Inaugural dissertation

for

obtaining the doctoral degree

of the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of the

Ruprecht - Karls - University

Heidelberg

Presented by

M.Sc. Liang Xue

born in: Shandong, China

Oral examination: April 7th, 2022

In-cell structural biology

of protein synthesis

in *Mycoplasma pneumoniae*

Referees: ……*Prof. Dr.* Irmgard Sinning……

……*Dr.* Martin Beck……

Abstract

To structurally and quantitatively delineate molecular processes inside the cell with atomic detail is the ultimate goal of structural biology. Cryo-electron tomography (cryo-ET) and sub-tomogram analysis provide currently the most promising approach to achieve this goal. In this thesis, I aim to push the limits of in-cell structural biology and achieve its full potential, using the small genome-reduced bacterium *Mycoplasma pneumoniae* as a cell model.

Translation, the fundamental process of protein synthesis, is mediated by ribosomes in all living cells. With comprehensive ribosome sub-tomogram datasets extracted *in silico* from tomograms of intact *M. pneumoniae* cells, I determined the in-cell ribosome structure at 3 Å resolution, the highest for complex structures resolved directly inside cells to date. The built atomic model represents the first ribosome model for mycoplasma species, revealing distinct ribosomal protein extensions. I then performed extensive sub-tomogram classification, and resolved over 13 translation intermediate structures at 4-10 Å resolutions. The analysis not only visualizes structural dynamics of translation elongation in native cells but also provides unique insights into its energy landscapes. Moreover, the spatial organization of translating ribosomes revealed by mapping back structures into the cell volume unraveled a local elongation coordination mechanism within polysomes, which is mediated by the ribosomal protein L9. Furthermore, I resolved more than 17 ribosome intermediate structures in three antibiotic-treated datasets and demonstrated how the drug molecules fundamentally reshape the energy and spatial landscapes of translation inside the cell.

The two fundamental processes of gene expression, transcription and translation, can be coupled in prokaryotes. By combining cryo-ET, in-cell crosslinking mass spectrometry and integrative structural modeling, I determined the architecture of an actively transcribing and translating expressome complex inside the cell. In the integrative model, the transcription factor NusA was pinpointed at the interface between the NusG-bound RNA polymerase and the ribosome to mediate transcription-translation coupling. Inhibiting translation by antibiotics dissociated the expressome, whereas inhibiting transcription stalled and rearranged it, demonstrating that the active expressome assembly requires ongoing translation and transcription within the native cellular context.

These works provide unprecedented insights into molecular machines in action within the cell, and herald a new era in the emerging field of integrative in-cell structural biology. This thesis demonstrates its possibility to visualize dynamic molecular machineries, and eventually the molecular sociology, with atomic detail inside cells.

Zusammenfassung

Das ultimative Ziel der Strukturbiologie ist es, molekulare Prozesse in der Zelle strukturell und quantitativ mit atomaren Details zu beschreiben. Kryo-Elektronentomographie (Kryo-ET) und die Analyse von Subtomogrammen bieten derzeit den vielversprechendsten Ansatz, um dieses Ziel zu erreichen. In dieser Dissertation möchte ich die Grenzen der zellinternen Strukturbiologie erweitern und ihr volles Potenzial ausschöpfen, indem ich das kleine genomreduzierte Bakterium *Mycoplasma pneumoniae* als bakterielles Zellmodell verwende.

Die Translation, der grundlegende Prozess der Proteinbiosynthese, wird in allen lebenden Zellen durch Ribosomen vermittelt. Anhand umfassender Datensätze von Ribosomen Subtomogrammen, die *in silico* aus Tomogrammen intakter *M. pneumoniae*-Zellen extrahiert wurden, bestimmte ich die Struktur des Ribosoms in der Zelle mit einer Auflösung von 3 Å, die bisher die höchste für makromolekulare Strukturen ist, die direkt in Zellen bestimmt wurden. Das erstellte Atommodell ist das erste Modell für Mycoplasma-Ribosomen, das deutliche ribosomale Proteinerweiterungen zeigt. Anschließend führte ich eine umfangreiche Subtomogramm-Klassifizierung durch und bestimmte mehr als 13 Translations-Zwischenstrukturen mit Auflösungen von 4-10 Å . Die Analyse visualisiert nicht nur die strukturelle Dynamik der Translationselongation, sondern lieferte auch einzigartige Einblicke in ihre Energielandschaften in nativen Zellen. Darüber hinaus enthüllte die räumliche Organisation translatierender Ribosomen, die durch Zurückkartieren von Strukturen in das Zellvolumen aufgedeckt wurde, einen lokalen Koordinationsmechanismus innerhalb von Polysomen, welcher durch das ribosomale Protein L9 vermittelt wird. Darüber hinaus habe ich mehr als 17 Ribosomen-Zwischenstrukturen in drei mit Antibiotika behandelten Datensätzen aufgelöst und dabei demonstriert, wie diese Wirkstoffmoleküle die Energie- und räumlichen Landschaften der Translation innerhalb der Zelle grundlegend umgestalten.

Die beiden grundlegenden Prozesse der Genexpression, Transkription und Translation, können in Prokaryoten gekoppelt sein. Durch die Kombination von Kryo-ET, In-cell crosslinking Massenspektrometrie und integrativer Strukturmodellierung habe ich die Architektur eines aktiv transkribierenden und translatierenden Expressom-Komplexes innerhalb der Zelle bestimmt. Im integrativen Modell wurde der Transkriptionsfaktor NusA an der Schnittstelle zwischen dem NusG-gebundenen RNA-Polymerase und dem Ribosom lokalisiert, um die Transkriptions-Translations-Kopplung zu vermitteln. Die Hemmung der Translation durch Antibiotika dissoziierte das Expressom, während die Hemmung der Transkription es zum Stillstand brachte und neu anordnete, was zeigt, dass der aktive Aufbau des Expressoms eine fortlaufende Translation und Transkription innerhalb des nativen zellulären Kontextes erfordert.

Diese Arbeiten bieten beispiellose Einblicke in molekulare Maschinen, die in der Zelle aktiv sind, und läuten eine neue Ära im aufstrebenden Gebiet der integrativen Zellstrukturbiologie ein. Diese Arbeit demonstriert die Möglichkeit, dynamische molekulare Mechanismen und letztlich die molekulare Soziologie im Inneren von Zellen mit (fast) atomaren Details zu visualisieren.

Acknowledgement

First, I want to thank my supervisor Julia Mahamid, who gave me the chance to do a PhD at EMBL and has taught me many things. Julia's working spirit, pursuit of excellence, and humbleness always inspire me.

I am grateful to Martin Beck, Irmgard Sinning, Nassos Typas and Bernd Bukau in my thesis committee. They trusted me in the hard times. I appreciate Martin for his honest advices in many aspects.

I am thankful to Juri Rappsilber, and the wonderful people in his group, Francis O'Reilly, Swantje Lenz, Andrea Graziadei and Ludwig Sinn. This thesis started from a collaboration with their group.

The thesis would not be what it looks now without the great collaboration with Dimitry Tegunov and Patrick Cramer.

I am thankful to Nikolaus Grigorieff, Maria Zimmermann-Kogadeeva, Peer Bork, Jörg Stülke, Cedric Blötz, Xiangrui Zeng, Min Xu, and many other collaborators. Also thanks to Christian Spahn, Christoph Müller, Anne-Claude Gavin, Jan Kosinski, Luis Serrano and Richard Herrmann for offering their expertise.

This work would not be possible without the fantastic environment at EMBL. Especially I appreciate Xiaojie Zhang for her continuous help throughout the PhD. And I want to thank Ievgeniia Zagoriy, Herman Fung, Sara Goetz, Mauricio Toro-Nahuelpan and many other amazing people in the group. Thanks to Wim Hagen, Felix Weis, Thomas Hoffmann and others for the technical support. Also thanks to my master students Mahmoud Chaabou and Lucas Adam.

Last, but most importantly, I want to thank my parents and brother. Their understanding and support give me strength.

Table of Contents

Abstract ..... **I**
Zusammenfassung ..... **II**
Acknowledgement ..... **III**
List of Abbreviations ..... **VI**
Chapter 1: Introduction **1**
1.1 Towards structural biology inside the cell 1
1.2 *Mycoplasma pneumoniae* as a simple bacterial cell model 4
1.3 Translation and the ribosome 5
1.3.1 Structural investigation of ribosomes 5
1.3.2 Mechanistic understanding of translation 6
1.3.3 Structural studies of ribosomes inside the cell 9
1.4 Transcription-translation coupling 9
1.5 Cryo-ET and sub-tomogram analysis 12
1.5.1 Principles of electron tomography 12
1.5.2 Cryo-ET sample preparation and data collection 13
1.5.3 Cryo-ET data processing 15
1.5.4 Sub-tomogram analysis 16
1.6 Integrative in-cell structural approaches 19
1.6.1 In-cell crosslinking mass spectrometry 19
1.6.2 Additional methods for in-cell structural biology 20
1.6.3 Integrative modeling 21
Chapter 2: Materials and Methods **23**
2.1 Cryo-ET sample preparation and data collection 23
2.1.1 *Mycoplasma pneumoniae* cell cultivation 23
2.1.2 Cryo-EM grid preparation 23
2.1.3 Cryo-ET data collection 23
2.2 Cryo-ET data processing and high-resolution ribosome averaging 24
2.2.1 Initial tilt-series alignment 24
2.2.2 Ribosome localization by template matching 25
2.2.3 Image processing in Warp 26
2.2.4 Structure refinement in RELION and M 26
2.2.5 Ribosome atomic model building 27
2.2.6 Bioinformatic analysis of ribosomal protein extensions 28
2.3 Ribosome classification and spatial analysis 28
2.3.1 Classification of the translation states of ribosomes 28

2.3.2 Model building for classified ribosomes 30
2.3.3 Spatial analysis of ribosomes and polysomes 30
2.3.4 RELION classification of polysomes 31
2.3.5 Statistical analysis of translation elongation states in polysomes 32
2.3.6 Single cell clustering analysis 33
2.4 Integrative structural study of expressome 33
2.4.1 Classification and refinement of the expressomes 33
2.4.2 Whole-cell crosslinking mass spectrometry 35
2.4.3 Integrative modeling of the expressome based on cryo-ET and CLMS data 35
Chapter 3: High-resolution in-cell ribosome structures 37
3.1 Cryo-ET of *M. pneumoniae* cells 37
3.1.1 On-grid cultivation 37
3.1.2 Cryo-ET data collection and processing 38
3.2 Ribosome localization by template matching 39
3.3 High-resolution ribosome structures inside the cell 41
3.3.1 The Warp-RELION-M workflow 41
3.3.2 In-cell ribosome maps at 3 Å resolution 42
3.3.3 Factors behind high resolutions 44
3.4 Atomic modeling of ribosomes in cells 47
3.4.1 The first atomic model for mycoplasma ribosomes 47
3.4.2 Visualization of Cm binding to the ribosome inside the cell 48
3.4.3 Ribosome protein extensions in *M. pneumoniae* ribosome 50
3.4.4 Ribosome protein extensions exist throughput bacteria 53
Chapter 4: Translation landscapes in cells 55
4.1 Structural dynamics of the translation machinery 55
4.1.1 Considerations of the states of frozen cells 55
4.1.2 Extensive classification of translation states 55
4.1.3 Ribosome class structure refinement and modeling 58
4.1.4 Reconstructing the complete translation elongation cycle 61
4.1.5 Energy landscapes of translation elongation in cells 63
4.1.6 Ribosome classes involved in other translation phases 64
4.2 Spatial and functional organization of ribosomes 66
4.2.1 Mapping ribosomes back into 3D cellular volume 66
4.2.2 Detection of polysomes 66
4.2.3 Organization of ribosomes within the polysome 69
4.2.4 Coordination of translation elongation within the polysome 71
4.2.5 Polysome elongation coordination mediated by L9 74

4.2.6 Dynamics of L9 on ribosomes 77
4.3 Antibiotics reshape translation landscapes in cells 78
4.3.1 Ribosome classification in cells treated by chloramphenicol 78
4.3.2 Ribosome classification in cells treated by spectinomycin 80
4.3.3 Ribosome classification in cells treated by pseudouridimycin 82
4.3.4 Single cell clustering analysis based on translation profiles 84
4.3.5 Effects of antibiotics on spatial organization of ribosomes 85
4.3.6 Antibiotics fundamentally reshape translation in cells 87
Chapter 5: Transcription-translation coupling 89
5.1 In-cell structure of the RNAP-ribosome supercomplex by Cryo-ET 89
5.1.1 Focused classification on the mRNA entry region of ribosomes 89
5.1.2 Refinement of the RNAP-ribosome supercomplex 91
5.1.3 An unexplained density links the RNAP-ribosome supercomplex 93
5.2 In-cell CLMS revealed NusA links RNAP and the ribosome 95
5.3 In-cell architecture of the elongating expressome by integrative modeling 97
5.3.1 Preparation of building blocks for the expressome 97
5.3.2 Integrative modeling of the expressome 101
5.3.3 In-cell architecture of the elongating expressome 105
5.4 Dynamics of transcription-translation coupling in cells 107
5.4.1 Inhibiting translation decouples the ribosome from RNAP 107
5.4.2 Stalling RNAP by PUM reshapes transcription-translation coupling in cells 109
Chapter 6: Conclusions and Perspectives 113
6.1 In-cell structural biology with atomic detail 113
6.1.1 In-cell ribosome structure at 3 Å resolution 113
6.1.2 Bring atomic structures into the cell 114
6.2 Structural and functional landscapes of translation inside cells 115
6.2.1 Structural dynamics of the translation machinery 115
6.2.2 Spatial and functional organization of ribosomes in the cell 116
6.2.3 Antibiotics fundamentally reshape translation landscapes in cells 117
6.2.4 In-cell structural biology, from macromolecules to molecular sociology 118
6.3 Deciphering transcription-translation coupling in cells 120
6.3.1 A novel transcription-translation coupling mechanism mediated by NusA 120
6.3.2 New landmarks in transcription-translation coupling research 121
6.3.3 Understanding transcription-translation coupling inside the cell 123
6.4 A vision for future structural biology 124
List of main works and conclusions 125
List of publications and contributions 126

References 127

List of Abbreviations

2D	Two-dimensional
3D	Three-dimensional
30S	Prokaryotic small ribosomal subunit
50S	Prokaryotic large ribosomal subunit
Å	Angstrom
AI	Artificial intelligence
B. subtilis	*Bacillus subtilis*
℃	Celsius degree
CDS	Correlative double sampling
CLMS	Crosslinking mass spectrometry
Cm	Chloramphenicol
CNN	Convolutional neural network
Cryo-EM	Cryogenic electron microscopy
Cryo-ET	Cryogenic electron tomography
CTD	C-terminal domain
CTF	Contrast transfer function
Da	Dalton
DDD	Direct Detection Device
DNA	Deoxyribonucleic acid
DSS	Disuccinimidyl suberate
DSSO	Disuccinimidyl sulfoxide
e^-	Electron
EC_{50}	Half maximal effective concentration
E. coli	*Escherichia coli*
EF	Elongation factor
EF-G	Elongation factor G
EF-Tu	Elongation factor thermo unstable
EM	Electron Microscope
EMBL	European Molecular Biology Laboratory
EMDB	Electron Microscopy Data Bank
FDR	False discovery rate
FFT	Fast Fourier transform
FIB	Focused ion beam
FSC	Fourier shell correlation
FRET	Förster resonance energy transfer
GUI	Graphical user interface
IC_{50}	Half maximal inhibitory concentration
IMP	Integrative modeling platform
K^+	Potassium ion
K_D	Dissociation constant
kDa	1000 Dalton

LSU	Large ribosomal subunit
MOA	Mechanism of action
mRNA	Messenger RNA
MS	Mass spectrometry
ms	millisecond
M. p.	*Mycoplasma pneumoniae*
M. pneumoniae	*Mycoplasma pneumoniae*
nm	nanometer
NTD	N-terminal domain
PBS	Phosphate-buffered saline
PCA	Principal component analysis
PDB	Protein Data Bank
PPI	Protein-protein interaction
PUM	Pseudouridimycin
rRNA	Ribosomal RNA
RNA	Ribonucleic acid
RNAP	RNA polymerase
s	Second
Subtomo	Sub-tomogram
SNR	Signal-to-noise ratio
SPA	Single particle analysis
Spt	Spectinomycin
SSU	Small ribosomal subunit
StA	Sub-tomogram analysis/averaging
TEM	Transmission electron microscope
TM	Template matching
tRNA	Transfer RNA
TS	Tilt-series
µm	Micrometer

Chapter 1: Introduction

1.1 Towards structural biology inside the cell

Structural biology can provide the mechanistic understanding of life. Macromolecules and cells represent two pillars of modern biology on the microscopic scale. The cell is the basic structural and functional unit for all living organisms, and consists of the plasma membrane, small molecules, and macromolecules such as nucleic acids and proteins. Macromolecules and molecular interactions in the cell determine its living activities, which also depend on the native cellular context to function. Thus, to decipher structures of macromolecules and cells is central to appreciating life.

Traditional structural approaches, such as X-ray crystallography, nuclear magnetic resonance (NMR) spectrometry, and single particle cryo-electron microscopy (cryo-EM), focus on individual macromolecules and macromolecular complexes, which are usually isolated from the native cellular context and purified to generate homogeneous samples. X-ray crystallography is the most representative method in traditional structural biology, which not only helped reveal the double-helix DNA model that ushered molecular biology[1], but also determined the first protein structure[2,3]. To perform X-ray diffraction, a crystal comprised of highly ordered molecules is required, which poses the major technical limitation and often requires large amounts of purified samples[4]. Whether the structure solved in a crystal represents the native structure poses another major concern[4,5]. X-ray free electron laser has the potential to determine macromolecular structure in solution yet many issues need to be overcome[6]. NMR[7] can be used to study molecular structure and dynamics in solution but is limited to small size (<50 kDa). Cryo-EM and single particle analysis (SPA) emerge over the last decade as the most popular approach to determine atomic structures of large macromolecular complexes in the frozen-hydrated state[8,9]. With a large number of projection images of isolated macromolecules or complexes frozen in a thin amorphous ice layer, a three-dimensional (3D) reconstruction can be derived and used for atomic modeling. One advantage of single particle cryo-EM is that it can handle certain levels of heterogeneity, and computational classification can be performed to reveal structural dynamics of molecular machines[9]. The vast majority of reported structures in the protein data bank (PDB) are from the above methods, albeit a small number are derived from other approaches such as electron crystallography[10-13]. With the large PDB data and recent developments of deep learning methods, high-quality structure prediction has been demonstrated to be feasible[14-16], which possibly marks a new era of structural biology.

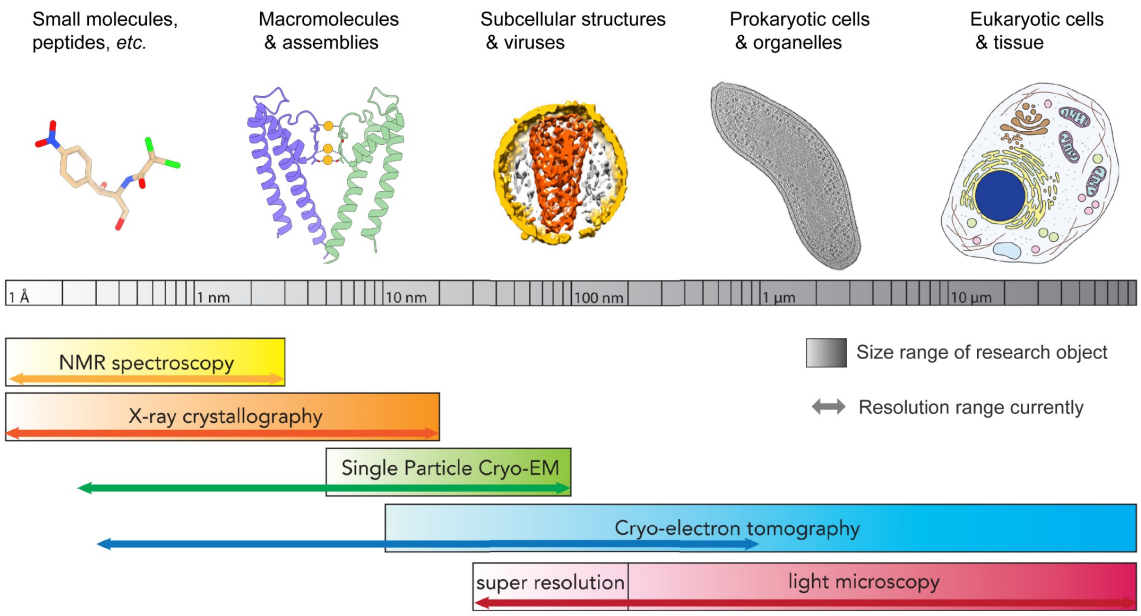

Figure 1.1 | Structural biology across scales.

Biological systems and objects on the microscopic scale range from about one nanometer (one amino acid) to tens of micrometers (a eukaryotic cell). Traditional structural approaches, such as X-ray crystallography and single particle cryo-EM usually, can achieve resolutions towards the atomic level, but are limited by size of the research object or special specimen preparation requirement. Traditional techniques in cell biology, such as light microscopy and super-resolution microscopy, are usually limited to spatial resolutions of several to tens of nanometers. Cryo-ET is an ideal approach to image samples of a large size range (large complexes, subcellular structures, intact cells, *etc*.) and holds the potential to bring high-resolution structural information into the cell, bridging traditional structural biology and cell biology. Left to right: chloramphenicol, KcsA ion channel (PDB 2QTO), HIV virus, a bacterial cell[17], a eukaryotic cell. Modified from the reference[18].

For structural biology, there are three critical considerations: resolution, dynamics and preservation of the native states. Different resolutions reveal different levels of information. The full atomic model can be built with resolutions better than 4 Å, which links the functions of macromolecules to chemical detail. Macromolecules and the formed complexes often undergo conformational and/or compositional changes when conducting their functions, and to capture the structural dynamics can facilitate mechanistic understanding. To what extent are native structures of biological systems maintained has been a major concern for traditional structural approaches, which require breaking down the native functional context - the cell. Although small proteins or stable complexes are possibly not affected, transient complexes that are maintained by weak interactions or supramolecular assemblies that require the intracellular context may not survive cell lysis. Often, protein engineering, recombinant expression, harsh purification process and specific sample preparation (*e.g.* crystallization) are performed to resolve *in vitro* structures,

causing further concerns of their representativeness of the native states in cells. To determine structures directly inside the cell avoids these concerns and can decipher how macromolecules function in their native functional context.

Between chemical reactions mediated by macromolecules and living activities of cells lies a gap of dimension and perception (Figure 1.1). Collective behaviors of macromolecules in a cell determine its activities, and it is necessary to understand them as highly connected molecular machineries and systems (molecular sociology)[19,20]. Structural biology should not be limited to macromolecules and stable complexes. To structurally appreciate life, it is necessary to extend it to the scale of a cell or higher.

In-cell structural biology empowered by cryo-electron tomography (cryo-ET) holds great potential to bridge the gap (Figure 1.1). Cryo-ET can image a wide range of biological systems across scales, from macromolecular complexes to intact cells or even larger objects like tissues[21-23]. Super-resolution microscopy has been shown to be able to obtain spatial resolutions of several nanometers and reveal overall shape of large assemblies in cells[24]. It, however, needs further improvement to reach the real molecular resolution (~1 nm). And lower resolutions in the axial (z) dimension remains an intrinsic issue[25,26]. NMR can be extended into the cell but is limited to small proteins[7]. Although cryo-ET data alone is limited to resolutions of several nanometers, sub-tomogram averaging or analysis (StA) of repetitive complexes in the cellular tomogram can generate high-resolution structures in the sub-nanometer resolution range or better[20,27,28]. It has the potential to bring atomic detail into the cell, and to provide a glimpse into collective behaviors of molecular machineries inside the cell.

In this thesis, I aim to push the limits of in-cell structural biology, and fulfill the great promise of cryo-ET and StA. The small bacterium *Mycoplasma pneumoniae* is chosen as a simple bacterial cell model. Ribosomes, one of the most ancient and essential molecular machines in all living cells, will be the focus of this thesis. Translation, the fundamental process of protein synthesis, will be probed inside native cells, as well as their changes upon antibiotic perturbations. With these objectives, this thesis is not only to push the resolution that can achieved directly inside cells, but also to reveal the structural dynamics and other unrevealed landscapes behind the translation process. To establish the feasibility and exploit the full potential of integrative in-cell structural biology, the cellular cryo-ET and sub-tomogram analysis results are combined with in-cell crosslinking mass spectrometry (CLMS) data from collaborators to decipher the transcription-translation coupling in *M. pneumoniae* cells.

1.2 *Mycoplasma pneumoniae* as a simple bacterial cell model

Mycoplasma pneumoniae belongs to the bacterial class Mollicutes that are characterized by small genomes and the lack of a cell wall. *M. pneumoniae* is the most common pathogenic mycoplasma species that can cause respiratory infections and atypical pneumonia[29,30]. Due to the genome simplicity, it has been used as a model organism for multi-omics and system-wide studies[31-36]. *M. pneumoniae* is one of the smallest self-replicating cells, in terms of cell and genome size[29]. Usually, one cell is half to several micrometers long and a few hundred nanometers wide (Figure 1.2). The most characterized feature in *M. pneumoniae* cell is a gigantic multi-protein assembly named the attachment organelle, which has been suggested to play important roles in cell adherence, gliding mobility and division[37-41]. The whole genome of *M. pneumoniae* contains approximately 816k base pairs and has simple linear structure[31,41]. According to the most recent annotation[42], it contains 694 open reading frames, 311 noncoding RNAs and 43 conventional RNAs (rRNAs, tRNAs, *etc.*). There are, however, more than 200 proteins in the proteome with unknown functions[36,43]. Many of these are lipoproteins that are highly abundant on the cell surface, and have been suggested to play roles in virulence or host immune response[44,45].

M. pneumoniae cells live a strict parasitic lifestyle. They depend on host cells or the environment to provide nutrients because of the absence of many essential genes[34,46]. The cells can be cultivated in rich medium *in vitro*[29], and tend to grow on surfaces of the culture dish (Figure 1.2B). *M. pneumoniae* cells show slow growth rate and long duplication time >8 hours, *in vivo* or *in vitro*[34,38,47]. Interestingly, it was suggested that the slow duplication rate of *M. pneumoniae* cells is due to the limited protein synthesis capacity, but not metabolic efficiency[34,46,48].

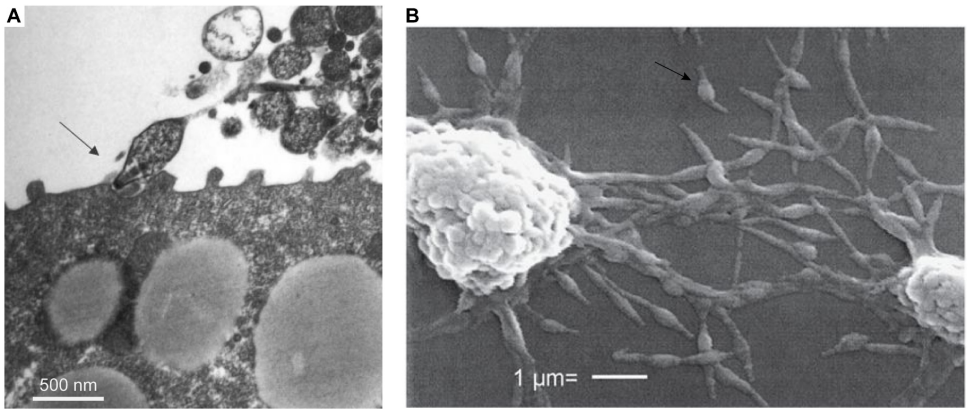

Figure 1.2 | *Mycoplasma pneumoniae* cells *in vivo* and *in vitro*.

(**A**) Attachment of a *M. pneumoniae* cell to the surface of epithelial cells, via its attachment organelle (arrow). The picture is a transmission electron microscopy (TEM) image of the section of the *M. pneumoniae*-infected hamster tracheal ring, which is modified from the reference[29]. (**B**) Scanning EM picture of *M. pneumoniae* cells cultivated *in vitro*. The cells tend to form large cell clusters or colonies, while in between there are individual cells (arrow). Modified from the reference[49].

The transcription and translation machineries, including various accessory factors, are retained in *M. pneumoniae* despite of the severe genome reduction[31,43]. Although with a simple genome, it was shown that *M. pneumoniae* has unanticipated complexity in transcription regulation and metabolic adaptability in response to environmental changes[33,34]. *M. pneumoniae* does not have a defined nucleoid region. Yet, it was found that the chromosome has defined structure and DNA supercoiling regulates its local organization[41]. Although many mycoplasma species have been used as templates for generating the minimal cell in synthetic biology[50-52], *M. pneumoniae* maintains a level of molecular and cellular complexity that is representative of the most conserved processes in bacterial cells.

The small cell and genome size make *M. pneumoniae* an ideal simple cell model to explore the limits of in-cell structural approaches combining cryo-ET and in-cell CLMS. Intact *M. pneumoniae* cells are electron transparent under cryo-EM, without the need for additional thinning process[32,39,53]. Previous cryo-ET studies[32,39,53] of *M. pneumoniae*, however, were limited to morphological descriptions or structures at low resolutions (>25 Å). The comprehensive multi-omics data available for *M. pneumoniae* provide valuable information to facilitate structural investigations. The small proteome also makes *M. pneumoniae* suitable for whole-cell CLMS analysis.

1.3 Translation and the ribosome

Central to all living organisms is translation of the genetic information via messenger RNAs into proteins, which is catalyzed by the translation machinery comprised of ribosomes and translation factors[54]. Ribosomes are among the most primordial molecular machines, and to understand how the translation machinery works is key to understand the origin and evolution of life[55-57].

1.3.1 Structural investigation of ribosomes

Structural determination of ribosomes, to some extent, reflects the development of structural biology, especially cryo-EM. In the very early electron microscopy studies of bacterial cell lysate, ribosomes were imaged before they are characterized[58-60]. With gradient centrifugation and negative staining techniques, images of isolated ribosome components such as the two ribosomal subunits and the fully assembled ribosomes offered the first glimpse of the translation machines[59,60]. The corresponding sedimentation coefficients 30S, 50S and 70S (Svedberg unit) are still used today to indicate the small subunit (SSU), the large subunit (LSU) and the full ribosome in prokaryotes. In 1976, the first 3D representation of the ribosome was made manually based on EM images, which could already show the overall shape and map some ribosomal proteins[61]. Computational image processing and 3D reconstruction techniques were first applied to negative staining EM images[62-64], and later in 1990s to cryo-EM data[65-70]. With the determined

ribosome structures at medium resolutions (usually >20 Å), structural insights can be gained on how ribosomes work, such as mapping of functional regions and different translation factors[66-68], and ratchet-like inter-subunit rotation[70]. In 2000, atomic models of ribosomes were first provided by X-ray crystallography[71-73], following the great success of ribosome crystallization and solving of the phase problem[74]. It marks the peak time of the X-ray crystallography method[5]. Later, X-ray structures revealed atomic detail of the full ribosomes and their functional association with mRNA, tRNAs and other translation factors[75-85]. Around the same time before 2014, the maturing single particle cryo-EM already demonstrated its advantage in capturing translation intermediate structures and reveal the structural dynamics, but the obtained resolutions (usually >6 Å) were not sufficient to derive atomic models[86-94]. Following the so-called resolution revolution[8,95-97], single particle cryo-EM has become the first choice for determining dynamic structures involved in different translation steps[98-101]. Currently, the best resolved ribosome structure is an *Escherichia coli* 70S ribosome at 2 Å resolution by single particle cryo-EM[102], followed by 2.1-2.4 Å resolutions by X-ray crystallography[103,104]. In these structures, atom-level information like rRNA modifications, ions, detailed antibiotic binding are clearly revealed. Recently, time-resolved cryo-EM enabled determination of rare intermediate structures during translation[105-110]. Cryo-EM studies contribute major milestones in the structural studies of ribosomes, and *vice versa*.

Besides the direct structural approaches like X-ray crystallography and single particle cryo-EM, other biophysical approaches such as single-molecule Förster resonance energy transfer (FRET) have been widely used to probe dynamics of ribosomes in real-time translation[111-115]. It should be pointed out that most of the structural and biophysical studies so far are based on isolated or purified ribosomes from model systems such as *E. coli* and *Thermus thermophilus*.

1.3.2 Mechanistic understanding of translation

The structural studies with isolated ribosomes have provided great insights into the translation machine and the translation process[79,113,116]. The ribosome in prokaryotes consists of the small and large subunits (SSU and LSU, or 30S and 50S, respectively), which assemble into the 70S ribosome as the translation machine (Figure 1.3). The small subunit contains one ribosomal RNA (16S rRNA) and about 20 different ribosomal proteins (r-proteins); the large subunit contains two rRNA sequences (23S rRNA and 5S rRNA) and about 30 different r-proteins. The small subunit can be further divided into “head” and “body”, and between them exists the decoding center where the codon-anticodon recognition happens. The large subunit contains the peptidyl transfer center (PTC) and the nascent peptide exist tunnel. Within the interface between subunits, there are three transfer RNA (tRNA) binding sites: the aminoacyl (A) site that accommodates the incoming aminoacyl-tRNA; the peptidyl (P) site that holds the peptidyl-tRNA which is connected to the nascent peptide chain before peptidyl transfer; and the exit (E) site that bears the deacylated tRNA

before it disassociates from the ribosome. Except for these major functional sites, there are other mobile parts such as the L1 stalk that interacts with the E-site tRNA[79] and the L7/L12 stalk that coordinates elongation factors (EFs)[87].

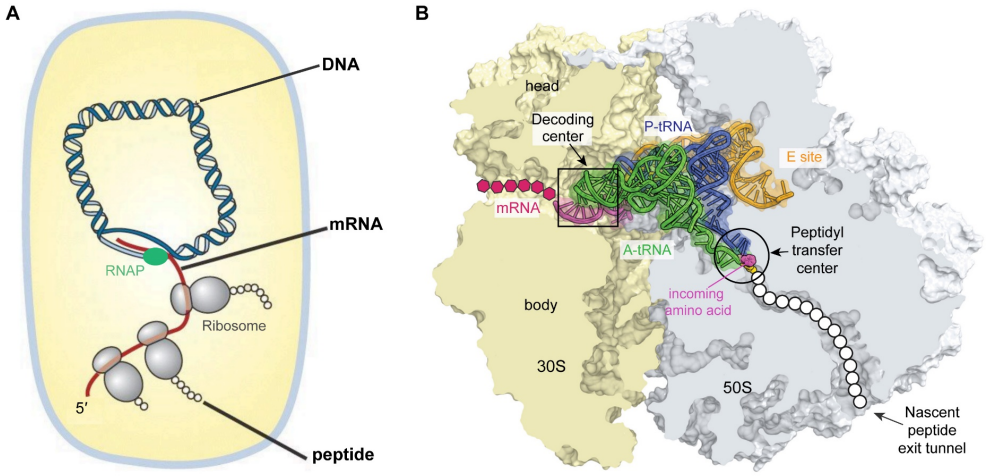

Figure 1.3 | Translation and the ribosome.

(**A**) Schematic of the central dogma of molecular biology in bacteria. Translation of genetic information via mRNA into the nascent peptide in ribosomes is the conserved protein synthesis process in all living cells. Modified from the reference[117]. (**B**) Illustration of the bacterial 70S ribosome structure and major active sites. Modified from the reference[118].

In prokaryotes, the translation process can be divided into initiation, elongation, termination and recycling phases (Figure 1.4). During initiation, 30S first binds to the ribosome-binding site which spans nucleotides -20 to +15 around the start codon, and then the 70S initiation complex forms after recruiting initiator tRNA, initiation factors and 50S. The initiation phase may take a few hundred milliseconds to few seconds before 70S starts elongation[105,113]. Elongation is the major translation phase which can add 10 - 20 new amino acids to the nascent peptide chain per second in bacteria[116,119]. Within one elongation cycle, the ribosome goes through various intermediate states to conduct decoding, peptidyl transfer and translocation to add one new amino acid to the elongating nascent peptide chain[113,116]. In the decoding step, correct anticodon-codon base paring between the cognate EF-Tu•GTP-tRNA ternary complex in the A site and the mRNA, leads to GTP hydrolysis and EF-Tu disassociation[107,116,120]. During initial selection and proofreading steps, there are a series of conformational changes of the 30S subunit, tRNA and EF-Tu within the decoding center to ensure translation fidelity[100,107]. The CCA tail (cytosine-cytosine-adenine sequence at the 3′ end) of the cognate A-site tRNA carrying the new amino acid is then accommodated into the peptidyl transfer center, followed by new peptide bond formation between this amino acid moiety and the nascent peptide chain hold by the P-site tRNA. In the next translocation step, the peptidyl-tRNA in the A site and the deacylated tRNA in the P site need to

move to the P and E sites respectively, and mRNA progresses by one codon to expose the next codon in the emptied A site for binding the next ternary complex. Along with the tRNA and mRNA movement, the ribosome undergoes large conformational changes including ratchet-like rotation between ribosomal subunits, SSU head swiveling relative to SSU body, L1 stalk opening, *etc*. Such conformational changes can happen spontaneously and exhibit many hybrid states from the pre-translocational "A, P" to the post-translocational "P, E" state[115,116,121]. Translocation is catalyzed by the translational GTPase EF-G, which can increase the reaction rate by four orders of magnitude[115]. The catalytic mechanism of EF-G (either as a molecular motor that propels the movement or as a pawl-like mechanism that biases the unidirectional spontaneous translocation) and the exact sequence of different structural rearrangements during translocation are not fully elucidated yet[108-110,115,121]. The elongation cycles of the translating ribosome continue until encountering a stop codon on the mRNA. It then requires additional release factors to terminate translation and release the nascent peptide, and the ribosome recycling factor and EF-G to disassociate the 70S ribosome[106,113,122].

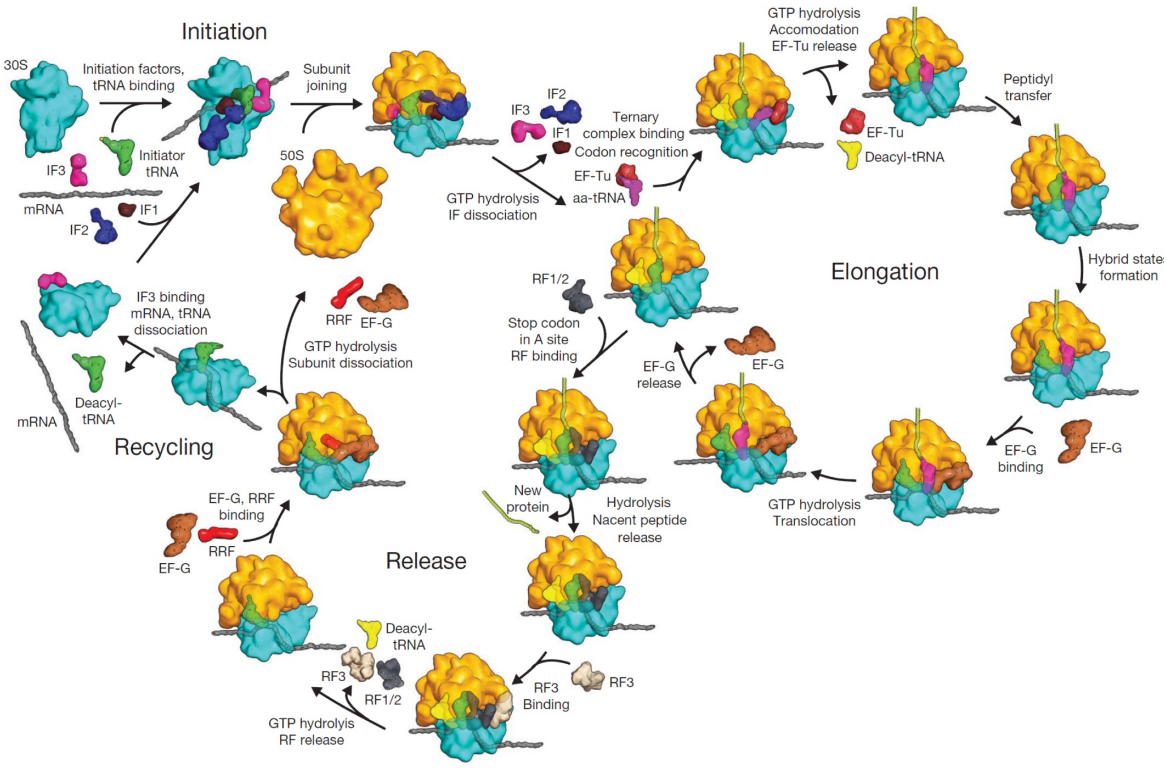

Figure 1.4 | Overview of the translation process in bacteria.

Major intermediate steps in the translation initiation, elongation, release and recycling phases are illustrated. IF, initiation factor; aa-tRNA, aminoacyl-tRNA; EF, elongation factor; RF, release factor. RRF, ribosome recycling factor. Figure adapted from the reference[79].

1.3.3 Structural studies of ribosomes inside the cell

There are certain limitations of *in vitro* approaches that prevent further understanding of the translation machinery in the native functional context. The current translation model is derived from various *in vitro* studies, each of which focused on one or several specific translation states[113,116]. Few studies performed systematic structural investigation of the translation process. The most comprehensive among these is the single particle cryo-EM study of isolated human polysomes, in which eleven translation intermediate structures were resolved[98]. However, in their reconstructed translation elongation cycle there are still several key steps missing. Moreover, *in vitro* structures are often resolved with mutations, antibiotics or GTP analogs to trap ribosomes in specific functional states. Most importantly, after cell lysis the original spatial organization of translation machines and their possible association with other processes are lost after cell lysis. Thus, to gain further mechanistic understanding of translation in living organism, it is vital to resolve structures directly inside cells.

A number of ribosome structures have been solved by cryo-ET and sub-tomogram averaging inside cells, including bacteria[32,123-125], protozoa[126-129], plant[130-132], nematode[133], and human[22,126,134,135]. However, most of these structures are limited to nanometer-level resolutions, prohibiting detailed investigation of the translation process within the native cellular context. Even though, previous cryo-ET works on polysomes, either *in situ*[134] or *in vitro*[136-138], showed that ribosomes have various arrangement patterns and the mRNA and nascent peptides are well arranged within polysomes. Mapping of the whole ribosome population into the cell with cryo-ET and template matching was pioneered by a few studies[32,123,139]. Despite the limited accuracy of the template-based mapping, they showed overall distribution of ribosomes in cells. A special case is the study of ribosomes in isolated mitochondria, which showed additional densities associated to the ribosome that may be contributed by unknown mechanisms[140-143]. Inside the cell, there may be unanticipated diversity and complexity involving ribosomes[144,145], which can be revealed by in-cell structural approaches.

1.4 Transcription-translation coupling

The two fundamental processes of gene expression, transcription and translation, can be spatially and temporally coupled in prokaryotic cells that lack a membrane-enclosed nucleus and the complex RNA splicing mechanisms[146,147]. While RNA polymerase (RNAP) is transcribing DNA into mRNA, ribosomes can load and start translation after exposure of the ribosome-binding site on the nascent mRNA. Transcription is known to be frequently interrupted by pausing and backtracking[148,149], and the predicted pause sequences are enriched in translation start sites[150,151]. It is therefore possible for the trailing ribosome to approach the transcribing RNAP and lead to transcription-translation coupling.

Early evidence of transcription-translation coupling came from observations of the transcription polarity and attenuation effect[152,153], and electron microscopy studies of *E. coli* cell lysate showing close proximity between DNA transcription site and ribosomes (Figure 1.5A)[154,155]. When translation is stalled or slowed down, it allows the Rho termination factor to bind to the exposed mRNA and lead to premature transcription termination of downstream genes[152,156]. Proshkin *et al*[157] found that the transcription elongation rate matches exactly the translation elongation rate in *E. coli* under different conditions, and the trailing ribosome strictly controls elongation rates of RNAP. They and following studies[157-159] suggested that the translating ribosome can stop RNAP from pausing or backtracking when they are coupled.

NusG, a conserved transcription factor[160,161], was proposed to link RNAP and the ribosome by Burmann *et al*[162] (Figure 1.5B). NusG binds to RNAP via its N-terminal domain (NTD), while its C-terminal domain (CTD) was found to interact with NusE, a protein shared by both transcription and translation machineries and is identical to the ribosomal protein S10. The association of NusG to the assembled 70S ribosome was revealed in later studies[163,164]. However, there was no reported structure of the possible RNAP-NusG-ribosome assembly until two recent publications in 2020[165,166]. Detailed introduction of the two studies is available in section 6.3 in Chapter 6. The interaction between NusG CTD and NusE/S10 has a K_D of 50 μM in *E. coli* [162], leaving questions on how NusG is able to tether the two gigantic gene expression machines during the dynamic elongation process. It should be noted that NusG is among many transcription factors that bind to RNAP and regulate transcription[148,167]. Both NusG and NusA are general factors that can continually associate with RNAP throughout transcription elongation[168]. The NusG-linking model is based solely on *E.coli* and *in vitro* results, and requires further validation.

Direct interaction between RNAP and the ribosome has been suggested by several *in vitro* studies and was suggested to contribute to transcription-translation coupling[158,169-171]. Among these is the cryo-EM structure of *E. coli* RNAP-ribosome supercomplex (termed "expressome") determined by Kohler *et al*[169]. They designed a special DNA-RNA hybrid to mimic the transcription bubble, and then incubated it with RNAPs, ribosomes and other translation factors to induce coupling. In the determined expressome structure, the RNAP mRNA exit regions form close contact with the ribosome mRNA entry regions, making a direct mRNA path from the transcription active center to the translation active center. Binding of NusG to RNAP is possible in the structure but it does not support the NusG-S10 interaction model. Binding of NusA is impossible in their model because of a predicted clash with the small ribosomal subunit. The *in vitro* expressome structure was the first to illuminate the transcription-translation coupling complex consisting of DNA, RNAP, mRNA and the ribosome.

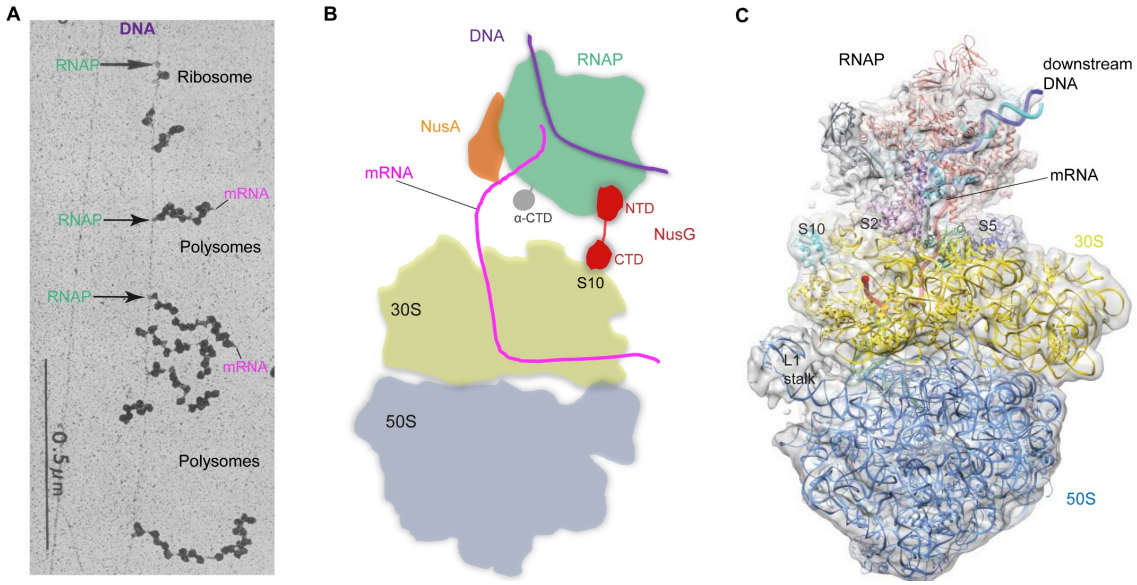

Figure 1.5 | Landmarks in transcription-translation coupling research.

(**A**) Visualization of gene expression in *E. coli* cell lysate under TEM with Miller's spreading technique. EM picture modified from the reference[155]. (**B**) The proposed NusG-linking model for transcription-translation coupling. NusA, a general transcription factor that binds to RNAP near the mRNA exit site, is also displayed, as well as one CTD of RNAP α subunits. Figure inspired by the reference[162]. (**C**) The first RNAP-ribosome supercomplex (named expressome by the author) structure obtained through *in vitro* reconstitution and single particle cryo-EM. Modified from the publication by Kohler *et al*[169].

The NusG-linking model and the direct interaction model were the main proposed models for transcription-translation coupling before this thesis was started. I will adopt the term "expressome" to refer to supercomplexes formed by RNAP, the ribosome and other possible auxiliary factors. The previous studies on transcription-translation coupling are based *in vitro* experiments, where temporospatial organization and completeness of potential auxiliary factors are absent[147,172]. Whether or how transcription and translation are arranged and coupled inside bacterial cells needs to be further investigated.

To be more precise, in this thesis, transcription-translation coupling only refers to physical or structural coupling arising from close proximity between RNAP and the ribosome. It is noted that, apart from the physical coupling, functional coupling or coordination between transcription and translation exist in cells[147,173]. For example, in *E. coli,* the alarmone (p)ppGpp involved in stringent response was suggested to play an important role in coordinating transcription and translation activities[174-177]. Functional coupling/coordination in cells may play important roles in creating the environment for the occurrence of physical coupling.

1.5 Cryo-ET and sub-tomogram analysis

Cryo-electron tomography emerges as the most promising approach to bring high-resolution structures into the cell[20,23,178-180]. Cryo-ET is suitable to image biological systems across scales, and the following sub-tomogram analysis enables obtaining high-resolution structures of repetitive complexes in the native functional context.

1.5.1 Principles of electron tomography

Electron tomography[181] shares the same 3D reconstruction theory[182] to single particle analysis, that projections of an object from different directions under the transmission electron microscope can be combined to derive a 3D reconstruction of the object (Figure 1.6). In practice, in electron tomography a series of projection images (tilt-series) from different directions is collected by tilting the specimen, which differs from single particle analysis that relies on a large number of identical particles with different orientations in the untitled specimen[183]. Macromolecules and complexes can be isolated and purified to perform single particle cryo-EM. For biological systems on larger scale, such as unique subcellular structures or intact cells, electron tomography is more suitable for obtaining their 3D reconstructions[22].

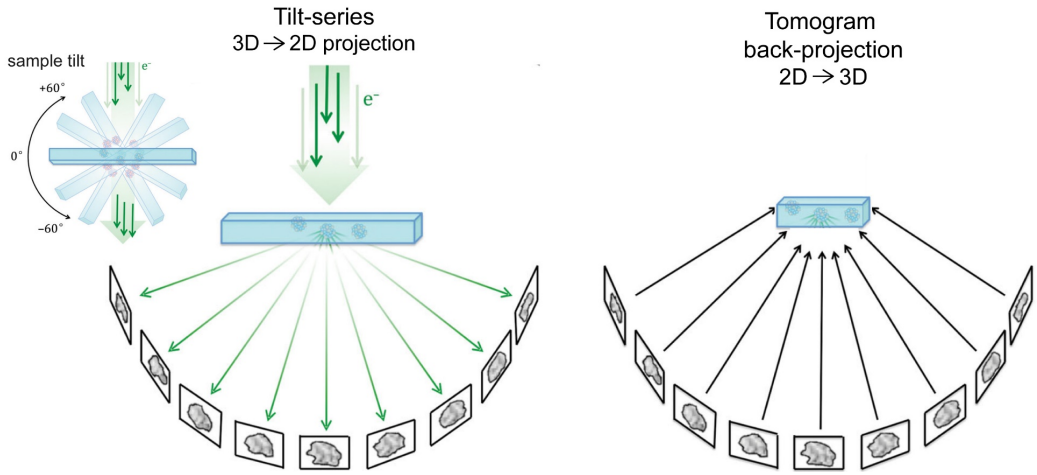

Figure 1.6 | Principles of electron tomography.

For electron tomography, the specimen is tilted under a transmission electron microscope, and a series of 2D projection images (tilt-series) of the 3D object (more specifically, the electrostatic potential of the specimen) is recorded. 2D images are combined to reconstruct a 3D volume (tomogram) representing the density distribution of the original 3D object, by back-projection or other reconstruction algorithms. Modified from the reference[184].

For cryo-electron tomography, projection images of the frozen-hydrated specimen under a cryogenic transmission electron microscope are acquired. As biological systems mostly consist of light atoms (C, H, O, N) similar to the surrounding water molecules, the projection images mainly have phase contrast arising from interference between elastically scattered and unscattered

electrons[183]. Larger specimen thickness causes more inelastic scattering that increases the noise and lowers the contrast. Spherical aberration of the microscope and out-of-focus imaging (typically defocus) can induce additional phase difference between the scattered and unscattered electrons, and in practice are used to increase contrast. Thus, the acquired projection image is modulated by the contrast transfer function (CTF) oscillating between -1 and 1 in the frequency domain. For a given microscope, CTFs are determined by the employed defocus in imaging. To faithfully reconstruct the object (specimen), CTF correction needs to be performed to demodulate the signal by the estimated corresponding CTF[27,183]. Another aspect especially important for cryo-ET is that the inelastic scattered electrons can ionize atoms in the specimen and cause radiation damage, which can result in beam-induced motion, specimen deformation, and other consequences to damage the high-resolution information in the projection images[183,185,186].

1.5.2 Cryo-ET sample preparation and data collection

Vitrification preserves biological samples in their close-to-native states, where water molecules form vitreous ice and other molecules are frozen in the hydrated state[187-189]. The plunge freezing method developed by Dubochet *et al*[187,188] make cryo-EM/ET possible, and it has been the routine procedure for specimen preparation. Usually, the grid with solution and/or cells is blotted with filter paper to only keep a thin solution layer, before plunging into liquid ethane or propane/ethane mixture[190] cooled by liquid nitrogen. For samples thinner than 1 μm, the vitrification time is about 0.1 ms and the freezing is fast enough to ensure amorphous ice[188]. For thicker samples like human cells, it has been shown that normal plunge freezing failed to fully vitrify cellular regions 5 μm below the surface[180,191]. Biological samples usually contain different components such as lipids, macromolecules, small molecules and ions, which make freezing of the cell different from pure water or simple solutions[189]. The penetration depth of vitrification with plunge freezing may vary depending on the sample, and needs to be carefully checked case by case. To achieve full vitrification for large eukaryotic cells, using cryoprotectants like glycerol has been suggested to be helpful[180,192]. There are, however, concerns on whether cryoprotectants influence the native states of living cells. High-pressure freezing has been used to vitrify specimens up to 200 μm thick, but whether the freezing rate is uniformly fast enough remains a major question[193,194].

The specimen thickness is one major factor to be considered for cryo-ET sample preparation. Most cells are too thick to be directly imaged under cryo-EM. In practice, <500 nm is typically needed for cryo-ET data collection, while 300 nm or thinner thickness is necessary to achieve good contrast and high-resolution sub-tomogram averaging[23,180]. Cryo-sectioning has been invented for a long time, but the induced artifacts and other technical issues hinders it from being a routine method[195]. Cryogenic focused ion beam (cryo-FIB) thinning technique provides an alternative method to thin samples to the thickness suitable for cryo-ET, especially with recent automation

developments[196-198]. The lift-out technique associated with cryo-FIB make it possible to handle even thicker samples like tissues and nematode worms[133,199-201].

To collect tilt-series, the specimen is usually tilted between -60° and +60° because of limitations of the microscope stage and the increased apparent thickness (Figure 1.7). The total electron dose applied to the biological samples is typically limited to 100-150 e^-/A^2 and it should be distributed over the tilt-series[183]. The dose for each tilt image is further fractionated to frames if a direct detection decide (DDD) camera is used. The signal-to-noise ratio (SNR) is therefore extremely low, and even lower for higher tilt angles due to the increased apparent thickness (Figure 1.7). With the cumulative radiation, high-resolution information deteriorate faster[185,202]. There may also be considerable deformations of the specimen because of the accumulated radiation damage. Thus, the data collection strategy should be decided based on the sample and purpose. For high-resolution structure determination, the dose-symmetric scheme has been demonstrated to give the best results[203,204], which maximizes the high-resolution information at lower tilt angles and lower dose that can be used for averaging. Additional developments in cryo-EM, such as microscopes with stable stages and zero-loss energy filters, DDD and frame motion correction[205,206], dose weighting[202] should be considered. Phase plate induces a phase shift between the scattered and unscattered electrons, and can boost contrast in the images[183]. The Volta phase plate (VPP)[207-209] has been widely used in cryo-ET data collection. However, due to issues such as charging and unstable phase shift, currently VPP is not recommended for high-resolution sub-tomogram averaging[204]. The laser phase plate[210] represents a promising direction in phase plate design.

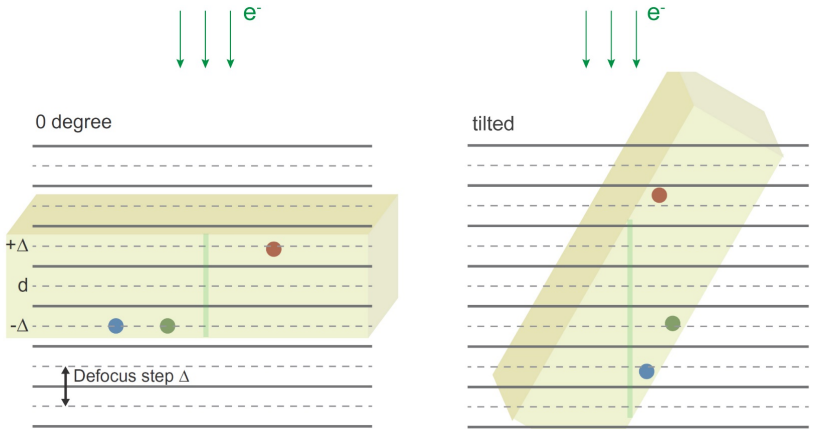

Figure 1.7 | Thickness and CTF issues due to specimen tilting.

The apparent thickness that electrons need to pass through (light green line) is increased for the tilted specimen. The number is doubled at ±60° tilt degree of the specimen. During tilt-series data collection, there will be a defocus gradient perpendicular to the tilting axis. In addition, different regions (*e.g.* the colored dots) within the relatively thick specimen will have additional defocus difference. Both cause different defocus values for different regions in the recorded 2D projection images. Modified from the reference[211].

1.5.3 Cryo-ET data processing

Tilt-series need to be accurately aligned before generating the 3D reconstruction and the alignment accuracy determines the quality of the tomogram (Figure 1.8). For dose-fractionated data collected with a DDD, frames for each tilt images should be aligned first, similar to the motion correction in single particle cryo-EM. Because of the low SNR, high-contrast markers like gold fiducials are usually applied to the specimen and used for tracking during tilt-series collection and alignment whenever possible. The most commonly used software for tilt-series alignment is Etomo in the IMOD package[212], which can use either gold fiducials or local regions (patches) for tracking, and is employed by most cryo-ET processing packages for initial tilt-series alignment.

The reconstructed tomogram is usually limited to 2-5 nm resolution because of low SNR and limited number of images in the tilt-series[22]. The limited tilt range causes the "missing-wedge effect", *i.e.* anisotropy of the tomogram along the beam direction. Moreover, as the projection images are modulated by CTFs, accurate defocus estimation and CTF correction are crucial to restore the structural information[27]. Due to specimen tilting and the influence of thickness, different regions in the recorded images will have different defocus values (Figure 1.7). 3D CTF correction that takes the specimen tilting and objects of different positions in the thick specimen into consideration is necessary to high-resolution sub-tomogram averaging[27,211].

