. Obtained peak files were analysed using xiSearch 1.6.746 (34) with the following settings: MS1/MS2 error tolerances 4 and 10 ppm, allowing up to 3 missing isotope peaks (Lenz et al., 2018), tryptic digestion specificity, carbamidomethylation on Cys as fixed and oxidation on Met as variable modification, losses: $-\text{CH}_3\text{SOH}$ / $-\text{H}_2\text{O}$ / $-\text{NH}_3$, crosslinker DSS (138.06807 Da linkage mass) or DSSO (158.0037648 Da linkage mass) with variable crosslinker modifications (“DSS-NH2” 155.09463 Da, “DSS-OH” 156.07864 Da, “DSSO-NH2” 175.03031 Da, “DSSO-OH” 176.01433 Da). For samples crosslinked with DSSO, additional loss masses were defined accounting for its cleavability (“A” 54.01056 Da, “S” 103.99320 Da, “T” 85.98264). Crosslink sites for both reagents were allowed for side chains of Lys, Tyr, Ser, Thr and the protein N-terminus. The database consisted of 687 Swiss-Prot annotated entries for *M. pneumoniae* strain M129 (ATCC 29342; taxon identifier 272634), 19 small open reading frames that were recently annotated (21), and annotated lysine acetylation sites (39). Decoy sequences were generated by reversing the protein sequences.

Spectral matches were filtered prior to FDR to crosslinked peptide matches having a minimum of four matched fragments per peptide. DSSO matches required an A, S or T peak for each peptide. Results were then filtered to an estimated false-discovery rate (FDR) of 5% on residue-pair-level and 5% on protein-protein interaction (PPI)-level using xiFDR (version 1.2.30.59) (40). Inter- and intra-crosslinks were handled separately for FDR estimation. The DSS datasets (datasets 1-3) were validated together for FDR estimation. Spectral matches in the DSSO dataset were prefiltered before FDR estimation to only those that had cleaved crosslinker peptide fragments for both peptides.

LC-MS/MS acquisition and data analysis for bottom-up proteomics

For identifying proteome abundances, we subjected the proteome to bottom-up proteomics. The Proteome was extracted as outlined above. Mass spectrometric analysis of native (non-crosslinked) *M. pneumoniae* proteome was performed on a Lumos Tribrid Mass Spectrometer (Thermo Fisher Scientific, San Jose, USA) connected to an Ultimate 3000 RSLCnano system (Dionex, Thermo Fisher Scientific, Germany). The proteome was analysed in triplicate, each with a total run time of 230 min. The chromatographic setup was identical to the one used for crosslinking acquisitions with the following changes on the LC gradient: The gradient started at

2% B to 5% B in 1 min, to 7.5% B in 5 min, then to 32.5% in 190 min, 40% B in 7 min, 50% B in 2.5 min followed by ramping to 90% B in 2.5 min and washing for 5 min. The settings of the mass spectrometer were as follows: 2 second cycle time; Data-dependent mode; MS1 scan at 120,000 resolution over 350 to 1,600 m/z; automatic gain control (AGC) target of $3e^6$ with maximum Injection time (IT) of 50 ms; MS2 triggered only on precursors with charge state between +2 and +7; 1.6 m/z isolation width; AGC target of $1e^5$ with 80 ms max. IT and min. AGC target of $2.5e^4$; fragmentation by HCD 29; MS2 scan was ion trap in Rapid mode; peptide match was set as preferred and dynamic exclusion was enabled upon single observation for 60 seconds.

Mass spectrometry raw data were processed using MaxQuant 1.6.1. with default settings with minor changes: three allowed missed cleavages; up to four variable modifications per peptide including oxidation on Met, acetylation on protein N-terminal. Carbamidomethylation on Cys was set as fixed. The database covered 706 proteins (similar to crosslinking search, see above). The ‘matching between runs’ feature was disabled with default settings. Protein quantification was based on two or more peptides using the iBAQ approach (41). Proteins are reported at 1% FDR (Table S2).

Bacterial two-hybrid assay (BACTH)

Primary protein-interactions were validated by bacterial two-hybrid analysis (42). All oligos and plasmids referred to are found in Table S3. The BACTH system is based on the interaction-mediated reconstruction of *Bordetella pertussis* adenylate cyclase (CyaA) activity in *E. coli*. Functional complementation between two fragments (T18 and T25) of CyaA as a consequence of the interaction between bait and prey proteins results in cAMP synthesis. The β -galactosidase reporter gene was fused to a cAMP-dependent promoter which allows indirect detection of the protein-protein interaction by assaying β -galactosidase activity. Plasmids pUT18/pUT18C and p25N/p25 allow the expression of proteins fused to the C- or N-terminal T18 and T25 fragments of CyaA, respectively. For these experiments, we constructed the plasmids pGP3280-3297, pGP295-300, pGP1451-GP1453. DNA fragments corresponding to the genes *mpn530*, *sigA* (*mpn352*), *rpoA* (*mpn191*), *rpoB* (*mpn516*), *rpoC* (*mpn515*), and *spxA* (*mpn266*) were obtained by PCR, digested and ligated into the vectors pUT18/pUT18C and p25N/p25 (Table S3). These plasmids were used for co-transformation of *E. coli* BTH101, and the protein-protein interactions were then analyzed by plating the cells on LB plates containing 100 μ g/ml ampicillin, 50 μ g/ml kanamycin, 80 μ g/ml X-Gal (5-bromo-4-chloro-3-indolyl- β -D-galacto-pyranoside), and 1 mM Isopropyl β - d-1-thiogalactopyranoside/Isopropyl β -d-1-thiogalactopyranoside (IPTG). The plates were incubated for a maximum of 36 h at 28°C.

Construction of NusA truncation mutants

To verify the essentiality of the C-terminal region of NusA, deletion of this region after positions I⁴¹⁸ or K⁴⁴⁰ was attempted using a recombineering approach (43). We replace the section of the gene downstream of the truncation sites with the antibiotic resistance gene, chloramphenicol acetyltransferase, to allow selection of successfully transformed cells with chloramphenicol. All oligos and plasmids referred to are found in Table S3.

To construct the recombination substrate, Gibson assembly was used. The common homologous downstream region (NS47, NS48, 380 bp) was amplified from genomic DNA of *M. pneumoniae* strain M129 with an overlapping flank on one end. Two different homologous

upstream regions (380 bp) corresponding to post-I⁴¹⁸ (NS49-50) and post-K⁴⁴⁰ (NS59-50) truncations were amplified in the same way. Finally, the antibiotic resistance genes - chloramphenicol acetyltransferase (cat) (from plasmid pGP2727, NS52-53/56) and tetracycline resistance (tetM) (from plasmid pGP2777, NS54-55) were amplified with flanking lox sites and P48 promoter upstream. These purified PCR products (QIAquick PCR Purification kit, QIAGEN) were fused together by running 10 PCR hybridization cycles (no primers) followed by adding the terminal primers for 25 PCR amplification cycles. These products were digested with XhoI and NotI, and cloned between the corresponding sites of the pBluescript II SK(-) vector backbone. A total of three plasmids were constructed, two for I⁴¹⁸ NusA truncation pGP3376 (cat) and pGP3377 (cat-tetM) and one for K⁴⁴⁰ pGP3378 (cat). As a control, plasmid pGP2694 (cat), a gene deletion plasmid for the non-essential gene mpn668 was used.

These plasmids were used to amplify double stranded DNA in which the two strands were tagged differently. The strand used for transformation of *M. pneumoniae* was protected by phosphorylation (NS57) while the other strand was biotinylated (NS58). The biotin tag allows the separation of the two strands with magnetic Streptavidin-Beads (Pierce). The protected ssDNA was purified, dried and re-suspended in electroporation buffer (8 mM HEPES, 272 mM sucrose, pH 7.4). The ssDNA was used as the transformation substrate for the deletion of the NusA C-terminal region in the strain GPM116 (expressing GP35 recombinase (mpn560)-arcA::GP35-Puro^R). For deletion of osmC, ssDNA was prepared as described above using the primer pair LK47/ LK48 for protection and biotinylation.

Transformation into *M. pneumoniae* was done using a standard electroporation protocol (31). Colonies were passaged and checked via PCR-based assays for the presence of the chloramphenicol resistance gene (NS73-74) and the correct location of the homologous recombination event (NS46, NS51 for the NusA plasmids, CB51, CB52 for the control). In case of the control, after verifying the deletion event had occurred, the resistance gene was excised leaving a loxE (lox-sites flank the resistance gene) scar. However, for both nusA constructs, no viable transformants were obtained.

M. pneumoniae cultivation and vitrification for cryo-electron microscopy (cryo-EM)

Wild-type *M. pneumoniae* strain M129 (ATCC 29342) were cultivated in cell culture flasks (Greiner Bio-One, Frickenhausen, Germany) with modified Hayflick medium (31) as described above. For cryo-EM grid preparation, 200 mesh gold (Au) grids with holey Quantifoil support (R2/1, Quantifoil Micro Tools, Jena, Germany) were glow discharged for 45 s and sterilized under UV irradiation in a laminar flow hood for 30 min, before immersing into culture medium. After inoculation, *M. pneumoniae* cells were grown at 37°C for less than 20 hours to ensure that cells were in the fast-growing phase. Vitrification was carried out using a manual plunger (Manufactured by the Max Planck Institute of Biochemistry, Martinsried, Germany). Grids with cells were quickly washed with PBS solution containing 10 nm protein A-conjugated gold beads (Aurion, Wageningen, Netherlands), blotted for 2 seconds from the back side, and plunged into a liquid ethane/propane mixture at liquid N₂ temperature. For antibiotic-treated datasets, either chloramphenicol (Cm; Sigma-Aldrich, USA) at a final concentration of 0.5 mg/ml or pseudouridimycin (PUM; Adipogen AG, Switzerland) at a final concentration of 0.4 mg/ml were added directly into the culture medium 15-20 minutes prior to vitrification. The frozen grids were stored in sealed boxes in liquid N₂ until further processing.

Cryo-electron tomography

Cryo-electron tomography data were collected on a Titan Krios transmission electron microscope (TEM) operated at 300 kV (ThermoFisher Scientific, Eindhoven, Netherlands) equipped with a field-emission gun, a Quantum post-column energy filter (Gatan, Pleasanton, CA, USA), a K2 Summit direct detector camera (Gatan) and a Volta phase plate (VPP). Images were recorded in dose-fractionation mode using acquisition procedures in SerialEM software v3.7.2 (44). Prior to tilt-series acquisition, montages of grid squares were acquired at 2.2 nm/pixel (Fig. S5A). Tilt-series were collected with a dose-symmetric scheme (45), Energy-Filtered TEM operated in zero-loss, nano-probe mode, magnification 81,000x, calibrated pixel size at the specimen level of 1.7 Å, defocus range 1.5 to 3.25 µm, tilt increment 3° with constant dose for all tilts, tilt range -60° to 60°, total dose of ~ 120 e⁻/Å². 14 tilt-series were acquired with the Volta phase plate (VPP) (Fig. 2A), primarily for generation of a *de novo* reference for template matching (described below). Alignment and operation of the VPP were essentially carried out as described previously, applying a beam tilt of 10 mrad for autofocusing (46). Tilt-series for high-resolution sub-tomogram analysis were acquired without phase plate, with a 70 µm objective aperture and a beam tilt of 4 mrad for autofocusing (Fig. S5B). In total, 500 tomograms were acquired: 352 for native, untreated cells, 65 for Cm-treated and 83 for PUM-treated cells. Tomograms were acquired on the following number of grids representing independent biological replicates: 6 for untreated cells, 1 for Cm-treated and 1 for PUM-treated cells.

Tomogram reconstruction

Prior to tilt-series alignment, movies from the K2 camera were corrected for beam-induced motion with the SerialEM plugin. Tilt-series alignment using fiducial gold beads and tomographic reconstruction were performed in etomo from the IMOD software package v4.9 (47). Aligned images were 4 times binned to a pixel size of 6.8 Å. For tomographic reconstruction, the radial filter options were left as default (cut off, 0.35; fall off, 0.05). Cell thickness was estimated using central YZ-slices of the tomograms.

Ribosome localization using template matching

To generate a data-driven template for correlation-based ribosome localization as implemented in pyTOM (48), 404 ribosomes were manually selected from 7 VPP tomograms using EMAN2 boxer (49). 4x binned sub-tomograms were extracted and refined in RELION (version 3.0) using a featureless sphere as an initial reference. The refinement converged into a ribosome density (Fig. S5D). The density was low-pass filtered to 30 Å resolution using TOM toolbox (50) implemented in MATLAB (MathWorks 2016), and used for template matching in 4x binned conventional defocus tomograms. For each tomogram, the 400 highest-scoring cross-correlation peaks were extracted, with a minimal Euclidean distance of 20.4 nm (30 voxels) between peaks to avoid multiple detection events for the same ribosome. The peaks were visually inspected to exclude obvious false positives, such as membranes and edges of the carbon support. With an initial dataset of 20 defocus tomograms and 4,668 sub-tomograms, a ribosome density of 25 Å resolution was obtained. The ribosome density was low-pass filtered to 30 Å and used as the new template for matching in all defocus tomograms. For each tomogram, the top 400 hits were visually inspected, and on average 307 hits were retained after removing obvious false positives

(Fig. S5E,F). In total, 108,501 ribosome sub-tomograms were extracted from the untreated cells tomograms, 21,299 from Cm-treated and 22,089 from PUM-treated datasets.

Due to its small cell size, 70%-90% of a *M. pneumoniae* cell could be sampled in one tomogram acquired at a pixel size of 1.7 Å/pixel. Approximately 300 ribosomes (including both 50S and 70S) could be localized in one tomogram, which equals to 300 to 500 ribosomes per *M. pneumoniae* cell, higher than those reported previously (15, 51). By comparing the proportion of ribosomes following classification with the correlation coefficient results from template matching (Fig. S5G), we found that the percentage of 70S ribosomes decreased with the correlation coefficient, and at the threshold used here for peak extraction from template matching (400 top scores), approximately 90% of all ribosomes could be retrieved. Therefore, the current sub-tomogram dataset retrieved the majority of 70S ribosomes inside the cell.

Sub-tomogram analysis

The sub-tomogram analysis workflow described below is summarized in Figure S6. All analysis steps were performed using Warp (version 1.0.6) and RELION (version 3.0), mostly following previously published protocols (52, 53).

Defocus parameters were determined using Warp on individual tilt images and consecutively refined on the whole tilt-series following tilt-series alignment in etomo. Phase modulations introduced by the contrast transfer function (CTF) and radiation damage were corrected using 3D CTF/wedge models reconstructed for each sub-tomogram in Warp. Ribosome sub-tomograms based on coordinates retrieved by pyTOM template matching were reconstructed at 2x binning (3.4 Å/pix) in Warp. These were subsequently sorted into different classes using RELION 3D classification (54). Sub-tomograms that could not be aligned and 50S ribosomes were excluded first (Fig. S7). 70S ribosomes were further classified to separate those arranged in a closely assembled polysome configuration (Fig. S7). Subsequent rounds of focused 3D classification with a mask at the 30S mRNA entry site enabled classifying the remaining 70S sub-tomograms into 3 major subclasses: ribosomes with well-defined RNAP density near the mRNA entry site, ribosomes with undefined additional density, and ribosomes without additional density (Fig. S7). At least 3 independent classification runs with different masks and settings were performed to validate the reproducibility of the classification results. Refinement using RELION 3D auto-refine focused on the ribosome density of 73,858 70S ribosomes attained a density at 10.3 Å resolution as determined by gold standard Fourier shell correlation (FSC) at 0.143. To improve the resolution, the sub-tomogram poses and tilt-series geometry were refined in a new tool associated with Warp, *M* (see details in the following section). With the same sub-tomogram dataset re-extracted at the original pixel size of 1.7 Å after *M* refinement, the final overall resolution improved to 5.6 Å (Fig. S8A, B). Local resolution map, angular distribution analysis, as well as 3D FSC analysis (55), validated the refinement results (Fig. S8 C-E), while densities corresponding to secondary structural elements confirmed the reported resolution (Fig. S8 F-H).

Different refinement strategies were performed with the 2,952 ribosome sub-tomograms exhibiting clear RNAP density. Refinement with an overall spherical mask generated a density of 8.5 Å overall resolution (Fig. S13A). However, the resolution of the RNAP density was in the range of ~ 12 Å (Fig. S13B). To avoid the expected influence of the ratcheting movement between the small and large ribosomal subunits, while keeping the RNAP-30S interface fully covered, a focused refinement with a local RNAP-30S mask was carried out (Fig. S13C). Although the

overall resolution dropped to 9.2 Å, the density quality of the RNAP improved (Fig. S13D). High flexibility of RNAP itself and the interface with the small ribosomal subunit hindered achieving higher local resolution. To overcome this, multi-body refinement as implemented in RELION was performed (56). Following one consensus refinement, 3 bodies corresponding to the three major parts of the supercomplex were segmented (i.e. RNAP and linker density, small and large ribosomal subunits), and a local mask was created for each body (Fig. S14A-D). Principal component analysis (PCA) on the relative orientations of all bodies was performed to characterise the most predominant motions (Fig. S14E and Movie S1).

Refinement and multi-body refinement of the 17,202 ribosomes with additional density did not produce better resolved structures, nor did further classification runs with different settings (local masks, class number, angular search range, etc.), suggesting that these sub-tomograms may have both high compositional heterogeneity and large conformational flexibility.

Similar sub-tomogram analysis, including classification, refinement and multi-body refinement were performed for the Cm-treated dataset (Fig. S25) and PUM-treated dataset (Figs. S26-27 and Movie S2). The resolution of all sub-tomogram averages, referred to as cryo-EM densities or maps, was estimated based on FSC of two independent data halves using FSC=0.143 as the cutoff criterion.

Disruption or alteration of the supercomplex architecture following either Cm or PUM treatment (Fig. 4B) indicated that assembly of the elongating expressome might be highly dynamic and transient inside the cell. This was further exemplified by the dramatic shifts observed in ribosome populations (Fig. 4A), following relatively short perturbation time (~15 minutes).

Warp and M processing pipeline for sub-tomogram analysis

To obtain higher resolution for the 3D averages, the particle (sub-tomogram) poses and tilt-series alignments were refined in a new software tool, *M*. Rather than following single particle analysis' central paradigm of treating each particle as a single, isolated entity, *M* implements a multi-particle analysis approach. All particles in a single tilt-series are treated as parts of the same physical system, which has its own set of hyperparameters that are optimized simultaneously with the individual particle poses.

In the case of tilt-series, the hyperparameters include the global rotation angles, coarse nonlinear image-space deformation to describe stage movement and beam induced motion, and the CTF models of individual tilt images, as well as a non-linear 3D deformation model of the tomographic volume describing its local changes as a function of the accumulated dose with a fine spatial, and coarse temporal resolution. Making use of the high amount of signal available for each ribosome particle, the particle poses themselves were modeled as a function of accumulated dose, with 3 optimizable sets of rotation and translation parameters per particle.

The non-linear deformation models were implemented as uniform 2D (image deformation in each tilt) and 4D (volume deformation over dose) grids of cubic splines, as described previously (53). The coarseness of the deformation models had a regularizing effect on the local divergence of individual 2D particle tilt image rotations and shifts, as previously formulated (57), and implemented in emClarity (58) and, for the special case of 2D data, Bayesian particle polishing in RELION 3.0 (59). Multi-particle refinement in *M* provides the first unified framework for performing reference-based particle alignment with physically plausible constraints for both 2D

and tilt-series data types in a single gradient descent optimization, and significantly improves the resolution of 3D reconstructions derived from such data.

Multi-particle refinement in M requires an initial tilt-series alignment, e.g. established using IMOD (47), as well as the particle poses, e.g. refined in RELION (60). At the beginning of each iteration, individual 2D particle tilt images are extracted from the original tilt-series. For extraction and subsequent refinement, 3D particle positions are related to their 2D positions in individual tilt images by applying the volume deformation model, the affine transform of the sample at the respective tilt angle, and the deformation model of the respective tilt image. The defocus value for each particle view is obtained by adding the Z coordinate of the transformed particle position to the global defocus value for the respective tilt image. Particle rotations are related to rotations in individual 2D images by applying the affine transform of the sample at the respective angle.

All per-particle parameters and the hyperparameters of the multi-particle system are optimized to maximize the cross-correlation between reference projections multiplied by the CTF and spectrally weighted according to accumulated dose and tilt angle, and the individual 2D particle tilt images summed over all particles and tilts. To achieve this in a gradient descent-like procedure, gradients must be calculated for all parameters with respect to the cost function in each optimization step. This calculation can be performed efficiently for particle poses using a finite differences numerical approach. However, most of the hyperparameters affect multiple particles, and a finite difference calculation must re-evaluate the cost function for all of these particles, making it computationally costly. Here, the same approach formulated in Warp (53) is employed: Before the optimization starts, partial derivatives are pre-computed for every affected 2D particle parameter with respect to each hyperparameter. During optimization, the hyperparameter gradients can then be efficiently calculated as a linear combination of the gradients of the affected 2D particle parameters.

