Technische Universität München
TUM School of Natural Sciences

Visualizing chaperonin function in protein folding by cryo-electron tomography

Jonathan Wagner

Vollständiger Abdruck der von der TUM School of Natural Sciences der Technischen Universität München zur Erlangung eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigten Dissertation.

Vorsitz: Prof. Dr. Johannes Buchner

Prüfer der Dissertation:

1. Hon.- Prof. Dr. Wolfgang Baumeister
2. Hon.- Prof. Dr. Franz-Ulrich Hartl
3. Prof. Dr. Rubén Fernández-Busnadiego

Die Dissertation wurde am 29.01.2024 bei der Technischen Universität München eingereicht und durch die TUM School of Natural Sciences am 28.03.2024 angenommen.

Abstract

Chaperonins are essential for protein folding across all domains of life. The bacterial chaperonin, the ATPase GroEL, mediates the folding of ~10% of newly-synthesized proteins in *Escherichia Coli* (*E. coli*). GroEL is composed of two homo heptameric rings, stacked back-to-back, which orchestrate protein folding in an enclosed, central, shielded chamber, in cooperation with its co-factor GroES.

Although *in vitro* analyses have provided considerable insights, there has been persistent debate about the precise nature of the chaperonin reaction cycle and the functional stoichiometry during protein folding. This thesis addresses these controversies through cutting-edge *in situ* cryo-electron tomography (cryo-ET) to visualize chaperonin function within the native cellular environment.

Our *in situ* observations reveal that GroEL predominantly binds GroES in an asymmetrical fashion (60-70% of cases), forming 'bullet' shaped complexes ($EL:ES_1$), with a smaller proportion existing as symmetrical 'football' shaped complexes ($EL:ES_2$), where the GroEL 14-mer is bound to two GroES 7-mers simultaneously. Further, by correlating *in situ* data with high-resolution *in vitro* structures, we observed a sizable proportion of folded substrate proteins within the GroEL/ES chambers, awaiting release into the cytosol.

Furthermore, we identified and described the *in situ* organisation of bacterial polysomes, which are hubs for protein translation. We used this *in situ* data to probe for a linkage of protein translation on the bacterial polysome complex with protein folding by the GroEL/ES system. Our data indicates that, potentially the buffering of unfolded proteins by co-translational chaperons as well as the DnaJ/K system alleviate the need for spatial orchestration of these two processes.

These insights provide a comprehensive picture of the *in vivo* GroEL/ES cycle, consisting of linked asymmetrical and symmetrical sub-reactions that facilitate protein folding. By situating the chaperonin action within its native cellular context, our research resolves longstanding disputes about the GroEL mechanism, shedding light on the fundamental processes of protein folding within the complex milieu of the cell. This work not only elucidates an essential biological process but also bridges the gap between *in vitro* studies and *in vivo* cellular biology, offering a deeper understanding of the chaperonin reaction cycle and stoichiometry.

Zusammenfassung

Chaperonine sind für die Proteinfaltung in allen Lebensbereichen unerlässlich. Das bakterielle Model-Chaperonin, die ATPase-GroEL, vermittelt das Falten von etwa 10% der neu synthetisierten Proteine in *Escherichia Coli*. GroEL besteht aus zwei homo-heptameren Ringen, die Rücken an Rücken gestapelt sind, und die in Zusammenarbeit mit dem Kofaktor GroES das Protein-Falten in einer abgeschlossenen und geschützten zentralen Kammer koordinieren.

Obwohl *in vitro* Analysen erhebliche Einblicke geboten haben, besteht eine anhaltende Debatte über die genaue Natur des Chaperonin-Reaktionszyklus und die funktionale Stöchiometrie des Komplexes während des Protein-Faltens. Diese Dissertation adressiert diese Kontroversen durch den Einsatz von *in situ* Kryo-Elektronentomographie (cryo-ET), um die Funktion von Chaperoninen in der nativen zellulären Umgebung zu visualisieren.

Unsere in situ Beobachtungen zeigen, dass GroEL vorwiegend auf asymmetrische Weise (in 60-70% der Fälle) GroES bindet und 'Bullet'-Komplexe bildet, während ein kleinerer Anteil als symmetrische 'Football'-Komplexe existiert. Weiterhin haben wir durch die Verbindung von *in situ* Daten mit hochauflösenden *in vitro* Strukturen gefaltete Substratproteine innerhalb der GroEL-ES-Kammern nachgewiesen, die auf ihre Freisetzung ins Zytosol warten.

Diese Erkenntnisse liefern ein umfassendes Bild des *in vivo* GroEL/ES-Zyklusmodells, das aus gekoppelten asymmetrischen und symmetrischen Teilreaktionen besteht, die das Protein-Falten erleichtern. Indem die Aktion der Chaperonine in ihren natürlichen zellulären Kontext gesetzt wird, löst unsere Forschung langjährige Streitigkeiten über den GroEL-Mechanismus und beleuchtet die grundlegenden Prozesse des Protein-Faltens im komplexen Milieu der Zelle. Diese Arbeit veranschaulicht nicht nur einen wesentlichen biologischen Prozess, sondern überbrückt auch die Lücke zwischen *in vitro* Studien und der zellulären Realität und bietet ein definitives Verständnis des Chaperonin-Reaktionszyklus und deren Stöchiometrie.

Darüber hinaus haben wir die *in situ* Organisation des bakteriellen Polysoms, einem Zentrum für die Proteintranslation, identifiziert und beschrieben. Wir nutzten diese umfassenden *in situ* Daten, um eine Verbindung zwischen der Proteintranslation am bakteriellen Polysomenkomplex und der Proteinfaltung durch das GroEL/ES-System zu untersuchen, wobei gezeigt wurde, dass das Abpuffern von ungefalteten Proteinen

durch kotranslationale Chaperone sowie das bakterielle DanJ/K-System den Bedarf an einer besonderen Orchestrierung dieser beiden Prozesse stark verringern.

Acknowledgments

As with almost everything, this thesis did not come to be in a vacuum. And even though we heavily rely on computers and machines to accomplish what we do, we even more so rely on the goodwill and support of the people around us! They are the ones that keep us going, provide us with inspiration, and the energy to keep going even when things are not working out as expected. Therefore, I would like to thank a lot of people who made this piece of work possible.

First of all, I would like to thank my academic mentors who supported me throughout my journey. Prof. Dr. Wolfgang Baumeister and Prof. Dr. Ulrich Hartl who fostered an environment of collaboration and exploration and gave me the freedom to truly make this project mine and feel ownership and pride in the work I did while still somehow managing to keep me on the right path. It was truly inspiring writing a manuscript with them and getting to learn how to translate results into a great story. I further have to thank Prof. Dr. Rubén Fernández-Busnadiego who not only inspired me to first start my journey into structural biology but also mentored me throughout these last 8 years and first convinced me to start this truly fascinating project.

As mentioned before, this work would not have been possible without the help of many others. You are what made me succeed and why my time at the Max Planck Institute for Biochemistry felt truly special and fun! Although not an exhaustive list, I would like to point out the contributions of a couple of individuals without whom I would have been truly lost in the process, in no particular order. Dr. Andreas Bracher, who helped me with model building, writing and was always happy to discuss my project and my problems. Florian Beck, who showed me first how to really use a computer and helped me to learn how to make sense of the amazing amounts of data that we produce. Dr. Roman Körner, who first showed me how a mass spectrometer works and greatly helped me with this part of my project. Alonso Izzat Carvajal Alvarez for close collaboration and shared projects. Dr. Stefan Bohn, who convinced me that single particle analysis was a really good idea and got me started on that part of the journey. Prof. Dr. William Wan, who really helped me understand how subtomogram averaging was supposed to work and how to be a better scientist overall. I also would like to acknowledge Prof. Dr. Jürgen Plitzko, who worked tirelessly in the background to keep all our amazing machines running and shape the culture I so enjoyed. I truly appreciated all the freely given time.

I also want to acknowledge my present and former colleagues that made it fun to come into the office every day and made the time fly by. Without all of you, this could have been far from fun, and I will miss our discussions over coffee or with a cup of tea.
It would be negligent to not mention my parents, Reiner and Margareth, at this point, who first instilled a sense of wonder and exploration in me as well as provided me with all the support along the way that one could wish for. I also want to thank my brothers, Joshua and Cecil, who not only emotionally supported me throughout this journey but also were always happy to help me along my way even at odd hours on weekends. Without all your love, I wouldn't be where I am right now.
And now, to the one person who got to see all the ups and downs of the last years and was there to celebrate the little wins along the way and pick me up after all the failures. Without your love and support, Julia, this would not have been possible. Thank you for always being there for me and going on this journey with me.
To all of you and all the others who have been part of this endeavor, I am eternally grateful.

Jonathan Wagner

Publications

Wan, W., Khavnekar, S., **Wagner, J**., Erdmann, P. and Baumeister, W., 2020. STOPGAP: a software package for subtomogram averaging and refinement. *Microscopy and Microanalysis*, *26*(S2), pp.2516-2516.

Jiang, W., **Wagner, J**., Du, W., Plitzko, J., Baumeister, W., Beck, F. and Guo, Q., 2022. A transformation clustering algorithm and its application in polyribosomes structural profiling. *Nucleic Acids Research*, *50*(16), pp.9001-9011.

Dietrich, H.M., Righetto, R.D., Kumar, A., Wietrzynski, W., Trischler, R., Schuller, S.K., **Wagner, J**., Schwarz, F.M., Engel, B.D., Müller, V. and Schuller, J.M., 2022. Membrane-anchored HDCR nanowires drive hydrogen-powered CO2 fixation. *Nature*, *607*(7920), pp.823-830.

Wagner, J., Carvajal, A.I., Bracher, A., Beck, F., Wan, W., Bohn, S, Körner, R., Baumeister, W., Fernandez-Busnadiego, R., Hartl, F.U., 2023 Visualizing chaperonin-mediated protein folding *in situ* by cryoelectron tomography. *In preparation*

1 Contents

1. Introduction into protein folding and molecular chaperons

Proteins are essential functional units of life. Most of the principal tasks in a cell are carried out by proteins: gene regulation by transcription factors and other regulatory proteins, most of catalysis by enzymes, the shaping of the cells by structural proteins like actin or tubulin and intracellular transport along the cytoskeleton by molecular motors. Proteins are also involved in communication with the environment through sensor proteins, changes in cell shape and cell motility. Moreover, the defense against viral and bacterial infections as part of the immune system are enabled by proteins and protein complexes.

Just as a key must be precisely shaped to unlock a door, proteins must fold into a specific structure to perform their functions. This transformation into a unique three-dimensional shape, known as protein folding, is vital for proteins to carry out their function within the cell, making it a central focus of biological and biotechnological research (Anfinsen 1972).

1.1 Protein Structure

Proteins are linear polymers of amino acids linked by peptide bonds. Amino acids are α-amino carbonic acids. All atoms beyond the C-α atom are called side chains. These side chains can be grouped into charged, polar and hydrophobic side chains according to their chemical properties. The sequence of amino acid is encoded by the DNA in an organism’s genome, the so-called primary structure of a protein. The sequence of amino acids, the chemical properties and spatial demands of the side chains defines the overall conformation (“structure”) of the protein.

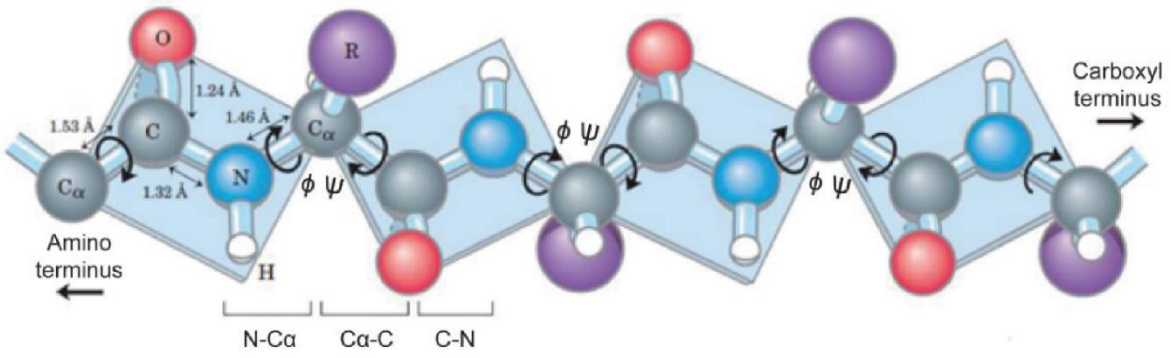

Figure 1: Protein backbone and its torsion angles.

The diagram presents the structure of a protein backbone, highlighting the torsion angles between peptide bond planes. Spheres designate atoms, sticks chemical bonds. Carbon atoms are depicted in gray, nitrogen in blue and the first atom of the side chain is shown in purple, labeled with an R. Notably, every C-N peptide bond is flat and lacks rotational capability. However, the bonds between N-Cα and Cα-C exhibit rotational freedom and are represented with torsion angles ϕ and ψ, respectively. Adapted from (Nelson and Cox 2008)

The process of information transfer begins in the cell nucleus, where DNA- the genetic code, is transcribed into messenger RNA (mRNA). This mRNA is then decoded by the ribosomes, the cell's protein factories, located in the cytoplasm. Here, the mRNA sequence is translated into a chain of amino acids, a process known as translation. The ribosome catalyzes the formation of peptide bonds between the amino acids, linking them together in a growing polypeptide chain.

Due to its partial (~40%) double-bond character, the peptide bond cannot easily rotate, locking the Cα-C-N-Cα atom repeats into a planar structure. The flexibility of the peptide backbone stems from the rotations around a single bond of N-Cα (angle ϕ) and Cα-C (angle Ψ) (Fig. 1). G. N. Ramachandran recognized that many combinations of φ and ψ angles are forbidden because of steric collisions between atoms. The allowed values can be visualized on a two-dimensional plot called a Ramachandran diagram (Fig. 2). Considering the energy contained in various pairs of ψ and φ angles, Ramachandran and his colleagues found three especially stable pairs, the so-called α and β conformations or a 3-10 helix. Successions of amino acids in α- and β-conformation give rise to α-helixes and β-strands respectively. Such stretches are called secondary structural elements. These two pairs of angles are found to almost exclusively occur in folded proteins. (Ramachandran, Ramakrishnan et al. 1963)

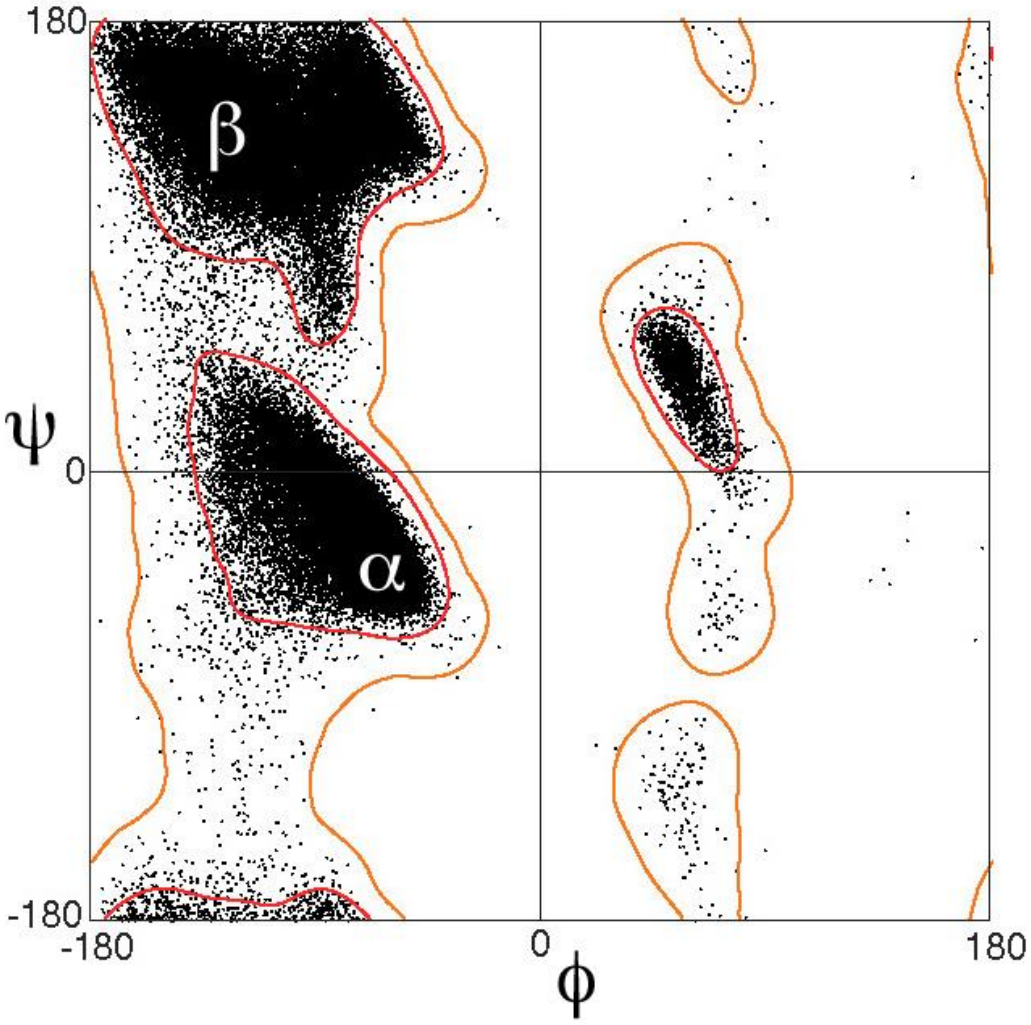

Figure 2: Ramachandran plot.

φ,ψ plot ("Ramachandran diagram") of the torsion angles that specify protein backbone conformation, with the alpha-helix and beta-strand regions labelled. The data illustrated is drawn from the study(Lovell, Davis et al. 2003) and showcases roughly 100,000 data points from high-resolution crystal structures of general amino acid types, excluding Glycine, Proline, and pre-Proline. Figure is reproduced form: (Richardson. 2023)

The amide groups and carbonyl oxygens involved in the peptide bonds are strongly polar and show a strong tendency to form hydrogen bonds with other polar groups. In secondary structure elements, the amide hydrogen bonds are intramolecular. Under physiological conditions, the polypeptide chain forms as many of these high-energy bonds as possible. Meanwhile, the polypeptide backbone and side chains must adopt a conformation to minimize their steric strain. As a result, two types of secondary structure are commonly found in proteins: the α-helix and the β-sheet. Concurrently, the hydrophobic core effect significantly influences protein folding by driving hydrophobic amino acids towards the interior of the protein structure, minimizing their

exposure to the aqueous environment. This behavior fosters a stable, functional structure, with hydrophilic amino acids primarily located on the protein's surface, facilitating the formation of a distinct hydrophobic core that not only stabilizes the protein but also critically impacts its functionality within the cellular milieu.

After establishing the secondary structures, the protein can further fold into its tertiary structure. This involves the spatial arrangement of α-helices, β-sheets and random coils, orchestrated by interactions between distant amino acid side chains. Hydrophobic interactions, hydrogen bonds, ionic interactions, and disulfide bonds play pivotal roles in this process.

Many proteins further assemble into homo- or hetero-oligomeric complexes, forming the quaternary structure. The forces keeping these subunits together are similar to those stabilizing the tertiary structure. Large proteins for more complicated functions are synthesized by assembling these multiple-subunit complexes.

1.2 Protein folding

Protein folding, the biophysical process by which a polypeptide chain transitions from a random coil into its native, three-dimensional conformation, is a central question in molecular biochemistry. The seminal experiments conducted by Christian Anfinsen on ribonuclease A *in vitro* demonstrated that in this no extrinsic factors were needed to quickly assume the enzymatically active structure after unfolding, i.e. the protein sequence encoded the tertiary structure. This implies that the native conformation of a protein is thermodynamically stable state, the global minimum in the free energy landscape of all kinetically accessible structures (Anfinsen 1973).

1.2.1 The protein folding paradox

Contrary to a stochastic process, protein folding follows specific pathways, a concept underscored by the Levinthal paradox. This paradox posits that a protein comprising 100 amino acid residues could theoretically adopt 10^{100} distinct conformations. If protein folding were a random process and most structures isoenergetic, the timescale required to explore the entire conformational space would far exceed any biological timescale. Consequently, proteins must navigate an energy gradient or a specific folding pathway to achieve their native conformation (Fig 3).

However, the trajectory to the native state is often convoluted. The free energy landscape of protein folding is typically rugged, indicating that polypeptides must overcome substantial kinetic barriers to attain their native states. Transiently populated partially folded metastable intermediates may manifest as kinetically trapped species.

These can either be disorganized globules, which maintain large configurational entropy but lack specific native interactions, or intermediates stabilized by non-native contacts. Because in natively folded proteins hydrophobic groups are mainly buried in the hydrophobic core, such folding intermediates and misfolded species are thought to expose hydrophobic groups, which are then available for intermolecular interactions, leading to aggregation (Fig. 3).

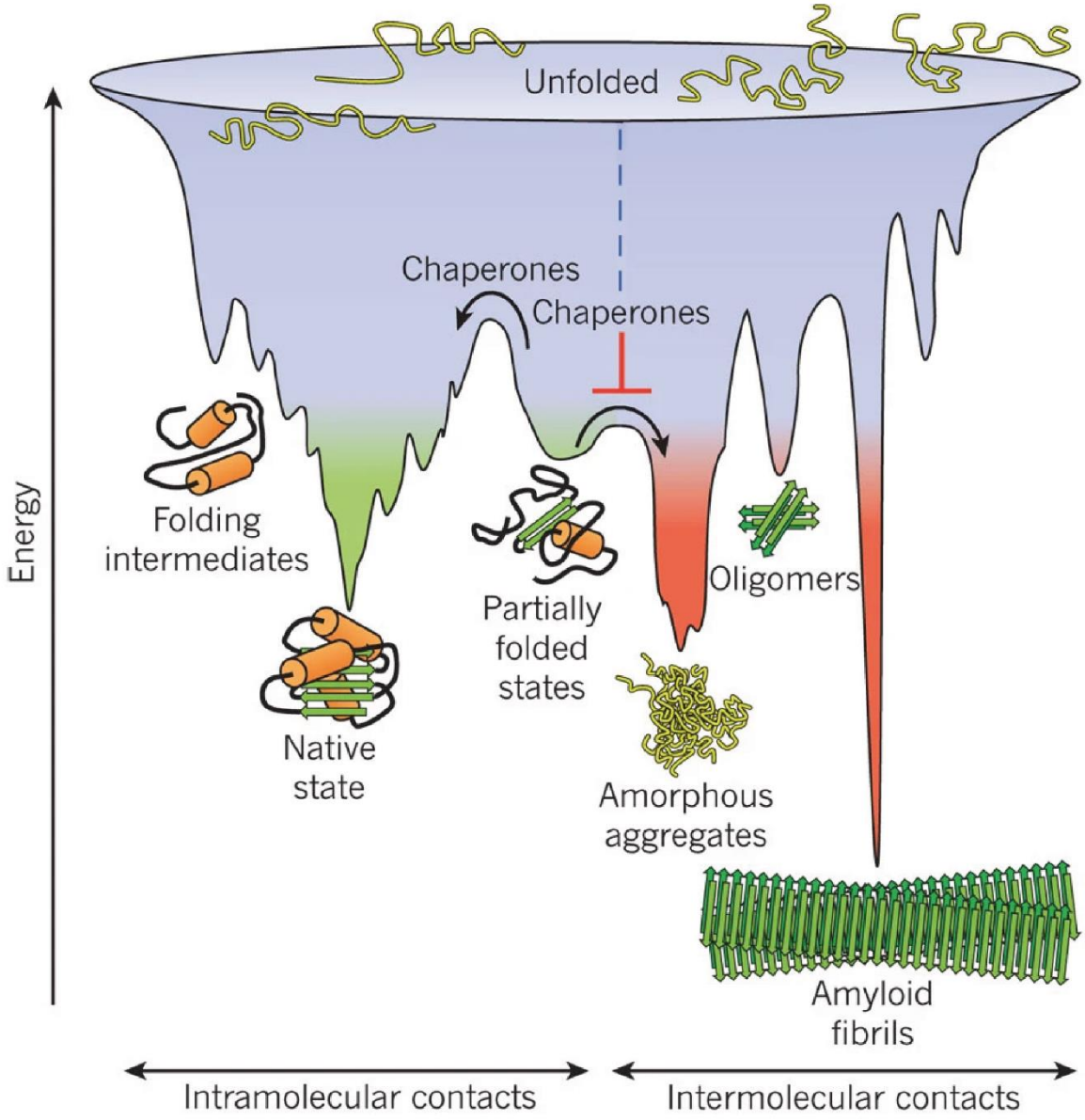

Figure 3: Energy landscape of protein folding and aggregation

This depicts the funnel-like free-energy surface that proteins traverse as they progress towards their native configuration (highlighted in green) through the establishment of internal molecular interactions. The free-energy landscape's irregularity can lead to the buildup of conformations that are kinetically stuck. To proceed towards the native state, energy barriers must be overcome. Within cellular environments, chaperones may expedite these steps. When multiple proteins fold concurrently in a shared compartment, their individual folding energy landscapes might intersect with aggregation landscapes, leading to the emergence of unstructured aggregates, detrimental oligomers, or structured amyloid fibrils (denoted in

red). Fibril aggregation typically follows a nucleation-dependent polymerization pathway. This can originate either from intermediates during the primary folding process or from the destabilization of the already established native structure, leading to semi-folded conformations. Such aggregation is generally prevented by molecular chaperones. (Hartl, Bracher et al. 2011)

1.2.2 Protein aggregation

These intermediates can aggregate in a concentration-dependent manner, predominantly forming amorphous structures. However, a subset of proteins can aggregate into amyloid fibrils, characterized by cross a β-structure that aggregates into a β-strand fibril. These fibrils, which are thermodynamically highly stable, are associated with numerous pathologies, particularly neurodegenerative disorders such as Alzheimer's disease, Parkinson's disease, Huntington's disease, and amyotrophic lateral sclerosis (ALS)(Hartl 2017). We have to discuss an additional layer of complexity: how proteins fold properly in the crowded intracellular environment and how this folding process is assisted by a chaperone protein, which function to help other proteins to fold.

Mutations in protein sequences that destabilize a protein can lead to an increased aggregation propensity in the highly crowded cytoplasm, as well as cause the loss of protein stability. Aggregation of misfolded proteins creates toxicity (toxic gain of function) and may lead to severe disease. For instance, retinitis pigmentosa mutations in the highly abundant photoreceptor protein rhodopsin affect its folding and transport and eventually result in photoreceptor cell death and blindness (Athanasiou, Aguila et al. 2018). Serious neurodegenerative conditions, including Alzheimer's disease, Parkinson's disease, Huntington's disease and prion disease, result from the aggregation of a diverse set of peptides and proteins converting into amyloid-like fibrillar assemblies. Another serious disease involving amyloids is type II diabetes(Labbadia and Morimoto 2015).

Although the structural and mechanistic bases of cytotoxicity remain under debate, evidence exists for membrane damage by oligomeric intermediates in amyloidogenesis, as well as the overload of protein quality control systems. In healthy individuals, these processes are prevented or rescued by a subset of protein known as molecular chaperones.

1.3 Protein folding *in vivo*

Protein folding *in vivo* comes with a unique set of challenges. The intracellular environment is highly crowded, with protein concentration reaching up to 400 mg/ml,

corresponding to a volume fraction of macromolecules of 20–40% of the total cellular volume. This leads to a volume exclusion or 'macromolecular crowding effect', which impacts the behavior of biopolymers inside a cell. One of the major consequences of the crowding effect is the increase in the intermolecular binding constants between partially folded states, which leads to an increased probability of aggregation during folding (Ellis and Minton 2006).

Conditions in the cell are likely to affect protein folding pathways and it has been speculated that protein structure could be dynamically 'tuned' by changing the microenvironment within the cell (Choi, Kwon et al. 2013). Different types of proteins might exploit properties of the cellular environment to increase folding efficiency. For example, the folding mechanism for a large protein is likely to be altered by folding vectorially (from N to C terminus) during translation or secretion(Yu, Dang et al. 2015). The maximum translation rates of ribosomes are typically 17–21 amino acid residues per second in procaryotes whereas eukaryotic cells typically exhibit a slower rate of about 6–9 amino acid residues per second (Ross and Orlowski 1982).

A nascent chain of average length spends about 15 seconds in the unfolded state on the ribosome, exposing hydrophobic residues, which has been suggested to increase the potential risk for nascent chain aggregation and misfolding. However, more recent studies revealed that the three-dimensional organization of individual ribosomes in polysomes maximizes the distance between nascent chains and reducing the probability of unproductive interactions between nascent chains (Brandt, Etchells et al. 2009).

1.4 Molecular chaperones

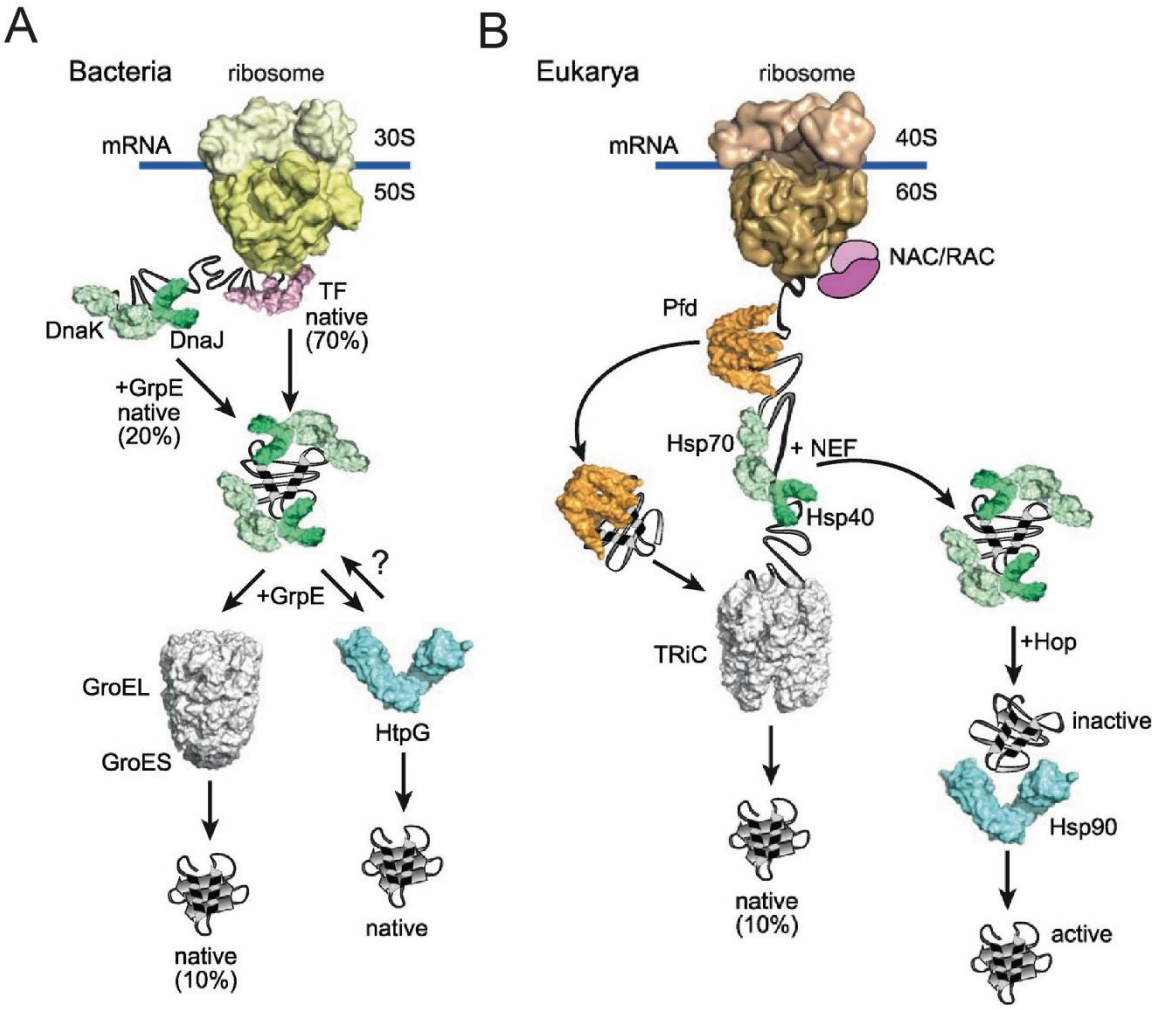

Figure 4: Chaperone assisted de novo protein folding pathways in the cytosol.

Around 70% of proteins in both bacteria (A) and eukaryotes (B) receive folding assistance from the ribosome and related factors, including the trigger factor (TF) in bacteria and the nascent chain–associated complex (NAC) and ribosome-associated complex (RAC) in eukaryotes. Post translational folding assistance by the chaperone network centers around Hsp70 (DnaK in bacteria). Partnering with Hsp40s (DnaJ in bacteria) and nucleotide exchange factors (NEFs; GrpE in bacteria), Hsp70 aids in folding approximately 20% of proteins. The remaining 10% are entrusted to chaperonins like GroEL/ES in bacteria and TRiC in eukaryotes. Some nascent proteins in eukaryotes are directly handed to TRiC through prefoldin (Pfd). This specific pathway is predominant in archaea, of which most do not possess Hsp70 proteins. Additionally, eukaryotes utilize the Hsp90 system to activate unstable proteins such as kinases and transcription factors. Proteins are transferred to Hsp90 via the Hsp70 system and the cochaperone Hop. In bacteria, the Hsp90 counterpart, HtpG, is believed to have a broader role in folding without known co-chaperones. (Balchin, Hayer-Hartl et al. 2016)

The crowded cellular environment significantly influences protein folding and stability.

The presence of other macromolecules, especially chaperones, can alter the folding

pathways and stability of proteins, leading to distinct behaviors *in vivo* compared to *in vitro* conditions (Fig. 4). Chaperones are molecules dedicated to facilitate the folding of other macromolecules. They can be specific to just one step in the assembly of one protein or broad general machines that don't rely on the recognition of specific sequences. Besides folding they are also implicated in preventing aggregation, assembly and disassembly of complexes, isoform switching, translocation and protein quality control (Balchin, Hayer-Hartl et al. 2016). Similar classes of molecular chaperones are fond in all branches of the tree of life, such as so-called, sHsp and AAA-proteins. While many use a conformation cycle regulated by ATP hydrolysis other are ATP-independent "holdases". Understanding these effects is crucial for a comprehensive understanding of protein folding, stability, and function in the cellular context. This can either be accomplished *in vitro* in simplified reconstituted systems by a reductionist approach or, ideally, by studying protein folding *in vivo* giving a holistic view.

1.4.1 Co-translational protein folding and ribosome associated chaperones

A significant portion of protein folding occurs co-translationally, as the nascent polypeptide chain is being synthesized and extruded from the ribosome. This process is not merely a matter of the proteins or protein portions ("domains") spontaneously finding their native structure. It involves a complex interplay between the translating ribosome, the emerging polypeptide, and a host of cellular machinery, including molecular chaperones. Molecular chaperones such as GroEL, GroES mainly stabilize and shield nascent polypeptides, preventing premature non-native interactions and thereby helping proteins reach their functionally active state, typically, by protecting exposed hydrophobic segments. This role is crucial as the nascent polypeptide chain is topologically restricted on the ribosome until the C-terminal region of the protein is released to engage in long-range interactions (Gloge, Becker et al. 2014).