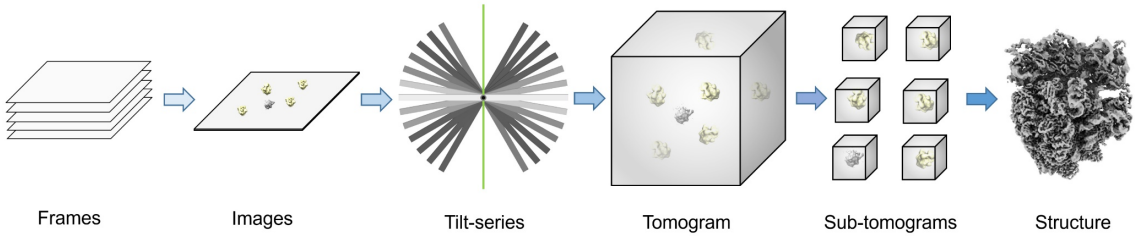

Figure 1.8 | Schematic of the cryo-ET and sub-tomogram analysis process.

First, for each tilt angle, dose-fractionated frames from DDD camera are aligned and summed to generate the 2D image. CTF (defocus) estimation is also performed in this step. The tilt-series are projection images of the specimen from different angles. The dose-symmetric scheme is illustrated, with grey colors (from lighter to darker) indicating the accumulated dose. With the aligned tilt-series, a 3D volume (tomogram) is reconstructed, which contains targets (*e.g.* ribosomes) that can be extracted as 3D sub-volumes (*i.e.* sub-tomograms). Sub-tomograms can then be aligned to generate high-resolution averages of the repetitive complexes. 3D classification of sub-tomograms is often needed to deal with structural heterogeneity. In the process, there are three alignment steps that determine the processing quality: frame alignment, tilt-series alignment and sub-tomogram alignment. Made with inspirations from the references[18,184].

Cryo-ET can be used for different purposes, and different data collection and processing strategies should be chosen dependently. If visualization is the main purpose, tilt schemes giving more images (*e.g.* larger range and smaller increment), higher defocus and Volta phase plate can be

considered, as well as tomogram denoising and "missing-wedge" restoration algorithms[213-215]. For high-resolution sub-tomogram averaging, the dose-symmetric scheme, relatively small defocus, and no phase plate are currently the best practice[204].

1.5.4 Sub-tomogram analysis

Repetitive complexes within the tomogram (*e.g.* ribosomes) can be extracted as sub-tomograms, and then averaged to increase SNR and resolution. This is typically termed sub-tomogram averaging. Here, I use the term sub-tomogram analysis (StA) to refer to the comprehensive analysis on the sub-tomogram level following tomogram reconstruction, which includes particle localization, averaging, classification, spatial analysis, *etc*. A summary of the major cryo-ET and StA workflows can be found in Table 1.1.

To localize the targets within the tomogram is the primary, yet the most challenging step. Currently, there is no gold standard method. Manual picking in the original tomogram of extremely low SNR is difficult and often requires additional filtering or denoising to boost the contrast. It is time-consuming, experience-dependent, and prone to be biased towards particles or orientations of particles that have higher contrast. Usually, manual picking is used to generate the initial template or training dataset for other approaches. For targets of certain symmetric shapes (*e.g.* filaments) or of special distributions (e.*g.* membrane-associated particles), geometry-based picking[216,217] can be helpful. Template matching (TM) has been demonstrated to be successful and efficient in detecting large complexes such as ribosomes and proteasomes[123,218,219]. In practice, the structure of the target, either from previous studies or from averages of the manually picked particles, is lowpass filtered and used as the template to search the tomogram. Cross-correlation between the template and the sub-regions of the tomogram are calculated and highest scoring regions indicate potential targets. Template marching is dominated by low-resolution features and prone to have false positives[27,218]. To detect smaller complexes, one needs to do template matching iteratively and mask out big complexes and other impossible regions (such as membrane) first[220]. Particle localization methods based on deep learning, especially convolutional neural network, have been developed and implemented recently[221-223]. However, to have a large and clean enough training data is still a bottleneck. Generalizability is another issue, meaning that model trained with one dataset cannot be directly used for other datasets. Unsupervised learning and self-learning methods that do not rely on training data represent the future direction[224].

Sub-tomograms (3D sub-volumes) containing the picked complexes are either extracted from the tomogram or reconstructed locally from the tilt-series[225]. The sub-tomograms are then subjected to classification and refinement (averaging). Sub-tomogram refinement iteratively calculates cross-correlations between sub-tomograms and the reference, and determines the rotations and shifts for each of the sub-tomograms[27]. The calculation is usually performed in Fourier space, and

a compensation for the "missing-wedge" is necessary to avoid alignment bias[27]. Classification of sub-tomograms is especially important for in-cell data with high structural heterogeneity. Different classification algorithms have been implemented in most cryo-ET processing packages (Table 1.1). RELION enables maximum-likelihood classification and provides the comprehensive toolbox. Handling of sub-tomograms has been incorporated into the RELION pipeline, with the introduction of 3D CTF and missing-wedge models[226,227]. The post-processing step that includes resolution estimation and density sharpening is performed after sub-tomogram refinement. Fourier shell correlation (FSC) between density maps refined independently from two randomly split half datasets is calculated to estimate the final resolution. Density sharpening is performed to restore the high-resolution features, and the related methods that were initially developed for single particle analysis can be used for sub-tomogram analysis[27,225].

Table 1.1 | Comparison of major cryo-ET processing software.

Software/ packages	IMOD /PEET	TOM /AV3	PyTom	Dynamo	EMAN2	RELION	emClarity	Warp/M
GUI	✓		✓	✓	✓	✓		✓
Raw image processing*	✓				✓			✓
Tilt-series alignment	Fiducial Patch	with Etomo	with Etomo	with Etomo	Auto	with Etomo	with Etomo	with Etomo
Particle picking			Template matching	Manual geometry	Manual semi-auto			Template matching
CTF correction	2D	**	2D	2D	3D	Per-particle	3D	3D
Missing-wedge representation	Binary wedge	Binary wedge	Binary wedge	Binary wedge	Fourier intensity threshold	3D CTF models	3D sampling function	3D CTF models
Sub-tomogram averaging	✓	✓	✓	✓	✓	✓	✓	✓
Average-based tilt-series refine					Subtilt refine		✓	✓
Average-based CTF refine								✓
Classification			#Multiple correlation optimization	PCA MRA	Multi-reference alignment	Maximum likelihood	Multi-scale PCA & clustering	With RELION

*including motion correction and CTF estimation. **3D-CTF correction is possible if combining with NovaCTF[211]. PCA, principal component analysis. MRA, multi-reference analysis. #similar to multi-reference alignment. For tilt-series alignment, most packages use results from IMOD/Etomo[212]. For the similar function or process, different software may have different implementations. Details and updates for the listed tools (IMOD/PEET[212,228], TOM/AV3[229,230], PyTom[218], EMAN2[231], Dynamo[216,217], RELION[226,227], emClarity[232], Warp[233]/M[234]) can be found on their web pages. The summary is based on updates until December 2021.

Among the available StA packages, the Warp-RELION-M workflow unifies major processing steps and provides comprehensive functions (Figure 1.9). Warp[233] was developed for real-time image processing in cryo-EM and can do the initial processing, such as raw frame alignment and CTF estimation. The dose-weighted tilt-series generated by Warp are aligned in Etomo and the alignment files are read by Warp to reconstruct tomograms. Warp provides TM-based method for

particle localization, or imports coordinate files obtained from other approaches. Sub-tomograms and per-particle 3D CTF models (combining CTF, missing-wedge and dose weighting information) generated by Warp are subjected to RELION for refinement and classification. M[234] can read the refinement results from RELION, and then perform average-based tilt-series refinement, simultaneously with CTF parameters and particle poses. The improved tilt-series and CTF parameters after M refinement can be utilized by Warp to re-extract better sub-sub-tomograms.

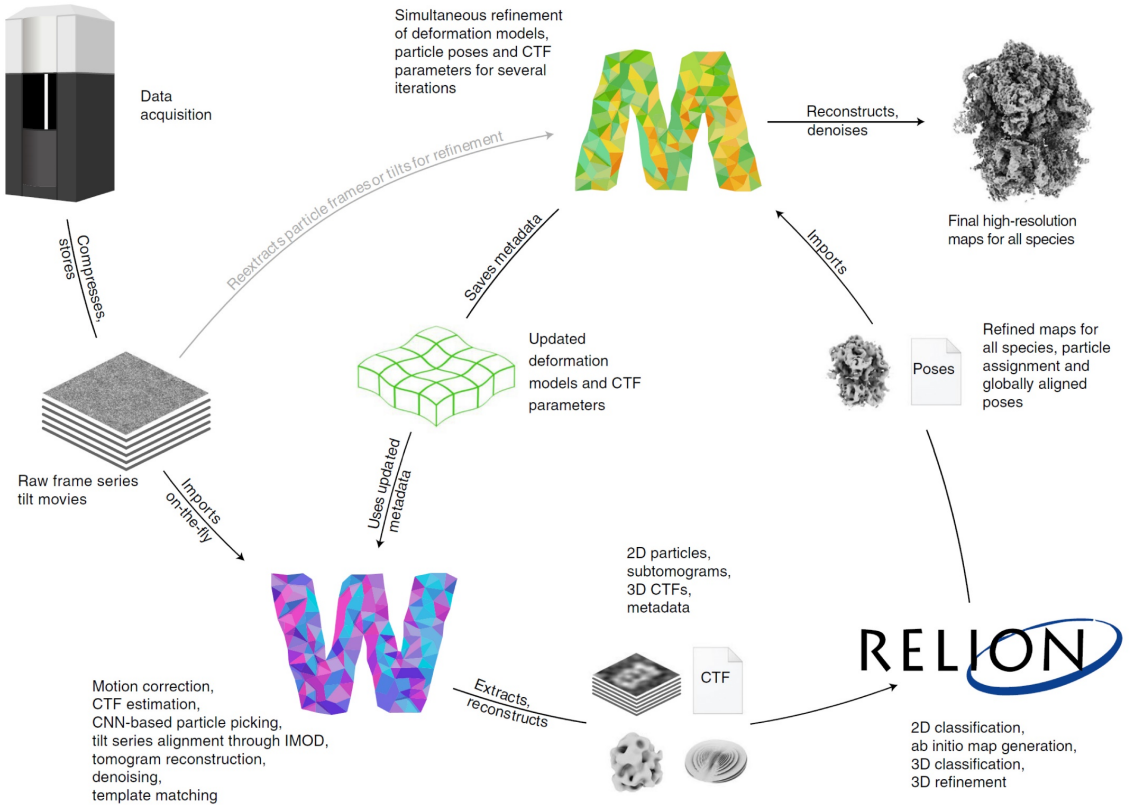

Figure 1.9 | Warp-RELION-M data processing workflow.

The cryo-EM/ET data processing workflow consists of Warp, RELION and the multi-particle refinement software M. The figure is made by Dimitry Tegunov for the collaborative publication[234].

Spatial analysis is aimed to obtain a comprehensive and quantitative description of the 3D distribution of analyzed macromolecules and complexes in the cell, beyond the visualization purpose. Mapping the determined structures back into the 3D cellular volume can be performed based on the coordinates and rotations determined for all particles during particle picking and structure refinement. However, most available StA packages do not include this function, and the mapping is usually done with TOM toolbox[229] or home-made scripts. Further analysis including overall distribution, local distribution, position and orientation relationships, functional organizations, *etc.* can be carried out to examine how macromolecules and complexes are organized in cells. Such analysis performed in several works[128,220,235-237] have yielded intriguing insights into molecular organizations in cells.

1.6 Integrative in-cell structural approaches

1.6.1 In-cell crosslinking mass spectrometry

Crosslinking mass spectrometry (CLMS) is a powerful method for studying protein-protein interactions (PPIs) and for structural investigations[238-242]. For CLMS, biological samples are first treated with the crosslinking reagent (crosslinker) that contains a flexible spacer and two end reactive groups that can form covalent bonds with specific amino acid residues (Figure 1.10). If two such residues are close enough in space (*e.g.* within the same complex), they are possible to be crosslinked by the same crosslinker molecule. Mass spectrometry and database search are then performed to detect the crosslinked residue-residue pairs (crosslinks). The length of the spacer (usually 10 - 15 Å) determines the maximum distance that may be probed. Hence, CLMS not only reveals PPIs with residue-level resolution, but also provides distance restraints between specified residues that can produce structural or topological information. Indeed, CLMS has been widely used to obtain complementary information for the modeling of challenging complexes such as the nuclear pore complex[243-245].

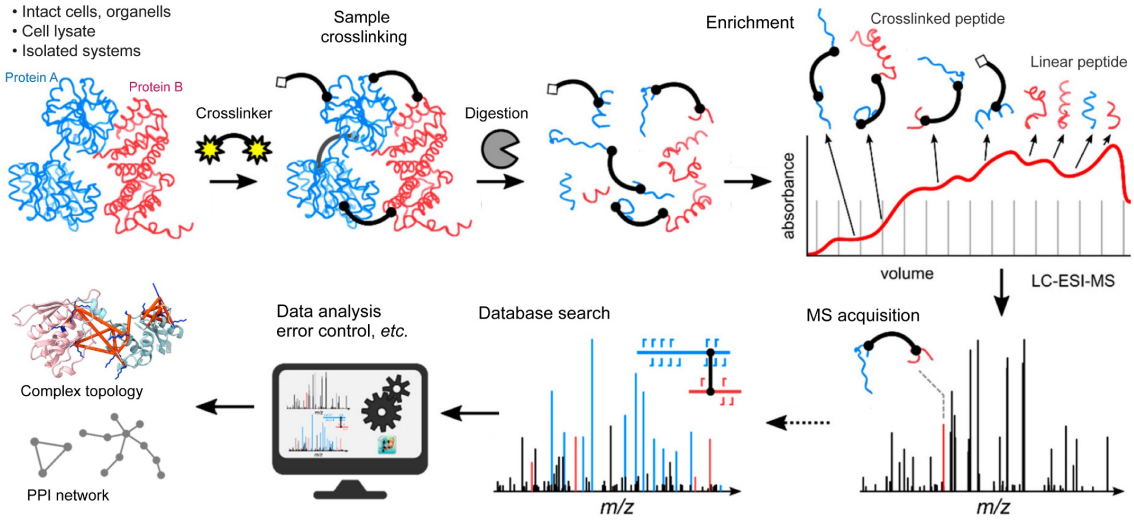

Figure 1.10 | General workflow for crosslinking mass spectrometry.

The samples for CLMS in principal can be any biological sample, ranging from isolated systems to intact cells. Here only peptide crosslinking workflow is presented, although peptide-nucleic acid crosslinking methods have been reported[246]. Different crosslinkers can be chosen or designed based on the sample and purpose of the study, while the data-driven approach may be preferred for in-cell CLMS[247]. For crosslinking whole cells, membrane-permeable crosslinkers can be used. After crosslinking, the following steps of protease digestion, peptide enrichment, MS acquisition and data analysis are performed. For in-cell CLMS, a scheme consists of different chromatographic approaches may be needed to enrich the crosslinked peptides. After database search, a list of specific residue pairs (with sequence position information) is obtained and can be visualized as PPI networks or topology maps. Modified from the references[241,248].

The high sensitivity and throughput nature of mass spectrometers makes CLMS possible for system-wide studies. In-cell CLMS emerges as a promising method that can systematically capture protein-protein interactions inside cells. Recently, it has been successfully applied to cell lysate[249-251], isolated organelles[252,253], intact cells[254-256] and even tissues[257]. Membrane-permeable crosslinkers such as disuccinimidyl suberate (DSS) and disuccinimidyl sulfoxide (DSSO) can enter the cell, and crosslink reactive residues (*e.g.* lysine, for most available crosslinkers) that are spatially close[241]. Proteins that form complexes or assemblies to carry out specific cellular processes can be crosslinked. In-cell CLMS does not break down the cell, giving it the advantage over other MS methods such as affinity-purification MS. What is noteworthy is that how crosslinkers diffuse in cells and how living cells are influenced by crosslinkers remain to be further elucidated[241,258]. Optimization of different experimental conditions should be performed to balance sufficient crosslinking and representation of the determined crosslinks. Error assessment and control of the identified crosslinks is necessary to confidently utilize crosslinks for data interpretation and model building, especially for in-cell CLMS[251,259,260]. The crosslinks yield valuable compositional and topological information for interpreting structures from cellular cryo-ET and sub-tomogram analysis.

Besides in-cell CLMS, many other mass spectrometry applications can be combined with cellular cryo-ET. Quantitative MS can provide basic information (*e.g.* abundance and stoichiometry) of proteins in cells and has been explored for the effort to achieve molecular sociology of the cell[20,139,261]. Quantitative CLMS can be used to reveal relative abundances of crosslinks in different interactions or conformations[241,262]. In cells, such quantification will be extremely valuable to decipher dynamics of protein interactions and facilitate integrative modeling. Other developments, such as native MS[239,263,264], single-cell MS[265], will establish MS as a major player toward system-wide biological research.

1.6.2 Additional methods for in-cell structural biology

Besides cryo-ET and in-cell CLMS, there are a number of methods than can facilitate structural investigation in cells. Fluorescence microscopy, especially super-resolution fluorescence microscopy, is currently the main imaging method for cell biology[25,26]. Fluorescent labelling enables high molecular specificity, and the low invasiveness of light allows live-cell imaging. Multi-color imaging opens the possibility to visualize protein co-localization and interaction in cells. Although in theory there is no resolution limit for super-resolution microscopy, practically the best spatial resolutions achieved by different applications are typically limited to ~10 nm[26]. Recently the MINFLUX[24] nanoscopy have attained 1-3 nm resolution in fixed and living cells, approaching the molecular resolution of ~1 nm. These resolutions enable revealing overall spatial

organization of macromolecules and in rare cases shape of large assemblies such as nuclear pore complexes and centrioles[24,26].

Correlative light and electron microscopy (CLEM) combines the advantages molecular specificity by fluorescent labelling and high-resolution structural details by electron microscopy, especially under cryogenic condition (cryo-CLEM)[266]. The fluorescent signal can be used to localize cells or subcellular regions of interest, guide cryo-FIB lamella preparation and cryo-ET data collection. The specific florescent labelling can also provide additional information to help identify subcellular structures or regions in the reconstructed tomogram[180]. For such purpose, super-resolution microscopy under cryo-temperature is required to be combined with cryo-ET. Recent developments showed the enormous potential although many technical issues need to be overcome[266-270].

FRET can be used as a "spectroscopic ruler" to measures distances and distance changes on the nanometer scale. In-cell single-molecule FRET has been developed to gain structural and quantitative information about protein dynamics in cells[271], In-cell NMR[7] and in-cell EPR (electron paramagnetic resonance)[272] can also reveal structures and dynamics of proteins in cells. They are suitable for probing structural and conformational dynamics of small proteins, but some intrinsic limitations hinder them from being general structural approaches[179]. A recent work explored combining solid-state NMR and cryo-ET to study membrane protein in the native membrane environment[273]. These approaches can generate additional contextual, structural and dynamic information for integrative in-cell structural biology.

1.6.3 Integrative modeling

Modeling is a key step to interpret results obtained from the experimental approaches. Broadly speaking, all modeling is an integrative method that endeavors to combine the available experimental (like cryo-EM/ET density maps) and existing information (like the protein sequence). Especially, to interpret cryo-ET maps often of large complexes and of relatively low resolutions (10-30 Å), integrative modeling that combines the obtained density map, available atomic models of the components of the resolved complex, and restraints from other experimental approaches is necessary[274,275].

Modeling strategies depend on resolutions. There is a variety of choices in the modeling toolbox that can be utilized to build models into cryo-EM/ET densities[274]. Most tools are consolidated in platforms such as PHENIX[276] and CCP-EM[277]. For maps with better than 4 Å resolution, atomic models can be built directly into the density, with backbones and side chains traced. For cryo-EM/ET maps of 4-10 Å resolutions, pseudo-atomic models can be derived by flexible fitting of available atomic models. Secondary structures are mostly visible in this resolution range, and

density interpretation is unambiguous. Complementary information such as crosslinks is useful to assist flexible fitting and validate the model. For maps of relatively low resolutions from ~10 to 30 Å, usually only domain or subunit density is resolved, and fitting other models into the density can have some ambiguity. To confidently interpret the density, additional information from different approaches such as CLMS is required to perform integrative modeling. Especially for in-cell or system-wide structural studies, to ambiguously determine component(s) within the obtained density is even challenging due to compositional diversity and uncertainty.

Integrative modeling can be performed using programs such as IMP[278], HADDOCK[279,280] and Rosetta[281]. Typically, there are three major information sources for integrative modeling: the cryo-EM/ET density map as volume restraints, atomic models for components of the complex, and crosslinks from CLMS as distance restraints[275,282]. Restraints and information from other approaches such as NMR and small-angle X-ray scattering can also be included. Atomic models of the components can be downloaded from the PDB database if they are experimentally solved previously. For proteins without reported models, there are a few ways to generate the starting models. Homology modeling used to be a widely used approach, but it highly depends on the template. Deep learning based structure prediction may provide a much better option[14,283], and proteome-wide prediction[16] has been made possible. Although the predicted structures have been shown to have high accuracy, they may still need experimental validation before being used for or after integrative modeling[284]. Leveraging deep leaning methods to systematically investigate PPI networks and assisting complex modeling represents another promising direction[285-287], especially when combining with systematic data from in-cell CLMS and cryo-ET.

Available experimental data is typically represented as different restraints in the integrative modeling program[275]. Atomic models can be coarse-grained for computational efficiency. Thorough structural sampling (*e.g.* Monte Carlo method) is then carried out, and the scoring function is defined to evaluate how the sampled models meet the given restraints. Lastly, model assessment and validation are performed to estimate the uncertainty in the integrative models. Often, the integrative model presented in the publication may only represent the highest-scoring one among many possible modeling solutions. Recently, there is consensus in the field that all the relevant experimental data, modeling and validation results need to be deposited in public databases such as EMPIAR and PDB-Dev[288].

Chapter 2: Materials and Methods

2.1 Cryo-ET sample preparation and data collection

2.1.1 *Mycoplasma pneumoniae* cell cultivation

The wild-type *M. pneumoniae* strain M129 (ATCC 29342) were received from Jörg Stülke's group in the University of Göttingen. Cells were cultivated at 37°C in cell culture flasks with the modified Hayflick medium[289]: 14.7 g/L Difco PPLO (Becton Dickinson, USA), 20% (v/v) Gibco horse serum (New Zealand origin, Life Technologies, USA), 100 mM HEPES-Na (pH 7.4), 1% (w/w) glucose, 0.002% (w/w) phenol red and 1000 U/mL freshly prepared penicillin G.

2.1.2 Cryo-EM grid preparation

M. pneumoniae cells were directly grown on cryo-EM grids. Quantifoil gold grids with holey carbon film support (200 mesh, R2/1; Quantifoil Micro Tools, Germany) were sterilized under Ultraviolet irradiation for 30 minutes, glow discharged for 45 seconds and finally sterilized for 10 minutes, before placing into a cell culture dish with the modified Hayflick medium. After inoculation, cell culture in dishes with grids placed on the bottom was performed at 37 °C. The culture time was controlled to be less than 20 hours to make sure cells are in the fast-growing phase before vitrification. Plunge freezing was performed with a manual plunger manufactured in the Max Planck Institute of Biochemistry in Germany. Grids with cells were quickly washed with PBS solution with protein A-conjugated gold beads (10 nm, Aurion, Netherlands), blotted for ~2 seconds from the back side, and plunged into the liquid ethane/propane mixture at the liquid nitrogen temperature.

For antibiotic-treated samples, the procedure was the same as above, except that drugs were added into the culture medium 15 to 20 minutes before plunge freezing. The final concentration of chloramphenicol (Cm; Sigma-Aldrich) in the medium is 0.5 mg/ml. The final concentrations for pseudouridimycin (PUM; Adipogen AG, Switzerland) and spectinomycin (Spt; Sigma-Aldrich) were both 0.4 mg/ml.

2.1.3 Cryo-ET data collection

Cryo-ET data were collected on a Titan Krios 300 keV transmission electron microscope (ThermoFisher Scientific) with a K2 or K3 direct detector camera (Gatan). The SerialEM software[290] was used for data collection. Before tilt-series acquisition, grid and grid square maps were acquired for cell position selection, at 135x and 2250x magnifications, respectively. Tilt-series were collected with the dose-symmetric scheme[203]. Images were recorded in dose-fractionation counting mode and raw frames were saved. Frames were motion-corrected with the

SerialEM plugin *alignframes* on-the-fly to generate tilt-series and mdoc files were created to store relevant information. For different tilt-series datasets, the used camera and imaging settings differed.

For tilt-series of the untreated cells, Cm-treated cells (session I 20181109) and PUM-treated cells, the used camera and settings are: Gatan K2 summit camera, magnification 81,000x, calibrated pixel size on the specimen 1.7005 Å, targeted defocus range 1.5 to 3.5 µm, tilt angle range -60° to 60°, tilt increment 3°, constant dose of ~3 $e^-/Å^2$ for all tilts, total dose ~120 $e^-/Å^2$. All tilt-series were acquired without phase plate, except for 14 tilt-series of untreated cells with the Volta phase plate. The VPP data were primarily used for visualization and generation of data-driven references for template matching. Alignment and operation of VPP were carried out as described previously[208].

For tilt-series of the Spt-treated cells, the camera and settings are: Gatan K3 camera in non-CDS mode, magnification 81,000x, pixel size 1.053 Å, targeted defocus range 1.25 to 3.25 µm, tilt range -60° to 60° with 3° increment, constant dose of 3.4 $e^-/Å^2$ for all tilts, total dose ~140 $e^-/Å^2$.

For a second Cm-treated dataset (session II 20201209), the camera and settings are: Gatan K3 camera in non-CDS mode, magnification 64,000x, pixel size 1.329 Å, targeted defocus range 1 to 3.25 µm, tilt range -60° to 60° with 3° increment, constant dose of 3.34 $e^-/Å^2$ for all tilts, total dose ~137 $e^-/Å^2$.

In total, 356 untreated, 65 Cm-treated, 86 PUM-treated, and 70 Spt-treated tilt-series were analyzed in this work, after excluding data with ice contamination or bad tracking during acquisition. For the expressome study (Chapter 5), 4 untreated and 3 PUM-treated tomograms were not included because they caused crashes in the prototype Warp/M software (1.0.6). It is possibly due to some values are beyond the program limit, and the crash disappeared in Warp/M versions after 1.0.7b. All tilt-series were acquired on the following number of grids from different preparations: 6 for untreated cells, 2 for Cm-treated, 1 for PUM-treated and 2 for Spt-treated cells.

2.2 Cryo-ET data processing and high-resolution ribosome averaging

2.2.1 Initial tilt-series alignment

All tilt-series alignment were based on fiducial gold beads in the Etomo software in the IMOD package (v4.9 or later)[212]. Due to the long time spanning between acquisitions of different datasets, the processing procedures are slightly different. For the untreated datasets, the first Cm-treated dataset and the PUM-treated dataset, the tilt-series were assembled from motion-corrected frames

generated by SerialEM plugin *alignframes*. For the Spt-treated and the second Cm-treated dataset (K3 camera), motion correction was performed in Warp. In Etomo, the binning used during alignment was 4 for K2 data and 6 for K3 data. Usually, one tilt-series had more than five fiducials that could be used for tilt-series alignment, except for the Spt-treated dataset. For the Spt-treated dataset of 70 tilt-series, 25 had no more than three fiducials suitable for alignment. For 21 of those, patch tracking was used instead. Tomograms used for template matching were reconstructed in Etomo software, with default settings (radial filter cut off 0.35, fall off 0.05). Tomograms that have big ice contaminants or more than three bad tilt views were excluded, which constitute less than 5% of all collected tilt-series. Cell thickness was estimated by measuring the distance from the top boundary of a cell to the bottom boundary in the central tomogram slice in the x-z view.

2.2.2 Ribosome localization by template matching

To generate a data-driven ribosome template for correlation-based particle localization in PyTom[218], 404 ribosomes were manually picked from seven VPP tomograms using EMAN2 *e2spt_boxer*. Sub-tomograms were extracted at the voxel size of 6.802 Å using the TOM toolbox[229] implemented in MATLAB 2016b (MathWorks). The averaging was done using RELION 3.0 with a feature-less sphere as the initial reference and it converged into a 70S ribosome. The density was low-pass filtered to 30 Å resolution and used for template matching in 20 defocus-only tomograms. In each tomogram, the top 400 peaks with highest cross-correlation scores were extracted. The Euclidean distance between peaks was set to at least 20 nm (30 voxels) to avoid duplicate extraction for the same ribosome. The raw hits were manually inspected to remove obvious false positives, such as plasma membranes, small extracellular particles and carbon edges. 4,668 clean sub-tomograms were extracted and averaging these resulted in a ribosome density at 25 Å resolution. The new ribosome density was low-pass filtered to 30 Å resolution and used as the template for all tomograms. During template matching, the angle search option was using the 'angles_19.95_1944.em' file available in PyTom and the score type is 'FLCFScores'.

Template matching is computationally heavy and all jobs were run on the high-performance cluster at EMBL. Tomograms were binned at 4 times for K2 data and 6 for K3 data. For each K2 tomogram, the 400 highest-scored peaks were manually inspected and obvious false positives including carbon edges, plasma membrane and other extracellular features were excluded. For the Spt-treated tomograms, the top 400 hits were extracted and no manual inspection was performed. For the second Cm dataset, the top 600 - 900 hits were inspected depending on cell size covered in the tomogram. The coordinates (x, y, z) of the center of all clean hits were saved for sub-tomogram extraction. In total, 109,990 untreated, 21,299 Cm-treated (first dataset), 31,547 Cm-

treated (second dataset), 23,014 PUM-treated, and 13,418 Spt-treated ribosome sub-tomograms were extracted in Warp.

2.2.3 Image processing in Warp

The datasets of untreated, Cm-treated (session I 20181109) and PUM-treated cells were processed in Warp[233]/M[234] version 1.0.7b, a prototype version which was later updated to version 1.0.9. For these, frame alignment used the on-the-fly results by SerialEM. Other image processing steps (sorting, CTF estimation, dose filtering, *etc*.) were done using Warp 1.0.7b. For the Spt-treated and the second Cm-treated datasets, all image processing was done in Warp 1.0.9. For all datasets, the tilt-series alignment was done in Etomo as described above. Warp can import the Etomo tilt-series alignment results (.xf and taSolution files) and reconstruct the tomograms. Tomograms can also be denoised using a neural network-based denoising tool. The VPP tomograms were denoised using a network trained by Dimitry Tegunov. Sub-tomograms were directly reconstructed in Warp after providing the coordinate files. Usually, 4x binned sub-tomograms were extracted for initial classification and refinement. Lower binning was used for high-resolution refinement after M.

2.2.4 Structure refinement in RELION and M

Initial 3D classification was performed in RELION 3.0.7 or 3.0.8[291] to further exclude false positives and classify 70S and free 50S. 70S and free 50S were refined separately and only 70S were used to further refine the tilt-series parameters in M. Initial refinement on 70S was performed in RELION, usually with a sphere mask (~320 Å in diameter). The average and star file from RELION were read by M, which can perform average-based multi-particle refinement of the tilt-series (geometric alignments, warping, CTF fittings, tilt movie) and refinement of structure (particle pose, *etc*.). The process of a typical M refinement is provided in Table 2.1. The geometric parameters were first refined for 1-2 rounds, followed by 2-3 rounds including CTF and other parameters. After overall 70S refinement, focused refinement (only particle pose) on 30S and 50S separately with solvent masks was performed. After M refinement, new sub-tomograms were extracted in Warp, with the better refined geometric and CTF parameters automatically read by Warp. Fourier shell correlation (FSC) calculation, local resolution estimation, and other post-processing operations were either done in M or in RELION.

Table 2.1 | Illustration of the M multi-particle refinement process.

Refinement round	#0*	#1	#2	#3	#4	#5
Resolution (Å)	10.5	5.6	4.4	3.6	3.3	3.2
Settings for the refinement run in Warp/M						
Image warp grid 7 x 5		✓	✓	✓	✓	✓
Volume warp grid 3x4x2x10		✓	✓	✓	✓	✓
Stage angles		✓	✓	✓	✓	✓
Particle pose		✓	✓	✓	✓	✓
CTF				✓	✓	✓
Tilt movie refine				✓	✓	✓

* The initial RELION refinement result before M.

2.2.5 Ribosome atomic model building

To obtain the first atomic model for *M. pneumoniae* ribosomes, the small and large subunit were built separately based on the focused refined 30S and 50S maps from the Cm-treated dataset. To start, homology models for ribosomal RNAs and proteins were generated. Sequences of rRNAs (REFSEQ NC_000912) were first aligned to the *E. coli* ribosome (PDB 4YBB) with the online SINA server[292]. The software ModeRNA[293] was then used to generate rRNA homology models based on previous alignments. Homology models for all ribosomal proteins were generated using SWISS-MODEL[294] with a *Bacillus subtilis* ribosome structure as the template (PDB 3J9W), except for S21 (PDB 5MMJ), L9 (PDB 1DIV and 4V63) and L10 (PDB 1ZAV). The initial model was generated by fitting the homology models of rRNAs and r-proteins into the densities as rigid-bodies using Chimera[295]. Iterative refinement with real-space refinement in PHENIX[276] and manual adjustment in Coot[296] were performed to refine the model. The ribosomal protein L9, L10 and L11 were not changed due to limited local map quality. For L9, based on the results from focused classification, the flat conformation that resides on the ribosome surface was the predominant conformation and was built in the model. Sequence extensions for S6, L22 and L29 did not exist in the template and were built *de novo*. FSC between the model and the map were calculated for validation[297]. Final models were validated using MolProbity[298].

2.2.6 Bioinformatic analysis of ribosomal protein extensions

This analysis was performed by Maria Zimmermann-Kogadeeva in a collaborative project. Sequences of ribosomal proteins in *M. pneumoniae* and other representative bacterial species were compared. The brief procedure is summarized as below. First, protein sequences and the RefSeq genome annotation for *M. pneumoniae* M129 were imported from the NCBI (National Center for Biotechnology Information, USA). Second, eggNOG-mapper was used to further annotate protein sequences to assign the COG (Clusters of Orthologous Groups)[299]. Third, for all the annotated ribosomal proteins in representative bacterial species, the COG multiple sequence alignment results were directly read from the eggNOG database [300]. Because *M. pneumoniae* M129 is not listed as one of the representative species, MAFFT[301] was used to add its sequences to the multiple sequence alignments. Fourth, the sequences of *M. pneumoniae* M129 and the selected representative bacteria were compared to those of *E. coli* K-12 (substrain MG1655), and the length of sequence extensions before the N-terminus and after the C-terminus were calculated. Fifth, the ones with N- or C-terminal extensions of more than twenty residues were analyzed in iTOL based on NCBI taxonomy tree[302]. Sixth, for the 11 *M. pneumoniae* proteins with extensions of more than twenty residues, results from several resources were combined: i) crosslinks from the in-cell CLMS data[255] were mapped on the sequence, including self-crosslinks and crosslinks to other proteins; ii) essentiality sequences were labelled according to the previous transposon study[303]; iii) protein disorder and secondary structure predication were performed and the predication results were mapped on the sequence.

2.3 Ribosome classification and spatial analysis

2.3.1 Classification of the translation states of ribosomes

Maximum-likelihood 3D classification was performed in RELION 3.0.7 or 3.0.8 with the ribosome sub-tomograms extracted after M refinement on 70S ribosomes. A hierarchical and exhaustive classification workflow was developed to handle the high structural heterogeneity, with at least three tiers of RELION classification jobs (Chapter 4). A similar procedure was applied for the untreated and antibiotic-treated datasets. Below I will use the untreated dataset as an example to describe the full process.

In tier one, the 109,990 sub-tomograms extracted from untreated cells were first classified into 70S and free 50S. The 24,157 free 50S were further divided into two classes, with a focused spherical mask on the ribosome recycling factor binding site. Before extensive classification on the 70S ribosomes, visual inspection, multibody refinement and test classification runs were carried out to estimate the structural heterogeneity and determine classification strategies. To find

the best settings in RELION, I tested different masks (70S mask, focused 30S mask, spherical mask on the elongation factor binding region, solvent elongation factor mask, spherical tRNA path mask, solvent tRNA path mask, *etc.*), initial reference (70S ribosome, sphere, 'none'), angular search ranges (global, local, 'none'), and other optimization parameters (numbers of specified classes 2 - 16, regularization T values 2 - 10, specified iterations 25 - 40, 'limit resolution E-step' 5 - 10 Å, *etc.*). In practice, I mostly used a feature-less sphere with soft edges as the mask to avoid introducing any arbitrary information. The classification results using local masks focusing on the factors binding regions (EFs or tRNAs) appeared to be the most stable and consistent. The specified number of classes for one classification job was usually made higher than the number of distinct structures that could be classified.

In tier two, the 77,539 full 70S ribosomes were classified with a spherical mask (mask I, see Figure 4.1) that cover the A-P-E tRNA path. In the first rounds of classification jobs, sub-tomograms were divided into four major groups based on different tRNA occupancies: "P, E", "A**, P/E", "P", "A, P". In the following rounds, the "A**, P/E" group was further classified into two groups "A*, P/E" and "A/P, P/E", which showed slight difference for the tRNA near "A**" site. In addition, one new class with only one hybrid "P/E"-site tRNA was obtained. Two additional classes were not interpretable based on the density: 2,150 particles with poorly resolved “dim” 30S subunit and 1,484 particles with density near the P site.

In tier three, for each of the major classes from tier two, focused classification with a local mask (mask II, see Figure 4.1) covering the elongation factor and A/T tRNA binding regions was performed. The "P, E" class was further classified to 3,324 particles with EF-G, 4,634 particles with EF-Tu•tRNA, and 1,803 particles without any additional density. Class "A*, P/E" and "A/P, P/E" could be further classified to be with and without EF-G. For "P" class with only one P-site tRNA, 12,464 particles were classified to have additional EF-Tu•tRNA.

For every classification step, more than three RELION classification jobs with the same or slightly different settings (local search range, class number, *etc.*) were run in parallel. The most stable RELION classification job was chosen to sort sub-tomograms. After sub-tomogram sorting, following classification runs were done to validate the classification result. Particles that were "wrongly" classified were relocated to the correct class. The validation runs were repeated until convergence. The exhaustive approach was carried out for all classification steps. In the end, RELION refinement was done for every final class, which was followed by more classification runs with the newly refined structure as the reference. The loop was repeated until it did not generate any new structures, as a demonstration of exhaustiveness of the final classification.

2.3.2 Model building for classified ribosomes

To construct models for different ribosome classes, atomic models of the small and large ribosomal subunit were flexibly fitted into the densities to generate initial 70S structures. For elongation factor EF-G and EF-Tu, the initial structures were generated by homology modeling with PDB structure 4V7D and 4V5L as templates, respectively. The homology model of *M. pneumoniae* Phe-tRNA (REFSEQ NC_000912) was based on the tRNA from an *E. coli* ribosome structure (PDB 4V7C) as the template. The homology models for mRNA and nascent peptide chain were generated using templates from PDB structure 3J9W. Homologous structure for the ribosome recycling factor was generated using PDB structure 1EH1 as the template. For each class, the initial model was first fitted into the density in Chimera and then flexible fitting was performed using Namdinator[304]. Visual inspection and validation were done to ensure good fitting and interpretation of the density.

To quantify the rotation of 30S body, all class structures were aligned to the large subunit and then the rotation angles were measured with nucleotide eleven of 16S rRNA as the pivot point. Looking from the solvent side of the small subunit, counterclockwise rotations were indicated by positive numbers and clockwise rotations negative. To estimate rotations of the 30S head domain, structures of all classes were aligned by the 30S body and the rotation angles were measured with the rotation axis near the neck region. For both measurements, the rotation angles in the predominant "A, P" class were defined as zero degree. To describe dynamics of the L1 stalk, after aligning all classes based on the large subunit, the distance between L1 center (near 23S rRNA, nucleotide 2181) and the center of E site (a fixed point localized in class "P, E") was estimated.

2.3.3 Spatial analysis of ribosomes and polysomes

The spatial and functional map of ribosomes in the cell was obtained by projecting back the structures of ribosome classes into the cellular tomogram. The projection was performed using the TOM toolbox[229] after conversion of shifts and rotations from RELION refinement. Coordinates for the new center are "_rlnCoordinateX/Y/Z" (the coordinates used for particle extraction) minus "_rlnOriginX/Y/Z" in RELION data star. The rotation system for TOM is zxz while RELION is using zyz. Conversion of the three Euler angles was done using the TOM script tom_eulerconvert_xmipp. The 3D mapping were usually using 4x or lower binning (voxel size 6.8 Å or larger). For ribosome concentration estimation, the total number of ribosomes was divided by the estimated cellular volume in the tomogram.

A custom script was written in MATLAB 2016b for the polysome detection. The calculation considered the positions and orientations of every ribosome, which provides more accurate

annotation and the ribosome sequence information within the polysome. Coordinates of the mRNA entry and exit sites for all ribosomes were calculated with the rotation and translation information in the data star file from RELION refinement. Distances from the mRNA exit site of one ribosome (as the possible preceding ribosome) to the mRNA entry sites of all neighboring ribosomes (as the potential following ribosomes) were calculated. A distance cutoff of 7 nm was selected to define whether two ribosomes belong to the same polysome. A unique number identifier was assigned to each polysome. The calculation and polysome definition were performed for all 70S ribosomes, along with the sequential number assignment.

The main part of the polysome detection script can be found in the GitHub repository, https://github.com/xueliang4906/polysome_detect/.

The distribution of relative positions (x, y, z) of adjacent ribosomes within the polysome, *i.e.* the relative position of the following ribosome (i+1) to the preceding ribosome (i), was analyzed after normalizing the relative position vector with rotations of the preceding ribosome. The relative rotations between adjacent ribosomes was represented as Euler angles (ψ, θ, φ in XYZ system) of the rotation of the following ribosome. Using the k-means clustering function in MATLAB 2016b, I determined two major arrangement configurations for adjacent ribosomes (how the following ribosome rotates relative to the preceding ribosome), which resemble the previously reported "top-top" and "top-back" configurations[136].

2.3.4 RELION classification of polysomes

In parallel to the polysome detection based on spatial information as described above, *de novo* RELION classification was performed to classify di-ribosomes (*i.e.* structure of the two adjacent ribosomes in polysomes). For this purpose, the sub-tomogram for each ribosome were re-extracted with a larger box size that can accommodate its close neighboring ribosomes. Initial classification was performed with a spherical mask (diameter 740 Å), with either a ribosome or a sphere as the reference. Two classes of di-ribosome could be identified. In one class, the ribosome in the center is the preceding one, while in another class it is the following. They are structurally identical if rotating 90 degrees, but the large rotation possibly made RELION identify them as two separate classes. Both classes were considered as polysomes when comparing with the polysomes detected spatial mapping. Refinement of the di-ribosomes only resulted in medium-resolution (~30 Å) because of the high flexibility between two ribosomes.

To better refine the ribosome-ribosome interface in polysomes, the di-ribosomes (polysomes) were further classified. One of the two classes identified above was selected and the corresponding sub-tomograms were re-extracted after centering on the interface rather than on one ribosome.

After alignment on one ribosome, classification was performed with a spherical mask focusing on the second ribosome without angular search. 8 to 16 classes were specified to maximize the possibility to classify fine rotation difference of the following ribosome relative to the central one. Models for the resulting classes with better resolutions (10-15 Å) were built by fitting the ribosome structures. For the ribosomal protein L9 in the preceding ribosome, the homology model with the stretched conformation (PDB 4V63) was fitted as a rigid-body.

2.3.5 Statistical analysis of translation elongation states in polysomes

This part was done in collaboration with Maria Zimmermann-Kogadeeva. All processing were with the native untreated cells if not specified otherwise.

First, we compared the occurrence frequencies of different translation elongation states across 356 cellular tomograms in polysomes against all ribosomes and mono-ribosomes. This comparison indicated whether there is a difference in the overall state distribution in polysomes. Statistical significance was assessed with a two-sided Wilcoxon-Mann-Whitney test in MATLAB 2019b and its *ranksum* function. Next, we divided polysomes into ribosome pairs (di-ribosomes, with a continuous sampling box shifting from the first ribosome to the second last ribosome in the polysome) and calculated the experimental occurrence frequency of each ribosome pair. The theoretical probabilities of random ribosome pair, *i.e.* if there is no cross-influence between adjacent ribosomes within the polysome, were calculated based on the overall state frequencies of elongation states across all ribosomes. The experimental pair frequencies were divided by the theoretical random pair probabilities to test whether there is a difference in the occurrence of specific state pairs.

To test the statistical significance of the difference in occurrence frequencies of the experimental pairs against theoretical probabilities, permutation analysis was performed. The sequence of elongation states in polysomes were represented as a matrix: each row indicates one polysome and each column indicates a ribosome position in the sequence. In each unit of the matrix is a unique number representing the elongation state of the ribosome. As polysomes are of different length, blank positions were given the 'NaN' (not-a-number) value. The polysomes from all tomograms were combined into one matrix (8,641 polysomes in total). In the permutation performed with the *randperm* function in MATLAB 2019b, all elements of the polysome matrix were randomly shuffled for ten thousand times. For each row, the elements were sorted after shuffling to make sure the blank 'NaN' units are always in the last positions in each row and do not affect analysis. Rows with sequence less than two ribosomes were not included. The frequencies of shuffled elongation state pairs were calculated in the same way as the experimental pairs. Permutation *p*-value for each pair was calculated as the minimum between the number of

permutations that have the pair frequency not larger than the experimental frequency divided by the total number of permutations, and the one minus this value (equations see Figure 4.15). Permutation *p*-values were adjusted for multiple hypotheses testing with the Benjamini-Hochberg procedure using the *mafdr* function ('bhfdr', 'true') in MATLAB 2019b.

To further compare the influence of different polysome detection cutoffs on the elongation analysis, new matrices were generated by using distance cutoffs from 3 nm to 10 nm. As described above, the distances from the mRNA exit site of one ribosome and the mRNA entry site of another ribosome were measured. The fractions of elongation state pairs were calculated in the same way as above. For each distance cutoff and each state, we calculated the ratio between the number of state pairs in which the preceding ribosome (i) is of this state and the number of state pairs in which the following ribosome (i+1) is of this state.

2.3.6 Single cell clustering analysis

This part of work was performed together with Maria Zimmermann-Kogadeeva for a collaborator project. The distribution of translation elongation states of 70S ribosomes in the four major datasets (356 untreated, 65 Cm-treated, 70 Spt-treated and 86 PUM-treated cells) were used to perform clustering analysis. As tomograms usually cover 60% - 90% of the cell, the distribution of ribosome classes in each tomogram is expected to be representative of the distribution within the whole cell. Ribosomes identified in one elongation state (class) were connected to the cellular tomogram they belong to. Each class identified in antibiotic-treated cells was assigned with the same class identifier of its closest class in the untreated cells. The percentages of the elongation states (classes) within one cellular tomogram were calculated, and these numbers form the translation elongation profile for one cell. Hierarchical clustering analysis using the *clustergram* function in MATLAB 2016b was performed (see Figure 4.24).

2.4 Integrative structural study of expressome

2.4.1 Classification and refinement of the expressomes

To identify possible assemblies formed by RNA polymerase and the ribosome during transcription-translation coupling, extensive classification focusing on the region above the ribosome mRNA entry site was carried out with the ribosome sub-tomograms reconstructed at a large box size (~680 Å). This part of the work was first done in the version 1.0.6 of Warp and M and results were based on that if not specified otherwise. The results can be reproduced in Warp/M 1.0.9. To keep consistence with the publication[255], most results presented here are using those from the Warp/M 1.0.6. The improved CTF refinement in the updated Warp/M didn't make any

difference to structure determination of the expressomes, as the intrinsic flexibility of the complex appears to constitute the main limiting factor. There may be slight difference in particle numbers in comparison to the publication because several tomograms that failed in Warp/M 1.0.6 were added to the high-resolution ribosome analysis in Warp/M 1.0.9.

First, ribosome sub-tomograms that belong to the class of free 50S, and one class of closely assembled polysomes were excluded. The remaining 70S ribosome sub-tomograms were subjected to focused classification with a local mask near the mRNA entry site on top of the small ribosomal subunit. After testing different strategies and settings, these particles were sorted into three major classes: ribosomes with clear RNA polymerase density, ribosomes with undefined additional density and ribosomes without significant density above the mRNA entry site. At least three parallel RELION classification jobs with slightly different settings or masks were carried out to validate the final classification result.

For refinement of the 2,952 ribosome sub-tomograms with clear RNAP density, different strategies were undertaken to get the best local resolutions for data interpretation, especially for the RNAP density. RELION refinement with a global mask covering the whole complex generated a density with 8.5 Å overall resolution. The local resolution of RNAP, however, was only around 12 Å because of high flexibility and the dominance of alignment by the large ribosomal subunit. To make sure the RNAP-30S interface is determined with the best resolutions, focused refinement with a local RNAP-30S mask was performed. Though the final overall resolution slightly decreased to 9.2 Å, the local density quality of the RNAP part was improved.

Multibody refinement in RELION[305] was performed to further refine the RNAP density and to examine the flexibility of the complex. Following the consensus refinement with a global mask covering the whole complex, three local masks corresponding to the three major parts of the complex (the RNAP, the small and large ribosomal subunits) were provided for RELION to refine the three bodies separately. Following structure refinements, principal component analysis based on relative orientations of the three bodies was done to characterize the relative motions.

For the 17,202 ribosomes with additional blurred density, further classification or different refinement strategies did not result in any interpretable density, possibly due to both high compositional heterogeneity and conformational flexibility.

Similar classification, refinement and multibody refinement were carried out for sub-tomograms from the antibiotic-treated cells.

FSC calculation, B factor sharpening and local resolution estimation were done following standard post-processing procedure in RELION. Plots of angular distribution of particles were generated using scripts as described before[306].

2.4.2 Whole-cell crosslinking mass spectrometry

This part of work was done by the collaborators Francis O'Reilly and Ludwig Sinn in the group of Juri Rappsilber for the collaborative expressome project. The detailed protocol can be found in the Materials and Methods in the collaborative paper[255]. Briefly, *M. pneumoniae* M129 cell suspension was incubated with either DSS or DSSO. The incubation was done at 25°C for 30 minutes with gentle agitation. For proteome extraction, the cell pellet was lysed with sonication in the denaturing solution. Four separate datasets were collected, with different peptides enrichment scheme (see Figure 5.6). The mass spectrometry data acquisition and processing was done as described previously[255]. Final results were filtered to an estimated false-discovery rate (FDR) of 5% on the residue pair level and on the protein-protein interaction level using xiFDR[259]. For FDR estimation, inter-protein and intra-protein crosslinks were handled separately. To identify proteome abundances, mass spectrometry data acquisition and analysis for bottom-up proteomics was also performed as described[255].

2.4.3 Integrative modeling of the expressome based on cryo-ET and CLMS data

This part of work was done by collaborators Andrea Graziadei and Swantje Lenz in the lab of Juri Rappsilber for the collaborative expressome project. The detailed protocol can be found in the Materials and Methods in the collaborative paper[255]. Below is a brief summary of the procedure.

The integrative modeling platform (IMP, v2.12)[278] was used to build the integrative model of the *M. pneumoniae* expressome based on in-cell CLMS and cryo-ET data. The building blocks for integrative modeling are summarized in the 'Table S4' in the publication[255]. The ribosome and RNAP-NusG models were first rigid-body fitted into the cryo-ET density thanks to their relatively large size. Then, parts of the model (30S subunit, NusG-bound RNAP, RNAP δ subunit and NusA domains) were coarse-grained and fitted as rigid bodies. The crosslinks from whole-cell CLMS (377 DSS and 144 DSSO crosslinks within RNAP, RNAP δ subunit, NusG, NusA and the ribosomal proteins) were represented as distance restraints in IMP.pmi. The inflection point in the scoring function for both DSS and DSSO was set to 25 Å Cα-Cα distance. The multibody refined RNAP+ density of the complex (after repositioning to the consensus refinement) was represented as a Gaussian mixture model, and encoded as a Bayesian restraint using the IMP.BayesianEM function in IMP[307]. Theoretical densities of rigid bodies and flexible regions were represented as Gaussian mixture models as well. Sampling was done by the 'Replica Exchange Gibbs sampling' in IMP. During the sampling, core regions of RNAP (α subunits not including the flexible CTD, β subunit not including flat-tip helix, β' subunit and NusG) and ribosomal proteins were kept fixed because the rigid-body fitting already ambiguously determine their positions and they should not be affected by random sampling.

The scoring was done following previous studies[308,309]. The replica exchange samplings were selected with generous crosslink cutoffs (distance < 35 Å) and stringent cutoffs in BayesianEM score and excluded volume score, resulting in a subset of 20,765 models (out of 14,400,000 sampled models). The sampling exhaustiveness was assessed by splitting the solutions into two and cross-comparison. The rmsd-based clustering of NusA in different cutoff models was used to examine the precision of sampling[308]. The precision of each cluster and subsequently the model precision were defined as the mean distance to the solution that is closest to center of the cluster (the centroid model). The most populated cluster (containing 97% of selected models) had a model precision of 11.54 Å. Only one solution cluster met the threshold based on the sampling precision. The centroid model and the corresponding localization probability densities were regarded as the representative solution from integrative modeling. Finally, five sets of jobs (two independent runs with the same parameters as the main) after randomly excluding 20% of the crosslinks to test the modeling robustness. Crosslinks were visualized with XlinkAnalyzer[310].

Structure visualization and preparations for figures and movies were done in Chimera[295] and ChimeraX[311] if not specified otherwise. Plots were generated in either MATLAB 2016b or 2019b.

Chapter 3: High-resolution in-cell ribosome structures

3.1 Cryo-ET of *M. pneumoniae* cells

3.1.1 On-grid cultivation

To maintain *M. pneumoniae* cells in the native states and avoid potential perturbations, cells were directly grown on gold cryo-EM grids with holey carbon film support. When cultivated *in vitro*, *M. pneumoniae* cells tend to grow on culture flask surfaces rather than in suspension[38,53]. Taking advantage of this property, I developed the on-grid cultivation workflow which can reduce the operation time from taking grids out of the culture medium to the final vitrification to less than 10 seconds. Within this time, quick wash with buffer containing gold beads and blotting with filter paper from the back side were carried out. In the following plunge freezing step, native cells on the grid are rapidly (0.1 millisecond[188]) frozen in the hydrated state.

As *M. pneumoniae* cells mostly grow on the surface, they do not have a typical growth curve like suspension culture. In my measurement, *M. pneumoniae* cells have a fast-growing phase for about 80 hours after inoculation before reaching the stationary phase. The time for on-grid culture was controlled below 20 hours to keep cells in the fast-growing states. *M. pneumoniae* cells on the grid exhibited typical morphologies of vividly growing colonies (Figure 3.1), aligning with the observations under light microscopes[38] and scanning electron microscopes[29,49]. Individual cells and relatively small cell clusters extend from big colonies on the carbon support and in the holes, which are demonstrations of cells in the fast-growing phase. Hence, after rapid plunge freezing, *M. pneumoniae* cells were preserved in close-to-native states with minimal perturbations.

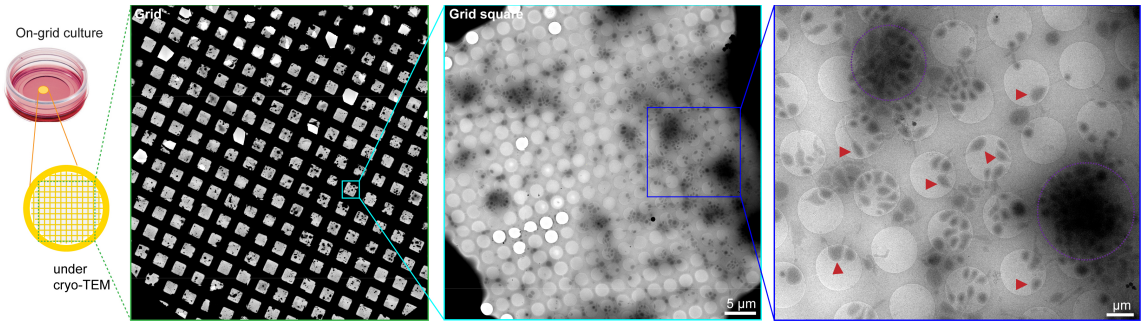

Figure 3.1 | *M. pneumoniae* cell cultures under the cryo-EM.

M. pneumoniae cells were cultivated in the cell culture dish with cryo-EM grids placed on the bottom. Grids with cells growing on the surface were taken to plunge freezing, and then transferred into cryo-EM. The grid was imaged at different magnifications to map and select individual cells (red arrowheads) suitable for tilt-series data collection. Big cell colonies (the two dark regions in the last panel) were too thick for cryo-ET data collection.

3.1.2 Cryo-ET data collection and processing

The data collection strategy was optimized for high-resolution sub-tomogram averaging. First, the ice thickness for regions containing individual cells can be made 100-200 nanometers by carefully optimizing the blotting conditions prior to freezing. The data collection procedure is straightforward and high-throughput without the need for FIB thinning. Second, I adopted the dose-symmetric scheme and the well-established SerialEM scripts for automated data collection at EMBL[203]. With the dose-symmetric scheme, high-resolution information at low exposure dose and low tilt angles is maximized for structure refinement. For each tilt-series, the tilting range was set to -60° to 60° with 3° increment, and a total dose of 120-140 $e^-/Å^2$. Third, the magnification was chosen to give a small pixel size on the sample and meanwhile to image most regions of one *M. pneumoniae* cell. Generally, the tomogram can cover 60% - 90% of a typical *M. pneumoniae* cell, which provides a good representation of the whole cell (Figure 3.2A-D).

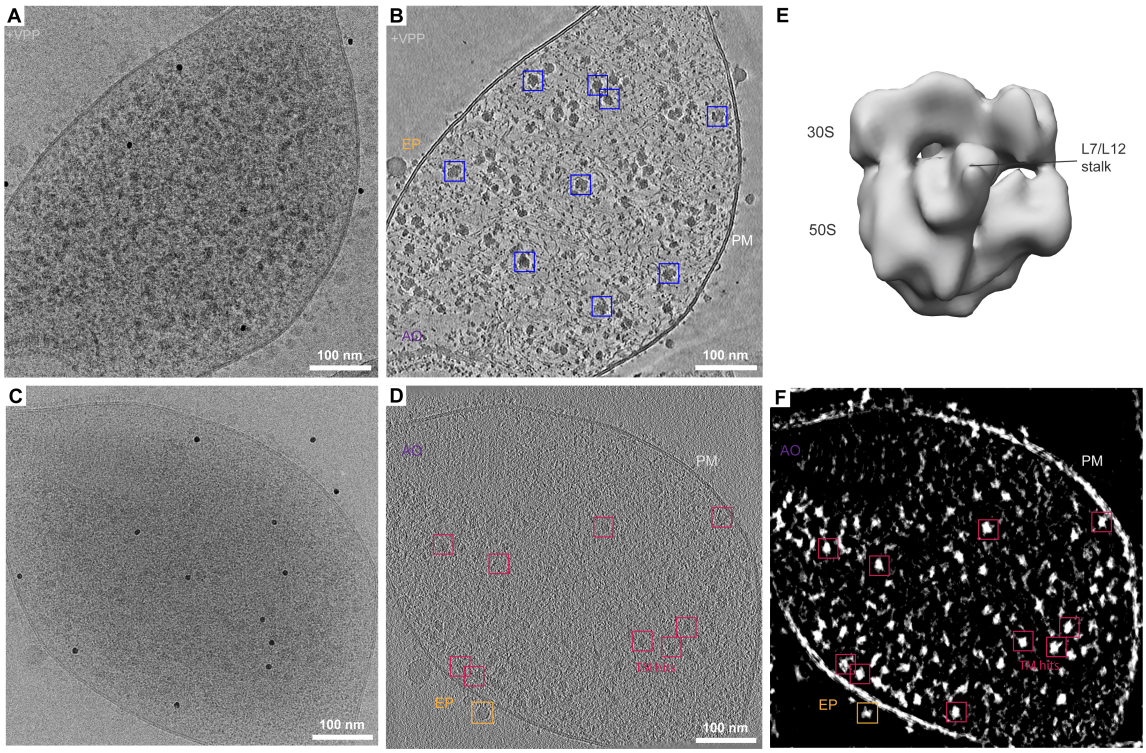

Figure 3.2 | Tomograms of *M. pneumoniae* cells and ribosome template matching.