The optimization is typically performed for 20 steps. At the end of the refinement of each tilt-series, individual 2D particle tilt images are back-projected into 3D volumes through Fourier-inversion (61), weighted by their local CTF and the dose- and angle-dependent spectral weights. Once refinement of all tilt-series is finished, reconstruction with Fourier-space gridding and amplitude correction is performed using the procedure described in RELION 1 (60) to obtain the final 3D maps. This approach, like emClarity (58), omits the lossy intermediate step of reconstructing individual sub-tomograms for each particle, and obtaining the final reconstructions by averaging these volumes, as implemented in other tools (50, 60, 62, 63). The improved reconstructions can be used in another refinement iteration to further improve the resolution. As all parameters are optimized simultaneously in each iteration, the procedure's convergence is rather quick, taking 3 to 4 iterations. At last, sub-tomograms and their corresponding 3D CTF/wedge models can be re-generated after M refinement, with the updated hyperparameters taken into calculation inside Warp.

Maximum-likelihood classification and other functions implemented in RELION enabled resolving the large heterogeneity and flexibility of the ribosomes and associated complexes. The obtained 5.6 Å 70S ribosome density following Warp/ M refinement shows the potential of this newly introduced sub-tomogram analysis workflow for resolving macromolecular structures at the level of secondary structures.

Visualization and analysis of sub-tomogram analysis results

RELION post-processing was used to calculate FSC and perform B factor sharpening. UCSF Chimera (64) and ChimeraX (65) were used for visualization and figure generation. Movies were generated using the “Volume Series” function in ChimeraX. Angular distribution maps were made by scripts described in cryo-EF (66). 3D FSC calculation and plots were performed in the online 3D FSC server (55). Sphericity values were calculated in 3D FSC analysis to measure directional resolution anisotropy resulting from preferential angular distribution. Quantification and plots were done in MATLAB (R2016b, The Mathworks, Inc. Massachusetts, USA).

Homology modeling and structure prediction

Sequence alignments were generated using clustal omega (67), disorder prediction was performed with IUPred2A (68) and secondary structure prediction using Jpred (69). Results were visualized in JalView (70).

The *M. pneumoniae* ribosome homology model was generated in Swiss-Model (71) using the *B. subtilis* MfiM-stalled ribosome (PDB 3J9W, (72)) as a template for the ribosomal proteins, while retaining the rRNAs from *B. subtilis*. The validity of the homology models was verified by CLMS, as well as using rigid-body fitting into the available densities (Fig. S9). Independent structural validation of crosslink data quality was performed by mapping crosslinks on homology models of additional complexes detected by CLMS, yielding a good agreement between crosslinks and predicted structures (Fig. S3), further confirming the native state of complexes in DSS- and DSSO-treated cells. For the ribosomal proteins L22 and L29, CLMS information was also used to guide the assignment of structured regions unique to *M. pneumoniae* that were observed in the density (Fig. S10-11). Disordered regions not localised in the density and presenting large crosslink violations were removed from the model. For Cm-treated cells, the ribosome homology model was based on the *T. thermophilus* chloramphenicol-bound ribosome (PDB 6ND5). The state of the ribosome in the stalled expressome from PUM-treated cells was determined by observation of the position of the L1 stalk, of EF-G and tRNA states (Fig. S28 (73)). The homology model was based on the viomycin-bound *E. coli* pre-translocation ribosome (PDB 4V7D, (30)).

Homology models of the *M. pneumoniae* RNAP were generated in Swiss-Model using the NusA-bound paused elongation complex from *E. coli* as template (PDB 6FLQ, (23)). Highly flexible loops and areas not covered in the sequence alignment were removed from the model and kept fully flexible in subsequent modeling steps (Table S4, “fully flexible coarse-grained regions”; Fig. S15). The NusG model was generated in Swiss-Model using NusG in the *E. coli* elongating RNAP complex as template (PDB 6C6U, (74)), and placed relative to the RNAP accordingly, generating a combined RNAP-NusG model.

The homology models of the ribosome and RNAP were fitted into densities for the RNAP-ribosome complex, ribosome alone and the PUM-stalled complex using the fitmap tool in UCSF Chimera (64). A homology model of NusA was generated using MODELLER (75) based on an alignment with known NusA structures for the N-terminal, S1 and KH domains in multi-domain assembler (MDA) (76). Due to the low sequence homology and predicted disorder, residues 366-540 were left fully flexible and not modelled as domains (see supplementary text and Fig. S18). Additionally, a homology model for the structured region (residues 1-80) of the firmicute-specific RNAP δ subunit, observed crosslinked to the RNAP core in a region with unassigned density, was generated using MODELLER. Homology model accuracy was assessed as reported in Table S4.

Integrative structure modeling: representation and sampling

The building blocks described above and in Table S4 were used to build an integrative model of the *M. pneumoniae* RNAP-ribosome supercomplex using the integrative modeling platform (IMP, version 2.12) (26). The overall modeling workflow is described in Fig. S19. Prior to modeling, RNAP-NusG and the ribosome were fitted into densities and kept fixed in subsequent steps. Unambiguity of fit was assessed by deriving a Benjamini-Hochberg-corrected p-value as in (77), and goodness-of-fit was judged by correlation around the mean. The p-value for the ribosome fit to the expressome density is $6.9e^{-13}$ (N=50,000), while the p-value for the RNAP density to the multi-body-refined expressome density is $1.29e^{-12}$ (N=50,000).

In a first step, the models were coarse-grained as rigid bodies comprising the 30S subunit, including the *B. subtilis* 16S RNA and P-site tRNA, the NusG-bound RNAP, each of the NusA domains and the RNAP δ subunit. Within each rigid body, regions that were previously removed (i.e. disordered and flexible regions, as well as regions for which no homology model could be generated) were coarse grained as beads and kept fully flexible, which allowed unassigned density in the RNAP and NusG regions to be filled according to the multi-body refined cryo-EM map (Fig. S14) and CLMS information. The rigid body boundaries and coarse-graining level are reported in Tables S4 and S5.

377 DSS and 144 DSSO-derived crosslinks involving the NusG-bound RNAP, NusA, the RNAP δ subunit, and the ribosomal subunits crosslinking to these proteins were used. The crosslinks were represented as distance restraints using the Bayesian CrossLinkingMassSpectrometryRestraint function in IMP.pmi, with both psi (crosslink nuisance) and sigma (positional uncertainty) to be sampled. The inflection point in the scoring function was set to 25.0 Å C α -C α distance for both DSS and DSSO. The cryo-EM density from multi-body refinement comprising the RNAP and NusA regions of the complex was approximated as a gaussian mixture model (GMM), using 700 gaussians (cross-correlation 0.885) in IMP, and encoded as a Bayesian restraint using the IMP.BayesianEM function (78). The theoretical densities for the rigid bodies and flexible regions used in the modeling were also represented as GMMs, with numbers of residues per gaussians reported in Table S5. Upon analysis of the cryo-EM density, fitting was not imposed for one RNAP α -CTD, as well as the coarse grained residues 366-540 of NusA. In addition, excluded volume restraints, at a resolution of 10 residues, and connectivity restraints were added to the scoring function. The weight of connectivity and excluded volume restraints was set to 2. To compensate for the different number of crosslinks in each dataset, the weight of DSS links was set to 5 and the weight of DSSO links to 10.

Sampling was performed by Replica Exchange Gibbs sampling in IMP 2.12, using 24 replicas in a temperature range between 1.0 and 20 in 20 independent runs with randomised initial configurations. A model was saved every 10 sampling steps, with each sampling step allowing for a 0.1 radians rotation and a 6 Å translation of each bead and rigid body. The number of frames was set to 30,000. During the calculations, the RNAP core, comprising the non-coarse-grained regions of RNAP α NTDs, β (excluding flat-tip helix), β' and NusG, as well as ribosomal proteins (Table S4) were kept fixed, as they could be unambiguously fit by prior rigid body fitting into the density.

Integrative structure modeling: scoring and analysis

The scoring was performed as reported previously (79, 80). A subset of 20,765 models of the sampled 14,400,000 configurations generated in replica exchange sampling were selected based on loose cutoffs in crosslink satisfaction (minimum 90% crosslinks < 35 Å distance), but stringent cutoffs in BayesianEM score and excluded volume score, as described in Table S6. In order to assess sampling exhaustiveness, the solutions were split into two samples, which were then compared to each other (Fig. S21B). The sampling precision was determined by r.m.s.d.-based clustering of NusA in the models at different thresholds (Fig. S21C), leading to an overall sampling precision estimate of 21.8 Å. All precision and r.m.s.d. calculations are carried out on the parts of the system that are sampled throughout the calculation - i.e. excluding the fixed RNAP and ribosome regions (see Table S7). Here, we defined sampling precision as in (80), i.e. as the r.m.s.d. threshold at which representative models are no longer different from each other in a statistically significant manner ($p > 0.05$), or with a small effect size ($V < 0.1$) and more than 80% of models fall in clusters in the subsequent r.m.s.d.-based clustering procedures. The solutions were clustered based on $C\alpha$ root-mean-squared-deviation using a 21.8 Å cutoff, derived from the overall sampling precision. The precision of each cluster and therefore the model precision was defined as the average distance to the solution closest to the cluster center, i.e. to the centroid model. The top-populated cluster yields a model precision of 11.54 Å and the cluster comprises 97% of selected models (Fig. S21D). At the clustering threshold derived from sampling precision, only one cluster of solutions is found. The centroid model, with the associated localization probability densities, is taken as the representative model for the integrative modeling solutions. Localization probability densities were defined as the probability of any voxel (here, $5 \times 5 \times 5 \text{ \AA}^3$) being occupied by a specific region in model densities over the selected cluster, each of which is obtained by convolving superposed models with a Gaussian kernel (here, with a standard deviation of 20.0 Å). In order to assess model robustness, 5 additional sets of runs were performed with 20% each of DSS and DSSO crosslinks discarded. Each set comprised 2 independent runs performed with the same parameters as the main integrative modeling procedure. Models were selected and analyzed as in the main runs. The resulting localization probability densities are displayed in Fig. S21I. Models were visualised using PyMOL, UCSF Chimera (64) and UCSF ChimeraX (65). Crosslinks were visualized with XlinkAnalyzer (81).

Supplementary Text

Crosslinking-derived protein interaction network

53% of the PPIs identified in the in-cell CLMS have previous evidence in *M. pneumoniae* or homologues, as covered by the STRING database or reported by AP-MS in *M. pneumoniae* (15) (Fig. S1C). We identified interactions between the subunits of well-annotated complexes including the ribosome, ATP synthase, DNA gyrase and the condensin complex (see Table S1 and Fig. S2). For the ribosome, we found several interactions with known associated proteins, including initiation factors IF-1 and IF-3, elongation factor G, and trigger factor. Moreover, ribosomal protein S10 (NusE) was crosslinked to the transcription antitermination factor NusB, an interaction previously reported in *E. coli* (82) (Fig. S2A).

The detection of crosslinks is biased towards the abundant proteome due to the relative low abundance of crosslinked peptides compared to their unmodified counterparts (13). Accordingly,

13

* This manuscript has been accepted for publication in Science. This version has not undergone final editing. Please refer to the complete version of record at <http://www.sciencemag.org/>. The manuscript may not be reproduced or used in any manner that does not fall within the fair use provisions of the Copyright Act without the prior, written permission of AAAS

interactions with highly abundant proteins, such as the chaperone proteins DnaK, Tuf and GroEL, account for 150 of the binary PPIs identified (26%) and appear as “hubs” in the network. Other proteins that have more than 15 unique PPIs each are enolase, glyceraldehyde-3-phosphate dehydrogenase, the uncharacterized integral membrane protein Y376 (Mpn376), phosphotransacetylase, pyruvate dehydrogenase E1 component subunit alpha, the uncharacterized lipoprotein Y052 (Mpn052), and L-lactate dehydrogenase. The interactions in which these proteins participate account for another 155 PPIs (27%) identified and are all in the top order of magnitude of protein abundance (Table S1).

In-cell CLMS identified PPIs involving membrane proteins and other difficult-to-solubilise parts of the proteome. Lipidated proteins or proteins with transmembrane domains were involved in 41% of the newly identified interactions (interactions not annotated in the STRING database (83), see Fig. S1). These include interactions between the high molecular weight cytoadherence proteins Hmw1-Hmw2 (29 crosslinked residue pairs) and Hmw2-Hmw3 (26 crosslinked residue pairs), for which there are no structural models. *M. pneumoniae* possesses reduced biosynthetic capacity, as these bacteria rely on the uptake of most building blocks for metabolism. Accordingly, we detected ABC transporter complexes, which transport peptides (OppA, OppC, OppF) and phosphate (PstA, PstB, PstS, PhoU) (Fig. S2B). RNA turnover is important for controlling gene expression in bacteria (84). Many bacteria contain the exonucleolytic 5'-3' RNase J. In *B. subtilis* and other bacteria, this RNase consists of a heterotetramer of two paralogous enzymes, RNase J1 and J2. For *M. pneumoniae*, so far only one RNase J protein has been annotated. This protein interacts with the membrane-bound protein MPN621 which we identified as the missing paralog of RNase J (Fig. S2B). Moreover, the essential membrane-bound endoribonuclease RNase Y crosslinked to the essential membrane protein MPN262 (8 crosslinked residue pairs), which is specific for *M. pneumoniae* and its close relative *M. genitalium*, suggesting they form a complex (Fig. S2B).

In total, the crosslinking-derived protein interaction network contains 301 proteins, of which 76 are only poorly functionally annotated or not at all.

Aspects of the expressome model

In *E. coli*, NusA has an N-terminal domain, an S1 domain, two KH domains and two terminal acidic-repeat (AR) domains (Fig. S18). The AR domains are not present in *B. subtilis* NusA, which terminates after KH2. Similarly, *M. pneumoniae* NusA has a low-complexity region in place of AR2, and low sequence and charge conservation for AR1, which indicates that the fold and function of this region may not be conserved between *E. coli* and *M. pneumoniae* (Fig. S18A). In this study, we chose to employ homology models for the *M. pneumoniae* NusA NTD and SKK domains, and to fully coarse-grain regions after the second KH domain (Fig. S18, S23 and Table S4). Moreover, due to their predicted disorder, these regions were not fitted to the cryo-EM density, but rather left fully flexible.

In the cryo-EM map of the active expressome from untreated cells, a density corresponding to a small domain connected by a long stretch of density connecting back to the RNAP α subunit N-terminal domain was observed (Fig. S18E). We suggest that this density does not belong to the C-terminal region of NusA due to its predicted disordered nature and the lack of crosslinks between the C-terminal region of NusA and RNAP. Thus, one of the α -CTDs was also represented as gaussians, and used to fit the cryo-EM density in the IMP protocol. In the cluster center solution

selected by the protocol, this α -CTD was positioned above the second KH domain of NusA and occupies the density connecting NusA KH2 and the α subunit core (Fig. S23).

While NusA interacts with nascent mRNA via its SKK domains, the interaction between NusA and RNAP has been proposed to occur via multiple mechanisms. The structure of the paused elongation complex suggests binding between one α -CTD and AR2, consistent with previous studies (85, 86), and an interaction between NusA NTD and the second α -CTD, besides the direct interaction between NusA NTD and the β flat-tip helix in the antitermination complex (23, 87). The resolution in this study was not sufficient to unambiguously position the flat-tip helix. However, its position, as indicated by CLMS, is consistent with its proposed interaction with NusA. Similarly, density for one α -CTD was observed in proximity of KH2, while the second α -CTD could not be localized to a high precision. However, no density was observed for α -CTD binding to the NusA NTD as in the *E. coli* paused elongation structure. Nevertheless, the overall position of the NusA NTD relative to RNAP closely resembles that of the paused elongation complex (Fig. S22). The model is also consistent with interactions between mRNA and the SKK domains, which may guide the mRNA into the entry site on the 30S ribosome (Fig. S23E).

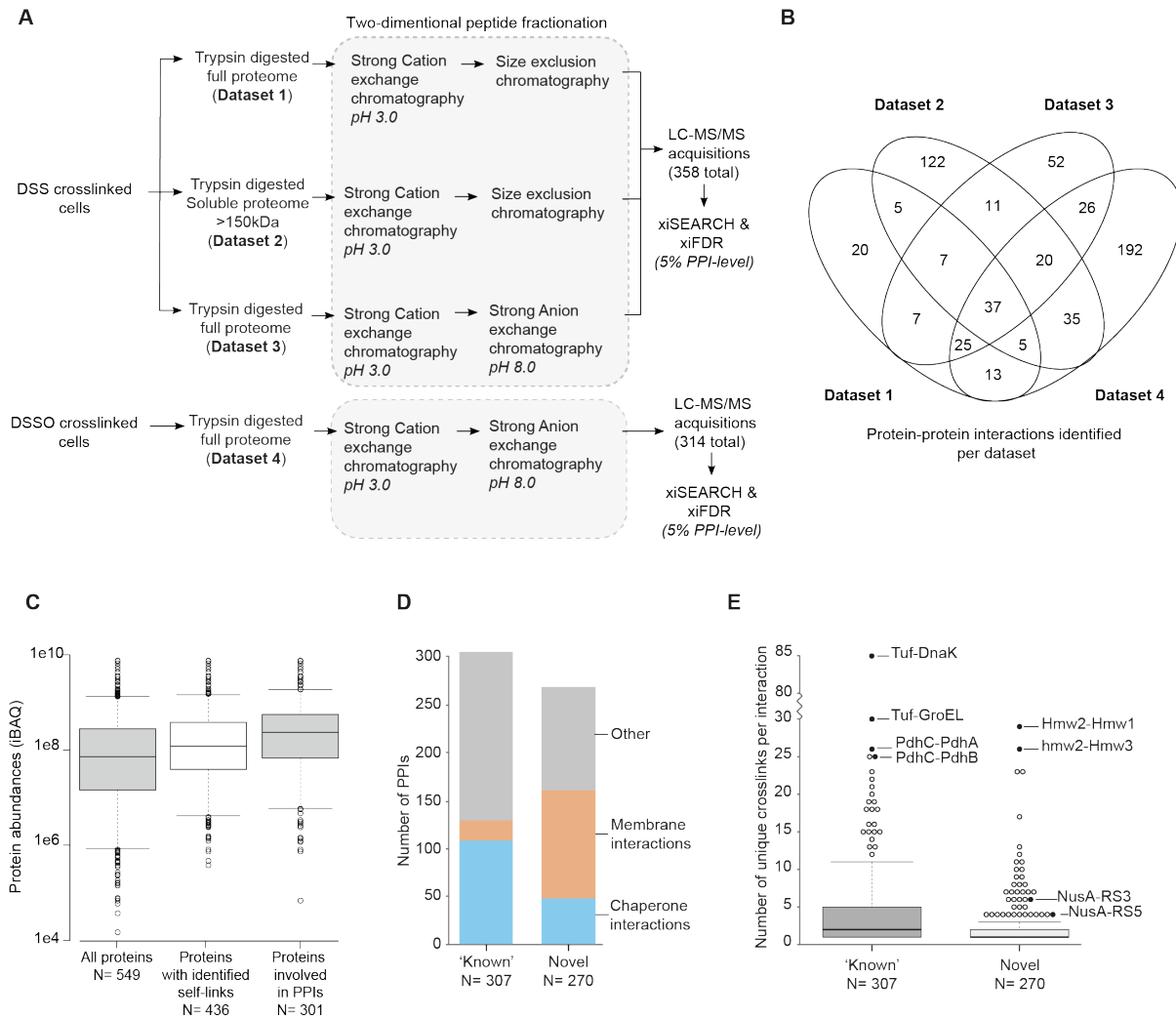


Fig. S1.

Whole cell CLMS dataset. (A) Cells were crosslinked with either DSS or DSSO and the crosslinked proteome was extracted and digested using Trypsin. For one dataset of DSS crosslinked cells, only the high molecular weight part of the soluble proteome (>150 kDa) was used (Dataset 2). Crosslinked peptides were sequentially enriched and fractionated in two dimensions by either strong cation exchange chromatography followed by size exclusion chromatography, or strong cation exchange chromatography followed by strong anion exchange chromatography. (B) Venn diagram showing overlap of PPIs identified in each dataset. (C) 549 proteins were detected from *M. pneumoniae* and 83% of these were identified with CLMS. As crosslinked peptides are generally low-abundance in comparison to linear peptides, they tend to be detected in proteins present at higher copy numbers. The median intensity of proteins identified by shotgun proteomics is $7.4e^7$, but for proteins with identified intra-protein or inter-protein crosslinks it is $1.2e^8$ and $2.4e^8$ respectively. (D) More 'known' interactions than unknown interactions (according to STRING-DB v11 (83)) are reported. Of the 'novel' interactions, 59% involve either membrane-associated proteins or chaperones. (E) Known interactions tend to have

16

* This manuscript has been accepted for publication in Science. This version has not undergone final editing. Please refer to the complete version of record at <http://www.sciencemag.org/>. The manuscript may not be reproduced or used in any manner that does not fall within the fair use provisions of the Copyright Act without the prior, written permission of AAAS

higher numbers of unique crosslinked residue pairs, suggesting higher abundance of these interactions/protein assemblies in the cell. Interactions between the high molecular weight cytoadherence proteins (hmw2-hmw1; hmw2-hmw3) or NusA with 30S ribosomal (RS3 and RS5) proteins are not reported in STRING.