In eubacteria, the chaperone trigger factor (TF) plays a significant role in co-translational folding. TF has an elongated three-domain structure and binds to the ribosome in a 1:1 stoichiometry using the ribosomal protein L23 at the exit site as the major docking site. By localizing directly at the ribosomal exit site, TF binds to hydrophobic stretches in nascent chains, presumably delaying chain collapse and keeping them in folding-competent states. As a result, TF slows down the rate of co-translational folding but increases the folding yield. For longer nascent chains, TF

cooperates with the downstream DnaK/DnaJ system. The release of nascent chains from TF is not ATP-dependent but is governed by the propensity to bury hydrophobic segments during translation (Fig. 4) (Kaiser, Chang et al. 2006).

In eukaryotes, the ribosome-associated complex (RAC), and the nascent chain-associated complex (NAC) are thought to fulfil a similar role to that of TF in nascent chain folding. RAC, consisting of the specialized Hsp70 Ssz1 and Hsp40 Zuo1 in budding yeast, assists nascent chain folding in cooperation with the ribosome-binding Hsp70 isoforms Ssb1 and Ssb2. NAC, a heterodimer consisting of α and β subunits, docks on the ribosome via the β subunit and binds short nascent chains. Recent findings show that NAC is not only important for co-translational folding of nascent chains, but is also required for proper intracellular protein sorting (Preissler and Deuerling 2012, Deuerling, Gamerdinger et al. 2019).

While fast folding of small, single domain proteins at the ribosome is thought to be in quasi equilibrium, co-translational folding of more complex protein structures is a non-equilibrium process, the outcome of which is dependent on the interplay between the rate of protein folding and the rate of translation by the ribosome. The kinetics of protein chain elongation can vary in a sequence-specific manner encoded in the genome. Through a combination of these effects, co-translational folding pathways may be tuned to maximize the efficiency of protein biosynthesis (Waudby, Dobson et al. 2019).

1.4.2 The Hsp70/40 system: The major sorting hub for protein folding

In eubacteria and eukaryotes, a significant part of the proteome, more than 30%, cannot achieve their native conformation solely with the assistance of ribosome-associated chaperones. For these proteins, the so-called Hsp70 system facilitates the next stage of assisted folding. Acting as a central node in the cytosolic chaperone network, Hsp70, or DnaK in bacteria, is involved in a variety of cellular processes. These include protein folding, refolding, disaggregation, and protein trafficking to different cellular compartments or to the proteolytic machinery (Calloni, Chen et al. 2012).

The molecular chaperone Hsp70 (a heat shock protein of 70 kDa) is composed of two domains connected by a hydrophobic linker: an N-terminal nucleotide-binding domain (NBD, approximately 40 kDa) and a C-terminal substrate-binding domain (SBD, approximately 30 kDa). The SBD is made up of a β-sandwich domain, which contains the substrate binding site, and an α-helical lid segment. A peptide substrate, typically a 5- to 7-amino acid motif rich in hydrophobic residues, binds in a groove of the β-

sandwich domain. The NBD contains the ATP binding pocket and ATP-hydrolysis controls the conformational cycling of the molecular machine Hsp70 (Fig. 5 A) (Rosenzweig, Nillegoda et al. 2019).

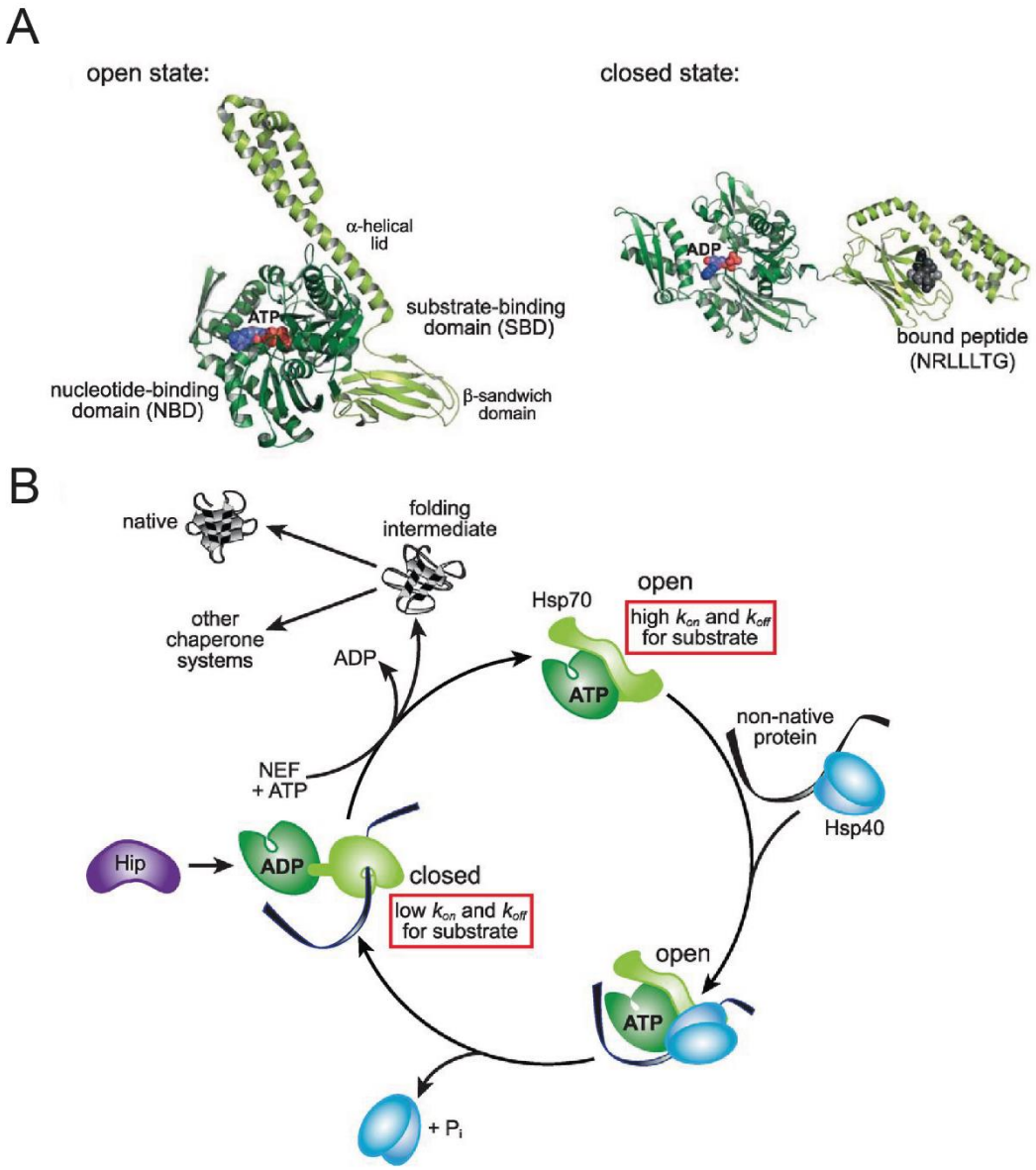

Figure 5: Mechanism and structure of bacterial Hsp70/ DnaK.

(A) Structural states of Hsp70/DnaK. The nucleotide binding domain (NBD) (dark green) undergoes significant conformational shifts in Hsp70 upon ATP binding and hydrolysis. The ATP-bound state leads to an open configuration of Hsp70 (left; PDB 4B9Q), where the α-helical lid of the substrate binding domain (SBD) associates with the NBD. Post ATP hydrolysis, the structure transitions from an open state (high substrate exchange rates) to a closed state (right; PDB 2KHO). Here, the NBD and SBD are apart, with the α-helical lid capping the peptide binding groove, resulting in low substrate exchange rates. The nucleotides within the structures have their adenosine in blue and phosphate in red. The bound peptide, represented by the sequence NRLLLTG, is showcased in a space-filling model. (B) Substrate interaction cycle. Initially, the non-native polypeptide substrate is stabilized by binding to Hsp40 and subsequently gets transferred to the ATP-bound Hsp70 in its open state. This interaction induces ATP hydrolysis in Hsp70, facilitating its transition

to the closed state. The nucleotide exchange factor (NEF) catalyzes the release of ADP, and the subsequent ATP re-binding initiates the release of the substrate for folding or potential handover to subsequent chaperones. Image adapted from (Kim, Hipp et al. 2013, Balchin, Hayer-Hartl et al. 2016).

The Hsp70 reaction cycle is tightly regulated by the co-chaperone Hsp40 and nucleotide exchange factors (NEFs). Hsp40 cochaperones, also known as J-domain proteins in bacteria, recognize and deliver substrates to Hsp70 in its ATP-bound state. In this state, the hydrophobic linker and the α-helical lid of the SBD interact with the NBD, causing the SBD to adopt an open conformation with high substrate on and off rates. The interaction of Hsp40 with the Hsp70 NBD and the hydrophobic linker segment greatly stimulates ATP hydrolysis in the NBD. The hydrolysis of ATP to ADP triggers an allosteric rearrangement that closes the α-helical lid, thereby tightly trapping the bound SP with low on and off rates. The subsequent binding of a NEF, such as GrpE in bacteria, to the NBD catalyzes the exchange of ADP for ATP, which results in the opening of the α-helical lid and the release of the SP for folding or transfer to downstream chaperones or the degradation machinery (Fig. 5 B). In metazoans, additional factor Hip stabilizes the ADP state, thereby postponing substrate release. Extended binding of a non-native protein to Hsp70 might promote its degradation. Through repeated cycles of high and low SP binding affinity, Hsp70 prevents off-pathway aggregation and corrects misfolding by removing abnormal long-range interactions present in the folding intermediate. This is enabling the distribution of folding over times and can be described as a kinetic partitioning mechanism. A subset of proteins still cannot reach its native fold and is handed over for further processing to the chaperonin system or Hsp90 (Balchin, Hayer-Hartl et al. 2016).

1.4.3 The role of Hsp90 in proteostasis

Heat shock protein 90 (Hsp90) is an indispensable molecular chaperone in eukaryotic cells, serving as a proteostasis hub that orchestrates numerous essential signaling pathways (Taipale et al., 2010). Beyond its conventional role in protein folding, Hsp90 aids in the conformational maturation and maintenance of a wide array of signaling molecules, including proto-oncogenic kinases, transcription factors, and steroid hormone receptors (Balchin et al., 2016; Taipale et al., 2012). This multifaceted role in protein homeostasis allows Hsp90 to buffer destabilizing mutations in its client proteins, presenting an evolutionary advantage by facilitating the emergence of innovative traits (Lindquist, 2009; Rutherford and Lindquist, 1998).

Structurally, Hsp90 operates as a dimer. Each subunit is compartmentalized into an N-terminal nucleotide-binding domain, a C-terminal dimerization domain, and a middle domain that bridges the two (Fig. 6 A) (Ali et al., 2006; Karagoz and Rudiger, 2015; Verba et al., 2016). This dimeric architecture of Hsp90 undergoes dynamic conformational shifts through an ATP-driven reaction cycle. These changes, transitioning from an open state to a closed conformation, are influenced by several factors such as nucleotides, client proteins, and various cochaperones. Interestingly, although ATP hydrolysis and ADP release are intrinsic to this cycle, the rate-limiting step lies in the expansive conformational transition from the open to the closed state (Fig. 6 B) (Hessling et al., 2009).

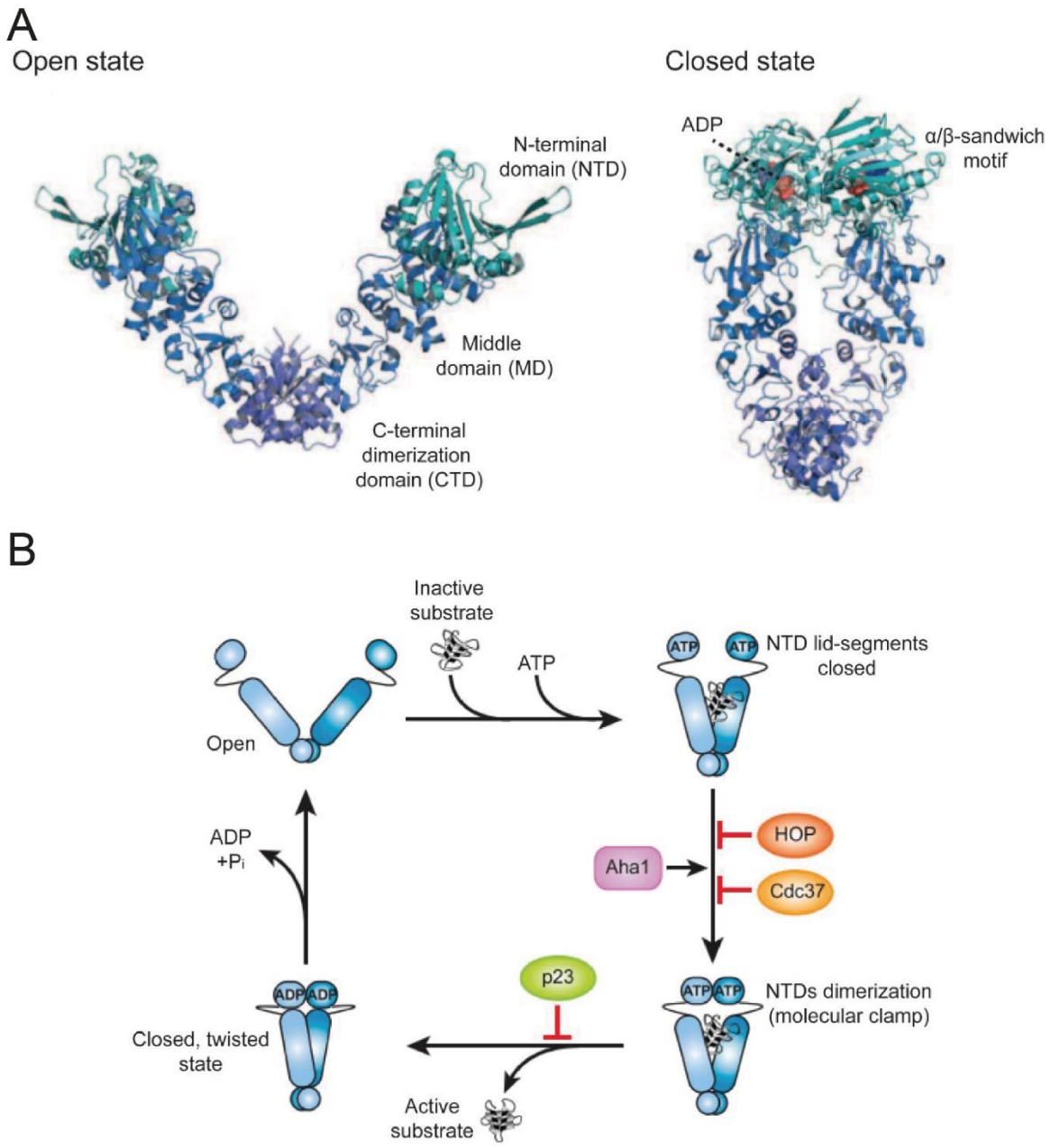

Figure 6: The Hsp90 reaction cycle.

(A) The crystal structures of bacterial Hsp90 in its open configuration is shown on the left (PDB 2IOQ) and the yeast Hsp90 in its closed form on the right (PDB 2CG9). Hsp90 functions as a dimer. Each monomer of Hsp90 is segmented into an N-terminal domain for nucleotide binding (NTD), an intermediary domain (MD), and a C-terminal domain that facilitates dimerization (CTD). (B) The functional cycle of Hsp90. An unfolded client protein associate with the Hsp90 dimer. ATP binding shifts the chaperone from its open configuration to a closed one, a transition characterized by the dimerization of the NTD. This metastable state is committed to ATP hydrolysis, inducing a yet further closed, twisted state. The eventual release of ADP and inorganic phosphate (Pi) restores the Hsp90 open conformation. Several co-chaperones, such as Hop, Cdc37, Aha1, and p23, interact with Hsp90, modulating its ATPase activity and facilitating client recruitment. Adapted from (Balchin, Hayer-Hartl et al. 2016) (A) and (Hartl, Bracher et al. 2011) (B).

Cochaperones play a pivotal role in modulating Hsp90's function. Many, like Hop and Cdc37, employ tetratricopeptide repeat domains to bind to Hsp90 (Scheufler et al., 2000). Hop facilitates the transfer of client proteins from Hsp70 to Hsp90, while Cdc37 acts as an adapter for kinase proteins. Both stabilize Hsp90 in its open configuration (Scheufler et al., 2000; Verba et al., 2016). In contrast, Aha1 accelerates Hsp90's transition to its closed state by binding asymmetrically to the chaperone, thus augmenting ATPase activity (Retzlaff et al., 2010). Additionally, p23 acts later in the cycle, enhancing client maturation by stabilizing Hsp90's closed conformation before ATP hydrolysis (Li et al., 2012).

In bacterial cells, the homolog of Hsp90 is known as HtpG (high-temperature protein G). While eukaryotic Hsp90 plays a central role in cellular processes and proteostasis, the function of HtpG in bacteria is less well understood. HtpG is generally induced under stress conditions, such as elevated temperatures, similarly to other heat shock proteins. Despite the structural similarities between Hsp90 and HtpG, the exact roles of HtpG in bacterial physiology and stress response are not as well-defined as those of its eukaryotic counterpart. It's worth noting that not all bacteria possess HtpG, and among those that do, its abundance and importance can vary. However, studies have suggested that HtpG, like Hsp90, may assist in the folding and stabilization of specific client proteins, particularly those involved in signal transduction and adaptive responses (Wickner, Nguyen et al. 2021).

1.5 Chaperonin assisted protein folding

Chaperonins are essential in most eubacteria as well as in eukaryotes and archaea. The function of all of them is tied to the formation of an enclosed chamber inside which client proteins can fold shielded from unproductive, off pathway intermolecular interactions. They are separated in two groups according to their molecular architecture. The group I chaperonins are composed of two seven membered Hsp60 (GroEL in bacteria) rings stacked back-to-back and complemented by the co-factor Hsp10 (GroES in bacteria) which forms a lid. They are present in most eubacteria as well as in eucaryotic organelles presumably derived from endosymbionts, mitochondria (Hsp60) and chloroplasts (Cpn60) (Balchin, Hayer-Hartl et al. 2020, Horwich and Fenton 2020). The group II chaperonins form hexadecameric or octadecameric assemblies, with two eight membered rings stacking back-to-back
and are functional without the addition of co-chaperones, having an inbuilt lid.

Each subunit can be structurally divided into an equatorial ATPase domain, an intermediate hinge-like domain, and an apical domain. The equatorial domain facilitates intra- and inter-ring interactions, while the apical domain is responsible for binding non-native SP and the cochaperonin GroES (Hayer-Hartl, Bracher et al. 2016).

1.5.1 TRiC, the eucaryotic chaperonin complex

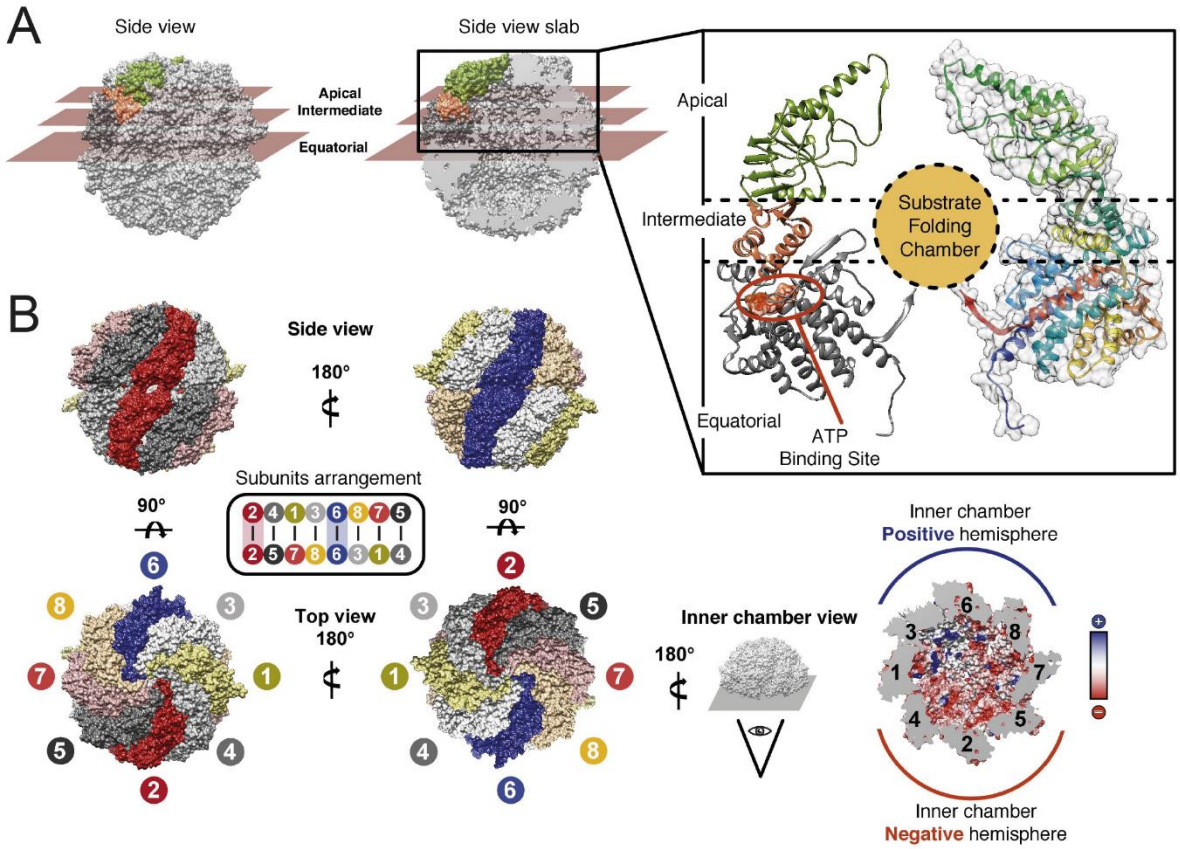

Figure 7: Overview of TRiC's structure and subunit organization.

(A) The illustrated closed TRiC configuration showcases the three distinct domains within a single subunit: the equatorial domain in dark gray, which binds ATP; the apical domain in green, handling substrate binding and providing an in-built lid; and the hinge domain in salmon, translating ATP-related changes into apical domain movements. The domain's distribution is shown through a planar cross-section and is also depicted in a ribbon representation of a single subunit.
(B) The subunit layout of TRiC is displayed, presenting rings aligned in a back-to-back manner. The positioning shows CCT2 (in red) and CCT6 (in blue) directly on top of one another. In the bottom right of section (b), the chamber's inner electrostatic potential is illustrated, revealing charge asymmetry. One side (in blue) is covered with positive charge side chains (+), while the opposite side (in red) features negative charge side chains (−). The structural details are based on the PDB file: 4V94. Reproduced from (Gestaut, Limatola et al. 2019)

In eukaryotes, the TCP-1 ring complex (TRiC), also known as chaperonin containing tailless complex polypeptide 1 (CCT) in archaea, is involved in the folding of a significant portion of cytosolic proteins, notably the essential cytoskeletal proteins actin and tubulin. Structurally, TRiC is a hexadecameric assembly composed of eight distinct paralogous subunits (CCT1–CCT8). This unique structure facilitates the recognition of nascent proteins at the apical domain, which is equipped with a built-in lid. The ATP

binding and hydrolysis in the equatorial domains drive the lid's closure, facilitating an open to close conformational transition. The evolutionary divergence into eight paralogous subunits is thought to allow TRiCs fine-tune its substrate specificity for various client proteins (Fig. 7) (Gestaut, Limatola et al. 2019).

The fact that it consists of eight different subunits suggests a multifaceted mechanism governing client recognition and subsequent action. The subunits possess charged and hydrophobic residues in their apical domains, which are the canonical substrate binding sites. These interactions are believed to be mediated by specific contacts between particular apical domain residues in TRiC subunits and specific domains of the client protein.

1.5.1.1 TRiC cochaperones

Furthermore, TRiC collaborates with a vast array of chaperones and co-chaperones. The co-chaperones that assist TRiC are prefoldin and the phosducin-like proteins (PhLP).

Prefoldin is a hexameric co-chaperone complex that is thought to play a crucial role in the early stages of protein folding. Prefoldin acts on nascent polypeptides emerging from ribosomes. It recognizes and stabilizes these newly synthesized polypeptides, especially those with complex domain topologies, and directs them towards TRiC. This handover ensures that these polypeptides are efficiently folded by TRiC, preventing potential misfolding and aggregation. By acting as an intermediary, Prefoldin not only increases the efficiency of protein folding but also ensures that a broader range of substrates can be accommodated by TRiC (Gestaut, Zhao et al. 2022, Horovitz, Reingewertz et al. 2022, Kelly, Tranter et al. 2022).

The phosducin-like proteins were proposed to assist the folding of G-proteins α-subunit (PhLPx), actin (PhLP2) and tubulin (Plp1). PhLP2 plays a nuanced role in the TRiC -mediated protein folding process. Unlike Prefoldin, which acts as a substrate delivery system, PhLP2 modulates the activity of TRiC within the context of a ternary complex comprising TRiC, the substrate, and PhLP2 itself. This co-chaperone seems to act as a molecular bridge or strut between the substrate and the TRiC chamber, birding both chambers and potentially influencing the folding dynamics within the chamber. The exact mechanisms by which PhLP2 modulates TRiCs activity are still a subject of research (Han, Jin et al. 2023).

Even though type II chaperonins and type I chaperonins are similar in overall shape, the mechanism of action seems to be different, as indicated by the inability to fold each

other's obligate substrates, such as actin or tubulin. The mechanism, structure and substrates of GroEL, the bacterial chaperonin and focus of this study, will be explained in detail in the next chapter.

1.5.2 The bacterial chaperonin: GroEL/ES

Within the realm of molecular chaperones, the GroEL chaperonin from *Escherichia coli* stands as a paradigmatic example of group I chaperonins. Distinct from eukaryotic counterparts such as TRiC, GroEL has been extensively studied.

GroEL is characterized by a tetradecameric assembly, comprising two heptameric rings of ~57 kDa GroEL subunits stacked back-to-back.

The combined action of GroEL and GroES provides a protective nano-cage, ensuring that protein folding occurs in isolation, free from aggregation risks. This system is thought to rely on intra-ring positive allosteric cooperativity and inter-ring negative cooperativity.

1.5.3 Structure of GroEL/GroES

Over the years the GroEL/ES system has been subject of intensive structural biology investigations, as indicated by the over 200 deposited structures in the PDB at the time of writing.

Early work established the overall shape of the EL:ES complex and found two different architectures: A symmetrical football shaped EL:ES_2 complex one and an asymmetric bullet shaped EL:ES_1 complex (Fig. 8).

Investigations into substrate binding to the "naked" EL 14-mer showed significant reorientations along the whole complex. It also established that the helices αI and αH in the apical domain act as the primary binding sites for substrate. Particularly, hydrophobic residues exposed to the central cleft, facilitate the binding of molten globule-like folding intermediates (Wang and Chen 2003).

Different cryo electron microscopy (cryo-EM) studies established a movement of the apical domains as part of the ATP cycle of GroEL. Clare et al established that, during the ATPase cycle reorientation of the apical domains of EL change the diameter of the chamber opening of EL (Clare, Vasishtan et al. 2012).

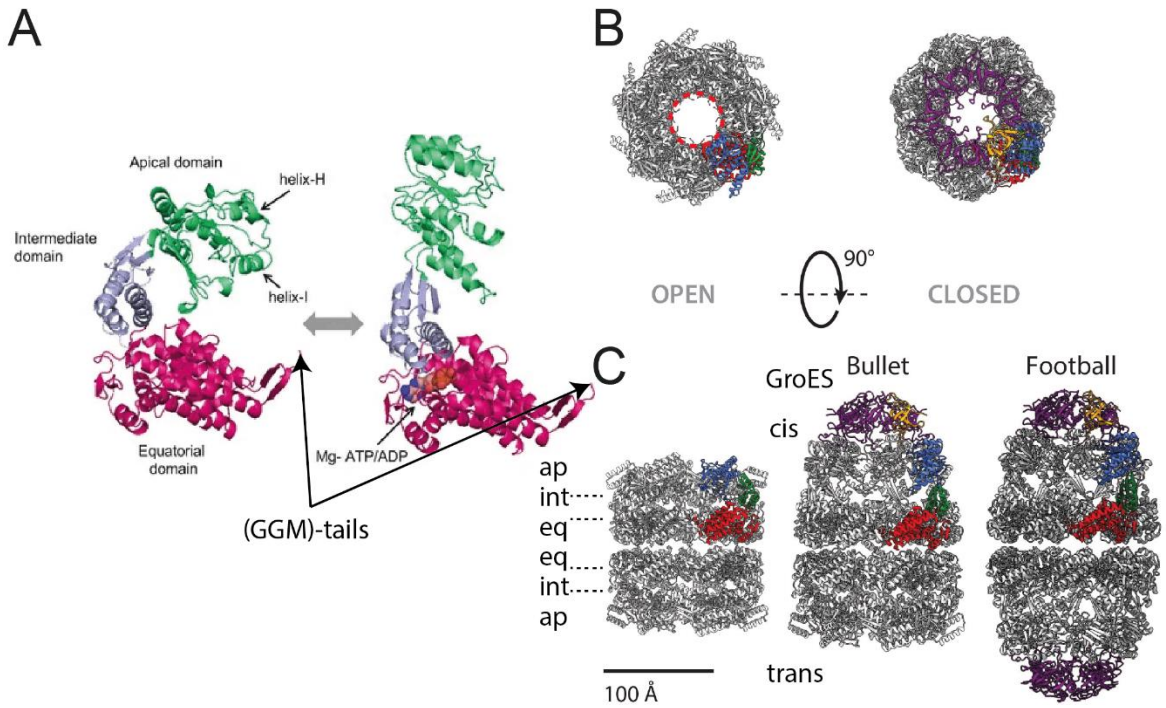

Figure 8: Architecture of chaperonin.

(A) Crystal structures of GroEL subunits, both with and without a bound nucleotide. The binding site for the substrate protein is located at helices αH and αI. With nucleotide binding, the apical and intermediate domains move upward, leading to a substantial expansion of the cavity on the *cis*-side. This movement aids in the encapsulation of the substrate protein into the Anfinsen's cage. The starting points for the flexible GGM repeat tails are indicated with a black arrow. (B) Two distinct views of the GroEL atomic structure in its unbound form and its distinctive bullet- or football-like complexes with GroES. The latter is formed by the GroES attachment (GroES colored in purple, one of its subunits highlighted in orange) either to one ring (yielding an asymmetric or 'bullet' shape) or to both rings (creating a symmetric or 'football' shape). The respective atomic models are based on GroEL (PDB: 1XCK), the EL:ES_1 complex (PDB: 1AON), and the EL:ES_2 structure (PDB: 4PKO). The red dashed circle in (B) highlights the apical domains helices αH and αI, crucial for engagement with protein substrates. The domain structure in indicated on the left with eq standing for the equatorial domain, ap for the apical and int for the intermediate domain. Adapted from (Ishii 2017) and (Horovitz, Reingewertz et al. 2022)

1.5.4 The GroEL/ES reaction cycle

The reaction cycle of GroEL relies on large conformational transitions. Upon ATP binding, GroEL undergoes allosteric changes, facilitating the binding of the dome-shaped GroES, a heptameric ring of ~10 kDa subunits. This interaction sequesters the substrate protein within the chamber. The basic cycle consists of substrate binding, GroES-mediated encapsulation, ATP hydrolysis-driven folding and subsequent release (Fig. 9).

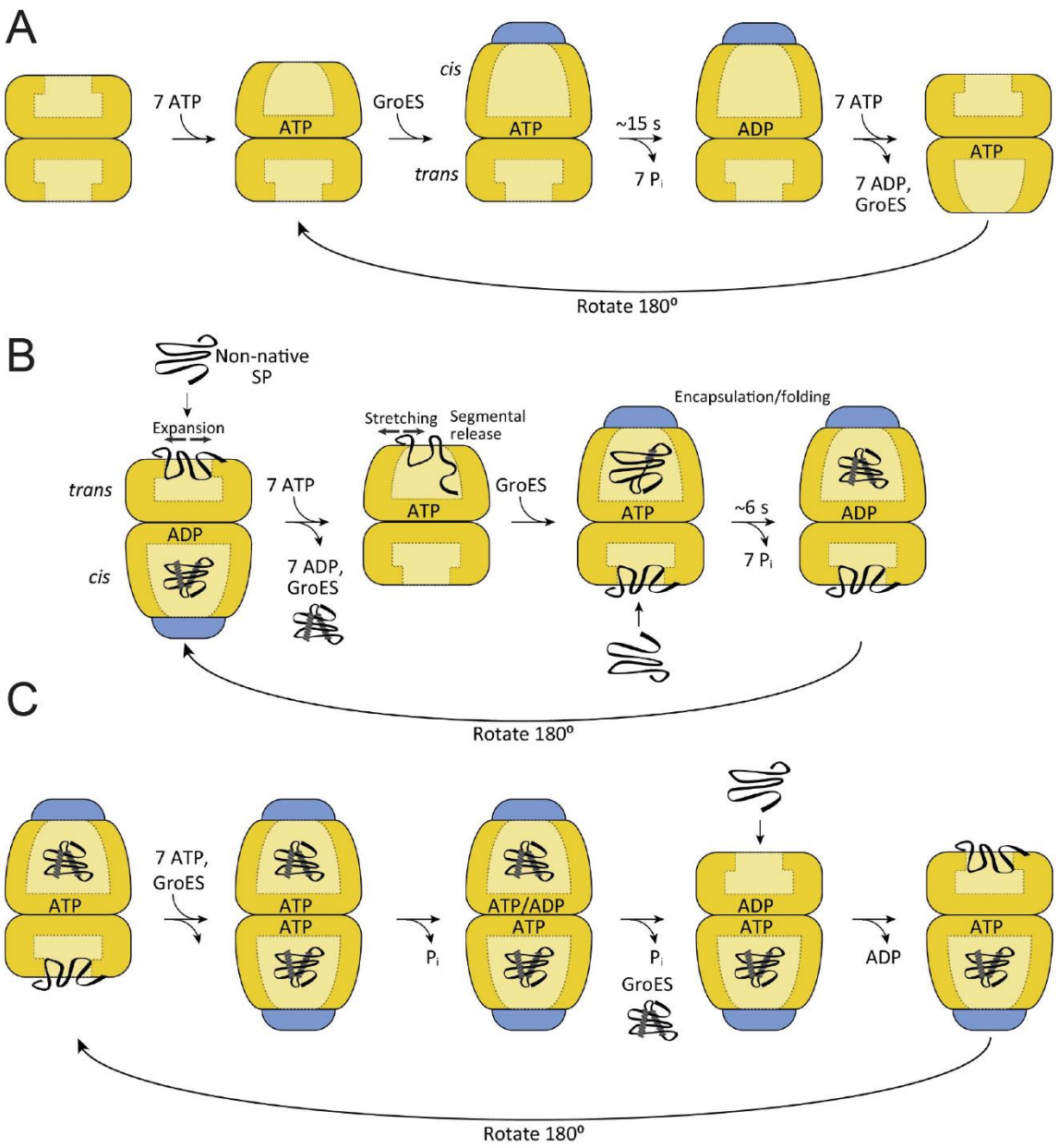

Figure 9: Possible reaction cycles of GroEL and GroES.

(A) Asymmetric reaction cycle of GroEL/ES when SP is absent. GroEL is expected to be consistently bound to nucleotide and GroES *in vivo.* (B) & (C) Asymmetric (B) and symmetric GroEL/ES interaction cycles (C) when SP is present, respectively. The conformational changes of SP are highlighted in (B). It's crucial to point out that while encapsulation folding isn't deemed mandatory in the model presented in (C), for the sake of clarity, only in-cage folding is illustrated. (Hayer-Hartl, Bracher et al. 2016)

Despite a multitude of studies, a consensus on several aspects of the reaction cycle remains elusive. This ambiguity stems from the inherent complexity of the GroEL machinery, which can undergo a multitude of parallel and sequential kinetic processes, cycling through various conformational and ligand-bound states. Adding to the complexity, the GroEL-mediated folding mechanism can exhibit variability based on the identity of the substrate and the specific experimental conditions chosen *in vitro*

(Hayer-Hartl, Bracher et al. 2016, Horwich and Fenton 2020, Horovitz, Reingewertz et al. 2022).