(**A**) Image of a representative *M. pneumoniae* cell at zero degree in a tilt-series collected with the Volta phase plate and 2 µm defocus. The dark black dots are applied gold beads. (**B**) Central slice of the reconstructed tomogram of the cell in (A), with gold beads masked out. The tomogram was denoised using Warp (Materials and Methods)[233,234]. The ribosomes (blue boxes) are visible. PM, plasma membrane; AO, attachment organelle. (**C-D**) Tilt image and central tomographic slice of a representative cell imaged with defocus only. Several ribosome TM hits are highlighted by red boxes in (D). EP, unknown extracellular particles. (**E**) The initial ribosome template generated as described in section 3.2. (**F**) Cross-correlation coefficient map corresponding to the tomogram in (D) after TM. Brighter voxels indicate higher TM scores.

Processing of raw frames and images was primarily performed in SerialEM and IMOD packages[212,290], which was later re-processed with the Warp/M software package[233,234]. Alignment of the tilt-series was using gold beads in Etomo[212] and the reported alignment residual errors were usually kept below 6 Å. Tomograms (*i.e.* reconstructed 3D cellular volumes) were first generated in Etomo, which were used for visualization and template matching. Cellular tomograms of low quality (not enough gold beads, large residual errors, more than four "bad" tilt views, ice contaminations, *etc.*) were excluded.

In this thesis, four major data groups were acquired depending on treatment conditions: native untreated, chloramphenicol (Cm)-treated, pseudouridimycin (PUM)-treated and spectinomycin (Spt)-treated. Two Cm-treated datasets were collected with different cameras and pixel sizes. Details for the major datasets are summarized in Table 3.1.

3.2 Ribosome localization by template matching

The 3D template matching function in PyTom[218] was used to localize ribosomes inside the *M. pneumoniae* cellular tomograms. To avoid template biases, a data-driven ribosome template was generated by averaging ~400 manually picked ribosomes from a small dataset of VPP tomograms (Figure 3.2E). The ribosome density was then used as the template to search in defocus-only tomograms. The template matching function calculated cross-correlation coefficients between the 3D template rotating over 1,944 combinations of Euler angles with the sampling step of about 20° (details see Materials and Methods) and a sub-region of the tomogram, and then the best-matched coefficient value was assigned for the center voxel of the sub-region. After the thorough search for all sub-regions, a coefficient map of the same dimensions of the tomogram was generated (Figure 3.2F), in which the value of each voxel is the best-matched coefficient. Based on the coefficient map, coordinates of the center of the highest-scored regions (*i.e.* ribosome hits) could be calculated and used to derive a list of raw TM hits. The estimated numbers of ribosomes in one *M. pneumoniae* cell varies in different publications, with an average number of 300[32,53,312]. Combining the reported numbers and the preliminary estimation, I extracted the 400 highest-scored hits from each tomogram for the datasets collected with Gatan K2 camera. For the Spt-treated dataset and the second Cm-treated dataset collected with a K3 camera, cutoffs of 600 or higher were tested because of different camera and magnifications.

The 3D template matching method is dominated by low frequency information (*e.g.* overall shape) and prone to false positives that have higher contrast than the cell interior (*e.g.* membranes)[218]. The most common false positives in *M. pneumoniae* data are plasma membranes, extracellular particles/vesicles, edges of the grid carbon support, ice contaminations and gold beads (Figure 3.2F). For most datasets, obvious false positives outside the cell were manually excluded to avoid

their potential interference with sub-tomogram averaging and classification. For the Spt-treated dataset, 3D classification was tested instead to clean up raw TM hits and remove false positives. This approach worked, but took much longer time to ensure true positives were not discarded due to large classification variations.

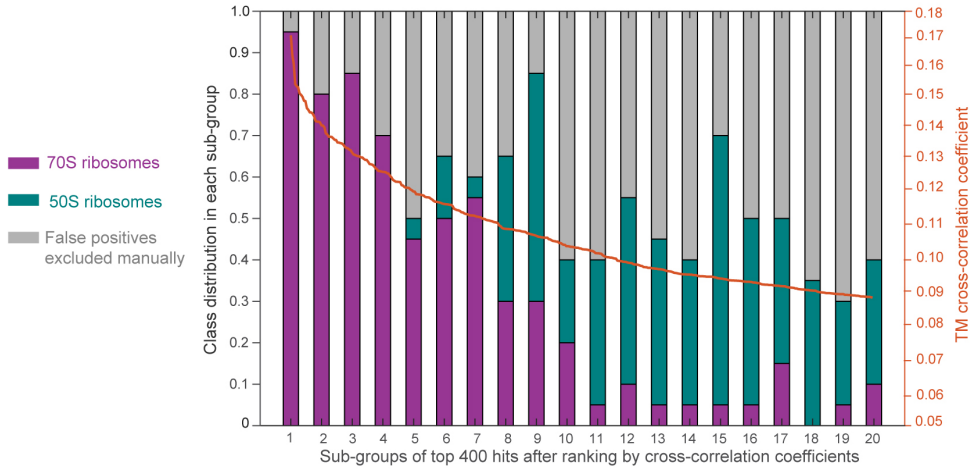

Figure 3.3 | Template matching cutoff selection.

Top 400 hits in a representative K2 tomogram were ranked by cross-correlation scores (orange curve) and divided into 20 sub-groups. After visual inspection and initial classification in RELION, percentages of 70S ribosomes (magenta), free 50S (green) and false positives (grey) were calculated within each sub-group. The percentages of 70S ribosomes are considerably low in sub-groups with lower scores, suggesting that the cutoff of 400 for one tomogram can recover the vast majority (estimated >90%) of 70S ribosomes in the data. Larger cutoffs have been tested, but did not significantly recover more 70S from the tomograms.

According to the preliminary results, ribosome localization by template matching gave high recovery rate and accuracy. Initial classification in RELION was performed to sort the extracted sub-tomograms into free 50S and full 70S ribosomes. As the primary aim of the study is focused on translating 70S ribosomes, free 50S were regarded as false positives. The initial classification results suggested that the vast majority (>90%) of the targeted 70S ribosomes and part of free 50S within the cellular tomograms were retrieved (Figure 3.3). As shown in Table 3.1, large ribosome sub-tomogram datasets were obtained from differently treated cells, providing the data basis for high-resolution structure determination.

3.3 High-resolution ribosome structures inside the cell

In this section, I will focus on refinements of 70S sub-tomograms to achieve high-resolution ribosome structures. Details about initial classifications that distinguish 70S and free 50S, and following classification of 70S can be found in the next Chapter 4.

3.3.1 The Warp-RELION-M workflow

The image processing and sub-tomogram analysis were performed in the newly established Warp-RELION-M workflow we tested and validated in a collaboration with the software developer, Dimitry Tegunov[233,234] . Warp was used for image processing, CTF estimation, and reconstruction of tomograms, sub-tomograms and the corresponding 3D CTF models. Classification and refinement of sub-tomograms were performed in RELION. The software M imported the RELION refinement results and the Warp image processing files to perform average-based multi-particle refinement that can refine both tilt-series (alignments, deformation models, CTFs, *etc.*) and particle poses in a gradient manner (Figure 1.9; also see Materials and Methods).

The similar Warp-RELION-M processing procedures were performed for all major cryo-ET datasets (*M. pneumoniae* cells of different conditions) and the final resolutions for 70S ribosome averages obtained with these datasets are summarized in Table 3.1.

Table 3.1 | Statistics of major cryo-ET and ribosome sub-tomogram datasets.

	Native untreated	**Cm-treated I, 20181109**	**Spt-treated**	**PUM-treated**	**Cm-treated II, 20201209**
Tilt-series number	356	65	70	86	137
Camera	K2	K2	K3	K2	K3
Pixel size (Å)	1.7005	1.7005	1.053	1.7005	1.329
Etomo residual error Mean±SD (Å)	4.7±1.6	4.5 ±1.7	7.1±4.1	5.1±1.5	4.6±1.2
Estimated thickness Mean±SD (nm)	157±28	164±30	140±35	143±38	144±42
clean TM hits	109,990	21,299	13,418	23,014	31,547
free 50S	24,157	-	483	5,671	773
70S	77,539	18,987	12,935	15,314	30,774
70S resolution before M (Å)	9.6	11.5	13	10.6	10.5
70S resolution after M (Å)	3.5	3.5	4.4	4.0	3.0

Resolution estimation based on "gold-standard" FSC=0.143 in RELION or M.

3.3.2 In-cell ribosome maps at 3 Å resolution

The three datasets that provided the best resolved ribosome structures will be discussed in details. These are the native untreated dataset, the first and the second Cm-treated datasets (Table 3.1).

Refinement of the 77,539 ribosome sub-tomograms from 356 untreated cellular tomograms generated a 70S ribosome density with 3.5 Å overall resolution. Local resolutions for core regions of the large ribosomal subunit are close to the Nyquist limit of 3.4 Å for this dataset. (Figure 3.4A-C). The small ribosomal subunit is of lower resolutions (4-8 Å) because of LSU's dominance of the alignment and SSU's high flexibility during translation[70,113].

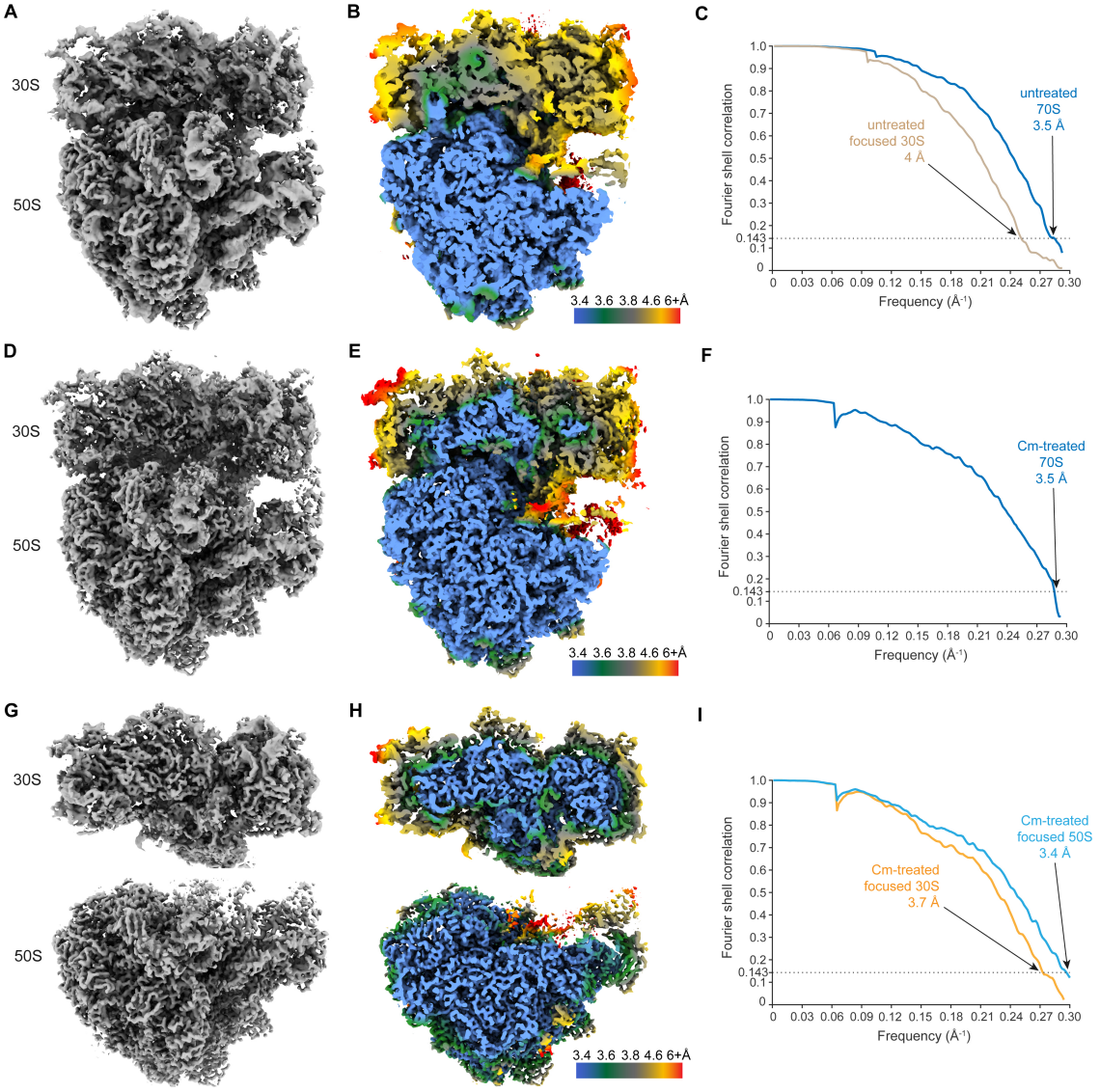

Figure 3.4 | 3.5 Å in-cell ribosome densities from K2 datasets.

(**A**) 70S ribosome density averaged from 77,539 ribosome sub-tomograms in untreated *M. pneumoniae* cells. (**B**) Local resolution map for (A). (**C**) FSC curves for the global 70S refinement and the focused refinement on SSU. (**D**) 70S ribosome averaged from 18,987 sub-tomograms in the first Cm-treated dataset (dataset I 20181109). (**E**) Local resolution map for (D). (**F**) FSC curve for the Cm-treated 70S ribosome. (**G**) The

focused refined 30S and 50S densities of the Cm-treated ribosome. (**H**) The corresponding local resolution maps for (G). (**I**) FSC curves for focused refinements of the Cm-treated 70S on 30S and 50S respectively. The untreated and the first Cm-treated datasets were collected with the Gatan K2 camera.

The 70S ribosome density averaged from 18,987 sub-tomograms in 65 Cm-treated cells (dataset I 20181109) was also determined at 3.5 Å overall resolution. Thanks to the reduced structural heterogeneity upon Cm treatment, the small ribosomal subunit showed higher local resolutions (Figure 3.4D-F). Focused refinement of the Cm-treated ribosomes with a SSU mask resulted in higher resolution (3.7 Å) compared to the untreated (Figure 3.4G-I). For both untreated and the first Cm-treated datasets collected with a Gatan K2 camera (pixel size on the sample 1.7 Å), core regions of the ribosome were already at the Nyquist limit of 3.4 Å.

To push the limit of cryo-ET and sub-tomogram averaging further, I collected the second Cm-treated dataset with the Gatan K3 camera. The pixel size on sample is 1.33 Å while the imaging scope still cover 60% - 90% of a *M. pneumoniae* cell. With the 30,774 extracted sub-tomograms, the Cm-treated 70S ribosome was determined at 3.0 Å overall resolution (Figure 3.5A-C). Focused refinements on 30S and 50S generated the subunit densities with 3.2 Å and 2.9 Å overall resolutions respectively (Figure 3.5D-F). For 50S core regions, local resolutions reached 2.8 Å.

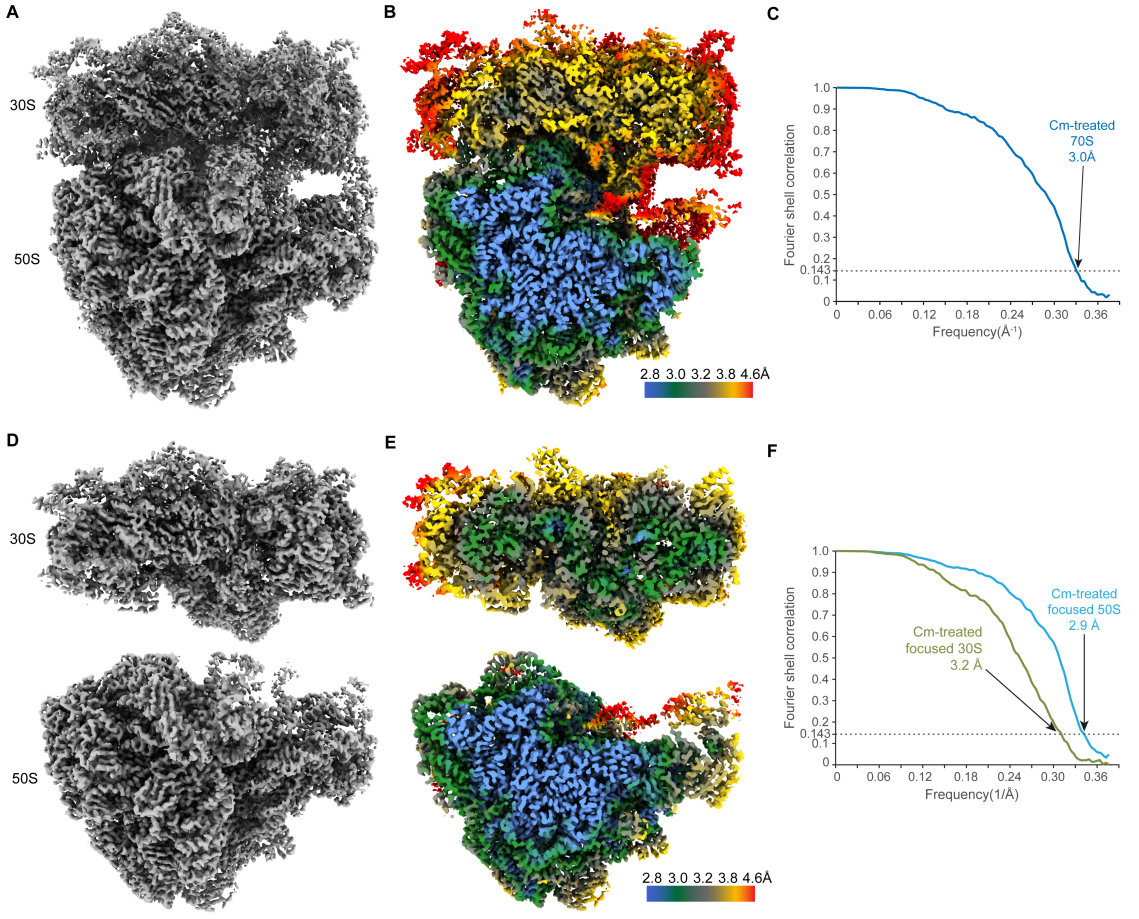

Figure 3.5 | Cm-treated ribosome determined at 3.0 Å inside the cell.

(**A**) 70S ribosome map averaged from 30,774 sub-tomograms in the second Cm-treated dataset collected with the Gatan K3 camera (dataset II, 20201209). Pixel size for the data is 1.33 Å. (**B**) Local resolution map, and (**C**) FSC curve for the refinement of (A). (**D**) Focused refinements of the Cm-treated 70S map on SSU and LSU respectively. (**E**) Local resolution maps, and (**F**) FSC curves corresponding to the focused refinements as shown in (D).

3.3.3 Factors behind high resolutions

The B factor in cryo-EM image processing is derived to quantify decay in high-resolution information in the averaged structure, and the apparent $B_{overall}$ factor can be used to evaluate general quality of the data and data processing[186,313]. Different considerations throughout the experiment, including sample heterogeneity, stability, imperfect imaging and inaccurate alignments, can affect the calculated $B_{overall}$ factor. In the post-processing step, B factor based sharpening that uses a negative number to reweight and restore high-resolution features is widely used in single particle cryo-EM[313,314], and has been adopted in sub-tomogram averaging[27,225]. However, there is no systematic evaluation and comparison of B factors in sub-tomogram averaging. Taking advantage of the large and comprehensive data, I calculated the $B_{overall}$ factors in four representative cases, by plotting the obtained resolutions against the numbers of particles used for averaging[313]. The comparison highlighted several experimental factors enabling the high resolutions achieved in this thesis and hinted on future improvements in sub-tomogram averaging.

For sub-tomogram averaging before multi-particle refinement in M, the estimated $B_{overall}$ factor is as large as 1,290 $Å^2$, which is predicted to limit the final resolution to ~10 Å, regardless of large numbers of sub-tomograms used for averaging (Figure 3.6). This prediction aligns well with the experimental resolutions obtained for native untreated and antibiotic-treated datasets before doing M (Table 3.1). $B_{overall}$ factors dropped drastically after performing multi-particle refinement in M, meaning that much better resolutions can be obtained with the same number of sub-tomograms. $B_{overall}$ for the native untreated dataset was estimated to be 126 $Å^2$, with which the nominal resolution of 3.5 Å was reached by averaging approximately eighty thousand sub-tomograms (Figure 3.4 and 3.6). For the first Cm-treated dataset with the same Gatan K2 camera and microscope settings, $B_{overall}$ was estimated to be 90 $Å^2$ and the same 3.5 Å overall resolution was available with less than twenty thousand sub-tomograms. Such improvement is possibly related to the reduced structural heterogeneity and flexibility upon Cm treatment (Figure 3.4). For the second Cm-treated dataset, which was acquired with a smaller pixel size using the Gatan K3 camera, $B_{overall}$ decreased slightly to 86 $Å^2$ and the resolution was improved to 3.0 Å with about thirty thousand sub-tomograms.

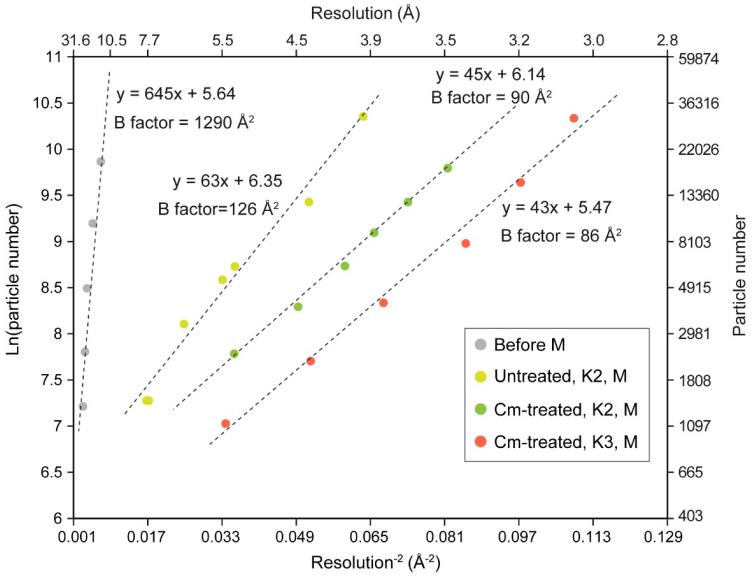

Figure 3.6 | Plotting of resolutions against numbers of averaged sub-tomograms.

Four representative cases were selected to calculate the $B_{overall}$ factors. The primary x axis shows the squared reciprocal of the resolution of the 70S ribosome, while the secondary x axis plots the resolution in Å. The primary y axis shows the natural logarithm of the number of sub-tomograms used for averaging, while the secondary y axis plots the number. Linear regression and slop calculation were done in MATLAB 2016b. The $B_{overall}$ factor equals to two times the slop of the fitted line[313]. The "before M" is using results from the first Cm-treated dataset, which performed similarly to other datasets before M. The $B_{overall}$ factor for the first and the second Cm dataset are close, but with the same number of particles, the resolution obtained with the second Cm dataset is higher. The possible reason is that the second dataset has a smaller pixel size and is not limited by the Nyquist limit, while density core regions in the first Cm dataset is at the Nyquist limit.

Multi-particle refinement implemented in M, which performed average-based refinement of tilt-series and particle related parameters simultaneously, significantly decreased $B_{overall}$. Tilt-series refinement (including tilt-series alignment, image and volume warping, anisotropic pixel size, *etc.*) resulted in the largest resolution improvement from ~10 Å to <5 Å and provided the basis for additional refinements such as CTF parameters (Table 2.1; also see the apoferritin results in the collaborative publication[234]. For sub-tomogram averaging, there are three tiers of alignments: frame alignment, tilt-series alignment and sub-tomogram alignment. Refinement of frame alignment in M (tilt movie) didn't significantly change the resolution, possibly because the SNR is not enough to support the refinement or the high-contrast gold fiducials already make frame alignment as accurate as it can be. Alignment of sub-tomograms in RELION, as suggested by Bharat *et al* who developed RELION's subtomo module, may not be the limiting factor for achieving high resolutions[226]. Therefore, average-based tilt-series refinement that improves the accuracy of the registration from 2D images to the 3D volume is the key to high resolutions.

The average-based tilt-series refinement not only improves tilt-series alignment parameters (tilt angles, *etc.*), but also models and corrects the specimen deformation (image and volume warping, *etc.*). Due to the increased specimen deformation with dose[185,234], it is no longer accurate enough to assume the tilt-series as a series of projections of the same object. One way to handle this problem is to divide the tilt-series into regions and each region is aligned separately. Such local alignment is available in Etomo but it requires a large number of gold fiducials[212]. For cryo-ET of isolated virus, gold fiducial particles can be mixed with the sample at high concentrations[315,316]. This, however, is not practically feasible for most in-cell studies where FIB thinning is required. Even for small *M. pneumoniae* cells, to have enough fiducials to perform local alignment is difficult because most fiducials are within a very thin solution layer on the cell surface. Average-based tilt-series alignment that use particles as "internal markers" was first successfully implemented in emClarity, and has been suggested to confer the major resolution improvement[232]. As M refinement is directly associated with the warping concept in Warp, it can achieve the local alignment effect similar to emClarity. Additionally, M's multi-particle system carefully models different image and volume warping[234], and M can simultaneously refine these with CTF and other parameters with a gradient-descent optimization, which makes it performs better than separate refinements[234].

Average-based refinement is a local optimization process and requires the initial tilt-series alignment to have good accuracy as the starting point. In the Spt-treated and PUM-treated datasets of relatively thinner ice but higher residual errors during initial tilt-series alignment in Etomo, the final resolutions they gave after M refinement are worse when compared to the untreated and the Cm-treated datasets (Table 3.1). It is also possible that due to the thinner ice or larger tension, the Spt-treated and PUM-treated datasets has higher specimen deformation. Nevertheless, the initial 2D-to-3D registration in these two datasets may not be efficiently corrected as other datasets.

The best $B_{overall}$ obtained in this study is comparable with routine single particle cryo-EM, which usually have $B_{overall}$ factors of 50-100 Å^2, and in the most ideal cases ~30 Å^2 to date[317,318]. Similar to single particle analysis, to have a lower $B_{overall}$ factor will be the key to improve resolutions in sub-tomogram averaging[313]. For abundant complexes like ribosomes, it is possible to achieve better than 3 Å resolutions with tens of thousands of particles at the $B_{overall}$ of 86 Å^2. For targets of the similar $B_{overall}$ factor, 8 Å resolution that is sufficient to do flexible fitting can be gained with less than 500 particles. Although it is anticipated that $B_{overall}$ for smaller or flexible complexes cannot reach the same level, it is possible to use ribosomes as the biological markers to first refine the tilt-series and CTF. For tomograms without enough ribosomes, to refine the tilt-series with other natural or engineered high-contrast features has been shown to be helpful[231]. The better refined tilt-series and CTF parameters can be used for extracting sub-tomograms of smaller or more flexible complexes within the same tomogram.

3.4 Atomic modeling of ribosomes in cells

3.4.1 The first atomic model for mycoplasma ribosomes

There was no reported atomic model of ribosomes of *M. pneumoniae*, or of any mycoplasma species within the Mollicutes class prior to this thesis. With the high-resolution ribosome densities resolved inside cells, I worked together with Swantje Lenz to build the first full atomic model of *M. pneumoniae* ribosome.

First, the best refined ribosomal subunit densities from focused refinements of ribosomes in Cm-treated cells were used to build atomic models for 30S and 50S respectively. The initial model was generated by combining *M. pneumoniae* homology models based on *E. coli* and *B. subtilis* ribosome models (see Materials and Methods). Iterations of PHENIX real-space refinement[276] and Coot manual adjustment[296] were performed to refine the 30S and 50S models. For most regions, except for the intrinsically flexible parts like L1 stalk, rRNA bases and protein side chains could be modelled with high quality (Figure 3.7).

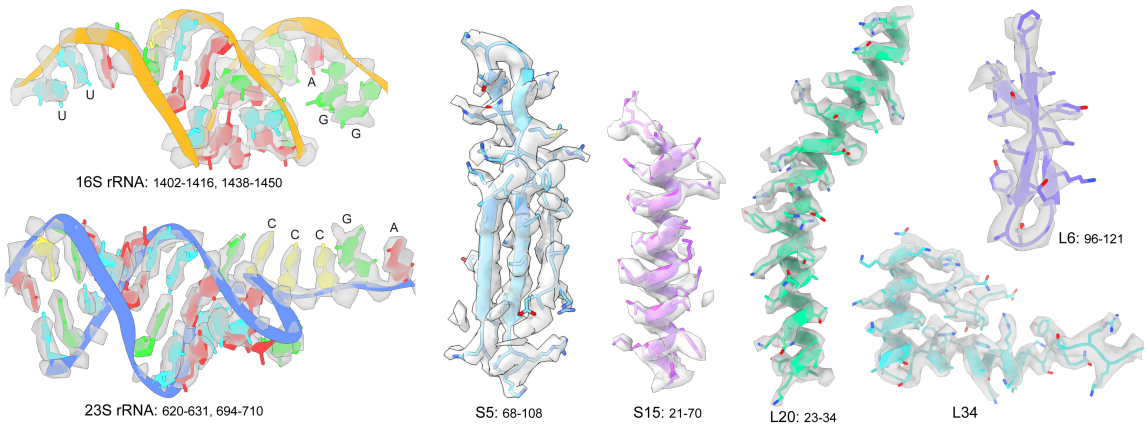

Figure 3.7 | Regions of the atomic model built in the density.

Examples of regions of the built atomic model fitted into the cryo-ET density maps of 30S and 50S. For most regions of rRNAs, bases (A, U, C, G) are nicely resolved in the density, as well as base pairing in the rRNA helices. For most ribosomal proteins, side chains can be clearly modelled into the density map, including regions of alpha helices, beta sheets and some loops.

The full atomic model of *M. pneumoniae* ribosome was built into the 70S density in either the Cm-treated or the untreated cells using the previously built 30S and 50S models as the starting model (Figure 3.8). Finally, 49 of all the 52 ribosomal proteins annotated in *M. pneumoniae* were built in the 70S ribosome model, except for the highly flexible proteins (*e.g.* L1 and L7/L12) that were not clearly resolved in the density. The ribosomal protein L9 mostly resides on the ribosome surface and was modeled in this flat conformation[99]. Ribosomal proteins S1, L25 and L30 are not found in *M. pneumoniae* proteome and consistently were not observed in the cryo-ET densities[36].

In general, the *M. pneumoniae* ribosome showed high structural similarity to the reported 70S ribosome models from other bacteria such as *E. coli* and *T. thermophilus*. Interestingly, new structural features that have not been revealed previously were discovered for the first time with the cryo-ET densities and built in the final model (detailed discussion can be found in the later section 3.4.3).

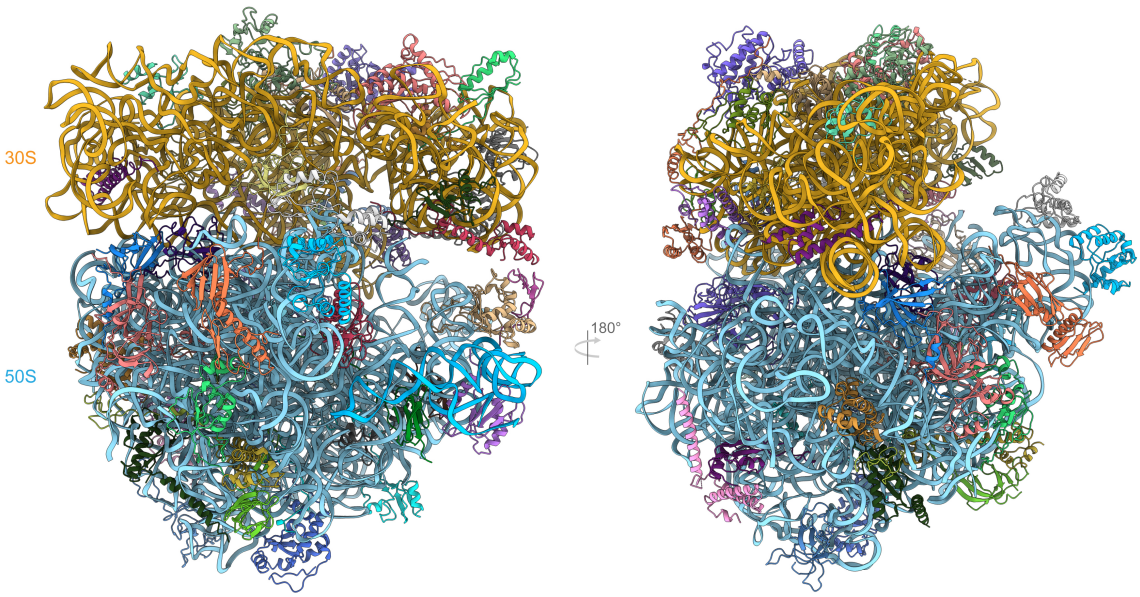

Figure 3.8 | Atomic model of *M. pneumoniae* ribosome.

Overview of the built *M. pneumoniae* 70S ribosome model based on the density map averaged from sub-tomograms computationally extracted from cellular tomograms. Different ribosomal RNAs and proteins are colored differently. Flexible ribosomal proteins L1 and L7/L12 were not resolved in the average density and thus not built in the model. The atomic modeling was performed together with Swantje Lenz with equal contributions.

3.4.2 Visualization of Cm binding to the ribosome inside the cell

The Cm molecule was clearly resolved in the 70S ribosome density in Cm-treated *M. pneumoniae* cells, with the obtained local resolution better than 2.9 Å (Figure 3.9A, B). The Cm binding site in *M. pneumoniae* ribosome is identical to its canonical site near the peptidyl transfer center as reported by previous *in vitro* structural studies[319-321]. In *M. pneumoniae* ribosome, bases A2459, C2460, G2468 and U2512 of the 23S rRNA in the large subunit form the main binding pocket, and the π–π stacking between the base of C2460 and Cm's nitrobenzene ring provides the main interaction (Figure 3.9C-E). These findings are consistent with the mutations that were found to confer chloramphenicol resistance[320].

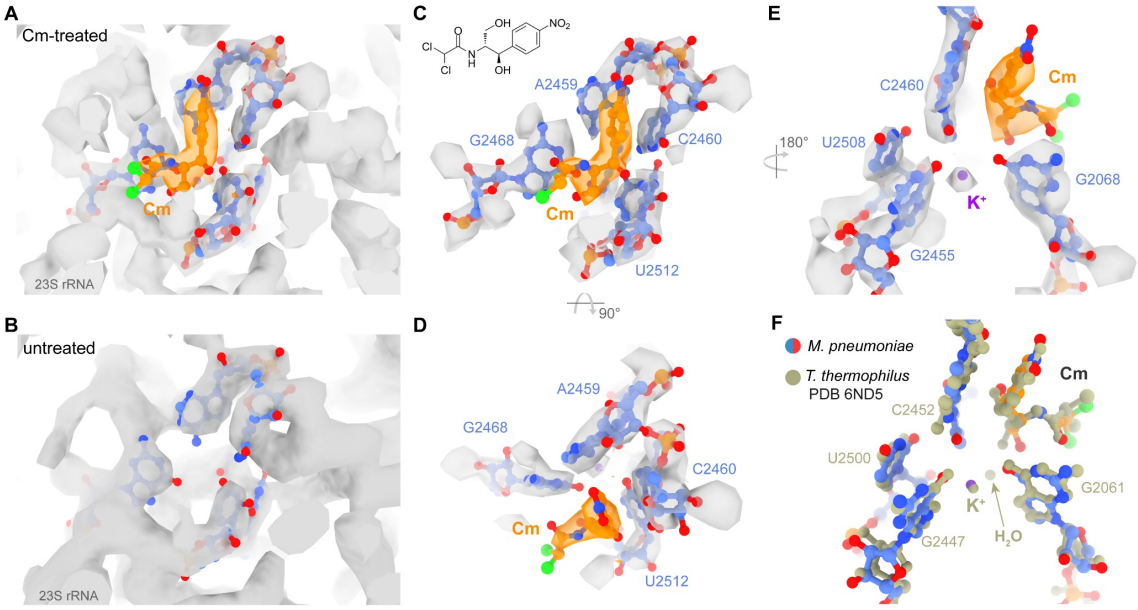

Figure 3.9 | Visualization of Cm binding to the ribosome inside cells with atomic detail.

(**A**) Local densities showing Cm bound to the PTC center in the ribosome in Cm-treated cells. The Cm molecule is mostly colored in orange, except for the oxygen atoms in red and the chlorine atoms in green. The density corresponding to Cm is also colored in orange. The surrounding bases of 23S rRNAs are colored in blue except for the oxygen atoms that are colored in red. Only the surrounding rRNA bases are shown. (**B**) Same region in the untreated ribosome as control. Only the surrounding rRNA bases as in (A) are shown. (**C-E**) Illustration of the Cm binding site. Insert: the chemical structure of Cm. The surrounding RNA bases (G2468, A2459, C2460, U2512) that form the main binding pocket are colored as before. In particular, as shown in (E), a potassium ion was resolved in the density. Bases of C2460, G2455 and G2068 are close enough (~2.8Å) to form coordination bonds to the potassium ion. The hydroxyl group of Cm was suggested[319] to form the coordination bond with the K^+, but it appears too far (~4.7Å) in the in-cell structure. (**F**) Superposition of the in-cell *M. pneumoniae* ribosome-Cm complex model with another 2.6Å *T. thermophilus* model by X-ray crystallography[321]. A water molecule modelled in the *T. thermophilus* structure (PDB 6ND5) was suggested to form coordination bonds with the potassium ion and the hydroxyl group of Cm, and such model fits well the in-cell *M. pneumoniae* model.

Strikingly, a K^+ ion was clearly resolved between the surrounding rRNAs and the Cm molecule, which is coordinated by bases of C2460, G2455, G2068 and possibly the hydroxyl group of Cm (Figure 3.9E). The potassium ion was first demonstrated in a 3.2 Å structure by Bulkley *et al* [319], and later confirmed by a 2.6 Å structure by Svetlov *et al* [321], both based on X-ray crystallography of isolated *T. thermophilus* ribosomes. The difference between the two *T. thermophilus* structures is that Bulkley *et al* proposed a direct coordination bond between Cm's hydroxyl group and the potassium ion (PDB 4V7W), while Svetlov *et al* modelled a water molecule between them (PDB 6ND5). The *M. pneumoniae* in-cell structure agrees more with the latter, supporting a water molecule exists in between (Figure 3.9F). In *E. coli* ribosome, the potassium ion was not modelled

in one X-ray cryptography structure (PDB 4V7T)[320], but the corresponding density appears clear in another cryo-EM map of 2.9 Å resolution (EMD-13180)[322]. Thus, the dependence of K^+ for Cm binding may represent a common feature in bacterial ribosomes[319,323,324]. Different from *in vitro* experiments that have concerns with regard to different ion types and concentrations in the buffer, the *M. pneumoniae* structure unambiguously demonstrated the role of a K^+ ion for Cm binding in cells. This is the first time to directly visualize antibiotic binding with atom-level information inside bacterial cells.

3.4.3 Ribosome protein extensions in *M. pneumoniae* ribosome

The new feature that was found in the *M. pneumoniae* ribosome is the sequence extensions of ribosomal proteins. The ribosomal protein extensions were first found in the cryo-ET density, which showed unexplained density after fitting the *B. subtilis* homology model[255]. The extra amino acids that are absent in the homology model, especially those of ribosomal proteins S6, L22 and L29 that form secondary structures and showed clear density in the map, were first built in the *M. pneumoniae* ribosome model (Figure 3.10).

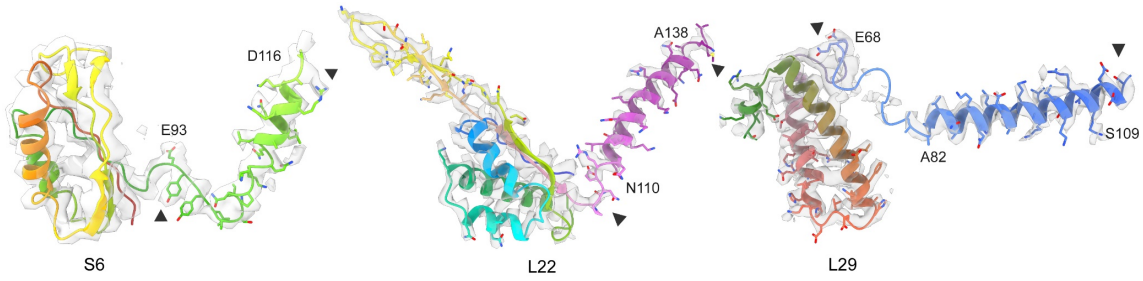

Figure 3.10 | Ribosomal protein extensions form secondary structures.

Part of the extended sequence in the ribosomal proteins S6, L22 and L29 can form secondary structures. The helices are clearly visible in the density map and were built in the model. The extended sequence for each protein is highlighted between two black arrowheads.

In addition, we found that 11 of the total 52 ribosomal proteins in *M. pneumoniae* have sequence extensions when compared to *E. coli* (10 when compared to *B. subtilis*; see Figure 3.11). The extensions were found in C- and/or N-termini of these ribosomal proteins, which usually consist of more than twenty additional residues (Figure 3.11). The extensions in most ribosomal proteins, except for S6, L22 and L29, were predicted to be disordered, and consistently do not contribute any significant densities in the map due to large flexibility. For the flexible extended sequences, only a few additional residues following the conserved structured part can be traced and modelled in the cryo-ET map.

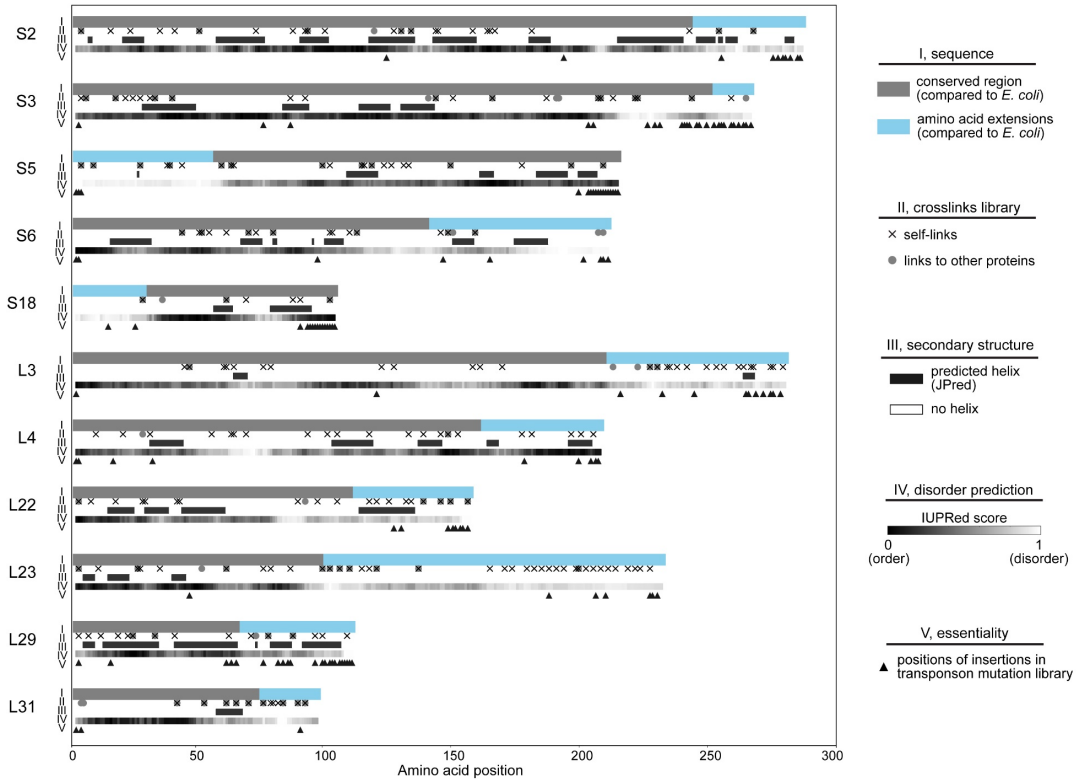

Figure 3.11 | Sequence extensions in *M. pneumoniae* ribosomal proteins.

Eleven ribosomal proteins in *M. pneumoniae* M129 were found to have extra amino acids (colored in light blue) in either the C-terminal or the N-terminal, compared to *E. coli* K-12. For these proteins, crosslinks from in-cell crosslinking mass spectrometry[255], secondary structure prediction, sequence disorder prediction, and transposon insertion mutation screening[42] results are displayed along the sequence. If compared to *B. subtilis*, 10 proteins also have extensions, except for L4. The analysis and figure were produced together with Maria Zimmermann-Kogadeeva within a collaborative project.

Interestingly, the C-terminal extension of S6 could be largely traced in the density map, which extends from the conserved part of S6 near the 30S-50S subunit interface to the mRNA entry site on top of the small subunit (Figure 3.12). Although the extension sequence beyond the helix-forming part (after residue D116) appeared to have weaker density, it showed a continuous density when using lower threshold for map visualization (Figure 3.12A). The modeling was confirmed by in-cell crosslinking mass spectrometry data (see section 5.2), which revealed several crosslinks to other ribosomal proteins along the traced path (Figure 3.12). Additionally, residue D167 was found to extend near the mRNA entry site, and crosslinked to the δ subunit of RNA polymerase, possibly because of transcription-translation coupling (for details about RNAP δ, see Figure 5.7 and 5.13 in Chapter 5)[255].

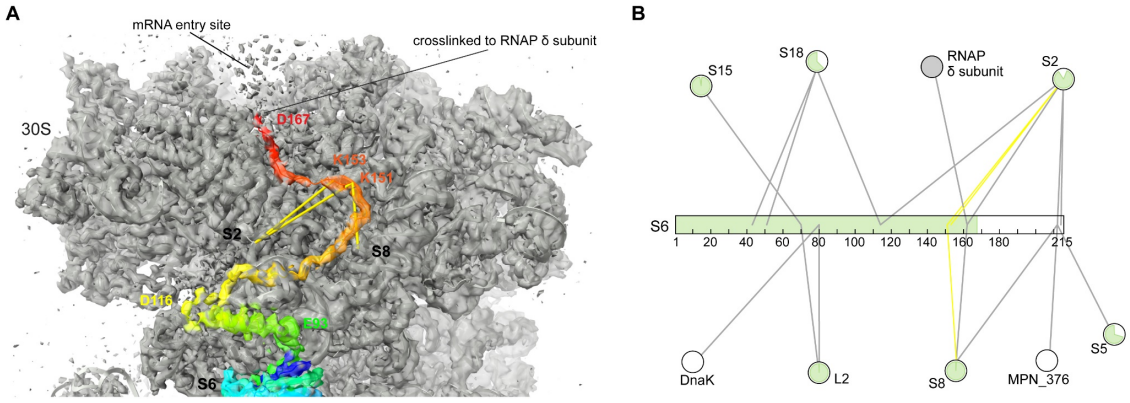

Figure 3.12 | Extension of the ribosomal protein S6.

(**A**) The C-terminal extension of S6 modelled in the density map. Residues D151 and D153 of S6 were found to be crosslinked to the 30S proteins S2 and S8 (yellow lines) according to the crosslinking data[255]. The C-terminal loop extends toward the mRNA entry site, and one residue D167 is crosslinked to RNAP δ subunit. (**B**) The crosslink network of S6. Details about the in-cell crosslinking mass spectrometry can be found in section 5.2 in Chapter 5. Figure prepared together with Swantje Lenz for a collaborative manuscript.

All extensions are solvent-exposed on the ribosome surface and placed away from the active sites (Figure 3.13). However, the eleven ribosomal proteins with extensions are mostly localized near critical regions such as the mRNA entry site or the nascent peptide exit site. On the small subunit, the ribosomal proteins S2, S3 and S5 surrounding the mRNA entry site all have extensions of various lengths (Figure 3.11). The conserved part of S6 is positioned near the subunit interface, but its C-terminus extend towards the ribosome mRNA entry site and was found to crosslink to RNA polymerase (Figure 3.11 and 3.12). These four ribosomal proteins form the main contact interface between the ribosome and RNAP during transcription-translation coupling in *M. pneumoniae*[255]. The extensions of these ribosomal proteins may have functional roles near the mRNA entry site of the ribosome. On the large subunit, three (L22, L23 and L29) of the four ribosomal proteins surrounding the nascent peptide exit site have sequence extensions, but the extensions are positioned away from the exit tunnel (Figure 3.11 and 3.12). The helix-forming extensions of L22 and L29 mainly contact with 23S rRNA sequence on the peripheral regions and seem unable to interact with the newly synthesized peptide during translation. It remains unclear what may be the functional relevance for these extensions near the nascent peptide exit site.

Crosslinks between the sequence extensions and other proteins were mapped according to the in-cell crosslinking mass spectrometry data, but there is no clear hint on any specific functions except for the transcription-translation coupling (Figure 3.11; also see materials associated to the publication[255]). Moreover, a previous transposon insertion mutation study showed that disruption of most extensions can affect survival or fitness of *M. pneumoniae* cells[303]. Some disordered extensions of ribosomal proteins in other bacterial species have been suggested to play roles in

ribosome biogenesis or assembly[325-328]. It is, however, impossible to conclude whether the essentially in *M. pneumoniae* is related to the disrupted ribosome assembly, because a new sequence is inserted in the transposon mutation study[303]. Interestingly, ribosomal protein extensions were suggested to be the main driving force for the formation and coevolution of the protein interaction network in ribosomes[326]. It is possible that the ribosomal protein extensions are not directly involved in the translation process but they may play certain accessory roles inside cells.

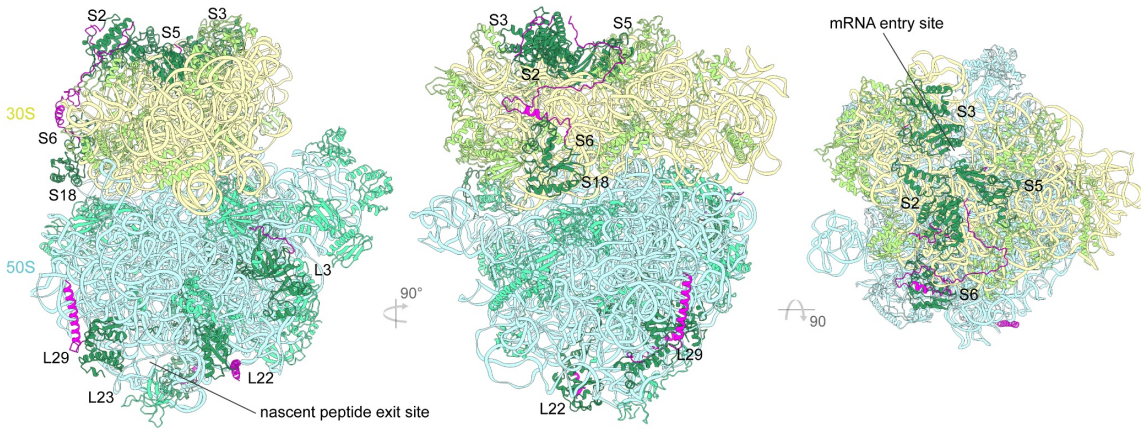

Figure 3.13 | Distribution of the ribosomal proteins with extensions.

Ribosomal proteins with sequence extensions are highlighted in the *M. pneumoniae* ribosome model, with the conserved part (dark green) and the traceable extension sequence (magenta) colored differently. Other ribosomal proteins are colored in lighter colors. For most extensions, except for those of S6, L22 and L29, only a few extra amino acids following the structured part could be modelled.

3.4.4 Ribosome protein extensions exist throughout bacteria

Sequence extensions in ribosomal proteins were found throughout the bacteria domain (Figure 3.14), possibly constituting a source of ribosome diversity in bacteria[54,329]. For the eleven ribosome proteins found to have sequence extensions in *M. pneumoniae*, bioinformatics analysis of 4,396 representative bacterial strains showed that the percentages of strains having extensions (at least 20 more amino acids compared to *E. coli* K-12; for details see Materials and Methods) varies largely. For L4, most (78%) bacterial strains were found to have extensions compared to *E. coli*. The C-termini of L4 in Gammaproteobacteria to which *E. coli* belongs appear different from other bacteria, which resulted in multiple sequence alignment that identified extensions in most bacteria. For the other ten ribosomal proteins with extensions, the percentages of species with extensions varies from 1% to 44% (Figure 3.14). Thus, ribosomal protein extensions represent a common feature across bacteria.

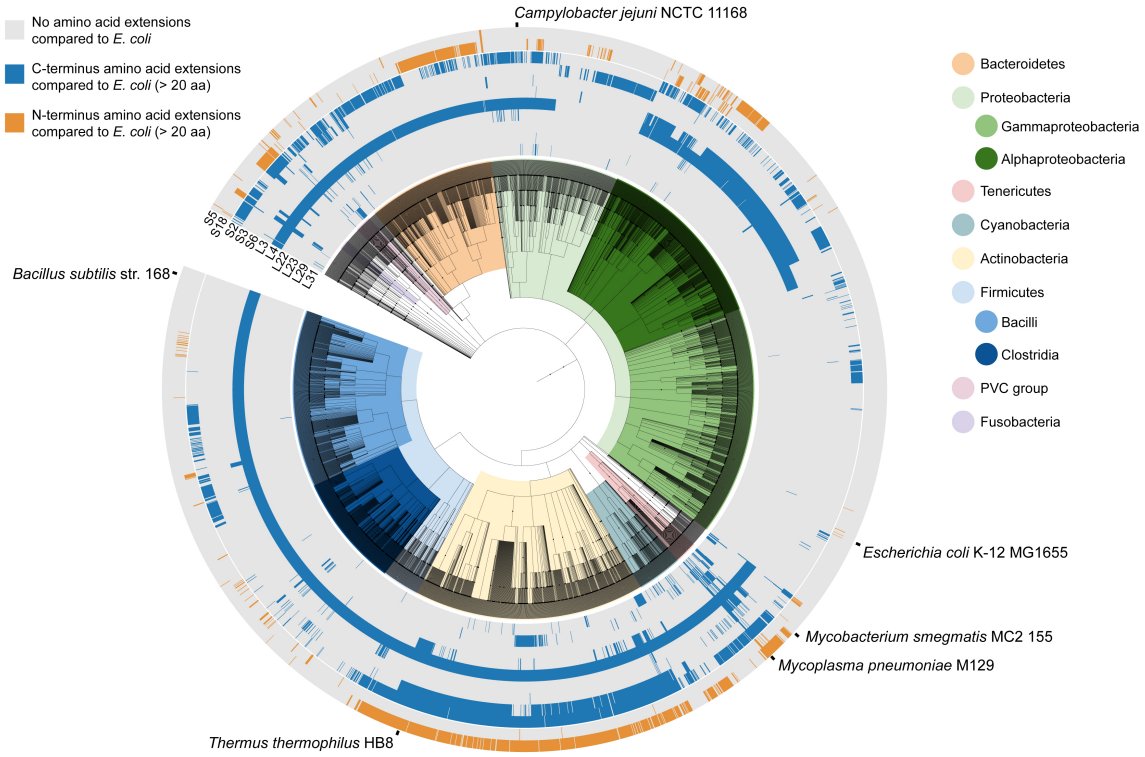

Figure 3.14 | Ribosomal protein extensions widely exist throughout bacteria.

The analysis and figure were made by Maria Zimmermann-Kogadeeva within a collaborative project. 4,396 representative bacterial stains selected from the bacterial domain were included and their sequences of the eleven proteins found to exhibit extensions in *M. pneumoniae* were compared to *E. coli* K-12.

The identified protein extensions are not unique to any bacterial sub-groups (Figure 3.14). Their occurrence appeared to be more frequent in the Mollicutes class or the Tenericutes phylum that contains *M. pneumoniae* and other mycoplasma species. On the other hand, ribosomal protein extensions in some bacteria such as *Mycobacterium smegmatis*, are not restricted to the eleven proteins identified in this work[330-332]. Beyond bacteria, parasitic protozoans are reported to have ribosomal protein extensions in comparison to other eukaryotes, and it is assumed that the extensions are related to their parasitic lifestyle[333]. In *M. pneumoniae* with severe genome reduction, the benefits of having these extensions remain unclear. It is hypothesized that the extensions may be related to *M. pneumoniae*'s parasitic lifestyle as a human pathogen[29].

Chapter 4: Translation landscapes in cells

Translation carried out by ribosomes and translation factors is one of the core processes in all living cells. To structurally dissect the dynamic process inside cells, I performed extensive classification to sort out ribosome intermediate structures during translation. The large and complete 70S ribosome sub-tomogram datasets provided the unique opportunity to investigate compositional and conformational changes of the translation machinery in cells.

4.1 Structural dynamics of the translation machinery

4.1.1 Considerations of the states of frozen cells

Structural dynamics of ribosomes along their functional trajectory, as well as the distribution of intermediate states, are expected to be well-preserved in frozen-hydrated *M. pneumoniae* cells. The sample preparation was optimized to avoid potential perturbations (see section 3.1.1). Cells were grown on grids at 37 °C in rich medium and were kept in the fast-growing phase. Active translation activities are going on inside the cells before plunge freezing. A manual plunger without an environmental chamber was used, and the grids were unavoidably exposed to room temperature (20-25 °C) during the blotting step. Whether this short (~5 seconds) exposure causes a temperature change on the grid is indeterminable. The ice thicknesses of *M. pneumoniae* cell samples is between 100 and 200 nm, a range suitable for rapid freezing. The vitrification time (0.1 millisecond)[188] is much shorter than major know translation steps (millisecond scale)[113,334,335].

4.1.2 Extensive classification of translation states

I designed a hierarchical, exhaustive classification workflow (Figure 4.1; also see Materials and Methods), based on the maximum-likelihood classification approach implemented in RELION[226,314]. The study aimed not only to classify as many ribosome intermediate structures as possible, but also to determine their distribution in a statistically reliable way. Classification settings in RELION were extensively tested and optimized for this purpose. Focused classification with a spherical mask based on the presence of associated factors gave more consistent results than classifying on the continuous conformational changes between the small and large ribosomal subunit. Parallel classification runs were performed to mitigate variations of individual RELION jobs (Figure 4.2A-G). The 101,696 ribosome sub-tomograms extracted from 356 native untreated cellular tomograms were first sorted into two tracks: the full 70S (76%) and the free 50S (24%). The free 50S were further divided based on the association of the ribosome recycling factor (RRF). The 70S ribosomes were further classified with two more tiers of focused classifications on the tRNA pathway and elongation factor binding region, respectively, resulting in 13 ribosome classes (Figure 4.1). Each class was assigned a unique class name indicating the elongation factor and

tRNA occupancy states and a unique number identifier. For example, class "EF-G, P, E" means that the ribosome contains EF-G, and P- and E-site tRNAs.