Protein complexes and their interactors detected in the dataset. (A) Networks showing all intra- and inter-protein crosslinks identified on known complexes. Any proteins found crosslinked to these are also shown. Circle diameter indicates relative protein size. Each edge represents one or more crosslinks. (B) Crosslink-based protein networks of previously uncharacterised interactions involving membrane-bound subunits. Brown lines on the sequence indicate transmembrane regions. The endoribonuclease RNase Y (Rny) does not have a previously annotated interaction partner.

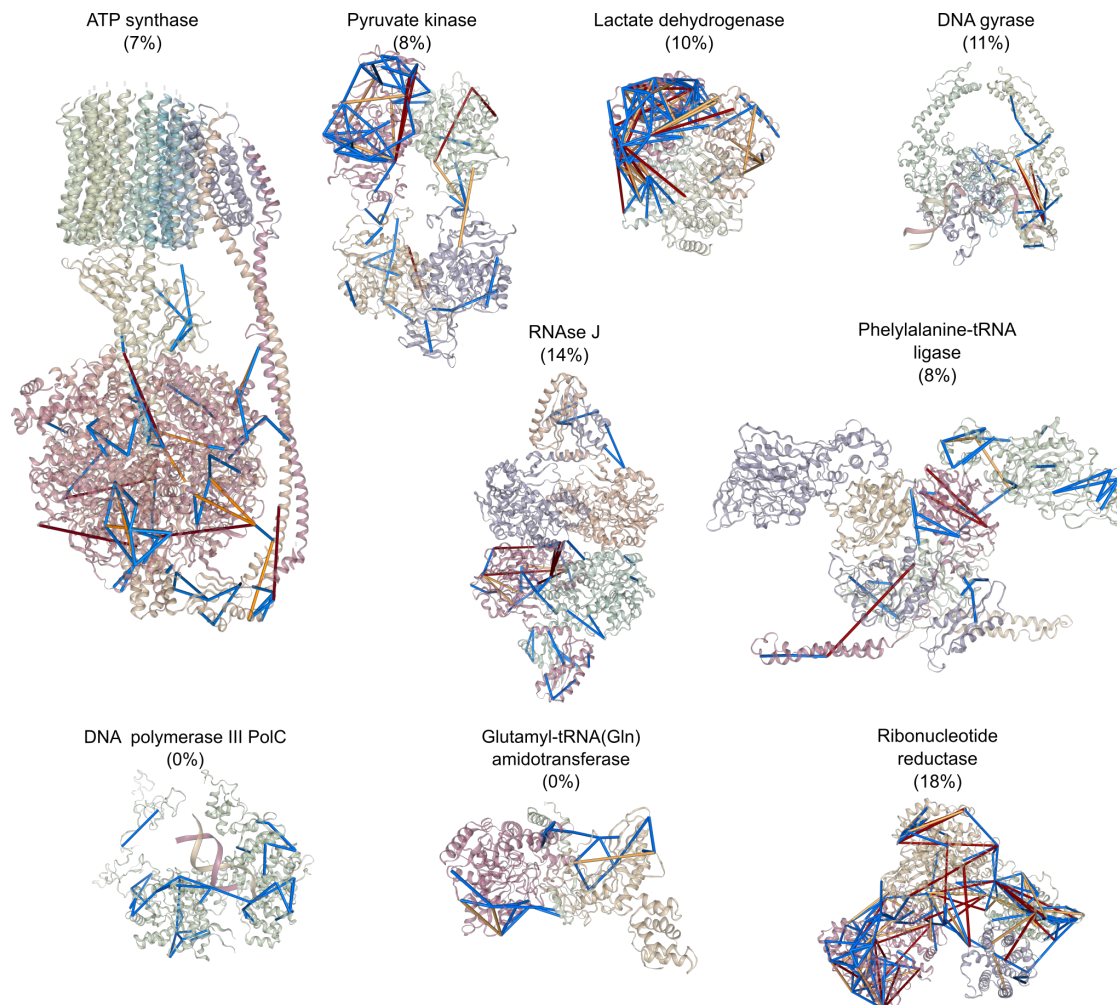


Fig. S3.

Mapping of crosslinks on homology models of known complexes in *M. pneumoniae*.

Crosslinks mapped on homology models of detected protein complexes generated with Swiss-Model. Satisfied links (<28 Å) are displayed in blue, links near the violation threshold (28-35 Å) in yellow and violated links (>35 Å) in red. The percentage of violated links is shown in parentheses.

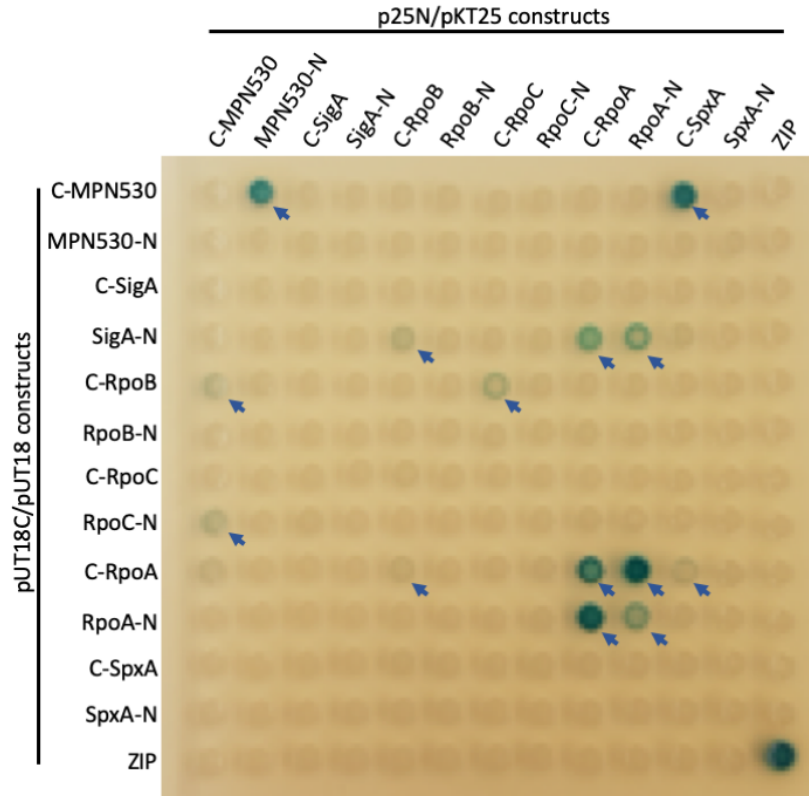


Fig. S4.

Bacterial two hybrid screen confirmed interaction of MPN530 and RNAP. To confirm novel interactions identified by in-cell CLMS between MPN530 and the subunits β and β' of the RNAP, the binary interactions were studied by using a bacterial two-hybrid screen. In this system, interacting proteins reconstitute the *B. pertussis* adenylate cyclase, resulting in cAMP synthesis and subsequent activation of β -galactosidase synthesis. Colonies in which interactions were observed (β -galactosidase was expressed) are marked by blue arrows. MPN530 and SpxA showed a strong interaction. Moreover, MPN530 and RNAP α (RpoA) exhibited strong self-interaction. We detected a weak, but significant, interaction between MPN530 and β (RpoB) as well as β' (RpoC) (left column). We also confirmed the interaction between SigA with α (RpoA) and β (RpoB) and the interaction of SpxA with α (18). As expected, all tested subunits of the RNA polymerase interact.

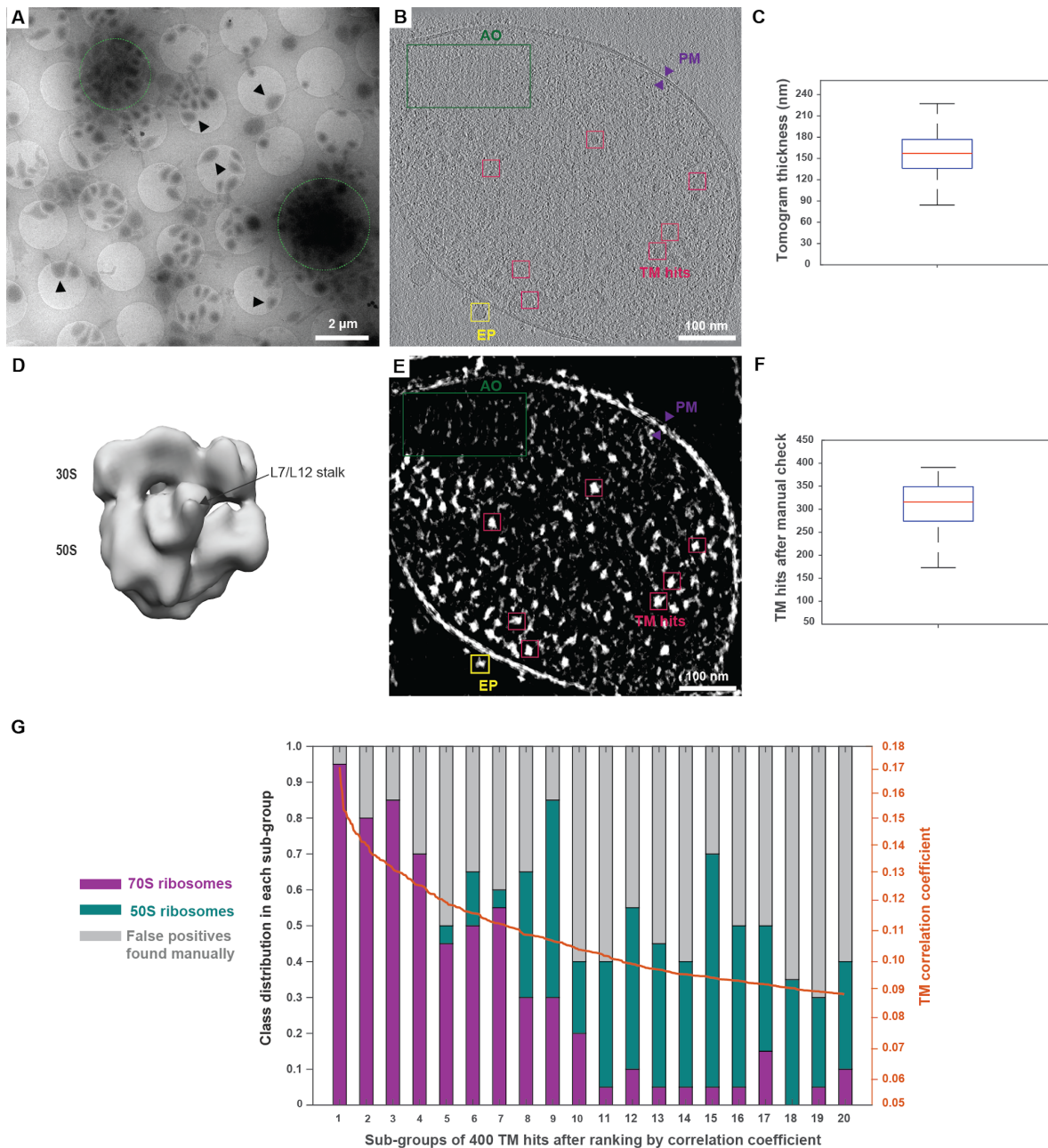


Fig. S5.

Cryo-EM of *M. pneumoniae* and ribosome localization. (A) Individual *M. pneumoniae* M129 cells (arrowheads) and cell colonies (green dashed circles) were grown on cryo-EM grids. cryo-ET data were collected on individual cells. (B) XY slice of a conventional defocus tomogram. Annotation of representative ribosomes localized by template matching (TM, red boxes), the attachment organelle (AO, green box), the plasma membrane (PM, purple arrowheads) and unknown extracellular particles (EP, yellow boxes). (C) Distribution of cell thickness estimation

from tomographic reconstructions. The thickness ranges from 84 nm to 284 nm (black bars), with a median value of 158 nm (red line). Blue box represents the first and third quartile. **(D)** *De novo* ribosome sub-tomogram average reconstructed from 404 manually picked ribosomes from VPP tomograms. Density was low-pass filtered to 30 Å and used as reference for template matching. **(E)** Cross correlation coefficients of the ribosome reference with the tomographic volume in **(B)**. Brighter pixels indicate higher cross correlation values. Obvious false positives such as membranes, carbon edges and extracellular particles were manually excluded. **(F)** Number of ribosomes picked per tomogram after template matching and visual validation. The median number is 316 (red line), while most tomograms contained between 274 to 349 ribosomes (blue box, first to third quartile). **(G)** The top 400 template matching scores from a representative tomogram were ranked by correlation coefficients (yellow curve) and divided into 20 sub-groups. After visual inspection and 3D classification of the sub-tomograms, percentages of obvious false positives (grey), 50S (green) and 70S (magenta) ribosomes within each sub-group were calculated and plotted, showing that a cutoff of 400 peaks per tomogram retained most ribosomes in the data.

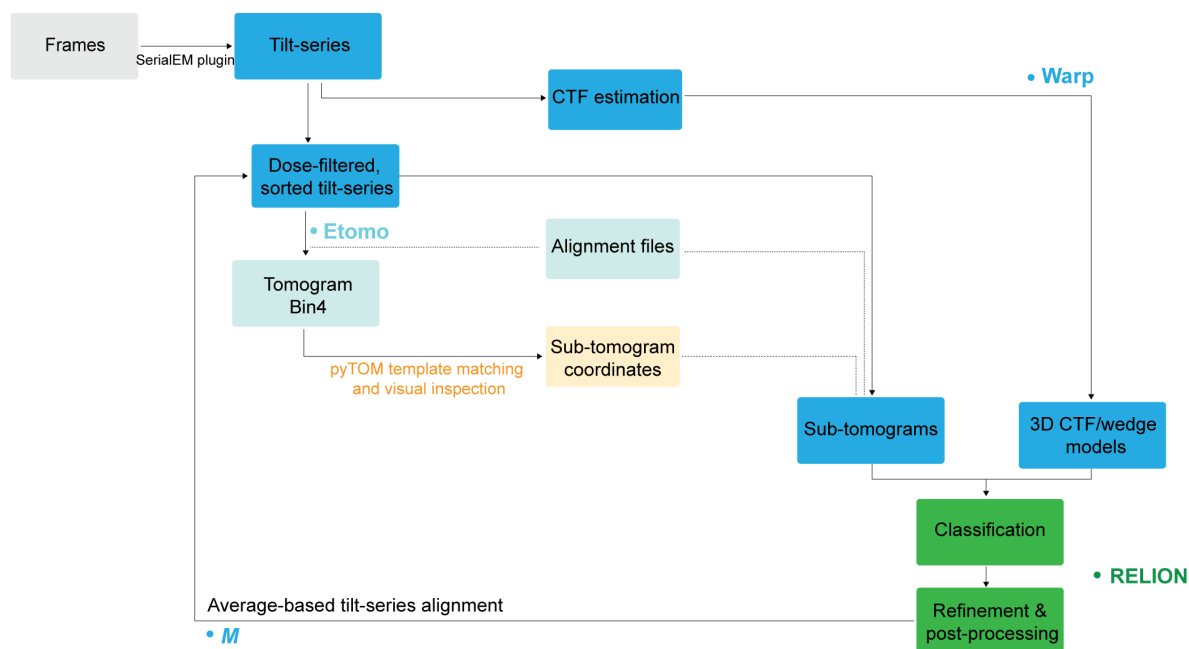


Fig. S6.

Sub-tomogram analysis workflow. Frames were aligned on-the-fly with the IMOD alignframes plugin in SerialEM. Dose-symmetric tilt-series were pre-processed in Warp for CTF estimation, sorting of the tilt images and dose filtering. Tilt-series were aligned in Etomo with 4x binning using gold fiducials. Template matching was performed in pyTOM, followed by visual inspection using MATLAB/TOM toolbox. Tilt-series, alignment files, CTF files and particle coordinates were utilized by Warp to reconstruct sub-tomograms, as well as per sub-tomogram 3D CTF/wedge models containing both CTF, dose filtering and missing wedge information. Sub-tomograms were processed in RELION 3.0, including 3D classification, 3D refinement, multi-body refinement and post-processing. RELION refinement and Warp processing results were used in *M* for 3D average-based (i.e. ribosomes) alignment for each tilt-series. Sub-tomograms and 3D CTF/wedge models were re-generated after *M*, with the refined alignment parameters for further analysis.

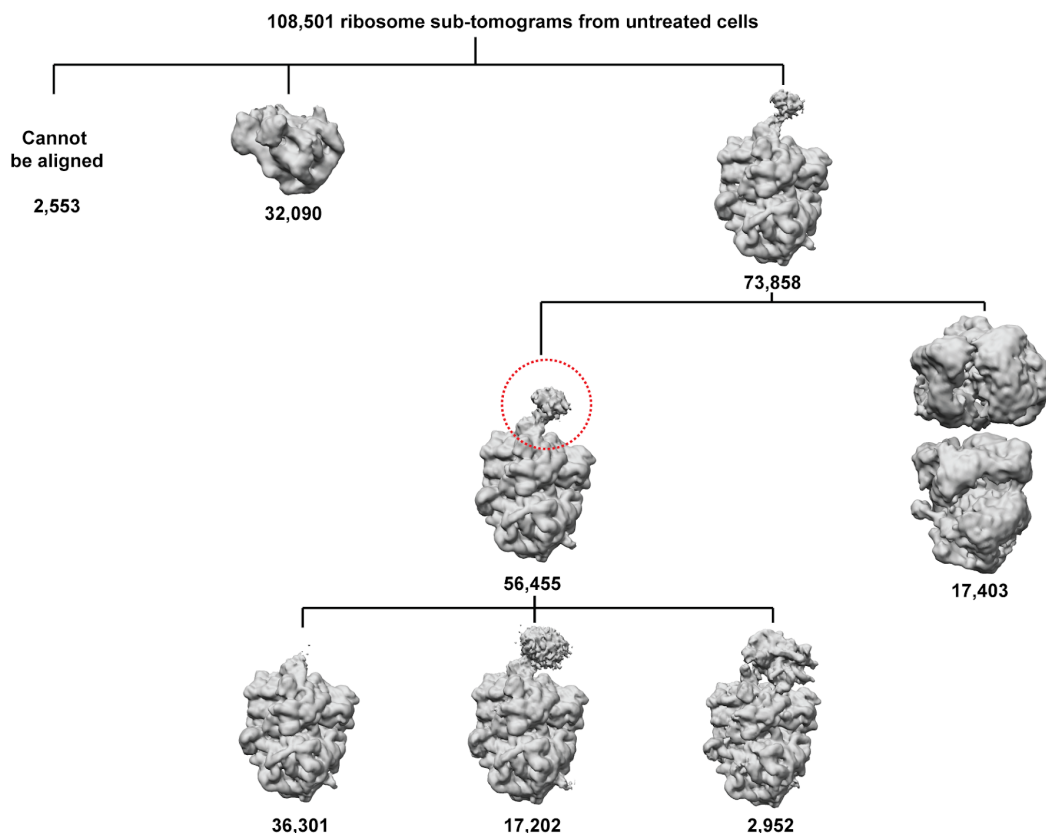


Fig. S7.

Maximum likelihood-based classification of ribosome sub-tomograms from untreated *M. pneumoniae* cells. In total, 108,501 ribosome sub-tomograms were reconstructed and classified in RELION. Except for 2,553 sub-tomograms that could not be aligned, 2 distinct classes of ribosomes were identified: 50S ribosomes (32,090) and 70S ribosomes (73,858). Prior to subsequent focused classification near the mRNA entry site (red dashed circle), 70S ribosomes in closely assembled polysomes (17,403) were sorted out (24). The rationale to exclude closely-assembled ribosomes is that neighboring ribosomes could interfere with further focused classification and are unlikely to interact with RNAP. 2,952 sub-tomograms exhibit a well-defined RNAP density near the mRNA entry site, while 36,301 ribosomes do not have significant additional density. For the class of 17,202 ribosomes, the additional density near the mRNA entry site is significant and of similar size to RNAP. Multi-body refinement of the this class did not produce better resolved structures, nor did further classification runs with different settings (local masks, class number, angular search range, etc.), suggesting that these sub-tomograms represent supercomplexes with high compositional heterogeneity and large conformational flexibility formed during transcription-translation coupling.