Even in the absence of substrate protein (SP), GroEL and GroES undergo ATP-regulated binding and release. This results in an asymmetry between GroEL rings, especially in relation to nucleotide presence and GroES binding. The GroEL subunits exist in an equilibrium between the T-state (tens) (low affinity for ATP) and the R-state (relaxed) (high affinity for ATP). Cooperative binding of ATP to one GroEL ring stabilizes the R-state, inducing a movement in the apical domains that facilitates GroES binding to the cis-ring in about 200 milliseconds. Following ATP binding, GroES binds, leading to the formation of the GroEL–GroES complex. ATP molecules are then hydrolyzed, leading to the release of ADP and GroES, resetting the system for another cycle.

During this whole cycle, the negative cooperativity between the two rings stays intact. This cycle can be potentially modified by the presence of substrate protein.

The nucleotide-free ring (so-called *trans*-ring) of the GroEL/ES complex is thought to be the acceptor state for a non-native SP (Figure 8 and 9). Collapsed folding intermediates in the “molten globule” state, lacking stable tertiary elements and thus exposing hydrophobic segments, usually engage two or more apical domains within one ring for efficient binding (Hartl 1996, Sharma, Chakraborty et al. 2008). Binding to GroEL is accompanied by a rapid conformational expansion of collapsed substrates (t1/2 ~100 ms) as measured by fluorescence resonance energy transfer (FRET) (Lin, Madan et al. 2008) (Sharma, Chakraborty et al. 2008).

Upon ATP binding, the *trans*-ring undergoes a significant conformational change, aiding in further expanding the SP and releasing weakly bound segments. This ATP-triggered expansion is transient and is immediately followed by GroES binding. GroES binds to the cis-ring, displacing the SP into the enclosed chamber. The SP is then free to fold inside this chamber for a duration, dependent on temperature, typically ranging from 2 to 7 seconds, while ATP hydrolysis occurs. This leads to the release of ADP and GroES. The SP, if not yet folded, is rapidly recaptured by GroEL for another folding attempt (Hayer-Hartl, Bracher et al. 2016).

In the complex with GroES, the inner wall of the GroEL/ES cage is hydrophilic and becomes net-negatively, providing an environment permissive for folding (Xu, Horwich et al. 1997). In this state the substrate binding hydrophobic groove is buried. The

GroEL/ES cage (~175 000 $Å^3$) is approximately twice the size of that of an apo GroEL ring without GroES, and is sufficiently large to encapsulate SPs up to ~60 kDa in size. In the chaperonin cycle described above, the two rings of GroEL fold SPs sequentially, and mainly asymmetrical GroEL:GroES complexes are populated. The *trans*-ring can bind ATP and GroES only after the cis-ring has hydrolyzed its bound ATP (Rye, Roseman et al. 1999). This asymmetry is due to the negative allosteric coupling of the GroEL rings, with communication between the two rings being mediated by critical interactions at the inter-ring interface of the equatorial domains (Gruber and Horovitz 2016), (Horovitz and Willison 2005) (Saibil, Fenton et al. 2013) (Yan, Shi et al. 2018). However, symmetric EL:ES_2 complexes with GroES binding simultaneously to both GroEL rings, have also been reported in the absence or presence of SPs (Sameshima, Ueno et al. 2008) (Schmidt, Rutkat et al. 1994) (Yang, Ye et al. 2013, Ye and Lorimer 2013). The functional importance of the symmetric complexes has been debated (Horwich and Fenton 2020). Based on the symmetric EL:ES_2 complexes, a non-sequential model has been proposed (Fig. 9), in which GroES units bind simultaneously to both GroEL rings and dissociate stochastically upon ATP hydrolysis, with SP accelerating nucleotide exchange (Hayer-Hartl, Bracher et al. 2016) (Yang, Ye et al. 2013). SP binding to the *trans*-ring would stimulate the rate-limiting ADP dissociation and thereby allow fast ATP and GroES binding before the *cis*-ring has hydrolyzed ATP (Ye and Lorimer 2013). Subsequent studies applied the same calibrate FRET pair measurement to EL:ES_2 complexes but generated different results depending on the fluorophore pairs for labelled GroEL and GroES. A more robust study applying novel dual-color fluorescence cross-correlation spectroscopy and avoiding GroEL labeling, strongly suggested that symmetric EL:ES_2 is disfavored in the presence of foldable SPs and at a physiological ATP:ADP ratio *in vitro* (Haldar, Gupta et al. 2015).

Differences in these studies might also be due to differences in concentrations of the different components of the system. The interplay of a large number of factors, GroEL, GroES, ATP/ADP ratios and the concentration of ions in the medium as well as the concentration and characteristics of substrate protein make this intricate system difficult to reconstitute *in vitro*. Factors like molecular crowding, additional chaperones and co-translational folding further complicate the situation *in vivo*.

Another open topic is the exact mechanistic pathways of how GroEL aids in protein folding. Three mechanistic models have been proposed and will be discussed in the

following: The passive cage model, the active cage model and the iterative annealing model.

In the passive cage model, GroEL functions as a protective microenvironment. According to this idea, GroEL provides conditions analogous to an infinitely diluted solution, where proteins can fold shielded from aggregation ("Anfinsen cage"). Obligate SPs, which have been observed to fold exclusively within the GroEL/ES cage, lead to the proposition of this model. Contrary to the model prediction empirical evidence shows GroEL's ability to accelerate protein folding, which contradicts a purely passive role. To resolve this contradiction proponents of this model suggested the existences of reversible aggregation that would be prevented inside of the chamber. Although theoretically possible these species have never been observed.

The active cage model states that the properties of the GroEL chamber lead to a further rate enhancement beside the pure passive prevention of aggregation. The specific structural attributes of the GroEL cavity, encompassing its volume, charged walls, and C-terminal GGM repeats sequence, are posited to play an important role in enhancing protein folding kinetics. While theoretical frameworks align with this active mechanism, suggesting GroEL’s entropic confinement optimizes the folding landscape, the model's universal applicability remains a matter of debate. Not all proteins manifest accelerated folding kinetics in GroEL's presence, possible due to their smoother folding energy landscape.

The iterative annealing model adds a dynamic dimension, emphasizing ATP-driven interactions between GroEL and substrate proteins. This model suggests GroEL periodically refines protein conformations by 'resetting' misfolded states, thus offering them repeated opportunities to achieve their native structures. Some studies indicate that a singular interaction with GroEL can achieve the observed accelerated folding, challenging the necessity for continuous iterative cycles. Moreover, mounting evidence underscores the essential nature of the GroEL/ES cage in achieving optimal folding kinetics (Hayer-Hartl, Bracher et al. 2016, Horovitz, Reingewertz et al. 2022).

1.5.5 Substrates of GroEL/ES

GroEL/ES is indispensable for *E.coli* growth across all conditions, underscoring its role in protein folding (Georgopoulos 2006). An in-depth study identified a set of approximately 250 proteins, encompassing 67 essential ones, that interact with GroEL during *in vivo* translation. This subset represents around 10% of *E. coli's* total cytosolic proteome (Kerner, Naylor et al. 2005, Niwa, Fujiwara et al. 2016). The majority of these

GroEL substrates fall within the size range of ~35–60 kDa, compatible with the volume of the GroEL/ES cage (Fig. 10). However, certain larger substrates have been identified, suggesting that they might leverage GroEL primarily to prevent aggregation. These larger proteins might undergo *trans*-ring binding for their folding, bypassing the need for complete encapsulation within the GroEL/ES cage (Chaudhuri, Verma et al. 2009).

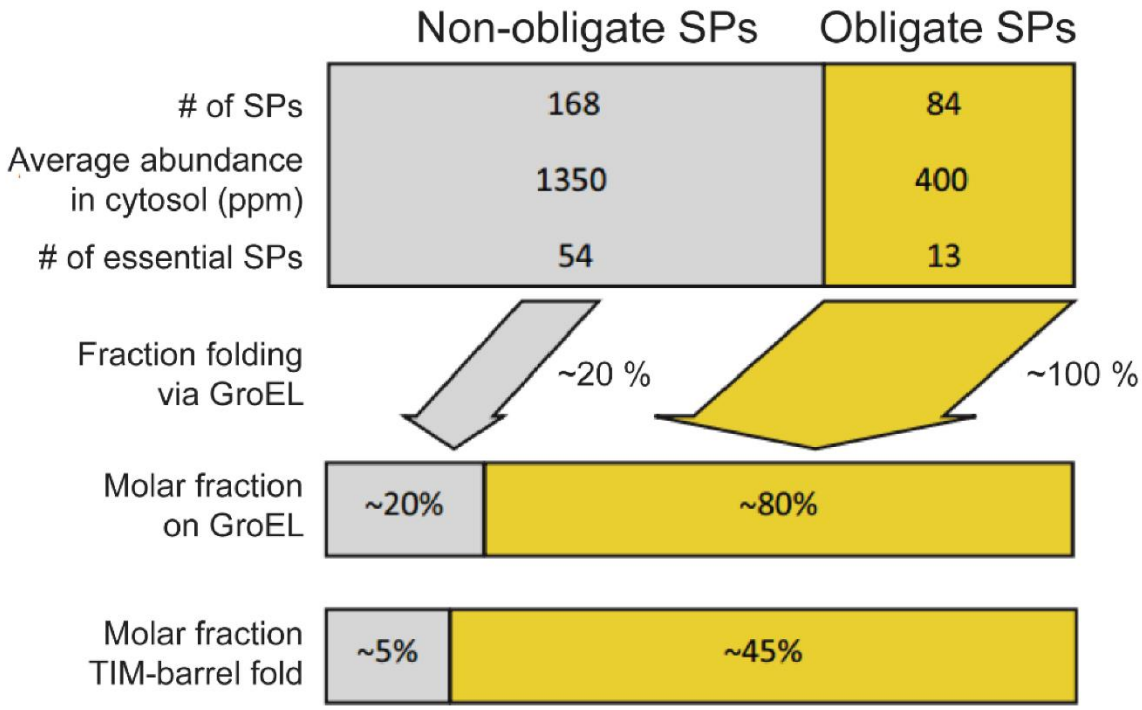

Figure 10: *In vivo* substrates of GroEL/ES.

Number of non-obligate versus obligate GroEL SPs and the proportion they occupy in terms of GroEL's molar capacity (based on data from (Kerner, Naylor et al. 2005)). Typically, non-obligate SPs employ upstream chaperones for their folding processes. (Hayer-Hartl, Bracher et al. 2016)

Delving deeper into the GroEL interactors, a specific subset of around 80 proteins, including 13 essential ones, has been found to be strictly reliant on the GroEL/ES system for their folding. This subset occupies a significant portion, approximately 75-80%, of the total chaperonin capacity (Kerner, Naylor et al. 2005). In experiments where GroEL was deleted, about 50 of these proteins were validated as obligate chaperonin substrates (Fujiwara, Ishihama et al. 2010). A striking feature of these obligate substrates is the prevalence of α/β folds or α+β domain topologies. A notable enrichment was observed for substrates with the (β/α)-TIM barrel topology. These proteins, stabilized by extensive long-range contacts, in sequence space, in their native conformations, are slow-folding and prone to forming kinetically trapped

intermediates during their folding journey. Such proteins face challenges navigating their folding energy landscapes, often getting trapped in non-native conformations (Dobson, Sali et al. 1998, Hayer-Hartl, Bracher et al. 2016). This propensity underscores the indispensable role of GroEL/ES in assisting these proteins to achieve their native, functional states. Despite success in the substrate identification, predicting a protein's folding dependence on GroEL based solely on its sequence or structure remains challenging, necessitating further exploration into additional determining features.

1.5.6 GroEL-substrate complexes

Early work by crystallography established the substrate binding of GroEL via peptides(Wang and Chen 2003). Subsequent cryoEM studies showed by Chen et al showed how the folding of a Rubisco, an obligate substrate of the GroEL system, alters the overall architecture in an $EL:ES_1$ complex. It also established the location of the substrate inside the EL chamber near the bottom and allowed first hints of possible interacting residues based on a medium resolution structure (Chen, Madan et al. 2013). Further work by Kim et al found that RuBisCo could be located in both chambers of an *in vitro* reconstituted $EL:ES_2$ complex and confirmed the location of the substrate near the bottom of the chamber as well as the possibility that the residues Phe 281, Tyr 360 as well as Phe 44 could from potential contacts of the EL chamber wall (Kim, Park et al. 2022). Closer characterization of these potential interactions as well as a confirmation of the folding state of the substrate protein inside the EL chamber was not possible in this work due to the limited resolution, especially in the substrate. The lack of clear secondary structure elements in the substrate density also made it difficult to establish how the substrate breaks the 7-fold symmetry of the EL chamber, due to the fact that the resulting structures could still be averages of multiple conformations.

Despite the extensive prior structural work, the structure of an EL:ES complex containing a folded substrate protein still remains elusive.

2 Introduction into cryo electron tomography

Technological, medical and biological progress was always driven in part by our ability to see things that were previously hidden from us. To see things with a new set of eyes can be taken literally when it comes to microscopy. Progress since the invention of the first magnifying glasses in the 13th century allowed humanity to see things that were either too small or too far away to see with the naked eye. While the invention of the first telescopes in the 17th century enabled the study of the universe and greatly enhanced our understanding of physics, the invention of the microscope enabled the first observation of the microcosm around us.

Progress since then has opened up new worlds for us to explore. From galaxies in the early universe to the structures of our own cells, a plethora of imaging technologies has been developed to visualize matter at every scale.

2.1 Microscopy techniques and resolution

Light microscopy, which employs visible light to illuminate samples, has been a foundational tool in biomedical research since the 19th century. It has allowed scientists to observe everything from cellular structures to the intricate details of tissue samples. It operates on the principle of optical magnification to visualize minute structures, employs visible light which passes through, reflects off, or is emitted by the sample being observed. This light is then magnified through a series of lenses. Initially, a light source such as a bulb or LED illuminates the specimen. Following the path of light, the condenser lens focuses the light from the source onto the specimen, ensuring that the light covers the viewed area evenly. The primary lens in this setup, the objective lens, magnifies the image of the specimen with different lenses offering varying degrees of magnification, commonly ranging from 4x to 100x. Further magnification occurs in the eyepiece lens, where the magnified image formed by the objective lens is enlarged further. The total magnification of the microscope is the product of the magnifications of the objective and eyepiece lenses. Focusing mechanisms are integral to this process, adjusting the distance between the objective lens and the specimen, or between the eyepiece and the objective lens, ensuring a sharp image is obtained. The stage holds the specimen in place and can usually be moved in the X, Y, and Z planes to position the specimen precisely. Finally, the microscope generates a magnified, real image which can be viewed through the eyepiece or projected onto a screen or sensor for digital imaging.

However, as with all scientific tools, both light microscopy and its contemporary offspring, fluorescence microscopy, come with their set of limitations. The resolution of light microscopy is inherently constrained by the wavelength of the light used, as defined by the Abbe's diffraction limit. The smallest distance *d* resolvable by a given wavelength of light (λ) is dependent on refractive index *n* and the half-angle α:

$$d = \frac{\lambda}{2n\,sin\alpha} = 0.5 \cdot \frac{\lambda}{NA}$$

Another way to define resolution, the ability to distinguish between two closely spaced objects, is defined by the Rayleigh criterion, which relates resolution (d) to the wavelength (λ) of the imaging medium and the numerical aperture (*NA*) of the microscope:

$$d = 0.61 \cdot \frac{\lambda}{NA}$$

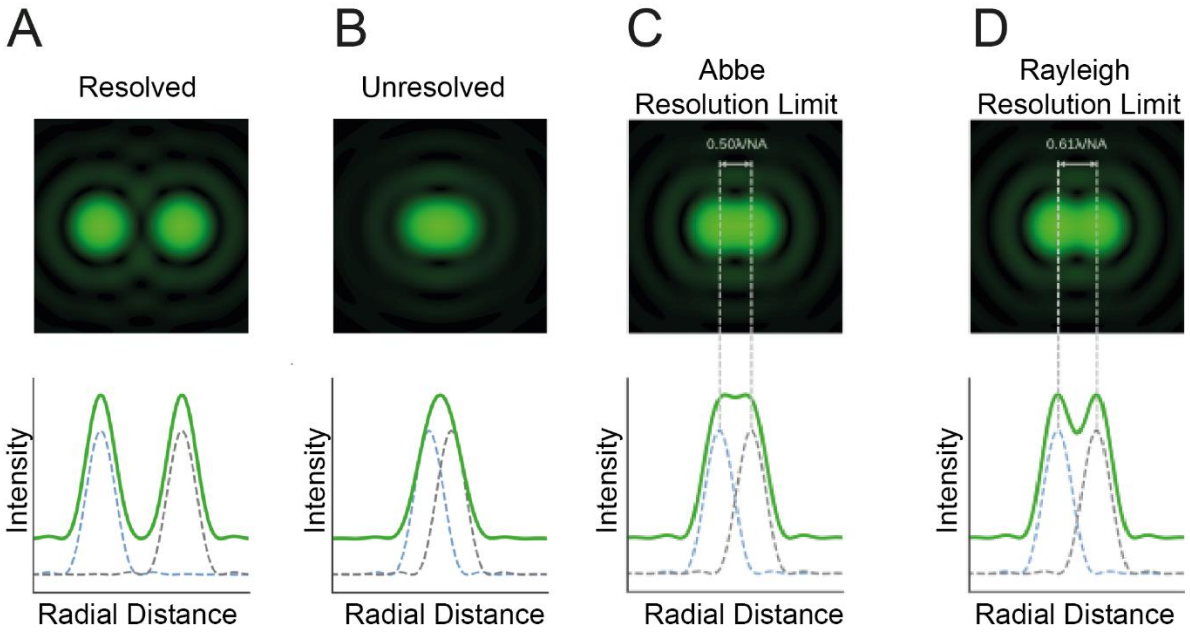

Figure 11: Interplay of Airy Discs from two converging point sources, the Abbe and Rayleigh criterion's role.

When the peaks of two-point sources marginally overlap (A), they remain distinguishable. However, when their combined signals merge to form one distinct peak (B), differentiation becomes impossible. Two widely accepted Limits to resolution are the Abbe (C) and Rayleigh criteria (D). These criterions are defined as 0.5·λ / NA and 0.61·λ / NA respectively, where NA is given by n sin α. Adapted from (Instruments 2023).

Irrespective of the resolution criterion chosen, it is clear that light with a smaller wavelength result in a higher resolution. For the use of visible light, this limits normal optical microscopy to a resolution of around 200 nm. Theoretical improvements in resolution can be achieved using X-rays due to their shorter wavelengths. However, X-ray manipulation presents challenges, such as focusing difficulties and potential sample damage (WA 1959).

To achieve even higher resolutions, we must look beyond photons. The de Broglie equation allows us to calculate a wavelength for any particle with mass (m) and momentum (p):

$$\lambda = \frac{h}{p} = \frac{h}{m \cdot v}$$

With the Plank constant *h*, and the momentum *p* of a particle being defined by its mass *m* and velocity *v*.

Considering this, neutrons could be a viable choice for imaging due to their high mass. However, the low neutron flux densities from available sources make them impractical to date. Electrons emerge as the preferred choice (Henderson 1992). At an acceleration voltage of 100 keV, electrons have a wavelength of around 0.037 Å. With current technological advancements, high-energy electrons can achieve resolutions below the atomic level, bridging the gap between x-ray crystallography and the limits of light microscopy (Fig. 12) (Kohl and Reimer 2008).

While advancements in light microscopy, particularly super-resolution techniques like stimulated emission depletion (STED), have bypassed Abbe's diffraction limit to achieve resolutions which are higher by orders of magnitudes, they still rely on fluorescently labeled structures. This means that visualization remains confined to these labeled structures, obscuring other details within the cell.

In contrast, transmission electron microscopy (TEM) offers a label-independent approach, enabling researchers to image any cellular structure, from larger organelles to macromolecules (Bond, Santiago-Ruiz et al. 2022).

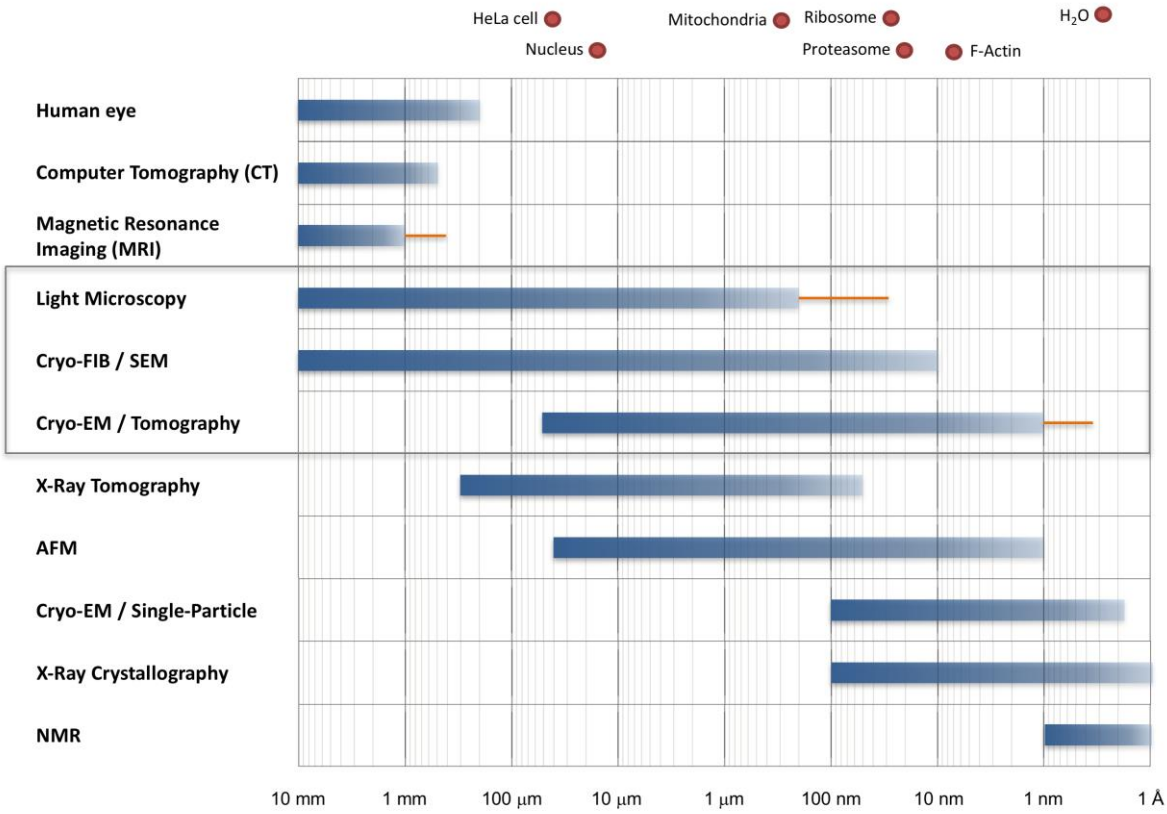

Figure 12: Resolution ranges of different imaging techniques whit a biological context.

General overview of imaging techniques applied to biological specimens and their respective resolution ranges. Blue bars represent the resolution capabilities of each technique, while orange lines highlight recent innovations that surpass earlier resolution limits for that specific method. Above the graph, red circles mark the typical sizes of biological structures. Adapted from (Bäuerlein 2018)

2.2 The Transmission electron microscope

The TEM was pioneered by Ernst Ruska and Max Knoll in the 1930s. TEM provides 2D projections of 3D specimens, with image contrast derived from the differential transmission of electrons through the sample.

Modern TEM systems are complex assemblies, with each component, from the electron source to the detector, having an influence on the final image quality. As the mean free path of an electron beam in air is only on the order of 100 nm (Kang et al.) the whole system interior has to be under high vacuum where the electrons have a mean free path in the km range. To understand the system better, we follow the path of an electron through this lens system (Fig. 13).

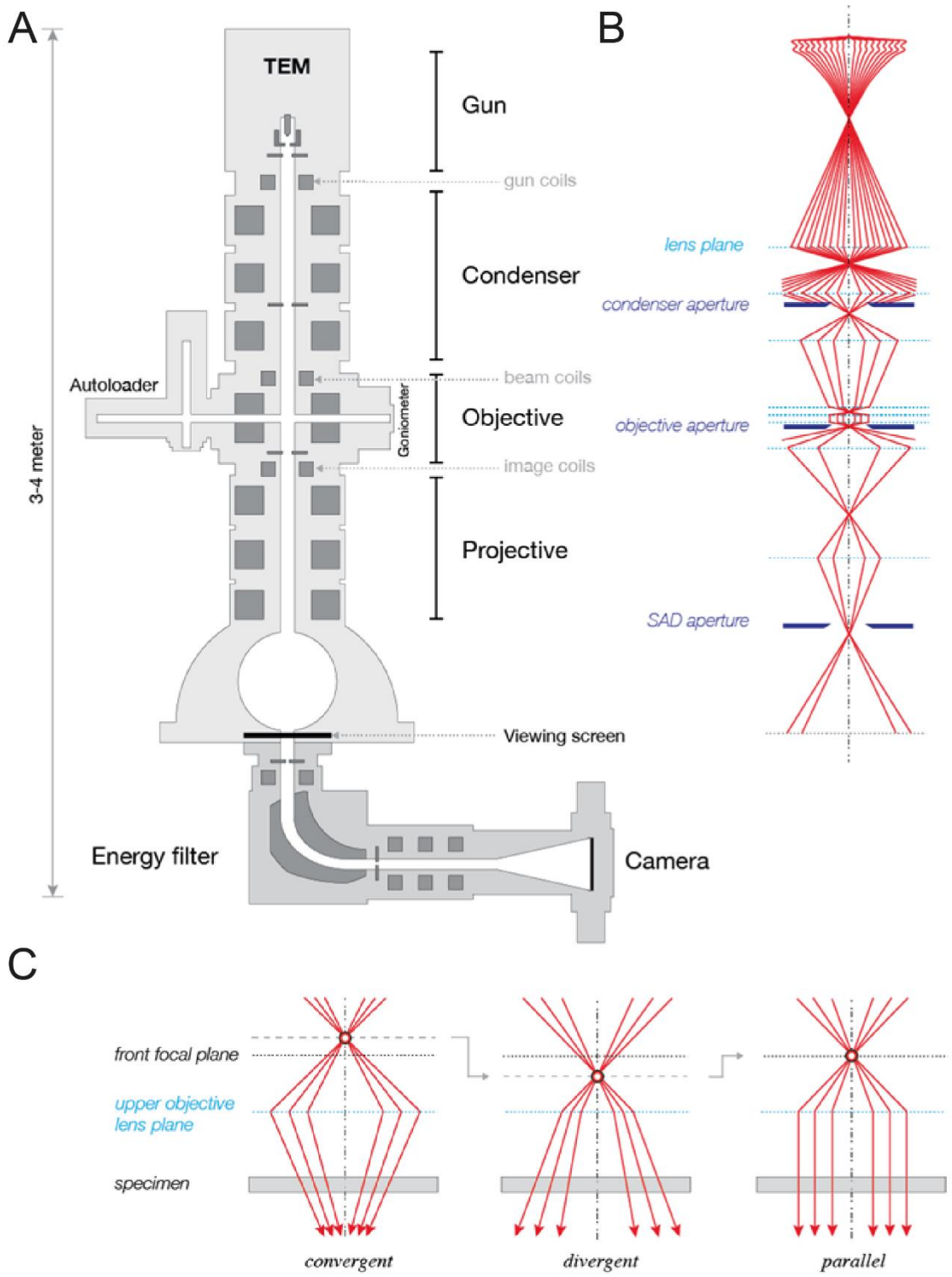

Figure 13: Schematical cross-section of a Transmission Electron Microscope (TEM).

(A) Design and positioning of magnetic lenses within the TEM. The electron source (“gun”) is on top. Positioned before the specimen, there is an adjustable condenser aperture that restricts the illuminating beam. Following the specimen, an objective aperture is placed, which filters out strongly diffracted electrons, thereby creating the scattering contrast. At the image plane, visualization can be done on a retractable fluorescent viewing screen (traditional approach). Situated beneath, there is either a high-sensitivity camera

(contemporary approach) for direct image capture, or, prior to reaching the camera, there's an energy filter designed to first eliminate inelastically scattered electrons. (B) Conceptual ray diagram detailing the path of the electron beam as it navigates through the primary lens systems of the TEM. (C) Schematic ray diagrams for convergent, divergent and parallel illumination. Adapted from (Plitzko and Bollschweiler 2024)

2.2.1 Illumination

The electron source is needs to generate a high-brightness electron beam with both high temporal and spatial coherence for high-resolution imaging. In modern TEMs, the preferred electron source is a field emission gun (FEG). This type of gun, equipped with a sharp tungsten tip, employs a strong electric field to facilitate the emission of high-energy electrons through quantum tunneling, even down to room temperature. This process ensures a bright, spatially and temporally coherent beam with a minimal energy spread.

Historically, thermic emission guns, which utilized either a bent tungsten filament or a LaB_6 crystal, were heated via an electric current to facilitate the release of electrons resulting in an inferior coherence and brightness compared to the (cold) FEG's.

After being emitted from the FEG, the electrons are accelerated by a series of anodes, ensuring a uniform and parallel electric field. This results in a parallel beam of accelerated electrons, which is crucial for the subsequent imaging processes in the TEM. The high voltage applied for acceleration necessitates a robust ultra-high vacuum to prevent contamination and potential short-circuits (Williams, Carter et al. 2009).

2.2.2 Lens systems, apertures and specimen stage

H. Busch showed in 1926 that axially symmetric electrostatic or magnetic fields could focus electron beams. This is still the foundational principle of how the beam is manipulated today. In modern TEMs, the electron beam is guided and shaped by a series of electromagnetic lenses and apertures, ensuring a defined illumination of the specimen area (Fig 14).

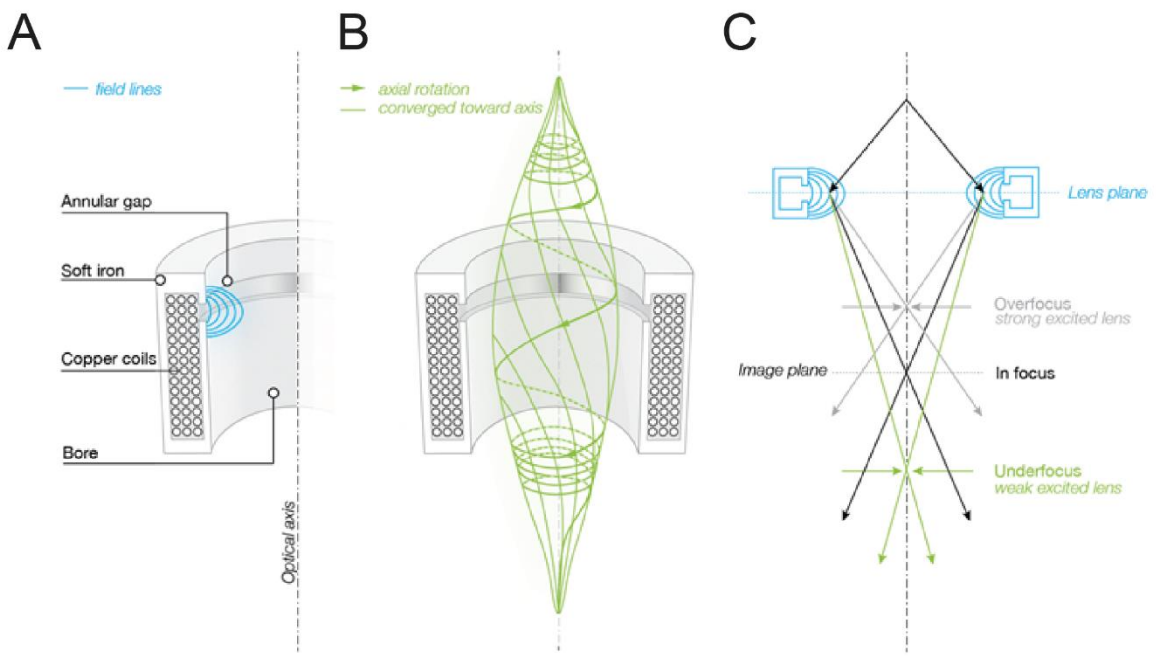

Figure 14: Overview of the Electromagnetic Lens.

(A) Schematic diagram highlighting the primary components of a rotationally symmetric electromagnetic lens. (B) Electron trajectory through the lens, showcasing both helical and converged paths. (C) Various lens settings, including in-focus, overfocus, and under focus configurations. Adapted from (Plitzko and Bollschweiler 2024)

TEMs typically employs two condenser lenses (C1 and C2) and their corresponding apertures to optimize illumination. Advanced systems, such as the Titan Krios FEG (Thermo Fisher Scientific), which was used in this study, incorporate a third condenser lens (C3) for parallel illumination. The specimen is mounted on a goniometer, a specialized stage, allowing precise translation and tilting for tomographic acquisition. For cryo-EM this sample stage has to be continuously cooled by liquid nitrogen.

After interacting with the specimen, the electron beam is focused by the objective lens in the back focal plane. Here, an objective aperture can filter out highly scattered electrons, enhancing the contrast. The beam is then spread by intermediate and projection lenses, resulting in the final magnification on the detector.

Each lens system is comprised of deflectors, the lenses, apertures, and sometimes a stigmator. These components ensure beam precision, with deflectors adjusting the beam's trajectory and stigmators correcting for astigmatism. Apertures, strategically placed metal discs, catch scattered electrons to reduce image background noise and enhance contrast in some cases. A further reduction in noise and improvement in contrast can be achieved by use of an energy filter.

2.2.3 Energy filter and detectors

The quality of the final image is significantly influenced by the energy filters and detectors used in the system.

Inelastically scattered electrons contribute to noise in the final image. To address this, post-column energy filters are employed to identify and filter out lower energy electrons, thereby enhancing amplitude contrast. The core component of this filter is a prism spectrometer, which deflects electrons of varying energies to different positions on the 'energy dispersive plane'. By placing a slit in this plane, electrons within a specific energy range can be selected, while others are blocked. In the 'zero-loss mode', electrons within a range of 5-20 eV around the acceleration energy (typically 300 kV) are chosen for detection, effectively filtering out most inelastically scattered electrons (Fig. 15) (Henkelman and Ottensmeyer 1974).