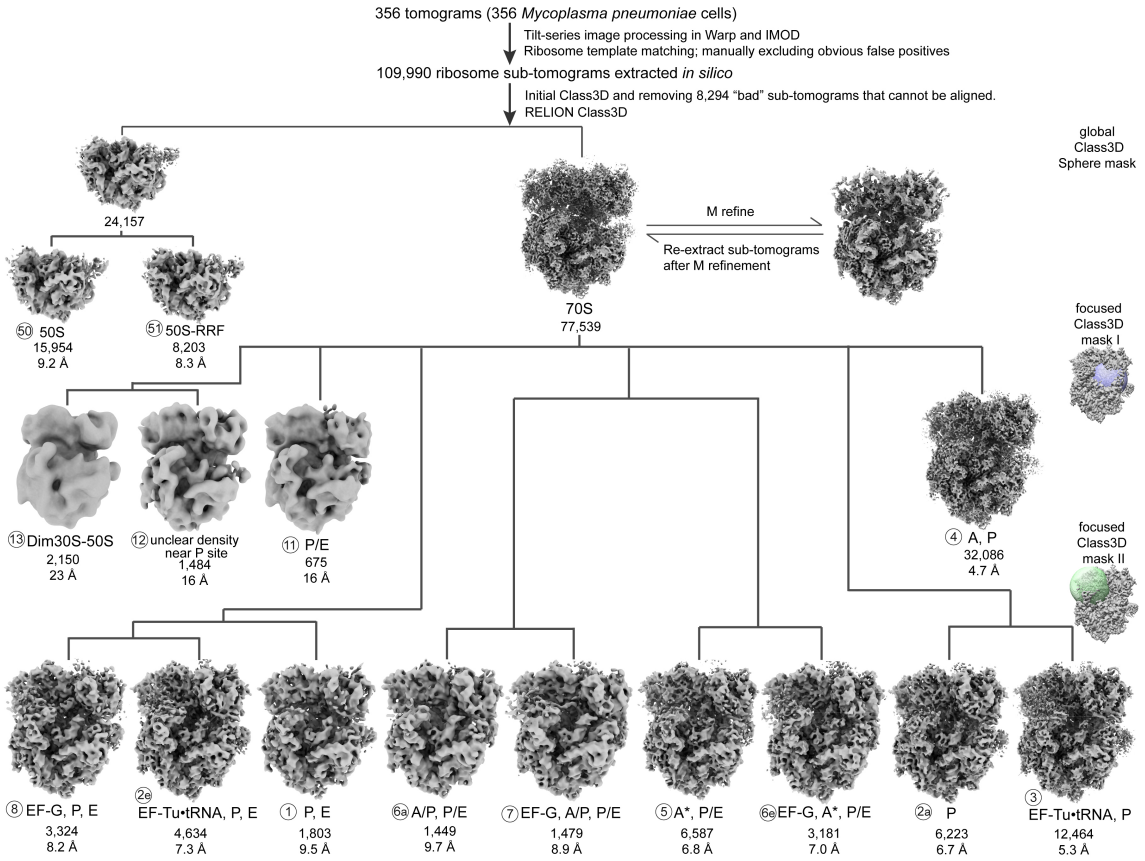

Figure 4.1 | Classification and refinement of ribosomes in native untreated cells.

The image processing and sub-tomogram analysis workflow of the native untreated cells dataset is summarized. The ribosome classification process can be divided into three tiers: global classification (sphere mask), focused classification on the tRNA pathway (mask I), and focused classification on the binding region of elongation factors (mask II). For every class, a unique number identifier, a class name indicating the occupancy states of elongation factors and tRNAs ("EFs + tRNAs"), the number of sub-tomograms in the class, and the final resolution are provided.

I further validated the unbiased representation of the extracted ribosome sub-tomograms and robustness of the whole classification process by plotting the class distribution against the cross-correlation scores determined in template matching (Figure 4.2H, I). The extracted ribosome sub-tomograms retrieved the vast majority of 70S ribosomes and a proportion of free 50S in the cellular tomograms (Figure 4.2H). The ribosome localization by template matching did not to introduce bias for 70S ribosome classification (Figure 4.2I). Therefore, the obtained classes and their relative abundance provided an accurate view of the translating ribosome population in native *M. pneumoniae* cells.

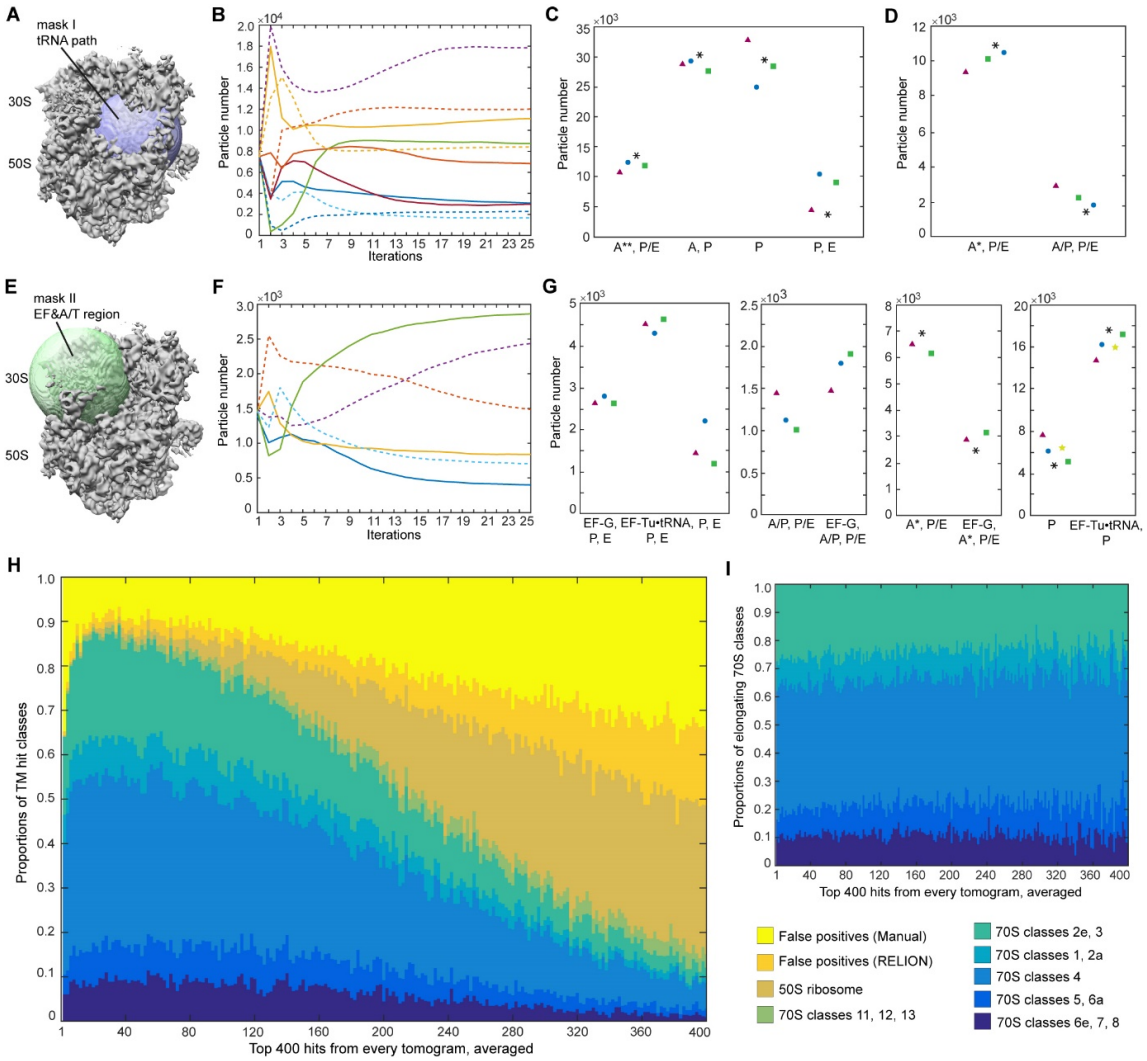

Figure 4.2 | Validation of the ribosome classification and detection.

(**A**) Mask I for focused classification on the path of tRNAs. (**B**) Illustration of one representative classification job with mask I. Numbers of sub-tomorgams in each class were monitored for 25 iterations. Classes were first grouped according to tRNA occupancies ("A**, P/E", "P, E", "P", "A, P"). (**C**) Results of four parallel classification jobs as in (B). (**D**) Results of the following jobs to classify the "A**, P/E" class into "A*, P/E" and "A/P, P/E" classes. (**E**) Mask II for focused classification on elongation factors and A/T-site tRNA binding region. (**F**) Numbers of particles in each class against classification iterations in one representative job with mask II. (**G**) Results of three parallel classification jobs with mask II. (**H**) Distribution of ribosome classes against the cross-correlation coefficients from template matching. For each of the 356 untreated tomograms, the top 400 hits were extracted and ranked by their cross-correlation scores. False positives include membranes, carbon edges and extracellular hits and bad particles that could not be aligned. For better visualization, 70S classes with similar overall structures were grouped, as indicated by different colors. (**I**) Same plotting as in (h), but showing only classes in the elongation phase. The percentages of 70S classes stay stable against cross-correlation coefficients, indicating that the classification is not biased by template matching.

4.1.3 Ribosome class structure refinement and modeling

Structure refinement and post-processing of the ribosome classes were performed in RELION (Materials and Methods). Detailed results are summarized in Figure 4.2 and 4.3. In total, thirteen 70S ribosome classes were determined, mostly at resolutions ranging from 4 to 10 Å; two free 50S classes were determined at 8-9 Å resolutions. For the 70S classes, two classes have relatively low resolutions and could not be reliably interpreted. Ten well-resolved 70S ribosome classes were determined, which were later interpreted to be intermediate structures during translation elongation.

Judging from the local resolution maps (Figure 4.3D-G), structural variance within each class still exists, especially for flexible regions like the L7/L12 stalk. However, no additional classes could be resolved with further exhaustive classification. The classes that were captured possibly represent states that are more populated or more structurally distinct, and may be averages of several close intermediate structures[334]. It is also recognized that rare states, such as those of the initiation and termination phases, are not be identified as distinct classes possibly due to their low abundancies[79,113].

Models for all classes were built by flexible fitting of the previously built atomic models of the two ribosomal subunits and homology models of other new factors (Figure 4.4; Materials and Methods). Most densities in each class could be nicely explained except for the intristically flexible regions, such as the L7/L12 stalk.

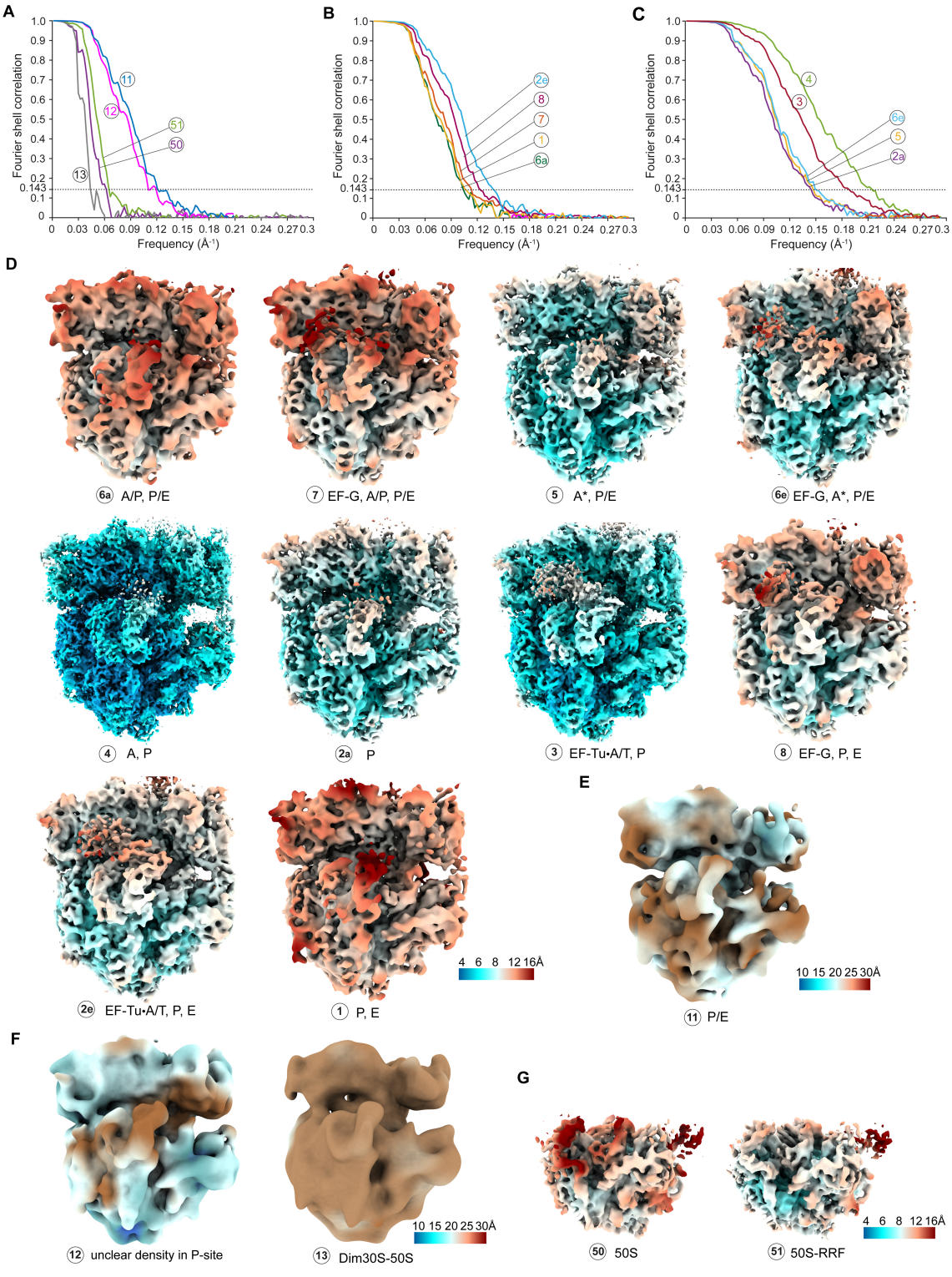

Figure 4.3 | FSC curves and local resolution maps for classes in native untreated cells.

(**A-C**) FSC curves for refinements of all 15 classified structures in native untreated cells. (**D-G**) Maps of the classes colored by local resolutions.

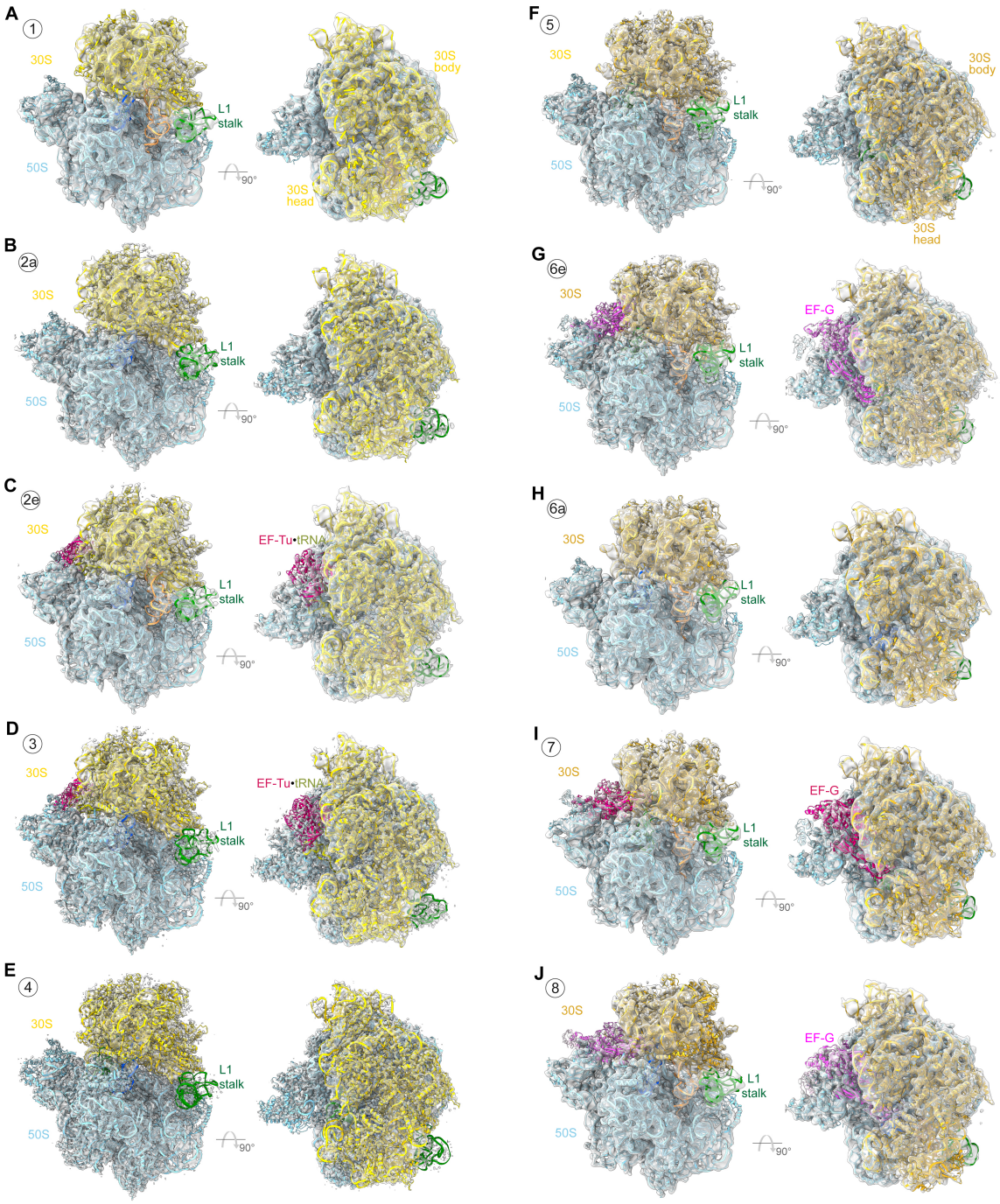

Figure 4.4 | Model building for ribosome classes in native untreated cells.

(**A-J**) Models built into the ten ribosome classes in the elongation phase. Models were built by flexible fitting using the Namdinator server[304]. I and Swantje Lenz contributed equally to this model building part of this work. For each class, the unique number identifier (as introduced in Figure 4.1) is displayed for tracking. The flexible L1 stalk is colored in green.

4.1.4 Reconstructing the complete translation elongation cycle

The ten 70S intermediate structures identified in the elongation phase could reconstruct the complete translation elongation cycle (Figure 4.5A). Major steps of the decoding, peptidyl transfer and translocation in the elongation cycle, characterized by different elongation factor and tRNA binding states and conformational changes of the ribosome (especially the small subunit), were clearly resolved. The elongation cycle reconstructed using the in-cell dataset is consistent with the translation model proposed previously based on various *in vitro* studies[113,116,121]. The ten elongation classes account for 72% of all extracted ribosomes, in agreement with the expectation that most 70S ribosomes are engaged in the translation elongation phase. The elongation time for translation of one protein can be tens of seconds or longer, while the time of other translation phases (initiation, termination and recycling) is in the range of milliseconds to several seconds[105,113,119,122].

The structures demonstrated continuous compositional and conformational changes of the translation machine, which include 30S body rotation, 30S head swiveling, tRNA movement through the A-P-E sites, L1 stalk dynamics, and elongation factor coordination (Figure 4.5). The "A, P" class is the most populated state and was determined at the highest resolution, showing clear densities for mRNA, A- and P-site tRNAs, and the nascent peptide chain (Figure 4.5B). In the following states, I observed the oscillation of A- and P-site tRNAs into the hybrid A/P and P/E state coupled with 30S rotations (Figure 4.5C, D), similar to previous reports[92,111,112,336,337]. Before the full hybrid state ("A/P, P/E"), a partial hybrid "A*, P/E" state was determined (Fig. 2a), where only A-site tRNA marginally moves toward the P site[92,336]. Interestingly, EF-G was found to bind to the ribosomes with either partial hybrid or full hybrid tRNAs. In the former case ("EF-G, A*, P/E"), EF-G is less extended and its domain IV does not overlap with the A site compared to the latter ("EF-G, A/P, P/E"), possibly representing an intermediate state transiting to the fully extended states[115,121]. Two EF-Tu-associated structures were determined, with and without the E-site tRNA (Figure 4.4 and 4.5). This suggested that disassociation of the E-site tRNA and binding of EF-Tu•tRNA are uncoupled. Taken together, the classified structures captured major steps during translation elongation in cells.

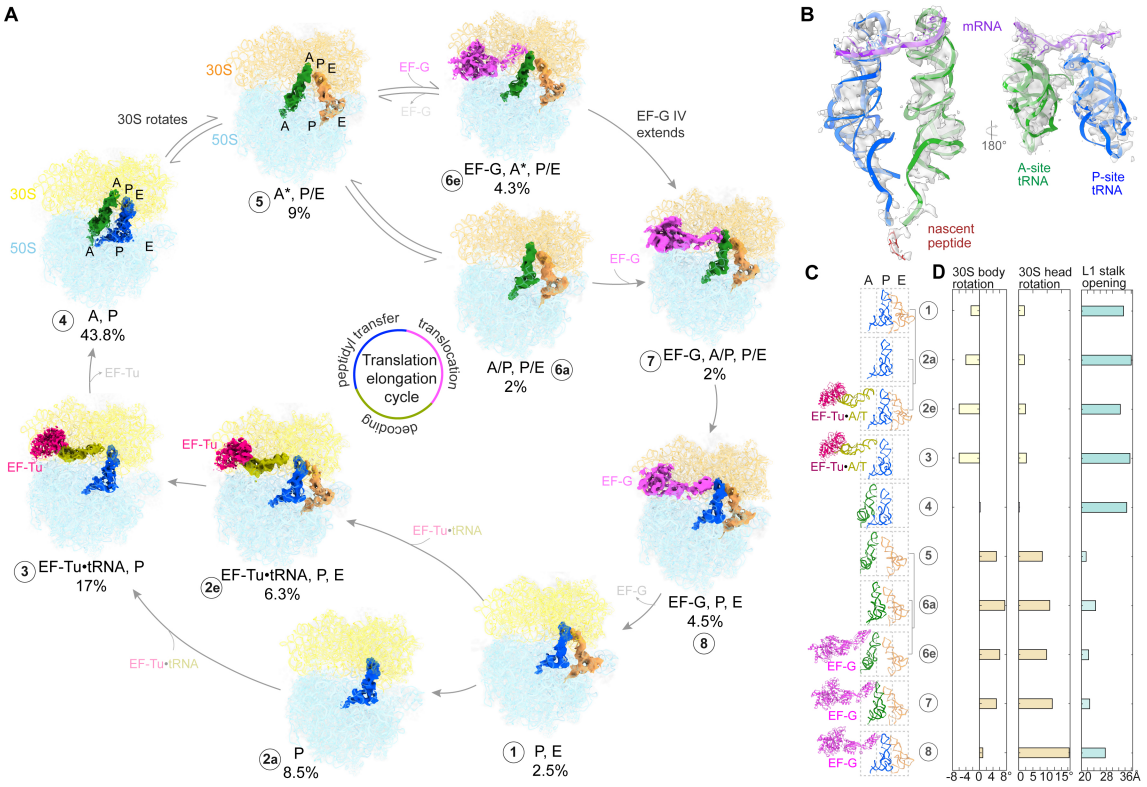

Figure 4.5 | Ribosome classes reconstruct the translation elongation cycle.

(**A**) Structures of the ten 70S classes in the translation elongation phase, characterized by different occupancy states of elongation factors and tRNAs. The proportions of different states were calculated based on the number of particles in the classes. Each class/state has the unique number identifier and class name for tracking. (**B**) The mRNA, tRNAs and the nascent peptide chain were well-resolved in the major "A, P" class. (**C**) Trajectory of the association of elongation factors EF-Tu and EF-G, and movement of tRNAs through the A-P-E sites. (**D**) Quantification of major conformational changes of the ribosome along translation cycles: 30S body rotation, 30S head swiveling, and L1 stalk dynamics.

The classification is pushed to the limit with the currently available methods. The "A, P" class could not be further classified to distinguish the states before and after peptidyl transfer. It is impossible to determine the state based on the density alone, although the density for the nascent peptide chain around the A site seems stronger than the P site (Figure 4.5B). No class with all the A-, P- and E-site tRNAs could be classified from the in-cell data, different from some *in vitro* results[77,78,107]. This does not exclude the possibility of ribosome intermediates with all A-, P- and E-site tRNAs. These, however, may be transient states with considerably low occurrence inside cells. This observation agrees well with one report showing that ribosomes are rarely (1.7%) occupied by three tRNAs during active translation[338]. Moreover, I did not determine any ribosome class with EF-G binding to the non-rotated state, consistent with the previous suggestion that it may be a less energetically favorable state[115]. For the two ribosome classes with EF-Tu, the blurred

local densities in the EF binding sites suggested flexibility of EF-Tu during the decoding step[100,107]. Classification tries with different settings, however, did not generate any new classes.

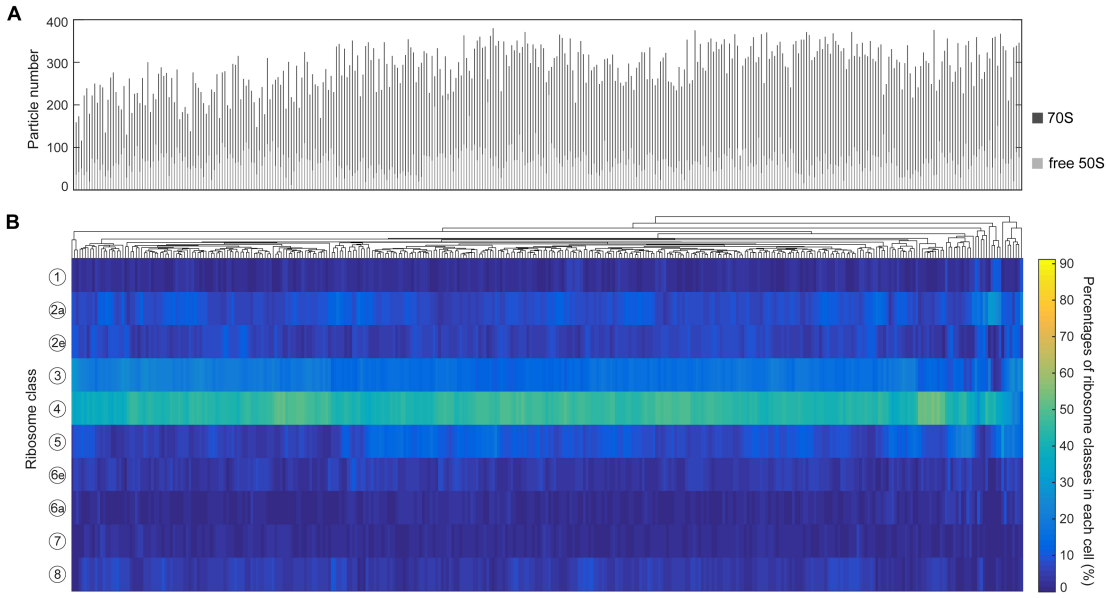

Figure 4.6 | Ribosome class distribution across 356 native untreated cells.

(**A**) Distribution of 70S and free 50S in single cells across the 356 tomograms of native untreated *M. pneumoniae* cells. (**B**) Percentages of 70S ribosome classes in the elongation phase in each of the 356 cellular tomograms were calculated and used for clustering analysis. The 356 cells exhibited similar translation elongation profiles.

4.1.5 Energy landscapes of translation elongation in cells

The ribosome classification not only determined intermediate structures in the translation elongation cycle, but also revealed their distributions along the functional trajectory inside cells. A living cell within a period of time can be considered as a non-equilibrium open system reaching the steady state[339]. In the case of translation, the steady state can be indicated by occurrence frequencies of different elongation states. For each cell, it is possible to obtain its translation elongation profile based on the proportions of the resolved ribosome classes. The finding of similar translation elongation profiles across the whole population of the 356 native untreated cells suggested that these cells reached similar steady states under the same experimental conditions (Figure 4.6). As discussed above, the cells were kept in the fast-growing phase in the rich medium. *M. pneumoniae cells* are known to have slow growth rate and long duplication time of more than eight hours[32,34,48]. These properties potentially contribute to the observation of the relatively low variations across the 356 native untreated cells.

The "A, P" class was found to be the most populated and may represent the most stable state within the translation elongation cycle in cells. The occurrence frequency decreases from the non-rotated

"A, P" state to the rotated partial hybrid ("A*, P/E") and full hybrid states "A/P, P/E", coinciding with the large conformational changes of the small ribosomal subunit. Binding of EF-G to the partial hybrid state decreases its occurrence and may promote transition to the full hybrid state[115]. Following translocation, the "P, E" class is of relatively low abundance, suggesting that the E-site tRNA is not stable and tends to disassociate quickly. This finding is consistent with previous studies showing rapid release of the E-site tRNA[338,340] , and can contribute to resolving the controversy of its disassociation time point[116,341].

Previous single-molecule fluorescence and single particle cryo-EM studies suggested energy landscapes of translation to be rugged[92,98,112,334-336,342]. The thermodynamic and kinetic properties of the ribosome can be influenced by many factors, including temperature[92,335], Mg^{2+} concentrations[88,89], EF-G[115,334]. Inside the cell, I hypothesize that the revealed translation landscapes result from both intrinsic properties of the translation machinery comprised of ribosomes and accessory factors, and the intracellular environments such as GTP and ion concentrations.

In *M. pneumoniae*, the concentrations of ribosomes, translation factors and substrates are considerably lower compared to these in *E. coli*. Through dividing the ribosome number by the cell volume, I estimated the concentration of 70S ribosomes in *M. pneumoniae* cells to be 7,400 ± 1,600 μm^{-3}. The number is significantly lower than the reported numbers in other bacterial species, for example, 15,000 μm^{-3} in *Spiroplasma melliferum*[123] and 27,000 - 60,000 μm^{-3} in *E. coli*[343]. Moreover, concentrations of other factors in *M. pneumoniae* (EF-Tu, ~40,000 μm^{-3}; EF-G, ~10,000 μm^{-3}; total tRNAs, ~14,500 μm^{-3}, according to database[36,344] updated till 2021) are lower than these in *E. coli* (EF-Tu, ~500,000 μm^{-3}; EF-G, ~75,000 μm^{-3}; total tRNAs, ~500,000 μm^{-3})[345-347]. Thus, it is possible that there are various translation landscapes for different cell types and conditions[92,98,133]. Application of the analysis introduced in this work to other cell types or conditions can help deepen the understanding of translation in the cellular context.

4.1.6 Ribosome classes involved in other translation phases

Three additional classes (70S with single P/E-site tRNA, free 50S and 50S in complex with the ribosome recycling factor) were interpreted to be intermediate structures in other translation phases (Figure 4.7). In the "P/E" class, one hybrid P/E-site tRNA could be resolved (Figure 4.7A), which possibly represent an intermediate state in the termination and recycling phases[79,348]. No other intermediate structures in these phases was captured. These results suggested that the next step that requires EF-G or RRF binding to the ribosome of "P/E" state to proceed is one of the rate-limiting steps in ribosome recycling, consistent with a previous study[122].

Free 50S could be classified into two classes based on the binding states of RRF (Figure 4.7B, C). In the "50S-RRF" class, domain I of three long helices was clearly resolved, but not domain II. There seem to be some densities around the E site, but they are too weak to be further classified or interpreted. Disassociation of RRF is supposed to be the last step in ribosome recycling[348]. The free 50S were not the target for the initial template matching and are therefore expected to be not fully recovered in the sub-tomogram dataset, I cannot confirm whether this step is one of the limiting steps in ribosome recycling.

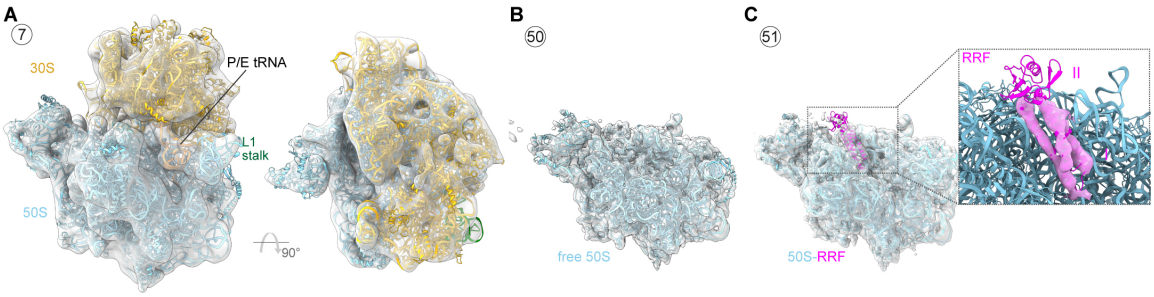

Figure 4.7 | Three other distinct classes in untreated cells.

(**A**) The ribosome with hybrid P/E-site tRNA. (**B**) Free 50S. (**C**) Free 50S in complex with the ribosome recycling factor. Domain I of RRF was clearly resolved while domain II was not, possibly due to its high flexibility during the interaction with EF-G.

The remaining two classes with low resolutions, *i.e.* a "Dim30S-50S" class and the 70S with "'unknown density in P-site", could not be unambiguously interpreted (Figure 4.1 and 4.3). The small ribosomal subunit in the "Dim30S-50S" class seems slimmer than normal 30S and there is no clear tRNA density between the two subunits. In another class, the density near the P site in the 70S ribosome looks different from a normal tRNA density.

4.2 Spatial and functional organization of ribosomes

4.2.1 Mapping ribosomes back into 3D cellular volume

One major advantage of cellular cryo-ET and sub-tomogram analysis over traditional structural approaches is that the generated structures can be mapped back into the tomogram (3D cellular volume). One tomogram in this work typically cover 60% to 90% of the cellular volume of a *M. pneumoniae* cell, and as described above, the vast majority of 70S and part of the free 50S ribosome in the cellular tomogram were retrieved. Combing the classification and refinement results, I obtained the spatial and functional organization of the ribosome community in the cell, with unprecedented completeness, resolution and functional detail.

The number of 70S ribosomes in one *M. pneumoniae* cell was estimated to be 300 to 500, which is slightly higher than the numbers estimated in previous cryo-ET studies[32,53]. The concentration of translating ribosomes is 7,400 ± 1,600 μm^{-3}, in line with *M. pneumoniae*'s long duplication time (as discussed in section 4.1.5). Ribosomes distribute throughout the cell, except for their exclusion from the attachment organelle region (Figure 4.8). The exclusion of ribosomes in regions of the attachment organelle was expected as it is a gigantic, multi-subunit protein assembly[40]. The central nucleoid region in bacterial cells such as *E. coli* was suggested to be largely ribosome-excluding[345,349]. Such separation, however, was not significant in *M. pneumonia* cells. The genome of *M. pneumoniae* is one of simplest among living cells and has been suggested to lack a defined nucleoid region[41,350].

4.2.2 Detection of polysomes

Ribosomes translating along the same mRNA form polysomes (also referred as poly-ribosomes) that can assemble closely in space[134,136,155]. Based on the spatial information (coordinates and rotations) obtained from ribosome localization and refinement, it is possible to detect the closely assembled polysomes. The general definition of polysomes is all ribosomes translating on one mRNA molecule. However, only when two ribosomes within the same polysome are close enough may they have direct influence on each other. In this sense, the detected polysome based on position proximity can reflect potential functional implications of polysome organization.

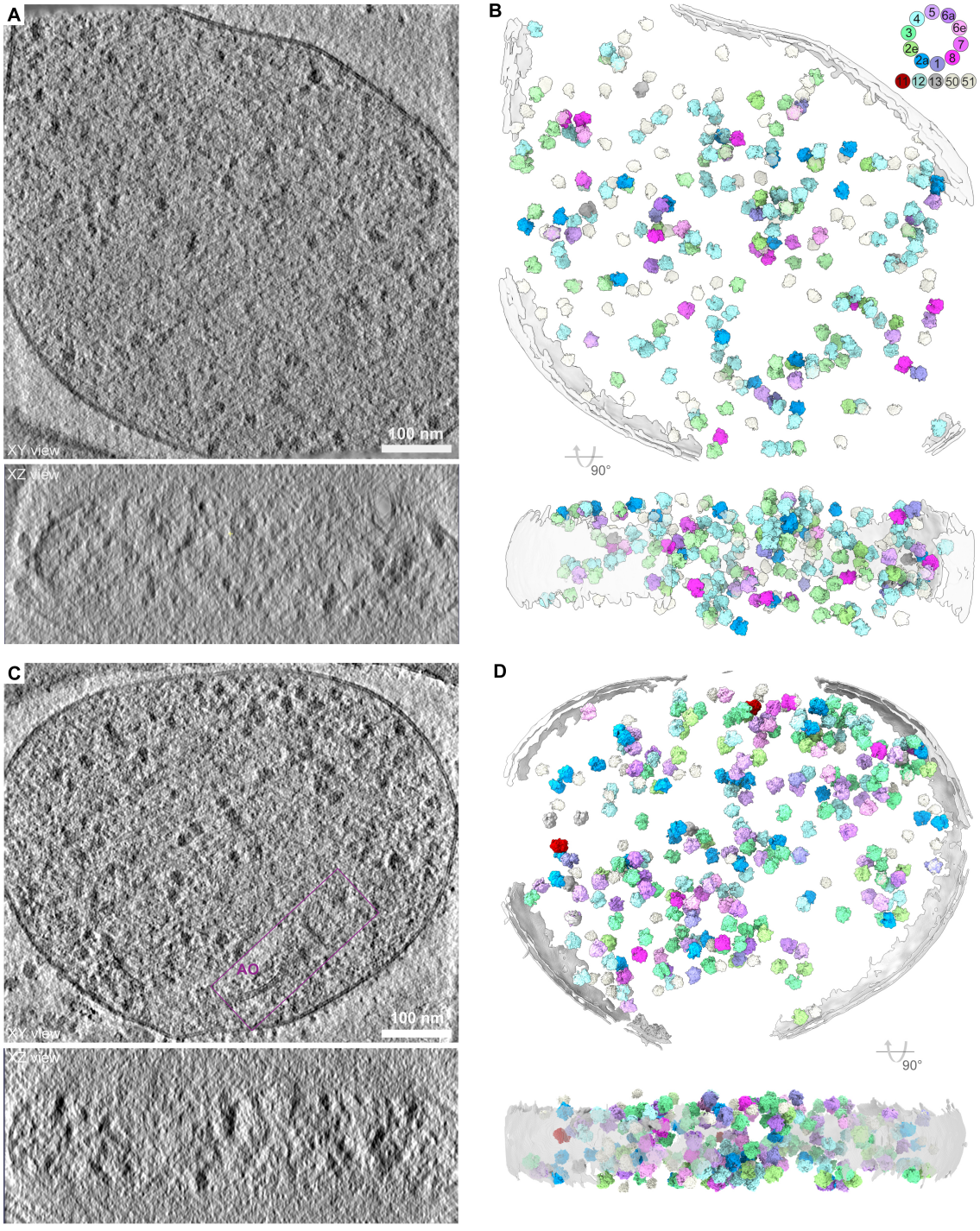

Figure 4.8 | Mapping ribosomes in two representative cells.

(**A-B**) Tomogram slice and 3D ribosome map of one *M. pneumoniae* cell. Different colors indicate different ribosome classes (insert). Each class is represented by the unique number identifier, as described in Figure 4.5. Cell thickness equals to ~200 nm. (**C-D**) Tomogram slice and ribosome map of another cell of relatively small cell size and ice thickness (~160 nm). AO, attachment organelle.

I detected polysomes based on a distance cutoff from the mRNA exit site of one ribosome to the mRNA entry site of the following ribosome with a custom-made MATLAB script (Materials and Methods). An example is provided in Figure 4.9 to illustrate the polysome detection in one cell. Compared to the ribosome center-to-center distance criterion, the script further considers relative orientations and sequence of ribosomes with the polysome. To select the most appropriate distance cutoff for polysome detection, I first calculated the distribution of distances between every ribosome and its neighboring ribosomes. The peak resulting from polysome formation was estimated to be between 3 and 6 nm (Figure 4.10A, B). Next, I tested different distance cutoffs ranging from 2 to 10 nm (Figure 4.10C), and visually checked the detection results in 3D. The distance cutoff was chosen to be 7 nm, which provided the best detection rate and accuracy.

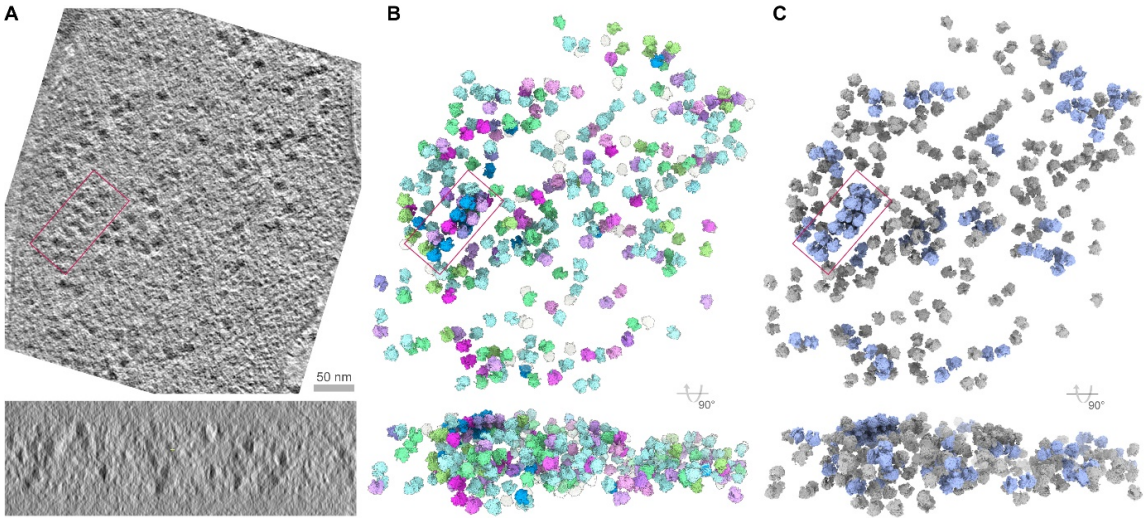

Figure 4.9 | Mapping of ribosomes and detected polysomes in one *M. pneumoniae* cell.

(**A-B**) Tomogram slice and 3D ribosome map of one *M. pneumoniae* cell. Different colors indicate different ribosome classes. (**C**) The ribosomes detected as polysomes are colored in light blue, while others in grey. One long polysome is highlighted by the red box.

The ribosomes annotated as polysomes account for 26.2% of all 70S ribosomes, consistent with the knowledge that polysomes in bacteria are usually of low abundance[351,352]. Most polysomes detected in *M. pneumoniae* cells are rather short, consisting of two to four ribosomes. This number range is in good agreement with the estimated numbers by a fluorescence microscopy work monitoring real-time translation activities in living cells[353]. Indeed, long compacted polysomes may not be economic in cells, especially small bacterial cells like *M. pneumoniae*, because they require high concentrations of elongation factors, tRNAs and GTPs in the local region to ensure efficient translation elongation. And ribosomes conducting translation within the compacted polysomes would have higher possibility to clash with each other and induce mistranslation, if not properly coordinated.

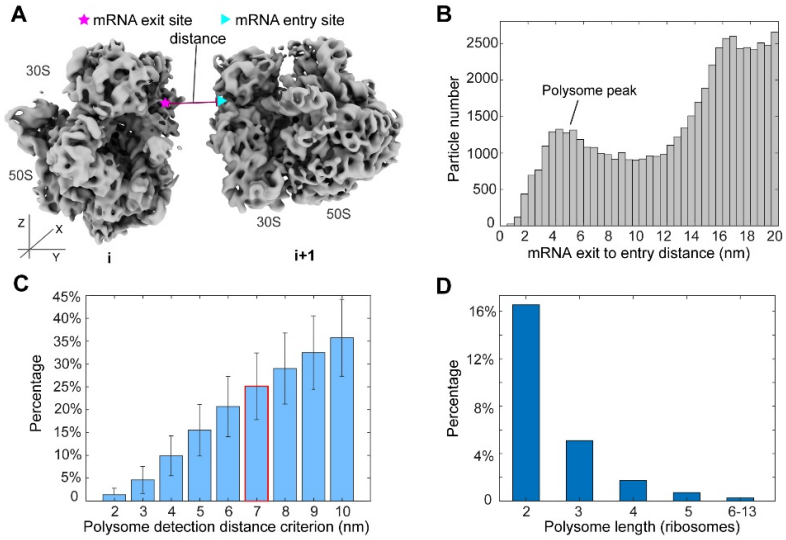

Figure 4.10 | Illustration of the polysome detection.

(**A**) The distance between the mRNA exit site of one ribosome (i) and the mRNA entry site of the following ribosome (i+1) was calculated and used to define whether the two ribosomes belong to one polysome. (**B**) Distribution of the distances between the mRNA exit site to the mRNA entry sites of neighboring ribosomes. The peak around 5 nm is attributed to polysome formation. (**C**) Percentages of 70S ribosomes annotated as polysomes with different distance cutoffs. Means and standard deviations were calculated across 356 cells. (**D**) Distribution of polysome lengths across the 8,641 identified polysomes in 356 native cells.

4.2.3 Organization of ribosomes within the polysome

For the detected polysomes, two basic arrangement patterns between adjacent ribosomes could be defined (Figure 4.11A, B), which are identical to the so-called "top-top" ("t-t", 78.5%) and "top-back" ("t-b", 21.5%) configurations described in the early work by Brandt *et al*[136]. There is a high degree of movement freedom between two adjacent ribosomes in the polysome. Relative to the mRNA exit site of the preceding ribosome (i), the following ribosome (i+1) may have different distances and orientations. The most suitable approach to represent such flexibility and to define the two configurations are rotations of the following ribosome relative to the preceding one (Figure 4.11C). The average distance between the mRNA exit site (i) and the mRNA entry site (i+1) was estimated to be 4.2 ± 1.4 nm for ribosomes pairs with the "t-t" configuration (N=9,100), and 5.4 ± 1.5 nm for "t-b" (N=2,491). Accordingly, ribosomes with the "t-t" configuration are more compacted than those with the "t-b" configuration.

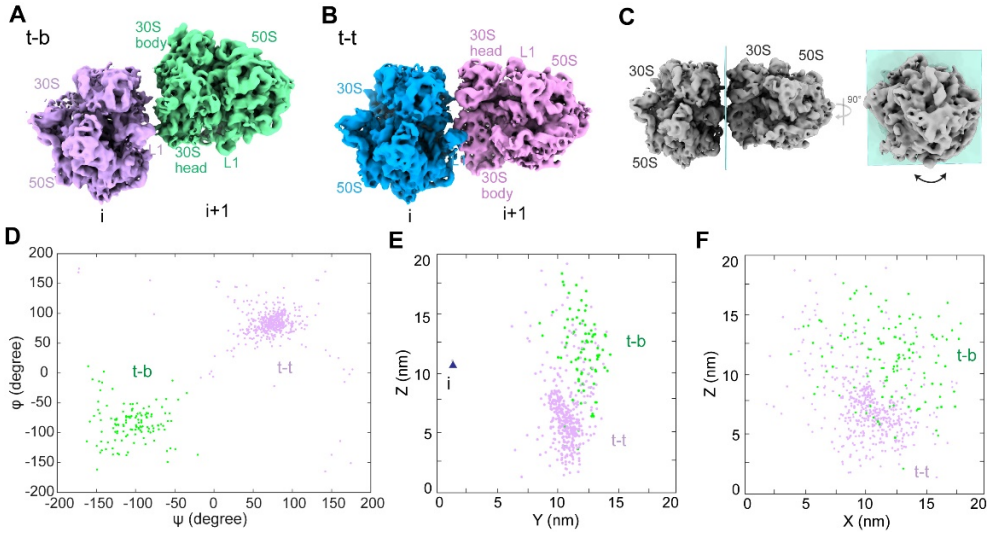

Figure 4.11 | Two arrangement patterns for adjacent ribosomes in the polysome.

(**A-B**) Two assembly patterns of adjacent ribosomes in polysomes. Here I adopt the names "t-t" and "t-b" from a previous publication[136] to name the two configurations. (**C**) The major flexibility of the following ribosome (i+1) relative to the preceding ribosome (i) is the rotation (up to 360°) around a plane perpendicular to the mRNA exit site of the preceding ribosome. (**D**) Rotations of the following ribosome relative to the preceding ribosome in polysomes. This rotation was used to define the "t-t" and "t-b" configurations. Results from 18 randomly selected tomograms are shown in this plot. (**E-F**) Positions of the following ribosomes (i+1, colored by different configurations as defined in (D)) relative to the preceding ribosome (i, triangle is the mRNA exit site) in polysomes.

Different topologies could be observed for long polysomes, from relatively loose association to tight packing with a helix-like configuration (Figure 4.12). The putative mRNA paths within the polysomes could be drawn and used to illustrate different topologies. The polysome architectures are identical to previous reported structures based on *in vitro* or *in situ* samples[134,136,137]. In the highly compacted polysome with a helix-like arrangement, all adjacent ribosomes are of the "t-t" configuration. By contrast, both "t-t" and "t-b" configurations could be found in polysomes that are less compacted. In polysomes of various topologies, the main contact was found to be within two adjacent ribosomes (i and i+1) and no direct interaction between other neighboring ribosomes (i and i±N, N>1).

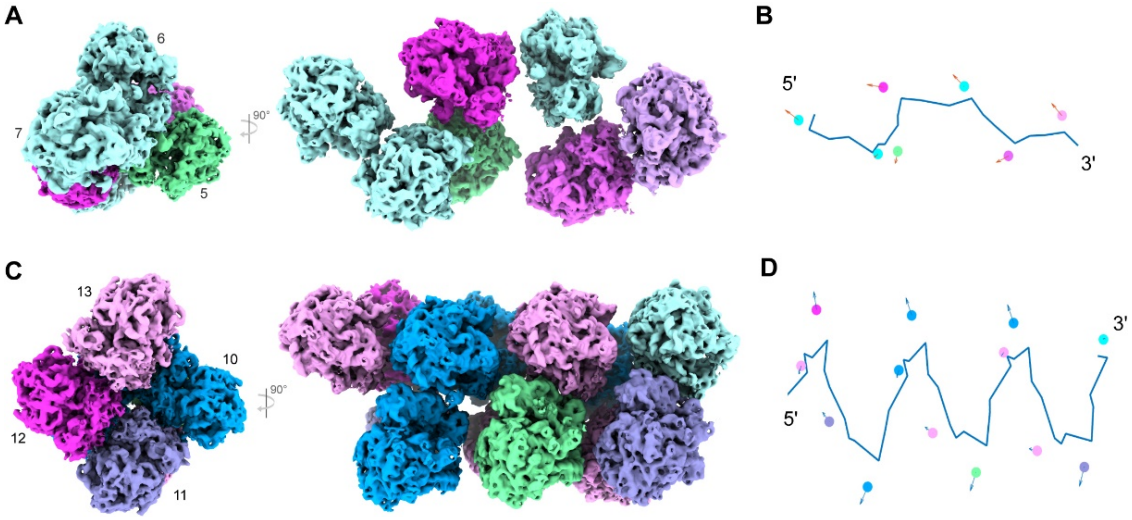

Figure 4.12 | Representative long polysomes of different topologies.

(**A-B**) One representative long polysome with loose configuration; and the putative mRNA path and nascent peptide vectors. (**C-D**) One tightly packed polysome with the helix-like topology, and the putative mRNA path and nascent peptide vector. Different colors indicate different elongation states as shown in Figure 4.8.

4.2.4 Coordination of translation elongation within the polysome

The polysome detection and the previous ribosome classification results enabled addressing a long-standing question in translation study: whether translation elongation is synchronized or coordinated within the polysome. First, the translation state distributions of ribosomes engaged within polysomes were compared to the total ribosomes. Though most states have equal occurrence frequencies, the state "A*, P/E" and "A/P, P/E" were found to be more frequent in polysomes compared to all ribosomes or mon-ribosomes (Figure 4.13). This suggested that organization into polysomes have has certain influence on translation elongation of the ribosomes.

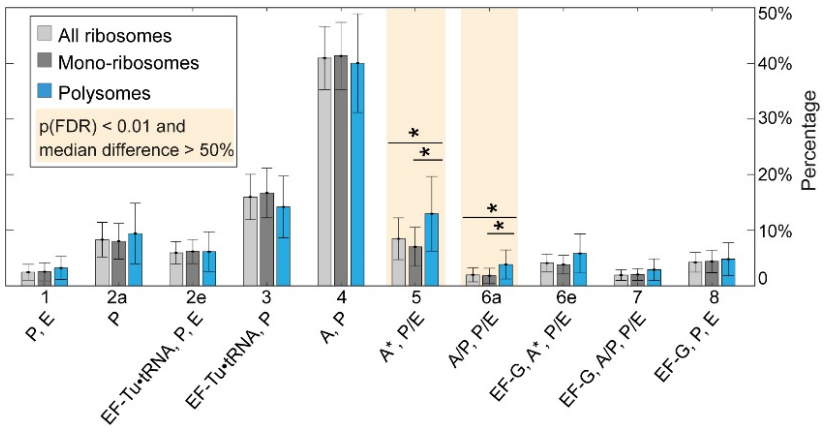

Figure 4.13 | Distribution of translation elongation states in polysomes.

The translation elongation states in polysomes were compared to all ribosomes and mono-ribosomes. Bar plots and whiskers are mean and standard deviation across 356 native cells. Highlighted in light yellow are

states for which the number in polysomes differs by more than 50% compared to all ribosomes or mono-ribosomes, and FDR-adjusted p-value calculated with Wilcoxon rank sum test < 0.01. The analysis and figure preparation were done together with Maria Zimmermann-Kogadeeva.

To examine possible coordination of ribosome elongation within the polysome, the frequencies of elongation state pairs of all adjacent ribosomes (the preceding ribosome i and the following i+1) in polysomes, were compared to the theoretical frequencies of random elongation state pairs calculated from the bulk distribution. The experimental frequencies of ribosome state pairs showed similar patterns to the theoretical frequencies of random state pairs (Figure 4.14A, B). This indicated that there is no global-level synchronization within the polysome, which is also seen from examples of long polysomes (Figure 4.12). By dividing the experimental numbers by the theoretical, it was found that the occurrence of elongation states in the preceding ribosome (i) and the following ribosomes (i+1) is asymmetric: states that requires elongation factor binding to proceed (states 1, 2a, 5 and 6a) occurs more frequently as the following ribosome (Figure 4.12C).

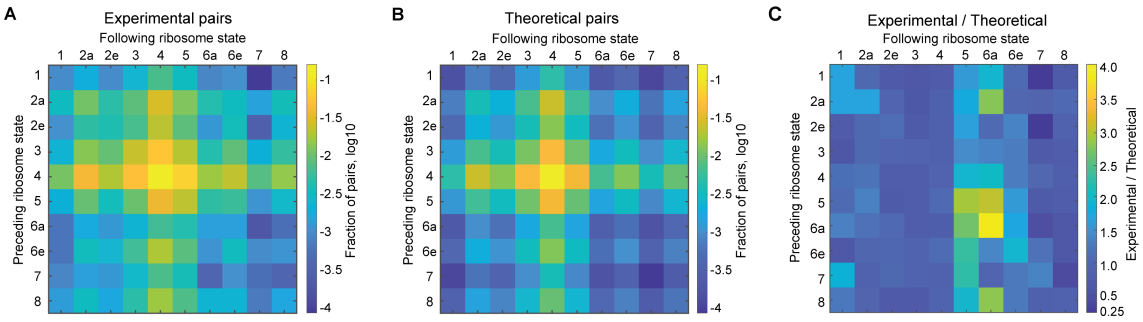

Figure 4.14 | Coordination of translation elongation states in polysomes.

(**A**) Occurrence frequencies of translation elongation state pairs of adjacent ribosomes in polysomes (the preceding i and the following i+1) calculated from the experimental data. (**B**) Theoretical frequencies of random elongation state pairs if there is no coordination within the polysome. (**C**) Comparing the experimentally observed frequencies of state pairs of adjacent ribosomes in polysomes against the theoretical probabilities of random pairs. The following ribosomes are more likely to be found in states with identifier 5 and 6a ("A*, P/E" and "A/P, P/E" respectively), compared to random organization. Translation states 1 and 2a ("P, E" and "P" respectively) also seem to have higher frequency, but are less significant.

A permutation test was then performed to validate the statistical significance of the detected differences (Figure 4.15; also see Materials and Methods). A polysome shuffling procedure enabled estimating the statistical significance of the differences in pair frequencies. The p-value represents the fraction of random permutations in which the state pair frequency was different from the experimental one (Figure 4.15A, B). The frequencies of shuffled state pairs were calculated from the polysome matrices after shuffling, which represent the distribution if different elongation states occur randomly in polysomes (Figure 4.15C). Comparing the experimental and the shuffled pairs, 22 state pairs for which the experimental values are significantly different from

the shuffled values were found (Figure 4.15B). Notably, 20 of these are related to states that require the binding of elongation factors (states 1 and 2a for EF-Tu, as well as states 5 and 6a for EF-G) to proceed in the elongation cycle. For four states (1, 2a, 5 and 6a), the number of pairs in which the ribosome of these states is a preceding ribosome is significantly lower than the expected number in random distribution, and the number of pairs in which the ribosome of these states is a following ribosome is significantly higher than random distribution (Figure 4.15B, D). In brief, the shuffling analysis proved the statistical significance of the asymmetric state distribution within polysomes.