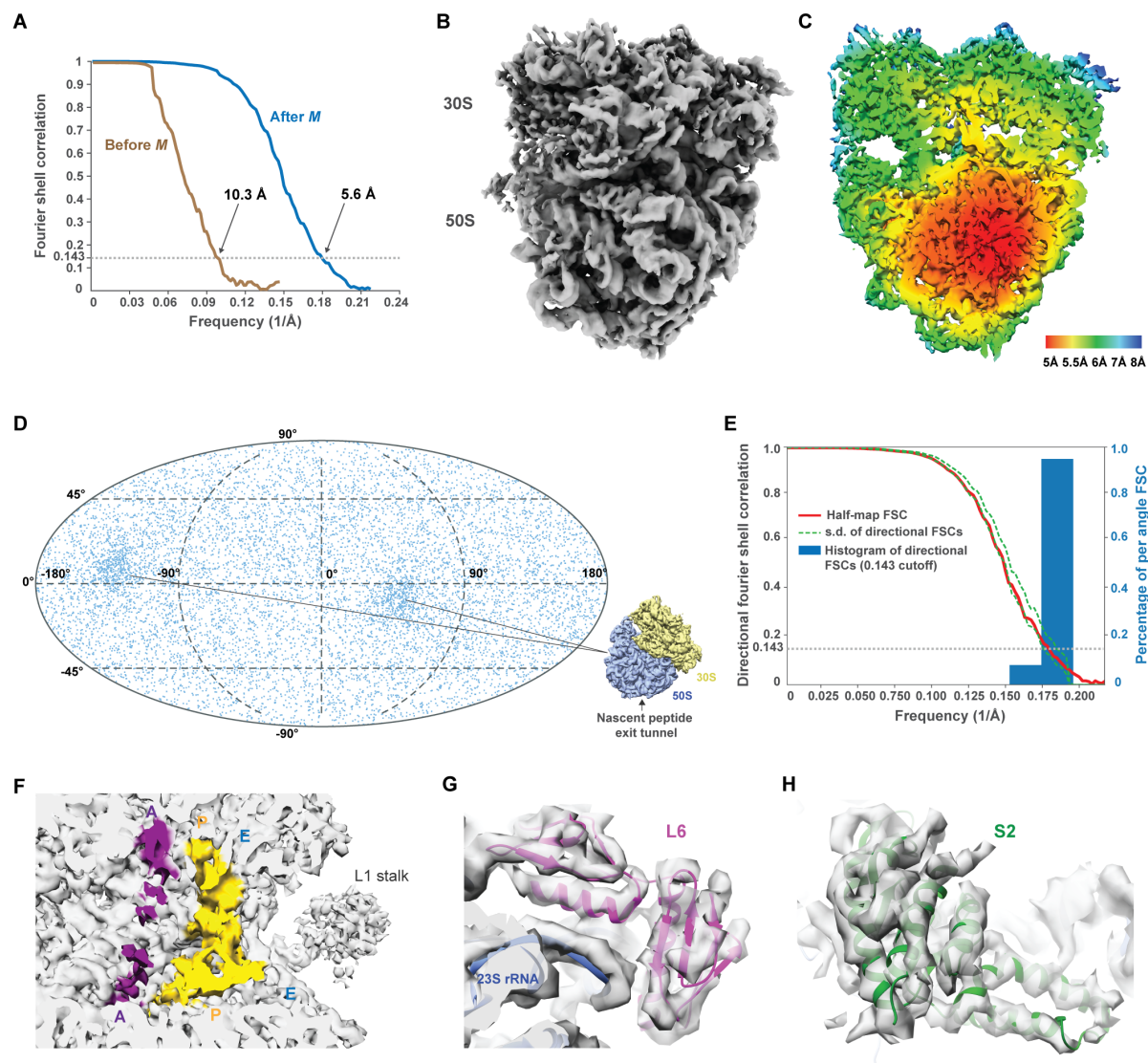


Fig. S8.

5.6 Å in-cell *M. pneumoniae* ribosome structure obtained by sub-tomogram averaging. (A) Refinement of all 70S ribosome sub-tomograms generated a map of 5.6 Å resolution (FSC=0.143). Refinement of the tilt-series in *M* resulted in a significant resolution improvement. (B) Surface representation of the 5.6 Å in-cell ribosome density. (C) Central slice of the local resolution map. (D) Angular distribution plot shows that the dataset covered all particle orientations, but exhibited weak preferential angular distribution. As indicated by the ribosome structure (bottom right), these two peaks represent ribosomes oriented parallel to nascent peptide exit tunnel in the data. (E) 3D FSC analysis. Calculated sphericity of 0.985 indicated minimal directional resolution anisotropy of the density. (F) Densities corresponding to tRNAs in A and P site. No significant density for E site tRNA was detected. (G, H) Homolog structures (based on PDB 3J9W) of two ribosomal proteins, L6 (G) and S2 (H), docked into the 70S density.

* This manuscript has been accepted for publication in Science. This version has not undergone final editing. Please refer to the complete version of record at <http://www.sciencemag.org/>. The manuscript may not be reproduced or used in any manner that does not fall within the fair use provisions of the Copyright Act without the prior, written permission of AAAS

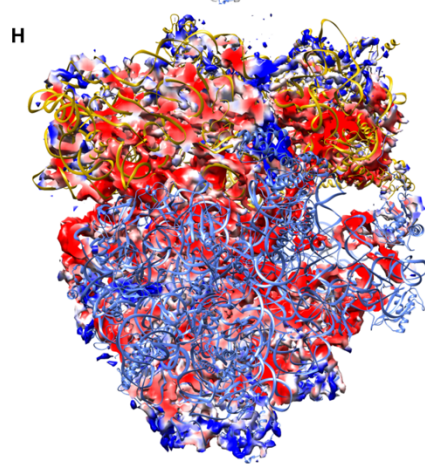
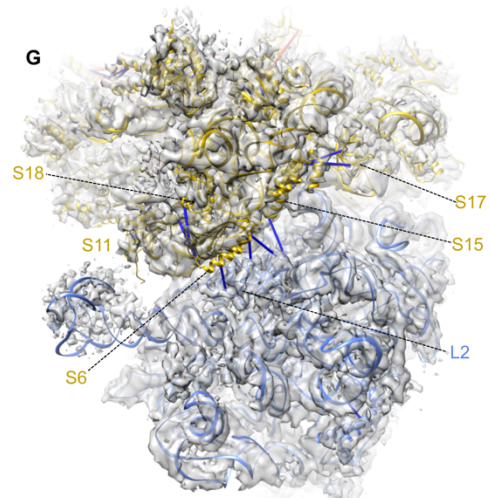
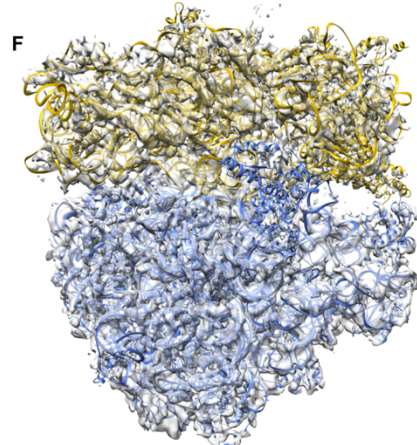
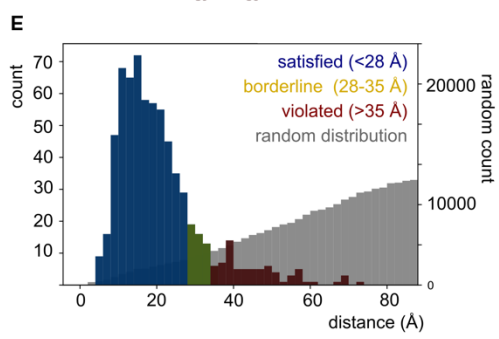
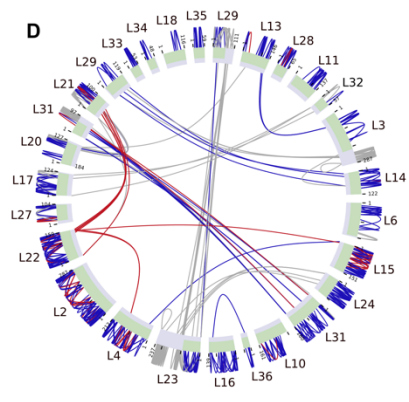
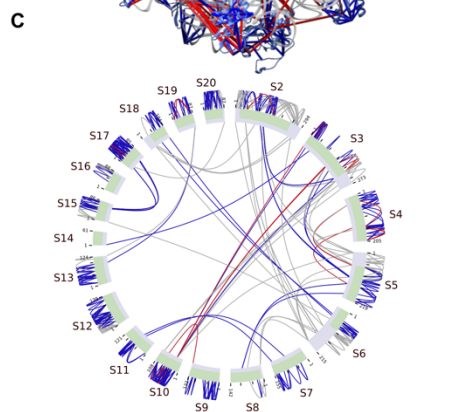
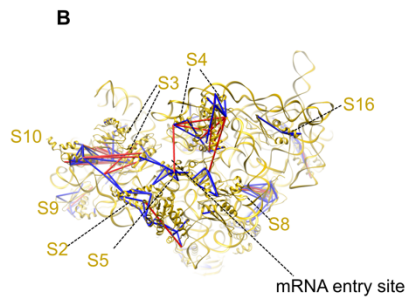
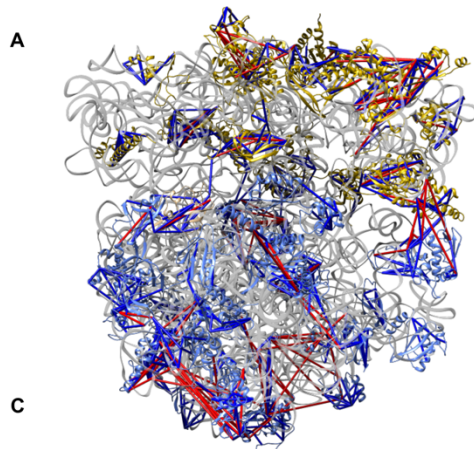


Fig. S9.

The *M. pneumoniae* ribosome homology model. (A) Crosslinks mapped on homology models of the ribosome based on the *B. subtilis* Mfim-stalled ribosome (PDB 3J9W) generated using Swiss-Model. Subunits not found in the density (L1, L7/L12, L9, S21) and fully flexible regions have been removed. The rRNA from *B. subtilis* has been retained, as the resolution was not sufficient to build *M. pneumoniae* rRNA into the map. The largest violations were observed in the peptide exit tunnel region (proteins L21, L22). Satisfied links (<35 Å) are displayed in blue, and violated links (>35 Å) in red. (B) View of the top of the 30S subunit in the homology model with crosslinks labelled by distance as in (A). (C) Circle representation of the crosslinks on the 30S subunit labelled by distance as in (A). (D) Circle representation of the crosslinks on the 50S subunit labelled by distance as in (A). (E) Distribution of 683 crosslink distances obtained from the ribosome model compared to a theoretical random distribution of lysine-lysine and lysine-N-termini crosslinks from the same model. (F) Fitting of the *M. pneumoniae* ribosome homology model to the 5.6 Å in-cell ribosome density. (G) View of the 30S-50S interface with inter-subunit crosslinks and density displayed. (H) Cross-section of the in-cell ribosome density colored by local cross-correlation (CC) to a theoretical 6 Å density derived from the ribosome homology model. Blue: $CC < 0.7$; White: $0.7 < CC < 0.9$; Red: $CC > 0.9$.

between L22 and L21, L4, as well as L15 in residues ~141 to 159 are consistent with this region being highly flexible. Crosslinks shorter than 35 Å are shown in blue, while those longer than 35 Å are in red.

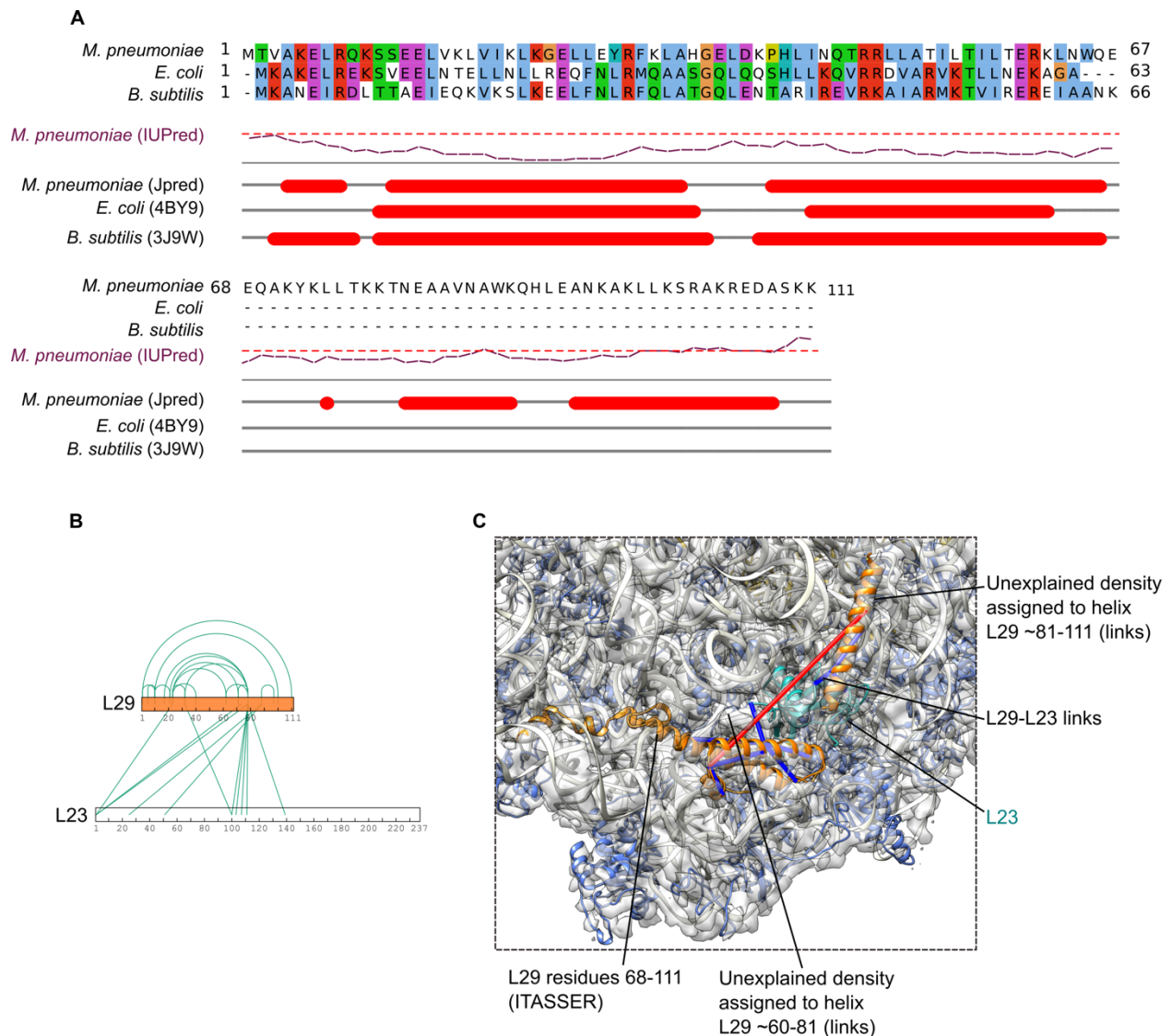


Fig. S11.

The C-terminal region of ribosomal protein L29 in *M. pneumoniae*. (A) Sequence alignment of *M. pneumoniae*, *E. coli*, and *B. subtilis* L29 proteins. For *E. coli* and *B. subtilis*, the secondary structure as observed in PDBs 4V7D and 3J9W is displayed. For *M. pneumoniae*, the secondary structure was predicted with Jpred. β -sheets are shown as arrows, α -helices as cylinders. The long-range disorder prediction of IUPred is also shown, with the threshold of 50% disorder likelihood displayed as a dashed line. *M. pneumoniae* presents a unique C-terminal extension from residue 68, not found in other sequences. (B) The *in-cell* crosslinking network of L29, confirming its placement near L23. (C) Placement of the L29 (orange) C-terminal region according to the cryo-EM density and crosslink data. The original prediction for the L29 C-terminal region, obtained with I-TASSER (shown in transparent orange), lay outside the EM density map and did not agree with crosslinking information. Using crosslinks, we could assign the unexplained density to the C-

terminal region of L29, although the resolution was not high enough to determine helical orientation and boundaries. The density, secondary structure prediction and crosslinks to L23 are consistent with a loop connecting the N-terminal region of L29 to a long C-terminal α -helix up to residue ~111. Crosslinks shorter than 35 Å are shown in blue, while those longer than 35 Å are in red.

L29 C-terminal helix found in *M. pneumoniae* (see Fig. S11). Crosslinks shorter than 35 Å are shown in blue, while those longer than 35 Å are in red. L24 residues cross-linked to the disordered region of L23 are shown as red spheres.

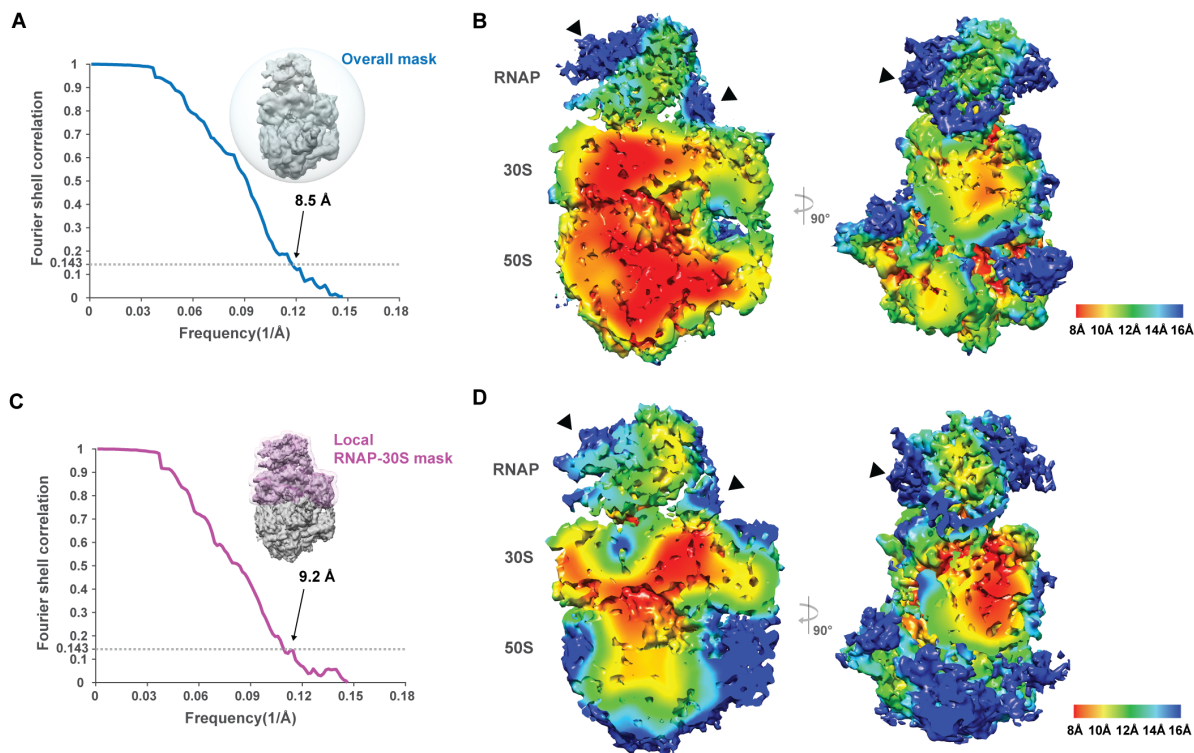


Fig. S13.

In-cell RNAP-ribosome supercomplex determined at sub-nanometer resolution. 2,952 subtomograms from the class representing RNAP-ribosome supercomplexes in Fig. S7 were refined with different masks. **(A)** Refinement with a spherical mask (insert) produced a density with overall 8.5 Å resolution. **(B)** Local resolution map showed that the ribosome core has higher resolution, while the RNAP density is of relatively lower resolution, especially for the flexible clamp region and the density between RNAP and the 30S (arrowheads). **(C)** Focused refinement with a RNAP-30S mask (insert) generated a map of overall resolution 9.2 Å. **(D)** The 30S is of higher local resolution, while the 50S is blurred, likely due to ratcheting. Local resolution for RNAP improved compared to the refinement with a spherical mask as in (B).