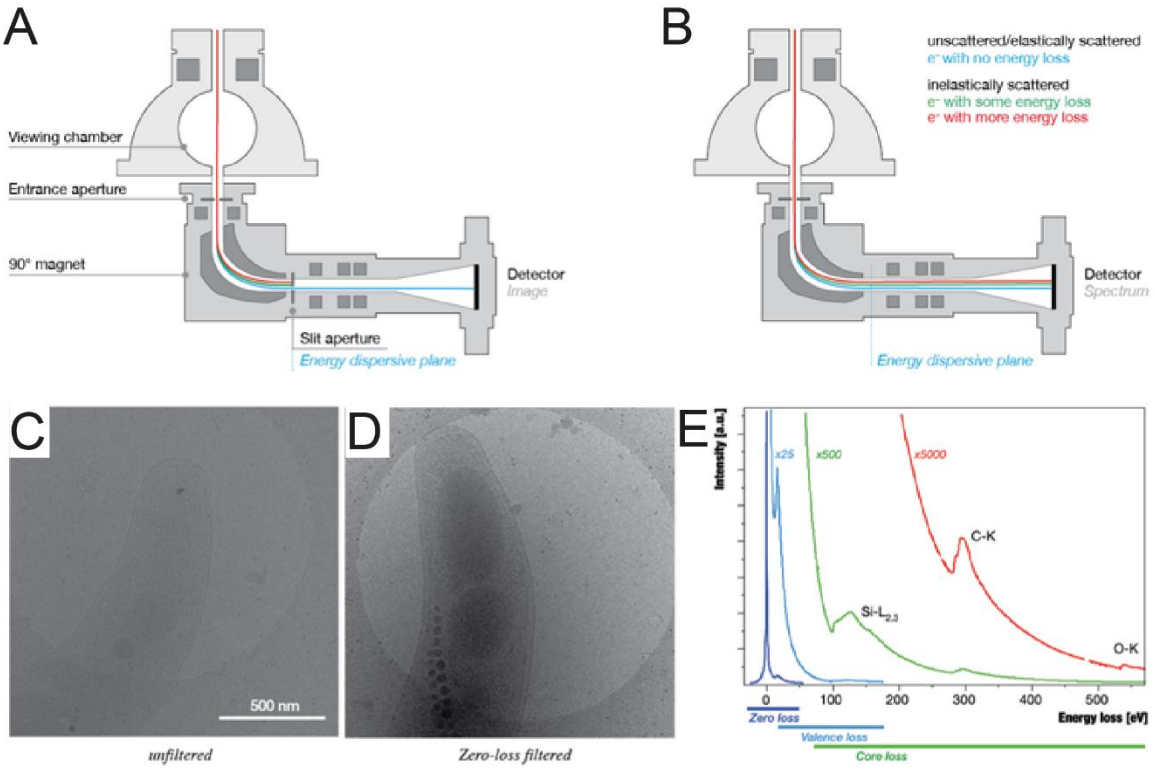

Figure 15: Energy Filter.

(A) Diagram illustrating the zero-loss imaging mode in comparison to a diagram representing the spectrum mode (B).
(C) Micrograph without filtering compared to a zero-loss filtered micrograph (D) showcasing magnetotactic bacteria (*Magnetospirillum gryphiswaldense*).
(E) Depiction of the Electron Energy Loss spectrum, highlighting different energy loss regions. Adapted from (Plitzko and Bollschweiler 2024)

Initially, photographic film was the primary medium for capturing images formed by the electron beam. Photographic film was replaced by charge-coupled device (CCD)

detectors. Even though CCDs provide inferior image quality, they offered a significant advantage over film, providing immediate access to images for data acquisition and analysis, eliminating the need for time-consuming film development and scanning. CCDs transform electrons into photons using a scintillator, and the photons are then guided to the CCD sensor. This process, while effective, introduces noise and reduces the detector quantum efficiency (DQE) (Fig. 16).

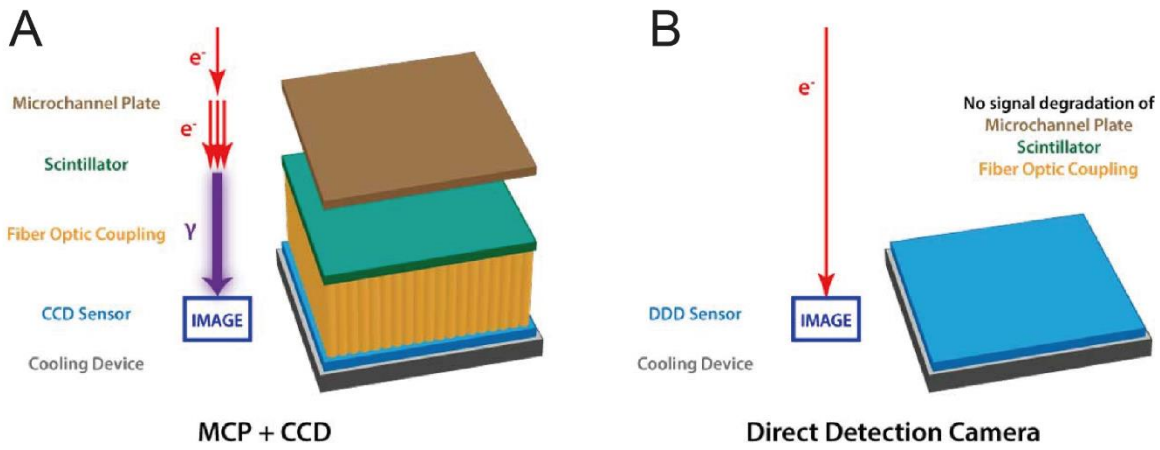

Figure 16: Comparing CCD and Direct Electron Detectors (DED).

(A) The CCD detector undergoes various intermediate processes from the time of the incident electron's arrival to its digitization on the chip, accompanied by a substantial supporting layer beneath. These aspects result in a diminished DQE and increased backscattering noise. (B) The back-thinned DED counters these challenges. The thinning process, although not visible since it is beneath the chip, contributes to a notably improvement in DQE and faster frame rates. Adapted from (Electron 2023).

The introduction of direct electron detectors (DED) marked a significant advancement in TEM imaging. DEDs, based on complementary metal-oxide-semiconductor (CMOS) technology, allow for the direct detection of single incident electrons in a thin detector. DEDs offer a higher DQE, a more localized signal, and a faster readout speed. This speed allows for the separation of images into frames and the subsequent computational correction of beam induced sample motion (Kühlbrandt 2014).

2.3 Image formation in a transmission electron microscope

After we discussed the makeup of a TEM let us take a look at how electrons interact with matter and how the image is formed.

2.3.1 Elastic and inelastic scattering

In electron microscopy, the nature of interactions between incident electrons and the specimen is fundamental for image formation. Two primary modes of interactions are observed: elastic and inelastic scattering (Fig. 17).

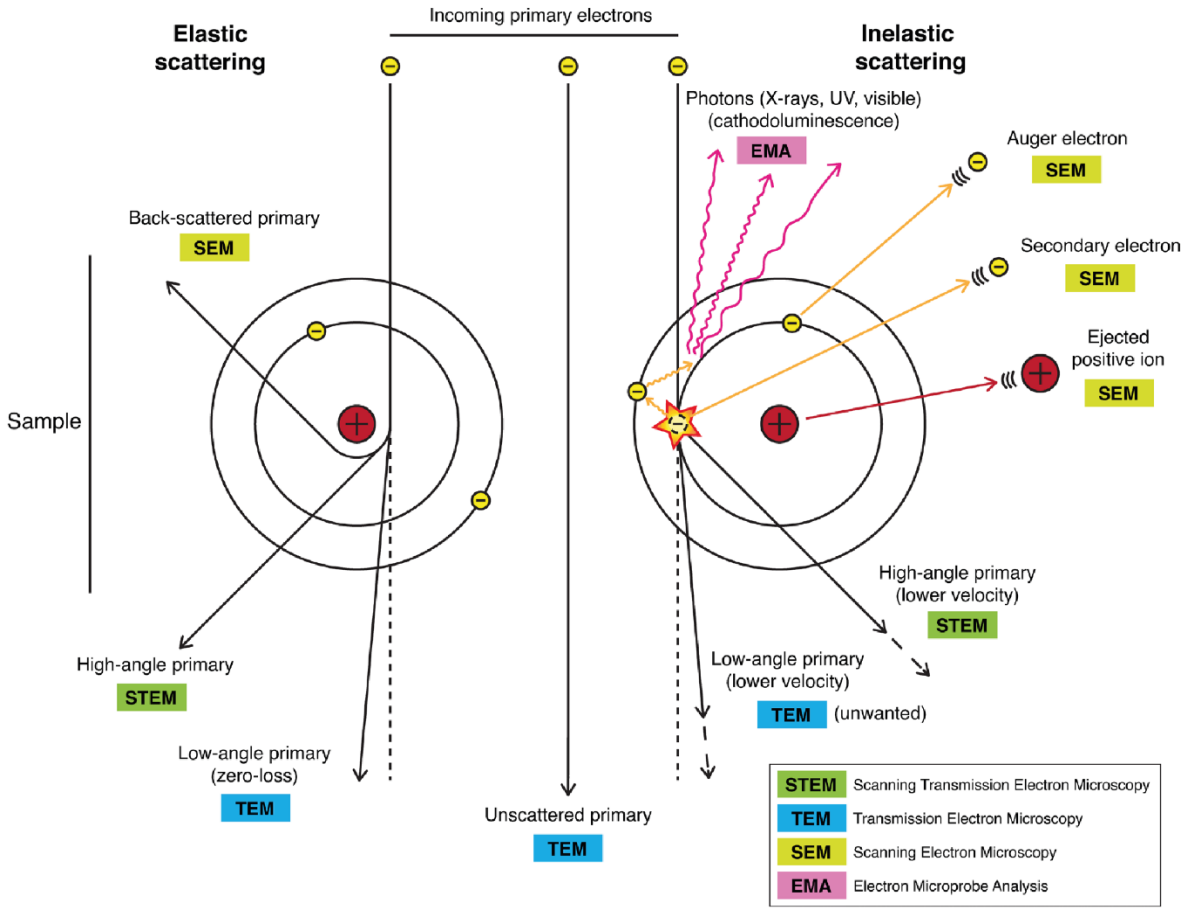

Figure 17: Comparing elastic and inelastic electron scattering mechanisms.

Typical electron-sample interactions observed in elastic (depicted on the left) and inelastic (shown on the right) scattering events. Furthermore, it highlights the various analytical techniques that exploit these respective interactions to study the samples. Adapted from (van den Hoek 2022)

In elastic scattering, incident electrons are deflected due to the Coulomb interactions with an atom without any consequential loss in their kinetic energy. This interaction, predominantly mediated by the atomic nucleus, results in a broad angular distribution of scattered electrons. The proximity of the incident electron's trajectory to the nucleus determines the magnitude of the scattering angle and the associated coherence.

Elastic scattering is instrumental in imparting contrast in biological cryo-electron microscopy (cryo-EM).

Conversely, inelastic scattering describes interactions where the incident electron transfers energy to the specimen, primarily via interactions with the atomic electron cloud. Such interactions can manifest in diverse phenomena, including X-ray emission, ejection of electrons from the valence band, and excitation of collective electron oscillations (plasmons) or lattice vibrations (phonons). Inelastically scattered electrons, due to their dispersed interaction volume and energy deposition, predominantly contribute to background or low-resolution signals. Given the potential deleterious effects of inelastic interactions on biological specimens, post-column energy filters are employed to selectively removed these electrons based on their energies divergent, thereby enhancing image fidelity in normal bright field TEM.

Furthermore, by imparting significant amounts of energy to the sample, inelastic scattering limits the amount of exposure for radiation sensitive samples.

The likelihood of a scattering event is determined by the cross-section, denoted as *s*. The average distance an electron travels between such events is represented by the mean free path length, *L*, which is the inverse of the cross-section: $L = \frac{1}{s}$. When considering both inelastic and elastic scattering, their combined cross-sections yield a total cross-section, s_{tot}. Given *N* atoms in the specimen, each with a cross-section s_{atom}, the total cross-section is expressed as:

$$s_{tot} = N * s_{atom}$$

Subsequently, the relationship between the mean free path length, the sample's thickness *t*, Avogadro's constant N_0, the atomic weight of the scattering atoms *A*, and the density *r* can be approximated to be:

$$L = \frac{A}{N_0 * s_{atom} * r * t}$$

This equation highlights that the mean free path length is inversely related to the thickness of the sample. This inverse relationship underscores the necessity for thin samples in transmission electron microscopy. For instance, under cryogenic conditions, the mean free path for inelastic scattering of 300 keV electrons has been observed to be ~320 nm (Martynowycz, Clabbers et al. 2021).

The propensity for elastic versus inelastic scattering is contingent on the atomic number of the constituent elements of the specimen. The ratio has been experimentally determined to be:

$$\frac{\sigma_{inel}}{\sigma_{el}} = \frac{20}{z}$$

With the atomic number *z* (Egerton).

In biological contexts, where elements with lower atomic numbers (e.g., carbon, nitrogen, oxygen) predominate, inelastic scattering events are more frequent. These inelastic scattering events, which limit exposure dose and thereby resolution in TEM, are useful for other techniques such as scanning electron microscopy (SEM) or electron spectroscopy. Scanning Transmission Electron Microscopy (STEM) is a specialized mode of electron microscopy wherein a focused electron beam is raster-scanned across the specimen, and transmitted electrons are collected to form an image. Unlike traditional TEM, where a broad beam illuminates the entire specimen, STEM's focused probe allows for localized analysis with high spatial resolution. One of the unique advantages of STEM is its ability to exploit inelastically scattered electrons. By employing electron energy loss spectroscopy (EELS) in conjunction with either TEM or STEM, it becomes feasible to analyze the energy loss of transmitted electrons, which is indicative of their inelastic interactions with the specimen. This energy loss provides valuable information about the local electronic structure, chemical composition, and even bonding states of the specimen at the nanoscale. Thus, in STEM, inelastically scattered electrons are not merely considered background noise but are used to extract detailed and localized spectroscopic information. Unfortunately, due to the high dose required, STEM is mostly used for fixed embedded samples and has so far only had limited applications for cryo preserved biological samples due to their high radiation sensitivity. With recent developments in instrumentation, especially detectors, there is a renewed interest in biological cryo-STEM.

Another technique that is extensively used in biological contexts now is Scanning Electron microscopy (SEM) as also used in the FIB-SEM instrument described in section 2.5.2.

In Scanning Electron Microscopy (SEM), the interaction of the primary electron beam with the specimen yields a plethora of signals, each revealing distinct facets of the sample. Backscattered electrons, elastically scattered from the specimen back in the beam direction, offer compositional insights due to their proportionality with atomic number. In contrast, secondary electrons (SE), ejected from surface atoms by the incident beam, enable high-resolution topographical images. Additionally, inner-shell electron ejections give rise to Auger electrons and characteristic X-rays, instrumental

in deducing surface elemental composition. Collectively, these diverse signals in SEM can provide a holistic view of the specimen's morphology, composition, and other intrinsic properties. For our purposes, in sample preparation mostly BSE are detected (Kohl and Reimer 2008).

2.4 The contrast transfer function

There are two predominant mechanisms of contrast, and therefore image formation in the cryo-TEM: amplitude and phase contrast. The fidelity of the images is contingent on the underlying contrast mechanisms.

In a TEM, amplitude contrast arises from two primary mechanisms. First, when elastic scattering takes place at wider angles, the deflected electrons are obstructed by the objective aperture. Second, inelastic scattering results in electrons being absorbed by the sample. Mathematically, the magnitude of contrast *C* is proportional to increasing sample thickness *t* and atomic number *Z* of the specimen:

$$C \propto tZ$$

In cryo-electron microscopy, amplitude contrast constitutes merely about ~10% of the total imaging contrast (Langmore and Smith 1992) but limits the dose that can be used to image a sample due to the high amount of energy deposited into the sample in the process.

The other, dominant, form of contrast formation is phase contrast due to inelastic scattering. The scattered waves undergo a phase shift, which, when superimposed with the unscattered waves, leads to interference. This interference is the foundation of phase contrast.

The scattered wave, $\Psi_{\text{scattered}}$, can be described as the incoming wave, Ψ_0, modified by a phase shift $\Phi(r)$, which is introduced by the scattering Coulomb potential *V* with the electron traveling along the z-axis:

$$\Psi\text{scattered} = \Psi_0 \, e^{i\Phi(r)}$$

For biological samples, which typically have a low atomic number, the scattered beam undergoes only a minor phase shift $\Phi \ll 1$. Given this weak phase shift, the scattered wave can be simplified using the Taylor expansion that is normally truncated at the second term:

$$\Psi(r) = \Psi_0(1 + i\Phi - \frac{1}{2}\Phi(r)^2 + \cdots)$$

This equation reveals that the new scattered wave is an incoherent sum of the incoming wave Ψ_0 and a wave $\Phi(r)$ with a small amplitude and a phase shift $\pm\frac{\pi}{2}$. The

intensity of the scattered wave is predominantly determined by the initial beam intensity I_0 and is nearly invisible in phase contrast.

To achieve noticeable contrast, additional phase differences are introduced by defocusing the objective lens and as a byproduct of the spherical aberrations of the instrument. For weak phase objects, the image distortions resulting from defocus and spherical aberration can be effectively described using a contrast transfer function (CTF) of the wave $\Psi_{scattered}$. The phase shift $\Phi(r)$ can be described in a first approximation as follows:

$$\phi(r) = \pi \Delta z \, r^2 \, \lambda + \frac{1}{2} \pi \, C_s \, \lambda^3 r^4$$

From this equation follows, that the phase shift introduced by defocus (Δz) can be used to compensate for the phase shift by the spherical aberration of the microscope C_s (Fig. 18 A-C).

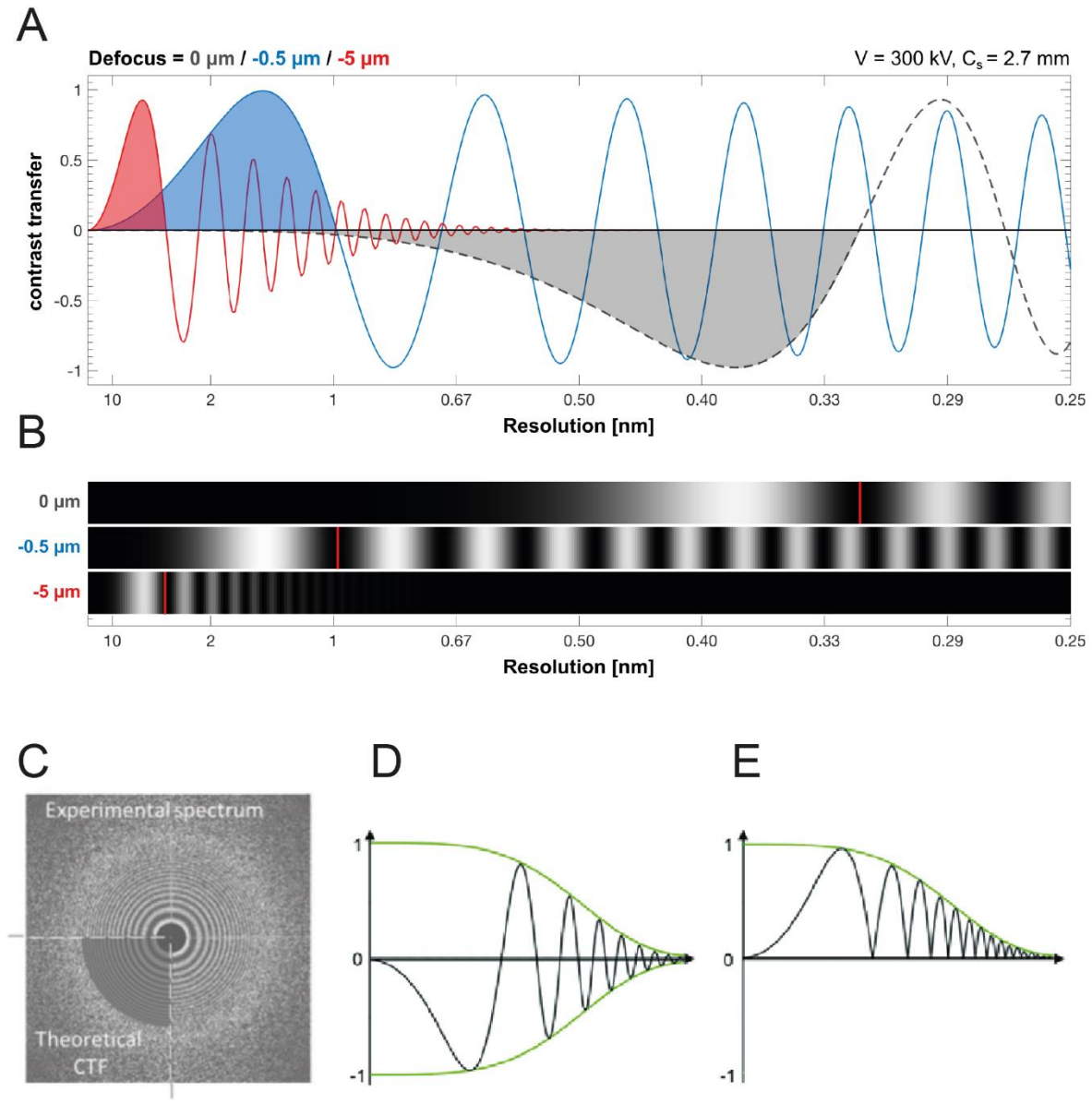

Figure 18: The role of the contrast transfer function in imaging.

(A) The Contrast Transfer Function (CTF) illustrates the way in which a microscope alters an actual image based on defocus and spatial frequency. Essentially, spatial frequency denotes the size of the object intended for imaging. Characterized by an oscillating sinusoidal pattern, the CTF efficiently transfers contrast at its peak value of 1. However, when the CTF dips to 0, no information is conveyed. If the CTF takes on a negative value, there is a contrast inversion. The point at which the CTF first hits zero defines the point-resolution. Beyond this point, contrast varies, leading to inconsistent information transfer. Defocusing can induce phase contrast. This figure delineates three defocus values: 0 µm (shown in grey), -0.50 µm (in blue), and -5.0 µm (in red). A defocus of 0 µm (grey zone) results in a significant loss of low frequencies, leaving mostly high frequencies, which produces a considerably blurred image. Elevating defocus amplifies contrast for lower frequencies (evident in the blue and red zones), but this comes at the cost of resolution. This is attributable to the enhanced attenuation of the CTF for elevated frequencies, which is clearly noticeable in the red curve (defocus -5.0 µm) where oscillations diminish rapidly. (B) Information transfer by the CTF. Here, the likelihood of information transfer is represented as an intensity gradient, where a spectrum from 0 to 1 translates to a color gradient from black to white in relation to spatial frequency. Frequencies shown in white or grey shades

are conveyed, whereas the ones in black signify lost information. The red demarcations indicate the point-resolution boundary that diminishes as defocus increases, showing values of 0.32 nm, 1.0 nm, and 3.1 nm for defocus readings of 0 µm, -0.50 µm, and -5.0 µm respectively. (C) 2D Correction Using the CTF. This figure provides an illustrative example of 2D-CTF correction. Panels (C) display plots of the radially averaged CTF as presented on the left. Once a close alignment is established with the CTF of the genuine tilt, the phases of the Thon rings (as depicted in (D)) are adjusted, resulting in a power spectrum that features solely positive contrast contributions, showcased in (E). Panels A-B were adapted from (Bäuerlein 2018) and panels C-E from (Costa, Ignatiou et al. 2017).

In reality, other higher order aberrations, such as second and third order astigmatism and axial coma also influence the CTF. All of them together result in coherent aberrations that can be computationally corrected for and are present even in an idealized beam (Plitzko and Bollschweiler 2024).

However, it is essential to account for incoherent aberration contributions, limiting the amount of information that can be transferred (green envelope in Fig 18 D, E). The spatial and temporal inconsistencies, stemming from the electron source, lead to a decline in the CTF as spatial frequency rises. Temporal coherence is influenced by chromatic aberration, which results from chromatic abirritations and the energy dispersion of the electron gun. The effects of spatial and temporal incoherence can be addressed by multiplying the CTF with dampening envelope functions resulting in a limit in the achievable resolution.

2.5 Sample preparation for high resolution TEM of biological samples

The TEM can be used to visualize a variety of samples across different size scale, from macromolecules in solution, 2D and 3D crystals, to cells and even tissue sections. In traditional TEM, biological samples undergo processes like chemical fixation, dehydration, and resin embedding, followed by staining with heavy metals to enhance contrast and make them vacuum stable. These procedures can introduce structural artifacts and are unsuitable for high-resolution imaging of biological samples. The vacuum environment of TEM could cause volatiles, such as water, from soft biological samples to evaporate, losing their native hydrated state. Cryo-electron microscopy offers a solution by preserving proteins and cells in their aqueous environment, ensuring that biological samples remain close to their native hydrated form. In the following we will first discuss different aspects of vitrification followed by how ticker biological samples such as cells are thinned for TEM.

2.5.1 Vitrification, plunge freezing and other aspects of sample preparation

Vitrification is a cornerstone technique in cryo-electron microscopy, preserving biological samples in a near-native state by rapidly freezing them. This process transforms liquid water into vitreous ice, a glass-like solid instead of ice crystals, which would damage the sample in their formation. Cooling rates between 10^4 to 10^5 K/s are needed to immobilize water molecules quickly enough. Typically, a mixture of liquid ethane and propane, cooled to just above their freezing points, serves as the cryogen for vitrification by plunge freezing.

For single particle analysis (SPA), the vitrification process begins by placing a small volume of a purified sample onto an EM grid covered with a hydrophilic holey film support, typically treated by glow discharging or plasma cleaning to ensure hydrophilicity. Excess solution is then removed by blotting it away with filter paper. The EM grid is then swiftly plunged into liquid ethane or an ethane-propane mixture (Dubochet, Adrian et al. 1988, Dubochet 2012).

Achieving optimal vitrification is difficult even for thin objects like solutions of purified protein. Factors such as sample stability, blotting time, and grid substrate can influence the quality and reproducibility of the vitrified specimen. The goal for SPA (see section 2.6.1) is to produce grids where macromolecules are randomly oriented and embedded in the thinnest possible layer of vitreous ice. Given that biological specimens scatter electrons minimally more than the surrounding buffer, ice thickness is intrinsically linked to image contrast and signal-to-noise ratio (SNR). This is especially crucial for SPA studies of smaller complexes, which have fewer scattering atoms contributing to the signal.

A significant challenge in specimen preparation for proteins in solution is the specimen's adsorption to the hydrophobic air-water-interface (AWI). This can lead to sample denaturation or a preferential orientation relative to the AWI (Fig. 19). While tilted collection strategies can address the latter, completely preventing sample adsorption to the AWI remains a challenge(Glaeser 2018, Noble, Wei et al. 2018). Recent strategies to counteract AWI effects include using surfactants, affinity grids, or monolayer supports like graphene. Automated vitrification robots, such as Spotiton (Dandey, Wei et al. 2018) and VitroJet (Weissenberger, Nijpels et al. 2020), offer potential solutions by reducing the time between sample dispensing and plunging, potentially mitigating AWI interactions.

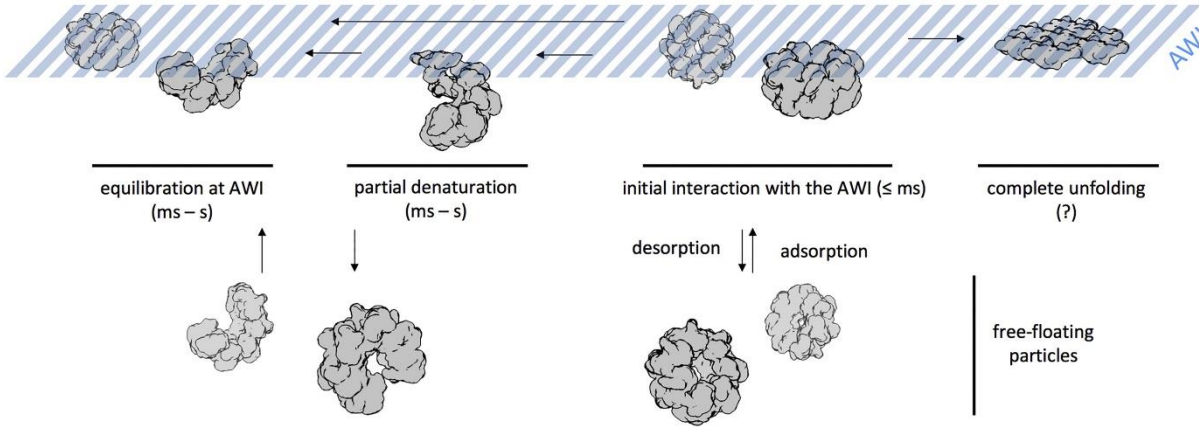

Figure 19: The air water interface.

The initial contact between the protein and the AWI (adsorption) occurs quickly. However, the system reaching a state of equilibrium takes longer due to processes like desorption, partial or complete denaturation. These processes, happening over different timescales (from milliseconds to seconds), are believed to be significantly influenced by the specific protein involved (e.g., the desorption from the AWI, and the rate at which unfolding occurs). (Klebl, Gravett et al. 2020)

Vitrification also offers significant advantages over fixation for cellular specimens. Cells are either cultivated directly on EM grids immersed in their culture medium or, for non-adherent cells, a suspension is administered to the grid just prior to plunging. Excess medium is blotted away to optimize sample thickness before the grid undergoes a plunge into a cryogen. This rapid freezing technique effectively halts all cellular activities, encapsulating cells in their native, aqueous environment. The result is a fully hydrated, unstained cellular sample, suitable for high-resolution imaging of cellular structures and functions (Mahamid, Pfeffer et al. 2016).

However, plunge freezing is often insufficient for complete vitrification when samples are too thick, such as cells or tissue. The depth of vitrification, for instance, is typically limited to about several micrometers due to the cooling rate achievable near the center of the sample. While the addition of cryo protectants such as glycerol can alleviate some of the limitations, they introduce problems such as changes in osmolarity and have to be optimized for every single cell type and to a lesser extent tailored around the biological process of interest (Bäuerlein, Renner et al. 2021).

High-pressure freezing can extend the maximum vitrification depth to around 100 µm but comes with its own set of challenges. After high-pressure freezing, the sample is embedded into up to ~100 µm of ice and has to be specifically targeted by a combination of volume imaging and cryo light microscopy or precut using a cryo

ultramicrotome. The removal of excess sample by cryo-FIB milling (see chapter 2.5.2) for thick samples has been demonstrated but remains low throughput and challenging (Kelley, Raczkowski et al. 2022).

2.5.2 Cryo Focused Ion Beam milling in a Scanning electron microscope

For high-resolution imaging using 300 kV TEMs, the sample thickness should ideally be in the range of 100-250 nm as discussed in 2.3.1. For the imaging of most cellular samples, which are not extremely thin by nature, like Mycoplasma or certain parts of bigger cells like the ciliary tail, the thickness of the sample is the limiting factor. Given that eukaryotic cells are significantly thicker, most of the cell's volume remains inaccessible by cellular cryo-EM. Therefore, the key to overcoming this limitation lies in preparing sufficiently thin samples (Rigort, Bauerlein et al. 2012).

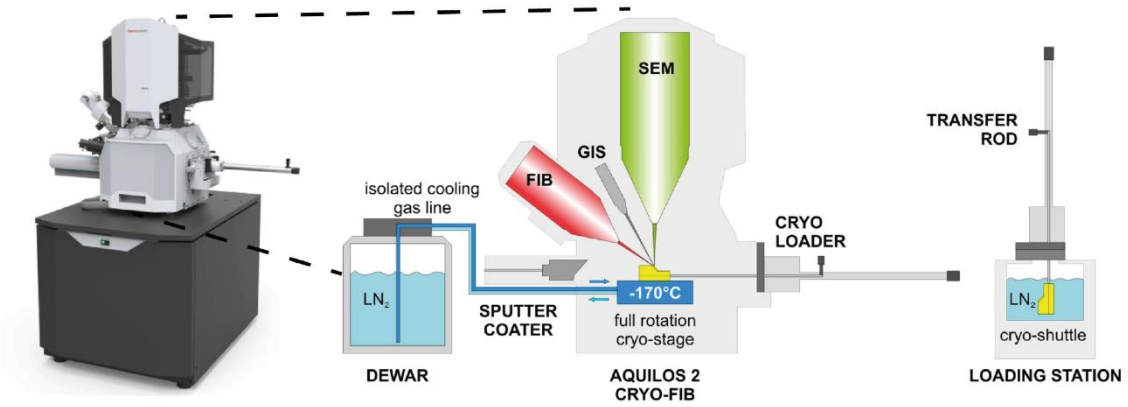

Figure 20: Cryogenic configuration of a FIB-SEM microscope.

Cryogenic assembly tailored for a Focused Ion Beam (FIB) microscope. In an external loading station, EM “autogrids” in a vitrified state are secured onto a cryo-shuttle. The setup is cooled using liquid nitrogen. This shuttle is then fetched by a specialized transfer tool equipped with a shutter. Before its insertion into the high chamber vacuum of the dual-beam microscope, the transfer rod is initially evacuated by a standard roughing pump. Upon equilibrating the pressure, the shuttle is moved to a cryostage within the FIB setup, cooled cold via nitrogen gas. External cooling to the cryostage, achieving temperatures near −170 °C, is facilitated by a dedicated liquid-nitrogen dewar. The ion beam is used for the “milling” of the specimen in its frozen-hydrated state. Adapted from (Rigort, Villa et al. 2012)

Historically, the approach to address the thickness problem was cryo-ultramicrotomy, which involved creating 50-200 nm thin sections of the sample by cutting it with a diamante knife. This technique, known as "cryo-EM of vitreous sections" (CEMOVIS), however, introduced significant artifacts due to mechanical cutting forces, such as compression, scatter, and knife marks. These artifacts distorted the cellular structures in the resulting sections (Al-Amoudi, Chang et al. 2004).

A solution to this problem came with the introduction of cryo-focused ion beam (cryo-FIB) technology (Fig. 20). This non-mechanical approach uses a focused ion beam to prepare thin vitreous sections. Initially demonstrated in prokaryotes, the cryo-FIB technology was later developed for single cells on EM-grids, becoming a routine preparation technique for *in situ* studies in both eukaryotes and prokaryotes. The FIB/SEM Dual-Beam Microscope, integral to this technology, is equipped with a SEM, a FIB, a cryo-transfer system, a cryo-stage, and a nitrogen-gas cooling system. While the SEM is used for targeting and intermittent imaging to assess the progress, the FIB is used to precisely ablate specific parts of the cell, typically using gallium ions, resulting in an electron-transparent lamella containing the structure of interest. This lamella, often referred to as a "window into the cell," provides a thin slice of the cell for detailed imaging (Bäuerlein and Baumeister 2021).

Even though the FIB milling process is becoming more accessible for non-expert users through better automatization and UI, it remains relatively low throughput. Before ablation can start, a protective layer, such as organic platinum or sputter coated platinum, is applied to the sample's surface to reduce erosion damage from stray gallium ions. The milling is conducted in stages with consecutively lower acceleration currents of the ions, starting with rough milling to remove bulk material, followed by fine milling, and finally polishing to achieve the desired lamella thickness. The resulting lamella should ideally be between around 100 nm in thickness as recently described in (Lucas and Grigorieff 2023, Tuijtel, Kreysing et al. 2023).

Recent advances in sample preparation, such as micropatterning, automation, as described in (Toro-Nahuelpan, Zagoriy et al. 2020) and (Klumpe, Fung et al. 2021), and in instrumentation have improved throughput to a point where, under ideal circumstances, up to 50 lamellas can be prepared in a day with minimal user intervention. With FIB milling becoming more routine and data acquisition being largely automated, as discussed in the following chapter, a wider range of projects is becoming feasible.

Improvements in the experimental protocol of cryo-lift out have sped up the process and made it more reproducible, throughput still remains an order of magnitude below the one achievable for isolated cells (Dung, Perone et al. 2023, Schiøtz, Kaiser et al. 2023). Furthermore, acceptable success rates in sample preparation require extensive training of the operating personal due to the technical difficulties. Overall, this remains an active arear of research with potential for further improvements.

2.6 Imaging modalities: tomography and single particle analysis

For the most part, imaging of biological samples in a cryo-TEM can still be roughly divided into two main imaging modalities: Single particles analysis (SPA) and cryo-ET. SPA describes the processes first imaging a purified protein or complex and then derive its structure, while cryo-ET aims to go one step closer to the native biological system by imaging bigger intact assemblies up to cells or even cellular tissue.