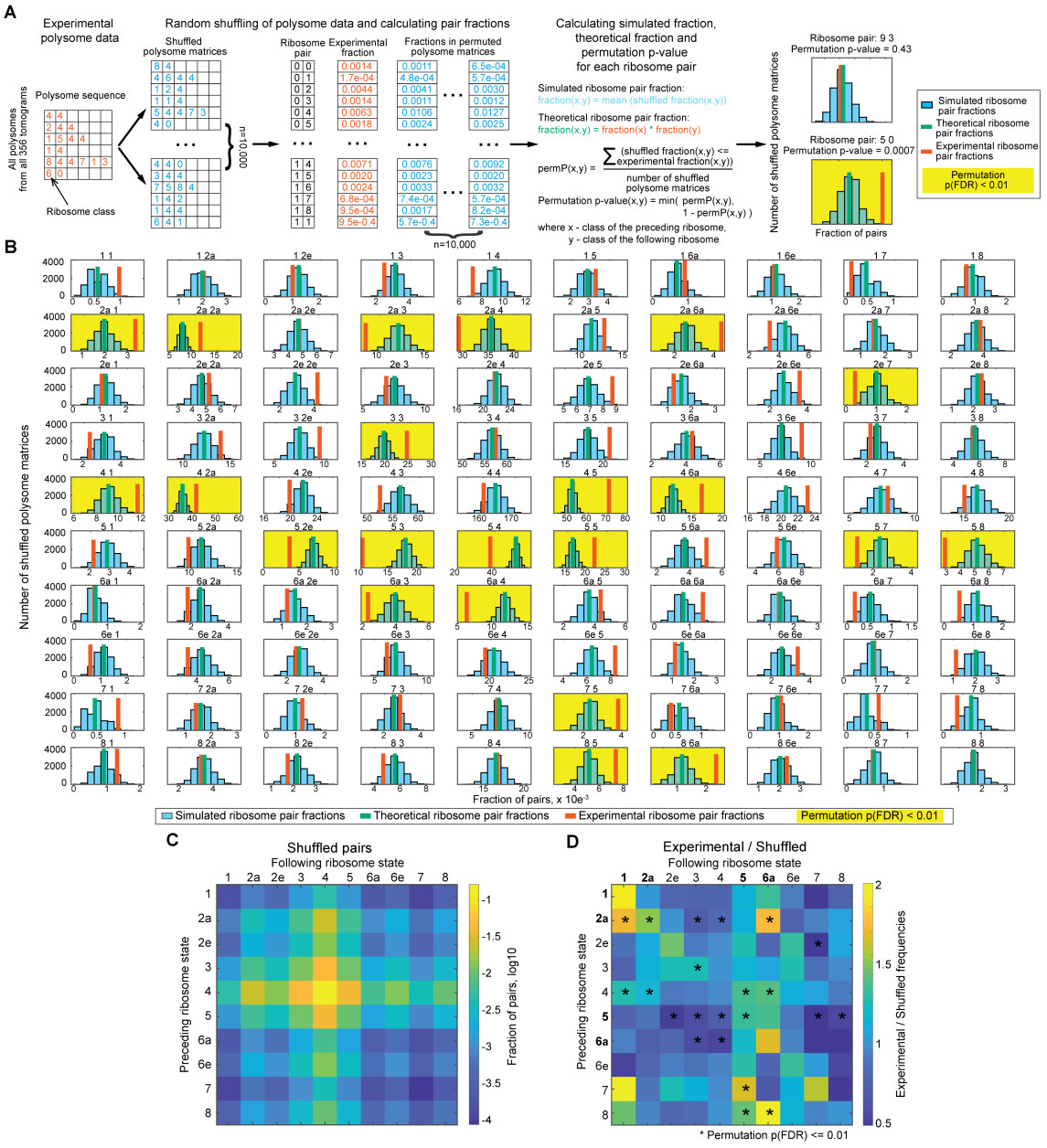

Figure 4.15 | Polysome shuffling analysis to validate the detected coordination.

(**A**) Illustration of the polysome shuffling experiment and permutation p-value calculation. (**B**) Comparison of experimental state pair fractions and the pair frequencies after shuffling. (**C**) Frequencies of elongation

state pairs calculated after random shuffling of the experimental data. (**D**) Difference between the experimental and shuffled state pair frequencies. Ribosomes in states that need the binding of elongation factors to proceed in the elongation cycle (states 1 and 2a for EF-Tu, and states 5 and 6a for EF-G) are more frequently engaged as the following ribosomes. The analysis and figure preparation were done by Maria Zimmermann-Kogadeeva within a collaborative project.

It was also found that the asymmetric elongation state distribution between adjacent ribosomes in polysomes becomes more significant as the distance between the two ribosomes decreases (Figure 4.16). In polysomes detected with decreasing distance cutoffs, the ratio between the number of pairs in which the state is a preceding ribosome to the number of pairs in which the state is a following ribosome decreases for states 1, 2a, 5 and 6a, but increases for the other states. These results suggested that elongation coordination within polysomes is a local mechanism that requires close proximity between two adjacent ribosomes in polysomes. As the states 1, 2a, 5 and 6a need elongation factor (EF-G or EF-Tu) to proceed in the elongation cycle, it is hypothesized that such local coordination is achieved by preventing elongation factor binding to the next ribosome within the polysome.

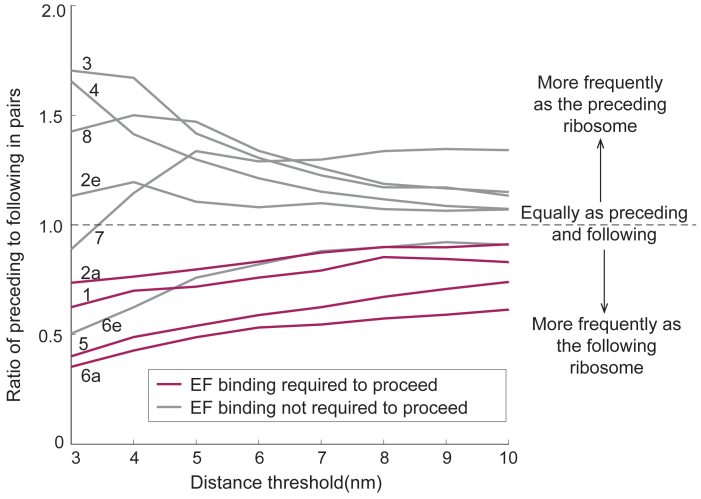

Figure 4.16 | Asymmetric state distribution is more profound as distance decreases.

Ratio of the frequencies of ribosomes in each state to be the preceding one against the following one in polysomes when using different cutoffs in polysome detection. If there is no difference in the state occurrence for the preceding and following ribosome, the ratio equals to 1. The analysis and figure preparation were done by Maria Zimmermann-Kogadeeva within a collaborative project.

4.2.5 Polysome elongation coordination mediated by L9

To provide further structural evidence for the local coordination hypothesis and to investigate the structural basis behind it, I did *de novo* classification of polysomes (di-ribosomes; *i.e.* adjacent ribosome pairs) in RELION to better resolve the ribosome-ribosome interfaces (Figure 4.17A). With the box size large enough to accommodate two closely assembled ribosomes, RELION 3D

classification was able to sort the ribosome pairs (di-ribosomes) in polysomes. Further classification focusing on the relative orientations within ribosome pairs reduced the high flexibility between adjacent ribosomes and resulted in several di-ribosome classes with the interface better resolved. According to these structures, the ribosomal protein L9 of the preceding ribosome (i) can adopt a stretched conformation within the interface, and its C-terminal domain protrudes into the elongation factor binding site of the following ribosome (i+1) (Figure 4.17B-D). When superposing ribosome structures with elongation factors, CTD of the stretched L9 has a steric clash with both EF-G and EF-Tu binding to the following ribosome (Figure 4.18A, B). Except for L9, the following ribosome did not appear to interfere with L1 stalk dynamics or tRNA disassociation in the preceding ribosome. These structures suggested that L9 within the ribosome interface may change conformation and interfere with elongation factor association in the following ribosomes.

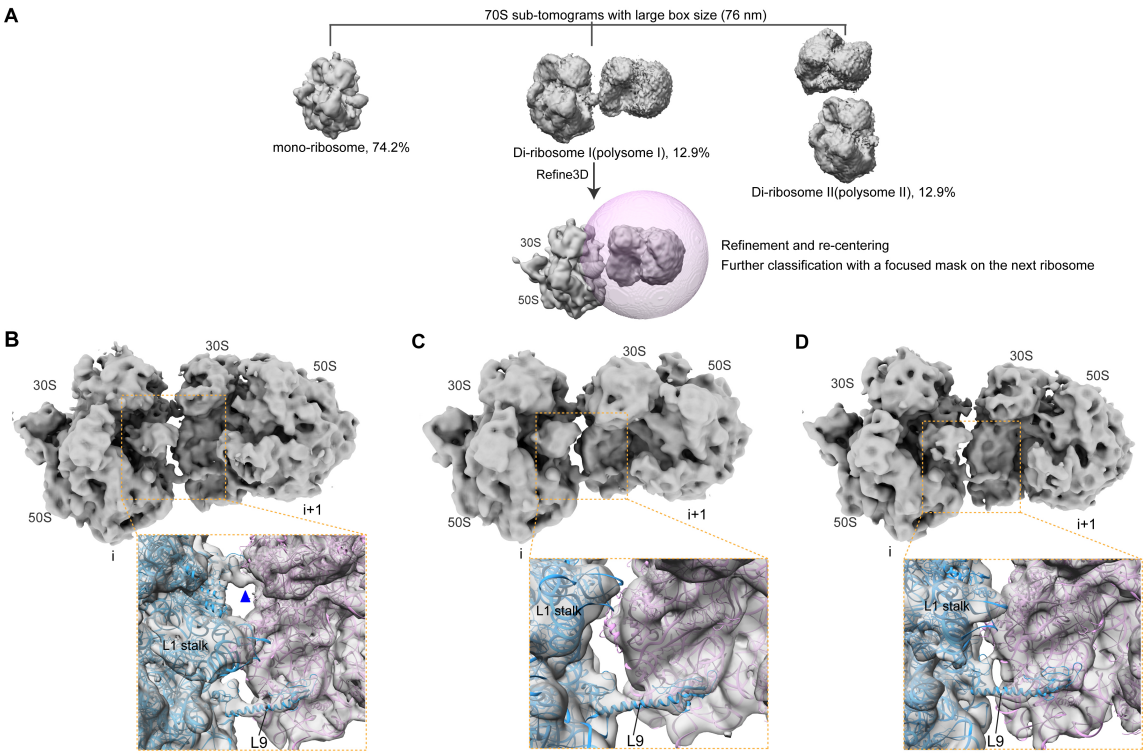

Figure 4.17 | RELION classification of polysomes.

(**A**) *De novo* RELION classification to detect polysomes. The ribosome sub-tomograms were re-extracted with a large box size that can accommodate potential neighboring ribosomes, and then the polysomes were classified as ribosome pairs (di-ribosomes). The two different di-ribosome classes just represent different orientations, but the large rotations made RELION regards them as separate classes. When aligning on one ribosome, the density of the next ribosome is blurred due to high flexibility. Further focused classification on the next ribosome allowed to refine several classes with better structures. (**B-D**) Three representative di-ribosome classes with the best resolved ribosome-ribosome interface. The stretched L9 of the preceding

ribosome is highlighted. In panel (B), the possible intervening mRNA density is also indicated by blue arrowhead.

L9 is conserved across the bacteria kingdom, but is absent in archaea and eukaryotes[354,355]. Structural studies of L9 showed that it is highly flexible, with two RNA binding domains connected by a long helix[99,356,357]. In previous X-ray crystallography studies, it was found that L9 can compete with translational GTPases binding to the adjacent ribosome and hinders crystallization of 70S ribosomes in complex with elongation factors[357,358]. Conversely, in single particle cryo-EM structures of isolated ribosomes, L9 has a flat conformation residing on the ribosome surface[99]. While L9 was reported to be nonessential, mutations in L9 result in increased frameshifting in translation and ribosomes compaction by one codon[355,359].

In the polysome (di-ribosome) classes I determined, L9 CTD is in close proximity with the 16S rRNA of the following ribosome (Figure 4.17B-D). The steric clash of L9 with both EF-G and EF-Tu can block certain steps in the elongation cycle of the following ribosomes (Figure 4.18). As demonstrated in Figure 4.16, the local translation elongation coordination depends on close distance between two adjacent ribosomes in the polysome. Moreover, the enrichment of states prior to EF-G binding (5 and 6a) is more significant than that of states prior to EF-Tu binding (1 and 2a). EF-G mediated translocation is the elongation step when the following ribosome moves one codon forward along mRNA and is thereby more likely to interact with the stretched L9 of the preceding ribosome. Therefore, I propose that within tightly assembled polysomes, or the compacted regions of polysomes, the stretched L9 of one ribosomes can influence elongation factor binding to the next following ribosome, and transiently block its elongation. This local elongation coordination mediated buffers adjacent ribosomes and helps maintain translation fidelity by avoiding harsh collision.

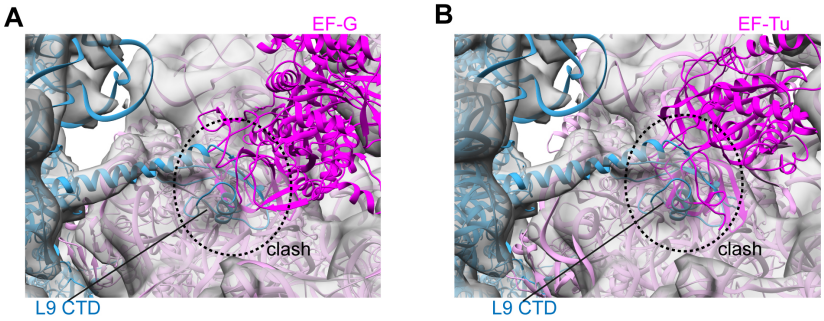

Figure 4.18 | Steric clash between the stretched L9 and elongation factors.

(**A**) Steric clash between L9 CTD of the preceding ribosome (i) and the superposed EF-G in the following ribosome (i+1) according to the "EF-G, A/P, P/E" structure. (**B**) Steric clash between CTD of the stretched L9 and EF-Tu according to the "EF-Tu, P" structure.

4.2.6 Dynamics of L9 on ribosomes

Independent focused classification on L9 provided additional evidence for the proposed coordination mechanism. L9 was found in most (89.1%) ribosomes inside cells. It mostly (68.9%) adopts the flat conformation residing on the ribosome surface, while the stretched conformation accounts for 20.2% of all 70S ribosomes (Figure 4.19A). Ribosomes with the stretched L9 largely overlap with the ribosomes detected as polysomes by either spatial analysis or RELION classification (Figure 4.19B). Especially, the overlap is more pronounced between ribosome pairs with a stretched L9 in the middle (*i.e.* the preceding one has stretched L9) and ribosome pairs with tighter "t-t" arrangement (Figure 4.19C). These results suggest that L9 tends to have a flat conformation in single ribosomes and to adopt the stretched conformation within tightly packed polysomes. This may also explain previous contradictory findings of the stretched L9 in X-ray crystallography structures and the flat L9 in single particle cryo-EM structures, because the ribosome packing in crystals recapitulates configurations of compacted polysomes[357]. The case of L9 is a great example showing the advantage of solving structures inside native cells where the molecular process is actively ongoing and the original spatial organization is maintained.

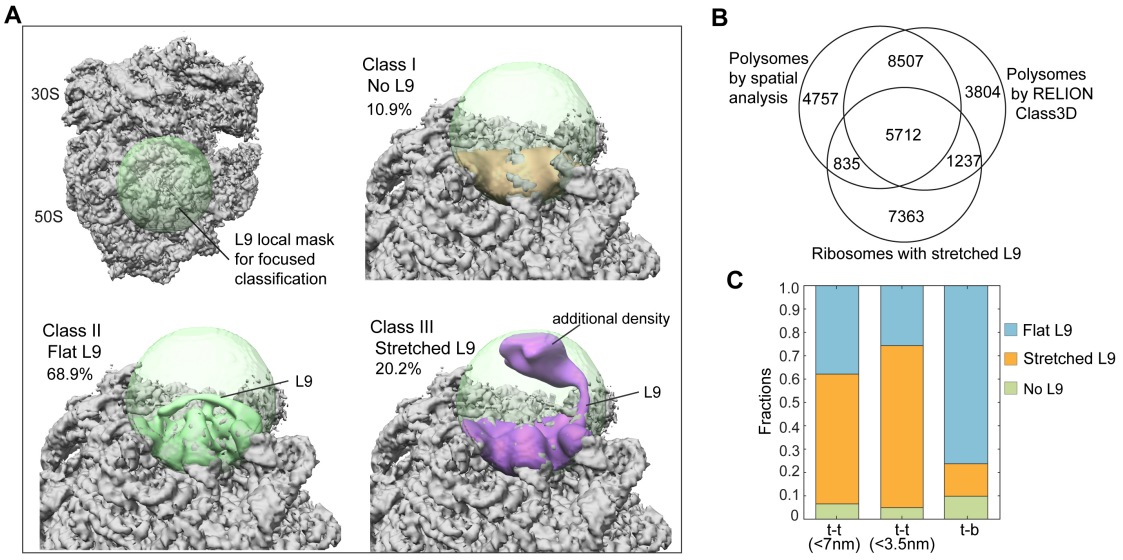

Figure 4.19 | Overlap between ribosomes with stretched L9 and polysomes.

(**A**) Focused classification with a spherical mask (light green) on the L9 region. Local densities in the three final classes are shown on the ribosome surface: no L9 (class I, 10.9%), flat L9 (class II, 68.9%) and stretched L9 (class III, 20.2%). The additional density connecting to the stretched L9 possibly comes from the neighboring ribosome. (**B**) Overlap between polysomes detected by spatial organization, di-ribosomes (polysomes) identified by RELION 3D classification, and the ribosomes with stretched L9. (**C**) Distribution of different L9 conformations in ribosome pairs with different polysome arrangements. The stretched conformation of L9 is more frequently observed in "t-t" pairs, especially those tightly packed ones, whereas the flat conformation is predominant in "t-b" pairs.

4.3 Antibiotics reshape translation landscapes in cells

The translation machinery is one of the most important targets for antibiotics. Many of the ribosome-targeting drugs are known to stall certain translation steps[360,361]. Structural studies of isolated ribosomes in complex with different drugs improved the understanding of antibiotic's mechanism of action (MOA), which also facilitated the investigation of the translation processes. However, recent studies suggest that how antibiotics function inside cells is more complex than previously expected[362-364]. The in-cell ribosome analysis performed previously enabled deciphering MOAs of antibiotics inside cells. More interestingly, using different drugs to treat *M. pneumoniae* cells, it is possible to probe how the whole ribosome community respond to various perturbations as a functional machinery.

I performed the cryo-ET and sub-tomogram analysis with *M. pneumoniae* cells treated by two representative ribosome-targeting antibiotics: chloramphenicol (Cm) that binds to 50S peptidyl transfer center and inhibits peptide bond formation[319-321], and spectinomycin (Spt) that binds to 30S neck and inhibits translocation[365,366]. The third antibiotic pseudouridimycin (PUM) that specifically inhibits transcription was tested to examine how the translation machinery responds to nonspecific perturbations[367]. The concentrations of antibiotics were made more than 10 times higher than the reported saturating concentrations[365,367,368]. The treatment time was controlled to 15 - 20 minutes to better reflect the immediate responses on the cell level, which is a compromise between the fast effect on the molecule-level and the long-time effect that may already terminate most cellular actives.

4.3.1 Ribosome classification in cells treated by chloramphenicol

In cells treated with Cm, ribosome sub-tomogram analysis including classification and refinement was performed following the pipeline I developed for the native untreated cells (Figure 4.20A-C). Due to different data collection and analysis time, only the first Cm-treated dataset was included for the detailed classification. The vast majority of extracted sub-tomograms were found to be fully assembled 70S ribosomes, possibly owing to Cm's inhibition of ribosome dissociation mediated by EF-G and RRF[369]. I obtained four interpretable 70S ribosome classes after extensive classification. Consistent with the inhibition of peptide bond formation by Cm, 72% of all 70S ribosomes are trapped in the "A, P" state prior to peptidyl transfer. Surprisingly, 28% of 70S ribosomes were found in other elongation states, which are right after or before the major "A, P" state (Figure 4.20A). The Cm molecule was clearly resolved in the major "A, P" class in its canonical binding site[321], but not in the other three classes determined at relatively low resolutions (Figure 4.20D). Whether these minor classes have Cm binding remains unclear, as the intracellular concentration of the drug is impossible to determine. Translation factors were clearly resolved in all the four 70S classes, especially for the "A, P" class where the densities for mRNA and the

nascent peptide were nicely observed (Figure 4.20E). For the first time, the translation inhibition mechanism of Cm was directly demonstrated in cells: Cm binds near the CCA tail of A-site tRNA, sterically hinders the formation of new peptide bond and blocks peptidyl transfer (Figure 4.20F).

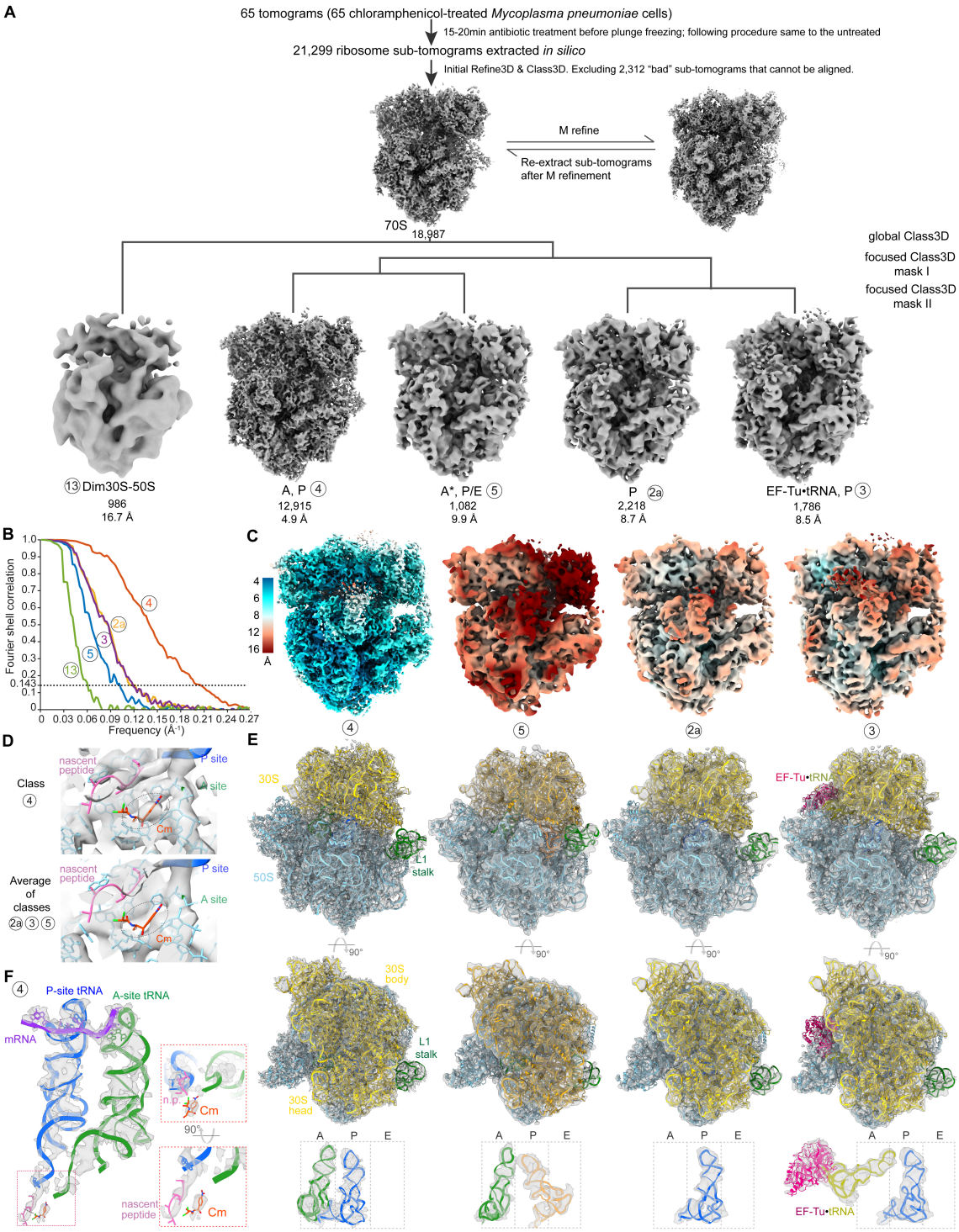

Figure 4.20 | Ribosome classes in Cm-treated cells.

(**A**) The processing workflow for the Cm-treated dataset. Only the first dataset was included for this analysis. Sub-tomogram classification and refinement are the same as the untreated. (**B**) FSC curves for RELION refinements of all classes. (**C**) Class densities colored by local resolutions. (**D**) The Cm molecule is resolved

in the major "A, P" class, but not in the other minor classes. (**E**) Models for the four interpretable classes in Cm-treated cells. Panels below show the occupancy states of elongation factors and tRNAs. (**F**) In the "A, P" class, mRNA, A- and P-site tRNAs, part of the nascent peptide, and the Cm molecule are clearly resolved. The nascent peptide chain has a clear continuous density linking to the P-site tRNA. Cm binds near the A-site tRNA and stops it from receiving the peptide.

4.3.2 Ribosome classification in cells treated by spectinomycin

Spt binds to the 30S neck region and blocks translocation by inhibiting head swiveling of the small ribosomal subunit[110,365,366,370]. In Spt-treated cells, five 70S classes and one free 50S class were obtained with a similar classification procedure as before (Figure 4.21). 67.5% of all 70S ribosomes are stalled in the "EF-G, A/P^{Spt}, P/E" state, which is similar to the pre-translocational "EF-G, A/P, P/E" state in untreated cells. The Spt molecule is clearly determined near the helix 34 of 16S rRNA (Figure 4.21D), consistent with the reported major binding site in the small ribosomal subunit[110,365]. In addition to the major "EF-G, A/P^{Spt}, P/E" class, ribosomes in other elongation states were detected with relatively low frequencies (Figure 4.21E). The Spt density in the four minor classes could not be unambiguously determined due to high flexibility and low resolutions.

The "EF-G, A/P^{Spt}, P/E" class is the first in-cell structure of 70S ribosome in complex with spectinomycin, mRNA, tRNAs and the nascent peptide. It ratifies the proposed mechanism of Spt that the drug binds to the interface between 30S head and body and blocks mRNA-tRNA translocation by inhibiting 30S dynamics. In the structure, Spt was determined in close contact with the lysine residue K81 (corresponds to K26 in *E. coli*) on loop 2 of the ribosomal protein S5. This agree well with previous publications showing that mutations in S5 loop 2 can cause resistance to Spt and affect translation fidelity[371,372].

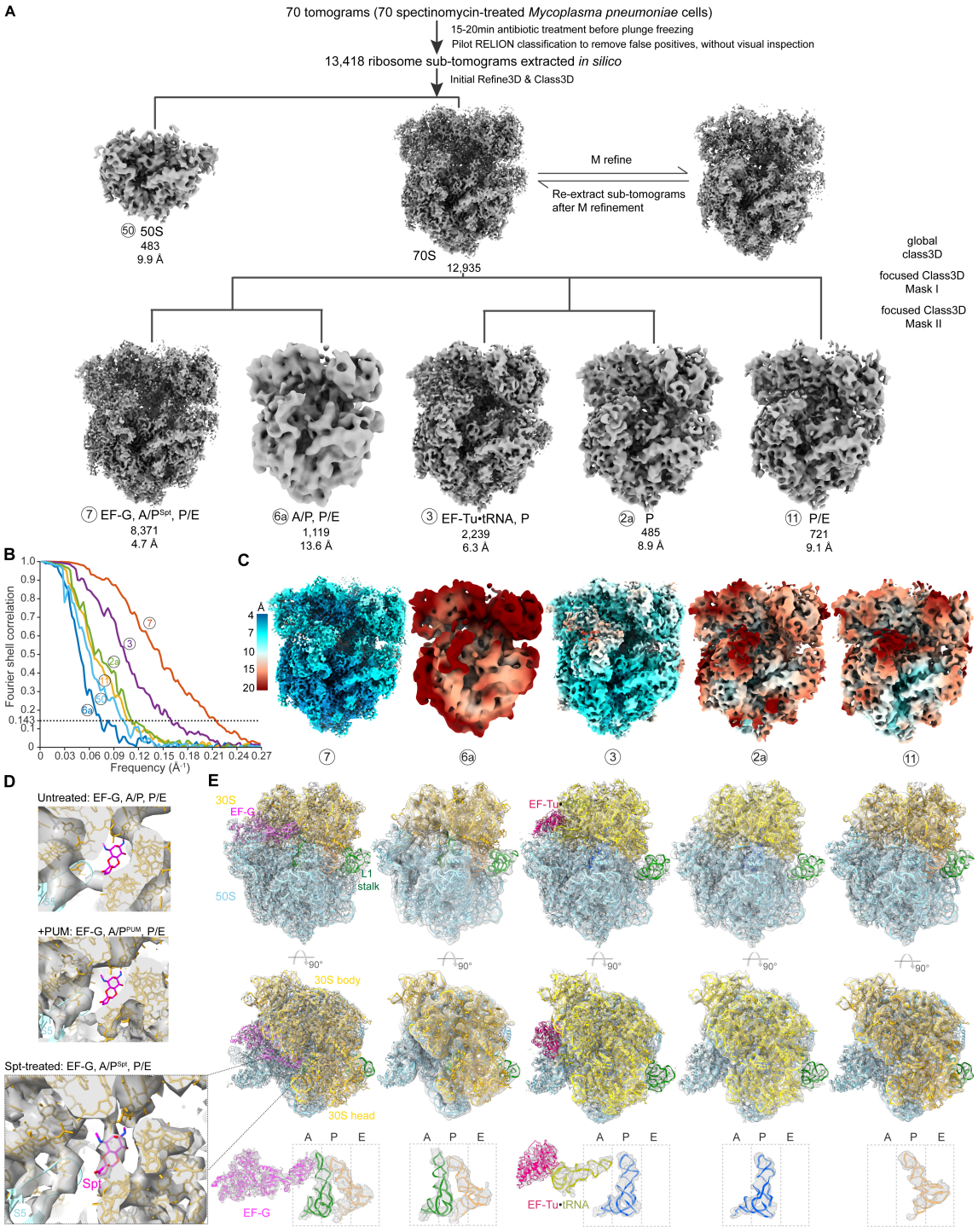

Figure 4.21 | Ribosome classes in Spt-treated cells.

(**A**) The processing workflow for the Spt-treated dataset. (**B**) FSC curves for all classes. (**C**) Class densities colored by local resolutions. (D) Spt (magenta) is clearly resolved in the major "EF-G, A/P^{Spt}, P/E" class. Due to flexibility of 30S, Spt is not resolved in other minor classes. Instead, the corresponding classes in untreated and PUM-treated (described below) cells were used to validate the Spt density. (**E**) Models of 70S classes and the occupancy states of elongation factors and tRNAs in Spt-treated cells.

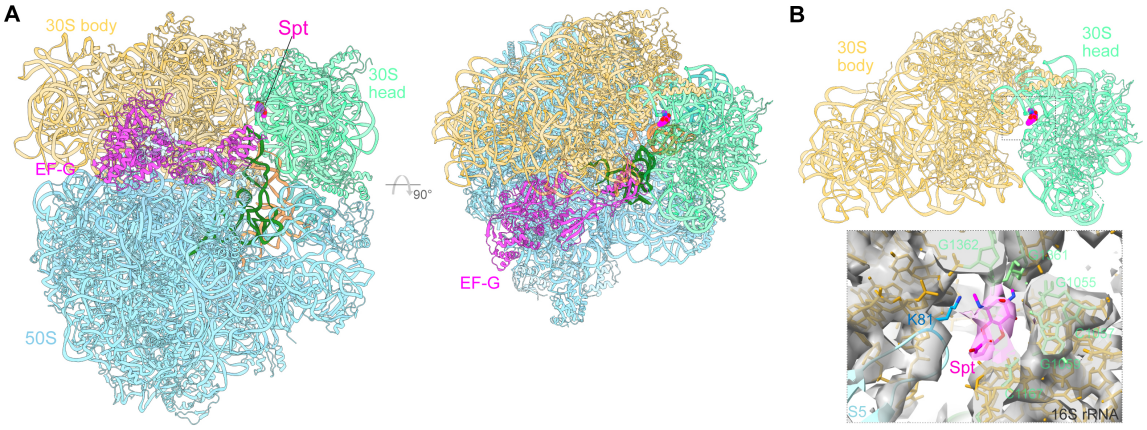

Figure 4.22 | Spt inhibits dynamics of the small ribosomal subunit.

(**A**) The "EF-G, A/P^{Spt}, P/E" structure showing the ribosome in complex with spectinomycin, EF-G, and tRNAs. The head and body domains of the small ribosomal subunit are colored differently. (**B**) The Spt molecule (magenta) binds to the neck region between 30S head and body, surrounded by several 16S rRNA bases (G1055, C1057, G1059, C1167, G1361 and G1362). This region is near one rotation hinge between 30S head and body, confirming Spt's inhibition of 30S dynamics. The drug molecule is in close proximity to loop 2 of the ribosomal protein S5, especially the residue K81.

4.3.3 Ribosome classification in cells treated by pseudouridimycin

PUM is a specific RNA polymerase inhibitor and the PUM-stalled RNAP can physically block mRNA translocation in the ribosome that collides with it[255] (also see section 5.4.2 in the next chapter). Through extensive classification (Figure 4.23A), I found that 59.5% of all 70S ribosomes are in the "EF-G, A/P^{PUM}, P/E" state, which largely overlaps with the PUM-stalled RNAP-ribosome supercomplexes[255] (Figure 4.23B). Five additional 70S ribosome classes in other elongation states were obtained with relatively low percentages. Interestingly, the major "EF-G, A/P^{PUM}, P/E" class in PUM-treated cells is similar to the "EF-G, A/P, P/E" class in untreated cells and the "EF-G, A/P^{Spt}, P/E" class in Spt-treated cells (Figure 4.24B). The observation that physical blocking of mRNA movement by PUM-stalled RNAP and chemical interference of 30S dynamics by Spt lead to similar pre-translocational structures, provides direct evidence for the coupling between mRNA translocation and 30S head swiveling[113,115,121].

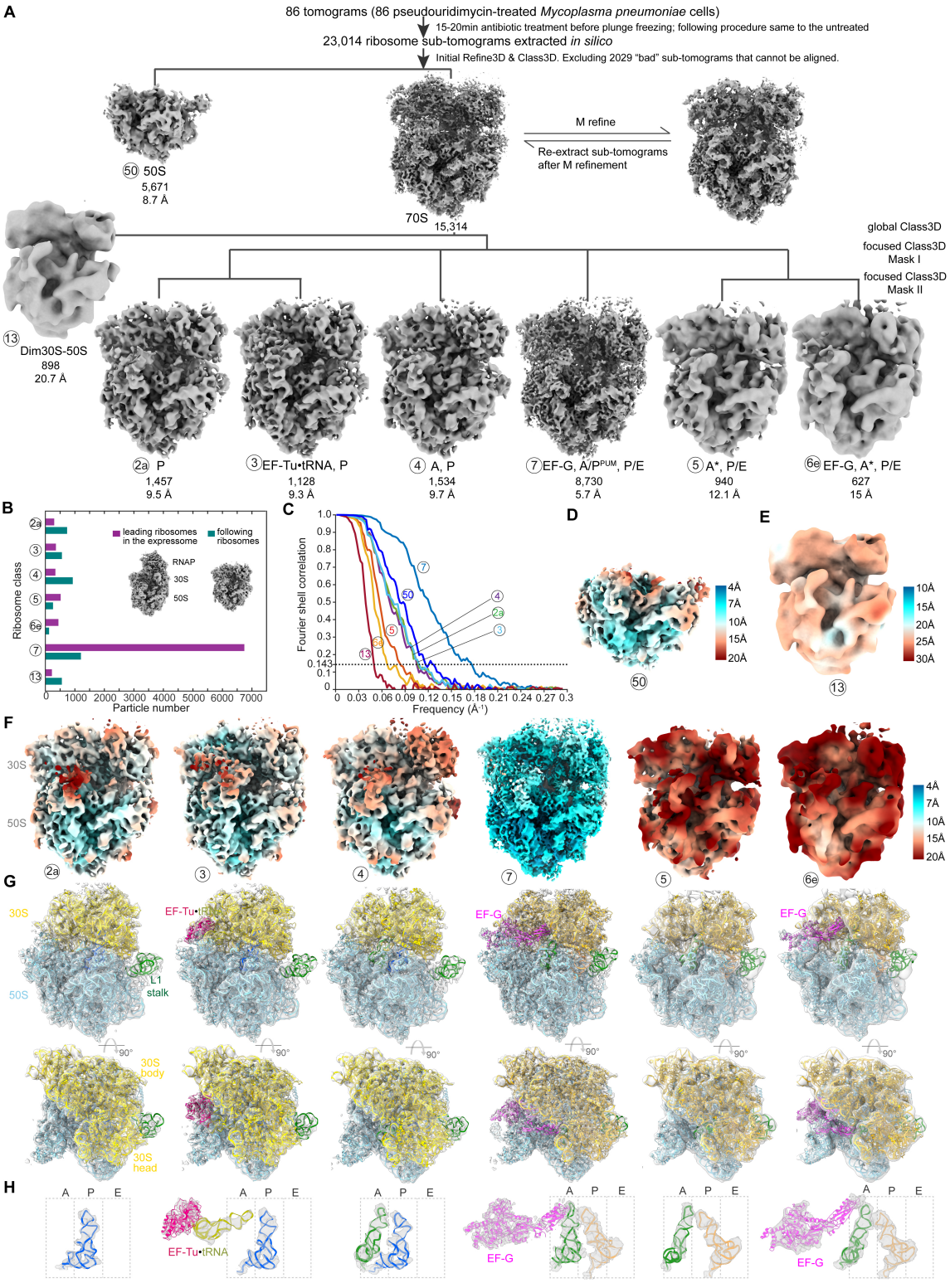

Figure 4.23 | Ribosome classes in PUM-treated cells.

(**A**) The processing workflow for the PUM-treated dataset. (**B**) Ribosome state distribution for individual ribosomes (green) and the ribosomes clashing with RNAP (purple) during transcription-translation coupling. Details can be found in section 5.4.2 in the next chapter. (**C**) FSC curves for refinement of all classes. (**D-F**) Class densities colored by local resolutions. (**G**) Models for 70S classes in PUM-treated cells. (**H**) The states of elongation factors and tRNAs corresponding to the classes in (G).

4.3.4 Single cell clustering analysis based on translation profiles

Comparison of the ribosome state distributions in the native untreated and antibiotic-treated datasets showed that translation landscapes are completely changed by antibiotics (Figure 4.24A). Each ribosome class in antibiotic-treated cells was compared to the identical class in untreated cells. Some fine conformational difference could be found for these comparable classes in different cell conditions. For example, the "EF-G, A/P, P/E" classes in untreated, Spt-treated and PUM-treated cells largely align, but there is a small positional difference for the A/P-site tRNA (Figure 4.24B). For the ribosome-specific antibiotic Cm and Spt, I found a variety of intermediate structures during translation, despite the majority of ribosomes are stalled in the expected states. The finding of these unexpected classes prompted the effort to examine the difference between single cells in response to antibiotics. Hierarchical clustering analysis based on translation elongation profiles was performed for the 577 cellular tomograms (one tomogram = one cell; 356 native untreated, 65 Cm-treated, 70 Spt-treated and 86 PUM-treated). The cells clustered into four major groups, which are in perfect alignment with the four experimental conditions (Figure 4.24C). Within each experimental group, all cells respond similarly to the antibiotic treatment, albeit some cell-to-cell variability. It is possible that within the relatively short treatment time of 15-20 minutes, cells did not develop or exhibit phenotypic differences (*e.g.* antibiotic resistance).

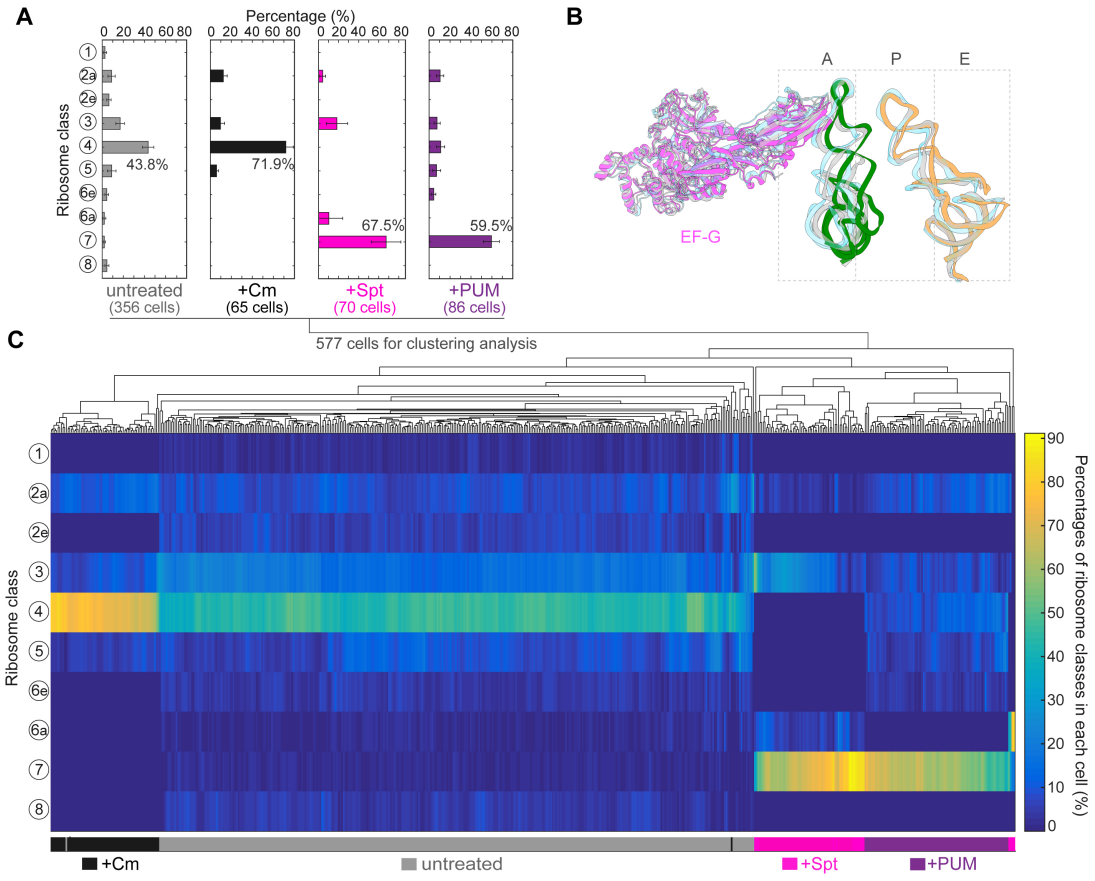

Figure 4.24 | Single cell clustering analysis based on translation profiles.

(**A**) Antibiotics induced distinct translation elongation profiles in cells. Ribosome classes in antibiotic-treated cells are placed and named according to their counterparts in untreated cells. +Cm, chloramphenicol-treated; +Spt, spectinomycin-treated; +PUM, pseudouridimycin-treated. (**B**) Comparison of the pre-translocational "EF-G, A/P, P/E" states from different datasets. The "EF-G, A/P^{Spt}, P/E" class in Spt-treated cells is colored as before (EF-G, magenta; hybrid A/P-site tRNA, green; hybrid P/E-site tRNA, yellow). For comparison, the "EF-G, A/P^{PUM}, P/E" class in PUM-treated cells (light blue) and the "EF-G, A/P, P/E" class in untreated cells (light gray) are shown. (**C**) Single cell clustering analysis based on translation elongation profiles of the 577 cells under different conditions. This panel C was generated by a modified script from Maria Zimmermann-Kogadeeva. Each thin vertical line corresponds to one cell. The bottom bar indicates the treatment (represented by different colors).

4.3.5 Effects of antibiotics on spatial organization of ribosomes

After antibiotic treatment, no significant change was observed for the overall spatial distribution of ribosomes in cells, possibly due to the simplicity (no defined nucleoid) of *M. pneumoniae* cells. However, closer analysis of the organization of ribosomes into polysomes, revealed that antibiotics profoundly influenced the organization of ribosomes inside cells. The percentages of detected polysomes and the mRNA entry-to-exit distances between adjacent ribosomes within polysomes under the four different cell conditions are summarized in Table 4.1.

Table 4.1 | Polysome analysis in untreated and antibiotic-treated cells.

	untreated	**Cm-treated**	**Spt-treated**	**PUM-treated**
Polysome percentage*	26.2%	20.63%	12.83%	7.73%
Distance mRNA exit to entry "t-t" (nm)	4.2±1.4 (9,100)	3.2±1.5 (1,959)	4.3±1.5 (580)	2.3±0.9 (635)
Distance mRNA exit to entry "t-b" (nm)	5.4±1.5 (2,491)	5.3±1.1 (186)	5.1±1.2 (326)	4.9±1.4 (16)

* the ribosomes engaged in polysomes against all 70S ribosomes.

In Cm-treated cells, the percentage of ribosomes detected as polysomes is close to the number in untreated cells. This agrees with literature showing that Cm treatment has limited impact on the formation of polysomes in *E. coli*[352,373]. It is possible that there are still context-dependent translation activities upon Cm treatment and the formation of polysomes are not prohibited[363]. In the detected polysomes, the ratio of "t-t" to "t-b" pairs increased in Cm-treated cells. For "t-t" pairs, the mean distance from the mRNA exit site of the preceding ribosome (i) to the the mRNA entry site of the following ribosome (i+1) is approximately 1 nm (about the length of one codon of three bases) shorter compared to untreated cells. These indicated that polysome organization becomes about one codon tighter upon Cm treatment. This may be related to Cm's specific inhibition of

the peptide transfer step but not translocation (relative mRNA movement). In the presence of Cm, the ribosome can still conduct translocation and move a step further along the mRNA[363,374,375].

In Spt-treated cells, the percentage of polysomes is lower than native untreated cells. More polysomes have the "t-b" arrangement, while the mRNA exit to entry distances for "t-t" and "t-b" pairs remain unchanged in comparison to the native untreated cells. It is interesting that the two ribosome-specific antibiotics caused significantly different polysome organizations. Spt specifically inhibits the 30S dynamics and mRNA translocation[365], which avoids the possibility for polysomes to become more compacted. This may also lead to less ribosomes that can be engaged in polysomes.

The more intriguing results come from the PUM-treated cell. The percentage of detected polysomes upon PUM treatment is considerably lower compared to native untreated cells (26.2% *vs* 7.73%). Moreover, the detected polysomes almost all adopt the "t-t" configuration with significantly short distances (2.3±0.9 nm) between adjacent ribosomes. The PUM-stalled RNAP acts as a physical barrier for the leading ribosome during transcription-translation coupling (see section 5.4.2)[255]. Clash between the leading translating ribosome and the PUM-stalled RNAP makes the leading ribosome arrested in the pre-translocational state (Figure 4.23B), while the following ribosomes translating on the same nascent mRNA continue elongating until collision with the leading ribosome. This collision explains why polysomes in the presence of a RNAP-specific inhibitor are more compacted. Considering the relatively short lifetime (several minutes in average) of mRNA[376] and the drug treatment time of 20-25 minutes, it is expected that most free mRNAs in *M. pneumoniae* cells are degraded. This means most translating ribosomes are elongating on the nascent mRNAs that are being transcribed by RNAPs. If more and more ribosomes start translating on the nascent mRNAs and then collide with preceding ribosomes, most ribosomes would be detected as tightly packed polysomes. However, the detected percentage is only 7.73% and there are more than 30% individual 70S ribosomes (*i.e.* 70S ribosomes that are either not colliding with RNAP or not within detected polysomes). One possible explanation is that the collided ribosomes already being released through ribosome rescue pathways[377]. As PUM does not directly influence ribosomes, the released ribosomal subunits can again load on the nascent mRNA, and start translating and then collide. Such futile loop continues until the cell cannot sustain. Although it is unclear if this assumption is true or what are really happening there, it is clear that there are unraveled complexity and dynamics inside cells upon antibiotic treatment.

Taken together, the data demonstrates that both the energy landscapes of translation and the spatial organization of the ribosome community are profoundly reshaped by antibiotics.

4.3.6 Antibiotics fundamentally reshape translation in cells

The results clearly demonstrated that antibiotics can thoroughly reshape the translation machinery and translation landscapes inside cells, rather than only blocking specific translation steps. The finding of unexpected translation elongation states at lower frequencies in antibiotic-treated cells is reminiscent of the context-dependent inhibition reported for certain antibiotics[362,363], and recent studies showing low but ongoing translation in bacterial cells treated by various antibiotics including Cm[375,378]. The inhibition effect of Cm can be influenced by the nascent peptide chain and the incoming amino acid[362,363]. So far, such effect has not been reported for spectinomycin, but the action of a similar antibiotic pactamycin that binds to 30S and inhibits translocation has been suggested to depend on the A-site substrate[363].

This thesis provides insights into the structural and functional landscapes of translation in cells under different conditions across scales, from the ribosome structures with atomic detail, to the organization of the whole translation machinery as a functional system. Most antibiotics, including Spt and Cm, are bacteriostatic drugs that only inhibit cell replication, but do not immediately kill bacterial cells[364]. After antibiotic molecules enter into the cell, specific binding to molecular machines and interfering with their dynamics make the primary MOA on the molecule level. I found that in cells treated with antibiotics, the translation machinery are profoundly reshaped due to the temporospatial organization of the altered molecular machines, which may make the cell-level MOA. It is possible that the antibiotic-reshaped translation machinery leads to imbalance or disorder in gene expression in bacterial cells, which eventually results in detrimental consequences for the cell. For PUM, the influence on translation may represent a "side effect", but I expect besides its specific inhibition of transcription the induced translation disorder is a major cause for cell lethality.

Chapter 5: Transcription-translation coupling

To structurally decipher transcription-translation coupling inside the cell, I performed extensive sub-tomogram classification of ribosomes to determine potential coupling complexes. To establish the feasibility of integrative in-cell structural biology, the cryo-ET maps I attained were combined with in-cell crosslinking mass spectrometry data from Juri Rappsilber's group. Integrative modeling was performed to derive models of the supramolecular assemblies (also referred as expressomes) formed by RNA polymerase, the ribosome and auxiliary factors. Most of the results presented in this chapter were published in a collaborative paper in Science[255].

5.1 In-cell structure of the RNAP-ribosome supercomplex by Cryo-ET

5.1.1 Focused classification on the mRNA entry region of ribosomes

The in-cell ribosomes provided ideal anchors for detecting potential supercomplexes formed during transcription-translation coupling, as ribosomes' large size and high contrast make them relatively easy to localize in cellular tomograms. As discussed in the previous chapters, the completeness and unbiased representativeness of the ribosome sub-tomogram datasets provided the data basis for the discovery of potential transcription-translation coupling complexes. When the transcription-translation coupling happens, RNAP is expected be positioned near the mRNA entry site of the translating ribosome[157,169]. Correspondingly, in the averaged density of ribosome sub-tomograms re-extracted with a large box size that can accommodate possible interacting proteins or complexes, I found a weak density above the ribosome mRNA entry site which possibly indicates an associated RNAP (Figure 5.1).

Following on the observation, I performed extensive focused classification on the additional density above the mRNA entry site in RELION[227,314] (Figure 5.1). The additional density may come from mRNA, neighboring ribosomes if the ribosome is engaged in polysomes, coupled RNAP, or other macromolecules that interact with the ribosome during translation. First, closely assembled polysomes were excluded to avoid the interference from neighboring ribosomes (Figure 5.1). Then, rounds of focused classification with a local mask above the mRNA entry site were carried out. The sub-tomograms were finally sorted into three groups based on additional densities associated with the mRNA entry site: ribosomes without significant extra density (36,301), ribosomes with undefined extra density (17,202), and ribosomes with a well-resolved RNAP density (2,952).

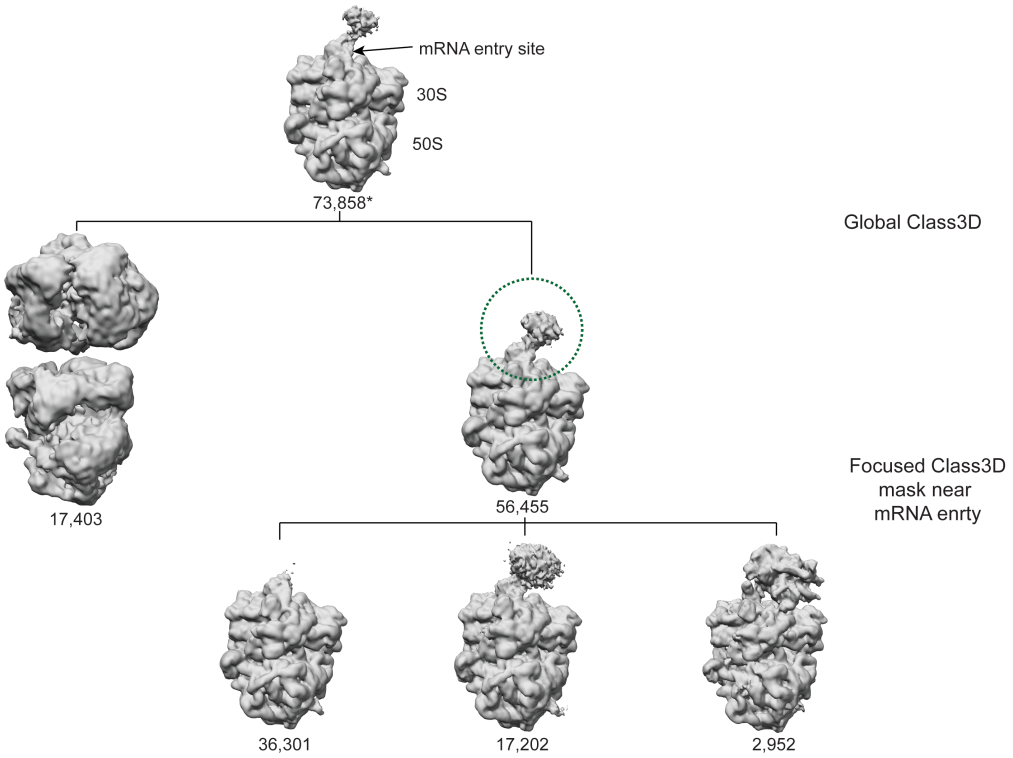

Figure 5.1 | Classification of expressomes in native untreated cells.

The additional density above the ribosome mRNA entry site implies a coupled RNAP. Except for it, no other additional densities around the ribosome were found to potentially contain a RNAP. Focused classification around the mRNA entry site of 70S ribosomes was performed to obtain the possible RNAP-ribosome supercomplexes. The total number of ribosomes from native untreated cells used here is slightly different from the number in the last chapters, because of different processing time (see Materials and Methods). In the ribosome average, the only additional density that may indicate an interacting RNAP is above the mRNA entry site (arrow). During classification, 70S ribosomes identified within closely assembled polysomes based on global classification (17,403) were first excluded. After focused classification on the mRNA entry site (green dashed circle), 2,952 ribosomes with a clear RNAP density were obtained, while 36,301 ribosomes did not appear to have significant density above the mRNA entry site. The remaining 17,202 ribosomes have the additional density, but could not be further classified or refined to any interpretable density. To make the additional density more obvious, visualization thresholds are made lower in Chimera.

The 2,952 well-resolved RNAP-ribosome supercomplexes may represent a relatively stable conformations during transcription-translation coupling. Despite of different classification strategies, the class of 17,202 sub-tomograms with unclear additional density could not be further classified or refined to generate an interpretable density. As demonstrated in Figure 5.2, the density above the ribosome mRNA entry site exhibited high structural heterogeneity, possibly including both compositional heterogeneity and conformational flexibility. These densities may

represent other structurally dynamic or flexible assemblies during transcription-translation coupling, or any other macromolecules and complexes interacting with the ribosome.

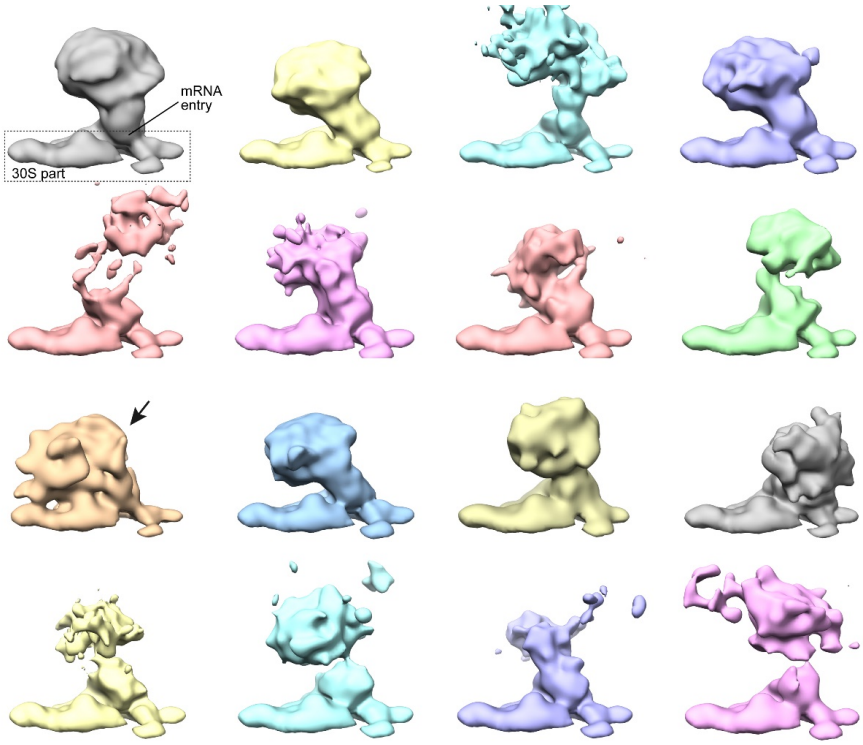

Figure 5.2 | Structural heterogeneity above the ribosome mRNA entry site.

A glimpse of the high heterogeneity above the mRNA entry site of ribosomes. The shown densities are classes of one focused classification run with a local mask covering the additional density above the mRNA entry site (as indicated by the green dashed circle in Figure 5.1, but from the back view). The mRNA entry site and part of the 30S density are labelled in the first class, as an example. Only one class has a defined RNAP density (indicated by arrow; identical to the 2,952 class in Figure 5.1), while all other classes did not converge or result into any interpretable density (grouped as the 17,202 class). Lower thresholds were used to visualize the additional density.

5.1.2 Refinement of the RNAP-ribosome supercomplex

To better refine the 2,952 classified RNAP-ribosome supercomplexes, especially the interface between RNAP and the ribosome, different RELION refinement strategies were performed (Figure 5.3). The best local density came from a focused refinement with a mask covering RNAP and 30S, which reduced the influence of the ratcheting between the small and large ribosomal subunit, and meanwhile kept the important interface between RNAP and 30S as an integral part during refinement. The final density has an overall resolution of 9.2 Å, while the local resolution for RNAP is slightly lower (Figure 5.3C, D).

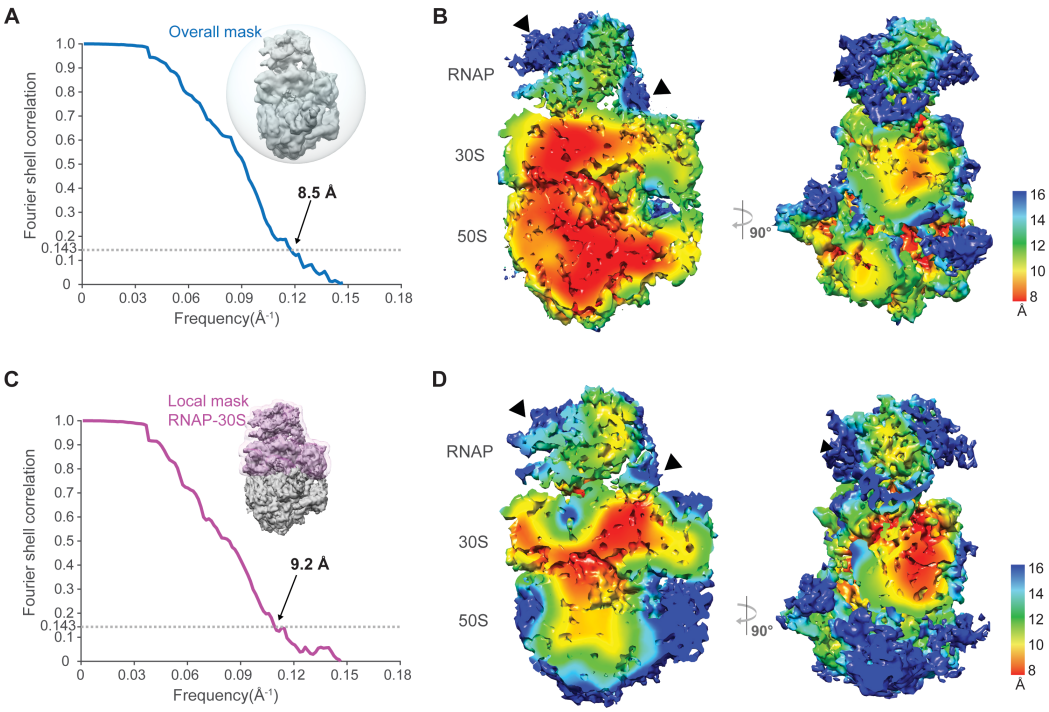

Figure 5.3 | The RNAP-ribosome supercomplex determined at sub-nanometer resolutions.