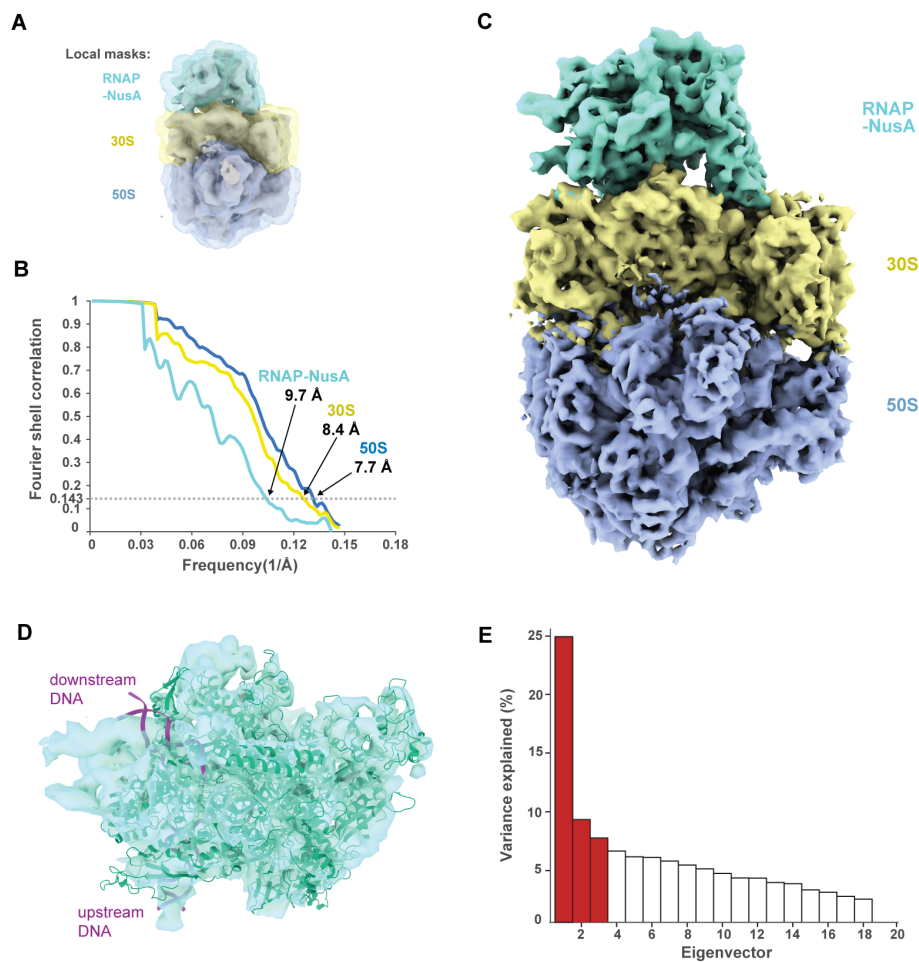


Fig. S14.

Multi-body refinement revealed high conformational flexibility of the in-cell expressome. To better resolve local density and to assess the flexibility of the 2,952 sub-tomograms used to obtain the density in Fig. S13, multi-body refinement implemented in RELION was employed. **(A)** The consensus refinement result was segmented into 3 bodies and correspondingly, three local masks were generated, i.e. RNAP-NusA, 30S, and 50S. **(B)** FCS curves for the 3 bodies were calculated separately. **(C)** Superposition of densities of the 3 bodies. **(D)** Top view of the RNAP density with homology model fitted. **(E)** Principal component analysis on the relative orientations of all bodies was performed to characterize the most predominant motions, represented by different eigenvectors. Relative movements corresponding to the 3 most dominant eigenvectors (red) are shown in Movie S1.

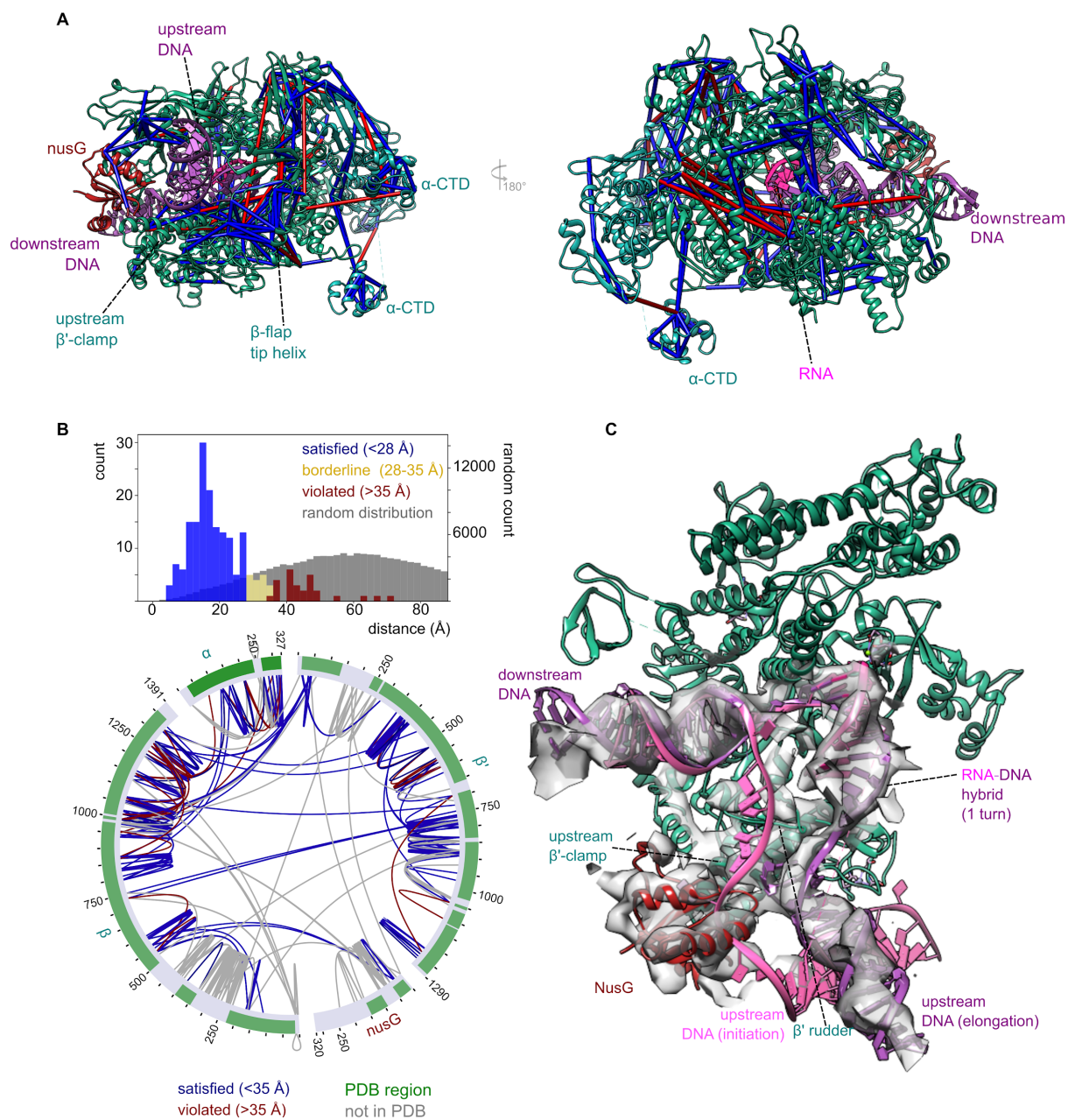


Fig. S15.

The *M. pneumoniae* RNAP homology model. (A) Crosslinks mapped on homology models of the core RNAP based on the *E. coli his*-paused elongation complex with nucleic acids included (PDB 6FLQ) generated using Swiss-Model. The regions which were kept fully flexible in the integrative modeling or are disordered in the homology model, have been removed. The α -CTDs and β flap-tip helix are structured, but flexible with respect to the core RNAP. NusG was placed

38

* This manuscript has been accepted for publication in Science. This version has not undergone final editing. Please refer to the complete version of record at <http://www.sciencemag.org/>. The manuscript may not be reproduced or used in any manner that does not fall within the fair use provisions of the Copyright Act without the prior, written permission of AAAS

according to the EM density in the same binding mode as in the *E. coli* RNAP elongation complex (PDB 6C6U, (74)). Satisfied links (<35 Å) are displayed in blue, and violated links (>35 Å) in red. **(B)** Top: distribution of 194 crosslink distances obtained from the RNAP model compared to a theoretical random distribution of lysine-lysine and lysine-Ntermini crosslinks from the same model. Bottom: circle view of the crosslink distances in the model. Regions missing in the model are shaded in gray. **(C)** The β' , DNA and RNA-DNA hybrid. The density in the RNAP region of the multi-body-refined expressome map corresponding to the DNA, RNA-DNA hybrid and upstream β' clamp are shown. For comparison, the DNA path in the initiation RNAP holoenzyme (PDB 5VT0, (88)) is shown in light pink. Moreover, the RNAP homology model showed a good fit to the EM density. No density was observed for single-stranded or folded mRNA near the exit tunnel.

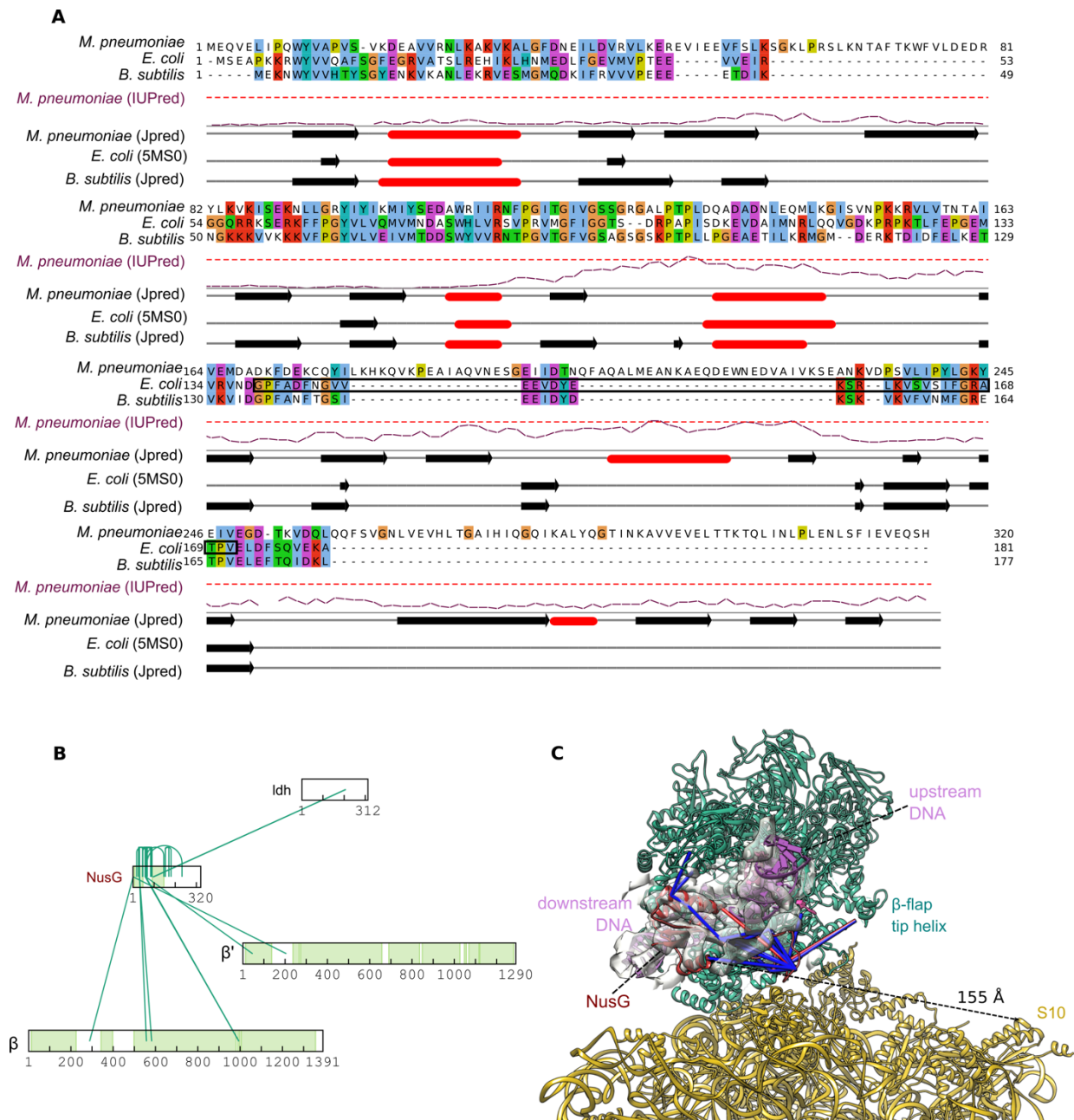


Fig. S16.

NusG in *M. pneumoniae*. (A) Sequence alignment of *M. pneumoniae*, *E. coli*, and *B. subtilis* NusG proteins. For *E. coli*, the secondary structure as observed in the λ phage antitermination complex (PDB 5MS0) is depicted. For *M. pneumoniae* and *B. subtilis*, the secondary structure was predicted with Jpred. β -sheets are shown as arrows, α -helices as cylinders. The long-range disorder prediction of IUPred is also shown, with the threshold of 50% disorder likelihood displayed as a dashed line. The residues involved in the *E. coli* NusG:S10 interaction, as reported in (8) are

40

* This manuscript has been accepted for publication in Science. This version has not undergone final editing. Please refer to the complete version of record at <http://www.sciencemag.org/>. The manuscript may not be reproduced or used in any manner that does not fall within the fair use provisions of the Copyright Act without the prior, written permission of AAAS

highlighted in the black box in the sequence. *M. pneumoniae* presents several inserts between the NTD (residues 1-150) and the putative S10 binding region. **(B)** Crosslink-derived interaction network of NusG. Structured regions of RNAP and NusG, not coarse grained during the modeling procedure, are shown in green. **(C)** NusG (in red) in the active expressome. Satisfied (<35 Å) crosslinks are shown in blue and violated crosslinks in red. The insertion loop in the NTD (residues 48-91) was positioned below the upstream DNA, in agreement with previous reports (74), and showed crosslinks to the β flat-tip helix. The distance between the end of the NTD (residue 146) and S10 on the ribosome is marked by a dashed line, which is too far to have direct interaction in our model

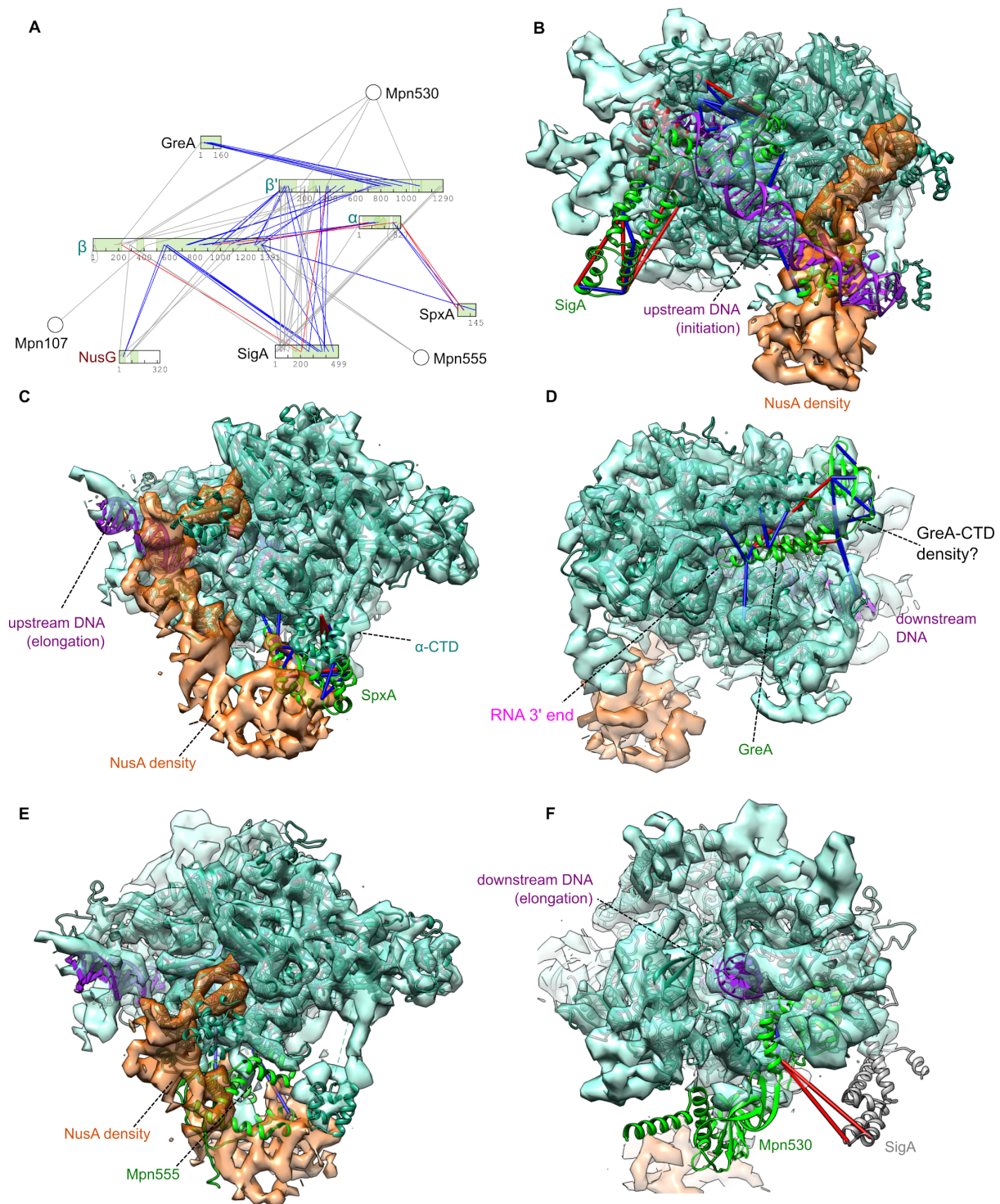


Fig. S17.

* This manuscript has been accepted for publication in Science. This version has not undergone final editing. Please refer to the complete version of record at <http://www.sciencemag.org/>. The manuscript may not be reproduced or used in any manner that does not fall within the fair use provisions of the Copyright Act without the prior, written permission of AAAS

Additional RNAP regulators found in the crosslink network are not constitutive components of the elongating expressome. (A) Interaction network of the core RNAP α (RpoA), β (RpoB), and β' (RpoC) subunits showing all binders of RNAP detected by in-cell crosslinking. The areas shaded in green are those present in the homology models used in this figure. Satisfied crosslinks (<35 Å) are shown in blue, while those longer than 35 Å are in red. (B) Homology model of the sigma factor SigA and DNA bound as in the RNAP holoenzyme (gray, template PDB 5VT0, (88)). The cryo-EM density assigned to RNAP is also displayed in green. The crosslinks between SigA and the core polymerase are consistent with sigma factor binding in the same mode as in the *E. coli* 70S holoenzyme, albeit with flexibility in its N-terminal region (of which residues 139-320 are present in the homology model). However, both the path of the nucleic acid and the density around the protein are not consistent with SigA being present in the expressome. (C) The regulatory protein SpxA, which is known to interact with the RNAP α -CTD, placed according to the *B. subtilis* model by Swiss-Model (template PDB 3GFK, (89)). It displays crosslinks to the α -CTD and the residues 1316-1362 of the β CTD. Its size, shape and position are not consistent with the expressome density. (D) Elongation factor GreA, responsible for resolving elongation arrest by stimulating the RNAP activity of RNAP. The binding is modelled on the structure of GreB-bound RNAP in the pre-cleavage state (PDB 6RIN, (90)). The crosslinks confirm the binding mode of this protein in *M. pneumoniae*. While some unassigned density could represent the GreA CTD, the lack of density around the NTD may indicate low occupancy, or a highly dynamic nature of the protein in the elongating expressome. Nevertheless, this density is filled by β -lobe 2 residues in the modeling results (Fig. S22A). (E) Tentative placement of the uncharacterised protein Mpn555 near the mRNA exit site according to CLMS data. The model of Mpn555 was generated using I-TASSER. (F) Tentative placement of the uncharacterised protein Mpn530 near the downstream DNA according to CLMS data. The model of Mpn530 was generated using I-TASSER.