Even though both aim to elucidate the 3D structure of a biological object the differ at the stage where 2D information is combined into a 3D image. In SPA we rely on the fact that we have the same object embedded in vitrified ice, ideally in random orientations, and can therefore image just at one angle and later reconstruct the 3D structure out of these 2D images by combining them.

In cryo-ET on the other hand, we take multiple images of the same sample at different angles and then reconstruct a 3D volume of our sample. If the goal is to solve a high-resolution structure of a protein of interest in this volume, one has to localize and average hundred to hundreds of thousands of copies of the same molecule due to the low SNR in the initial images. This process is called subtomogram averaging (STA) (Wan and Briggs 2016). In the following section, we first discuss data acquisition and image processing for both techniques and compare their advantages and disadvantages.

2.6.1 Data acquisition for single particle analysis

Initially the atomic structure of proteins could only be solved by x-ray crystallography. Nowadays however, thanks to the technical advances described in the chapter before, the number of structures solved by SPA has seen an exponential growth over the past decade (Fig. 21). This has been facilitated by a convergence of two trends in the realm of data acquisition, as well as data processing (discussed in the next chapter).

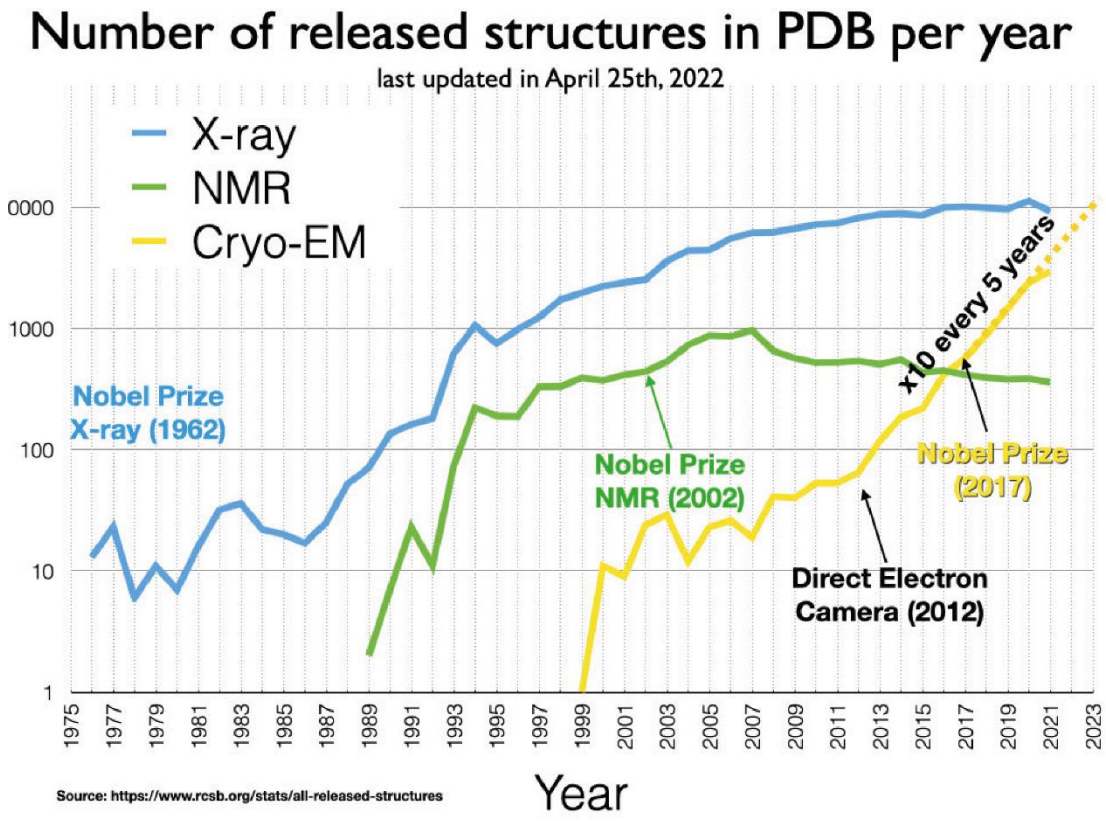

Figure 21: Number of structures deposited to the PDB per year split by technique.

The number of published high-resolution structures has seen a steady increase over the past decades. The blue line delineates the number of structures deposited to the PDB solved by X-ray crystallography, the green line those solved by nuclear magnetic resonance (NMR) and the yellow line the ones solved by cryo-EM. Important events such as Nobel Prizes are indicated. Adapted from: (Masahide 2022)

To obtain a high-resolution structure of a protein one needs many images of this particle. This relationship is captured by a ResLog plot (Stagg, Noble et al. 2014) (Fig. 22).

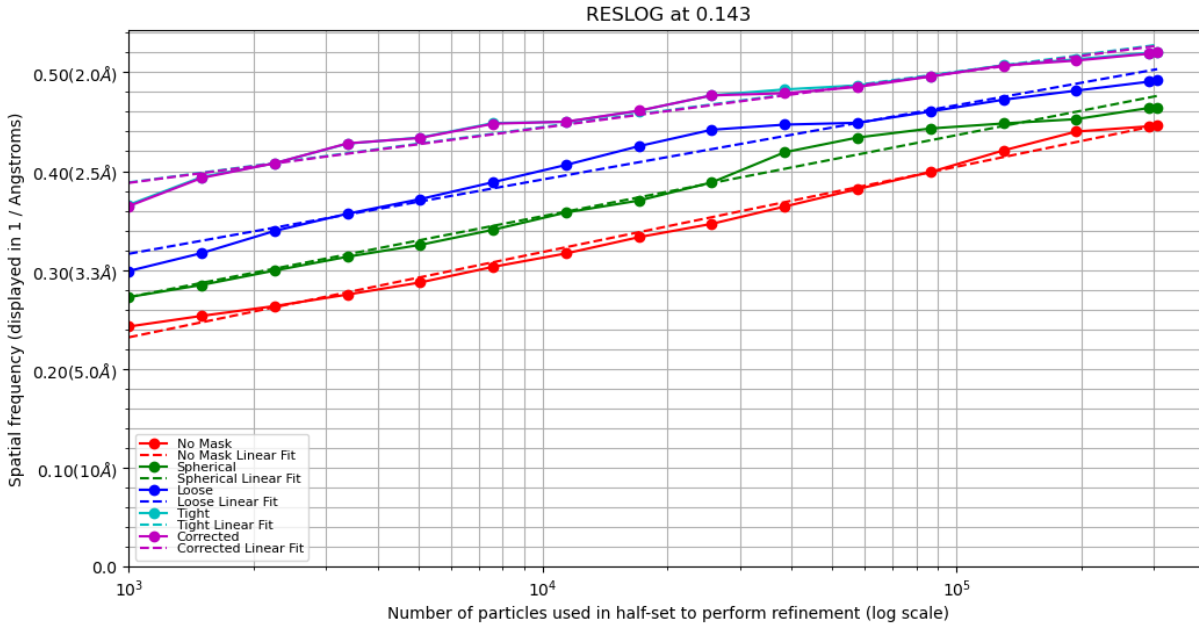

Figure 22: ResLog Plot.

Example of a ResLog plot characterizing the relationship between the number of particles and the achievable resolution. Colored dots correspond to the experimentally calculated resolution estimations with n particles per half set, and the dotted lines to the linear fit of all values. The colors correspond to different masking regimes: Red: no mask, Green: spherical mask, Blue: loose mask, Cyan: tight masking, Purple: tight masking corrected.

As both the speed of acquisition as well as the data quality have been improving rapidly, SPA emerged as the premiere technique to solve structures of biological complexes.

Initially, the datasets in cryo-EM comprised thousands of electron micrographs, often accumulated over several days. These micrographs yield a vast number of particle images, sometimes reaching into the millions (Danev, Belousoff et al. 2021).

The past decades have seen a shift towards automation in cryo-EM SPA for both image acquisition and processing. Software packages such as Leginon, SerialEM, and EPU have been developed to interface with TEMs, facilitating automated navigation and targeted exposures. The efficiency of image acquisition has been further augmented through strategies like beam-image shift, which eliminates the need for mechanical stage movements that disturb sample stability, thus reducing settling times until an equilibrium has been reached. Historically, this method was sidestepped for high-resolution imaging due to concerns about off-axis coma aberrations and phase shift errors from beam tilt. However, recent advancements have shown that these aberrations can be minimized and even corrected *in silico* (Zivanov, Nakane et al. 2018, Zivanov, Nakane et al. 2020, Zivanov, Oton et al. 2022). The number of images that can be taken without stage movement has been recently further increased by the

introduction of fringe free imaging (Weis and Hagen 2020), allowing for the placement of more exposures per hole in the holey carbon support film. As a point of reference, we reached around 1000 images per hour for latest dataset we acquired in 2023.

Microscope alignments during data collection can be time-consuming, and there is still a reliance on user expertise to pinpoint optimal grid areas for imaging. The future likely holds the integration of machine-learning algorithms to automate acquisition targeting and real-time image analysis, as recently demonstrated by CryoSPARC live (Punjani 2021) and further enhancing data collection speed and quality. As we look ahead, the interplay between hardware and software will continue to shape the trajectory of image acquisition in cryo-EM, with each leap in one domain potentially revealing bottlenecks in the other.

2.6.2 Data processing in single particle analysis

Alongside the improvements in imaging, changes in data processing have facilitated an unprecedented ease of use alongside the ability to resolve atomic resolution structures of biological macromolecules within days.

Data processing in SPA is a multi-step procedure that transforms raw micrographs into high-resolution 3D structures. In the following, a generalized workflow is described and recent advances in software are highlighted. This workflow can be significantly more complicated for more challenging, heterogenous samples but should serve to give a quick general overview (Fig. 23).

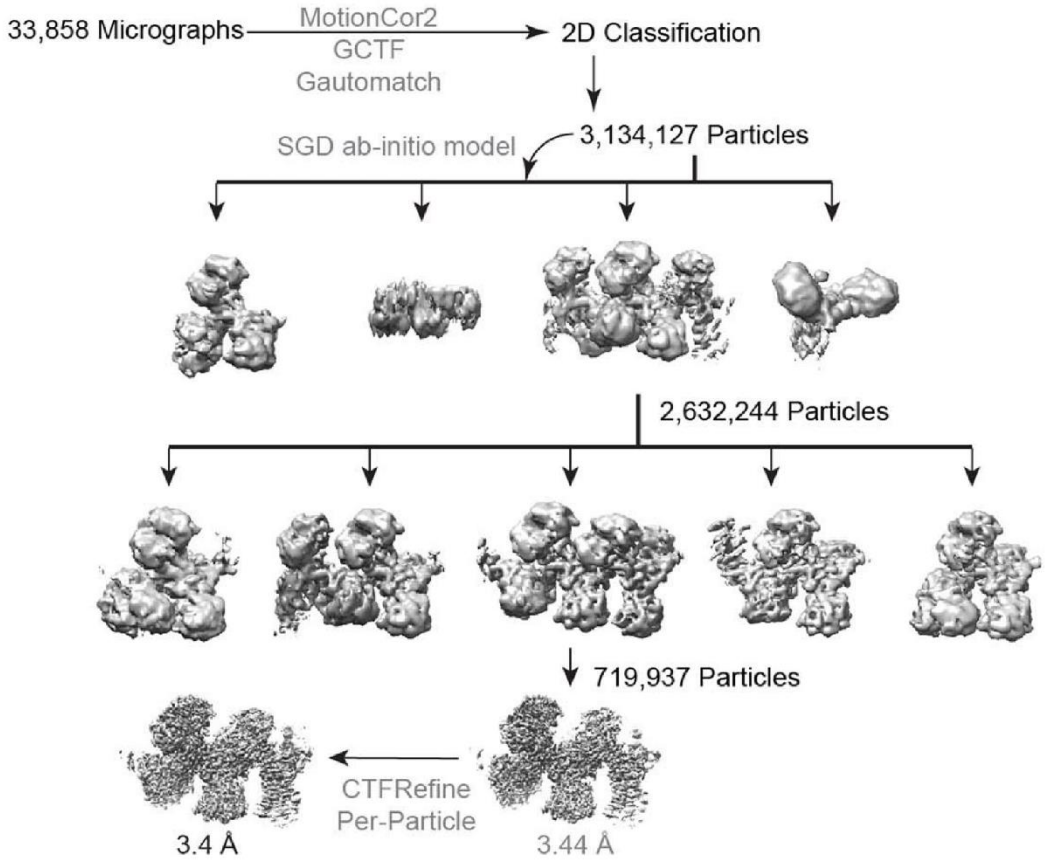

Figure 23: Example case of a single particle workflow.

Illustration shows an exemplary workflow for a single particle project on the case of HDCR. Figure is adapted from (Dietrich, Righetto et al. 2022)

The workflow begins with preprocessing, where the raw micrographs undergo motion correction to account for drift during image acquisition (Zheng, Palovcak et al. 2017), and CTF estimation and correction, to correct for the imaging aberrations (Rohou and Grigorieff 2015, Zhang 2016). Both these steps are now routine and have been shifted more and more to an “on the fly” process (Suloway, Pulokas et al. 2005, Biyani, Righetto et al. 2017, Cheng, Negro et al. 2021).

Following preprocessing, particle picking is executed, either manually or using automated algorithms, to identify and extract individual particle projections from the corrected micrographs. To date a variety of machine learning based software packages, such as CryoYOLO (Wagner, Merino et al. 2019), are able to pick particles with high accuracy for standard samples with minimal user input.

These extracted particles are then subjected to 2D classification, grouping similar projections together, which aids in removing outlier particles and obtaining averaged 2D class averages (Scheres 2016).

With these class averages, *ab initio* reconstruction generates an initial 3D model without the need for prior structural information (Punjani, Rubinstein et al. 2017, Reboul, Eager et al. 2018). Alternatively for challenging samples it is possible to use prior models such as the ones generated by AlphaFold2 *in silico*, as a starting point. Recently it has be demonstrated in CryoSPARC that both steps can be accomplished in a real time, on the fly manner, allowing for continuous monitoring of data acquisition by the user(Punjani 2021).

After a first round of angular and positional refinements, the model can serve as a reference for 3D classification, where particles are grouped based on their 3D conformations, allowing for the separation of distinct structural states and compositional heterogeneity. Here the recent introduction of machine learning based methods such as cryoDRGN (Zhong, Bepler et al. 2021) as well as 3DVA (Punjani and Fleet 2021) and 3D Flex (Punjani and Fleet 2022) within the cryoSPARC software package allow for the modelling of continuous motion of the individual molecules as well as the separation of compositional heterogeneity.

After further rounds of 3D refinement, per particle aberration and motion corrections are applied to account for any residual microscope aberrations and beam induced motion, ensuring the final 3D structure is of the highest possible resolution and accuracy, enabling de novo model building. The gold standard for this still remains Bayesian Polishing and CTF refinement within the RELION software package (Zivanov, Nakane et al. 2019).

The continuous innovation in software allows the elucidation of more and more challenging macromolecular complexes. This is further aided by the recent introduction of AlphaFold2 (Jumper, Evans et al. 2021), enabling the modelling of intermediate resolution structural elements (4-7 Å) based on their genetic sequence within bigger complexes.

2.6.3 Introducing 3D context: Tomography

Tomography aims to provide a three-dimensional perspective of non-repetitive structures by capturing a series of two-dimensional projection images, termed a tilt series, at varying angles (Fig. 24).

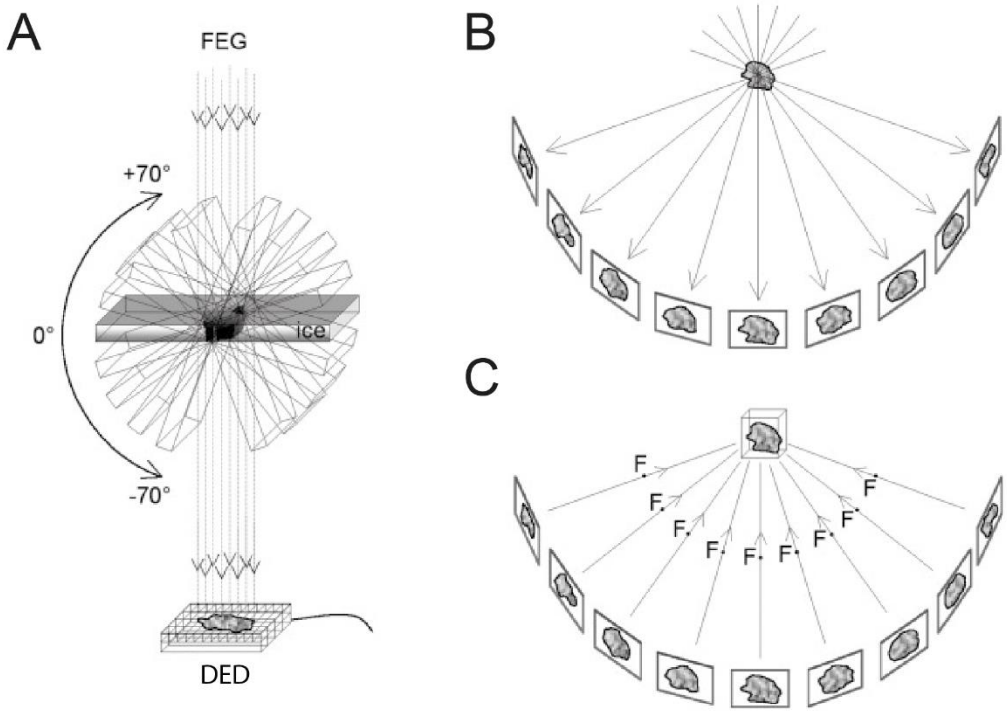

Figure 24: Overview of Electron Tomography.

(A) A specimen is secured within an electron microscope and tilted at small angles between up to -70° and +70° relative to the beam of electrons. For each tilt angle, a 2D image is captured using a DED. (B) These images, each representing a different viewing angle of the specimen, are then precisely aligned with one another. Finally, through a computational process known as 'weighted back-projection,' these aligned images are fused to create a 3D density map, or tomogram, which represents the internal composition of the specimen (C). (Grunewald, Medalia et al. 2003)

The foundational principle, the Radon transformation, suggests that a 3D model or tomogram can be reconstructed from its projections. Each projection represents a slice through the Fourier transform of the volume, and the tomogram is reconstructed by populating the 3D Fourier space with these slices.

To obtain a tomogram one has to obtain projections of the sample under a variety of angles. In cryo-ET, the sample stage is tilted incrementally and the same region is imaged at various stage angles. Typically, due to mechanical constraints, the tilt range is confined to about a maximum of ±70°, resulting in a "missing wedge" of information. This missing data leads to distortions and elongations of structures along the beam axis.

TEMs equipped with a dual-axis stage can record tilt series in two orthogonal directions, reducing the missing information to a pyramid, thus yielding a more isotropic resolution (Mastronarde 1997). This approach has largely been abandoned to date due to the increased difficulty of aligning the tilt series, the prolonged process of data

acquisition and reduced dose that can be put into each tilt, leading to lower SNR values for every tilt, further reducing alignment quality. Furthermore, from an instrumentation perspective, a single axis tilt stage has a superior tilting stability, further increasing the difference in acquisition speed.

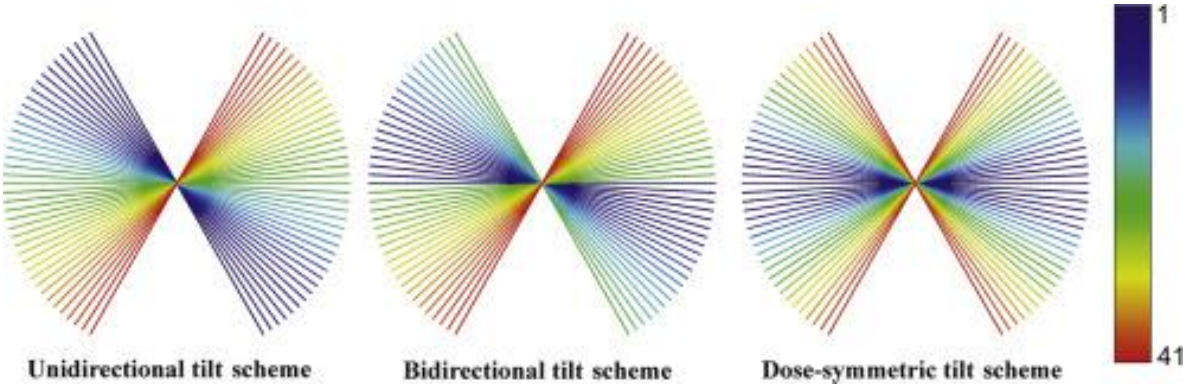

Figure 25: Tilt schemes.

Illustration of three methods of collecting tilt series in electron tomography: unidirectional, bidirectional, and dual-walkup schemes. The tilts range typically from -60 to +60 degrees, increasing in 3-degree intervals, totaling 41 tilts. Shades or colors indicate the sequence in which each tilt is collected, as denoted by the color key. In cases where each tilt is exposed to the electron beam for the same duration, the sequence of tilt acquisition affects the cumulative electron dose for the images. The unidirectional approach collects tilts in a linear fashion from one extreme angle to the other. The bidirectional method starts in the middle, goes to one extreme, then reverses to cover the other extreme, showing a break in sequence at the switch point. The dose-symmetric approach collects tilts from the middle towards both extremes, resulting in a more evenly distributed electron dose across the series. (Wan and Briggs 2016)

For each tilt in the series, four steps are consistently repeated: tilting, centering, focusing, and recording. The centering of the area of interest (x-y plane) and the focusing (z direction), are necessary to compensate for imperfections in the sample stage movement. Various tilt schemes, mono-, bidirectional and dose symmetric, are employed to distribute the electron dose optimally (Fig. 25). The dose-symmetric tilt scheme, which alternates tilt directions starting at 0° and progressing to the maximum tilt, is now widely adopted. Initiating data collection at minimal angles, where electrons traverse the shortest path through the specimen, primarily serves to preserve high-resolution details from radiation damage. Furthermore, distributing the electron dose in a nearly symmetrical manner curtails the alignment discrepancies commonly associated with bidirectional tilt strategies (Hagen, Wan et al. 2017).

Given the radiation sensitivity of frozen-hydrated biological specimens, electron beam exposure must be minimized. The optimal exposure hinges on the radiation sensitivity of the sample and the desired resolution. Distributing the total permissible electron

dose across a tilt series results in individual images with low SNR. This SNR is further diminished at higher tilt angles due to increased electron path lengths through the sample and the resulting multiple electron scattering events. The optimal approach for tilting and dose distribution is still an active field of research promising to further improve the data quality and speed of acquisition for cryo-ET.

Another detrimental factor for data quality in cryo-ET is the charging of the biological sample during acquisition. This problem seems to be more detrimental for bigger samples with more isolating bulk material remaining around the FIB milled lamella in the case of cellular tomography. Recent work has shown that this can be compensated for by sputter coating the final lamella with a thin film of inorganic platinum (Khavnekar, Vrbovská et al. 2022).

With the recent introduction of data acquisition strategies similar to multi shot exposure in SPA in the Tomo5 (TFS) software package and in the form of PACETomo (Eisenstein, Yanagisawa et al. 2023) for SerialEM (Mastronarde 2005), the acquisition time for a single tomogram has been significantly reduced. It is now possible to collect up to over 100 tilt series in one day, nearly an order of magnitude up from previous acquisition speeds. This will further facilitate a push for lower pixel size and higher resolution data in cryo-ET.

2.6.4 Analysis of cryo electron tomograms of cellular samples

To date, with the recent improvements in data acquisition, a majority of time, after the optimization of the biological sample, is now spent in data processing and analysis. An exemplary processing pipeline is outlined below that roughly follows the contours of the workflow used in this study (for details see Material and Methods).

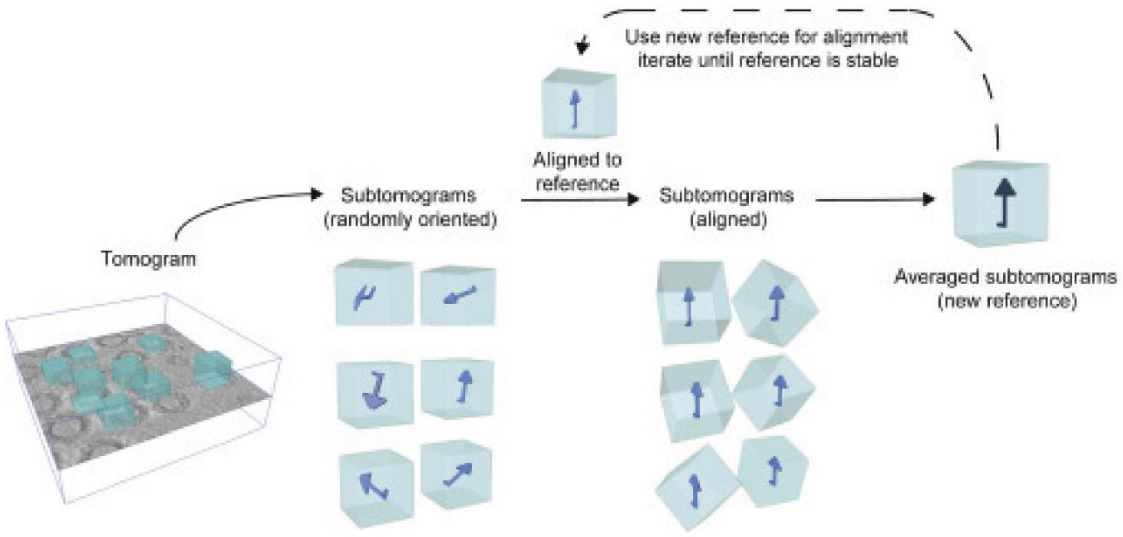

Figure 26: Outline of subtomogram averaging workflow used to analyze molecules within a tomogram.

Initially, subtomograms, which are essentially small cube-shaped sections, are cut out from the larger tomogram. Each of these contains one instance of the target molecule, oriented in a random direction. These subtomograms are then aligned to a reference model, and from this aligned set, a new, improved reference is created. The procedure is repeated—aligning subtomograms to this updated reference and then creating a new reference from them—until the references no longer change significantly with each iteration, indicating that the alignment converged. Adapted from (Wan and Briggs 2016)

Recent improvements will be highlighted and comments on the remaining bottlenecks made. This is by no means meant to be an exhaustive discussion of all possibilities for the analysis of cryo-ET data, but should serve as a primer about the underlying fundamentals and techniques used for this project.

2.6.4.1 Tomogram reconstruction and preprocessing

Tomogram reconstruction and preprocessing are essential steps in cryo-ET. The process begins with frame motion correction, analogous to what has been described for SPA.

As the tomogram is acquired, the sample accumulates electron dose, leading to radiation damage, especially during the later stages of the tilt scheme. Early tilts, with their low accumulated dose, retain high-resolution information, while later tilts do not. To address this, exposure dose filtering, such as with STOPGAPs implementation of the Grant & Grigorieff weighing scheme (Grant and Grigorieff 2015), apply a down-weighting factor on higher-resolution information based on the accumulated dose of each tilt.

Magnification is another pivotal factor in tomogram recording. While higher magnifications can theoretically achieve superior resolutions as per the Nyquist–Shannon sampling theorem, they also necessitate a proportional increase in beam intensity to maintain consistent electrons per pixel. Consequently, the electron exposure, measured in electrons per $Å^2$, rises quadratically with magnification. Striking a balance between optimal sampling and limiting cumulative exposure is crucial when selecting recording parameters and will depend on the individual project. If the project in question requires sub tomogram averring to increase the SNR of the target object, a further consideration is the quadratic reduction in field of view. If at a pixel size of 4 $Å^2$ one would need 100 tomograms to reach sufficient numbers for analysis and statistical testing downstream, a pixel size 2 $Å^2$ would necessitate the acquisition of 400 tomograms. This is to a certain extent alleviated by the better data quality at lower pixel sizes due to higher DQE, but still remains a consideration.

Recent work achieving resolutions that allow the *de novo* modelling of macromolecular complexes, mainly ribosomes, show that high magnifications, down to around a 1 $Å^2$ pixel size can be beneficial to obtain the highest resolution possible in cryo-ET and STA (Tegunov, Xue et al. 2021, Hoffmann, Kreysing et al. 2022, Xue, Lenz et al. 2022, Xue, Spahn et al. 2023).

Tilt alignment is the next step in tomogram reconstruction. Software tools like IMOD (Mastronarde and Held 2017) or Dynamo (Castaño-Díez, Kudryashev et al. 2012) are used to align the individual tilt images. One common method employs fiducial markers, often colloidal gold beads, added to the sample before plunge freezing or sputter coated on the finished lamella for cellular samples. These markers, ideally dispersed evenly around the sample, serve as reference points for precise tilt alignment. However, in situations where fiducial markers are not feasible, such as for most *in situ* tomography using FIB-milling, "patch tracking" is used (Mastronarde and Held 2017). This method divides tilt images into smaller patches and estimates tilt alignment by comparing these patches between tilts. Recently AreTomo (Zheng, Wolff et al. 2022) has been introduced as an automated GPU-based solution for tomogram reconstruction. IMOD also provides a robust automation interface so that a trend towards fully or partially automated tomogram alignment has become more common place. Even though reservations about the alignment quality remain, the increasing size of tomography datasets, now routinely in the 100s to over 1000 tomograms, will necessitate a further step into this direction.

Tomogram reconstruction algorithms include weighted back-projection (WBP), algebraic reconstruction techniques (ART), and simultaneous iterative reconstruction techniques (SIRT). While WBP gives higher-resolution information more weight than low-resolution information, ART and SIRT are iterative algorithms that converge to a 3D reconstruction by comparing the tomogram and its tilts. For high resolution STA WBP has been emerging as the de facto standard.

The next step in tomogram preprocessing is CTF correction. The CTF results in differential information transfer across frequencies. One way to address this is through 2D-CTF correction applied to individual tilts. This process involves estimating the CTF from the Fourier spectrum of the tilt, followed by "phase flipping" to produce a power spectrum with only positive values. 3D-CTF correction, such as with NovaCTF (Turonova, Schur et al. 2017), offers an improvement over the 2D approach by estimating the CTF across a gradient of defoci in a 3D tilted slab. Alternatively, CTF correction and refinement can be handled on a per particle basis as implemented in RELION4 (Zivanov, Oton et al. 2022).

Recent advances in deep learning have provided tools to enhance tomogram quality. Algorithms can be trained on tomograms containing noise (Buchholz, Jordan et al. 2019), correlating tilt images or even frames within tilts to discern and correct noise, resulting in a denoised tomogram. Additionally, methods like IsoNet (Liu, Zhang et al. 2022) apply deep learning algorithms to the missing wedge, attempting to fill in the missing information. These tools are especially helpful if the processing workflow relies on segmentation, but concerns about alteration of the underlying data prevent the routine use in STA.

2.6.4.2 Subtomogram averaging

STA allows the structure of a macromolecule of interest to be solved in its native cellular environment. As the signal in independent subvolumes, termed subtomograms, of the same molecule is correlated while the noise is uncorrelated, averaging of a large number of copies results in isotopically resolved structures with high SNR.

Especially when enhanced by the latest breakthroughs in structure prediction, like AlphaFold2, STA establishes a bridge between protein biochemistry and cellular proteomics. However, for effective STA, it is imperative to first pinpoint the particles of a target macromolecule within the tomograms. The inherent compositional

heterogeneity of especially *in situ* tomograms, compounded by the low SNR and the missing-wedge effect, makes this step non-trivial.

Traditional methods, such as manual annotation or template matching, have their limitations. Manual annotation is labor intensive, potentially introducing biases, while template matching, despite its emphasis on throughput, often compromises on picking accuracy, thereby constraining the effectiveness of STA.

For a subset of proteins, geometry-based picking offers a valid alternative (Castaño-Díez, Kudryashev et al. 2017). Unfortunately, this is only feasible for, for example, a relatively homogenous set of proteins embedded in a membrane, such as in the thylakoid stacks, or for virus capsids.

The advent of deep learning has ushered in a new era of particle picking tools, predominantly leveraging 3D-Unet convolutional neural network (CNN) architectures (Moebel, Martinez-Sanchez et al. 2021, de Teresa-Trueba, Goetz et al. 2023). Yet, these tools often lack generalization. For every protein of interest, users typically need to manually annotate a substantial number of particles in tomograms to train the neural network, a process that is not only tedious but also often unfeasible for a vast majority of the proteome. This limitation has kept traditional methods, like template matching, relevant in cryo-ET workflows (Chaillet, van der Schot et al. 2023, Cruz-Leon, Majtner et al. 2023).

The ideal solution would combine the accuracy of deep learning-based picking without the need for exhaustive manual annotations for each protein. One promising approach is deep metric learning, as recently introduced by TomoTwin (Rice, Wagner et al. 2023), where data is encoded into high-dimensional representations or embeddings. During training, the model is optimized to cluster each class in a distinct region of the embedding space, ensuring similar classes are proximal and dissimilar ones are distant. Such models, by understanding similarity relationships, can generalize effectively, placing new data classes in the embedding space based on their similarity to known classes, all without the need for retraining. Downsides are the high computational cost of this method, constraining throughput. Furthermore, although promising, the technique is still unproven.

Once particles are picked, the alignment and averaging phase starts. Subtomograms are aligned to a reference, which can be either a manually-derived structure or a homologous structure from existing datasets. This alignment is iterative, progressively refining the reference until a convergence criterion is met. During this iterative process,

the missing-wedge effect, has to be accounted for. Contemporary software tools, including, RELION, STOPGAP and Dynamo, have been developed to facilitate subtomogram alignment and averaging, with a particular emphasis on algorithms that mitigate the missing-wedge effect (Castaño-Díez, Kudryashev et al. 2012, Scheres 2016, Wan 2020).

Some software packages, such as emClarity (Himes and Zhang 2018) and M (Tegunov, Xue et al. 2021), have introduced methodologies that leverage high SNR 2D reference projections to refine the title series alignment locally resulting in enhanced subtomogram alignment and superior map quality.

Given the intrinsic heterogeneity often present in biological specimens, classification techniques are indispensable. Applied either during or post alignment, these techniques segregate particles based on different conformational or compositional states, ensuring accurate alignment and preserving high-resolution details. This classification not only refines the averaged structures but also provides insights into the dynamic landscapes of macromolecular complexes.

2.6.4.3 The other things: contextual information in cellular cryo-ET

The advantage of *in situ* cryo-ET is that rich contextual information that is available on top of the 3D structure of the protein of interest. For example, in neurodegenerative diseases, cryo-ET studies revealed that the cellular recycling machinery, the proteasome, gets sequestered and stalled by certain types of aggregates, and is therefore no longer able to keep up with its normal function, unveiling a potential mechanism of cellular toxicity (Guo, Lehmer et al. 2018). Other recent examples include the analysis of the ribosomal conformational space in the cell at high resolution(Hoffmann, Kreysing et al. 2022, Xue, Lenz et al. 2022, Xue, Spahn et al. 2023), as well as work highlighting the roll of molecular organization of the proteasome both on the nuclear pore and the ER (Albert, Schaffer et al. 2017, Albert, Wietrzynski et al. 2020).

But this is only scratching the surface. With increasing speed in data acquisition as well as improving throughput for particle picking and subtomogram averaging as well as the introduction of reliable automated segmentation tools, it is increasingly feasible to map complex biological networks within a cell and understand the intricate relationships between different macromolecular machines as well as the sophisticated level of organization within a cell.

Besides remaining bottlenecks in data acquisition as well as data processing, a large part of the proteome might remain inaccessible due to low copy numbers or size and flexibility of the respective protein. Despite these constrains, recent work shows that cryo-ET can be used to drive a fundamental understand of cellular biology (O'Reilly, Xue et al. 2020, Zimmerli, Allegretti et al. 2021, Mosalaganti, Obarska-Kosinska et al. 2022, Gemmer, Chaillet et al. 2023).