The 2,952 ribosome sub-tomograms with RNAP were refined with different strategies in RELION. (**A**) Refinement with a global spherical mask (insert) generated a map of 8.5 Å overall resolution. (**B**) Local resolution map. The alignment was dominated by the 50S, which caused low resolutions for the RNAP density (arrowheads). (**C**) Focused refinement with a solvent RNAP-30S mask (insert) resulted in an overall resolution of 9.2 Å. A spherical RNAP-30S mask resulted in the same resolution (not shown). (**D**) Local resolution map showing that RNAP and the interface to the 30S (arrowheads) have improved local resolutions.

Multibody refinement in RELION[305] was first tested in the sub-tomogram analysis pipeline, which helped further improve local densities and assess structural flexibility within the RNAP-ribosome supercomplex (Figure 5.4). Separate refinements of the three segmented bodies of the supercomplex, namely RNAP+ (RNAP and the linking density), 30S and 50S, yielded regional densities of higher local resolutions, which enabled unambiguous rigid-body fitting of homology models into the obtained density map, especially the RNAP part (Figure 5.4D). Additionally, multibody refinement showed that the relative motion of RNAP to the ribosome is large and continuous (Figure 5.4E).

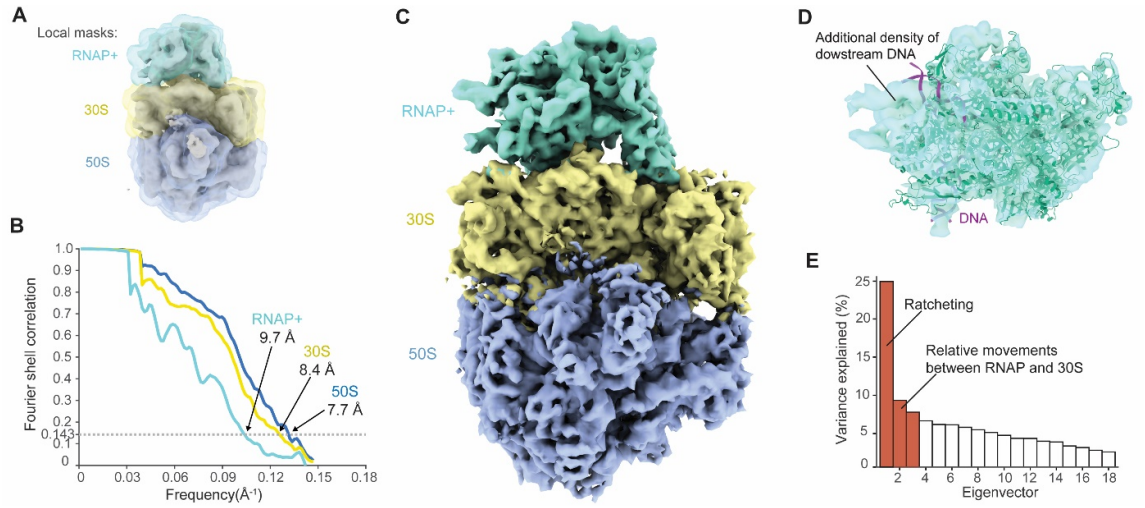

Figure 5.4 | Multibody refinement revealed high flexibility of the expressome.

(**A**) Solvent masks covering the three segmented bodies of the supercomplex, *i.e.* RNAP and linking density (RNAP+), 30S and 50S. (**B**) FCS curves for refinements of the three bodies. (**C**) Densities of the three bodies after multibody refinement. (**D**) Top view of the RNAP density with the fitted homology model (based on PDB 6FLQ). Only some regions near the downstream DNA and the flexible RNAP clamp region were not assigned. (**E**) Principal component analysis was performed to characterize relative movements between the three bodies. The major movements are ratcheting between ribosomal subunits, and relative movements between RNAP and 30S (see movie 1 in the publication[255]).

5.1.3 An unexplained density links the RNAP-ribosome supercomplex

Rigid-body fitting of homology models of RNAP and the ribosome into the obtained cryo-ET density map unambiguously explained most regions, except for the linking density between RNAP and the ribosome and a smaller density near the downstream template DNA and the flexible RNAP clamp (Figure 5.5). This is the first in-cell map of transcription-translation coupling supercomplex, and no similar structure has been reported before. Also because this is inside the cell, all proteins in *M. pneumoniae* proteome are in theory potential candidates to model the unexplained densities. Hence, additional information was needed to interpret the unexplained linking density in the cryo-ET map.

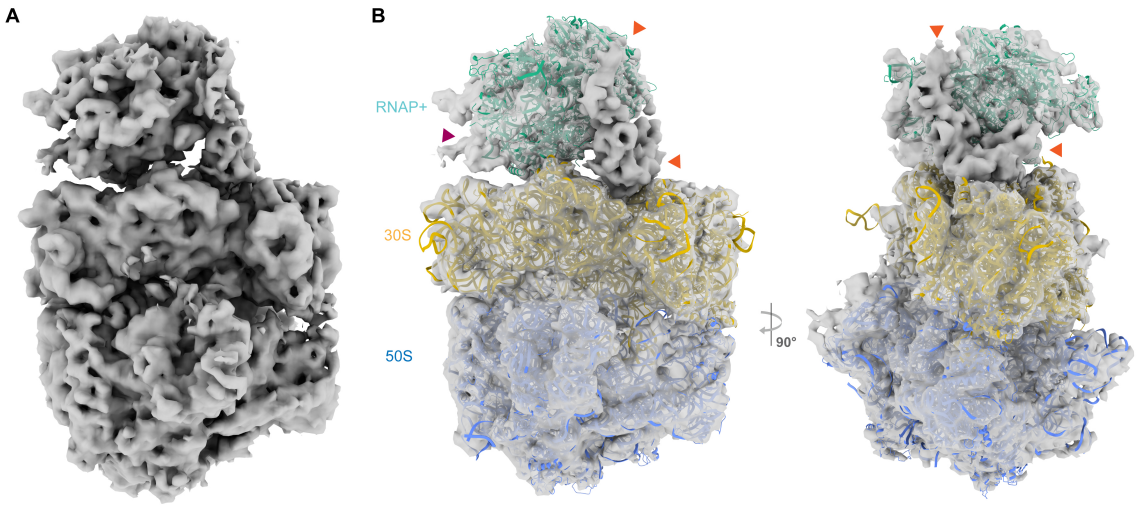

Figure 5.5 | In-cell structure of the RNAP-ribosome supercomplex.

(**A**) Cryo-ET density of the supercomplex determined from the 2,952 sub-tomograms. (**B**) The supercomplex density with fitted homology models of RNAP (based on PDB 6FLQ) and the ribosome (based on 3J9W). Most regions could be interpreted, except the region linking RNAP and 30S (orange arrowheads). Regions near the RNAP clamp and downstream DNA (red arrowhead) also have some unexplained density, which is partially because some flexible loops in RNAP clamp was not included in the homology model and a sequence of the template DNA was not modelled.

5.2 In-cell CLMS revealed NusA links RNAP and the ribosome

Crosslinking of intact *M. pneumoniae* cells with membrane-permeable crosslinkers can capture the protein-protein interaction networks inside cells[240], which can provide valuable compositional and topological information for interpreting the RNAP-ribosome supercomplex. The procedure for in-cell CLMS data collection and analysis can be found in detail in the collaborative paper with Juri Rappsilber's group[255], and is briefly summarized in Materials and Methods and illustrated in Figure 5.6A.

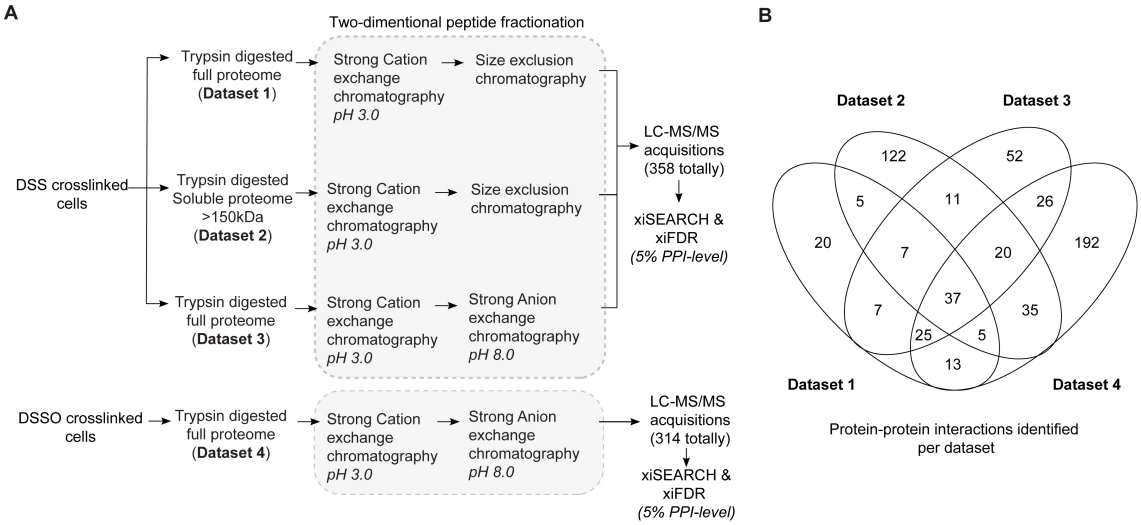

Figure 5.6 | Whole-cell CLMS of *M. pneumoniae*.

(**A**) The CLMS workflow for *M. pneumoniae* cells. Four separate datasets were collected, with different crosslinkers and peptide enrichment schemes. (**B**) Overlap of unique PPIs identified in the four datasets. The experiments, analysis and figure preparation were performed by Francis O'Reilly and Ludwig Sinn for the collaborative publication[255].

Through analysis of four in-cell CLMS datasets (Figure 5.6A), 10,552 crosslinks within the same protein (self-links) and 1,957 intra-protein crosslinks were identified with a residue-pair false discovery rate (FDR) of 5%. Among these, 577 distinct PPIs were identified at the 5% PPI-level FDR (Figure 5.6B). The identified crosslinks covered 83% of the detectable *M. pneumoniae* proteome, including membrane proteins, uncharacterized proteins and proteins in core complexes (DNA polymerase, RNA polymerase, ribosome, *etc.*). A comprehensive description of the in-cell CLMS data analysis and validation can be found in the collaborative publication[255]. Here, I only focus on crosslinks related to the transcription machinery and the translation machinery. Direct crosslinks between RNAP subunits and ribosomal proteins, which would be expected if direct RNAP-ribosome interaction happens as suggested previously[169-171], were not detected in the PPI network. Instead, NusA, an essential transcription factor involved in RNAP elongation, termination and antitermination [167,379-381], was found to link RNAP and the ribosome. The N-

terminal regions of NusA were crosslinked to several RNAP core subunits, which agrees with its reported binding site on RNAP[149,379-383]. The C-terminal sequence of NusA were crosslinked to several ribosomal proteins around the ribosome mRNA entry site on the small subunit (Figure 5.7B). Most known auxiliary factors of RNAP, including NusG, SigA, GreA, SpxA, and RpoE (also known as the RNAP δ subunit)[167,384,385] were identified in the network (Figure 5.7B), as well as two uncharacterized but essential proteins MPN555 and MPN530. Among these RNAP binders except for NusA, only the RNAP δ subunit showed one crosslink to the C-terminal loop extension of the ribosomal protein S6 (Figure 5.7B; also see Figure 3.12). The interaction between NusG and the ribosomal protein S10, the previously suggested transcription-translation linker in *E. coli*[162], was not detected in the network. Thus, in-cell CLMS results suggested that RNAP and the ribosome are linked by NusA in *M. pneumoniae*.

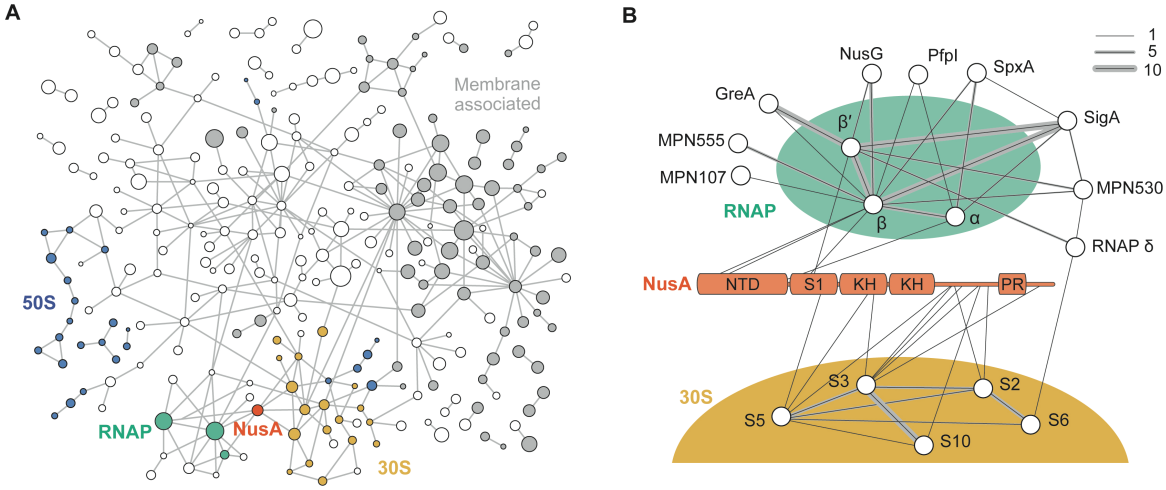

Figure 5.7 | Crosslink-based PPI network in *M. pneumoniae* cells.

(**A**) The network based on 577 distinct PPIs identified at 5% PPI-level FDR. Each circle represents one protein, with the diameter indicating relative protein size and the color indicating protein types. Grey: membrane-associated proteins. Green: RNAP subunits. Yellow: ribosomal proteins in 30S. Blue: ribosomal proteins in 50S. Orange: NusA. Each edge represents one or more crosslinks. For better visualization, interactions involving abundant chaperones were removed. (**B**) Part of the PPI network of RNAP, NusA and the ribosome. Only the interacting ribosomal proteins in the small ribosomal subunit are displayed. Crosslinks involving NusA domains (NTD, S1, KH1, KH2, and the proline rich region) are shown in details. The number of identified crosslinks per PPI is indicated by line thickness. This figure was made by Francis O'Reilly for the collaborative publication.

5.3 In-cell architecture of the elongating expressome by integrative modeling

Cryo-ET and in-cell CLMS both provided structural insights into the supramolecular assemblies during transcription-translation coupling inside native *M. pneumoniae* cells. However, one approach alone cannot bring about a complete picture of the architecture of the supercomplex. Hence, integrative structural modeling was employed to combine the cryo-ET and in-cell CLMS data (see Materials and Methods). Spatial restraints from the identified crosslinks were used to guide fitting of homology models into the cryo-ET density map and derive the integrative model of the expressome.

5.3.1 Preparation of building blocks for the expressome

To further validate the crosslinks and to prepare the starting models for integrative modeling, homology models of RNAP, the ribosome (before availability of the atomic model) and other auxiliary factors were first validated with densities and crosslinks. Fitting of the ribosome was confirmed first thanks to its large size, which could serve as the structural 'anchor' for the rest modeling (Figure 5.5). Mapping the 683 crosslinks within ribosomal proteins showed a good agreement with the model, except for several regions that are known to be intrinsically flexible (Figure 5.8A-C)[113].

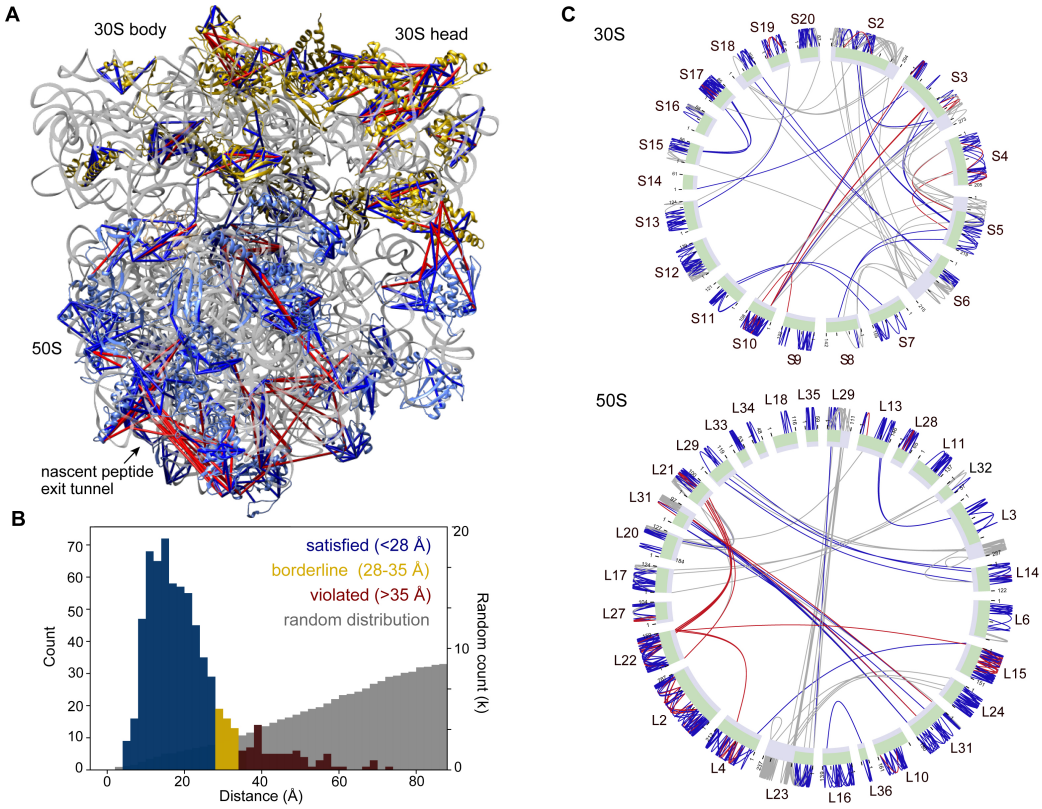

Figure 5.8 | Validation of crosslinks on the ribosome.

(**A**) Crosslinks mapped on the *M. pneumoniae* ribosome model. Links satisfying the distance restraint (<35 Å) are colored in blue, while violating (>35 Å) in red. Intrinsically flexible proteins (*e.g.* L1, L7/L12) and

unstructured regions are not shown. The regions that have most of the violating crosslinks are 30S head, intracellular interface, and the nascent peptide exit site, which are known to be flexible during translation. (**B**) Distance distribution of the 683 crosslinks within the ribosome compared to a theoretical random distribution (grey). (**C**) Crosslinks within ribosomal proteins of 30S and 50S respectively. Figure made by Andrea Graziadei for the collaborative paper[255].

Most crosslinks on RNAP also matched well with the homology model and the cryo-ET density (Figure 5.9A-C). Besides the RNAP core subunits (α, β, β'), many auxiliary proteins are known to form complexes with the RNAP core during different transcription phases[384,385] and were captured by in-cell CLMS (Figure 5.7B). After examining the identified RNAP-interacting proteins with the cryo-ET density and crosslink restraints, the binding of NusG, RNAP δ subunit and NusA was confirmed in the expressome supercomplex. The possibility of others RNAP-binding factors was excluded (details not shown here; can be found in the supplementary fig. S17 in the publication[255]).

The modeling of NusG in RNAP according to crosslinks and the cryo-ET density agrees with its reported binding site near the RNAP β' clamp and the upstream DNA[386,387]. As a general transcription factor, NusG has a relatively conserved NTD that interacts with RNAP, and a flexible CTD that has been shown to binds to various proteins[160,161]. In *E. coli*, NusG CTD interacts with NusE (a protein identical to the ribosomal protein S10), which was proposed to serve as a linker between RNAP and the ribosome during transcription-translation coupling[162]. In *M. pneumoniae*, NusG retains all conserved domains including the sequences involved in NusG-S10 interaction, yet its C-terminal sequence was predicated to have the extended sequence (Figure 5.9D). In the preliminary *M. pneumoniae* model based on both cryo-ET and CLMS data, NusG was placed away from S10 and their direct interaction appears impossible (Figure 5.9E). Additionally, both cryo-ET density and crosslinks supported the existence of the RNAP δ subunit[385], whose homology model was included as one module during integrative modeling. On top of these confirmed RNAP binders, NusA aligns best with the unassigned linking density between RNAP and the ribosome.

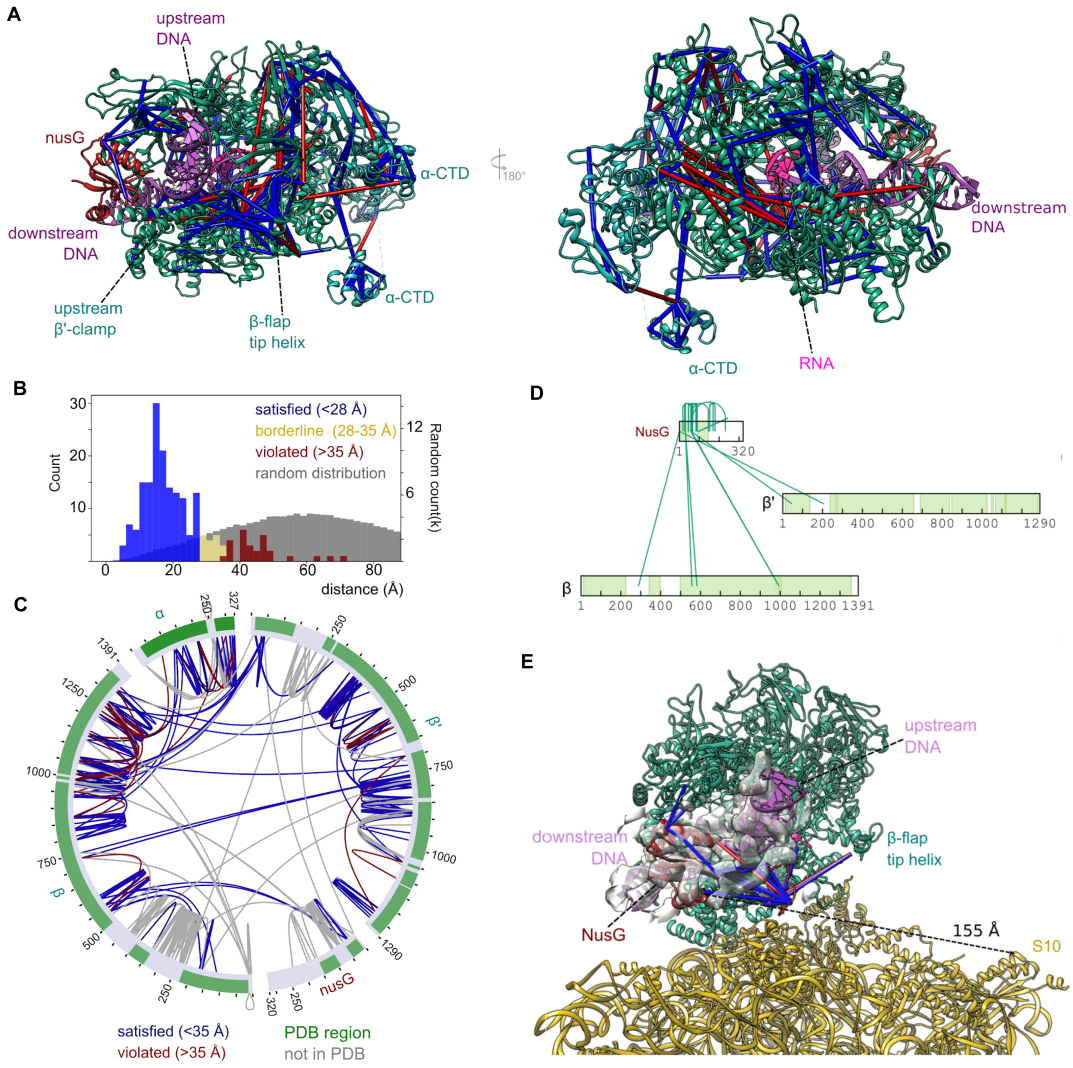

Figure 5.9 | Validation of crosslinks on RNAP and NusG.

(A) Crosslinks mapped on the RNAP homology model. NusG was placed in the canonical site according to a PDB structure (6C6U). Links satisfying the distance restraint (<35 Å) are colored in blue, while violating (>35 Å) in red. (**B**) Distance distribution of the 194 crosslinks in RNAP compared to a theoretical random distribution (grey). (**C**) All crosslinks on RNAP and NusG. (**D**) All crosslinks involving NusG. The structured parts of NusG was colored by light green, while its long C-terminal sequence (light grey) were predicated to be largely disordered. (**E**) The position of NusG showed good agreement between crosslinks, local density and the reported binding site, The previously reported NusG-S10 interaction is not possible (dashed line) in the model. Modified from figures by Andrea Graziadei for the collaborative paper.

NusA is a well-known RNAP binder involved in various transcription activities[167,379,380,388]. In *M. pneumoniae*, NusA has the conserved N-terminal regions compared to *E. coli* and *B. subtilis*, and the flexible C-terminus (Figure 5.10A, B). In *E. coli*, NusA can be divided into NTD, S1, KH1, KH2 and two acidic-repeat (AR) domains. NusA NTD is known to form the main contacts with RNAP core subunits, while the S.K.K. domains are shown to be able to interact with the nascent mRNA in several structures[379,380]. NusA in *B. subtilis* terminates after the KH2 domain and does

not have the AR domains. In *M. pneumoniae*, NusA's C-terminal sequence after KH2 was predicted to have low structural complexity and low conservation with the *E. coli* AR domains (Figure 5.10B). Most of it, however, was suggested to be essential by mutation experiments (see fig. S20 in the collaborative paper[255]). Thus, the C-terminal regions may represent structural and functional variations of NusA across bacteria species. The *M. pneumoniae* NusA homology model generated in this work contains only the conserved NTD and S.K.K. domains, which fitted well into the unassigned linking density between RNAP and the ribosome (Figure 5.10C, D). The C-terminal sequence after the KH2 domain was coarse-grained and left fully flexible.

After docking NusA into the unassigned density between RNAP and the ribosome, a small density extending to the RNAP α subunit was still not occupied (Figure 5.10D). It is assumed that this remaining density does not correspond to the C-terminal sequence of NusA because it was predicted to be mostly disordered and lacks crosslinks to RNAP core subunits (Figure 5.7A, B). Alternatively, the highly flexible but essential CTDs of RNAP α subunits[384], appeared to be the most plausible candidates based on the position and size. Hence, the flexible α-CTDs were included as separate modules for integrative modeling.

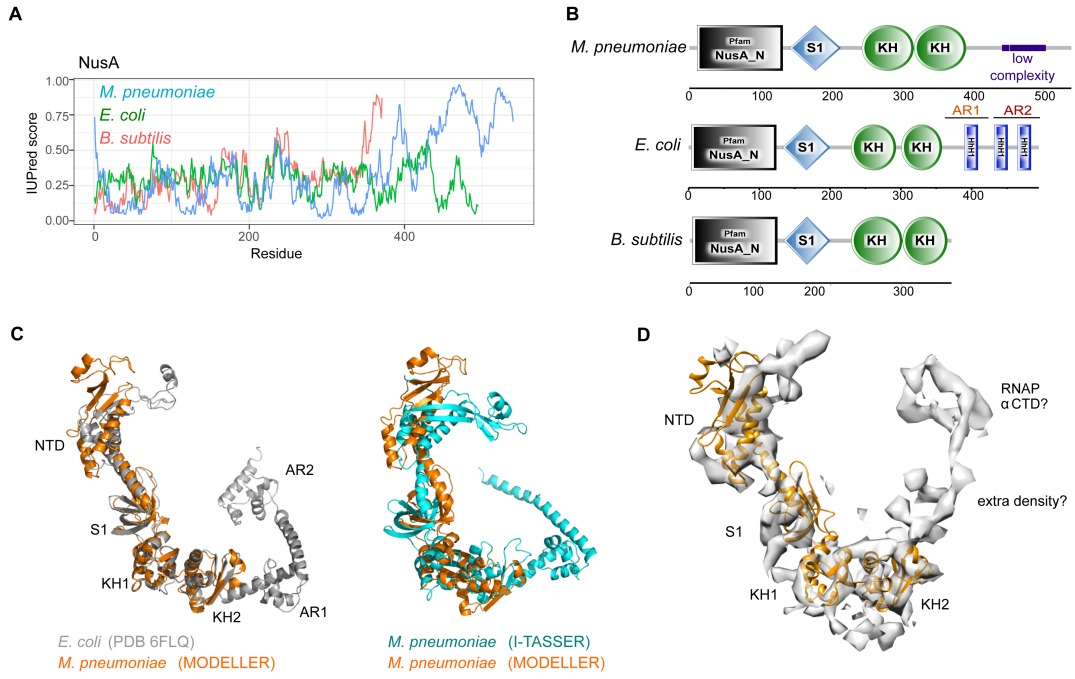

Figure 5.10 | NusA variations.

(**A**) Prediction of disordered regions for NusA using IUPred[389]. (**B**) NusA domains in *M. pneumoniae*, *E. coli* and *B. subtilis* according to the SMART database[390]. (**C**) *M. pneumoniae* NusA homology model by MODELLER was compared to the NusA in a reported *E. coli* transcription paused elongation complex (PDB 6FLQ) and to a predicted structure by I-TASSER[391]. (**D**) Rigid-body fitting of the NusA homology model into the linking density. The remaining part of the linking density may come from RNAP α-CTDs and NusA C-terminus. This figure was prepared by Andrea Graziadei for the collaborative paper.

5.3.2 Integrative modeling of the expressome

With the prepared homology models, the crosslinks and the cryo-ET density, integrative modeling of the expressome was performed using the integrative modeling platform[278] (Materials and Methods). As shown in Figure 5.11, different building blocks (RNAP core subunits, α-CTDs, NusG, RNAP δ subunit, NusA domains, 30S) were fitted into the cryo-ET density with distance restraints from crosslinks (Figure 5.12A). As the ribosome position was already determined by rigid-body fitting in the starting model, only the RNAP-NusA-30S part was included for computation and in most illustrations. The highest-scoring solutions in the integrative modeling were selected to derive the final expressome model. Here, I will describe the main modeling results of NusA and other important factors (Figure 5.12B-F).

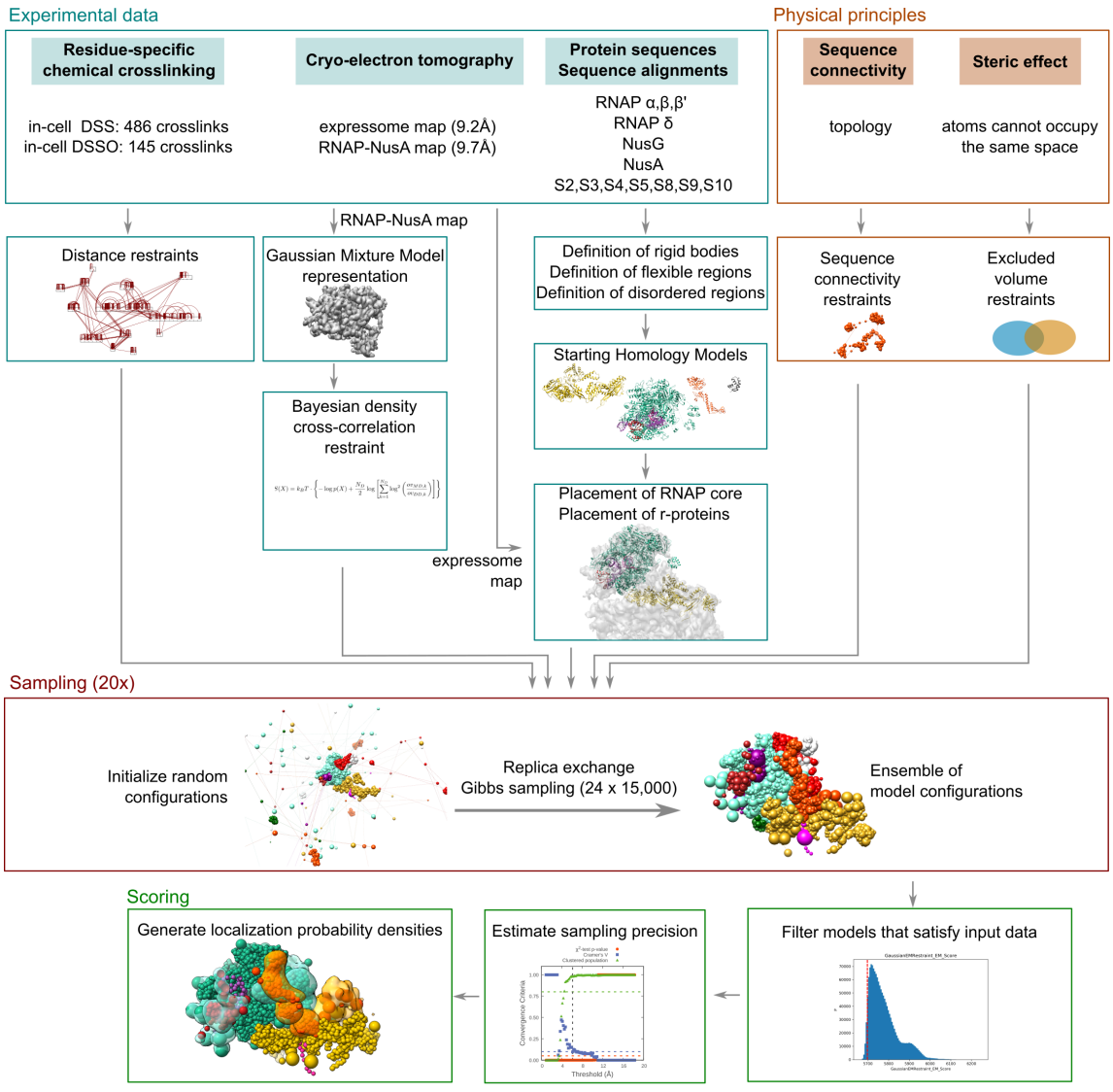

Figure 5.11 | Integrative modeling workflow.

The figure was prepared by Andrea Graziadei for the collaborative publication.

In the final model, NusA binds to RNAP in a similar manner to the *E. coli* RNAP paused elongation complex[380], with its NTD interacting with the β flat-tip helix and the S1 domain near the RNAP mRNA exit tunnel (Figure 5.12B, C). The KH1 and KH2 domains were positioned near the ribosomal protein S3, S4 and S5 around the ribosome mRNA entry site (Figure 5.12B). The KH domains adopt an orientation such that their main RNA-interacting surfaces face towards the putative mRNA path from the RNAP to the ribosome, and may chaperone the intertwining nascent mRNA between the two complexes[381,392]. The non-structured C-terminal sequence of NusA was placed in a large area on top of the 30S head, mainly in proximity to the ribosomal proteins S3, S9 and S10 (Figure 5.12B left panel and D). The integrative model clearly unraveled NusA as the linking part between RNAP and the ribosome in the expressome.

In the best modeling solutions, RNAP α-CTDs were found in a wide range of areas between NusA NTD and KH2 domains (Figure 5.12E, F). One α-CTD fits into the extra density between the NusA KH2 domain and the RNAP α subunit, which is close to NusA C-terminal sequences. This is reminiscent of the reported interaction between α-CTD and NusA AR2 in *E. coli*[393,394], which would place one α-CTD in the similar position. The second α-CTD was localized with relatively low precision, around NusA NTD and the RNAP mRNA exit site. It remains unclear whether there is an interaction between the second α-CTD and NusA NTD as indicated in the *E. coli* paused elongation complex[380]. Overall, our expressome model suggested that at least one of the two α-CTDs may play a role during transcription-translation coupling, which is consistent with the *in vivo* functional data in *E. coli*[169]. Since α-CTDs are known to be intrinsically flexible[384], it is expected that they are not fixed and can still have a certain level of flexibility within the expressome.

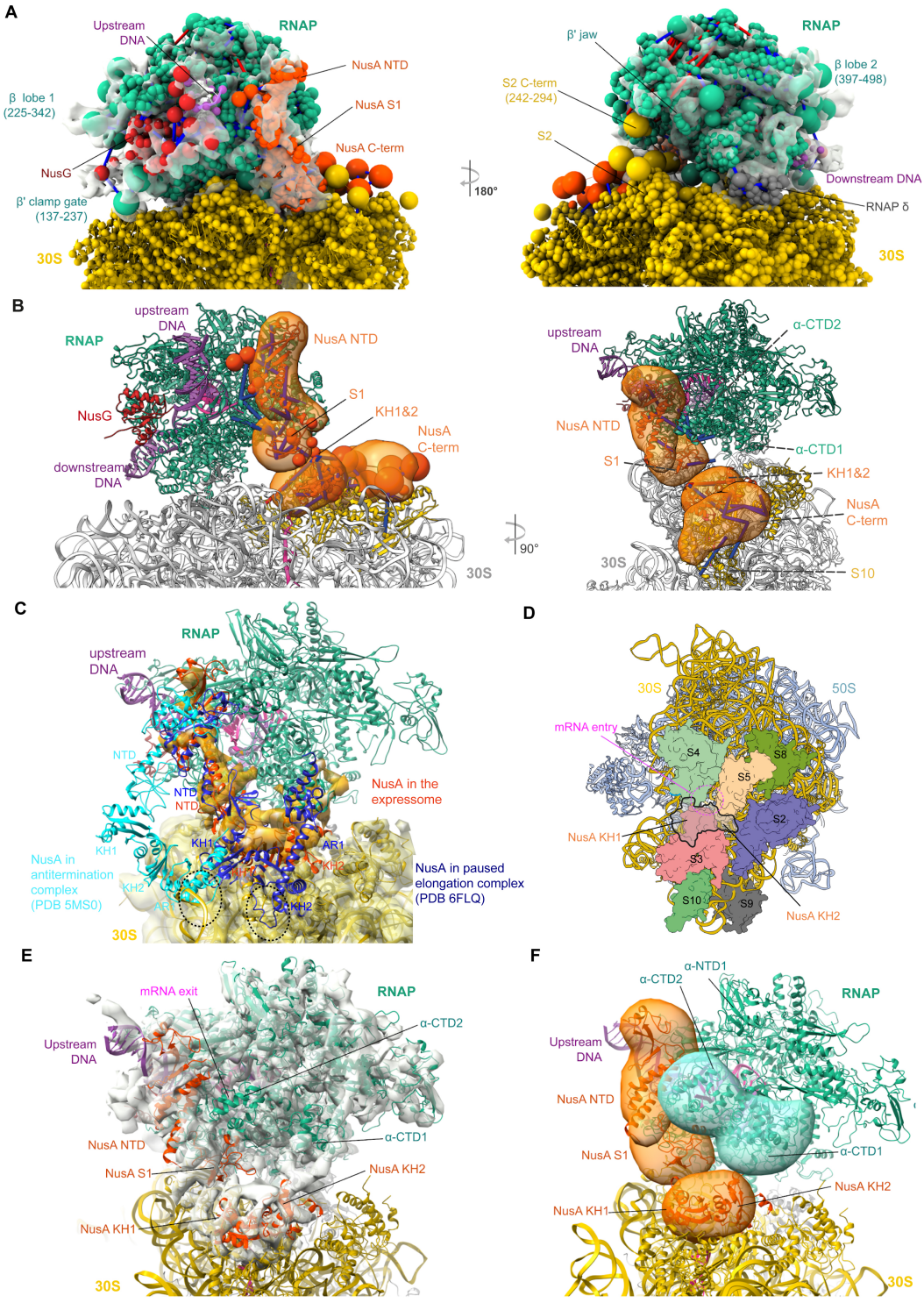

Figure 5.12 | Integrative modeling of NusA and other major components in the expressome.

(**A**) The centroid model showing the coarse-grained models (colored beads). Only RNAP, the RNAP-30S interface and part of 30S are displayed here. Cryo-ET density (transparent grey) and crosslinks (sticks; only partially visible) are also shown. (**B**) The localization probability map for NusA domains (orange). Crosslinks between NusA and components of RNAP and the ribosome are displayed (blue: satisfied, <35 Å. red: violated crosslinks). The flexible C-terminus of NusA is shown as beads in the left panel and removed in the right panel. (**C**) NusA in the *M. pneumoniae* expressome (orange) was compared to NusA

in two representative complexes: the λ-induced RNAP antitermination complex (cyan, PDB 5MS0) and the RNAP paused elongation complex (blue, PDB 6FLQ) in *E. coli*, after aligning the on RNAP. NusA in these complexes clashes with 30S regions (black dashed circles). (**D**) View of the RNAP-30S interface from RNAP. The main NusA contact region (black outline) is highlighted, close to the ribosomal proteins surrounding the mRNA entry site on 30S. (**E**) View focusing on the linking density between RNAP and the ribosome, with the modelled NusA (colored in orange) and RNAP α-CTDs. (**F**) The localization probability map for NusA domains (orange; not including the fully flexible C-terminus) and RNAP α-CTDs (light green). The structured domains of NusA shows good localization precision while RNAP α-CTDs are of relatively low precision. The α-CTDs were localized to a larger area between NusA NTD and the KH2 domain. Modified from figures prepared by Andrea Graziadei for the collaborative paper.

Additionally, the accessory RNAP δ subunit was positioned below RNAP β' CTD, near the downstream DNA binding cleft (Figure 5.13), in line with its reported roles in regulating RNAP-DNA interaction[385]. Our integrative model was the first in-cell structure showing RNAP with δ subunit, and the revealed binding site is consistent with recent studies in *B. subtilis*[395,396]. Similar to NusG and NusA, NTD of the δ subunit has defined structure and acts as the main RNAP-binding domain while the C-terminal regions are intrinsically flexible[397]. Interestingly, δ subunit was suggested to bias RNAP towards transcription of open-reading frames that have strong promoters or form more stable open complexes, and consequently favor expression of specific genome sequences[385]. Whether the observation of δ subunit in the expressome is related to the hypothesis that transcription-translation coupling preferentially happens in specific genes or sequences poses a very interesting question, which is, however, beyond the resolution and scope of this study.

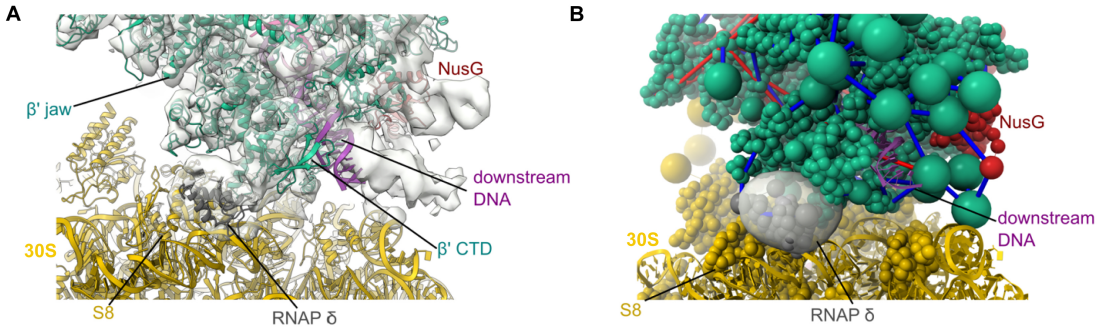

Figure 5.13 | Crosslink-assisted modeling of RNAP δ subunit in the expressome.

(**A**) The RNAP δ subunit (mainly the structured NTD) was placed near the RNAP β' CTD and the downstream DNA in the integrative model. The localization was supported by the extra density and several crosslinks by CLMS. (**B**) The localization probability map for RNAP δ. Crosslinks <35 Å are colored in blue and >35 Å in red. This figure was prepared by Andrea Graziadei for the collaborative publication[255].

5.3.3 In-cell architecture of the elongating expressome

Through integrative structural approaches, the first in-cell expressome structure in *M. pneumoniae* cells was determined. Components of the expressome including template DNA, RNAP, NusG, NusA, ribosome and translation factors were unambiguously modelled.

Transcription and translation are expected to be actively ongoing in native *M. pneumoniae* cells with minimal perturbations right before vitrification. For RNAP, the path of template DNA and about one turn of the DNA-RNA hybrid duplex were clearly resolved in the cryo-ET density, demonstrating that RNAP in the expressome is in an active elongating state (Figure 5.14A). Correspondingly, only dim density of single-stranded mRNA was observed, but no obvious folded mRNA near the exit site[398]. For the ribosomes within expressomes, their exact translation states were traced by combining results from the ribosome classification (as shown in Chapter 4), which showed that almost all (97%) of these ribosomes are in the translation elongation phase (Figure 5.14B). This indicated that the ribosomes in expressomes are actively elongating as well. These results demonstrated that the in-cell architecture represents an actively elongating expressome.

Intriguingly, the ribosome in the elongating expressome (*i.e.* the leading ribosome translating on mRNA that interacts with the transcribing RNAP) exhibited a different translation elongation state distribution profile compared to other ribosomes (Figure 5.14B). This difference may arise from: i) the interaction with RNAP/NusA or more generally cross-influence between the transcription and translation machines; ii) changes in local environment when such coupling happens; iii) different expressome assemblies exist in cells and the determined elongating expressome represents one of these. It is impossible to sort out these possibilities at this stage. More advanced sub-tomogram classification algorithms are be needed to further classify the 17,202 ribosomes with less-resolved additional density above the mRNA entry site (Figure 5.1 and 5.2).

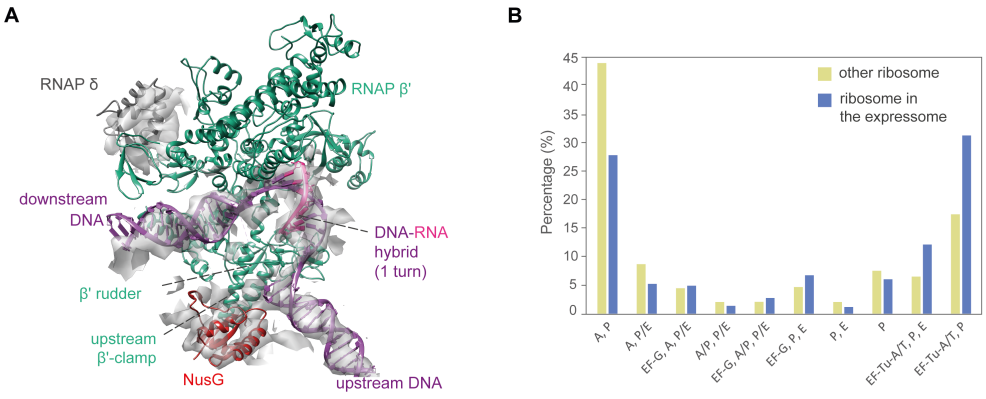

Figure 5.14 | In-cell expressome is actively elongating.

(**A**) DNA and one turn of hybrid DNA-RNA were nicely resolved in the expressome density, which agrees with the elongating conformation. RNAP δ subunit and NusG are also shown. (**B**) Translation elongation state distribution profiles for ribosomes in the expressomes in comparison to all other ribosomes. States

"EF-Tu•tRNA, P, E" and "EF-Tu•tRNA, P" are enriched for ribosomes within the expressomes, while the "A, P" state is underrepresented. Panel (A) was made by Andrea Graziadei for the collaborative paper. Panel (B) was made after combing the classification results of ribosomes and expressomes.

Both NusG and NusA were mapped in the expressome, which are general transcription factors that can continually bind to RNAP throughout transcription elongation[168]. Specifically, NusA was shown to bridge RNAP and the ribosome at the nexus of transcription-translation coupling, with its RNA-interacting S.K.K. domains positioned along the putative mRNA path (Figure 5.15A, B). As a versatile, multi-domain protein involved in transcription pausing, termination and antitermination[379-381,383,399], NusA has the potential to be reprogrammed by various factors to modulate transcription actives. Indeed, general RNAP binders like NusA, NusG and even the firmicutes-specific δ subunit share a similar pattern: they bind to critical regions on RNAP (DNA entry site, DNA exit site, mRNA exit site) via their conserved NTDs and have diverse C-termini to interact with diverse factors. In the expressome, NusA may act as a sensor of RNAP that detect the trailing ribosome and in return modulates transcription elongation through its interaction with the nascent mRNA[400]. The bridging of NusA keeps a certain level of movement freedom between the two gigantic molecular machines, protecting the RNAP from being harshly pushed by the trailing ribosome.

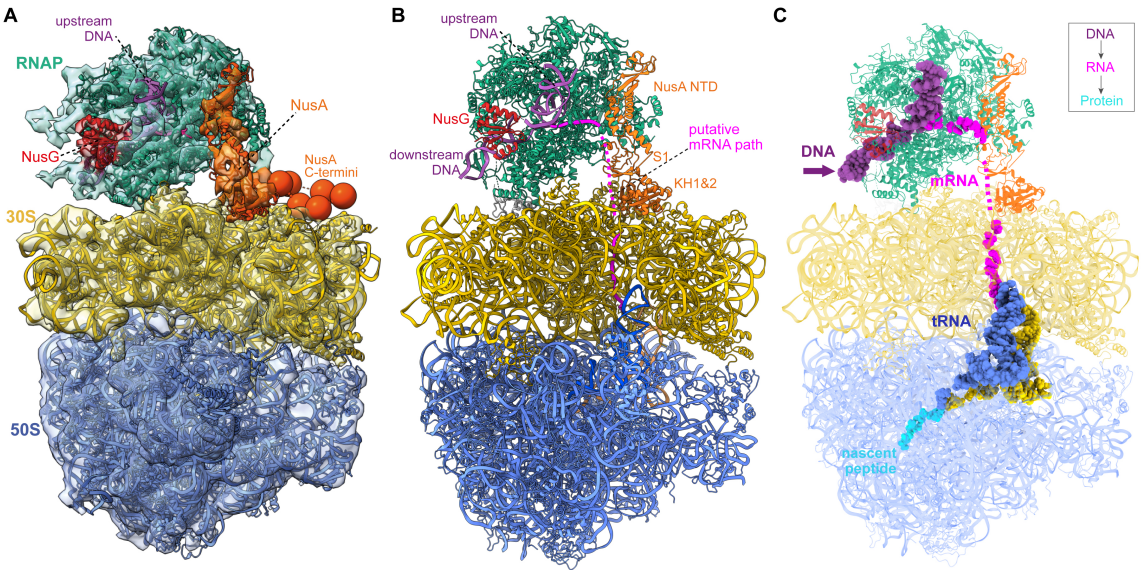

Figure 5.15 | Transcription-translation coupling mediated by NusA.

(**A**) In-cell architecture of the elongating expressome in native untreated *M. pneumoniae* cells. The integrative model is fitted in the cryo-ET density. Fully flexible regions of the RNAP clamp are removed. The flexible C-terminal sequence of NusA is shown as beads. (**B**) The elongating expressome model. The mRNA parts inside RNAP and the ribosome could be traced, while the intertwining mRNA was not resolved. The putative mRNA path (dashed magenta line) was drawn following the shortest distance between the RNAP exit site and the ribosome entry site. (**C**) Illustration of the genetic information flow in the elongating

expressome: DNA → mRNA → tRNA → nascent peptide. Panel (A) was made by Andrea Graziadei for the collaborative paper; panels (B-C) were made based on the integrative model.

To summarize, the elongating expressome structure solved directly in native cells revealed a novel transcription-translation coupling mechanism mediated by NusA (Figure 5.15). The genetic information flow from DNA to RNA (mRNA and tRNA, catalyzed by rRNA) to the protein, the central dogma of molecular biology, is vividly illuminated inside the cell.

5.4 Dynamics of transcription-translation coupling in cells

To probe dynamics of transcription-translation coupling in cells, ribosome sub-tomograms from *M. pneumoniae* cells treated with either a translation inhibitor (chloramphenicol or spectinomycin) or a transcription inhibitor (pseudouridimycin) were analyzed with the same classification procedure developed to identify the expressome in native untreated cells.

5.4.1 Inhibiting translation decouples the ribosome from RNAP

In *M. pneumoniae* cells shortly treated with Cm, an antibiotic that specifically binds to the large ribosomal subunit and inhibits peptide bond formation[361], no class of ribosomes showing potential RNAP density was detected (Figure 5.16A). Most ribosomes showed clear A- and P-site tRNAs and the state before peptidyl transfer (for detailed translation state classification see section 4.3.1), indicating a stalled state in most ribosomes. The finding of additional translation states (see section 4.3.1) and other reports showing low translation activity in Cm-treated bacterial cells[375,378], suggested that ribosomes in Cm-treated cells still can conduct translation but at relatively low rate. The stalled or slowed down ribosomes may not be able to elongate at the same rate as RNAPs.

Classification of ribosome sub-tomograms in Spt-treated cells did not generate any class with RNAP density above the mRNA entry site (Figure 5.16B). Upon the treatment of Spt which specifically inhibits 30S dynamics and the translocation step, most ribosomes were stalled in the expected pre-translocational state, while a minority were found in other translation elongation states (see section 4.3.2). Similar to Cm-treated cells, it is expected that there is low-level translation activity in Spt-treated cells.

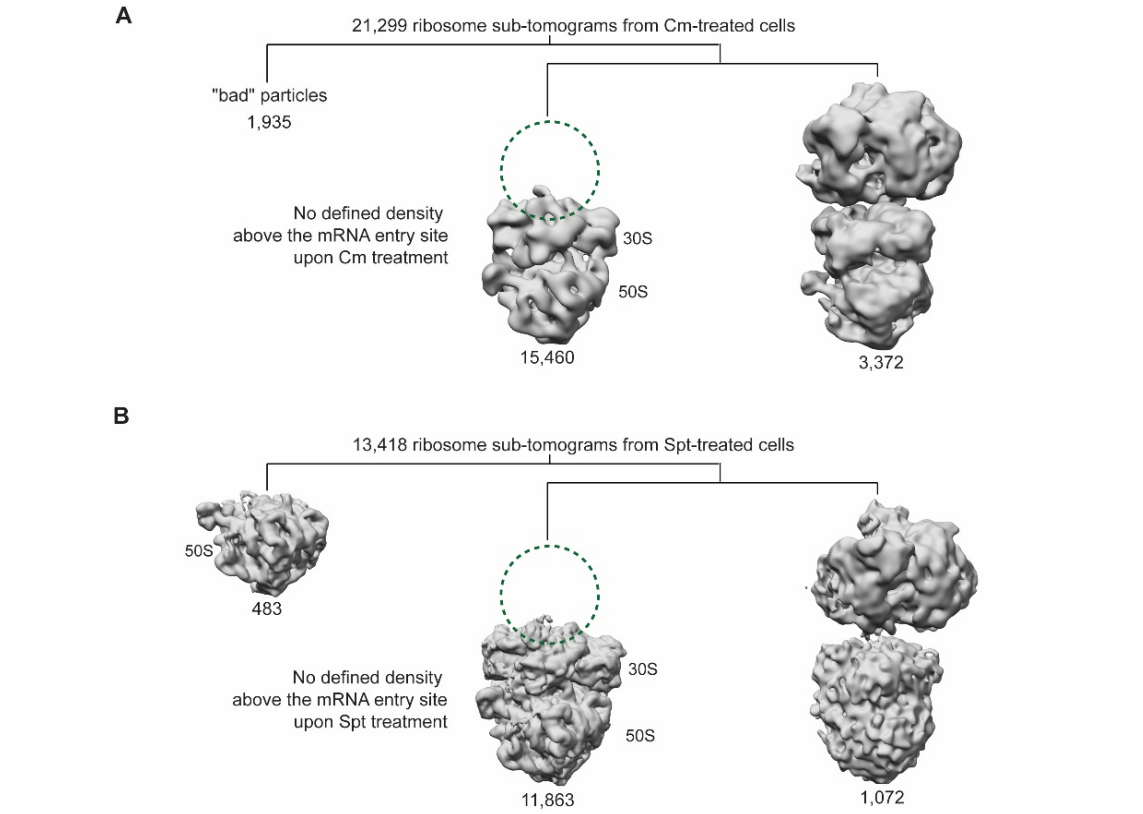

Figure 5.16 | Inhibiting translation decouples the expressome.

Focused classification around the mRNA entry site was performed for ribosome sub-tomograms in antibiotic-treated datasets. The ribosome numbers are slightly different from the numbers shown in previous chapters due to different processing time of the data. For both (**A**) chloramphenicol and (**B**) spectinomycin, no significant ribosome class with the RNAP density above the mRNA entry was classified, indicating that the elongating expressome architecture no longer forms.

These results demonstrated that inhibiting ribosomes, either stalling or slowing down translation, disrupts transcription-translation coupling in *M. pneumoniae* cells. The transcription machine comprised of RNAP and accessory factors can move forward and is no longer coupled to the lagging ribosome. This is reminiscent of other functional studies showing that transcription elongation is not affected by different ribosome antibiotics (Cm, and fusidic acid that inhibits translocation) in *E. coli*[176,177]. Although I cannot infer whether transcription elongation rate is changed or not upon treatment, it is clear that the interaction between the transcription and translation machines is insufficient to hold them together when translation is stalled or slowed down. Therefore, the elongating expressome assembly requires active translation in cells.

5.4.2 Stalling RNAP by PUM reshapes transcription-translation coupling in cells

How transcription-translation coupling responds to transcription inhibition was examined using cells treated with the RNAP-specific inhibitor PUM[367]. Surprisingly, the percentage of well-resolved expressomes significantly increased, from 4% of all 70S ribosomes in untreated cells to 60% in PUM-treated cells (Figure 5.17). This means that more than half of RNAPs in the *M. pneumoniae* cell are estimated to be directly coupled to the ribosome upon treatment, based on the reported number of 300 RNAPs[32] and the previous estimate of 300 to 500 ribosomes per cell (see section 4.2.1). Consequently, less polysomes were detected (~8% in PUM-treated *vs* ~26% in untreated) and the polysomes were found to be more compacted (Table 4.1). Moreover, the PUM-treated expressomes appeared to be more homogeneous and no other flexible class was detected.