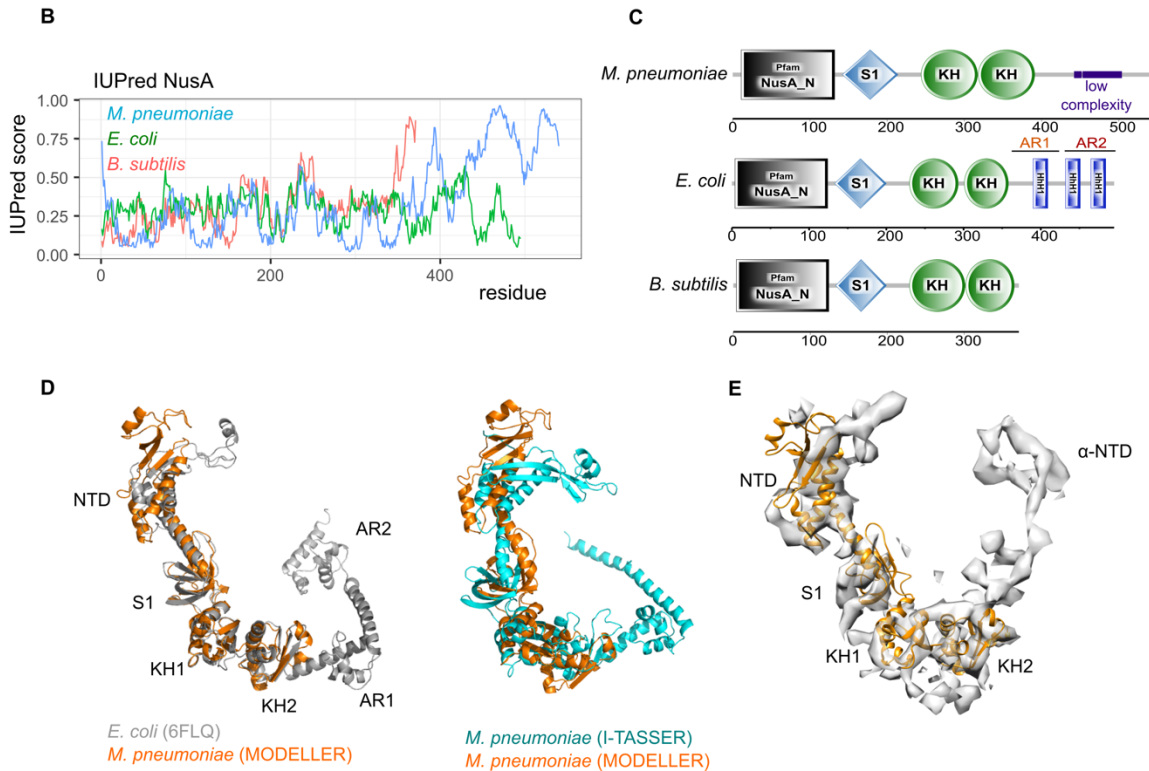
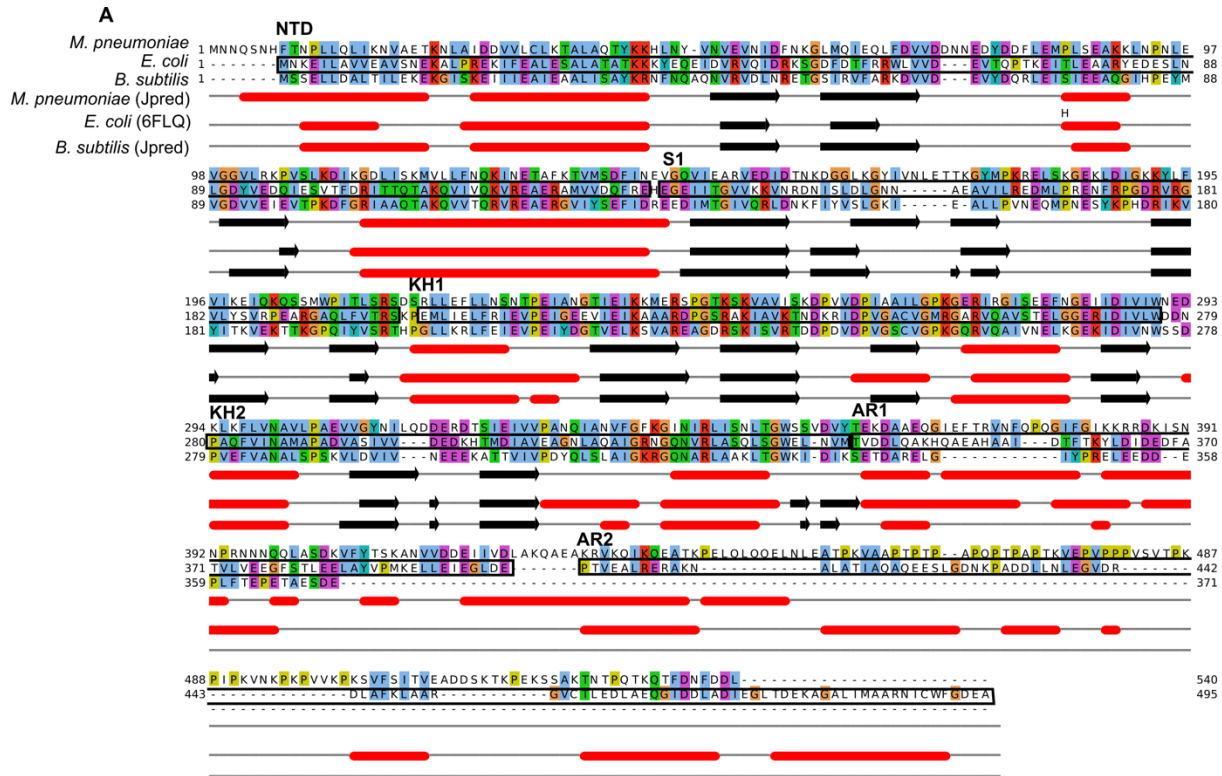


Fig. S18.

* This manuscript has been accepted for publication in Science. This version has not undergone final editing. Please refer to the complete version of record at <http://www.sciencemag.org/>. The manuscript may not be reproduced or used in any manner that does not fall within the fair use provisions of the Copyright Act without the prior, written permission of AAAS

NusA in *M. pneumoniae*. (A) Sequence alignment of *M. pneumoniae*, *E. coli*, and *B. subtilis* NusA proteins. For *E. coli*, the secondary structure as observed in the RNAP paused elongation complex was used (PDB 6FLQ). For *M. pneumoniae* and *B. subtilis*, the secondary structure was predicted with Jpred. β -sheets are shown as arrows, α -helices as cylinders. The *E. coli* NusA domain boundaries are shown in black boxes in the sequence. *M. pneumoniae* presents a highly divergent C-terminal extension, with no predicted AR2, and low sequence conservation in AR1. (B) Prediction of disordered regions for NusA. (C) The SMART database domain annotation for NusA. (D) Left: the *E. coli* NusA structure (gray; PDB 6FLQ) overlaid with the homology model of residues 1-364 *M. pneumoniae* NusA generated with MODELLER, used in this study (orange; rigid body boundaries in Table S4). Right: homology model of NusA compared with structure prediction using I-TASSER. I-TASSER generates a structure for residues 361-457 that resembles the *E. coli* AR1 domain followed by a long extension. No fold is predicted for residues 457-540 (not displayed). (E) Rigid body fitting of the homology model of NusA NTD and SKK domains to the region of cryo-EM density unaccounted by RNAP. The size and shape of the density is consistent with NusA. The predicted fold generated by I-TASSER could not be fitted into the density above the second KH domain. Therefore, the NusA C-terminal region was not fitted to the EM density in the integrative modeling (Table S4).

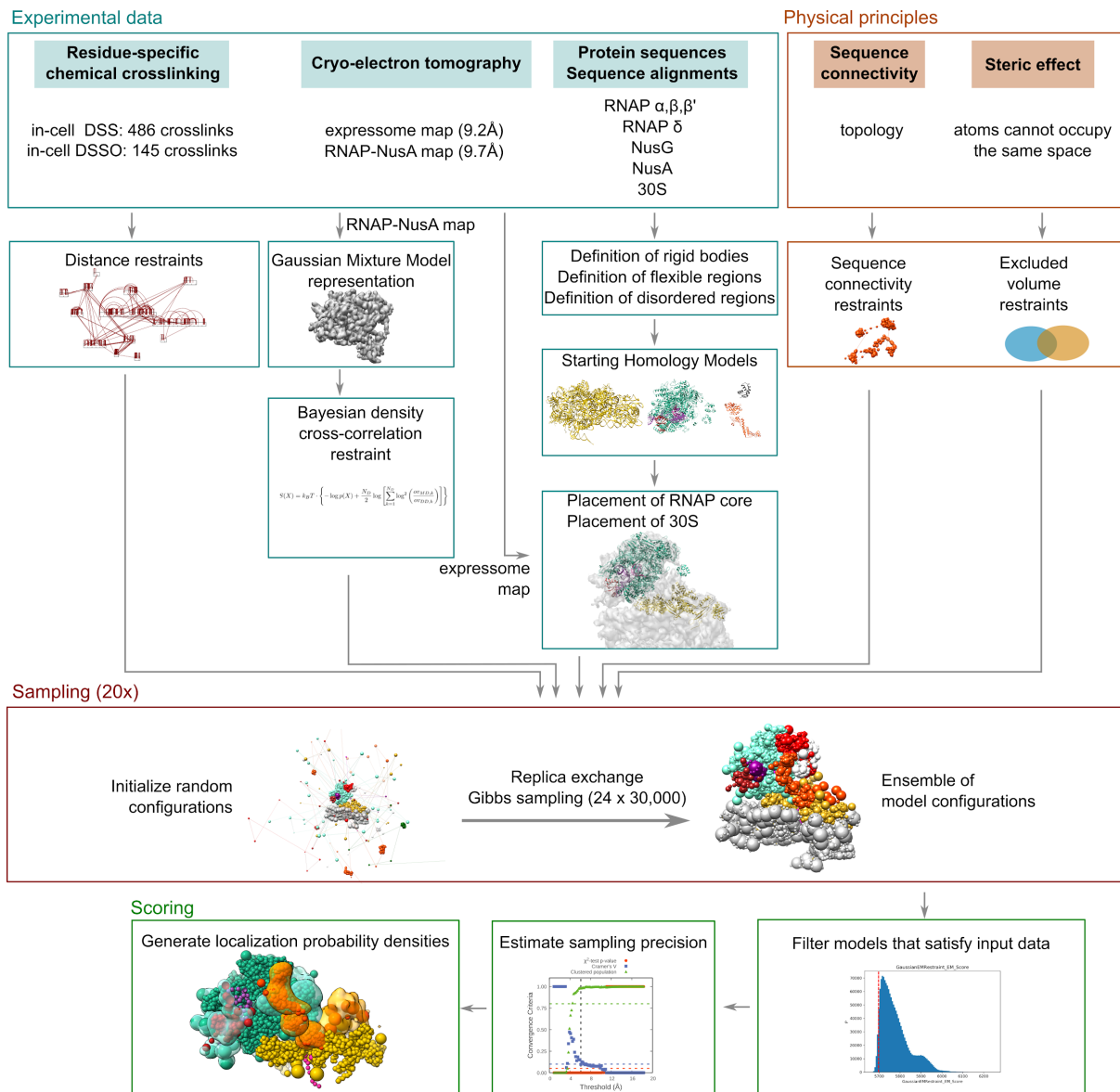


Fig. S19.
Integrative modeling workflow.

* This manuscript has been accepted for publication in Science. This version has not undergone final editing. Please refer to the complete version of record at <http://www.sciencemag.org/>. The manuscript may not be reproduced or used in any manner that does not fall within the fair use provisions of the Copyright Act without the prior, written permission of AAAS

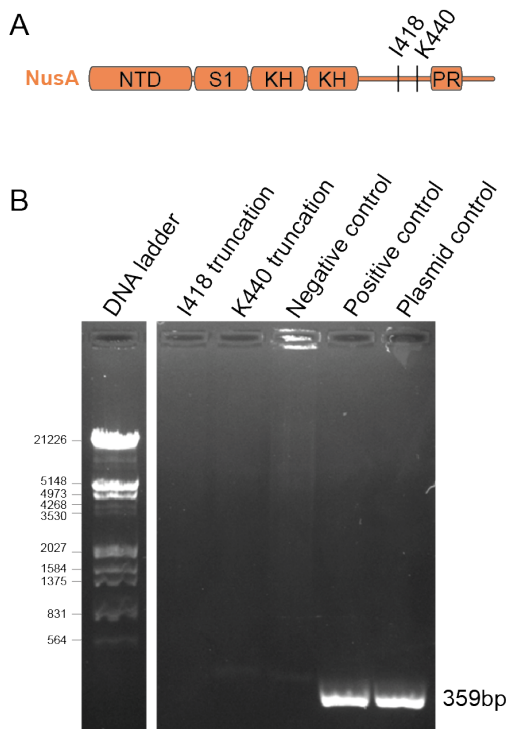


Fig. S20.

Genetic truncations showed that the C-terminal region of NusA is essential. (A) Scheme of *M. pneumoniae* NusA sequence with C-terminal truncation points highlighted; Ile⁴¹⁸ and Lys⁴⁴⁰. (B) Screening for C-terminal region mutants. Gel analysis of a PCR screen to detect presence of genome-inserted chloramphenicol resistance gene, chloramphenicol acetyltransferase, after the truncation site. Internal PCR probes NS73 and 74 yielded a product of 359 bp, i.e. the chloramphenicol resistance gene (Table S3). From left to right: PCR screening of transformants for both mutants showed no product in passaged colonies, indicating that the chloramphenicol resistance gene is absent and truncation is lethal. The negative control, consisting of the unmodified GP35 recombinase parent strain also showed no PCR product. The positive control, a deletion strain for *mpn668* (non-essential gene) showed the 359 bp PCR product. Another positive control, a plasmid containing the chloramphenicol resistance gene showed the same band.

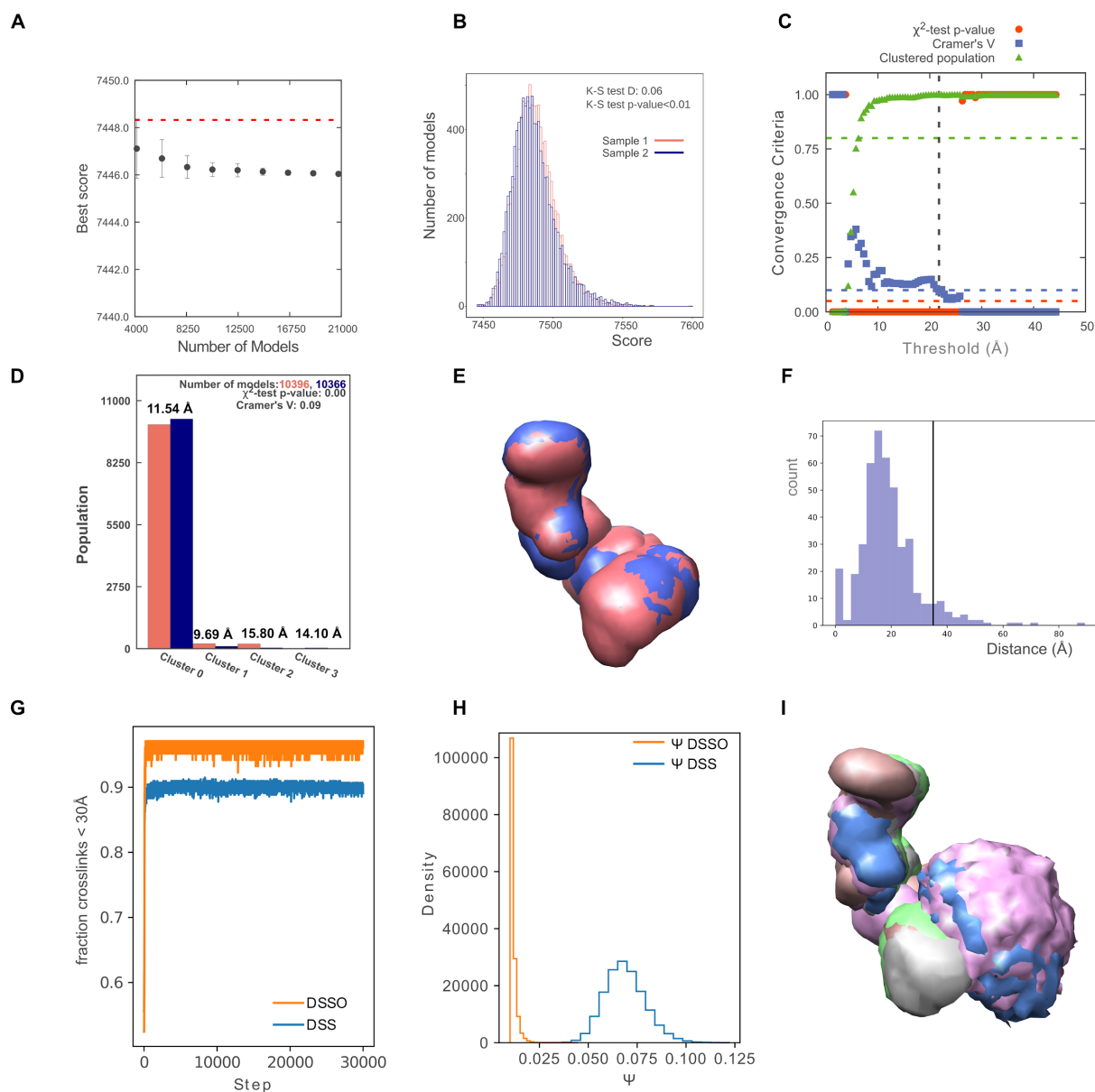


Fig. S21.

Sampling completeness and precision of the integrative modeling procedure. (A) Convergence of the model score in the 20,765 selected models. Minimum score of subsampling of random models is shown as mean +/- standard deviation (N=10). Score does not continue to improve as more models are sampled. (B) Comparison of score distributions for the selected models in (A). The distributions (N=10,399, N=10,366) display a negligible difference (D=0.06), despite having $p < 0.05$, indicating sampling completeness. Models in the first 10 modeling runs belong to sample 1, and models in the second 10 runs to sample 2. (C) Estimation of sampling

precision in the selected models as a function of r.m.s.d. sampling threshold. The vertical dashed line represents the threshold at which the p-value for homogeneity of proportions is satisfied ($p > 0.05$, or the effect size is small (Cramer's $V < 0.1$)) and $> 80\%$ of models fall within one of the selected clusters. This indicates an overall sampling precision of 21.8 \AA . This is the smallest clustering threshold at which each sample contributes models proportionally to its size. **(D)** Cluster populations obtained with an r.m.s.d. cutoff of 21.8 \AA based on the overall sampling precision. 97% of structures fall within one cluster, with a model precision (average r.m.s.d.-based distance to the cluster centroid model calculated over the ranges in Table S7) of 11.54 \AA . Colors refer to models originating from sample 1 or 2 in **(B)**. **(E)** Localisation probability densities for the NusA domains in models from the two random samples that fall within cluster 0, indicating the in-cluster variance. **(F)** Crosslink distance distribution in the cluster centroid model. The vertical line marks 35 \AA . The crosslink satisfaction is 93% ($89\% < 30 \text{ \AA}$). **(G)** Fraction of crosslinks having distance $< 30 \text{ \AA}$ in a representative modeling run ($N=20$) as a function of simulation frame. **(H)** Density plot for crosslink uncertainty (Ψ), roughly equivalent to the sampled false discovery rate, during the $30,000$ frames of a representative modeling run ($N=20$). DSS crosslinks sample Ψ values between 5 and 10% , while DSSO crosslinks lie $< 2.5\%$. **(I)** Localization probability densities for the NusA domains in models derived from 5 sets of 2 runs each in which 20% of the crosslinks were discarded at random. The structured region of NusA that are fitted to the cryo-EM density are robust to crosslink subsampling, while the NusA C-terminal region is highly variable.

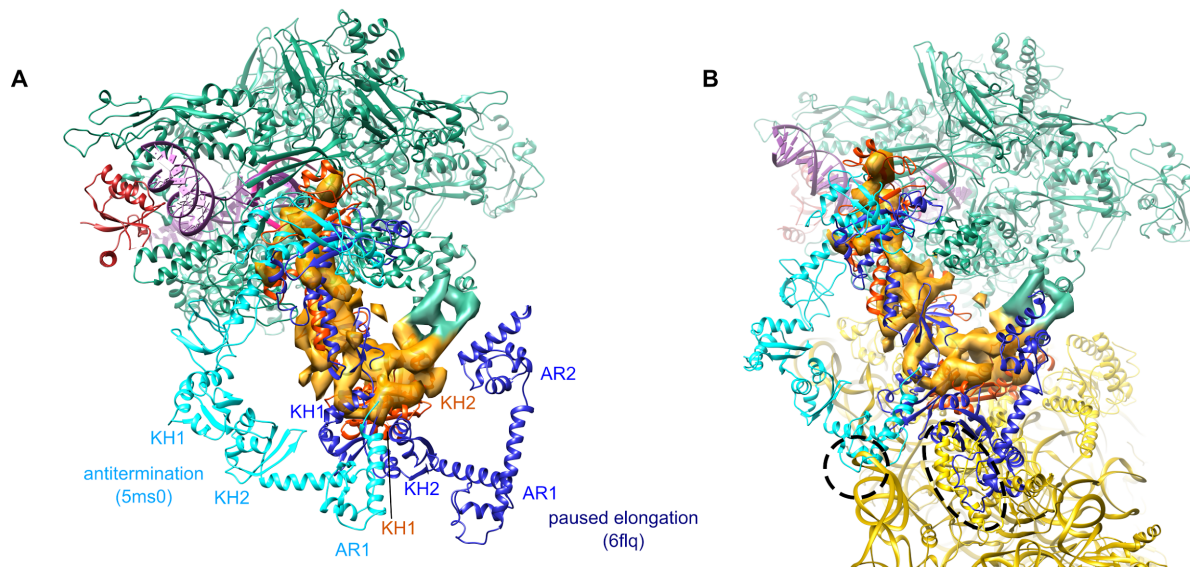


Fig. S22.