2.6.5 Comparison of SPA and cryo-ET

With recent work on viruses and on the photosystem, the line on what was feasible to do by SPA has been shifted. Both examples are showing that even in crowed environments it is possible to reconstruct the 3D structure of proteins by SPA. Furthermore, recent demonstrations of 2D template matching (Lucas, Himes et al. 2021) followed by SPA like reconstruction of proteins on lamellas point further in the direction that one should be able to structurally characterize a much wider range of proteins in cells by SPA-like approaches. Further developments in this direction greatly enhance the prospects of a post reductionist approach to structural biology.

This will by no means substitute tomography. Even though one can reach high resolutions this way, work by Tegunov et al. (Tegunov, Xue et al. 2021) shows that with the same number of particles there is little difference between the resolution achievable by SPA and STA. Furthermore, the real strength of *in situ* tomography does not exclusively lie in its ability to resolve high resolution structures in their native environment but in the context around these macromolecules. This context enables the formulation of new hypothesis about how our cells function and interact.

Recent studies demonstrated the great potential of combining high resolution SPA data with the cellular context exclusive to *in situ* tomography (Fig. 27) (You, Zhang et al. 2023).

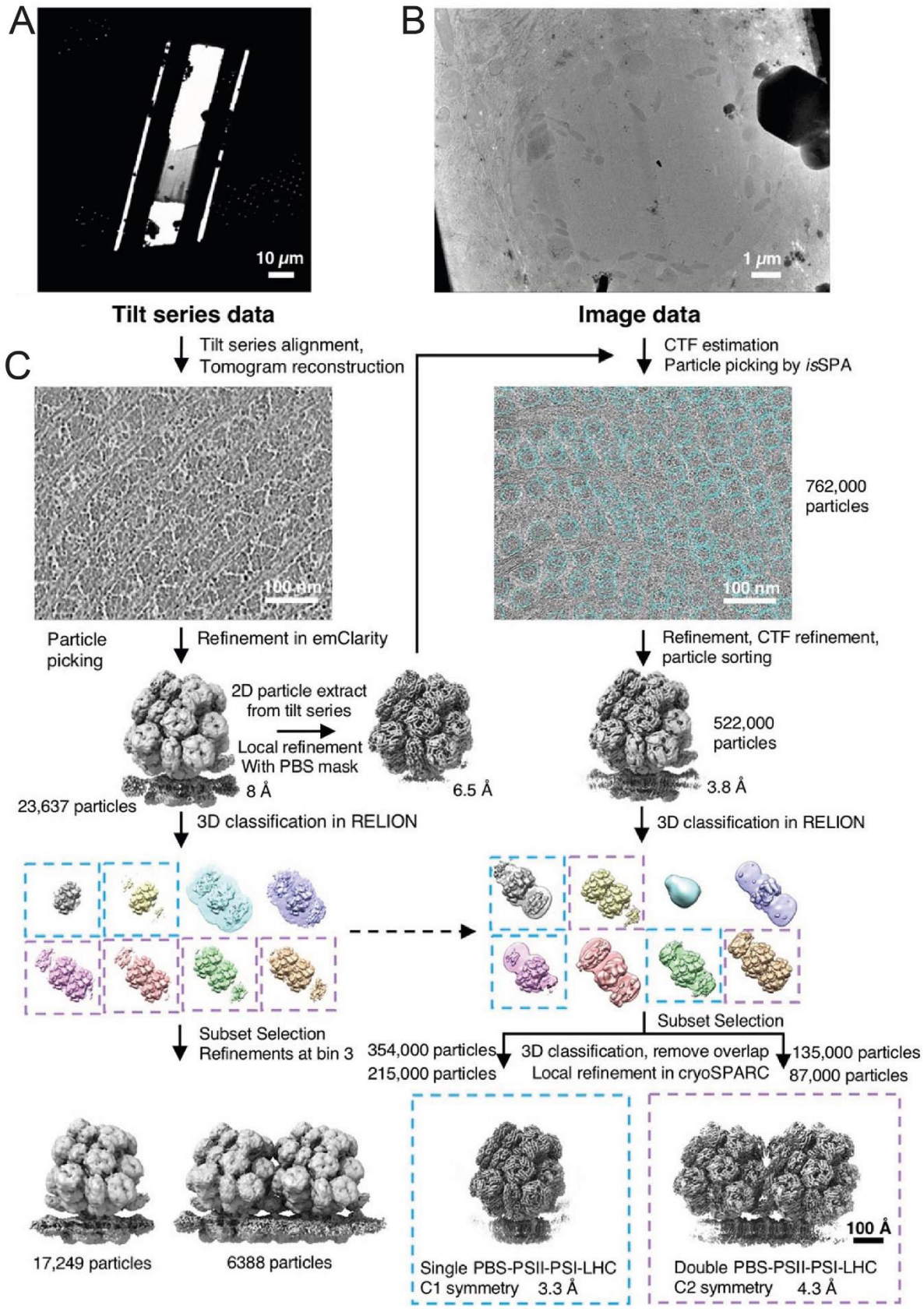

Figure 27: Example for a combined SPA and cryo-ET for red algal phycobilisome–PSII–PSI–LHC megacomplex.

(A) Low magnification EM image showcasing the structure of P. purpureum lamella. This image is one of 46 similar tilt series captured for analysis. (B) A higher magnification EM image of P. purpureum lamella. Chosen from the same set of 46 tilt series, highlighting finer

structural details. (C) Step-by-step cryo-EM analysis workflow: Starting with 46 tilt series, 23,637 sub-tomogram particles were identified and extracted using emClarity. These were then aligned and averaged, yielding an 8 Å resolution structure. Further 3D classification identified two unique shapes: the single and the double PBS-PSII–PSI-LHC supercomplexes. Subsequently, images from the tilt series corresponding to all subtomograms underwent local refinement in RELION using a soft-edge mask around the PBS, which refined the resolution to 6.5 Å. This high-resolution PBS structure served as a reference to spot potential particle candidates (762,000 in total) from a large set of high-dose images (2,245) using the isSPA technique. After additional rounds of particle selection, classification, and refinement, the final high-resolution structures of the single and double PBS-PSII-PSI-LHC supercomplexes were achieved, based on 215,000 and 87,000 particles, respectively. Adapted from (You, Zhang et al. 2023)

Dietrich et al. demonstrated that neither technique alone is sufficient to see the full picture and only a combination of both and other biochemical and biophysical techniques enable a holistic understanding of the biology around us (Dietrich, Righetto et al. 2022).

3 Aim of this study

This study aims to address two important open questions about the mechanism of GroEL/ES folding in the cell:

1. Does the substrate reach the native state within the folding chamber? and 2. Which species contributes to GroEL/ES-mediated protein folding *in vivo*?

Despite biochemical evidence that substrate proteins can fold inside the GroEL/ES chamber it was so far impossible to observe such a complex in molecular detail. A detailed structure of a (partially) folded substrate inside of the chamber would allow a better understanding of how the interactions between substrate protein and GroEL are able to facilitate folding of the substrate protein and allow for subsequent release.

The modulation of ATPase cycle by substrate protein and other co factors as described above leads to an ongoing debate about the folding active form of the GroEL/ES complex *in vivo*. As described in section 1.5.4, reconstitution systems are difficult due the complex nature and interplay of all different components.

To address these two open issues, we combined single particle cryo electron microscopy of GroEL-ES substrate complexes, *in situ* subtomogram averaging of GroEL/ES in focused ion beam milled *E. coli* cells with a series of biochemical experiments to elucidate the nature of the reaction cycle of this fascinating molecular machine and determine the substrate folding status inside the GroEL/ES chamber *in vitro* as well as *in situ*.

4 Results

In this section, we will address the objectives of the study in a structured sequence. Initially, evidence demonstrating that a substrate can attain a stable, near-native conformation within the GroEL:ES chamber will be presented. To explore this, we performed *in vitro* SPA on reconstituted GroEL:ES complexes, both alone and in conjunction with the obligate substrate protein MetK. The comparative analysis of the resulting high-resolution structures allowed us to elucidate the interaction between the GroEL:ES chamber and the near-native MetK molecule inside, as well as to observe how the folding of a large (~42 kDa) protein influences the conformation of the GroEL:ES chamber.

Secondly, we compared the *in vitro* structures to *in situ* structures acquired by subtomogram averaging of GroEL:ES in its native cytosolic environment. After evaluating various model systems, we selected *E. coli* for detailed study. Our findings indicate that the distribution of different GroEL:ES species remains relatively unchanged under various cellular disturbances, signifying that the protein folding mechanism is consistent across the tested physiological conditions.

4.1 SPA analysis of GroEL:ES complexes

4.1.1 High resolution cryo-EM structure of the EL:ES complex without substrate

There are a variety of structures of GroEL with and without cofactors and in different nucleotide states deposited in the PDB. However, there was no high-resolution SPA cryo-EM structure available for the different stoichiometries of the

EL:ES complex with the ATP transition state mimic ADP·BeF_x.

Therefore, as a starting point of the folding reaction of the EL:ES system *in vitro* we solved the molecular structure of the EL:ES_2 complex in the presence of ADP·BeF_x to a global resolution of 3.1 Å without the application of symmetry and to a resolution of 2.7 Å after the application of D7 symmetry (Fig. 28 A and C). In addition, we resolved the structure of EL:ES_1 complex to a global resolution of 3.0 Å in *C1* and 2.7 Å after the application of *C7* symmetry, respectively (Fig. 28 B and D).

We used these densities to build *de novo* models of both complexes and use these as reference points for further comparisons. Side chains in the hydrophobic core a well resolved as well as the nucleotide in the binding pocket (Fig. 28 insets).

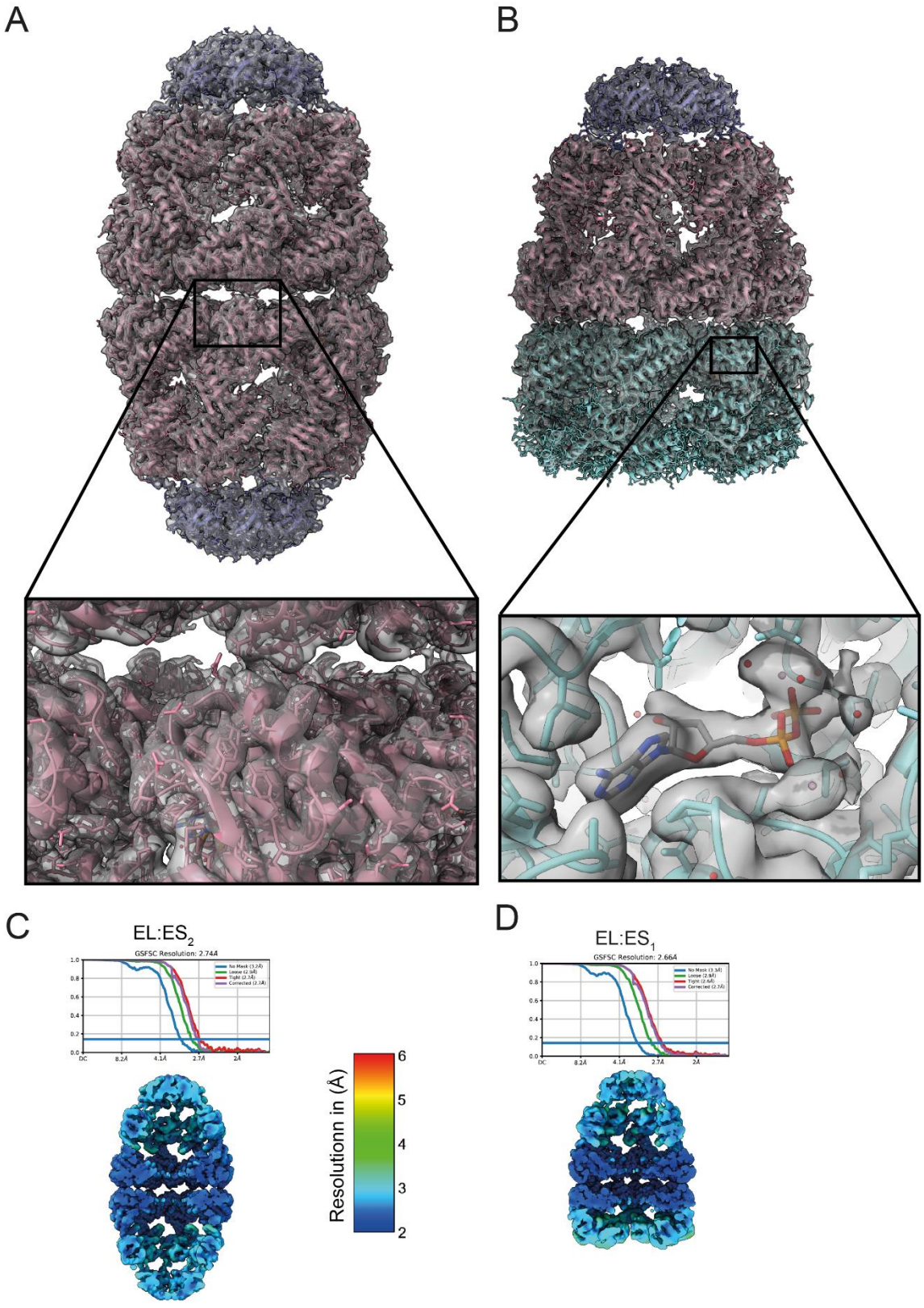

Figure 28: Structure of the EL:ES complexes in the presence of the ADP·BeF$_x$.

(A) Structure of the EL:ES$_2$ complex. Cryo-EM density of the EL:ES$_2$ complex with the corresponding model docked into the structure. The insets display the well resolved side chain densities in the equatorial domain.

(B) Structure of the EL:ES_1 complex. Cryo-EM density of the EL:ES_1 complex with the corresponding model docked into the structure. The insets display the well resolved density of the ATP transition state analogue ADP·BeF_x and surrounding water molecules.
(C-D) Local resolution maps and FSC curves. The local resolution maps and the corresponding FSC curves for the structures in (A) and (B) are depicted. Both structures have a global resolution of 2.7 Å after the application of *D7* and *C7* symmetry respectively.

4.1.2 Structure of protein folding intermediates inside the GroEL/ES chamber

Earlier cryo-EM studies have shown evidence for client proteins encapsulated inside the GroEL-GroES chamber (Chen, Madan et al. 2013, Kim, Park et al. 2022). With Rubisco, densities distinct from the GroEL-GroES cage were observed at the equatorial end of the cavity. These studies, however, did not resolve the precise molecular architecture of the substrate protein nor any detailed interactions within the EL:ES chamber with confidence.

To close this gap in our understanding of the GroEL:ES reaction cycle, we formed a stable, highly occupied GroEL:ES substrate complex, by heat-denaturing an excess of MetK in the presence of EL (Koike-Takeshita, Yoshida et al. 2008). Subsequently, ES and ADP·BeF_x were added to form and stabilize the ternary complex. The ternary complex was isolated by size exclusion chromatography (Fig. 29 A).

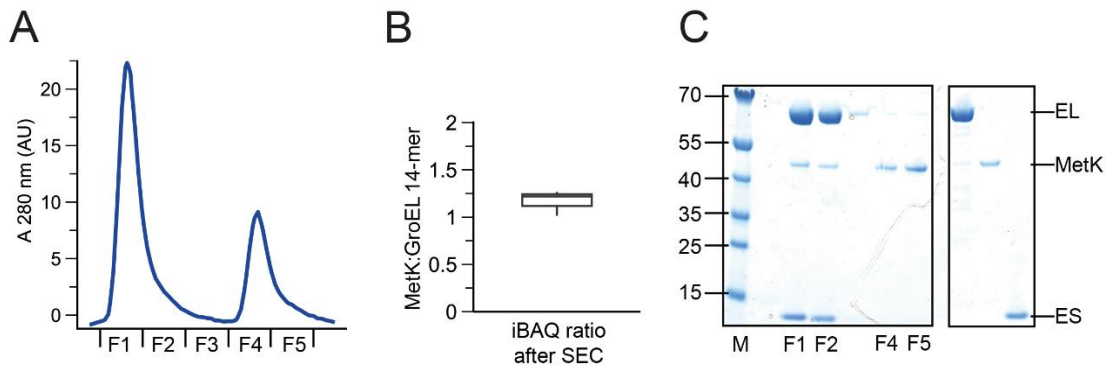

Figure 29: Biochemical characterization of the GroEL:ES:MetK complex.

(A) Size exclusion chromatogram. The stable GroEL:ES:MetK complex was prepared in the presence of ADP·BeFx, as outlined in Materials and Methods, and subsequently isolated by size exclusion chromatography. A representative chromatogram is shown. (B) Quantifications of EL:MetK ratios. The relative ratio of MetK to the GroEL 14-mer was estimated by using MS IBAQ values of the first peak fractions F1 and F2 (n = 3 independent replicates). (C) SDS-PAGE characterizations. All elution fractions as well as purified standards for GroEL GroES and MetK were characterized by SDS-PAGE followed by Coomassie staining. It was confirmed that the complex elutes in the first two fractions while the second peak (F4 and F5) only contains unbound MetK.

We confirmed the binding of MetK by mass spectrometry and quantified the relative substrate abundance using intensity-based absolute quantification (IBAQ) (Schwanhausser, Busse et al. 2011) (Fig. 29 B). The estimated MetK-GroEL tetra dodecamer molar ratio was ~1.2.

This sample was subsequently analyzed by cryo-EM (Fig 30). Reference-free 2D classification and *ab initio* reconstruction showed a ~ 27 % :73 % mixture of double and single capped complexes (for more detail see Material and Methods and Fig. 30 A-D). To simplify subsequent steps, we processed single ring unit substructures, i.e., EL_7:ES_7 units (Fig. 30 E-G).

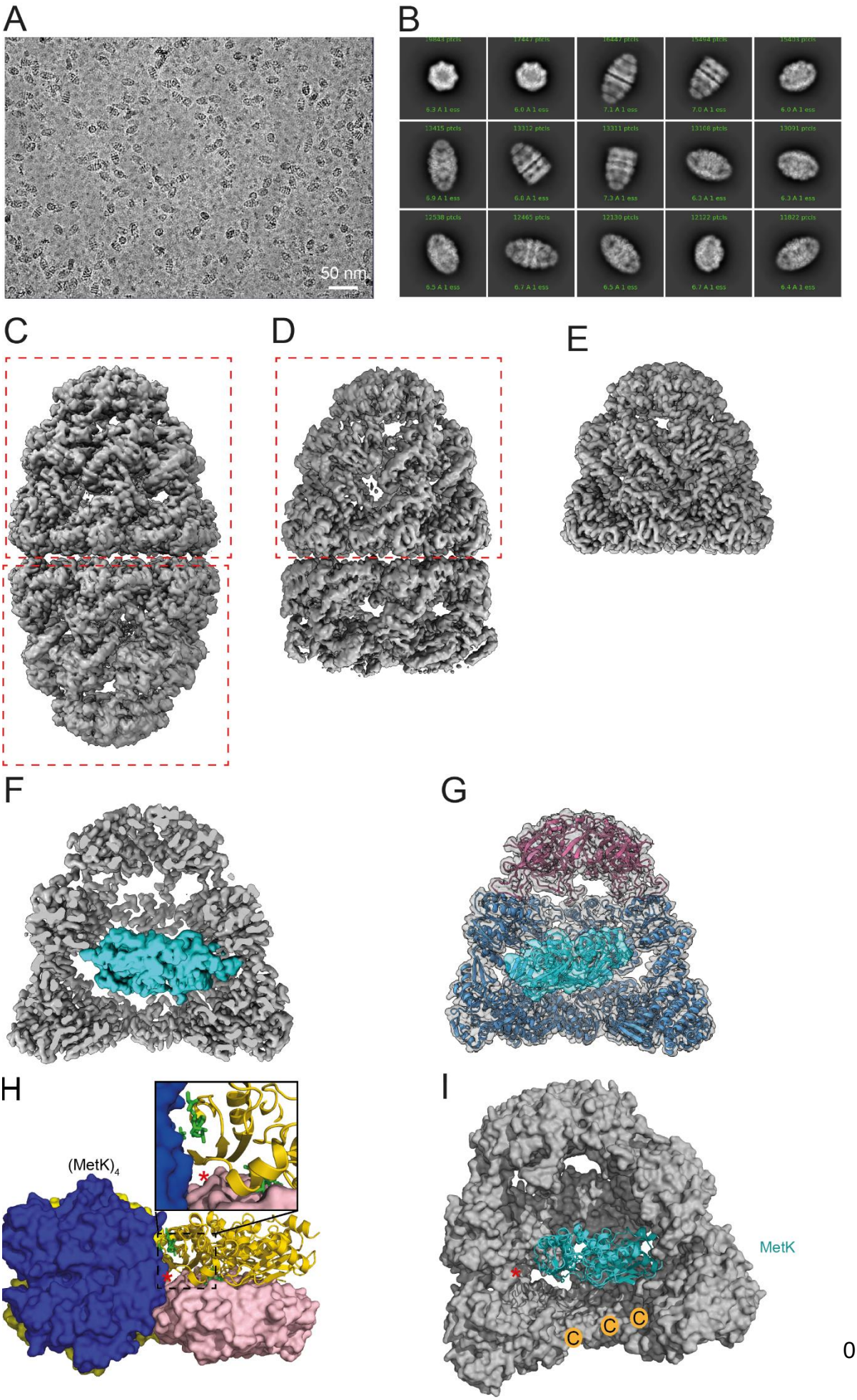
A
50 nm
B
6.3 A 1 ess
6.0 A 1 ess
7.1 A 1 ess
7.0 A 1 ess
6.0 A 1 ess
13415 ptcls
13312 ptcls
13311 ptcls
13108 ptcls
13091 ptcls
6.9 A 1 ess
6.8 A 1 ess
7.3 A 1 ess
6.3 A 1 ess
6.3 A 1 ess
12538 ptcls
12465 ptcls
12130 ptcls
12122 ptcls
11822 ptcls
6.5 A 1 ess
6.7 A 1 ess
6.5 A 1 ess
6.7 A 1 ess
6.4 A 1 ess
C
D
E
F
G
H
(MetK)4
I
MetK
C
C
C

Figure 30: Single particle analysis of cryo electron microscopy data of GroEL:ES:MetK complexes.

(A) Representative micrograph. Exemplary micrograph of GroEL:ES:MetK complexes. The GroEL:ES:MetK specimen at a magnification of 22,500x, captured on a Titan Krios G1 equipped with a Gatan K3 Summit detector. (B) 2D classification. A selection of the 2D classes of particles selected for further refinement. (C-D) Initial reconstructions. Surface representation of the densities for MetK-EL:ES double capped (C) and single capped complexes (D). The EL-ES chambers delineated with red boxes were further processed to resolve the structure of the GroEL:ES:MetK subcomplex shown in (E). (E) EL_7:ES_7 structure. Surface representation of the density for the refined MetK-EL:ES chamber unit. (F-G) EL:ES:MetK complexes. Depiction of a side view of the chamber interior in surface rendering (F) and with the superimposed structural model (PDB 8P4O) displayed in a ribbon format (G). (H) Crystal structure of the functional MetK tetramer (PDB 7LOO (Gade, Tan et al. 2021)). One subunit is highlighted as gold ribbon representation and the remaining three illustrated as molecular surfaces in tones of violet, blue, and yellow. An inset highlights the area proximal to the core loop, indicated by a red asterisk. The bound ligands, pyrophosphate and S-adenosylmethionine, are depicted in green stick models. (I) Relative orientation of MetK in the EL_7:ES_7 chamber. Cut-away visualization of the GroEL-ES chamber containing a folded MetK within (PDB 8P4O). MetK is rendered as ribbon in teal. Molecular surfaces represent the GroEL and GroES subunits. The MetK core loop is highlighted by a red asterisk, and the final ordered residue Pro525 in the GroEL subunits is indicated with 'C'. The disordered C-terminal GGM repeats (GroEL residues 536-548) could readily approach the exposed MetK interface regions, but no additional density was found on the surface of MetK.

Extensive alignment and classification (Fig. 31) showed that ~38 % of the chaperonin chambers exhibited density for an ordered substrate molecule with a defined orientation towards the cage wall (Fig. 30 F and G), whereas the remainder of the EL:ES chambers were either empty or contained a disordered MetK.

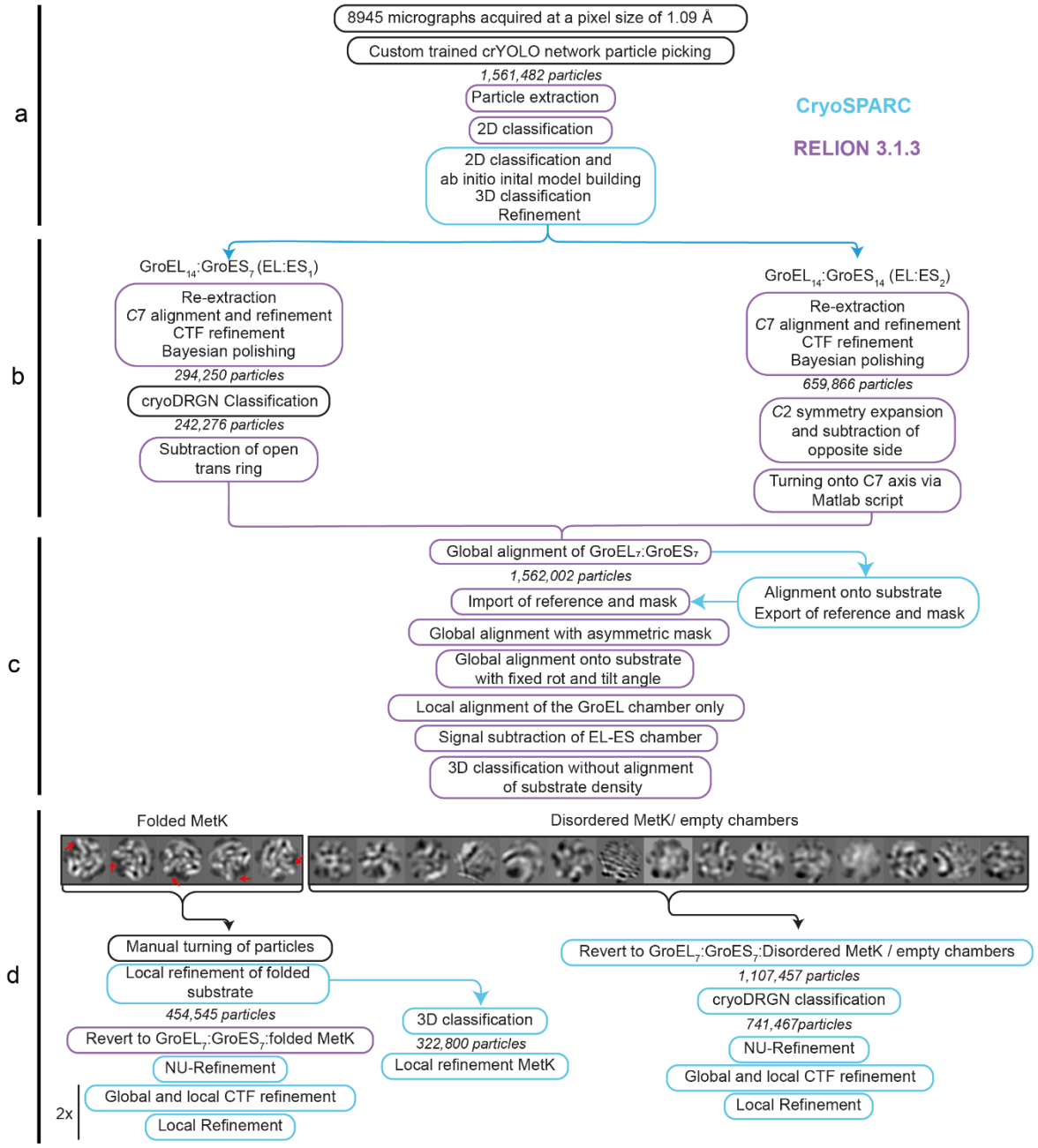

Figure 31: **Workflow for single particle analysis of EL:ES:MetK complexes.**

Single particle analysis data processing for EL:ES:MetK complexes. The flow diagram displays the workflow. Steps are demarcated by purple, light blue, and black borders, denoting usage of CryoSPARC, RELION 3.1.3, or the specified program, respectively. Following data collection, particle picking, and initial 2D classification (a), single capped EL:ES and double capped EL:ES complexes were processed separately, yielding 3D averages. (b). GroEL:ES chambers were extracted and combined for further processing. Subsequent to the subtraction of GroEL-ES density, the interiors of the chambers were separated via 3D classification without alignment (c). The row of images showcases central slices from the resulting 3D class averages, which aided the differentiation of folded MetK from disordered MetK or vacant chambers. All chambers in the disordered MetK/ empty selection show additional density inside of the chamber (d). Red arrows highlight MetK densities that vary by a 51.4° rotation in the GroEL-ES chambers. The other 3D class averages lacked discernible secondary structure elements. The GroEL:ES:MetK complexes that contained folded MetK were aligned and refined to a resolution of 3.0 Å. Local

refinement of MetK after signal subtraction led to a 3.7 Å resolution map. The final resolution for GroEL-ES chambers with disordered MetK or vacant chambers was determined to be 2.9 Å. For additional details, see Materials and Methods.

EL:ES chambers not containing a folded MetK displayed essentially *C*7 symmetry – likely at least partially due to the inability to align on small deviations of the GroEL chamber – and a slight expansion compared to empty chambers that were solved to high resolution as a reference structure (r.m.s.d. values of 0.28–0.40 Å, maximum Cα displacements 0.78–1.16 Å) (Fig. 32). All classes from 3D classification contained residual density in the chamber (Fig. 31), and further sub-classification attempts to find empty chambers or folding intermediates proved unsuccessful.

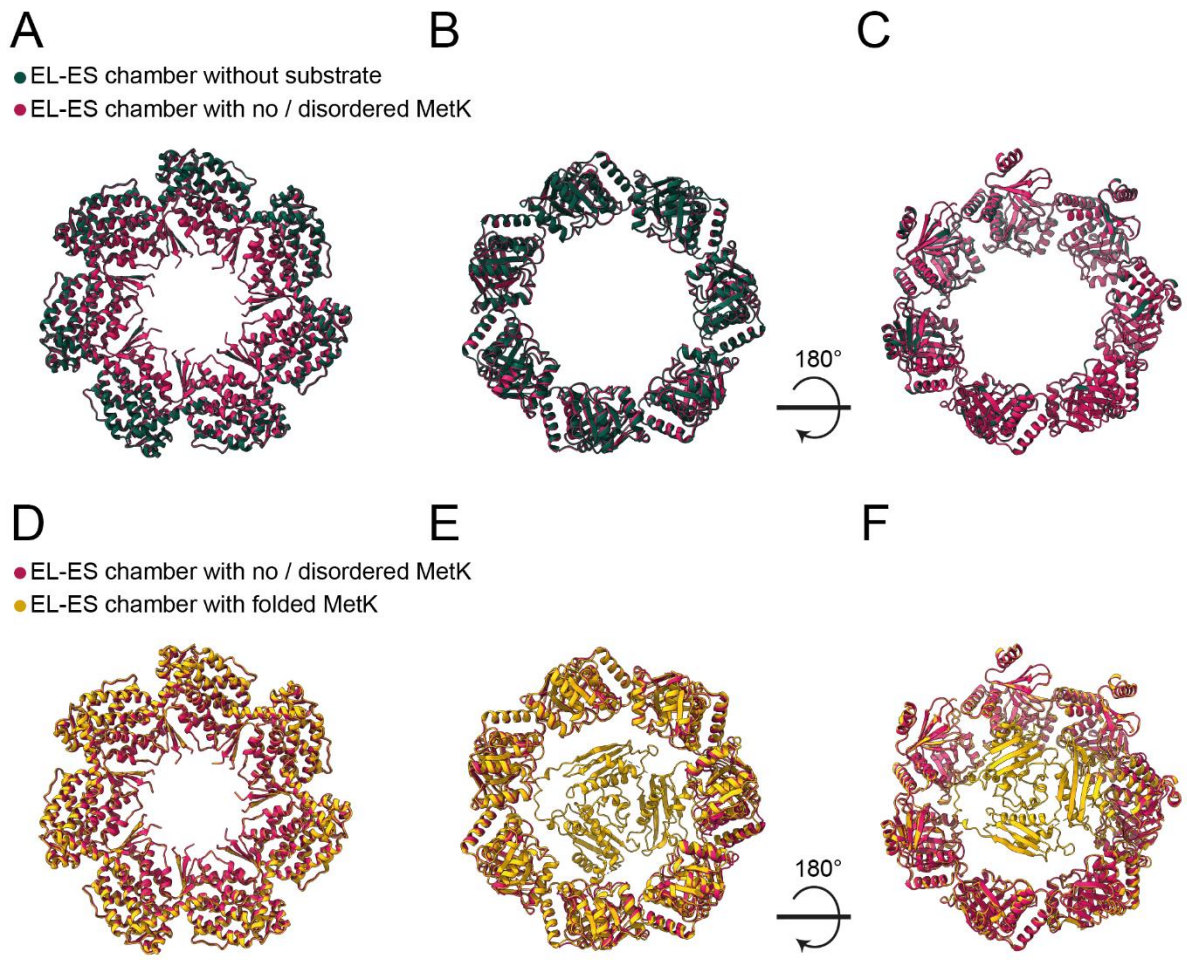

Figure 32: Conformational changes of the EL:ES chamber upon binding and folding of MetK

(A-C) Structural comparison. Comparison of the *C7*-symmetric model of the empty GroEL-ES chamber (dark green, PDB 8P4M) with its corresponding structure containing either disordered MetK or lacking substrate (pink, PDB 8P4N). This comparison shows structural changes at the level of the equatorial GroEL domains (A), the intermediate domains, and the hinge regions connecting equatorial and intermediate domains (B and C) viewed from both sides.
(D-F) Structural comparison. A comparative analysis of the model of the structure containing either disordered MetK or lacking substrate in the GroEL-ES chamber (pink, PDB 8P4N) with the chamber that accommodates folded MetK (gold, PDB 8P4O). This analysis offers insights into structural rearrangements at the equatorial GroEL domains (D), the intermediate domains, and the hinge regions that link equatorial and intermediate domains (E and F), viewed from bottom and top.

The ordered MetK had a nominal resolution of 3.7 Å, revealing sidechain density in its hydrophobic core (Fig. 31 F-I, Fig. 33, Fig. 34). Comparison with the crystal structure of the native MetK tetramer (PDB 7LOO (Gade, Tan et al. 2021)) showed that the ordered substrate had a native-like structure. The main deviation was observed in the

conformation of the core loop, residues 97-111, which interacted with the bound S-Adenosyl methionine and an adjacent subunit in the native MetK tetramer (Fig. 30 H, 33 F).