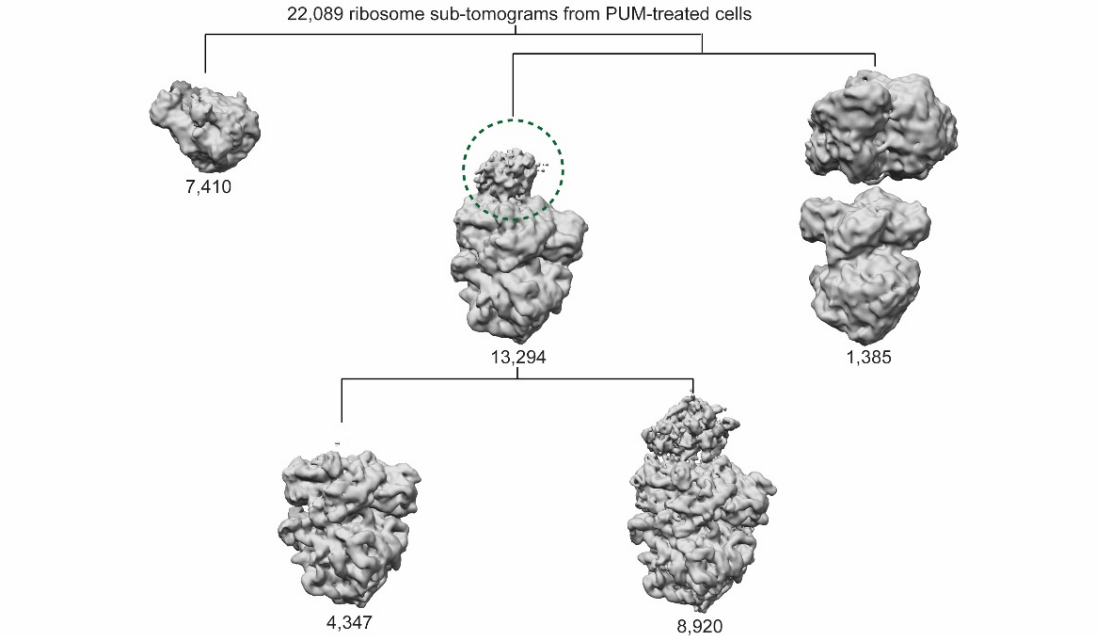

Figure 5.17 | Classification of the stalled expressomes in PUM-treated cells.

Similar focused classification around the mRNA entry site was performed for ribosomes in PUM-treated cells. The numbers here are slightly different from the numbers in previous chapters due to different data processing time.

Refinement of the PUM-treated expressome resulted in a structure of overall resolution 7.1 Å, showing direct interaction between RNAP and the ribosome, and the absence of NusA (Figure 5.18A-C). With multibody refinement, different parts of the PUM-treated expressome were better resolved, especially the RNAP density (Figure 5.18D-F). Long helices were clearly resolved in the RNAP density refined to 8.1 Å resolution (Figure 5.18E, G), which evinced that complexes on the scale of 450 kDa can be structurally determined at resolutions revealing secondary structure level information inside cells.

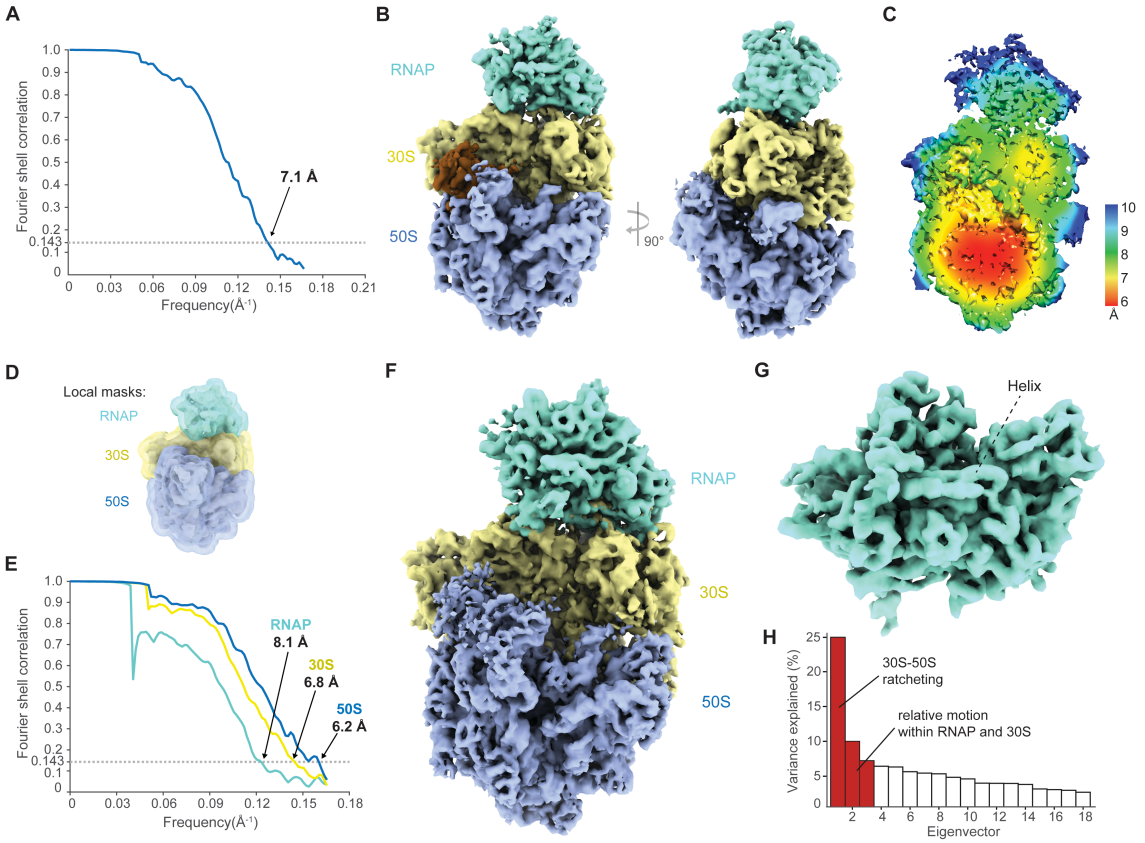

Figure 5.18 | Refinement of the stalled expressome in PUM-treated cells.

Global and multibody refinement of the PUM-treated expressomes were done in RELION. (**A**) FSC curves for the refinement with a global mask. (**B**) The refined expressome density, colored by different components. Brown: EF-G. (**C**) The local resolution map. (**D**) Multibody refinement with three local masks focusing on RNAP, 30S and 50S respectively. (**E**) FSC curves for the three parts after multibody refinement. (**F**) Illustration of the three separately refined bodies. (**G**) Top view of the RNAP. Some long helices are resolved. (**H**) Eigenvectors indicating the major relative movements within the expressome. Ratcheting between ribosomal subunits is the predominant motion and there are still motion between the closely associated RNAP and the ribosome.

Fitting of RNAP and ribosome models nicely explained the PUM-treated expressome density (Figure 5.19A, B). RNAP was found to form close contact with ribosomal protein S2, S3 and S5, with its mRNA exit site positioned right above the ribosome mRNA entry site. Though mRNA could only be partially traced inside RNAP and inside the ribosome, it is possible to draw a direct mRNA path from RNAP to the ribosome (Figure 5.19B). NusA was completely excluded in the structure, while the δ subunit and NusG remained bound to RNAP (Figure 5.19A, B). RNAP α-CTDs were not localized in the density, which would be placed right within the RNAP-30S interface and sterically clash with ribosomal proteins. Considering the flexible nature of RNAP α-CTDs[384], they may adopt other conformations outside the RNAP-30S interface. For the ribosome, the density for an extended EF-G together with hybrid A/P- and P/E-site tRNAs clearly indicated

that it is stalled in the pre-translocational state (Figure 5.19C). This was confirmed by combing results from the previous ribosome translation state classifications (Figure 5.19D), which demonstrated that the vast majority of ribosomes within the expressomes are in the "EF-G, A/P^{PUM}, P/E" state right before the mRNA movement occurs during translocation. Thus, both RNAP and the ribosome within the expressome are stalled upon PUM treatment. The ribosome is unable to complete the translocation step (specifically the relative mRNA movement), presumably due to physical obstruction by the PUM-stalled RNAP.

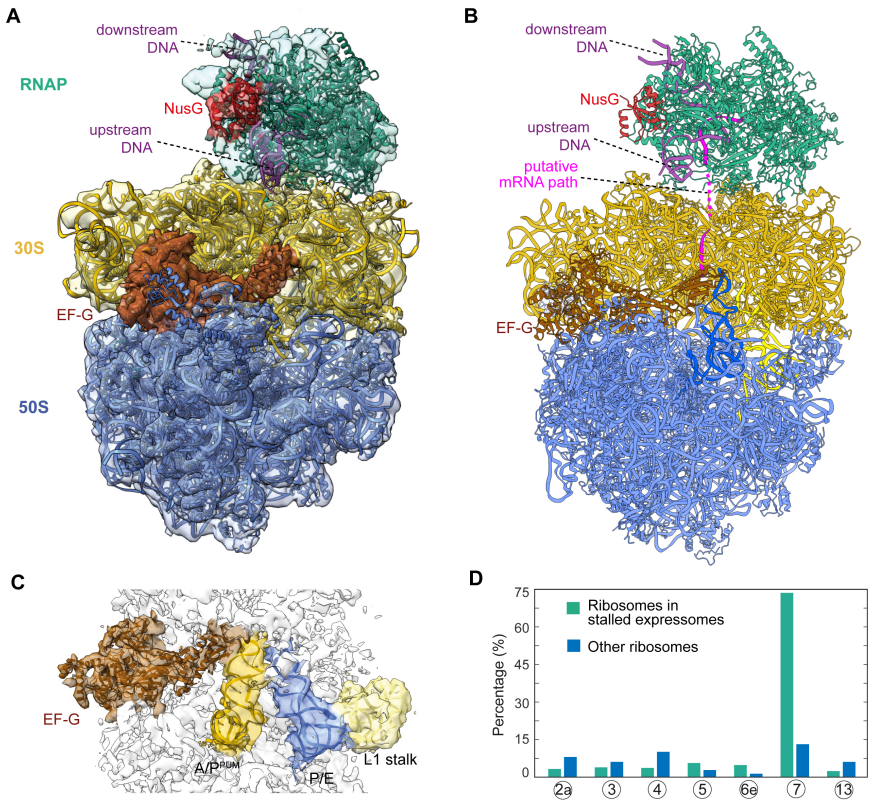

Figure 5.19 | Architecture of the stalled expressome upon PUM treatment.

(**A**) The stalled expressome in PUM-treated *M. pneumoniae* cells. The model is shown with the cryo-ET density. (**B**) Model of the stalled expressome. The dashed magenta line indicates the putative mRNA path from the RNAP exit site to the ribosome entry site. (**C**) EF-G, hybrid A/P- and P/E-site tRNAs were well-resolved in the stalled expressome. (**D**) Distributions of translation states for the ribosomes within the expressomes in comparison to the rest ribosomes, based on previous ribosome classification results (section 4.3.3). States were represented by number identifiers introduced before (see Figure 4.23). ⑦ is the pre-translocational "EF-G, A/P^{PUM}, P/E" state.

Strikingly, stalling RNAP by PUM fundamentally reshaped both transcription and translation. In addition to changing the expressome architecture, PUM rearranges transcription-translation coupling in cells. A large fractions of RNAPs (estimated over 50%) and ribosomes (60%) were found to be directly coupled. The PUM-stalled RNAP physically block translocation of the

collided translating ribosome, leading it to be arrested in the pre-translocational "EF-G, A/P, P/E" states. Furthermore, as discussed in the sections 4.3.3 and 4.3.5 in the previous chapter, there appears be a series of consequences arising from ribosome collision, including polysome compaction, potential ribosome rescue and other undiscovered processes, inside *M. pneumoniae* cells treated with PUM.

Chapter 6: Conclusions and Perspectives

6.1 In-cell structural biology with atomic detail

6.1.1 In-cell ribosome structure at 3 Å resolution

In this thesis, I determined the ribosome structure inside *Mycoplasma pneumoniae* cells at 3.0 Å resolution, the highest resolution for macromolecular complex obtained inside cells to date (Figure 6.1). The built atomic model represents the first ribosome structure for the genome-reduced mycoplasma species. The ribosome map determined in cells treated with the antibiotic Cm clearly showed the drug molecule in its canonical binding site. It not only provided the first structural evidence for a small molecule binding to its target inside the cell, but also revealed the atomic detail regarding its coordination and MOA. Specifically, a K^+ ion involved in Cm binding was unambiguously resolved in the large ribosomal subunit with local resolutions better than 2.9 Å. It is the first time to directly visualize a metal atom in ribosomes inside cells.

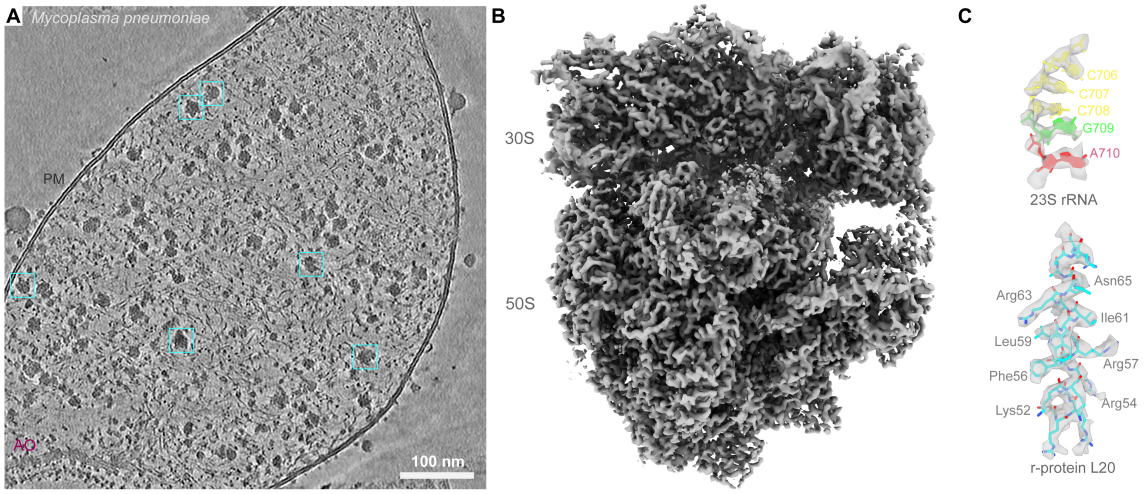

Figure 6.1 | Structural biology of *M. pneumoniae*: from the cell to atomic model.

(**A**) Tomographic slice of one *M. pneumoniae* cell revealed by cryo-ET. Several ribosomes are boxed out. (**B**) The 70S ribosome density map averaged from sub-tomograms extracted *in silico* from *M. pneumoniae* cellular tomograms. (**C**) Regions of the derived model fitted in the density map. This figure is a graphical summary of works in Chapter 3.

The *M. pneumoniae* ribosome model obtained in this work also highlighted ribosome diversity across bacteria. Distinct extensions in ribosomal proteins were found in ribosomes of *M. pneumoniae*, despite high structural similarity to other bacteria. Most extended sequences are disordered and are suggested to be not directly involved in the translation process. It remains to be clarified what are the possible roles of the ribosomal protein extensions, although they were found to be widely exist in the genome-reduced *M. pneumoniae* and other bacterial species.

Structural determination of ribosomes in non-model organisms like *M. pneumoniae,* will deepen our understanding of these ancient molecular machines, their evolution and diversity in adaptation to different lifestyles and environments.

6.1.2 Bring atomic structures into the cell

The in-cell ribosome structures obtained in this thesis pushed the resolution limit of cryo-ET and sub-tomogram analysis, and proved the feasibility of bringing atomic structures into the cell. To study molecular machines in the native functional context and link their functions with atomic detail have been the ultimate goals in structural biology[19,20,180]. This thesis highlights cryo-ET as the most promising approach for *in situ* structural biology and paves the way to achieving high resolutions inside cell.

I experimentally demonstrated the importance of B factors for high-resolution sub-tomogram averaging, which has not been systematically evaluated before. Data processing pipelines that existed at the beginning of this study resulted in $B_{overall}$ factors of over 1000 $Å^2$, which limit the attainable resolution to about 10 Å. Resolutions better than 3.5 Å were only practically possible when the $B_{overall}$ factor was reduced to below 100 $Å^2$ using a newly established image processing pipeline. Further analysis revealed that the main drawback originates from inaccurate registration from 2D images to 3D volume, mainly inaccurate tilt-series alignment and specimen deformation. High-resolution averaging requires particles to have lower B factors[313], that is, to maintain and maximize high-resolution information in the sub-tomograms. Structural heterogeneity is another factor that limits obtaining high resolutions, especially inside the cell where molecular machines undergo ensembles of conformations along their functional trajectories. Lastly, provided that the B factor is reduced to the lowest possible level, bigger data (larger numbers of particles) means larger possibility to achieve high resolutions. The data quantity is not an issue for *M. pneumoniae* ribosomes thanks to the high-throughput cryo-ET data collection[203] and relatively high abundance of ribosomes. For other rare or transient complexes, to have enough data is still a major bottleneck.

The Warp-RELION-M processing workflow that was first validated within this project provides the first comprehensive solution to tackle these limitations[234,255]. The multi-particle refinement software M developed by our collaborator enabled simultaneous average-based tilt-series and CTF refinement. Tilt-series, after refinement using ribosomes or other large complexes as "internal markers", have better alignments (global and local; including the deformation modeling) and CTF fitting, which can then be used to extracted sub-tomograms with lower B factors. Higher resolutions obtained in the averaging step also provided the basis for the following classification of compositional and conformational differences within the averaged sub-tomograms, which comprises the main advantage of in-cell structural studies. Maximum-likelihood classification implemented in RELION offers a comprehensive toolbox to deal with structural heterogeneity

within the sub-tomograms. Moreover, Warp and M can do distributed processing and combine cryo-ET datasets. Although such implementation still needs further testing, it provides a solution to do data pooling so that enough rare structures can be gained for averaging and classification. Recently, with the Warp-RELION-M pipeline or other pipelines with similar concepts, works by several other groups also showed the possibility to achieve 5 Å or better resolution *in situ,* including virus particles[401-404], bacteria surface S-layer[264], muscle thin filaments[405].

Atomic models can be brought into the cellular context with integrative structural approaches. The high abundance and large size of ribosomes make them an ideal case for sub-tomogram analysis, while other targets may have more difficulty in achieving comparable resolutions. The targets for cellular cryo-ET are usually complexes or assemblies of low abundance and high flexibility, and are thus less amenable to resolutions better than 4 Å that is required for *de novo* atomic modeling. On the other hand, atomic models of most proteins are available either by traditional structural approaches such as X-ray crystallography and single particle cryo-EM, or by structure prediction empowered by deep learning[14,15]. Integrative structural approaches combining in-cell maps of relatively low resolutions, existing atomic models of the components, and complementary information from methods such as in-cell CLMS can be used to derive atomic models of large complexes and assemblies[241,255,406]. Although there is no strict resolution limit for integrative modeling if enough additional information exists, it is still beneficial to push resolutions to 8 Å or better to enable unambiguous interpretation of the density map and high-quality modeling. A promising future direction is to use the experimental information (density maps, crosslinks, *etc.*) to guide deep learning based structure predication[407].

6.2 Structural and functional landscapes of translation inside cells

6.2.1 Structural dynamics of the translation machinery

Rapid freezing of the native *M. pneumoniae* cells that are supposed to maintain active cellular activities captured the translation machinery in action inside cells, and extensive classification of the complete ribosome sub-tomogram dataset extracted from cellular tomograms revealed structural dynamics of the translation process. The obtained ribosome classes represent structures of different intermediate states during translation. In particular, the complete translation elongation cycle can be reconstructed with ten of the thirteen distinct ribosome classes. They revealed continuous compositional and conformational changes of translation machines, including 30S body rotation, 30S head swiveling, elongation factor coordination, tRNA movement through the A-P-E sites, L1 stalk dynamics, *etc.* (Figure 6.2). The translation elongation cycle retrieved from one in-cell dataset complements current bacterial translation model derived from vast controlled *in vitro* studies. Moreover, the obtained high resolutions (4-10 Å) allowed resolving

bases of mRNA and tRNAs in the decoding center, and the associated nascent peptide chain in the peptidyl transfer center, providing a direct illustration of the information flow from the mRNA, via tRNAs, to the protein inside the cell.

Ribosome classification also revealed the distribution of different intermediate states along the functional trajectory, providing insights into energy landscapes of translation elongation inside cells. Such functional landscapes are assumed to result from both intrinsic properties of the translation machinery comprised of ribosomes and translation factors, and the intracellular environment like GTP concentrations. The work showed that individual native *M. pneumoniae* cells in the fast-growing phase exhibit relatively small variations across cells and tend to adopt a similar steady-state distribution of translation elongation states. Currently, only one previous work on isolated human polysomes[98] has comparable ribosome classification results, which revealed completely different translation state distribution from this in-cell study. Intriguingly, a recent cryo-ET study showed that the translation landscapes can be significantly different in two lift-out lamellas from *Caenorhabditis elegans* worms[133]. Therefore, it is possible that different translation energy landscapes exist for different organisms, cell types, cell conditions, and even sub-cellular regions for large eukaryotic cells. Application of the analysis introduced in this study to other organisms or conditions can help deepen the understanding of translation processes and their dynamics in the native functional context.

6.2.2 Spatial and functional organization of ribosomes in the cell

Mapping the classified ribosome structures back into the 3D cellular volume provided a glimpse of the spatial organization of translation inside the cell (Figure 6.2). Specifically, the organization of ribosomes into polysomes was analyzed, revealing two basic ribosome arrangement patterns and different topologies of polysomes. By combining the polysome detection with the translation state classification results, I found a local translation coordination mechanism mediated by the ribosomal protein L9. The stretched L9 of a ribosome within the polysome can interfere with elongation factor binding to the following ribosome and thus transiently block its elongation when they are too close. This local coordination mechanism maintains a certain distance between adjacent ribosomes and avoids harsh clash between elongating ribosomes within tightly assembled polysomes.

Spatial mapping of the determined structures is a major advantage of doing cryo-ET, and should be an integral part of sub-tomogram analysis beyond averaging and classification. The overall spatial precision of such mapping is on the same level of the resolution of the averaged structures (0.4-1 nm), which is one to two magnitudes higher than the spatial resolutions attainable with super-resolution microscopy[24,26]. Additionally, the orientations of each of the averaged particles are determined, rather than only position information. Thus, to some extent, cryo-ET and sub-

tomogram analysis can bring both structural and cell biology to a new level. This work demonstrated the power of such analysis, which reveals the spatial and functional organization of ribosomes and the undertaking translation process with unprecedented detail.

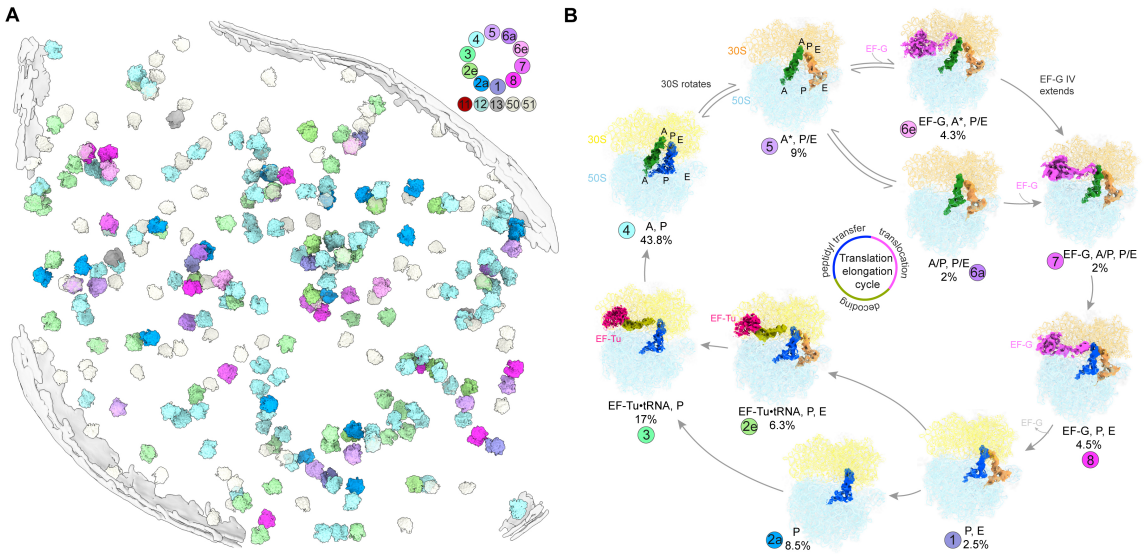

Figure 6.2 | Translation landscapes inside *M. pneumoniae* cells

This figure is a graphical summary of Chapter 4 on translation landscapes. (**A**) Mapping the classified ribosomes back into the 3D tomogram revealed spatial organization of ribosomes in thee *M. pneumoniae* cell. Different color means different state, along with the unique number identifier. (**B**) The reconstructed translation elongation cycle based on the ten classified 70S ribosome structures. The proportions of different intermediate states provide insights into energy landscapes of translation elongation in cells. For detailed results and discussions see Chapter 4.

6.2.3 Antibiotics fundamentally reshape translation landscapes in cells

The structural and quantitative profiling of ribosomes in cells treated with different antibiotics demonstrated how the translation machinery responds to different perturbations on the single cell level. The ribosome-specific chloramphenicol and spectinomycin stalled the majority of ribosomes in the specific translation states as expected based on previous *in vitro* studies. However, I found that there are additional classes of ribosomes in other translation intermediate states even under high concentrations of antibiotics. Moreover, the spatial organization of ribosomes, especially their formation of polysomes, was profoundly altered by antibiotics. In cells treated with the transcription inhibitor pseudouridimycin, both spatial organization and translation state distribution were significantly reshaped because of transcription-translation coupling. These results demonstrated that not only the energy landscapes but also the spatial organization of translation are fundamentally reshaped in cells upon various antibiotic treatment.

These results also provided new perspectives in understanding MOAs of antibiotics on the cell level. The traditional view of antibiotics as inhibitors of certain reaction steps may only constitute the primary MOA on the molecule level, while the reshaped machinery and related cellular processes make up the MOA on the cell level. This study and other recent reports[375,378] suggested low but ongoing translation in bacterial cells upon antibiotic treatment, possibly due to context-specific inhibition[363], mistranslation[408,409] or other unrevealed mechanisms. The antibiotic-reshaped translation machinery may lead to imbalance or disorder in gene expression in bacterial cells, which eventually results in detrimental consequences for the cell. It is noted that the final effect of a drug may depend on even higher levels, such as bacterial communities or populations[410]. Yet, a cell, the basic unit of life, is the key to comprehensively decipher the MOAs. The ribosome sub-tomogram analysis approach established in this thesis is possibly the most comprehensive and detailed method to probe translation profiles inside cells. Although translation profiles of individual *M. pneumoniae* cells analyzed in the single cell analysis exhibited low variations in response to different antibiotics, I expect that applying such analysis to different cells and conditions can expand the current MOA understanding of antibiotic drugs.

6.2.4 In-cell structural biology, from macromolecules to molecular sociology

Another conceptual advance this thesis contributes is that in-cell or *in situ* structural biology is to structurally describe the whole molecular machinery carrying out a specific cellular process, rather than only determining one structure within the cellular environment. On the macromolecule level, the built atomic models link the function with chemical detail. Complexes that are formed by macromolecules via relatively stable interactions, such as RNAP and the ribosome, can be isolated and studies with *in vitro* structural approaches. Supramolecular assemblies that are formed by relatively transient or weak interactions, such as the elongating expressome, may only be well studied inside the cell. Molecular machinery includes all molecular machines and accessory factors undertaking certain functions and processes in cells, and as revealed by the polysome study in this thesis, they have a higher level of complexity inside cells. Ultimately, molecular sociology aims to structurally map different molecular machineries within the cell[20]. Thus, in-cell structural biology is to attain and integrate information across scales to reveal the complete molecular landscapes in the cell.

This thesis shows a practical way toward such future: first, obtain high-resolution averages of the relatively stable and homogeneous parts of a molecular machinery; then, perform extensive classification to sort out the compositional and conformational heterogeneity, and resolve the flexible parts to interpretable resolutions; next, use integrative modeling to build models into the determined density maps; finally, map all detectable molecular machineries and gradually fill the cell with integrative models containing atomic detail. This in a way represents a computationally

"divide and conquer" strategy, in contrast to the traditional structural approaches that use the biochemically "divide and conquer" strategy. Sub-tomograms of complexes in the cell are computationally extracted from the cellular tomograms, aligned and averaged to generate the interpretable maps, and then the (pseudo-)atomic models are built in the maps. The major advance compared to traditional approaches is that the cell is intact and the obtained structures can be mapped back into the cell with high precision.

To achieve such goals, especially it is important to ensure the revealed structural and functional landscapes are representative of the native states in living cells. First, structures of complexes or assemblies, the distribution of structural ensembles along the functional trajectory and their spatial organization in native cells need to be maintained in the frozen-hydrated cells. During sample preparation, cells should be kept in the native environments and free from potential perturbation right before vitrification. Although molecular motions and landscapes are frozen after vitrification, it remains to be validated if the vitrification process is fast enough to capture the native states of cells in thicker samples like eukaryotic cells and tissues[188,189,194,411]. Second, the particle localization needs to be re-visited to ensure extracted sub-tomograms are representative and complete, because they determine the following sub-tomogram analysis. Third, the currently available sub-tomogram classification methods can only generate a number of discrete classes, which may not be able to reveal the full structural dynamics. To extract continuous structural dynamics information and quantitative data of the ensemble distribution, the classification also needs to be unbiased and exhaustive. It may be better to directly extract structural variability information from sub-tomograms, or most ideally, the original cellular tomograms. Emerging methods, especially those based on deep leaning, have been applied in single particle cryo-EM[412-415], yet not available for sub-tomogram analysis.

6.3 Deciphering transcription-translation coupling in cells

6.3.1 A novel transcription-translation coupling mechanism mediated by NusA

This thesis presents the first in-cell structure of supramolecular assemblies (expressomes) formed during transcription-translation coupling, by combing cryo-ET and in-cell CLMS data of *M. pneumoniae*. At the RNAP-ribosome interface, NusA, an essential and general transcription factor, was found to bridge the transcription and translation machines, following the path of the intertwining mRNA at the nexus of the expressome (Figure 6.3). NusA may enable RNAP to detect the approaching ribosome and responsively modulates transcription elongation. In the expressome, the bridging and mediating role of NusA can help avoid direct collision and maintain sufficient flexibility between the transcription and translation machines when they are actively elongating (base step *vs* codon step).

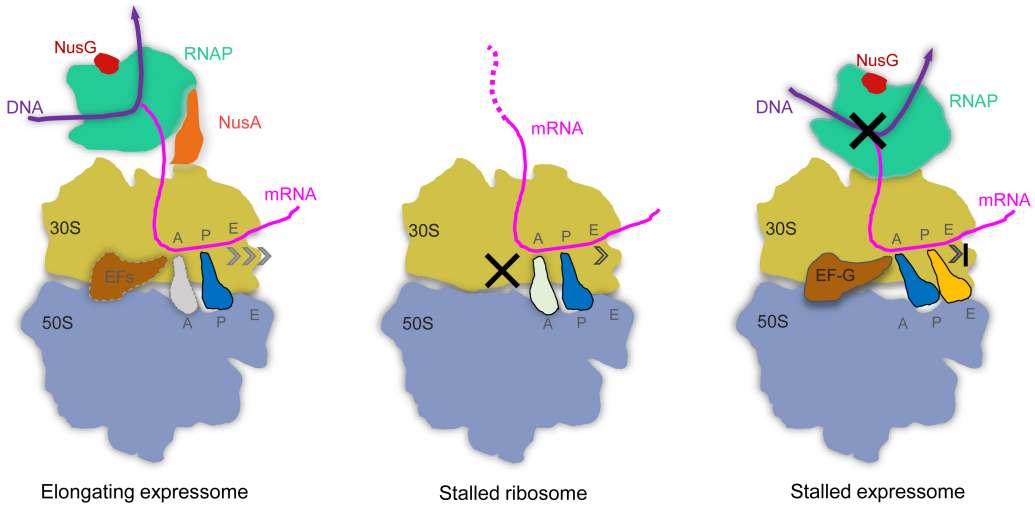

Figure 6.3 | Transcription-translation coupling is a dynamic process in cells.

A graphical summary of results and discussions of Chapter 5 on expressomes. Left to right: illustrations of transcription-translation coupling states in cells that are native untreated, ribosome antibiotic treated and RNAP antibiotic treated. In native untreated cells, the actively transcribing and translating expressome is mediated by NusA. The supercomplex is decoupled when the ribosome is inhibited by either Cm or Spt, suggesting its dependence on active translation. Stalling RNAP by PUM fundamentally reshaped transcription-translation coupling in cells, causing more ribosomes to collide with RNAP.

Moreover, this study demonstrate that transcription-translation coupling is a dynamic process in cells. Inhibiting translation decoupled the ribosome from the RNAP, while inhibiting transcription not only altered the expressome architecture but also completely reshaped landscapes of transcription and translation. Formation of the elongating expressomes requires both active transcription and translation within the native cellular context.

6.3.2 New landmarks in transcription-translation coupling research

Structural research of transcription-translation coupling has entered a new phase. The work presented in this thesis reported the first in-cell architecture of the *M. pneumoniae* expressome in 2020[255], which was followed by two single particle cryo-EM studies of *E.coli* expressomes[165,166]. A comparison of major expressome structures is shown in Figure 6.4.

There is no structural similarity between the in-cell *M. pneumoniae* expressome and the *in vitro* *E. coli* expressomes (Figure 6.4), except for confirming that the first *in vitro* expressome structure by Kohler *et al*[169] represents a collided model of RNAP and the ribosome. All *in vitro* structures are based on reconstituted systems consist of specially designed hybrid DNA-RNA scaffolds to mimic fixed transcription bubbles, and purified RNAPs and ribosomes from *E. coli*. Kohler *et al* did not specifically control the distance of spacing mRNA between RNAP and the ribosome. By contrast, ribosomes of expressomes in the two recent systems presented by Webster *et al*[166] and Wang *et al*[165] are fixed in certain positions on mRNA and their distance to RNAP is strictly controlled by the spacing mRNA sequence. It is thus possible that the ribosome in the system by Kohler *et al* is continuously translating along the mRNA until clashing with RNAP. This can be explained by our PUM-stalled expressome, whose overall structure is identical to the structure by Kohler *et al*. Webster *et al* and Wang *et al* also determined similar collided expressome structures when the spacing mRNA is short (Figure 6.4). The direct RNAP-ribosome interaction model may not be able to accommodate the dynamic elongation process and may only happen under certain circumstances.

By controlling length of the spacing mRNA and addition of NusG, Webster *et al* and Wang *et al* determined the similar NusG-linked *E. coli* expressome structures (Figure 6.4), which provided the structural basis for the previously proposed NusG-linking model[162]. Additionally, by adding additional NusA into the *in vitro* system, Wang *et al* solved a NusA-containing expressome architecture with which they suggested a role of NusA in flexibly maintaining the constant interaction between RNAP and 30S body. However, in their model NusA's conformation and interaction interface are completely different from the *M. pneumoniae* model. Interesting, in the expressomes by Webster *et al* the ribosomes all have A- and P-site tRNAs, while in the ones by Wang *et al* they all have P- and E-site tRNAs. These indicate the absence of active translation in both cases. Although the spacing mRNAs of different lengths may be able to mimic position changes between RNAP and the ribosome, the *in vitro* reconstituted systems cannot reflect the native cellular context where active transcription and translation are ongoing. Nevertheless, the *in vitro* structures revealed structural dynamics during transcription-translation coupling. Also, in this study, there is a class of ribosomes with unresolved additional densities above the ribosome mRNA entry site, which may result from structural heterogeneity during the dynamic coupling process.

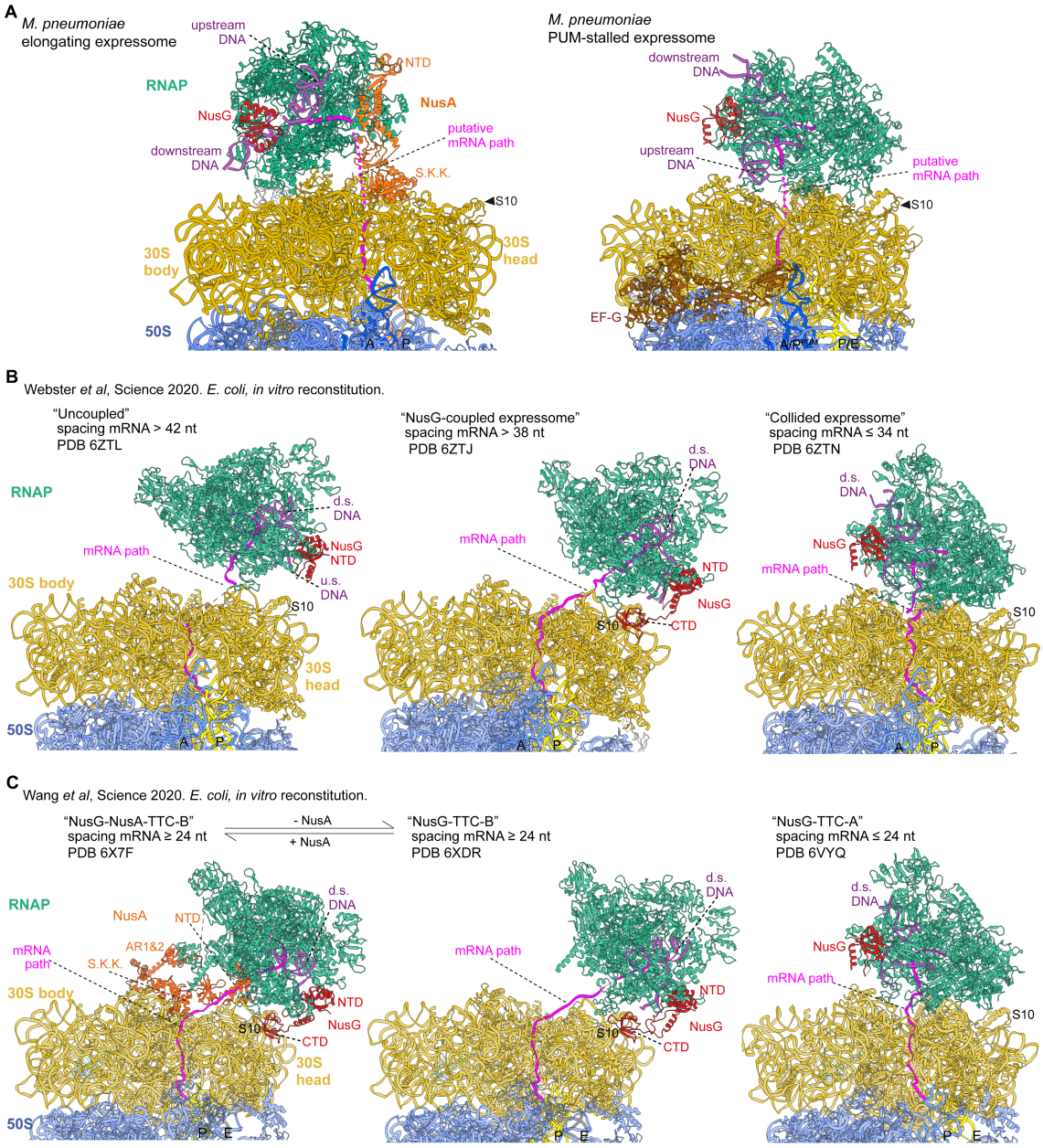

Figure 6.4 | Comparison of major expressome structures.

All structures are aligned on the ribosome part. (**A**) The in-cell *M. pneumoniae* elongating (left) and stalled (right) expressomes. (**B**) The *in vitro E. coli* expressomes by Webster *et al*. Left panel: if the spacing mRNA is >42 nucleotides, the expressome is flexibly linked by mRNA. Middle panel: when the spacer is >38 nt and NusG is present, the expressome is linked by NusG whose NTD binds to RNAP and CTD binds to 30S protein S10. Right panel: when the spacer is much shorter, it is a collided RNAP-ribosome. (**C**) The *in vitro E. coli* expressomes (termed TTCs by the authors) by Wang *et al*. Their spacing length definition is different from (B). Left & Middle panels: when the spacing mRNA is long and NusG is present, a similar structure (NusG-TTC-B) to the middle panel in (B) was determined. Additionally, when NusA is added, a new structure (NusG-NusA-TTC-B, left panel) was resolved. Right panel: when the spacer is short, a similar a collided model (NusG-TTC-A) was formed. The structure by Kohler *et al* (PDB 5MY1, not shown here) is similar to the stalled or collided structures (right panels in (A), (B) and (C)). The *E. coli* expressome structures used here are from the references[165,166].

This thesis and other recent works also highlight the species-related difference in transcription-translation coupling. The NusA-mediating model is based on *M. pneumoniae,* whereas the direct interaction model and the NusG-linking model are all based on *E. coli*. A NusG-linked expressome could not be detected in *M. pneumoniae* cells with deeper classification. Considering the high conservation of NusG and the interaction interfaces between RNAP and the ribosome among *M. pneumoniae*, *E. coli* and other bacterial species[161,165,166,169,255], I expect the *in vitro* assemblies may still occur inside bacterial cells albeit with low occurrence probabilities. On the other hand, it remains to be tested whether the NusA-mediating model represents a conserved feature across bacteria. In particular, high variations in the C-termini of NusA (Figure 5.10), may represent a source for differences across bacterial species. Interestingly, a recent publication by Johnson *et al* indicated that transcription and translation are fundamentally uncoupled in *B. subtilis*[416], raising the question whether coupled RNAP-ribosome is a universal feature across bacteria.

In summary, the transcription-translation coupling process are demonstrated to be extremely diverse and dynamic. Yet, the current models need more validation inside different cells, preferentially with in-cell structural approaches and cell biology approaches such as super-resolution microscopy.

6.3.3 Understanding transcription-translation coupling inside the cell

Beyond the discovery of new expressome structures, this thesis showed the importance and potential of studying dynamic molecular processes inside the cell. As gigantic molecular machines working on DNA and mRNA, whether and how RNAP and the ribosome are structurally coupled depend on temporospatial organization of the transcription and translation activities inside cells. It is superficial to view transcription-translation coupling as simple protein-protein interactions or the expressome as a fixed complex. The mechanisms to functionally coordinate transcription and translation and to form the physical coupling, are anticipated to be diverse, dynamic and transient, which may only be explicitly studied inside cells. It has been suggested that in big bacterial cells such as *E. coli*, transcription-translation coupling mainly happen in the peripheral regions around the ribosome-excluding nucleoid[345]. Although no specific pattern was identified for the expressome distribution in *M. pneumoniae* cells (data not shown) due to their simplicity (*e.g.* lack of a defined nucleoid), it is expected that the same cryo-ET and sub-tomogram analysis workflow applied to other bacterial cells will provide more insightful data. A future research direction following this thesis, through sub-tomogram analysis of all ribosomes and RNAPs, is to decipher landscapes of key molecular machineries of gene expression inside cells.

This thesis focused on transcription-translation coupling in *M. pneumoniae* cells in the fast-growing phase. It will be interesting to see how transcription-translation coupling change in response to different environments. This would be better combined with multi-omics techniques

such as transcriptomics and metabolomics to monitor levels of the proposed regulator (p)ppGpp and other relevant factors[173]. Whether transcription-translation coupling happens for all operons/genes or only specific sequences poses another interesting question to be addressed in the future. To reveal real temporospatial organization of transcription and translation in cells, I foresee that super-resolution microscopy[24,26], especially its combination with cryo-ET[266,268,270], will enable new opportunities in the study of transcription-translation coupling, or more generally, the coordination between transcription and translation in cells.

Transcription-translation coupling is a potential target for antibiotic drug development or at least a new perspective to understand MOAs of certain antibiotics in cells. For example, the newly developed antibiotic PUM is a RNAP specific drug but its impact on the translation machinery is probably a major cause to terminate a bacterial cell. Taking advantage of the high diversity of transcription-translation coupling mechanisms in bacterial cells, antibiotics targeting on it may achieve better specificity and lower toxicity.

6.4 A vision for future structural biology

Structural biology provides mechanistic understanding of life. Traditional structural approaches with isolated samples, as well as the rapidly evolving structure prediction, have enabled atomic models for most proteins. To bring these structures into the cell, the basic unit of life, is a main future direction. The methodological and conceptual advances presented in this thesis will further propel efforts toward this direction. This thesis proved the feasibility and enormous potential of integrative in-cell structural approaches in elucidating dynamic molecular machineries within the native functional context. It heralds the emergence of integrative in-cell structural biology.

Looking into the bright future of in-cell structural biology, I would propose there are four principals to follow: cellular, physiological, systematic and integrative. A cell is the basic unit of life and is also the hub to integrate different structural information. On one hand, it contains macromolecules and complexes; and on the other, it constitutes tissues, organs and organisms. To maintain the native physiological states of cells is not limited to avoiding potential perturbations. Ideally, cell samples should be directly taken from their native context, which means original environment (ocean, soil, guts, *etc.*) for single living cells, and the tissue and organ for multicellular organisms. Systematic structural description of molecular machineries or molecular sociology is necessary to understand the complexity of cells and biological systems on larger scales. Integrative structural approaches can tackle the complexity, and to efficiently combine experimental and existing information will be vital. System-wide structural approaches such as cryo-ET and in-cell CLMS, together with the powerful deep learning methods, lay the foundation for a golden age of structural biology.

List of main works and conclusions

- Collected and processed high-quality *M. pneumoniae* cryo-ET datasets of more than 850 cellular tomograms. These provide the basis for the works presented in this thesis, ongoing *M. pneumoniae* projects involving many groups at EMBL, benchmarking and method developments.
- Validated a newly developed Warp-RELION-M data processing workflow; provided the major experimental proof for methods developed by collaborators.
- Determined in-cell 70S ribosome at 3 Å resolution, the highest for cellular sub-tomogram averaging reported to date. Directly revealed antibiotics and ions in the ribosome inside the cell.
- Built the first atomic model of ribosomes inside the cell, which also represents the first ribosome model for mycoplasma species and enables discovery of distinct ribosome protein extensions.
- Determined over 13 translation intermediate structures (4-10 Å resolutions) that reconstruct the full elongation cycle and reveal structural dynamics of the translation machinery inside native cells
- Obtained the most comprehensive insights into the spatial and functional organization of the translation machinery in bacterial cells.
- Proposed a local translation elongation coordination mechanism within polysomes mediated by the ribosomal protein L9.
- Designed the antibiotic treatment experiments to probe translation dynamics; determined more than 17 ribosome intermediate structures in three antibiotic-treated cell datasets.
- Direct visualization of the antibiotic molecule chloramphenicol and spectinomycin binding to the translation machineries inside cells.
- Revealed how antibiotics fundamentally reshape the spatial and functional landscapes of translation in bacterial cells, which also provided novel insights into MOAs of antibiotics in cells.
- Determined the first in-cell architecture of an actively transcribing and translation expressome.
- Proposed a novel transcription-translation coupling mechanism mediated by the transcription factor NusA by combining cryo-ET and in-cell CLMS, which also serves as the primary proof-of-concept for the emerging field of integrative in-cell structural biology.
- Demonstrated that the native expressome requires active transcription and translation elongation within the cellular context, using antibiotic perturbation experiments
- Solved the in-cell structures of membrane-bound ribosomes, the ribosome-Sec translocon complex, and their changes upon antibiotic treatment. (ongoing works not shown; the same for below two).
- Obtained the first structural evidence for the transertion hypothesis.
- Discovered a cage-like complex on the cell surface. Determined a new supramolecular assembly formed by the ribosome, translocon and the surface cage-like complex upon antibiotic treatment.

List of publications and contributions

- F. O'Reilly*, **L. Xue***, A. Graziadei*, L. Sinn, S. Lenz, D. Tegunov, C. Blötz, N. Singh, W. Hagen, P. Cramer, J. Stülke, J. Mahamid, J. Rappsilber. "In-cell architecture of an actively transcribing-translating expressome". **Science** 2020.
  *equal contribution. Highlighted by Nature Reviews Microbiology, Faculty Opinions, *etc*.

Solely contributed the cryo-ET part, including sample preparation, data collection, image processing and map determination. Initiated the dynamics research using antibiotics. Equal contributions to manuscript writing. Obtaining the first evidence (>15 Å to ~5 Å) for high-resolution in-cell structure determination and the early experimental proof for Dimitry Tegunov's multi-particle refinement concept.

- W. Tichelaar, W. Hagen, T. Gorelik, **L. Xue**, J. Mahamid. "TEM bright field imaging of thick specimens: nodes in Thon ring patterns". **Ultramicroscopy** 2020.

My data validated Willem Tichelaar's observation of the node-like patterns in the power spectra of images of thick specimens with biological samples.

- D. Tegunov, **L. Xue**, C. Dienemann, P. Cramer, J. Mahamid. "Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 Å in cells". **Nature Methods** 2021.

I provided the Cm-treated *M. pneumoniae* tomography data to Dimitry Tegunov and helped early software testing. We together pushed the ribosome maps to 3.5 Å resolution, and the 3.5 Å ribosome structure in Cm-treated cells serves as the major breakthrough supporting the publication. I first experimentally validated the full Warp-RELION-M workflow. Contribution to the related manuscript writing.

- B. Lucas, B. Himes, **L. Xue**, T. Grant, J. Mahamid, N. Grigorieff. "Locating macromolecular assemblies in cells by 2D template matching with cisTEM". **Elife** 2021.

Collected a special EM+ET dataset and provided template matching results for Nikolaus Grigorieff to validate their new approach. Initiated and collected a 2D image dataset of *M. pneumoniae* cells, which serves as the main proof-of-concept. Contributed to figure preparation and manuscript writing.

- X. Zeng, A. Kahng, **L. Xue**, J. Mahamid, Y. Chang, M. Xu. "DISCA: high-throughput cryo-ET structural pattern mining by deep unsupervised clustering". Preprint on bioRxive; **Nature Communication**, in review.

Provided the tomography data and template matching results to validate the new tomogram mining method.

- **L. Xue**, S. Lenz, M. Zimmermann-Kogadeeva, D. Tegunov, P. Cramer, P. Bork, J. Rappsilber, J. Mahamid. "Visualizing translation dynamics at atomic detail inside a bacterial cell". Preprint on bioRxive; Top 5% influence by Altmetric. **Nature**, in review.

Lead author. Main contribution in project design, data collection, structure determination and manuscript writing; equal contributions in atomic model building.

References

1. Watson, J.D. & Crick, F.H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737-8 (1953).
2. Kendrew, J.C. et al. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**, 662-6 (1958).
3. Kendrew, J.C. et al. Structure of myoglobin: A three-dimensional Fourier synthesis at 2 A. resolution. *Nature* **185**, 422-7 (1960).
4. Shi, Y. A glimpse of structural biology through X-ray crystallography. *Cell* **159**, 995-1014 (2014).
5. Science's special section on Crystallography at 100 (7 March 2014). *Science* (2014).
6. Pellegrini, C. The development of XFELs. *Nature Reviews Physics* **2**, 330-331 (2020).
7. Luchinat, E. & Banci, L. In-cell NMR: a topical review. *IUCrJ* **4**, 108-118 (2017).
8. Kuhlbrandt, W. The resolution revolution. *Science* **343**, 1443-4 (2014).
9. Cheng, Y. Single-particle cryo-EM-How did it get here and where will it go. *Science* **361**, 876-880 (2018).
10. Henderson, R. & Unwin, P.N. Three-dimensional model of purple membrane obtained by electron microscopy. *Nature* **257**, 28-32 (1975).
11. Yin, C.C., Han, H., Wei, R. & Lai, F.A. Two-dimensional crystallization of the ryanodine receptor Ca2+ release channel on lipid membranes. *J Struct Biol* **149**, 219-24 (2005).
12. Wisedchaisri, G., Reichow, S.L. & Gonen, T. Advances in structural and functional analysis of membrane proteins by electron crystallography. *Structure* **19**, 1381-93 (2011).
13. Nannenga, B.L. & Gonen, T. The cryo-EM method microcrystal electron diffraction (MicroED). *Nat Methods* **16**, 369-379 (2019).
14. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
15. Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871-876 (2021).
16. Tunyasuvunakool, K. et al. Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590-596 (2021).
17. Oikonomou, C.M. & Jensen, G.J. A new view into prokaryotic cell biology from electron cryotomography. *Nat Rev Microbiol* **14**, 205-20 (2016).
18. Hutchings, J. & Zanetti, G. Fine details in complex environments: the power of cryo-electron tomography. *Biochem Soc Trans* **46**, 807-816 (2018).
19. Robinson, C.V., Sali, A. & Baumeister, W. The molecular sociology of the cell. *Nature* **450**, 973-82 (2007).
20. Beck, M. & Baumeister, W. Cryo-Electron Tomography: Can it Reveal the Molecular Sociology of Cells in Atomic Detail? *Trends Cell Biol* **26**, 825-837 (2016).
21. Lucic, V., Forster, F. & Baumeister, W. Structural studies by electron tomography: from cells to molecules. *Annu Rev Biochem* **74**, 833-65 (2005).
22. Mahamid, J. & Baumeister, W. Cryo-electron tomography: the realization of a vision. ***Microscopy and Analysis*** **26**, 45-48 (2012).
23. Turk, M. & Baumeister, W. The promise and the challenges of cryo-electron tomography. *FEBS Lett* **594**, 3243-3261 (2020).
24. Gwosch, K.C. et al. MINFLUX nanoscopy delivers 3D multicolor nanometer resolution in cells. *Nat Methods* **17**, 217-224 (2020).
25. Sahl, S.J., Hell, S.W. & Jakobs, S. Fluorescence nanoscopy in cell biology. *Nat Rev Mol Cell Biol* **18**, 685-701 (2017).
26. Sigal, Y.M., Zhou, R. & Zhuang, X. Visualizing and discovering cellular structures with super-resolution microscopy. *Science* **361**, 880-887 (2018).
27. Wan, W. & Briggs, J.A. Cryo-Electron Tomography and Subtomogram Averaging. *Methods Enzymol* **579**, 329-67 (2016).

28. Schur, F.K. Toward high-resolution in situ structural biology with cryo-electron tomography and subtomogram averaging. *Curr Opin Struct Biol* **58**, 1-9 (2019).
29. Waites, K.B. & Talkington, D.F. Mycoplasma pneumoniae and its role as a human pathogen. *Clin Microbiol Rev* **17**, 697-728, table of contents (2004).
30. Chaudhry, R., Ghosh, A. & Chandolia, A. Pathogenesis of Mycoplasma pneumoniae: An update. *Indian J Med Microbiol* **34**, 7-16 (2016).
31. Himmelreich, R. et al. Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae. *Nucleic Acids Res* **24**, 4420-49 (1996).
32. Kuhner, S. et al. Proteome organization in a genome-reduced bacterium. *Science* **326**, 1235-40 (2009).
33. Guell, M. et al. Transcriptome complexity in a genome-reduced bacterium. *Science* **326**, 1268-71 (2009).
34. Yus, E. et al. Impact of genome reduction on bacterial metabolism and its regulation. *Science* **326**, 1263-8 (2009).
35. Ochman, H. & Raghavan, R. Systems biology. Excavating the functional landscape of bacterial cells. *Science* **326**, 1200-1 (2009).
36. Wodke, J.A. et al. MyMpn: a database for the systems biology model organism Mycoplasma pneumoniae. *Nucleic Acids Res* **43**, D618-23 (2015).
37. Seto, S. & Miyata, M. Cell reproduction and morphological changes in Mycoplasma capricolum. *J Bacteriol* **180**, 256-64 (1998).
38. Hasselbring, B.M., Jordan, J.L., Krause, R.W. & Krause, D.C. Terminal organelle development in the cell wall-less bacterium Mycoplasmapneumoniae. *Proc Natl Acad Sci U S A* (2006).
39. Henderson, G.P. & Jensen, G.J. Three-dimensional structure of Mycoplasma pneumoniae's attachment organelle and a model for its role in gliding motility. *Mol Microbiol* **60**, 376-85 (2006).
40. Krause, D.C. et al. Electron cryotomography of Mycoplasma pneumoniae mutants correlates terminal organelle architectural features and function. *Mol Microbiol* **108**, 306-318 (2018).
41. Trussart, M. et al. Defined chromosome structure in the genome-reduced bacterium Mycoplasma pneumoniae. *Nat Commun* **8**, 14665 (2017).
42. Lluch-Senar, M. et al. Defining a minimal cell: essentiality of small ORFs and ncRNAs in a genome-reduced bacterium. *Mol Syst Biol* **11**, 780 (2015).
43. Dandekar, T. et al. Re-annotating the Mycoplasma pneumoniae genome sequence: adding value, function and reading frames. *Nucleic Acids Res* **28**, 3278-88 (2000).
44. Christodoulides, A. et al. The Role of Lipoproteins in Mycoplasma-Mediated Immunomodulation. *Front Microbiol* **9**, 1682 (2018).
45. Blotz, C., Singh, N., Dumke, R. & Stulke, J. Characterization of an Immunoglobulin Binding Protein (IbpM) From Mycoplasma pneumoniae. *Front Microbiol* **11**, 685 (2020).
46. Wodke, J.A. et al. Dissecting the energy metabolism in Mycoplasma pneumoniae through genome-scale metabolic modeling. *Mol Syst Biol* **9**, 653 (2013).
47. Dallo, S.F. & Baseman, J.B. Intracellular DNA replication and long-term survival of pathogenic mycoplasmas. *Microb Pathog* **29**, 301-9 (2000).
48. Maier, T. et al. Quantification of mRNA and protein and integration with protein turnover in a bacterium. *Mol Syst Biol* **7**, 511 (2011).
49. Bitnun, A. & Richardson, S. Mycoplasmal and Ureaplasmal Infections. in *Infections of the Central Nervous System* (eds. Marra, C.M., Scheld, M.W. & Whitley, R.J.) 420-433 (Wolters Kluwer Health, 2014).
50. Karr, J.R. et al. A whole-cell computational model predicts phenotype from genotype. *Cell* **150**, 389-401 (2012).
51. Hutchison, C.A., 3rd et al. Design and synthesis of a minimal bacterial genome. *Science* **351**, aad6253 (2016).
52. Rees-Garbutt, J. et al. Designing minimal genomes using whole-cell models. *Nat Commun* **11**, 836 (2020).
53. Seybert, A., Herrmann, R. & Frangakis, A.S. Structural analysis of Mycoplasma pneumoniae by cryo-electron tomography. *Journal of Structural Biology* **156**, 342-354 (2006).