NusA conformation compared to antitermination and paused elongation complexes. (A) Overlay of the arrangement of NusA relative to RNAP (green) in the *M. pneumoniae* expressome, the *E. coli* λ -induced antitermination complex (cyan) and the *E. coli* paused elongation complex (dark blue). The NusA density segmented from *M. pneumoniae* elongating expressome cryo-EM density is also shown (orange). While the arrangement of the NTD and S1 domains are similar to the paused elongation configuration, the KH domains are rearranged closer to RNAP in the active expressome. (B) The NusA arrangements as shown in (A) relative to the ribosome in the expressome structure. Both antitermination and paused elongation *E. coli* structures clash with the 30S subunit in the regions highlighted by circles. The segmented densities for the 30S subunit (yellow) and NusA (orange) are also shown.

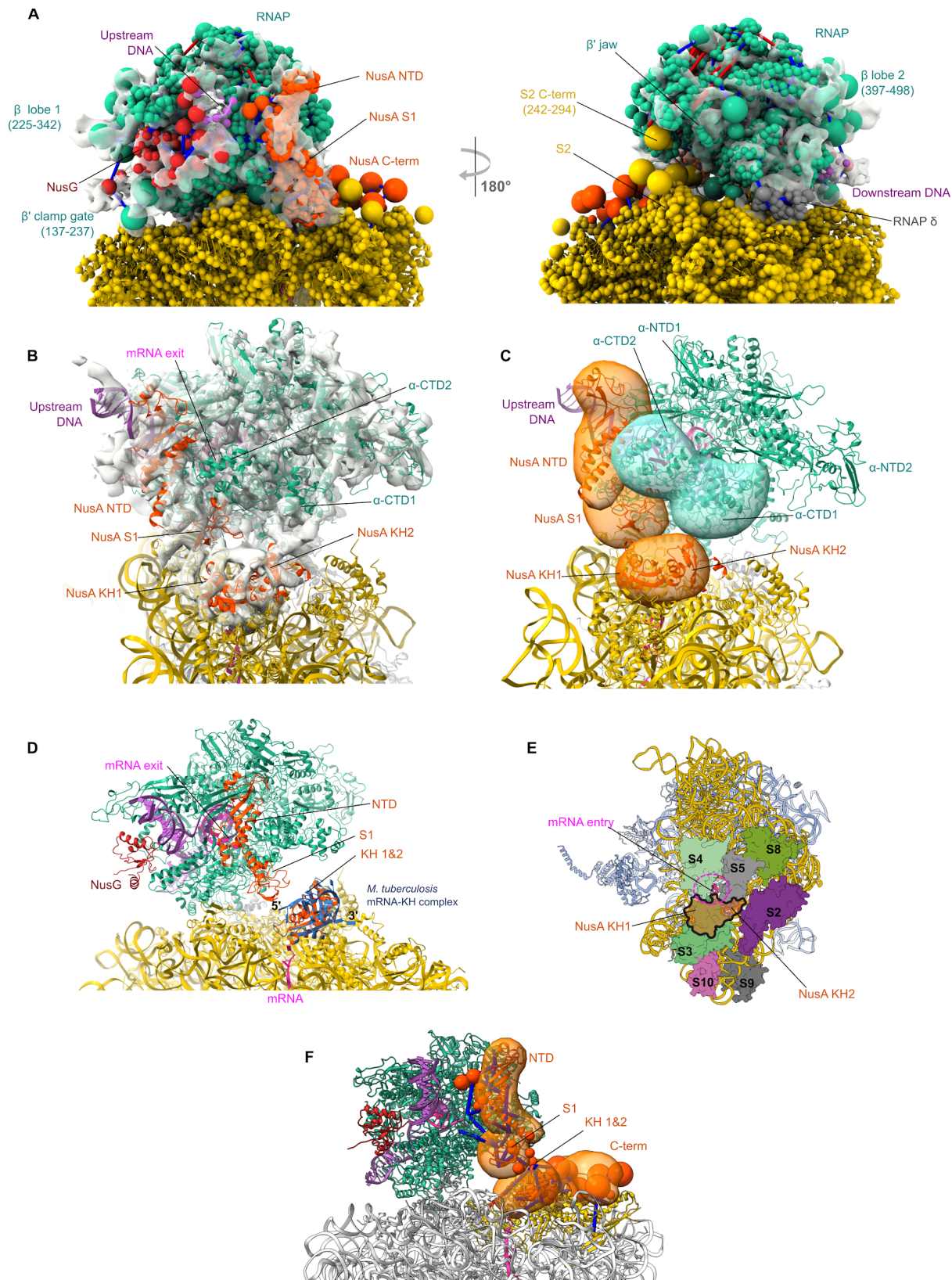


Fig. S23.

Crosslinking-assisted modeling of NusA in the cryo-EM density. (A) View of the centroid model of the *M. pneumoniae* elongating expressome in the multi-body-refined cryo-EM map with crosslink restraints displayed. Satisfied crosslinks (<35 Å) are shown in blue, while those longer than 35 Å are in red. (B) View of the mRNA exit channel face of RNAP with the density for NusA and the α -CTDs shown. (C) Localization probability densities for the NusA domains (orange) and the α -CTDs (green) displayed. The α -CTDs are localized at low precision, spanning the region between NusA NTD and the second KH domain. (D) Overlay of *M. pneumoniae* NusA in the elongating expressome model (orange) with the structure of the *M. tuberculosis* NusA KH domains bound to RNA (blue, PDB 2ATW, (27)). The putative mRNA path in the elongating expressome model is shown in magenta. The orientation of mRNA on the KH domains is consistent with RNA interacting with the KH domains near the mRNA entry site on the ribosome. (E) Top view of 30S ribosome with NusA footprint highlighted. (F) View of the centroid model of the *M. pneumoniae* elongating expressome with crosslinks involving NusA displayed. Satisfied crosslinks (<35 Å) are shown in blue, while those longer than 35 Å are in red. Localization probability densities for the NusA domains are shown in orange.

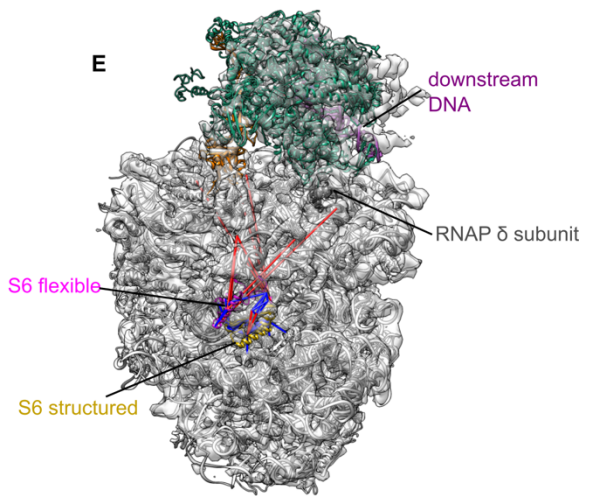
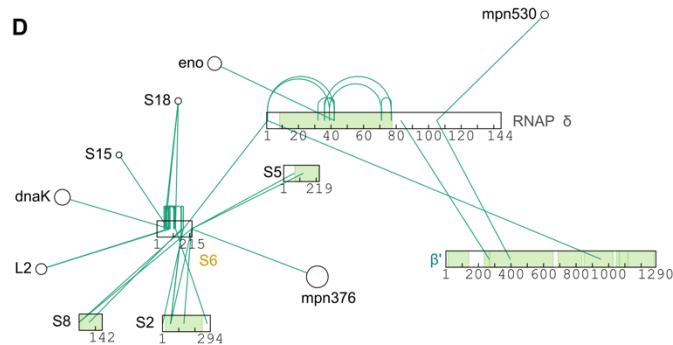
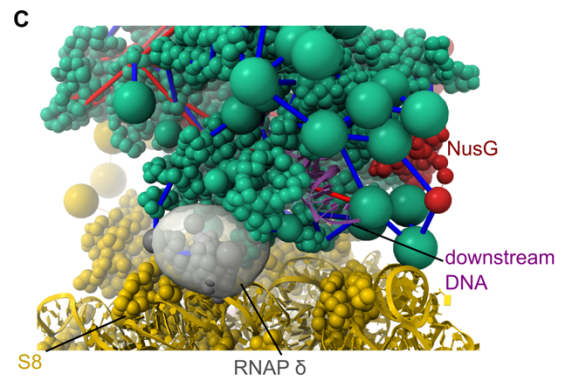
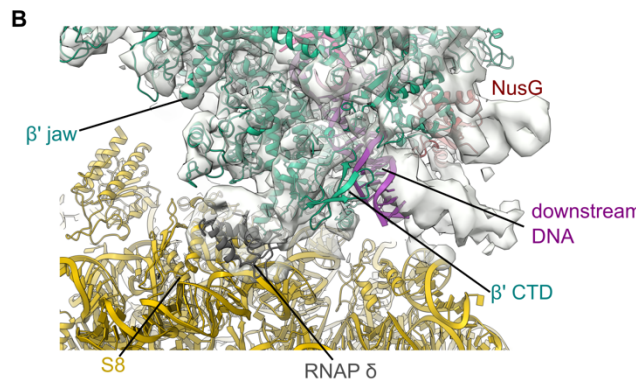
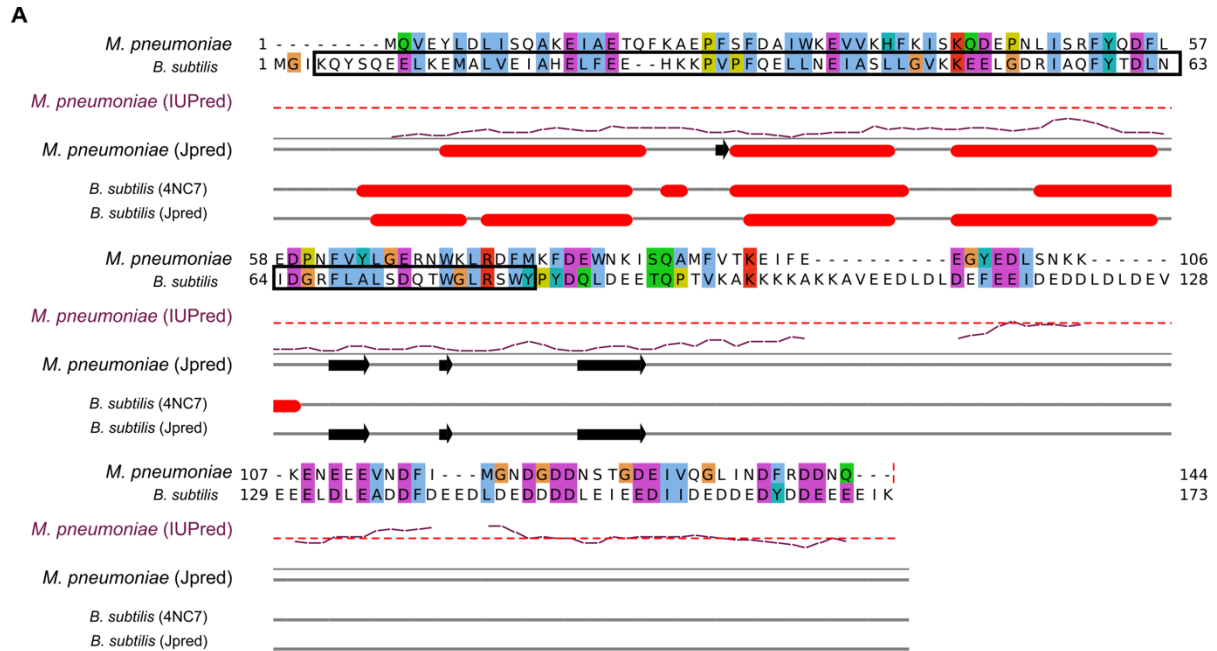


Fig. S24.

Localisation of the RNAP δ subunit. (A) Sequence alignment of *M. pneumoniae* and *B. subtilis* RNAP δ subunit (RpoE). The N-terminal domain, covered in the *B. subtilis* crystal structure (PDB 4NC, (91)) is highlighted by a black box. The secondary structure was predicted with Jpred. β -sheets are shown as arrows, α -helices as cylinders. The long-range disorder prediction of IUPred is also shown, with the threshold of 50% disorder likelihood displayed as a dashed line, indicating a disordered C-terminal region. (B) Placement of the NTD of the RNAP δ subunit in the elongating expressome integrative model in proximity of the downstream DNA, ribosomal protein S8 and the β' CTD. (C) Localisation density for RNAP δ in the most populated selected cluster displayed with crosslinks shown as lines. Satisfied crosslinks (<35 Å) are shown in blue, while those longer than 35 Å are in red. (D) The crosslink-derived interaction network of RNAP δ and ribosomal protein S6, highlighting the single link connecting the flexible RNAP δ N-terminus with the highly promiscuous S6 C-terminus. Structured regions included in the IMP protocols are shown in green. (E) While the structured NTD of S6 (yellow) is homologous to *B. subtilis* and is observed in the density, its C-terminal domain is likely to be very flexible (here, a prediction by I-TASSER is shown in magenta) and crosslinks to several 30S proteins far apart from each other (>60Å). The link to RNAP δ (gray) is equally overlength (110 Å), which may be due to the disordered nature of the S6 CTD, or to the fact that this interaction occurs outside the fully assembled ribosome.

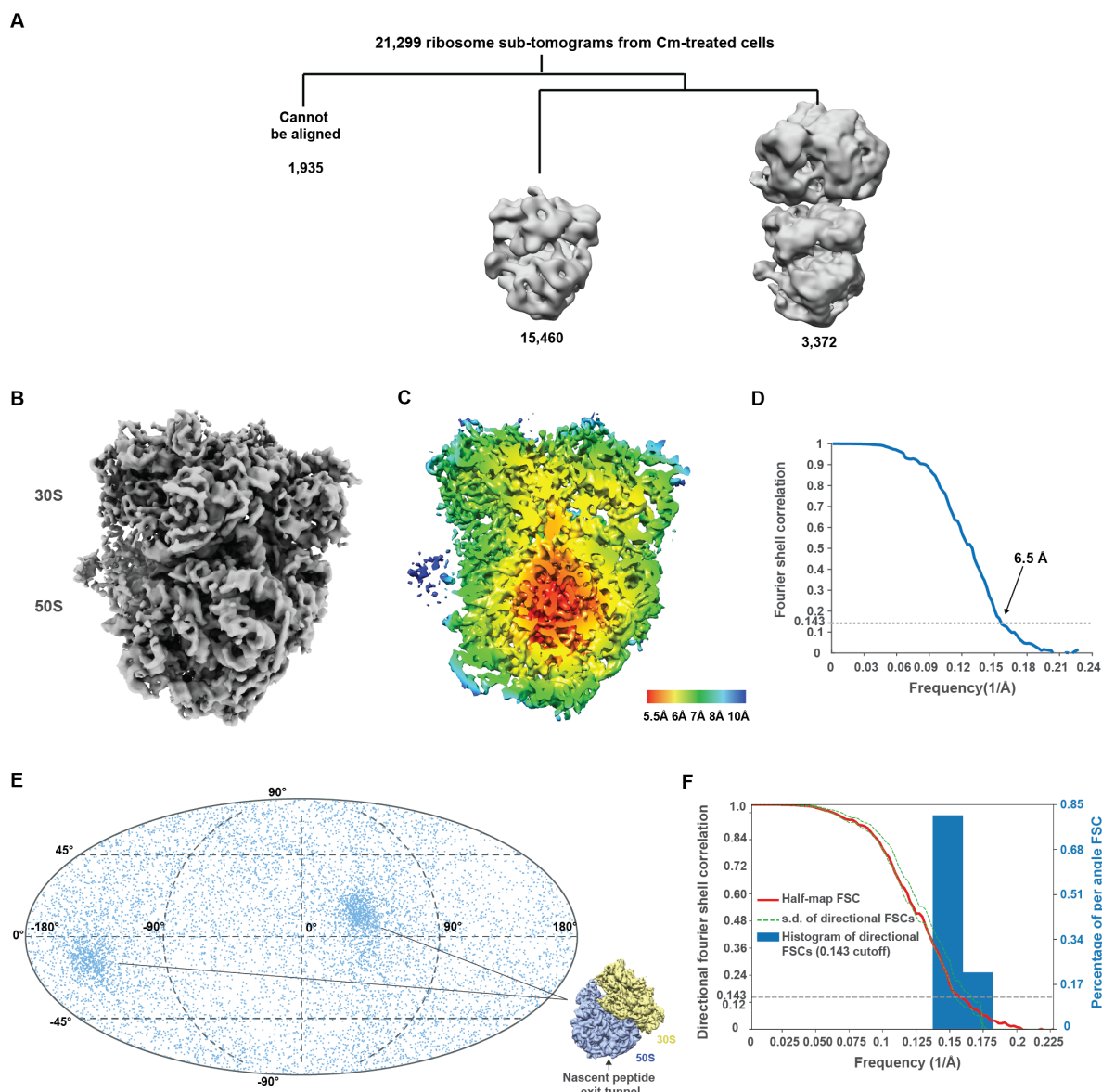


Fig. S25.

Inhibiting translation by chloramphenicol de-coupled RNAP and the ribosome. (A) Classification of 21,299 ribosome sub-tomograms from 65 tomograms of chloramphenicol (Cm)-treated *M. pneumoniae* cells. No significant density was detected despite focused classification efforts. (B) Refined Cm-treated 70S ribosome density. (C) Central slice of the local resolution map. (D) Overall resolution of the map is 6.5 Å (FSC=0.143). (E) Angular distribution map shows similar weak preferential angular distribution to the untreated dataset (Fig. S8). (F) 3D FSC analysis. Sphericity of 0.938 indicates low anisotropy.

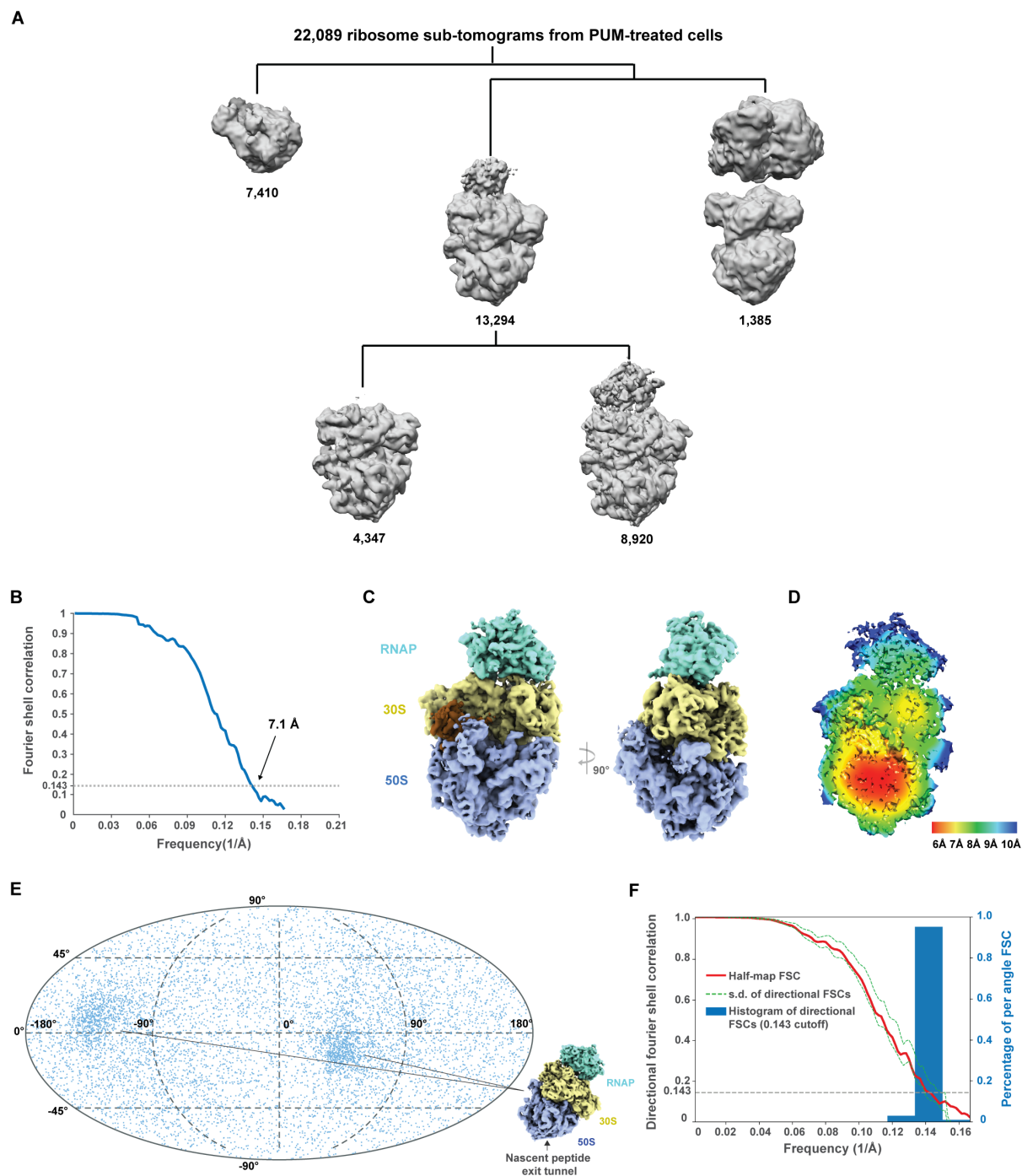


Fig. S26.