The disc-shaped MetK molecule, occupied a horizontal section near the cavity center, allowing ample space both above and below. Contacts to the chamber walls are potentially formed with GroEL residue Phe 44, which is located in the equatorial domain at the tip of a helical hairpin, in two GroEL subunits, and residues Phe 281 in five and Tyr 360 in three GroEL subunits, respectively (Fig. 33). These latter residues are located in the apical domain and protrude into the cavity as well. In the complex, these protrusions appear to pinch the bulky MetK molecule via van-der-Waals contacts. Many of the putative contact residues in MetK appear to have rather mismatched chemical properties, especially the carboxylate group bearing Asp 110, Glu 200, Glu 204, Glu 205 and Asp 378. It should be noted that the contact residue sidechains in the chaperonin-substrate were rather poorly defined; the peptide backbones were however clear (Fig. 33 A-F, Fig. 34 and Fig. 36)

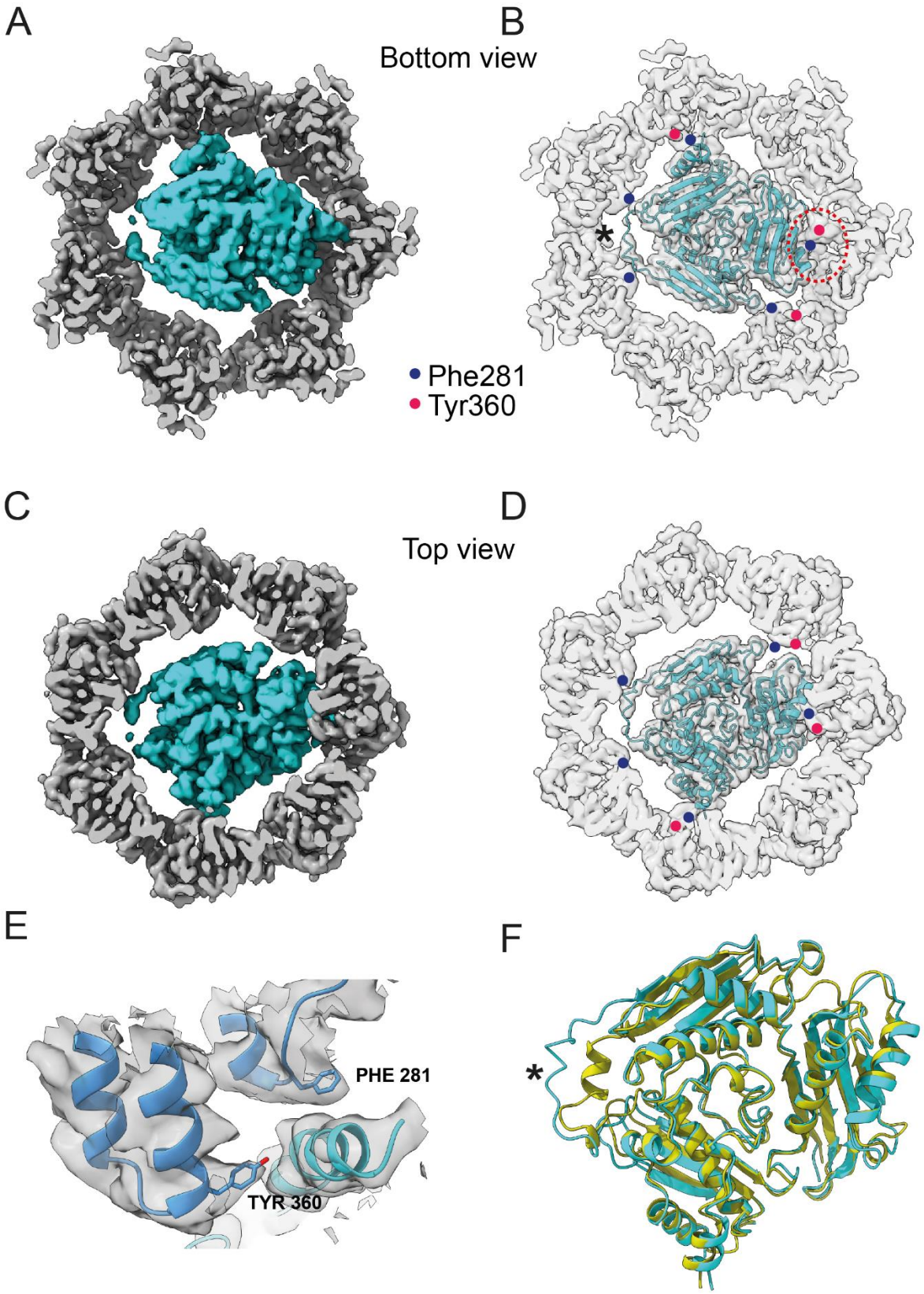

Figure 33: Interactions of the EL:ES chamber with the encapsulated MetK.

(A-D) Overall views of MetK contacts with a GroEL subunits. Cut-away depictions present a top view (A, B) and bottom view (C, D) of the density (A, C) and the superposed MetK model (teal) is represented in ribbon format (B, C). The contact residues Phe 281, and Tyr 360 of GroEL are highlighted by dark blue and pink colored dots, respectively (B, D). (E) Details of

a MetK-GroEL contact. A magnified view of the position highlighted by a pink ellipse in panels B and D is shown. The contact residues Phe 281 and Tyr 360 are displayed as stick models.
(F) Comparison of MetK structures. The superimposed structures of encapsulated MetK (teal) and a subunit in the MetK tetramer (PDB 7LOO, yellow) are displayed. The core loop of MetK is identified with an asterisk. MetK is in the same orientation as in (D).

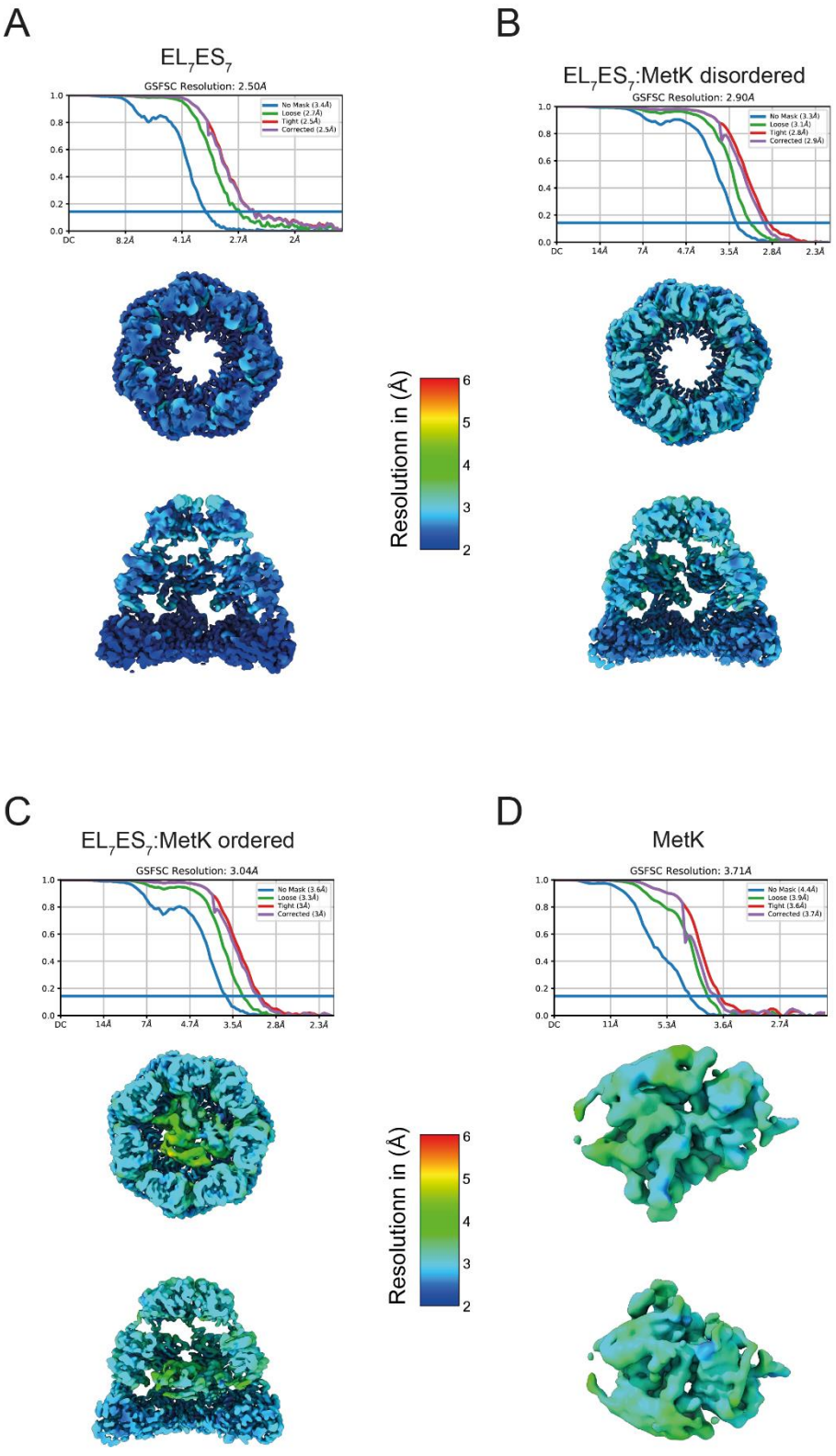

Figure 34: Local resolution and FSC curves of GroEL:ES:MetK complexes.

(A-D) FSC curves and the corresponding local resolution maps in a cut open top or side view of (A) EL_7:ES_7 (B) EL_7:ES_7 containing either disordered MetK or lacking substrate, (C) EL_7:ES_7 containing a fully MetK monomer, and (D) local refinement of the MetK density from

(C). Resolutions of 2.5 Å, 2.9 Å, 3.0 Å and 3.7 Å respectively were calculated. The color key of the local resolution maps is depicted besides panels (C) and (E) respectively.

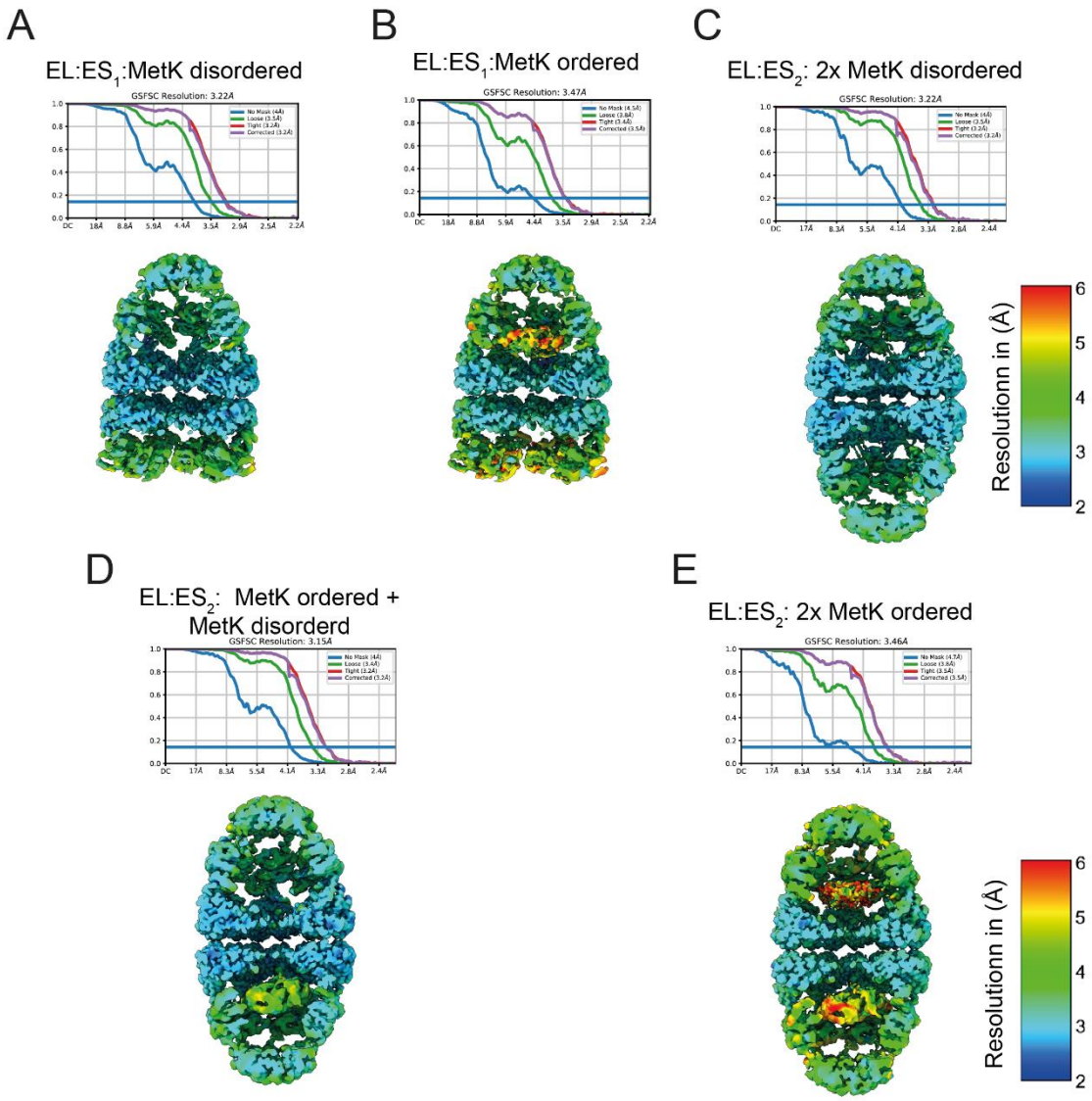

Figure 35: Local resolution and FSC curves of EL:ES:MetK complexes.

(A-E) FSC curves and the corresponding local resolution maps in a cut open side view of (A) EL:ES_1 containing disordered MetK or lacking substrate at a global resolution of 3.2 Å, (B) EL:ES_1 containing a fully folded MetK monomer at a global resolution of 3.5 Å, (C) EL:ES_2 containing disordered MetK or lacking substrate in both chambers at a global resolution of 3.2 Å (D) EL:ES_2 containing disordered MetK or lacking substrate in one chamber, and a fully folded MetK monomer in the opposing chamber at a global resolution of 3.2 Å and (E) EL:ES_2 with two fully folded MetK monomers in both chambers at a global resolution of 3.5 Å. One is blurred due to the rotational independence of the two substrate molecules. The color key of the local resolution maps is depicted on the right-hand side.

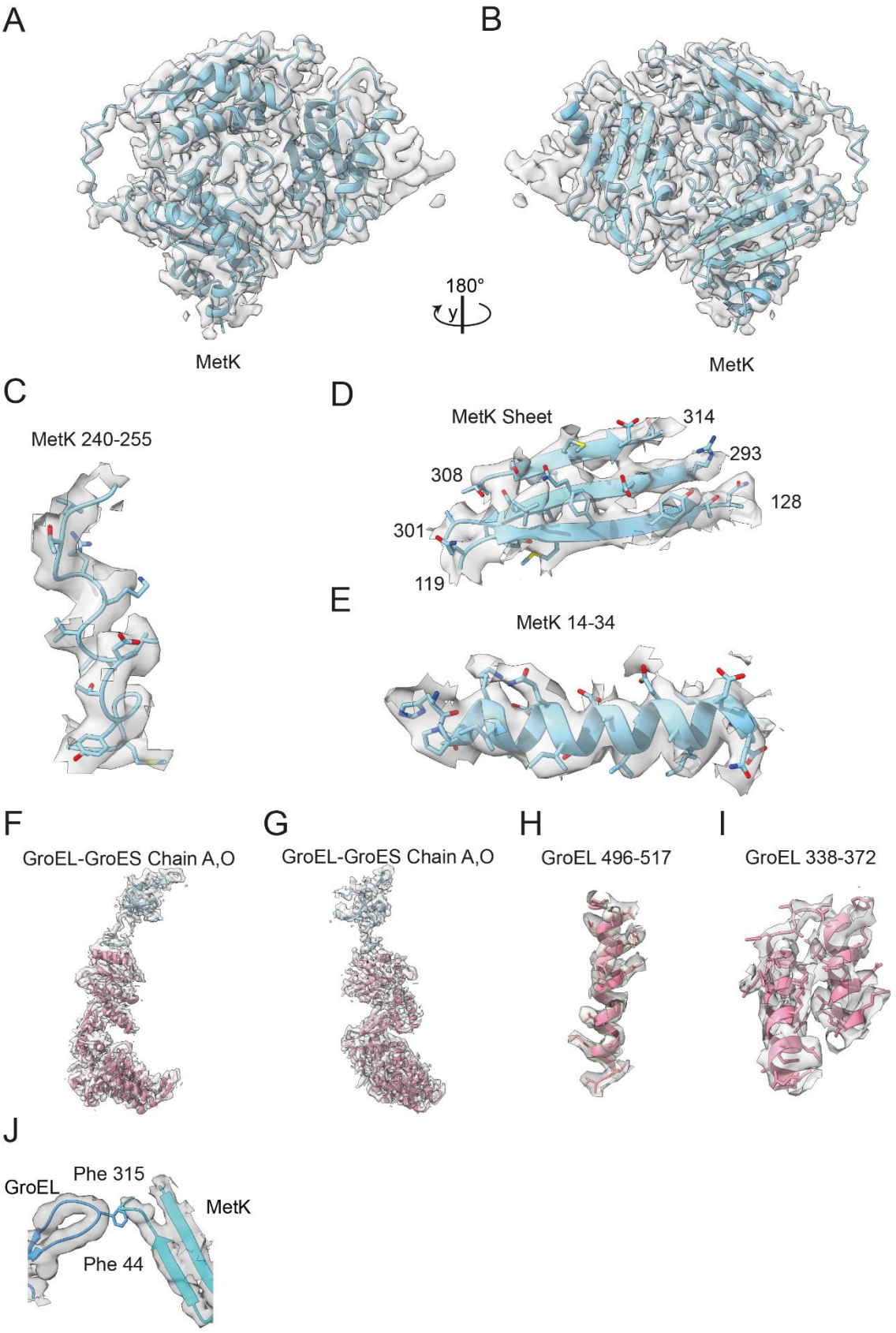

Figure 36: Structure model fit validation of the EL:ES:MetK complex.

Cryo-EM density of MetK with fitted structural model in ribbon representation. Two views for the entire protein are shown (A). Exemplary portions of the structure are shown below with side chains in stick representation (C-E). The respective residue ranges are indicated. Cryo-

EM density of EL:ES chains with fitted structural models in ribbon representation in two views (F, G) with zoomed in highlights of an alpha helix (H) and a hairpin loop (I). (J) shows the putative contact side between MetK and GroEL at the GroEL residue Phe44. The local resolution of this site does not allow a confident fitting of side chain conformations but the backbone is clearly visible.

The GroEL subunits contacting MetK show only minor conformational rearrangements, with r.m.s.d. of 0.52–0.96 Å and maximum Cα displacements of 1.96–3.37 Å compared to empty GroEL-GroES complexes (Fig. 32). Of note, the chaperonin cage does not mask any of the interface regions buried in the MetK tetramer. These appear exposed to solvent in the chamber. In summary, a fraction of the encapsulated MetK molecules appears to have assumed the native fold inside the chaperonin chamber, ready for assembly of the MetK enzyme complex. The mismatched interactions with the chamber walls suggest a small energetic penalty for substrate release, and may rather result in a considerable entropy gain upon proper solvation of the MetK contact residues.

The location near the bottom of the chamber would also allow for interactions of MetK with GroEL GGM repeat domains which have been found to be beneficial for folding (Hayer-Hartl, Bracher et al. 2016). Despite this we were not able to resolve any extra density near between the bottom of the MetK molecule indicating a more flexible interaction with the repeat domain.

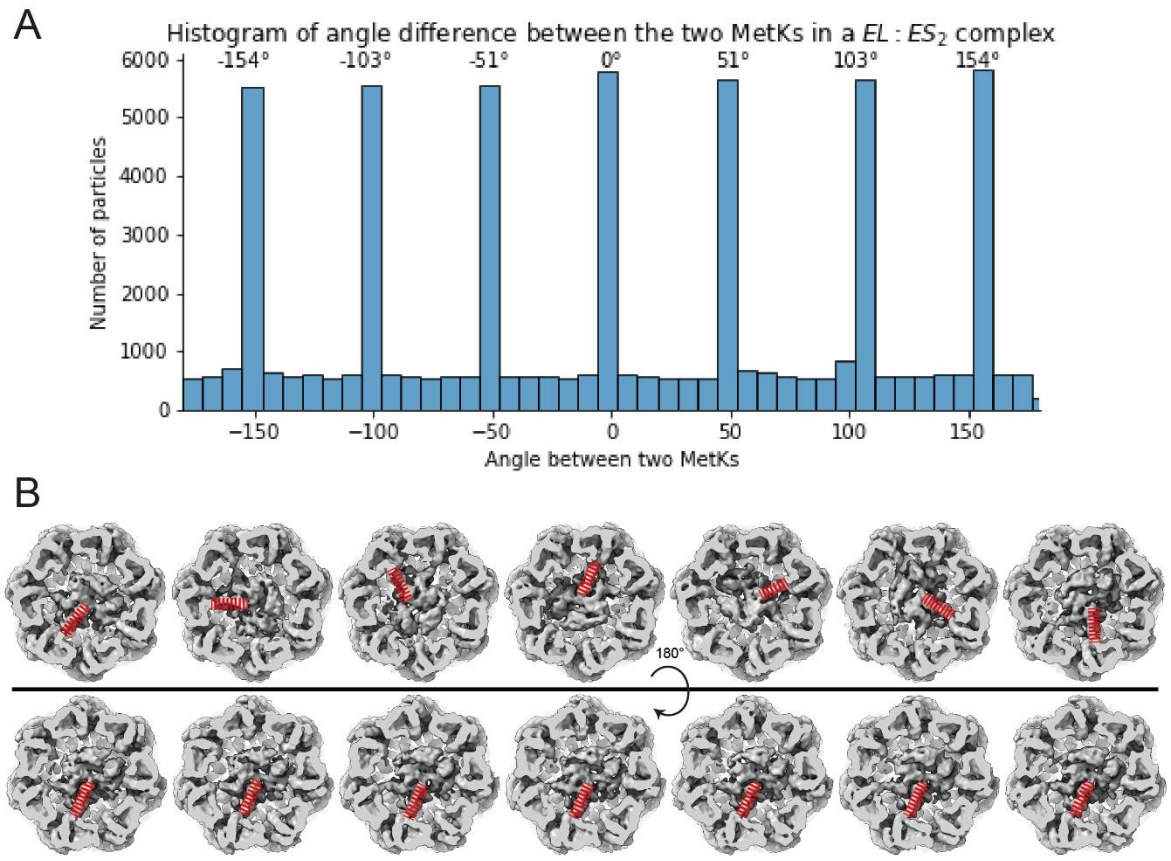

Figure 37: Torsion angles between two folded MetKs in EL:ES$_2$ complexes.

(A) Histogram of the relative angle between the two folded MetKs in the chambers of an EL:ES$_2$ complex containing two folded substrate molecules. The angle is indicated on the x-axis and the relative abundance on the y-axis. The calculated expected peaks around the *C7* symmetry axis of GroEL are indicated above. The experimental distribution shows that there is no coupling of the orientation between two folded molecules. (B) EM reconstructions. Reconstructions of the two halves of the MetK containing EL:ES$_2$ complexes are shown in B. the orientation of MetK is indicated by a red dashed line. GroEL was cut at the level of the intermediate domains to make MetK in the chamber visible.

Mapping back the substructure particles onto the original particles showed that all five possible ternary complexes were formed in presence of ADP·BeF$_x$: Bullets with ordered and disordered MetK and footballs with either one or two ordered MetK molecules, respectively (Fig. 35). The orientation of the MetK molecules in footballs with two ordered substrates encapsulated was stochastic, suggesting missing allosteric communication between the EL rings in the football under the given experimental conditions or EL ring-ring exchange as in the presence of ADP·BeF$_x$ (Fig. 37).

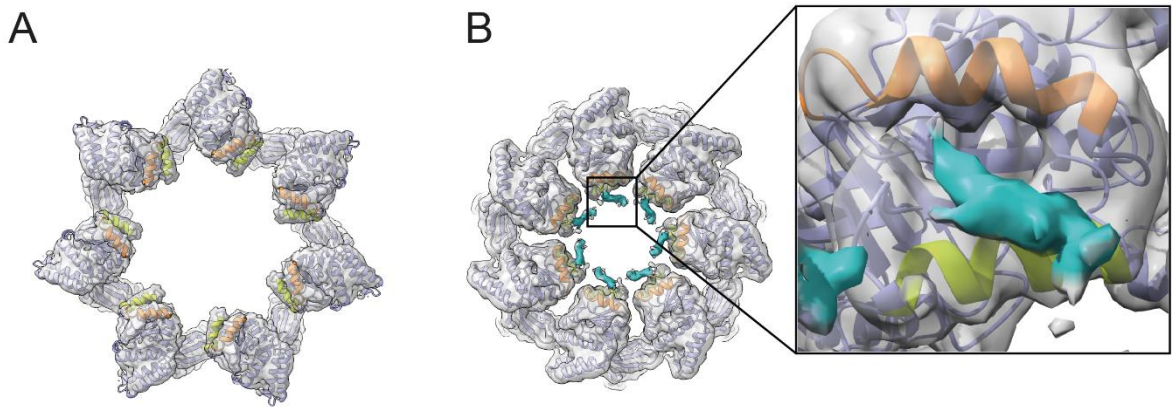

Figure 38: Different states of the apical domains in EL:ES$_1$.

(A) Density and superimposed model of the EL:ES$_1$ apical domains in the wide conformation. The density of the SPA structures is displayed in light semitransparent gray, the model in a dark blue. The helices indicated in substrate binding, αI and αH are highlighted in orange and yellow. No additional density is discernable even at high contour level at the substrate binding interface in the wide conformation. (B) Density and superimposed model of the EL:ES$_1$ apical domains in the narrow conformation. Densities and model the same as in (A) and additional density in the substrate binding cleft between helix αI and αH is highlighted in blue. The inset shows a magnified view of the potential substrate binding. Substrate density is only visible at relatively high couture levels indicating a symmetry averaged mixture of different substrate binding patterns on the seven available apical domains.

We further subdivided the EL:ES$_1$ particles by the opening diameter of the *trans*-ring by reference free 3D classification with a disc shaped mask on the apical domains of the *trans*-ring. Two different conformations were observed, referred to in the following as “narrow” and “wide” (Fig. 38). In the narrow state, the opening of the *trans*-ring has a diameter of ~50 Å (Fig.38 B), similar to the crystal structure of the GroEL:ADP$_7$:GroES complex (PDB 1AON (Xu, Horwich et al. 1997)). In contrast, the wide conformation shows a significant reorientation of the apical domains, extending the ring opening to ~60 Å (Fig.38 A), which would facilitate exit of larger SPs from the cavity such as folded MetK (~70 x 60 x 30 Å in size). A similar *trans*-ring conformation was previously observed in a cryo-EM structure of EL:ES$_1$ with ADP bound in both rings (PDB 7PBJ (Kudryavtseva, Pichkur et al. 2021)). Consistent with a role in facilitating discharge of folded SP, this state was interpreted as an intermediate in the asymmetrical chaperonin cycle after ATP-hydrolysis in the *cis*-ring and prior to ATP and GroES binding to the *trans*-ring (Fig. 9) (Kudryavtseva, Pichkur et al. 2021).

4.2 Understanding the biology of GroEL/ES *in situ*

To determine whether the structures identified in the previous chapter are representative of the GroEL-GroES reaction cycle *in vivo*, and to what extent, we employed *in situ* cryo-ET to determine the stoichiometry of various GroEL-GroES complexes within their natural environment by subtomogram averaging.

4.2.1 Model organisms for studying group I chaperonin complexes *in situ*

At the start of this work, we explored multiple species of bacteria for a variety of biological and technical criteria. For cryo-ET in general and this study in particular, there are several considerations when choosing a model system:

1) Imaging properties. Previous studies in *Chlamydomonas Reinhardtii* and a variety of human cell lines have indicated that a lower macromolecular concentration in the cytosol is advantageous for the identification and averaging of macromolecular complexes. Two possible explanations are the lower degree of overlap and lower background in the tomographic reconstruction leading to an improved signal to noise ratio (SNR).
2) Biological properties. There are a variety of different biological properties that are desirable for the investigation of the GroEL:ES complex *in situ*. Species with established molecular biological tools allow for the fast and easy manipulation of the protein folding system by perturbation. To fully understand a system beyond the initial observation, testing of different hypothesis by perturbation is essential. Another consideration is the ease of handling. Species with long doubling times and specialized growth conditions, lower cryo-ET throughput and thereby limit the set of parameters that can be explored. A special consideration for the GroEL/ES system in the context of low resolution cryo-ET subtomogram averaging is the possibility of an internal control. A subset of archaea has both group I and group II chaperonin systems. The group II chaperonin thermosome consist of two eight membered rings stacked back-to-back, in contrast to the two seven membered rings in the GroEL:ES system. The ability to differentiate between these two systems even at low to intermediate resolution, would provide confidence in our discrimination of single and double capped EL:ES complexes.

Therefore, we explored the following three species: *Escherichia coli*, *Methanosarcina mazei* and *Caulobacter crescentus*.

E. coli is the one of the most wildly studied model organisms. Most previous work on the bacterial chaperonin system has been done in this model system. It's the most convenient organism for this study in terms of handling and available tools and can be frozen in layers of a few µm thickness, suitable for FIB milling. The buffer surrounding the cells vitrified, potentially due to the high concentration of macromolecules with cryoprotectant properties in LB medium.

The initial rationale for exploring *C. crescentus* was that, previously published tomograms seemed to have lower cytosolic density then what was observed in *E. coli*. In addition, there are established molecular biological tools available for this organism and growth is relatively fast under laboratory standard conditions.

M. mazei was chosen for its interesting biological properties. While these cells normally form dense colonies, under certain conditions, the cells can be separated into individual cells allowing for plunge freezing without the need for cryo protectants and challenging lamella preparation protocols such as cryo lift out. While there has been a lot of research into this organism it is still rather challenging to work with it in a cryo-ET workflow due to its anaerobic growth requirements. Therefore, we established a low-tech way to handle cultures purchased or provided from the DSMZ and ensure plunge freezing under close to anaerobic conditions for an initial exploration of this organism.

4.2.1.1 Caulobacter Crescentus

To compare the imaging properties of *C. Crescentus* to the ones in *E.coli* we first cultured cells purchased from the "Deutsche Sammlung von Mikroorganismen und Zellkulturen" (DSMZ) using a standard protocol (DSMZ 2023). Plunge freezing was attempted in a similar way to the one optimized for *E. coli*: first the cells are concentrated by sedimentation and then plunge-frozen in a highly concentrated slurry in, ideally resulting in a ~4 µm thick, carpet of bacteria that is suitable for FIB milling.

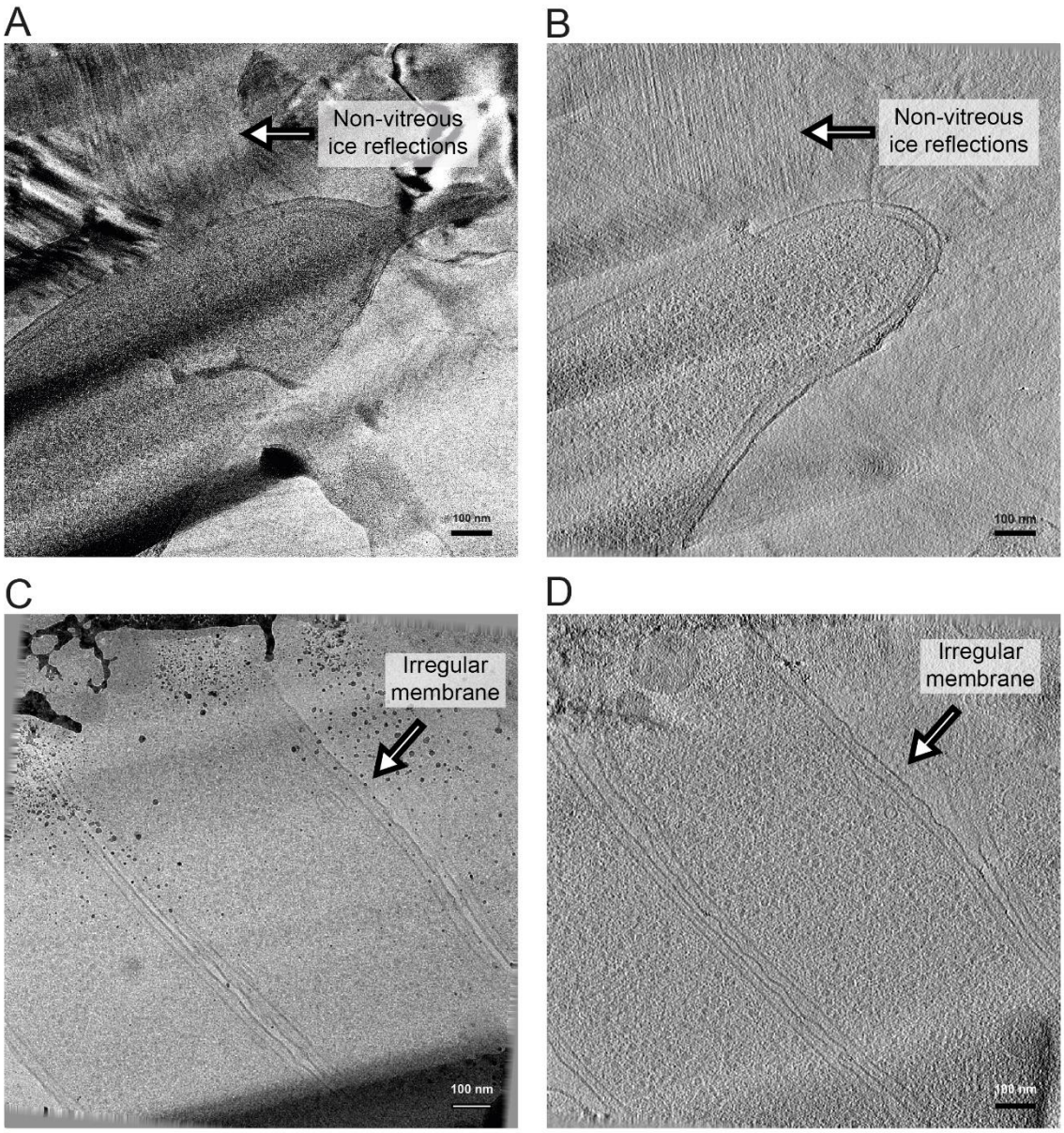

Figure 39: Cryo-ET of *C. crescentus.*

(A) Central tilt of a tilt series of *C. crescentus* plunge frozen in liquid ethane propane in growth medium without additives. The reflections around the cell correspond to non-vitreous ice that formed in the medium surrounding the cell. (B) The central slice of the tomographic reconstruction of the same tilt series shows clear distortions around the edge of the cells where the reflections from crystalline ice distorted the reconstruction. (C, D) Central tilt (C) and central slice of the tomographic reconstruction (D) of *C. crescentus* in the presence of 10 % glycerol. Upon the addition of 10% glycerol to the medium immediately before plunge freezing, the cells showed membrane deformations in the individual slices of the tilt series (C) as well as in the final tomographic reconstruction of a representativity tomogram (D), as indicated by the white arrows with black borders. The scale bar corresponds to 100 nm.

First attempts of cryo-ET of FIB milled *C. crescentus* are depicted in Fig. 39. One obvious initial problem was that the low macromolecular content of the growth medium results in patches of non-vitreous ice between cells. These regions, appearing as black

reflections across various tilts, create significant problems for the alignment of the tilt series. This, in turn, adversely affects the overall quality of the tomographic reconstructions, as depicted in Fig. 39.

To alleviate these issues, we tested if the addition of glycerol as a cryo protectant would prevent the formation of non-vitreous ice. This resulted in better vitrification, but led to the deformation of the outer bacterial membrane, indicating an osmotic event (Fig. 39 C and D). This pointed to a general problem with additives in the growth and plunging medium. Changes in osmolarity or composition could trigger a cellular stress response, altering the chaperonin cycle or prevalence. This would necessitate a careful characterization of the cellular stress response under the conditions used for plunge freezing with other omics techniques such as proteomics or transcriptomics, severely complicating the design of the study. Therefor *C. crescentus* was not further investigated as an alternative organism for this study.