54. Melnikov, S. et al. One core, two shells: bacterial and eukaryotic ribosomes. *Nat Struct Mol Biol* **19**, 560-7 (2012).
55. Petrov, A.S. et al. History of the ribosome and the origin of translation. *Proc Natl Acad Sci U S A* **112**, 15396-401 (2015).
56. Bernier, C.R., Petrov, A.S., Kovacs, N.A., Penev, P.I. & Williams, L.D. Translation: The Universal Structural Core of Life. *Mol Biol Evol* **35**, 2065-2076 (2018).
57. Bowman, J.C., Petrov, A.S., Frenkel-Pinter, M., Penev, P.I. & Williams, L.D. Root of the Tree: The Significance, Evolution, and Origins of the Ribosome. *Chem Rev* **120**, 4848-4878 (2020).
58. Luria, S.E., Delbruck, M. & Anderson, T.F. Electron Microscope Studies of Bacterial Viruses. *J Bacteriol* **46**, 57-77 (1943).
59. TISSIÈRES, A. & WATSON, J.D. Ribonucleoprotein Particles from Escherichia Coli. *Nature*, 778–780 (1958).
60. Huxley, H.E. & Zubay, G. Electron microscope observations on the structure of microsomal particles from Escherichia coli. *Journal of Molecular Biology* **2**, 10-IN8 (1960).
61. Lake, J.A. Ribosome structure determined by electron microscopy of Escherichia coli small subunits, large subunits and monomeric ribosomes. *J Mol Biol* **105**, 131-9 (1976).
62. FRANK, J. & GOLDFARB, W. Reconstruction of glutamine synthetase using computer averaging. *Ultramicroscopy* **3**, 283-290 (1978).
63. Radermacher, M., Wagenknecht, T., Verschoor, A. & Frank, J. Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of Escherichia coli. *J Microsc* **146**, 113-36 (1987).
64. Wagenknecht, T., Carazo, J.M., Radermacher, M. & Frank, J. Three-dimensional reconstruction of the ribosome from Escherichia coli. *Biophysical Journal* **55**, 455-464 (1989).
65. Frank, J. et al. A model of protein synthesis based on cryo-electron microscopy of the E. coli ribosome. *Nature* **376**, 441-4 (1995).
66. Agrawal, R.K. et al. Direct visualization of A-, P-, and E-site transfer RNAs in the Escherichia coli ribosome. *Science* **271**, 1000-2 (1996).
67. Stark, H. et al. Visualization of elongation factor Tu on the Escherichia coli ribosome. *Nature* **389**, 403-6 (1997).
68. Agrawal, R.K., Penczek, P., Grassucci, R.A. & Frank, J. Visualization of elongation factor G on the Escherichia coli 70S ribosome: the mechanism of translocation. *Proc Natl Acad Sci U S A* **95**, 6134-8 (1998).
69. Agrawal, R.K., Heagle, A.B., Penczek, P., Grassucci, R.A. & Frank, J. EF-G-dependent GTP hydrolysis induces translocation accompanied by large conformational changes in the 70S ribosome. *Nat Struct Biol* **6**, 643-7 (1999).
70. Frank, J. & Agrawal, R.K. A ratchet-like inter-subunit reorganization of the ribosome during translocation. *Nature* **406**, 318-22 (2000).
71. Ban, N., Nissen, P., Hansen, J., Moore, P.B. & Steitz, T.A. The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science* **289**, 905-20 (2000).
72. Wimberly, B.T. et al. Structure of the 30S ribosomal subunit. *Nature* **407**, 327-39 (2000).
73. Schluenzen, F. et al. Structure of Functionally Activated Small Ribosomal Subunit at 3.3 Å Resolution. *Cell* **102**, 615-623 (2000).
74. Doudna, J.A. Ribosomes: Cashing in on crystals. *Current Biology* **9**, R731-R734 (1999).
75. Schuwirth, B.S. et al. Structures of the bacterial ribosome at 3.5 A resolution. *Science* **310**, 827-34 (2005).
76. Korostelev, A., Trakhanov, S., Laurberg, M. & Noller, H.F. Crystal Structure of a 70S Ribosome-tRNA Complex Reveals Functional Interactions and Rearrangements. *Cell* **126**, 1065-1077 (2006).
77. Selmer, M. et al. Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**, 1935-42 (2006).
78. Yusupova, G., Jenner, L., Rees, B., Moras, D. & Yusupov, M. Structural basis for messenger RNA movement on the ribosome. *Nature* **444**, 391-4 (2006).

79. Martin Schmeing, T. & Ramakrishnan, V. What recent ribosome structures have revealed about the mechanism of translation. *Nature* **461**, 1234-42 (2009).
80. Gao, Y.G. et al. The structure of the ribosome with elongation factor G trapped in the posttranslocational state. *Science* **326**, 694-9 (2009).
81. Voorhees, R.M., Schmeing, T.M., Kelley, A.C. & Ramakrishnan, V. The mechanism for activation of GTP hydrolysis on the ribosome. *Science* **330**, 835-838 (2010).
82. Ratje, A.H. et al. Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites. *Nature* **468**, 713-6 (2010).
83. Dunkle, J.A. et al. Structures of the bacterial ribosome in classical and hybrid states of tRNA binding. *Science* **332**, 981-4 (2011).
84. Ben-Shem, A. et al. The structure of the eukaryotic ribosome at 3.0 A resolution. *Science* **334**, 1524-9 (2011).
85. Zhou, J., Lancaster, L., Donohue, J.P. & Noller, H.F. Crystal structures of EF-G-ribosome complexes trapped in intermediate states of translocation. *Science* **340**, 1236086 (2013).
86. Valle, M. et al. Locking and unlocking of ribosomal motions. *Cell* **114**, 123-34 (2003).
87. Diaconu, M. et al. Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation. *Cell* **121**, 991-1004 (2005).
88. Agirrezabala, X. et al. Visualization of the hybrid state of tRNA binding promoted by spontaneous ratcheting of the ribosome. *Mol Cell* **32**, 190-7 (2008).
89. Julian, P. et al. Structure of ratcheted ribosomes with tRNAs in hybrid states. *Proc Natl Acad Sci U S A* **105**, 16924-7 (2008).
90. Villa, E. et al. Ribosome-induced changes in elongation factor Tu conformation control GTP hydrolysis. *Proc Natl Acad Sci U S A* **106**, 1063-8 (2009).
91. Schuette, J.C. et al. GTPase activation of elongation factor EF-Tu by the ribosome during decoding. *EMBO J* **28**, 755-65 (2009).
92. Fischer, N., Konevega, A.L., Wintermeyer, W., Rodnina, M.V. & Stark, H. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature* **466**, 329-33 (2010).
93. Agirrezabala, X. et al. Structural characterization of mRNA-tRNA translocation intermediates. *Proc Natl Acad Sci U S A* **109**, 6094-9 (2012).
94. Anger, A.M. et al. Structures of the human and Drosophila 80S ribosome. *Nature* **497**, 80-5 (2013).
95. Bai, X.C., Fernandez, I.S., McMullan, G. & Scheres, S.H. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *Elife* **2**, e00461 (2013).
96. Liao, M., Cao, E., Julius, D. & Cheng, Y. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* **504**, 107-12 (2013).
97. Amunts, A. et al. Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343**, 1485-1489 (2014).
98. Behrmann, E. et al. Structural snapshots of actively translating human ribosomes. *Cell* **161**, 845-57 (2015).
99. Fischer, N. et al. Structure of the E. coli ribosome-EF-Tu complex at <3 A resolution by Cs-corrected cryo-EM. *Nature* **520**, 567-70 (2015).
100. Loveland, A.B., Demo, G., Grigorieff, N. & Korostelev, A.A. Ensemble cryo-EM elucidates the mechanism of translation fidelity. *Nature* **546**, 113-117 (2017).
101. Brown, A. & Shao, S. Ribosomes and cryo-EM: a duet. *Curr Opin Struct Biol* **52**, 1-7 (2018).
102. Watson, Z.L. et al. Structure of the bacterial ribosome at 2 A resolution. *Elife* **9**(2020).
103. Noeske, J. et al. High-resolution structure of the Escherichia coli ribosome. *Nat Struct Mol Biol* **22**, 336-41 (2015).
104. Polikanov, Y.S., Melnikov, S.V., Soll, D. & Steitz, T.A. Structural insights into the role of rRNA modifications in protein synthesis and ribosome assembly. *Nat Struct Mol Biol* **22**, 342-344 (2015).
105. Kaledhonkar, S. et al. Late steps in bacterial translation initiation visualized using time-resolved cryo-EM. *Nature* **570**, 400-404 (2019).
106. Fu, Z. et al. The structural basis for release-factor activation during translation termination revealed by time-resolved cryogenic electron microscopy. *Nat Commun* **10**, 2579 (2019).

107. Loveland, A.B., Demo, G. & Korostelev, A.A. Cryo-EM of elongating ribosome with EF-Tu*GTP elucidates tRNA proofreading. *Nature* (2020).
108. Carbone, C.E. et al. Time-resolved cryo-EM visualizes ribosomal translocation with EF-G and GTP. *Nature Communications* **12**(2021).
109. Petrychenko, V. et al. Structural mechanism of GTPase-powered ribosome-tRNA movement. *Nat Commun* **12**, 5933 (2021).
110. Rundlet, E.J. et al. Structural basis of early translocation events on the ribosome. *Nature* **595**, 741-745 (2021).
111. Blanchard, S.C., Kim, H.D., Gonzalez, R.L., Jr., Puglisi, J.D. & Chu, S. tRNA dynamics on the ribosome during translation. *Proc Natl Acad Sci U S A* **101**, 12893-8 (2004).
112. Belardinelli, R. et al. Choreography of molecular movements during ribosome progression along mRNA. *Nat Struct Mol Biol* **23**, 342-8 (2016).
113. Rodnina, M.V. Translation in Prokaryotes. *Cold Spring Harb Perspect Biol* **10**(2018).
114. Prabhakar, A., Puglisi, E.V. & Puglisi, J.D. Single-Molecule Fluorescence Applied to Translation. *Cold Spring Harb Perspect Biol* **11**(2019).
115. Rodnina, M.V., Peske, F., Peng, B.Z., Belardinelli, R. & Wintermeyer, W. Converting GTP hydrolysis into motion: versatile translational elongation factor G. *Biol Chem* **401**, 131-142 (2019).
116. Voorhees, R.M. & Ramakrishnan, V. Structural basis of the translational elongation cycle. *Annu Rev Biochem* **82**, 203-36 (2013).
117. Alberts, B. et al. *Molecular Biology of the Cell, 4th edition*, (2002).
118. Sergiev, P.V., Aleksashin, N.A., Chugunova, A.A., Polikanov, Y.S. & Dontsova, O.A. Structural and evolutionary insights into ribosomal RNA methylation. *Nat Chem Biol* **14**, 226-235 (2018).
119. Milo, R. & Phillips, R. *Cell Biology by the Numbers*, (2015).
120. Rodnina, M.V., Fischer, N., Maracci, C. & Stark, H. Ribosome dynamics during decoding. *Philos Trans R Soc Lond B Biol Sci* **372**(2017).
121. Ling, C. & Ermolenko, D.N. Structural insights into ribosome translocation. *Wiley Interdiscip Rev RNA* **7**, 620-36 (2016).
122. Borg, A., Pavlov, M. & Ehrenberg, M. Complete kinetic mechanism for recycling of the bacterial ribosome. *RNA* **22**, 10-21 (2016).
123. Ortiz, J.O., Forster, F., Kurner, J., Linaroudis, A.A. & Baumeister, W. Mapping 70S ribosomes in intact cells by cryoelectron tomography and pattern recognition. *J Struct Biol* **156**, 334-41 (2006).
124. Ortiz, J.O. et al. Structure of hibernating ribosomes studied by cryoelectron tomography in vitro and in situ. *J Cell Biol* **190**, 613-21 (2010).
125. Eisenstein, F., Danev, R. & Pilhofer, M. Improved applicability and robustness of fast cryo-electron tomography data acquisition. *J Struct Biol* **208**, 107-114 (2019).
126. Pfeffer, S. et al. Dissecting the molecular organization of the translocon-associated protein complex. *Nat Commun* **8**, 14516 (2017).
127. Delarue, M. et al. mTORC1 Controls Phase Separation and the Biophysical Properties of the Cytoplasm by Tuning Crowding. *Cell* **174**, 338-349 e20 (2018).
128. Jasnin, M. et al. The Architecture of Traveling Actin Waves Revealed by Cryo-Electron Tomography. *Structure* **27**, 1211-1223 e5 (2019).
129. Erdmann, P.S. et al. In situ cryo-electron tomography reveals gradient organization of ribosome biogenesis in intact nucleoli. *Nat Commun* **12**, 5364 (2021).
130. Rast, A. et al. Biogenic regions of cyanobacterial thylakoids form contact sites with the plasma membrane. *Nat Plants* **5**, 436-446 (2019).
131. Floris, D. & Kuhlbrandt, W. Molecular landscape of etioplast inner membranes in higher plants. *Nat Plants* **7**, 514-523 (2021).
132. Zhang, J. et al. VHUT-cryo-FIB, a method to fabricate frozen hydrated lamellae from tissue specimens for in situ cryo-electron tomography. *J Struct Biol* **213**, 107763 (2021).
133. Schaffer, M. et al. A cryo-FIB lift-out technique enables molecular-resolution cryo-ET within native Caenorhabditis elegans tissue. *Nat Methods* **16**, 757-762 (2019).

134. Brandt, F., Carlson, L.A., Hartl, F.U., Baumeister, W. & Grunewald, K. The three-dimensional organization of polyribosomes in intact human cells. *Mol Cell* **39**, 560-9 (2010).
135. Carter, S.D. et al. Ribosome-associated vesicles: A dynamic subcompartment of the endoplasmic reticulum in secretory cells. *Science Advances* **6**, eaay9572 (2020).
136. Brandt, F. et al. The native 3D organization of bacterial polysomes. *Cell* **136**, 261-71 (2009).
137. Myasnikov, A.G. et al. The molecular structure of the left-handed supra-molecular helix of eukaryotic polyribosomes. *Nat Commun* **5**, 5294 (2014).
138. Afonina, Z.A., Myasnikov, A.G., Shirokov, V.A., Klaholz, B.P. & Spirin, A.S. Conformation transitions of eukaryotic polyribosomes during multi-round translation. *Nucleic Acids Res* **43**, 618-28 (2015).
139. Beck, M. et al. Visual proteomics of the human pathogen Leptospira interrogans. *Nat Methods* **6**, 817-23 (2009).
140. Pfeffer, S., Woellhaf, M.W., Herrmann, J.M. & Forster, F. Organization of the mitochondrial translation machinery studied in situ by cryoelectron tomography. *Nat Commun* **6**, 6019 (2015).
141. Englmeier, R., Pfeffer, S. & Forster, F. Structure of the Human Mitochondrial Ribosome Studied In Situ by Cryoelectron Tomography. *Structure* **25**, 1574-1581 e2 (2017).
142. Gold, V.A., Chroscicki, P., Bragoszewski, P. & Chacinska, A. Visualization of cytosolic ribosomes on the surface of mitochondria by electron cryo-tomography. *EMBO Rep* **18**, 1786-1800 (2017).
143. Waltz, F. et al. How to build a ribosome from RNA fragments in Chlamydomonas mitochondria. *Nat Commun* **12**, 7176 (2021).
144. Simsek, D. et al. The Mammalian Ribo-interactome Reveals Ribosome Functional Diversity and Heterogeneity. *Cell* **169**, 1051-1065 e18 (2017).
145. Genuth, N.R. & Barna, M. Heterogeneity and specialized functions of translation machinery: from genes to organisms. *Nat Rev Genet* **19**, 431-452 (2018).
146. Conn, A.B., Diggs, S., Tam, T.K. & Blaha, G.M. Two Old Dogs, One New Trick: A Review of RNA Polymerase and Ribosome Interactions during Transcription-Translation Coupling. *Int J Mol Sci* **20**(2019).
147. Irastortza-Olaziregi, M. & Amster-Choder, O. Coupled Transcription-Translation in Prokaryotes: An Old Couple With New Surprises. *Front Microbiol* **11**, 624830 (2020).
148. Washburn, R.S. & Gottesman, M.E. Regulation of transcription elongation and termination. *Biomolecules* **5**, 1063-78 (2015).
149. Kang, J.Y., Mishanina, T.V., Landick, R. & Darst, S.A. Mechanisms of Transcriptional Pausing in Bacteria. *J Mol Biol* **431**, 4007-4029 (2019).
150. Larson, M.H. et al. A pause sequence enriched at translation start sites drives transcription dynamics in vivo. *Science* **344**, 1042-1047 (2014).
151. Saba, J. et al. The elemental mechanism of transcriptional pausing. *Elife* **8**(2019).
152. Yanofsky, C. Attenuation in the control of expression of bacterial operons. *Nature* **289**, 751-8 (1981).
153. Richardson, J.P. Preventing the synthesis of unused transcripts by rho factor. *Cell* **64**, 1047-1049 (1991).
154. Byrne, R., Levin, J.G., Bladen, H.A. & Nirenberg, M.W. The in Vitro Formation of a DNA-Ribosome Complex. *Proc Natl Acad Sci U S A* **52**, 140-8 (1964).
155. Miller, O.L., Hamkalo, B.A. & Thomas, C.A. Visualization of Bacterial Genes in Action. *Science* **169**, 392-395 (1970).
156. Kriner, M.A., Sevostyanova, A. & Groisman, E.A. Learning from the Leaders: Gene Regulation by the Transcription Termination Factor Rho. *Trends Biochem Sci* **41**, 690-699 (2016).
157. Proshkin, S., Rahmouni, A.R., Mironov, A. & Nudler, E. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* **328**, 504-8 (2010).
158. Klaholz, B.P. The Ribosome Holds the RNA Polymerase on Track in Bacteria. *Trends Biochem Sci* **42**, 686-689 (2017).
159. Stevenson-Jones, F., Woodgate, J., Castro-Roa, D. & Zenkin, N. Ribosome reactivates transcription by physically pushing RNA polymerase out of transcription arrest. *Proc Natl Acad Sci U S A* (2020).
160. Artsimovitch, I. & Knauer, S.H. Ancient Transcription Factors in the News. *mBio* **10**(2019).

161. Wang, B. & Artsimovitch, I. NusG, an Ancient Yet Rapidly Evolving Transcription Factor. *Front Microbiol* **11**, 619618 (2020).
162. Burmann, B.M. et al. A NusE:NusG complex links transcription and translation. *Science* **328**, 501-4 (2010).
163. Saxena, S. et al. Escherichia colitranscription factor NusG binds to 70S ribosomes. *Molecular Microbiology* **108**, 495-504 (2018).
164. Washburn, R.S. et al. Escherichia coli NusG Links the Lead Ribosome with the Transcription Elongation Complex. *iScience* **23**, 101352 (2020).
165. Wang, C. et al. Structural basis of transcription-translation coupling. *Science*, eabb5317 (2020).
166. Webster, M.W. et al. Structural basis of transcription-translation coupling and collision in bacteria. *Science*, eabb5036 (2020).
167. Sen, R., Chalissery, J. & Muteeb, G. Nus Factors of Escherichia coli. *EcoSal Plus* **3**(2008).
168. Mooney, R.A. et al. Regulator trafficking on bacterial transcription units in vivo. *Mol Cell* **33**, 97-108 (2009).
169. Kohler, R., Mooney, R.A., Mills, D.J., Landick, R. & Cramer, P. Architecture of a transcribing-translating expressome. *Science* **356**, 194-197 (2017).
170. Demo, G. et al. Structure of RNA polymerase bound to ribosomal 30S subunit. *Elife* **6**(2017).
171. Fan, H. et al. Transcription-translation coupling: direct interactions of RNA polymerase with ribosomes and ribosomal subunits. *Nucleic Acids Res* **45**, 11043-11055 (2017).
172. Weixlbaumer, A., Grünberger, F., Werner, F. & Grohmann, D. Coupling of Transcription and Translation in Archaea: Cues From the Bacterial World. *Frontiers in Microbiology* **12**(2021).
173. Chen, M. & Fredrick, K. RNA Polymerase's Relationship with the Ribosome: Not So Physical, Most of the Time. *J Mol Biol* **432**, 3981-3986 (2020).
174. Zhang, Y. et al. DksA guards elongating RNA polymerase against ribosome-stalling-induced arrest. *Mol Cell* **53**, 766-78 (2014).
175. Dai, X. et al. Reduction of translating ribosomes enables Escherichia coli to maintain elongation rates during slow growth. *Nat Microbiol* **2**, 16231 (2016).
176. Iyer, S., Le, D., Park, B.R. & Kim, M. Distinct mechanisms coordinate transcription and translation under carbon and nitrogen starvation in Escherichia coli. *Nat Microbiol* **3**, 741-748 (2018).
177. Zhu, M., Mori, M., Hwa, T. & Dai, X. Disruption of transcription–translation coordination in Escherichia coli leads to premature transcriptional termination. *Nature Microbiology* **4**, 2347-2356 (2019).
178. Oikonomou, C.M. & Jensen, G.J. Cellular Electron Cryotomography: Toward Structural Biology In Situ. *Annu Rev Biochem* **86**, 873-896 (2017).
179. Plitzko, J.M., Schuler, B. & Selenko, P. Structural Biology outside the box-inside the cell. *Curr Opin Struct Biol* **46**, 110-121 (2017).
180. Bauerlein, F.J.B. & Baumeister, W. Towards Visual Proteomics at High Resolution. *J Mol Biol* **433**, 167187 (2021).
181. Hart, R.G. Electron microscopy of unstained biological material: the polytropic montage. *Science* **159**, 1464-7 (1968).
182. De Rosier, D.J. & Klug, A. Reconstruction of three dimensional structures from electron micrographs. *Nature* **217**, 130-4 (1968).
183. Orlova, E.V. & Saibil, H.R. Structural analysis of macromolecular assemblies by electron microscopy. *Chem Rev* **111**, 7710-48 (2011).
184. Galaz-Montoya, J.G. & Ludtke, S.J. The advent of structural biology in situ by single particle cryo-electron tomography. *Biophys Rep* **3**, 17-35 (2017).
185. Glaeser, R.M. Specimen Behavior in the Electron Beam. *Methods Enzymol* **579**, 19-50 (2016).
186. Henderson, R. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q Rev Biophys* **28**, 171-93 (1995).
187. Dubochet, J. & McDowall, A.W. Vitrification of Pure Water for Electron Microscopy. *Journal of Microscopy* **124**, 3-4 (1981).

188. Dubochet, J. et al. Cryo-electron microscopy of vitrified specimens. *Q Rev Biophys* **21**, 129-228 (1988).
189. Dubochet, J. The physics of rapid cooling and its implications for cryoimmobilization of cells. *Methods Cell Biol* **79**, 7-21 (2007).
190. Tivol, W.F., Briegel, A. & Jensen, G.J. An improved cryogen for plunge freezing. *Microsc Microanal* **14**, 375-9 (2008).
191. Mahamid, J. et al. Visualizing the molecular sociology at the HeLa cell nuclear periphery. *Science* **351**, 969-72 (2016).
192. Bauerlein, F.J.B. et al. In Situ Architecture and Cellular Interactions of PolyQ Inclusions. *Cell* **171**, 179-187 e10 (2017).
193. Dubochet, J. High-pressure freezing for cryoelectron microscopy. *Trends in Cell Biology* **5**, 366-368 (1995).
194. Studer, D., Humbel, B.M. & Chiquet, M. Electron microscopy of high pressure frozen samples: bridging the gap between cellular ultrastructure and atomic resolution. *Histochem Cell Biol* **130**, 877-89 (2008).
195. Lucic, V., Rigort, A. & Baumeister, W. Cryo-electron tomography: the challenge of doing structural biology in situ. *J Cell Biol* **202**, 407-19 (2013).
196. Zachs, T. et al. Fully automated, sequential focused ion beam milling for cryo-electron tomography. *Elife* **9**(2020).
197. Tacke, S. et al. A streamlined workflow for automated cryo focused ion beam milling. *J Struct Biol* **213**, 107743 (2021).
198. Klumpe, S. et al. A modular platform for automated cryo-FIB workflows. *Elife* **10**(2021).
199. Mahamid, J. et al. A focused ion beam milling and lift-out approach for site-specific preparation of frozen-hydrated lamellas from multicellular organisms. *J Struct Biol* **192**, 262-9 (2015).
200. Parmenter, C.D. & Nizamudeen, Z.A. Cryo-FIB-lift-out: practically impossible to practical reality. *J Microsc* **281**, 157-174 (2021).
201. Kuba, J. et al. Advanced cryo-tomography workflow developments - correlative microscopy, milling automation and cryo-lift-out. *J Microsc* **281**, 112-124 (2021).
202. Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 A reconstruction of rotavirus VP6. *Elife* **4**, e06980 (2015).
203. Hagen, W.J.H., Wan, W. & Briggs, J.A.G. Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *J Struct Biol* **197**, 191-198 (2017).
204. Turonova, B. et al. Benchmarking tomographic acquisition schemes for high-resolution structural biology. *Nat Commun* **11**, 876 (2020).
205. Li, X. et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods* **10**, 584-90 (2013).
206. Scheres, S.H. Beam-induced motion correction for sub-megadalton cryo-EM particles. *Elife* **3**, e03665 (2014).
207. Danev, R., Buijsse, B., Khoshouei, M., Plitzko, J.M. & Baumeister, W. Volta potential phase plate for in-focus phase contrast transmission electron microscopy. *Proc Natl Acad Sci U S A* **111**, 15635-40 (2014).
208. Fukuda, Y., Laugks, U., Lucic, V., Baumeister, W. & Danev, R. Electron cryotomography of vitrified cells with a Volta phase plate. *J Struct Biol* **190**, 143-54 (2015).
209. Khoshouei, M., Pfeffer, S., Baumeister, W., Forster, F. & Danev, R. Subtomogram analysis using the Volta phase plate. *J Struct Biol* **197**, 94-101 (2017).
210. Schwartz, O. et al. Laser phase plate for transmission electron microscopy. *Nat Methods* **16**, 1016-1020 (2019).
211. Turonova, B., Schur, F.K.M., Wan, W. & Briggs, J.A.G. Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4A. *J Struct Biol* **199**, 187-195 (2017).
212. Mastronarde, D.N. & Held, S.R. Automated tilt series alignment and tomographic reconstruction in IMOD. *J Struct Biol* **197**, 102-113 (2017).

213. Bepler, T., Kelley, K., Noble, A.J. & Berger, B. Topaz-Denoise: general deep denoising models for cryoEM and cryoET. *Nat Commun* **11**, 5208 (2020).
214. Croxford, M. et al. Entropy-regularized deconvolution of cellular cryotransmission electron tomograms. *Proceedings of the National Academy of Sciences* **118**, e2108738118 (2021).
215. Liu, Y.-T. et al. Isotropic Reconstruction of Electron Tomograms with Deep Learning. *BioRxiv* (2021).
216. Castano-Diez, D., Kudryashev, M., Arheit, M. & Stahlberg, H. Dynamo: a flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments. *J Struct Biol* **178**, 139-51 (2012).
217. Scaramuzza, S. & Castano-Diez, D. Step-by-step guide to efficient subtomogram averaging of virus-like particles with Dynamo. *PLoS Biol* **19**, e3001318 (2021).
218. Hrabe, T. et al. PyTom: a python-based toolbox for localization of macromolecules in cryo-electron tomograms and subtomogram analysis. *J Struct Biol* **178**, 177-88 (2012).
219. Asano, S. et al. A molecular census of 26S proteasomes in intact neurons. *Science* **347**, 439-42 (2015).
220. Albert, S. et al. Direct visualization of degradation microcompartments at the ER membrane. *Proc Natl Acad Sci U S A* **117**, 1069-1080 (2020).
221. Che, C. et al. Improved deep learning-based macromolecules structure classification from electron cryo-tomograms. *Mach Vis Appl* **29**, 1227-1236 (2018).
222. Moebel, E. et al. Deep learning improves macromolecule identification in 3D cellular cryo-electron tomograms. *Nature Methods* (2021).
223. Gubins, I. et al. SHREC'19 Track: Classification in Cryo-Electron Tomograms. in *Eurographics Workshop on 3D Object Retrieval* (ed. S. Biasotti, G.L., B. Falcidieno, and I. Pratikakis) (2019).
224. Zeng, X. et al. DISCA: high-throughput cryo-ET structural pattern mining by deep unsupervised clustering. *BioRxiv* (2021).
225. Pyle, E. & Zanetti, G. Current data processing strategies for cryo-electron tomography and subtomogram averaging. *Biochem J* **478**, 1827-1845 (2021).
226. Bharat, T.A.M., Russo, C.J., Lowe, J., Passmore, L.A. & Scheres, S.H.W. Advances in Single-Particle Electron Cryomicroscopy Structure Determination applied to Sub-tomogram Averaging. *Structure* **23**, 1743-1753 (2015).
227. Bharat, T.A. & Scheres, S.H. Resolving macromolecular structures from electron cryo-tomography data using subtomogram averaging in RELION. *Nat Protoc* **11**, 2054-65 (2016).
228. Heumann, J.M., Hoenger, A. & Mastronarde, D.N. Clustering and variance maps for cryo-electron tomography using wedge-masked differences. *J Struct Biol* **175**, 288-99 (2011).
229. Nickell, S. et al. TOM software toolbox: acquisition and analysis for electron tomography. *J Struct Biol* **149**, 227-34 (2005).
230. Forster, F., Pruggnaller, S., Seybert, A. & Frangakis, A.S. Classification of cryo-electron sub-tomograms using constrained correlation. *J Struct Biol* **161**, 276-86 (2008).
231. Chen, M. et al. A complete data processing workflow for cryo-ET and subtomogram averaging. *Nat Methods* **16**, 1161-1168 (2019).
232. Himes, B.A. & Zhang, P. emClarity: software for high-resolution cryo-electron tomography and subtomogram averaging. *Nat Methods* **15**, 955-961 (2018).
233. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat Methods* **16**, 1146-1152 (2019).
234. Tegunov, D., Xue, L., Dienemann, C., Cramer, P. & Mahamid, J. Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 A in cells. *Nat Methods* **18**, 186-193 (2021).
235. Guo, Q. et al. In Situ Structure of Neuronal C9orf72 Poly-GA Aggregates Reveals Proteasome Recruitment. *Cell* **172**, 696-705 e12 (2018).
236. Fassler, F., Dimchev, G., Hodirnau, V.V., Wan, W. & Schur, F.K.M. Cryo-electron tomography structure of Arp2/3 complex in cells reveals new insights into the branch junction. *Nat Commun* **11**, 6437 (2020).

237. Dimchev, G., Amiri, B., Fassler, F., Falcke, M. & Km Schur, F. Computational toolbox for ultrastructural quantitative analysis of filament networks in cryo-ET data. *J Struct Biol*, 107808 (2021).
238. Leitner, A., Faini, M., Stengel, F. & Aebersold, R. Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines. *Trends Biochem Sci* **41**, 20-32 (2016).
239. Lossl, P., van de Waterbeemd, M. & Heck, A.J. The diverse and expanding role of mass spectrometry in structural and molecular biology. *EMBO J* **35**, 2634-2657 (2016).
240. O'Reilly, F.J. & Rappsilber, J. Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology. *Nat Struct Mol Biol* **25**, 1000-1008 (2018).
241. Graziadei, A. & Rappsilber, J. Leveraging crosslinking mass spectrometry in structural and cell biology. *Structure* (2021).
242. Piersimoni, L., Kastritis, P.L., Arlt, C. & Sinz, A. Cross-Linking Mass Spectrometry for Investigating Protein Conformations and Protein-Protein Interactions horizontal line A Method for All Seasons. *Chem Rev* (2021).
243. Kosinski, J. et al. Molecular architecture of the inner ring scaffold of the human nuclear pore complex. *Science* **352**, 363-5 (2016).
244. Kim, S.J. et al. Integrative structure and functional anatomy of a nuclear pore complex. *Nature* **555**, 475-482 (2018).
245. Akey, C.W. et al. Comprehensive structure and functional adaptations of the yeast nuclear pore complex. *Cell* (2021).
246. Dorn, G. et al. Structural modeling of protein-RNA complexes using crosslinking of segmentally isotope-labeled RNA and MS/MS. *Nat Methods* **14**, 487-490 (2017).
247. Belsom, A. & Rappsilber, J. Anatomy of a crosslinker. *Curr Opin Chem Biol* **60**, 39-46 (2021).
248. Gotze, M., Iacobucci, C., Ihling, C.H. & Sinz, A. A Simple Cross-Linking/Mass Spectrometry Workflow for Studying System-wide Protein Interactions. *Anal Chem* **91**, 10236-10244 (2019).
249. Kastritis, P.L. et al. Capturing protein communities by structural proteomics in a thermophilic eukaryote. *Mol Syst Biol* **13**, 936 (2017).
250. Kyrilis, F.L., Meister, A. & Kastritis, P.L. Integrative biology of native cell extracts: a new era for structural characterization of life processes. *Biol Chem* **400**, 831-846 (2019).
251. Lenz, S. et al. Reliable identification of protein-protein interactions by crosslinking mass spectrometry. *Nat Commun* **12**, 3564 (2021).
252. Liu, F., Lossl, P., Rabbitts, B.M., Balaban, R.S. & Heck, A.J.R. The interactome of intact mitochondria by cross-linking mass spectrometry provides evidence for coexisting respiratory supercomplexes. *Mol Cell Proteomics* **17**, 216-232 (2018).
253. Wittig, S. et al. Cross-linking mass spectrometry uncovers protein interactions and functional assemblies in synaptic vesicle membranes. *Nat Commun* **12**, 858 (2021).
254. Chavez, J.D., Schweppe, D.K., Eng, J.K. & Bruce, J.E. In Vivo Conformational Dynamics of Hsp90 and Its Interactors. *Cell Chem Biol* **23**, 716-26 (2016).
255. O'Reilly, F.J. et al. In-cell architecture of an actively transcribing-translating expressome. *Science* **369**, 554-557 (2020).
256. Leung, M.R. et al. In-cell structures of conserved supramolecular protein arrays at the mitochondria–cytoskeleton interface in mammalian sperm. *Proceedings of the National Academy of Sciences* **118**, e2110996118 (2021).
257. Chavez, J.D. et al. Chemical Crosslinking Mass Spectrometry Analysis of Protein Conformations and Supercomplexes in Heart Tissue. *Cell Syst* **6**, 136-141 e5 (2018).
258. Singh, J., Ponnaiyan, S., Gieselmann, V. & Winter, D. Generation of Antibodies Targeting Cleavable Cross-Linkers. *Anal Chem* **93**, 3762-3769 (2021).
259. Fischer, L. & Rappsilber, J. Quirks of Error Estimation in Cross-Linking/Mass Spectrometry. *Anal Chem* **89**, 3829-3833 (2017).
260. Leitner, A. et al. Toward Increased Reliability, Transparency, and Accessibility in Cross-linking Mass Spectrometry. *Structure* **28**, 1259-1268 (2020).

261. Beck, M. et al. Exploring the spatial and temporal organization of a cell's proteome. *J Struct Biol* **173**, 483-96 (2011).
262. Chen, Z.A. & Rappsilber, J. Protein Dynamics in Solution by Quantitative Crosslinking/Mass Spectrometry. *Trends Biochem Sci* **43**, 908-920 (2018).
263. Heck, A.J. Native mass spectrometry: a bridge between interactomics and structural biology. *Nat Methods* **5**, 927-33 (2008).
264. von Kugelgen, A. et al. In Situ Structure of an Intact Lipopolysaccharide-Bound Bacterial Surface Layer. *Cell* **180**, 348-358 e15 (2020).
265. Slavov, N. Unpicking the proteome in single cells. *Science* **367**, 512-513 (2020).
266. Tian, B., Xu, X., Xue, Y., Ji, W. & Xu, T. Cryogenic superresolution correlative light and electron microscopy on the frontier of subcellular imaging. *Biophysical Reviews* **13**, 1163-1171 (2021).
267. Moser, F. et al. Cryo-SOFI enabling low-dose super-resolution correlative light and electron cryo-microscopy. *Proc Natl Acad Sci U S A* **116**, 4804-4809 (2019).
268. Dahlberg, P.D. et al. Cryogenic single-molecule fluorescence annotations for electron tomography reveal in situ organization of key proteins in Caulobacter. *Proc Natl Acad Sci U S A* **117**, 13937-13944 (2020).
269. Hoffman, D.P. et al. Correlative three-dimensional super-resolution and block-face electron microscopy of whole vitreously frozen cells. *Science* **367**(2020).
270. Dahlberg, P.D. & Moerner, W.E. Cryogenic Super-Resolution Fluorescence and Electron Microscopy Correlated at the Nanoscale. *Annu Rev Phys Chem* **72**, 253-278 (2021).
271. Konig, I. et al. Single-molecule spectroscopy of protein conformational dynamics in live eukaryotic cells. *Nat Methods* **12**, 773-9 (2015).
272. Bonucci, A., Ouari, O., Guigliarelli, B., Belle, V. & Mileo, E. In-Cell EPR: Progress towards Structural Studies Inside Cells. *Chembiochem* **21**, 451-460 (2020).
273. Baker, L.A. et al. Combined (1)H-Detected Solid-State NMR Spectroscopy and Electron Cryotomography to Study Membrane Proteins across Resolutions in Native Environments. *Structure* **26**, 161-170 e3 (2018).
274. Malhotra, S., Trager, S., Dal Peraro, M. & Topf, M. Modelling structures in cryo-EM maps. *Curr Opin Struct Biol* **58**, 105-114 (2019).
275. Rout, M.P. & Sali, A. Principles for Integrative Structural Biology Studies. *Cell* **177**, 1384-1403 (2019).
276. Afonine, P.V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74**, 531-544 (2018).
277. Burnley, T., Palmer, C.M. & Winn, M. Recent developments in the CCP-EM software suite. *Acta Crystallogr D Struct Biol* **73**, 469-477 (2017).
278. Russel, D. et al. Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol* **10**, e1001244 (2012).
279. Karaca, E. & Bonvin, A.M. Advances in integrative modeling of biomolecular complexes. *Methods* **59**, 372-81 (2013).
280. Roel-Touris, J., Jimenez-Garcia, B. & Bonvin, A. Integrative modeling of membrane-associated protein assemblies. *Nat Commun* **11**, 6210 (2020).
281. Leman, J.K. et al. Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat Methods* **17**, 665-680 (2020).
282. Braitbard, M., Schneidman-Duhovny, D. & Kalisman, N. Integrative Structure Modeling: Overview and Assessment. *Annu Rev Biochem* **88**, 113-135 (2019).
283. Mirdita, M. et al. ColabFold - Making protein folding accessible to all. *BioRxiv* (2021).
284. Masrati, G. et al. Integrative Structural Biology in the Era of Accurate Structure Prediction. *J Mol Biol* **433**, 167127 (2021).
285. Evans, R. et al. Protein complex prediction with AlphaFold-Multimer. *BioRxiv* (2021).
286. Humphreys, I.R. et al. Computed structures of core eukaryotic protein complexes. *Science* **374**, eabm4805 (2021).
287. Pereira, J. & Schwede, T. Interactomes in the era of deep learning. *Science* **374**, 1319-1320 (2021).

288. Berman, H.M. et al. Federating Structural Models and Data: Outcomes from A Workshop on Archiving Integrative Structures. *Structure* **27**, 1745-1759 (2019).
289. Halbedel, S., Hames, C. & Stulke, J. In vivo activity of enzymatic and regulatory components of the phosphoenolpyruvate:sugar phosphotransferase system in Mycoplasma pneumoniae. *J Bacteriol* **186**, 7936-43 (2004).
290. Schorb, M., Haberbosch, I., Hagen, W.J.H., Schwab, Y. & Mastronarde, D.N. Software tools for automated transmission electron microscopy. *Nat Methods* **16**, 471-477 (2019).
291. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**(2018).
292. Pruesse, E., Peplies, J. & Glockner, F.O. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**, 1823-9 (2012).
293. Rother, M., Rother, K., Puton, T. & Bujnicki, J.M. ModeRNA: a tool for comparative modeling of RNA 3D structure. *Nucleic Acids Res* **39**, 4007-22 (2011).
294. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **46**, W296-W303 (2018).
295. Pettersen, E.F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-12 (2004).
296. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
297. Chen, S. et al. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24-35 (2013).
298. Williams, C.J. et al. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* **27**, 293-315 (2018).
299. Huerta-Cepas, J. et al. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol Biol Evol* **34**, 2115-2122 (2017).
300. Huerta-Cepas, J. et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* **47**, D309-D314 (2019).
301. Katoh, K., Rozewicki, J. & Yamada, K.D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* **20**, 1160-1166 (2019).
302. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* (2021).
303. Miravet-Verde, S., Burgos, R., Delgado, J., Lluch-Senar, M. & Serrano, L. FASTQINS and ANUBIS: two bioinformatic tools to explore facts and artifacts in transposon sequencing and essentiality studies. *Nucleic Acids Res* **48**, e102 (2020).
304. Kidmose, R.T. et al. Namdinator - automatic molecular dynamics flexible fitting of structural models into cryo-EM and crystallography experimental maps. *IUCrJ* **6**, 526-531 (2019).
305. Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S.H. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *Elife* **7**(2018).
306. Naydenova, K. & Russo, C.J. Measuring the effects of particle orientation to improve the efficiency of electron cryomicroscopy. *Nat Commun* **8**, 629 (2017).
307. Bonomi, M. et al. Bayesian Weighing of Electron Cryo-Microscopy Data for Integrative Structural Modeling. *Structure* **27**, 175-188 e6 (2019).
308. Viswanath, S., Chemmama, I.E., Cimermancic, P. & Sali, A. Assessing Exhaustiveness of Stochastic Sampling for Integrative Modeling of Macromolecular Structures. *Biophys J* **113**, 2344-2353 (2017).
309. Shi, Y. et al. Structural characterization by cross-linking reveals the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex. *Mol Cell Proteomics* **13**, 2927-43 (2014).
310. Kosinski, J. et al. Xlink Analyzer: software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *J Struct Biol* **189**, 177-83 (2015).

311. Pettersen, E.F. et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **30**, 70-82 (2021).
312. Catrein, I. & Herrmann, R. The proteome of Mycoplasma pneumoniae, a supposedly "simple" cell. *Proteomics* **11**, 3614-32 (2011).
313. Rosenthal, P.B. & Henderson, R. Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. *Journal of Molecular Biology* **333**, 721-745 (2003).
314. Scheres, S.H. Processing of Structurally Heterogeneous Cryo-EM Data in RELION. *Methods Enzymol* **579**, 125-57 (2016).
315. Schur, F.K. et al. An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. *Science* **353**, 506-8 (2016).
316. Mattei, S., Glass, B., Hagen, W.J., Krausslich, H.G. & Briggs, J.A. The structure and flexibility of conical HIV-1 capsids determined within intact virions. *Science* **354**, 1434-1437 (2016).
317. Nakane, T. et al. Single-particle cryo-EM at atomic resolution. *Nature* **587**, 152-156 (2020).
318. Yip, K.M., Fischer, N., Paknia, E., Chari, A. & Stark, H. Atomic-resolution protein structure determination by cryo-EM. *Nature* **587**, 157-161 (2020).
319. Bulkley, D., Innis, C.A., Blaha, G. & Steitz, T.A. Revisiting the structures of several antibiotics bound to the bacterial ribosome. *Proc Natl Acad Sci U S A* **107**, 17158-63 (2010).
320. Dunkle, J.A., Xiong, L., Mankin, A.S. & Cate, J.H. Structures of the Escherichia coli ribosome with antibiotics bound near the peptidyl transferase center explain spectra of drug action. *Proc Natl Acad Sci U S A* **107**, 17152-7 (2010).
321. Svetlov, M.S. et al. High-resolution crystal structures of ribosome-bound chloramphenicol and erythromycin provide the ultimate basis for their competition. *RNA* **25**, 600-606 (2019).
322. Su, T. et al. Structural basis of l-tryptophan-dependent inhibition of release factor 2 by the TnaC arrest peptide. *Nucleic Acids Res* **49**, 9539-9547 (2021).
323. Vogel, Z., Vogel, T., Zamir, A. & Elson, D. Correlation between the peptidyl transferase activity of the 50 s ribosomal subunit and the ability of the subunit to interact with antibiotics. *Journal of Molecular Biology* **60**, 339-346 (1971).
324. Bayfield, M.A., Dahlberg, A.E., Schulmeister, U., Dorner, S. & Barta, A. A conformational change in the ribosomal peptidyl transferase center upon active/inactive transition. *Proc Natl Acad Sci U S A* **98**, 10096-101 (2001).
325. Timsit, Y., Acosta, Z., Allemand, F., Chiaruttini, C. & Springer, M. The role of disordered ribosomal protein extensions in the early steps of eubacterial 50 S ribosomal subunit assembly. *Int J Mol Sci* **10**, 817-34 (2009).
326. Timsit, Y., Sergeant-Perthuis, G. & Bennequin, D. Evolution of ribosomal protein network architectures. *Sci Rep* **11**, 625 (2021).
327. Klein, D.J., Moore, P.B. & Steitz, T.A. The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *J Mol Biol* **340**, 141-77 (2004).
328. Lawrence, M.G. et al. The extended loops of ribosomal proteins uL4 and uL22 of Escherichia coli contribute to ribosome assembly and protein translation. *Nucleic Acids Res* **44**, 5798-810 (2016).
329. Melnikov, S., Manakongtreecheep, K. & Soll, D. Revising the Structural Diversity of Ribosomal Proteins Across the Three Domains of Life. *Mol Biol Evol* **35**, 1588-1598 (2018).
330. Shasmal, M. & Sengupta, J. Structural diversity in bacterial ribosomes: mycobacterial 70S ribosome structure reveals novel features. *PLoS One* **7**, e31742 (2012).
331. Hentschel, J. et al. The Complete Structure of the Mycobacterium smegmatis 70S Ribosome. *Cell Rep* **20**, 149-160 (2017).
332. Li, Z. et al. Cryo-EM structure of Mycobacterium smegmatis ribosome reveals two unidentified ribosomal proteins close to the functional centers. *Protein Cell* **9**, 384-388 (2018).
333. Jobe, A., Liu, Z., Gutierrez-Vargas, C. & Frank, J. New Insights into Ribosome Structure and Function. *Cold Spring Harb Perspect Biol* **11**(2019).
334. Munro, J.B., Sanbonmatsu, K.Y., Spahn, C.M. & Blanchard, S.C. Navigating the ribosome's metastable energy landscape. *Trends Biochem Sci* **34**, 390-400 (2009).

335. Frank, J. & Gonzalez, R.L., Jr. Structure and dynamics of a processive Brownian motor: the translating ribosome. *Annu Rev Biochem* **79**, 381-412 (2010).
336. Munro, J.B., Altman, R.B., O'Connor, N. & Blanchard, S.C. Identification of two distinct hybrid state intermediates on the ribosome. *Mol Cell* **25**, 505-17 (2007).
337. Lin, J., Gagnon, M.G., Bulkley, D. & Steitz, T.A. Conformational changes of elongation factor G on the ribosome during tRNA translocation. *Cell* **160**, 219-27 (2015).
338. Uemura, S. et al. Real-time tRNA transit on single translating ribosomes at codon resolution. *Nature* **464**, 1012-U73 (2010).
339. Qian, H. & Beard, D.A. Thermodynamics of stoichiometric biochemical networks in living systems far from equilibrium. *Biophys Chem* **114**, 213-20 (2005).
340. Chen, C. et al. Single-molecule fluorescence measurements of ribosomal translocation dynamics. *Mol Cell* **42**, 367-77 (2011).
341. Wilson, D.N. & Nierhaus, K.H. The E-site story: the importance of maintaining two tRNAs on the ribosome during protein synthesis. *Cell Mol Life Sci* **63**, 2725-37 (2006).
342. Bock, L.V. et al. Energy barriers and driving forces in tRNA translocation through the ribosome. *Nat Struct Mol Biol* **20**, 1390-6 (2013).
343. Bakshi, S., Siryaporn, A., Goulian, M. & Weisshaar, J.C. Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells. *Mol Microbiol* **85**, 21-38 (2012).
344. Stülke, J. MycoWiki database (2021).
345. Bakshi, S., Choi, H. & Weisshaar, J.C. The spatial biology of transcription and translation in rapidly growing Escherichia coli. *Front Microbiol* **6**, 636 (2015).
346. Schmidt, A. et al. The quantitative and condition-dependent Escherichia coli proteome. *Nat Biotechnol* **34**, 104-10 (2016).
347. Mustafi, M. & Weisshaar, J.C. Simultaneous Binding of Multiple EF-Tu Copies to Translating Ribosomes in Live Escherichia coli. *mBio* **9**(2018).
348. Zhou, D., Tanzawa, T., Lin, J. & Gagnon, M.G. Structural basis for ribosome recycling by RRF and tRNA. *Nat Struct Mol Biol* **27**, 25-32 (2020).
349. Castellana, M., Hsin-Jung Li, S. & Wingreen, N.S. Spatial organization of bacterial transcription and translation. *Proc Natl Acad Sci U S A* **113**, 9286-91 (2016).
350. Seto, S., Layh-Schmitt, G., Kenri, T. & Miyata, M. Visualization of the attachment organelle and cytadherence proteins of Mycoplasma pneumoniae by immunofluorescence microscopy. *J Bacteriol* **183**, 1621-30 (2001).
351. Noll, H. The discovery of polyribosomes. *Bioessays* **30**, 1220-34 (2008).
352. Qin, D. & Fredrick, K. Analysis of polysomes from bacteria. *Methods Enzymol* **530**, 159-72 (2013).
353. Morisaki, T. et al. Real-time quantification of single RNA translation dynamics in living cells. *Science* **352**, 1425-1429 (2016).
354. Lieberman, K.R. et al. The 23 S rRNA environment of ribosomal protein L9 in the 50 S ribosomal subunit. *J Mol Biol* **297**, 1129-43 (2000).
355. Herr, A.J., Nelson, C.C., Wills, N.M., Gesteland, R.F. & Atkins, J.F. Analysis of the roles of tRNA structure, ribosomal protein L9, and the bacteriophage T4 gene 60 bypassing signals during ribosome slippage on mRNA. *J Mol Biol* **309**, 1029-48 (2001).
356. Hoffman, D.W. et al. Crystal structure of prokaryotic ribosomal protein L9: a bi-lobed RNA-binding protein. *The EMBO Journal* **13**, 205-212 (1994).
357. Dunkle, J.A. & Cate, J.H.D. The packing of ribosomes in crystals and polysomes. in *Ribosomes* 65-73 (2011).
358. Selmer, M., Gao, Y.G., Weixlbaumer, A. & Ramakrishnan, V. Ribosome engineering to promote new crystal forms. *Acta Crystallogr D Biol Crystallogr* **68**, 578-83 (2012).
359. Smith, A.M., Costello, M.S., Kettring, A.H., Wingo, R.J. & Moore, S.D. Ribosome collisions alter frameshifting at translational reprogramming motifs in bacterial mRNAs. *Proc Natl Acad Sci U S A* **116**, 21769-21779 (2019).
360. Arenz, S. & Wilson, D.N. Bacterial Protein Synthesis as a Target for Antibiotic Inhibition. *Cold Spring Harb Perspect Med* **6**(2016).

361. Lin, J., Zhou, D., Steitz, T.A., Polikanov, Y.S. & Gagnon, M.G. Ribosome-Targeting Antibiotics: Modes of Action, Mechanisms of Resistance, and Implications for Drug Design. *Annu Rev Biochem* **87**, 451-478 (2018).
362. Marks, J. et al. Context-specific inhibition of translation by ribosomal antibiotics targeting the peptidyl transferase center. *Proc Natl Acad Sci U S A* **113**, 12150-12155 (2016).
363. Vazquez-Laslop, N. & Mankin, A.S. Context-Specific Action of Ribosomal Antibiotics. *Annu Rev Microbiol* **72**, 185-207 (2018).
364. Baquero, F. & Levin, B.R. Proximate and ultimate causes of the bactericidal action of antibiotics. *Nat Rev Microbiol* (2020).
365. Borovinskaya, M.A., Shoji, S., Holton, J.M., Fredrick, K. & Cate, J.H.D. A steric block in translation caused by the antibiotic spectinomycin. *ACS Chem Biol* **2**, 545-552 (2007).
366. Mohan, S., Donohue, J.P. & Noller, H.F. Molecular mechanics of 30S subunit head rotation. *Proc Natl Acad Sci U S A* **111**, 13325-30 (2014).
367. Maffioli, S.I. et al. Antibacterial Nucleoside-Analog Inhibitor of Bacterial RNA Polymerase. *Cell* **169**, 1240-1248 e23 (2017).
368. Bilgin, N., Richter, A.A., Ehrenberg, M., Dahlberg, A.E. & Kurland, C.G. Ribosomal RNA and protein mutants resistant to spectinomycin. *The EMBO Journal* **9**, 735-739 (1990).
369. Pathak, B.K., Mondal, S. & Barat, C. Inhibition of Escherichia coli ribosome subunit dissociation by chloramphenicol and Blasticidin: a new mode of action of the antibiotics. *Lett Appl Microbiol* **64**, 79-85 (2017).
370. Wasserman, M.R., Alejo, J.L., Altman, R.B. & Blanchard, S.C. Multiperspective smFRET reveals rate-determining late intermediates of ribosomal translocation. *Nat Struct Mol Biol* **23**, 333-41 (2016).
371. Ilina, E.N. et al. Mutation in ribosomal protein S5 leads to spectinomycin resistance in Neisseria gonorrhoeae. *Front Microbiol* **4**, 186 (2013).
372. Kamath, D., Gregory, S.T. & O'Connor, M. The Loop 2 Region of Ribosomal Protein uS5 Influences Spectinomycin Sensitivity, Translational Fidelity, and Ribosome Biogenesis. *Antimicrob Agents Chemother* **61**(2017).
373. Weber, M.J. & DeMoss, J.A. The inhibition by chloramphenicol of nascent protein formation in E. coli. *Proc Natl Acad Sci U S A* **55**, 1224-30 (1966).
374. Ermolenko, D.N., Cornish, P.V., Ha, T. & Noller, H.F. Antibiotics that bind to the A site of the large ribosomal subunit can induce mRNA translocation. *RNA* **19**, 158-66 (2013).
375. Volkov, I.L., Seefeldt, A.C. & Johansson, M. Tracking of single tRNAs for translation kinetics measurements in chloramphenicol treated bacteria. *Methods* **162-163**, 23-30 (2019).
376. Hui, M.P., Foley, P.L. & Belasco, J.G. Messenger RNA degradation in bacterial cells. *Annu Rev Genet* **48**, 537-59 (2014).
377. Muller, C., Crowe-McAuliffe, C. & Wilson, D.N. Ribosome Rescue Pathways in Bacteria. *Front Microbiol* **12**, 652980 (2021).
378. Aguirre Rivera, J. et al. Real-time measurements of aminoglycoside effects on protein synthesis in live cells. *Proceedings of the National Academy of Sciences* **118**, e2013315118 (2021).
379. Said, N. et al. Structural basis for lambdaN-dependent processive transcription antitermination. *Nat Microbiol* **2**, 17062 (2017).
380. Guo, X. et al. Structural Basis for NusA Stabilized Transcriptional Pausing. *Mol Cell* **69**, 816-827 e4 (2018).
381. Krupp, F. et al. Structural Basis for the Action of an All-Purpose Transcription Anti-termination Factor. *Mol Cell* **74**, 143-157 e5 (2019).
382. Yang, X. et al. The structure of bacterial RNA polymerase in complex with the essential transcription elongation factor NusA. *EMBO Rep* **10**, 997-1002 (2009).
383. Ma, C. et al. RNA polymerase-induced remodelling of NusA produces a pause enhancement complex. *Nucleic Acids Res* **43**, 2829-40 (2015).
384. Murakami, K.S. Structural biology of bacterial RNA polymerase. *Biomolecules* **5**, 848-64 (2015).
385. Weiss, A. & Shaw, L.N. Small things considered: the small accessory subunits of RNA polymerase in Gram-positive bacteria. *FEMS Microbiol Rev* **39**, 541-54 (2015).

386. Turtola, M. & Belogurov, G.A. NusG inhibits RNA polymerase backtracking by stabilizing the minimal transcription bubble. *Elife* **5**(2016).
387. Kang, J.Y. et al. Structural Basis for Transcript Elongation Control by NusG Family Universal Regulators. *Cell* **173**, 1650-1662 e14 (2018).
388. Lewis, P.J., Doherty, G.P. & Clarke, J. Transcription factor dynamics. *Microbiology* **154**, 1837-44 (2008).
389. Erdos, G. & Dosztanyi, Z. Analyzing Protein Disorder with IUPred2A. *Curr Protoc Bioinformatics* **70**, e99 (2020).
390. Letunic, I. & Bork, P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res* **46**, D493-D496 (2018).
391. Yang, J. & Zhang, Y. Protein Structure and Function Prediction Using I-TASSER. *Curr Protoc Bioinformatics* **52**, 5 8 1-5 8 15 (2015).
392. Beuth, B., Pennell, S., Arnvig, K.B., Martin, S.R. & Taylor, I.A. Structure of a Mycobacterium tuberculosis NusA-RNA complex. *EMBO J* **24**, 3576-87 (2005).
393. Mah, T.F., Kuznedelov, K., Mushegian, A., Severinov, K. & Greenblatt, J. The alpha subunit of E. coli RNA polymerase activates RNA binding by NusA. *Genes Dev* **14**, 2664-75 (2000).
394. Schweimer, K. et al. NusA interaction with the alpha subunit of E. coli RNA polymerase is via the UP element site and releases autoinhibition. *Structure* **19**, 945-54 (2011).
395. de Jong, L. et al. In-Culture Cross-Linking of Bacterial Cells Reveals Large-Scale Dynamic Protein-Protein Interactions at the Peptide Level. *J Proteome Res* **16**, 2457-2471 (2017).
396. Pei, H.H. et al. The delta subunit and NTPase HelD institute a two-pronged mechanism for RNA polymerase recycling. *Nat Commun* **11**, 6418 (2020).
397. Kuban, V. et al. Quantitative Conformational Analysis of Functionally Important Electrostatic Interactions in the Intrinsically Disordered Region of Delta Subunit of Bacterial RNA Polymerase. *J Am Chem Soc* **141**, 16817-16828 (2019).
398. Weixlbaumer, A., Leon, K., Landick, R. & Darst, S.A. Structural basis of transcriptional pausing in bacteria. *Cell* **152**, 431-41 (2013).
399. Strauss, M. et al. Transcription is regulated by NusA:NusG interaction. *Nucleic Acids Res* **44**, 5971-82 (2016).
400. Zhang, J. & Landick, R. A Two-Way Street: Regulatory Interplay between RNA Polymerase and Nascent RNA Structure. *Trends Biochem Sci* **41**, 293-310 (2016).
401. Sutton, G. et al. Assembly intermediates of orthoreovirus captured in the cell. *Nat Commun* **11**, 4445 (2020).
402. Sanchez, R.M., Zhang, Y., Chen, W., Dietrich, L. & Kudryashev, M. Subnanometer-resolution structure determination in situ by hybrid subtomogram averaging - single particle cryo-EM. *Nat Commun* **11**, 3709 (2020).
403. Mendonça, L. et al. CryoET structures of immature HIV Gag reveal six-helix bundle. *Communications Biology* **4**(2021).
404. Ni, T. et al. Structure of native HIV-1 cores and their interactions with IP6 and CypA. *Sci Adv* **7**, eabj5715 (2021).
405. Wang, Z. et al. Structure of the thin filament in native skeletal muscles reveals its interaction with nebulin and two distinct conformations of myosin. *BioRxiv* (2021).
406. Mosalaganti, S. et al. Artificial intelligence reveals nuclear pore complexity. *BioRxiv* (2021).
407. Terwilliger, T.C. et al. Improving AlphaFold modeling using implicit information from experimental density maps. (2022).
408. Parajuli, N.P., Mandava, C.S., Pavlov, M.Y. & Sanyal, S. Mechanistic insights into translation inhibition by aminoglycoside antibiotic arbekacin. *Nucleic Acids Res* **49**, 6880-6892 (2021).
409. Wohlgemuth, I. et al. Translation error clusters induced by aminoglycoside antibiotics. *Nat Commun* **12**, 1830 (2021).
410. Zimmermann, M., Patil, K.R., Typas, A. & Maier, L. Towards a mechanistic understanding of reciprocal drug-microbiome interactions. *Mol Syst Biol* **17**, e10116 (2021).

411. Bock, L.V. & Grubmuller, H. Effects of cryo-EM cooling on structural ensembles. *European Biophysics Journal with Biophysics Letters* **50**, 131-131 (2021).
412. Dashti, A. et al. Retrieving functional pathways of biomolecules from single-particle snapshots. *Nat Commun* **11**, 4734 (2020).
413. Punjani, A. & Fleet, D.J. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *J Struct Biol* **213**, 107702 (2021).
414. Zhong, E.D., Bepler, T., Berger, B. & Davis, J.H. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat Methods* **18**, 176-185 (2021).
415. Chen, M. & Ludtke, S.J. Deep learning-based mixed-dimensional Gaussian mixture model for characterizing variability in cryo-EM. *Nat Methods* **18**, 930-936 (2021).
416. Johnson, G.E., Lalanne, J.B., Peters, M.L. & Li, G.W. Functionally uncoupled transcription-translation in Bacillus subtilis. *Nature* **585**, 124-128 (2020).