Stalling RNAP by pseudouridimycin caused rearrangement of the expressome architecture.

(A) Classification of 22,089 ribosome sub-tomograms from 83 tomograms of pseudouridimycin (PUM)-treated *M. pneumoniae* cells. The percentage of classified RNAP-ribosome

56

* This manuscript has been accepted for publication in Science. This version has not undergone final editing. Please refer to the complete version of record at <http://www.sciencemag.org/>. The manuscript may not be reproduced or used in any manner that does not fall within the fair use provisions of the Copyright Act without the prior, written permission of AAAS

supercomplexes increased dramatically compared to untreated cells (as in Fig. S7), accounting for 40% of localized ribosomes. **(B)** Stalling RNAP by PUM reduced the heterogeneity and flexibility, and the supercomplex density was refined to 7.1 Å overall resolution (FSC=0.143). **(C)** Refined density of the RNAP-ribosome supercomplex. Density corresponding to NusA was not found. Identification of EF-G density (brown) associated with the ribosome indicates a pre-translocational state. **(D)** Central section of the local resolution map. **(E)** Angular distribution map of the 8,920 supercomplexes. Similar weak preferential angular distribution was observed as before (as in Fig. S8, 25). **(F)** 3D FSC analysis. Sphericity value is 0.942, indicating low anisotropy.

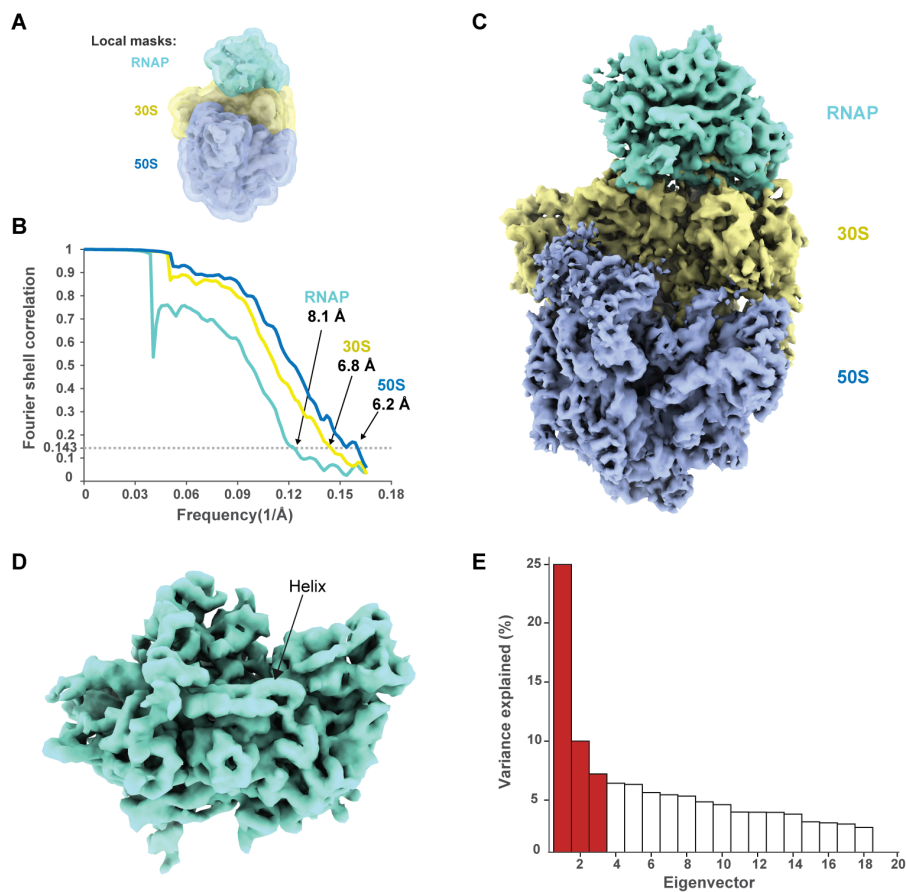
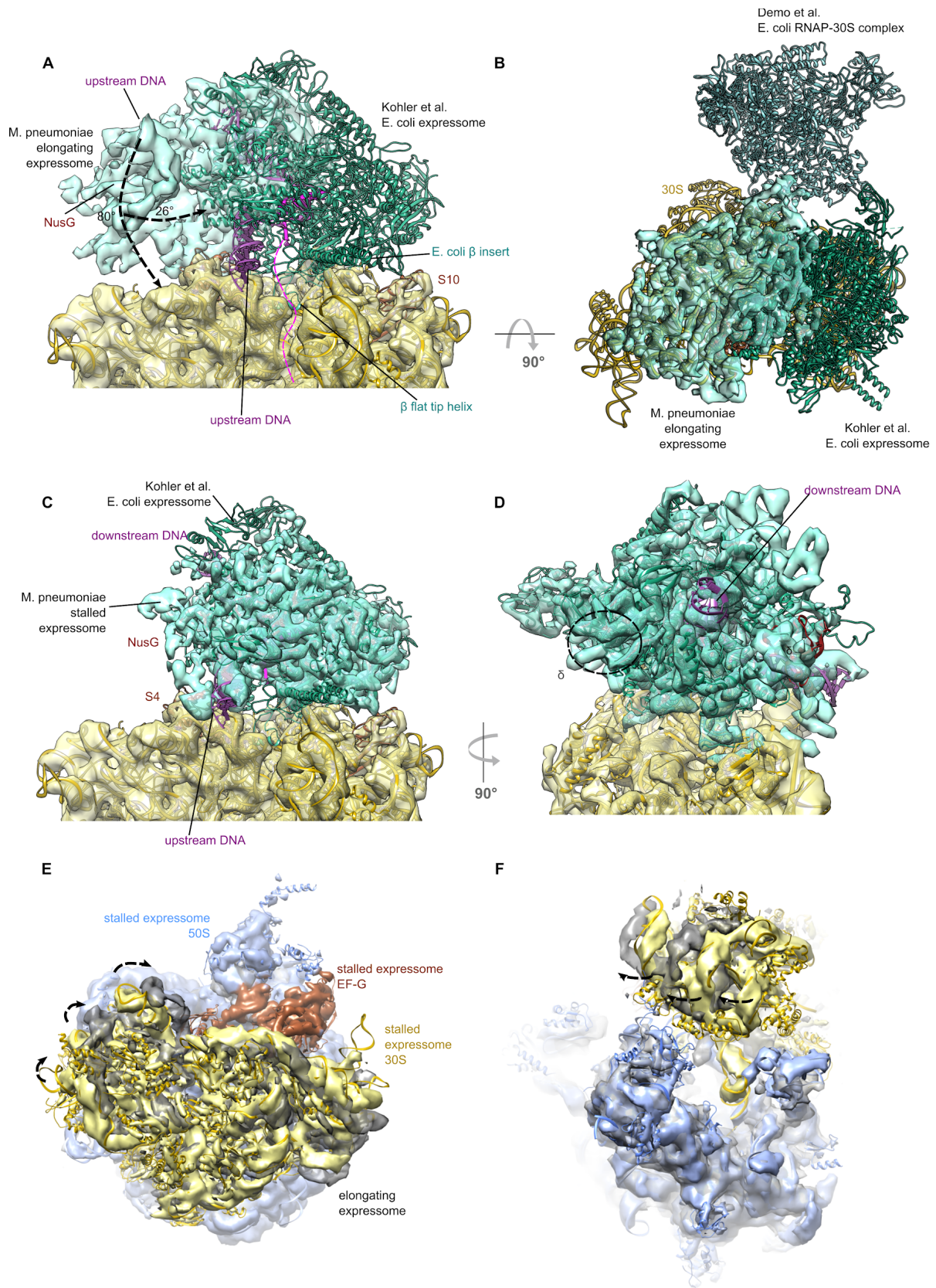


Fig. S27.

Multi-body refinement of PUM-induced, stalled expressomes. (A) Segmented bodies from the consensus refinement and their corresponding local masks, i.e. RNAP, 30S, and 50S. (B) Resolution estimation for each body. (C) Superposition of the three bodies/densities from multi-body refinement. The overall resolution of the RNAP density was estimated to be 8.1 Å (FSC=0.143), at which long helices could be clearly resolved (D). (E) Plot of variance contributions of all eigenvectors. Movements corresponding to the top three eigenvectors (in red) are displayed in Movie S2.



* This manuscript has been accepted for publication in Science. This version has not undergone final editing. Please refer to the complete version of record at <http://www.sciencemag.org/>. The manuscript may not be reproduced or used in any manner that does not fall within the fair use provisions of the Copyright Act without the prior, written permission of AAAS

Fig. S28.

Comparison of the untreated in-cell elongating expressome density and PUM-induced density to the expressome structure from Kohler *et al.* (A) The untreated *M. pneumoniae* expressome densities corresponding to RNAP and the 30S subunit overlaid with the cryo-EM structure of the *E. coli* expressome in cartoon representation (PDB 5MY1, (6)). The density corresponding to *M. pneumoniae* NusA has been omitted for clarity, and nucleic acids have been modelled in the *E. coli* expressome structure, the mRNA exit site on RNAP sits directly over the mRNA entry channel in the 30S, where the β flat-tip helix also resides, not leaving space for additional factors. Compared to the untreated *M. pneumoniae* RNAP orientation, RNAP tilts back towards ribosomal protein S10 and undergoes a rotation by which the upstream DNA points towards the ribosome rather than away from it. (B) Top view of the 30S subunit with comparison of the RNAP orientations in the untreated expressome in the present study with the orientations found in *in vitro* high resolution structures of the *E. coli* expressome (PDB 5MY1, (6); PDB 6AWD, (12)). The density corresponding to *M. pneumoniae* NusA has been omitted for clarity. (C) The PUM-treated *M. pneumoniae* expressome densities corresponding to RNAP and the 30S subunit determined in this study overlaid with the cryo-EM structure of the *E. coli* expressome in cartoon representation. Stalling *M. pneumoniae* RNAP with PUM leads to the same expressome configuration as observed in the *E. coli* expressome, indicating that this configuration may be the product of a ribosome reaching and “colliding” with a stalled RNAP. Nevertheless, *M. pneumoniae* RNAP retains NusG in the arrested state, as observed by the additional density present around the upstream DNA (Fig. 4). Interestingly, in the stalled expressome we did not detect any density corresponding to the α -CTDs. (D) Homology model of *M. pneumoniae* RNAP fitted into the multi-body refined RNAP cryo-EM density from PUM-treated cells. The circle highlights unexplained density at the position of δ in the elongating expressome near the β' CTD and the downstream DNA. (E) Superposition of the ribosome in the *M. pneumoniae* stalled expressome (colored) to the ribosome in the elongating expressome (gray). The densities are aligned on the 50S subunit. The stalled expressome density is fitted with the homology model derived from the pre-translocation ribosome stalled with viomycin (PDB 4V7D, (30)). The stalled expressome 30S rotation state is consistent with the pre-translocation state. (F) Side view of (E).

Table S4.
Homology models used in integrative modeling.

SUBUNIT (UniprotKb)	Template structure (PDB ID)	Modeling method (quality estimates)¹	IMP rigid bodies	Fully flexible coarse grained regions (residues per bead)
RNAP α/ rpoA (Q50295)	6FLQ	Swiss-Model (GMQE: 0.66; QMEAN: -4.52 Seq. identity 41%) ²	1-233; (core) 234-325 (CTD)	240-265 (5)
RNAP β/ rpoB (P78013)	6FLQ	Swiss-Model (GMQE: 0.66; QMEAN: -4.52 Seq. identity 41%) ²	1-984 + 1001-1356; (core) 985-1000 (flat-tip helix)	225-342; (25) 397-498; (25) 981-989; (5) 999-1005; (5)
RNAP β' /rpoC (P75271)	6FLQ	Swiss-Model (GMQE: 0.66; QMEAN: -4.52 Seq. identity 41.0%) ²	Whole sequence	137-237; (25) 266-272; (25) 658-690; (25) 836-848; (25) 1027-1047; (25) 1057-1063; (25) 1117-1120; (25)
NusG 1-160 (P75049)	6C6U	Swiss-Model (GMQE: 0.15 QMEAN: -5.91 Seq. identity: 22.1%)	1-150 (NTD)	48-91; (5) 147-150 (5)
RNAP δ/ rpoE 1-85 (P75090)	2KRC; 2M4K; 4NC7	MODELLER (GA341: 0.8911 z-DOPE: -1.183)	1-85 (structured region linked to RNAP)	81-85 (5)
S5/rpsE (Q50301)	3J9W	Swiss-Model (GMQE 0.61 QMEAN -1.44 Seq. identity 51.6%)	Whole sequence	1-69; (25) 213-219 (25)
S2/rpsB (P75560)	3J9W	Swiss-Model (GMQE 0.60 QMEAN -5.32 Seq. identity 35.3%)	Whole sequence	1-19; (25) 242-294 (25)
S10/rpsJ (P75581)	3J9W	Swiss-Model (GMQE: 0.70 QMEAN -4.93 Seq. identity 40.6%)	Whole sequence	1-11 (25)
S3/rpsC (P41205)	3J9W	Swiss-Model (GMQE: 0.61 QMEAN -3.14 Seq. identity 43.1%)	Whole sequence	207-273 (25)
S4/rpsD (P46775)	3J9W	Swiss-Model (GMQE: 0.76 QMEAN -2.71 Seq. identity 45.7%)	Whole sequence	203-205 (25)

S8/rpsH (Q50304)	3J9W	Swiss-Model (GMQE: 0.76 QMEAN: -1.24 Seq. identity: 50.8%)	Whole sequence	1-9; (25)
S9/rpsI (P75179)	3J9W	Swiss-Model (GMQE: 0.78 QMEAN: -3.32 Seq. identity: 56.9%)	Whole sequence	1-5;
S20/rpsT (P75237)	3J9W	Swiss-Model (GMQE: 0.66 QMEAN -1.85 Seq. identity: 30.4%)	Whole sequence	
S19/rpsS (P75576)	3J9W	Swiss-Model (GMQE: 0.75 QMEAN -3.83 Seq. identity: 55.8%)	Whole sequence	
S13/rpsM (Q50297)	3J9W	Swiss-Model (GMQE: 0.72 QMEAN -5.66 Seq. identity: 63.3%)	Whole sequence	
S7/rpsG (P75545)	3J9W	Swiss-Model (GMQE: 0.80 QMEAN -1.78 Seq. identity: 59.7%)	Whole sequence	
S11/rpsK (Q50296)	3J9W	Swiss-Model (GMQE: 0.77 QMEAN -2.83 Seq. identity: 48.7%)	Whole sequence	
S18/rpsR (P75541)	3J9W	Swiss-Model (GMQE: 0.56 QMEAN -1.91 Seq. identity: 41.4%)	Whole sequence	
S6/rpsF (P75543)	3J9W	Swiss-Model (GMQE: 0.26 QMEAN -2.08 Seq. identity: 21.7%)	Whole sequence	
S15/rpsO (P75173)	3J9W	Swiss-Model (GMQE: 0.78 QMEAN -0.81 Seq. identity: 44.1%)	Whole sequence	
S14(Z)/rpsN (Q50305)	3J9W	Swiss-Model (GMQE: 0.79 QMEAN -2.92 Seq. identity: 67.7%)	Whole sequence	
S17/rpsQ (Q50309)	3J9W	Swiss-Model (GMQE: 0.77 QMEAN -2.85 Seq. identity: 67.7%)	Whole sequence	
S12/rpsL (P75546)	3J9W	Swiss-Model (GMQE: 0.8 QMEAN - 3.97 Seq. identity: 76.1%)	Whole sequence	
S16/rpsP (P75131)	3J9W	Swiss-Model (GMQE: 0.77 QMEAN -0.56 Seq. identity: 52.5%)	Whole sequence	
16S RNA	3J9W	Original sequence	Whole sequence	
P-site tRNA	3J9W	Original sequence	Whole sequence	
NusA (P75591)	1HH2; 1L2F; 5LM7; 6FLQ; 6GOV; 5MS0; 6J9E; 5LM9; 4MTN;	MODELLER: (GA341 1.00 zDOPE: 0.42)	1-140 (NTD); 145-213 (S1 domain); 214-364 (KH1 and KH2)	1-11; 141-145; 212-216; 365-540

	1K0R; 2ASB; 2ATW			
--	------------------------	--	--	--

1: Homology model scores refer to the overall model quality prior to removal and coarse graining of any unstructured regions or regions not covered by alignment in subsequent integrative modeling steps.

Bold: Modelled as a single complex. Quality estimates and sequence identity refer to the combined α - β - β' model.

Table S5.

Coarse graining and gaussian mixture model (GMM) representations in integrative modeling. Each region was coarse grained at the resolutions described. The “rigid body” column indicates whether these proteins belong to the same or different building blocks (numbering is arbitrary). Building block 5 (RNAP) and 10 (30S) were fitted as single bodies with Chimera and were not allowed to move during modeling, except for the “fully flexible regions” annotated in Table S4.

Protein	UniProtID	Residues	Residues per Bead	Residues per Gaussian	Belonging to rigid body
NusA	P75591	1,142	5	10	1
NusA	P75591	143,213	5	10	2
NusA	P75591	214,364	5	10	3
NusA	P75591	365,END	25	0	4
RNAP α	Q50295.1	1,233	25	10	5
RNAP α	Q50295.1	234,END	5	10	6
RNAP α	Q50295.2	1,233	25	10	5
RNAP α	Q50295.2	234,END	5	0	7
RNAP β	P78013	1,984	25	20	5
RNAP β	P78013	985, 1000	5	5	8
RNAP β	P78013	1001,END	25	10	5
RNAP β'	P75271	11,063	25	20	5
RNAP β'	P75271	10,641,117	5	10	9
RNAP β'	P75271	1118,END	25	20	5
NusG	P75049	1,150	5	10	5
S5	Q50301	all	25	0	10
S2	P75560	all	25	0	10
S10	P75581	all	25	0	10
S3	P41205	all	25	0	10
S4	P46775	all	25	0	10
S8	Q50304	all	25	0	10
S9	P75179	all	25	0	10

S20	P75237	all	25	0	10
RS19	P75576	all	25	0	10
S13	Q50297	all	25	0	10
S7	P75545	all	25	0	10
S11	Q50296	all	25	0	10
S18	P75541	all	25	0	10
S6	P75543	all	25	0	10
S15	P75173	all	25	0	10
S14 (Z)	Q50305	all	25	0	10
S17	Q50309	all	25	0	10
S12	P75546	all	25	0	10
S16	P75131	all	25	0	10
<i>B. subtilis</i> 16S RNA	N.A. (PDB ID 3J9W)	all	25	0	10
<i>B. subtilis</i> P-site tRNA	N.A. (PDB ID 3J9W)	all	10	0	10
RNAP δ	P75090	1,85	5	10	11

Table S6.
Integrative modeling score cutoffs.

Energy function term	Mean score*	Standard Deviation*	Selected Cutoff
DSS crosslink score	625.7	13.3	Imposed minimum 90% crosslinks <35 Å
DSSO crosslink score	1293.5	24.4	Imposed minimum 90% crosslinks < 35 Å
Bayesian EM score	5848.3	55.2	5775.0
Excluded Volume	78.8	26.7	60

* Mean score and standard deviations reported after pre-filtering to only include models with >90% crosslink satisfaction.

Table S7.
Integrative model regions included in precision and clustering calculation

Region	Residue range
NusA NTD	14-141
NusA S1	142-210
NusA KH	211-362
NusA C-terminal extension	363-540
RNAP δ	Whole sequence
NusG	45-90
RNAP α flexible region & CTD	234-322
RNAP β loops	225-342;398-498
RNAP β' loops	137-237;658-690

* This manuscript has been accepted for publication in Science. This version has not undergone final editing. Please refer to the complete version of record at <http://www.sciencemag.org/>. The manuscript may not be reproduced or used in any manner that does not fall within the fair use provisions of the Copyright Act without the prior, written permission of AAAS

* This manuscript has been accepted for publication in Science. This version has not undergone final editing. Please refer to the complete version of record at <http://www.sciencemag.org/>. The manuscript may not be reproduced or used in any manner that does not fall within the fair use provisions of the Copyright Act without the prior, written permission of AAAS

Movie S1.

Movements between the 3 bodies (i.e. RNAP-NusA, 30S, 50S) corresponding to each of the three most predominant eigenvectors, shown successively, for the native elongating expressome.

Movie S2.

Movements between the 3 bodies (i.e. RNAP, 30S, 50S) corresponding to each of the three most predominant eigenvectors, shown successively, for the PUM-stalled expressome.