4.2.1.2 Methanosarcina mazei

M. mazei is an archaeal species that garnered a scientific interest for its biotechnological applications in methane production as well as for its potential role in the Permian–Triassic extinction event. For this study the availability of genetic tools, established protocols for cultivation and commercial availability made it an ideal test case for a species containing both type I and II chaperonin systems side by side. On top of this *M. mazei* (DSMZ 108209) can be separated into individual cells by a simple change in culturing conditions making it accessible for plunge freezing (Xun, Boone et al. 1988).

For initial testing we established a low-tech workflow for concentrating and plunge freezing *M. mazei* under near anaerobic conditions (for details see Material and Methods 6.16).

FIB milling of *M. mazei* was performed similar to what was described before for mammalian cells (Mahamid, Pfeffer et al. 2016). A first set of cryo electron tomograms showed that while the workflow per se was suitable and the cells amenable to vitrification and FIB milling, the resulting tomograms had the highest cytosolic density of macromolecules that the authors of this work have ever observed (Fig. 40). An initial test run trying to identify cellular ribosomes, the complex most amenable for STA, by template matching showed a significant higher false positive and false negative rate than the one observed for *E. coli* (Fig. 40). The rates for both types of error were estimated by manually comparing the score maps from template matching with the

underlying tomogram. Due to these challenging imaging properties *M. mazei* seemed unsuited to the established cryo-ET STA workflow at least at the current state of technical development.

Therefor we focused our efforts on *E. coli* with its combined unmatched availability of molecular biology tools and acceptable imaging properties.

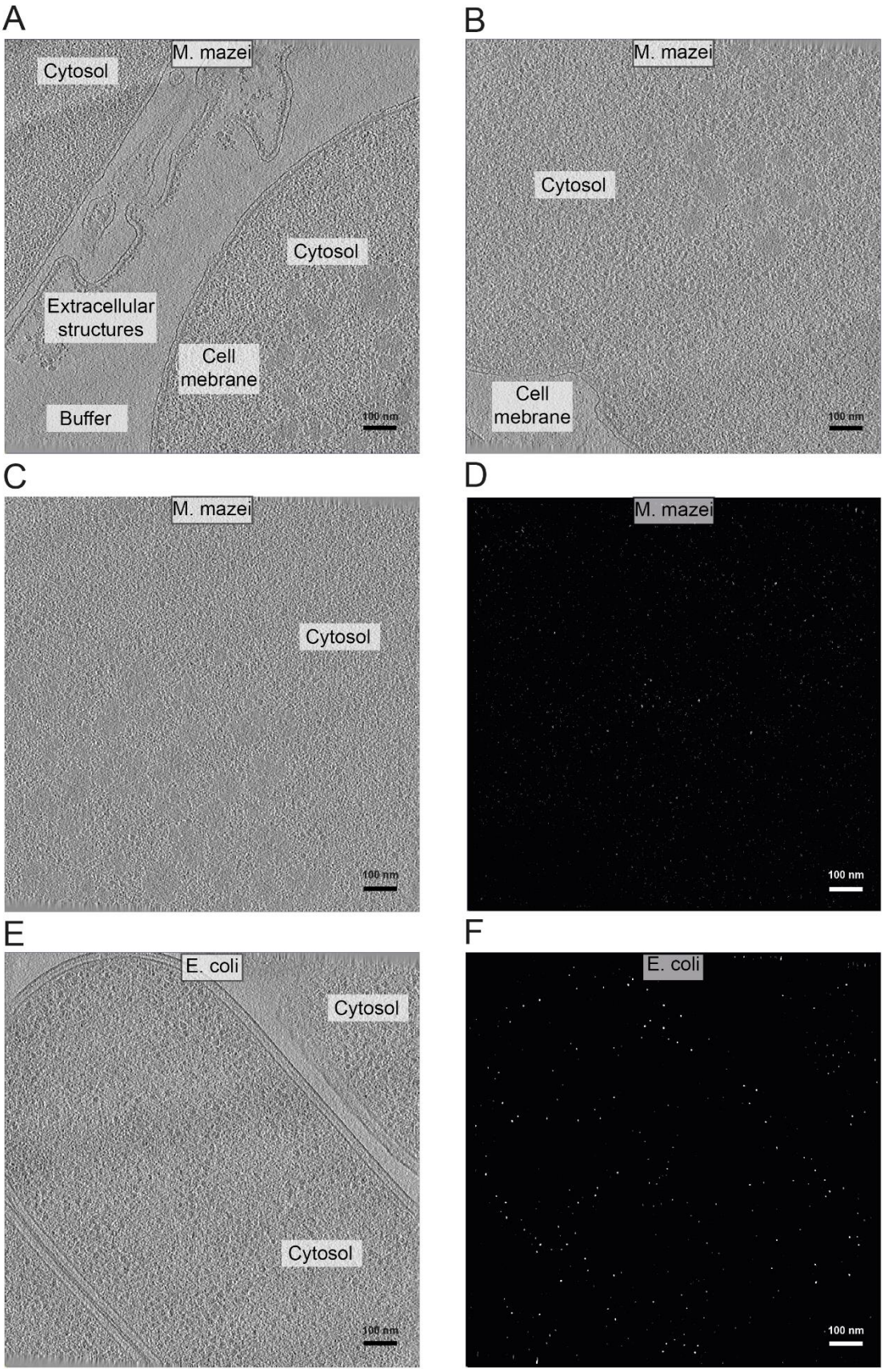

Figure 40: Cryo-ET of *M. mazei.*

(A, B) Tomographic reconstructions of tilt series acquired on FIB milled *M. mazei* cells. The cytosolic volume of *M. mazei* shows a very high density of macromolecules of unknow identity. Very dark regions with irregular shapes similar to what has been described as

glycogen storage granules in yeast (Prats, Graham et al. 2018) were observed. (C, D) Tomogram and cross correlation map of the ribosome after template matching. The tomogram was acquired in the middle of the cell body (C) and the corresponding score map results from template matching for cytosolic ribosomes. Withe spots indicate putative ribosomes (D). The high degree of background and bad separation of the peaks indicates that picking even ribosomes with high accuracy would be challenging. (E, F) *E. coli* tomogram and cross correlation map of the ribosome after template matching. The central slice of a tomogram of roughly equal thickness from a *E. coli* cell (E) and the corresponding score map form template matching for ribosomes (F). The stronger separation of peaks and background shows that template matching on these cells works seemingly better. All scale bars correspond to 100 nm.

4.2.2 *In situ* stoichiometries of different GroEL:ES complexes

To better understand the GroEL:ES reaction cycle *in vivo* we characterized the different stoichiometries and structures of the GroEL:ES complex under different cellular conditions. As a base line we imaged exponentially growing *E. coli* BL21(DE3) cells (hereafter 37 °C) at 37 °C in LB medium in the early exponential phase (OD_{600} 0.1-0.2) by cryo-ET. To characterize possible shifts under stress conditions we cultured *E. coli* BL21(DE3) cells at 46 °C for 2 hours prior to plunge freezing (hereafter HS). This leads to an upregulation of both GroEL and GroES and a higher, unspecific substrate load. The cells were still exponentially growing under these conditions and the abundance of EL:ES increases by ~3 fold compared to wild type by iBAQ mass spectrometry (Fig. 41 A-C and Fig 42 A).

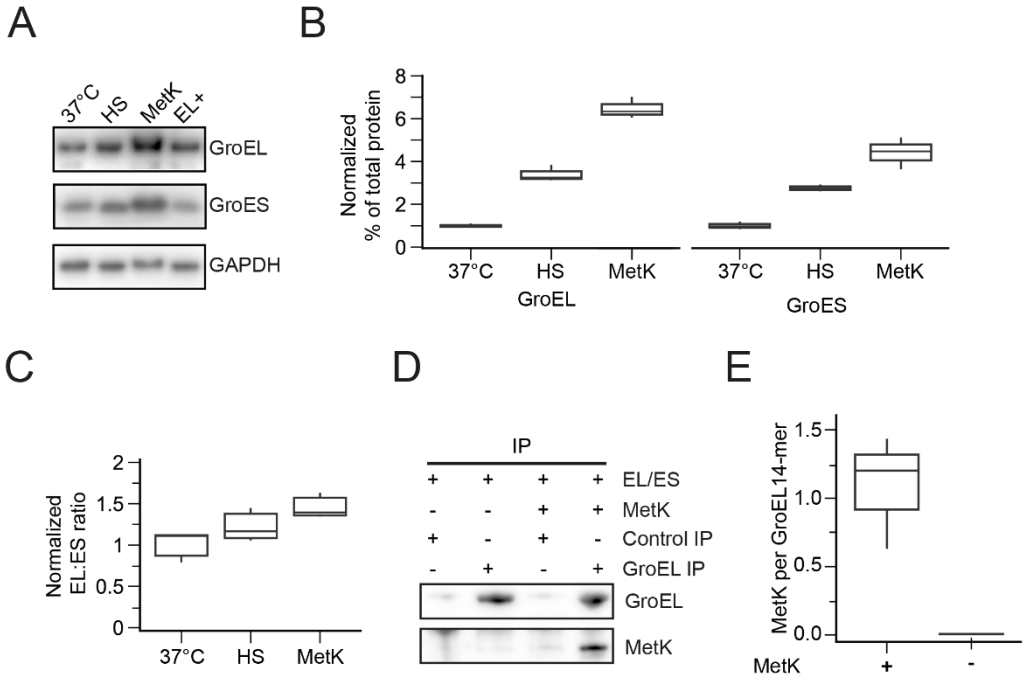

Figure 41: Biochemical characterization of EL:ES complexes.

(A) GroEL and GroES levels by western blot. Displayed here are representative immunoblot for GroEL and GroES under various growth conditions (37°C, HS, MetK, and EL+). Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) serves as the loading control. (B) GroEL and GroES quantification by MS. GroEL and GroES levels were quantified through label-free mass spectrometry analysis of cell lysates. The iBAQ values of GroEL and GroES from cells grown at 37°C were set to 1 and used for data normalization (n=3 independent replicates). In the boxplots, the median value is marked by the horizontal line; the boxes designate the upper and lower quartiles while the whisker caps denote the largest or smallest values within 1.5 times the interquartile range above the 75th percentile or below the 25th percentile, respectively.
(C) GroEL GroES ratios under different conditions. This chart represents the ratio of GroEL 14-mer to GroES 7-mer, based on the iBAQ values in (B). The EL:ES ratio in wild-type cells at 37°C was assigned a value of 1 and used for data normalization. (D) Quantification of MetK enrichment on GroEL by western blot. Quantification of the amount of MetK bound to GroEL complexes in cells at 37°C and in MetK overexpressing cells. Apyrase treatment was implemented post-cell lysis to halt the GroEL cycling. GroEL was then immunoprecipitated (IP) and subjected to immunoblotting with antibodies against GroEL and MetK. Antibodies against lactalbumin (LA) were used as a non-specific control. (E) Quantification of MetK enrichment on GroEL by MS. MetK:GroEL stoichiometry was quantified by MS in GroEL immunoprecipitations from (D). The fraction of MetK per GroEL 14-mer was calculated based on iBAQ values (n=3, independent replicates).

To identify if potential shifts were due to the substrate load or to changes in the cellular environment, we observed a third condition where GroEL:ES was mildly overexpressed under an arabinose inducible pBAD promotor. Subsequently, an obligate substrate protein, MetK, was overexpressed significantly under the viral T7 promoter. This would theoretically saturate the folding capacity of the ELS complex. We confirmed the enrichment of the EL:ES-MetK complex by IP of EL, followed by a western blot against MetK (Fig. 41 D) and quantified the ratio iBAQ mass spectrometry. We observed that ~1.3 MetK molecules were bound per GroEL complex on average, corresponding to >50% of GroEL rings containing MetK (Fig. 41 E).

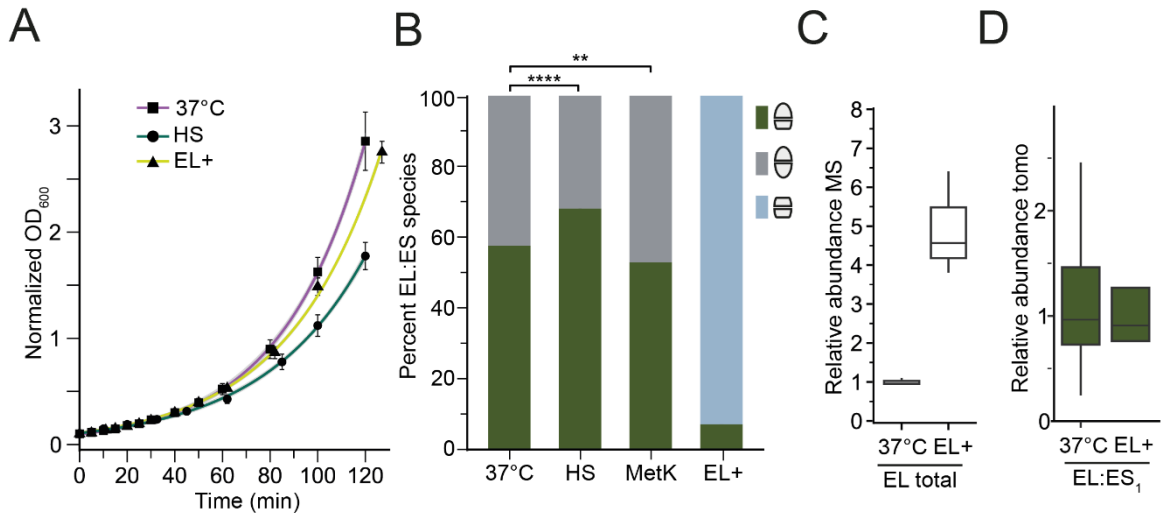

Figure 42: Growth curves of *E. coli* under different conditions and relative abundance of EL:ES complexes *in situ.*

(A) Exponential growth curves. Illustration of the growth of *E. coli* BL21(DE3) under different conditions: at 37 °C, under heat stress (HS) at 46 °C, or with GroEL overexpression (~4.5-fold, denoted as EL+) at 37 °C. The data points represent averages and the error bars show standard deviation (SD) for n=3 independent replicates. The light gray shadow delineates the variance of the fitted exponential growth curve. The growth curves have been standardized to a common starting point at an OD_{600} of 0.1. (B) EL:ES stoichiometries. The chart displays the relative abundance of EL:ES_1 (green), EL:ES_2 (gray), and EL (light blue) complexes in tomograms derived from cells grown under the aforementioned conditions, as well as under conditions of MetK overexpression at 37 °C (designated as MetK). Significant differences in relative abundance have been statistically validated, with p-values (derived from Wilcoxon rank sum test) of $4.9 \cdot 10^{7}$ for HS relative to 37°C and 0.0071 for MetK relative to 37°C. P-values were not adjusted for multiple testing (37°C, n=48; HS, n=58; MetK, n=60). (C) MS quantification of GroEL levels in EL+ cells. The graph showcases the cellular abundance of GroEL in cells at 37 °C and in EL+ cells. The total quantity of GroEL was assessed through label-free mass spectrometry using intensity-based absolute quantification (iBAQ). The data was normalized to a median value of 1 for 37°C (n=3, independent replicates). (D) Comparison of EL:ES_1 concentrations. The graph depicts the cellular abundance of EL:ES_1 in 37 °C and EL+ cells. The abundance of EL:ES_1 relative to ribosomes in tomograms was calculated to serve as an approximation for its cytosolic concentration, and was normalized to a median value of 1 for 37 °C.

Finally, we wanted to investigate whether the stoichiometry of different EL:ES complexes is governed by mass action or an intrinsic feature of the GroEL:ES recation cycle. Due to the operon organization of the groES and groEL genes (groELS) (Fayet, 1989 #167), GroES (7-mer) and GroEL (14-mer) are expressed at a ratio of 2:1.

For this we used another cellular system where GroEL was overexpressed, while ES levels remained endogenous (hereafter named EL+). This led to an imbalance in the relative abundance of both complexes, as shown by western blotting (Fig. 41 A). These cells were growing similar to wild type cells (Fig 42 A), while expressing significantly more GroEL (Fig. 42 C).

As a baseline we imaged 37°C cells. We vitrified these cells on EM grids, reduced their thickness via cryo-FIB milling, and subsequently imaged them in 3D using *in situ* cryo-ET (Fig. 44 A, B). For a detailed description of the process see Material and Methods 6.16 to 6.18.

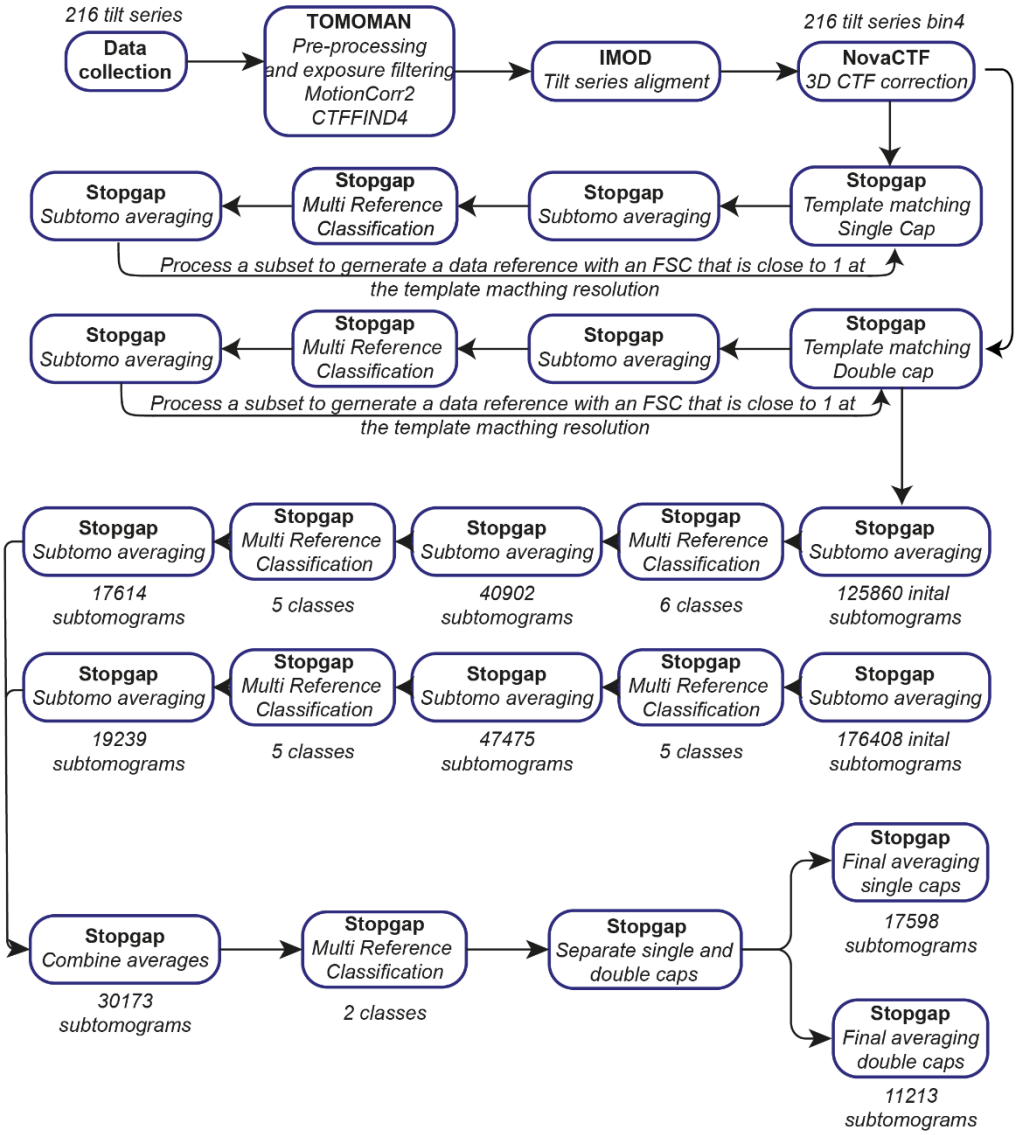

Figure 43: Workflow schematic for subtomogram averaging of EL:ES *in situ.*

Processing flowchart used for EL:ES$_1$ and EL:ES$_2$ subtomogram averaging *in situ*. See Materials and Methods 6.18-6.19 for details.

Utilizing template matching with both EL:ES$_1$ and EL:ES$_2$ references, followed by classification (Fig. 43 and Material and Methods for details), we were able to reveal the relative proportions and cellular distribution of these complexes. We found an EL:ES$_1$ to EL:ES$_2$ ratio of approximately 58% to 42% (Fig. 42 B) within these wild type cells. To ensure the validity of our template matching results, we compared the numbers of GroEL complexes we identified against the count of ribosomes, which are consistently discernible in cryo-ET due to their unique shape, size and high contrast (Hoffmann, Kreysing et al. 2022, Xue, Lenz et al. 2022, Xue, Spahn et al. 2023). We localized and

averaged virtually all cellular ribosomes to a resolution of about 8.7 Å (Fig. 44 C-E), revealing a GroEL-to-ribosomes ratio of 1:26 in wild-type cells at 37 °C (Fig. 44 E). The ratios from our intensity-based absolute quantification (iBAQ) mass spectrometry analysis were in good agreement with those from cryo-ET (Fig. 44 E), underscoring that our cryo-ET approach successfully identified the majority of GroEL complexes.

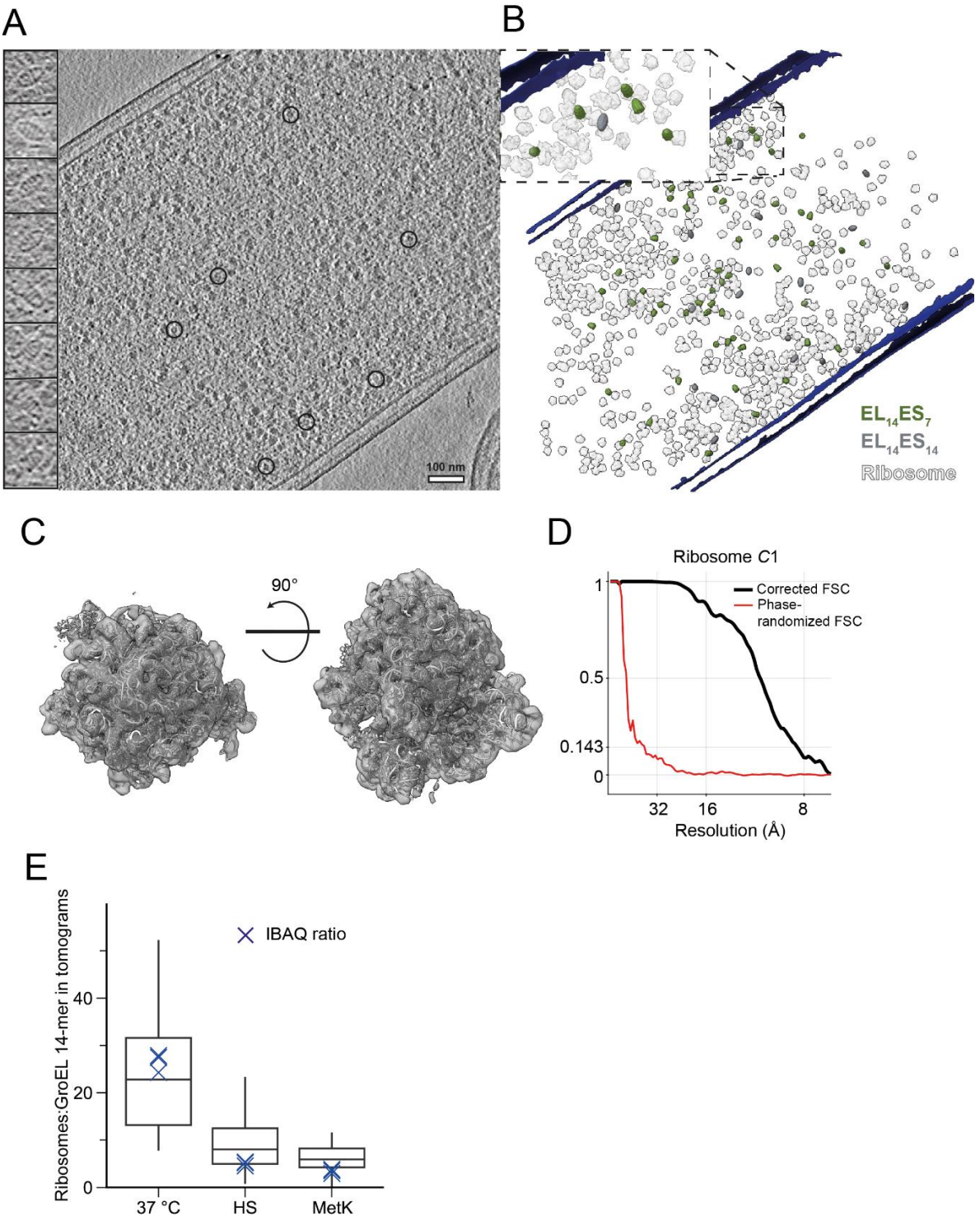

Figure 44: Cryo-ET of a *E. coli* cell and the localization of various macromolecular complexes.

(A, B) Tomogram with highlighted macromolecules and corresponding 3D rendering. (A) On the right: A representative z-slice from a tomogram of an *E. coli* cell subjected to HS is shown. EL:ES complexes are highlighted with black circles. The scale bar denotes 100 nm. On the left: A collection of central subtomogram slices is presented, displaying EL:ES_1 and EL:ES_2 complexes in a side view. (B) A 3D depiction of EL:ES_1 complexes (shown in green), EL:ES_2 complexes (depicted in silver), and ribosomes (represented in light gray) derived from the tomogram in (A). The cell membranes are illustrated in dark blue.
(C, D) STA of cellular ribosomes. Ribosomes from three different datasets (37°C, HS, MetK) underwent an averaging and refining process to achieve a global resolution of 8.7 Å. The resulting structure from the subtomogram and the overlaid model (PDB 4v4a (Vila-Sanjurjo, Ridgeway et al. 2003)), rendered in ribbon format, are illustrated (C). The corresponding Fourier Shell Correlation (FSC) curve is depicted (D).
(E) Estimation of template matching accuracy. Analysis of the ratio of ribosome to GroEL 14-mer in tomograms is represented by box plots (37°C n=48, HS n=58, MetK n=60 tomograms), and by MS indicated by blue crosses (n=3, independent repeats). In the boxplots, the median value is represented by the horizontal line; the boxes signify the upper and lower quartiles, while the whisker caps denote the 10th to 90th percentile, respectively. The MS measurements primarily fall within the range of the first to third quintiles of the tomographic data, suggesting that most EL complexes were identifiable *in situ*.

Certain environmental stimuli can cause protein unfolding, thus increasing the demand for cellular chaperone systems. The GroEL system's holdase function, which binds unfolded proteins preventing their aggregation, and its ability to refold these proteins, become particularly relevant during such stress periods. Holdase activity does not necessitate GroES binding and encapsulation, whereas refolding does. We probed whether a higher demand for the GroEL system due to heat stress would affect the single to double cap ratio. HS cells were plunge-frozen onto EM grids, revealing a single to double cap ratio of about 68% to 32% (Fig. 42 B). The cells were still exponentially growing under these conditions and the abundance of EL:ES increases by ~3 fold compared to wild type mass spectrometry using iBAQ values (Fig 41 A-C and 42 A).

To distinguish if this shift in single and double capped EL:ES complexes was due to the heat stress or the increased substrate abundance, the previously described MetK cells were imaged. Under these conditions, the single to double cap ratio was approximately 53% to 47% (Fig. 42 B). Differences in relative abundance are statistically significant with p-values (Wilcoxon rank sum test) of $4.9 \cdot 10^{7}$ for HS relative to 37 °C and 0.0071 for MetK relative to 37 °C. p-Values were not corrected for multiple testing (37°C, n=48; HS, n=58; MetK, n=60 tomograms).

Next, we investigated whether the double-capped EL:ES complex is not just sufficient but also necessary for complete reaction cycling. In the EL+ cells we imaged, only free

EL and EL:ES_1 complexes were observable by cryo-ET (Fig. 42 B), but no EL:ES_2. Importantly, subtomogram analysis revealed a similar abundance of single capped EL:ES complexes relative to ribosomes as in wild-type cells at 37 °C (Fig. 42 D). Thus, the increased GroEL abundance disfavors the formation of EL:ES_2 while supporting normal cell growth. This indicates that asymmetrical complexes are functionally active. In conclusion, EL:ES_1 and EL:ES_2 complexes coexist *in vivo*, with EL:ES_1 predominating over EL:ES_2 under various growth conditions, including high substrate protein (SP) load. Elevation of GroEL levels relative to GroES shifts the capped chaperonin population to only EL:ES_1, suggesting that EL:ES_2 formation is not inherently favored and is instead governed by mass action.

4.2.3 *In situ* structures of GroEL:ES complexes of varying stoichiometries

After subtomogram averaging, we reconstructed and superimposed structural models of the EL:ES_1 complex and the EL:ES_2 complex at an estimated resolution of 12 Å (symmetry applied) (Fig. 45 A-D).

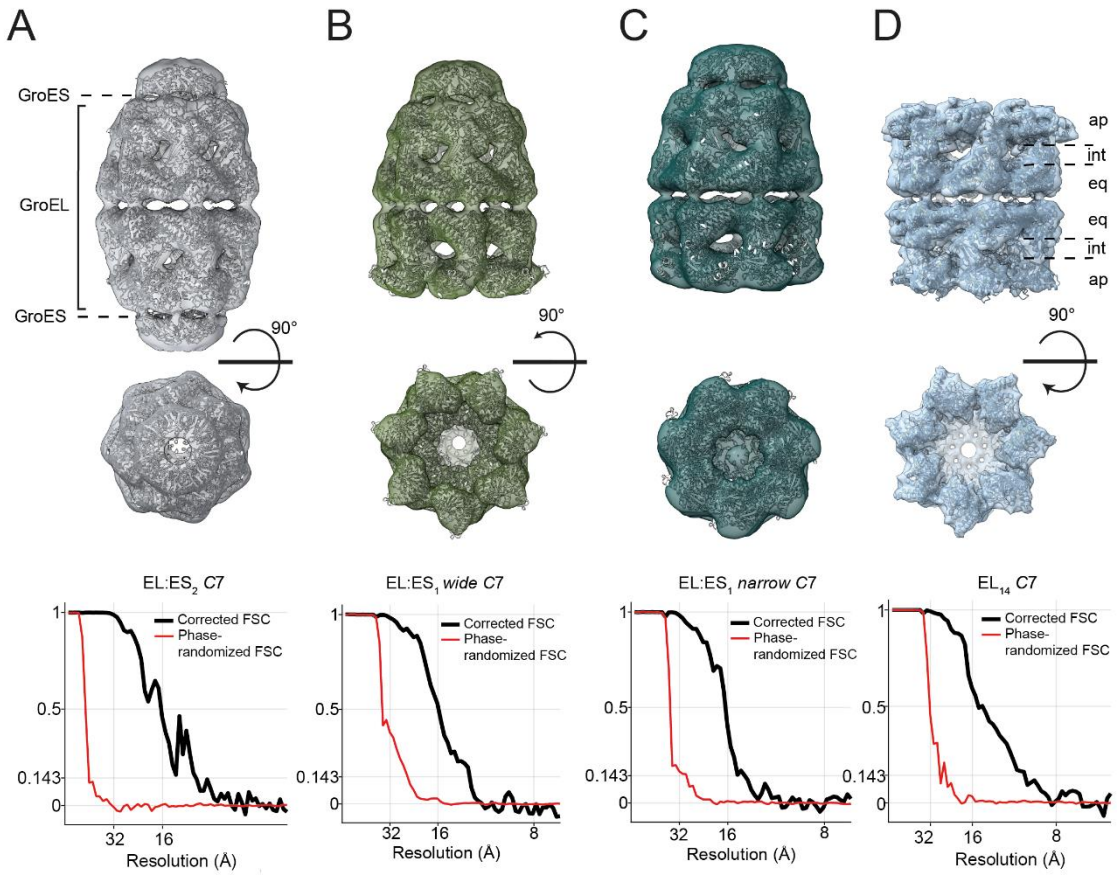

Figure 45: Subtomogram averages of different EL:ES species.

(A-D) Subtomogram averages of EL:ES_2 (A), EL:ES_1 wide (B), EL:ES_1 narrow (C) and EL_{14} (D) complexes (with symmetry applied) at an approximate resolution of ~10-13.5 Å. Both side and end views are shown. Ribbon representations of the models for open EL:ES_1, EL:ES_2, and open EL_{14} (with PDB codes 8P4S, 8P4R, and 8P4P, respectively) are superimposed. The EL:ES_1 and EL:ES_2 complexes originate from tomograms of cells grown at 37 °C, subjected to HS, and upon MetK overexpression. EL_{14} complexes are derived from tomograms of cells with GroEL overexpression (EL+). The positions of GroES and GroEL rings are indicated, in addition to the locations of the apical (ap), intermediate (int), and equatorial (eq) domains within GroEL rings. The corresponding FSC curves are shown below.

After sub-classification of the EL:ES_1 complexes based on their *trans*-ring opening, we identified the same two complexes that were also present in the SPA sample: EL:ES_1 wide and narrow with the former one being the predominant species, an effect most pronounced under HS conditions (Fig. 46).

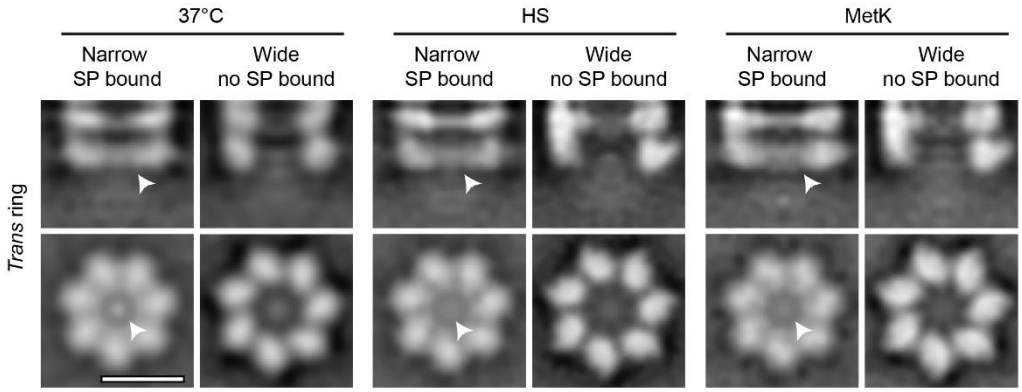

Figure 46: Different EL:ES_1 species.

Top view and central slice of a side view of the wide and narrow EL:ES_1 complexes. Both species are present under all tested cellular conditions. EL:ES_1 complexes with a narrow opening display addition density at the approximate location of the substrate binding cleft formed by the helixes αI and αH, indicated with a white arrow head. Scale bar: 10 nm.

The *trans*-ring of the single-capped EL:ES complex, which is uncapped by GroES, showed considerable variation from each other with regard to the position of the equatorial and apical domains of the *trans*-ring (Fig. 47 A-C). The narrow structure closely aligns with a previously reported 3.4 Å cryo-EM structure of a single-capped EL:ES complex with ADP bound in both rings as discussed in Chapter 4.1.2 (Kudryavtseva, Pichkur et al. 2021). It has been interpreted as an intermediate phase in the asymmetrical chaperonin cycle post-ATP hydrolysis in the *cis*-ring and prior to ATP association with the *trans*-ring (Fig. 45 A-C). Importantly, this unique orientation of the apical domains significantly enlarges the *trans*-ring cavity opening, with the inner diameter expanding from ~50 Å to ~60 Å (Fig. 45 A-C). This expanded aperture could

simplify the release of larger SPs, like the folded state of MetK (42 kDa; roughly 70 x 60 x 30 Å in size).

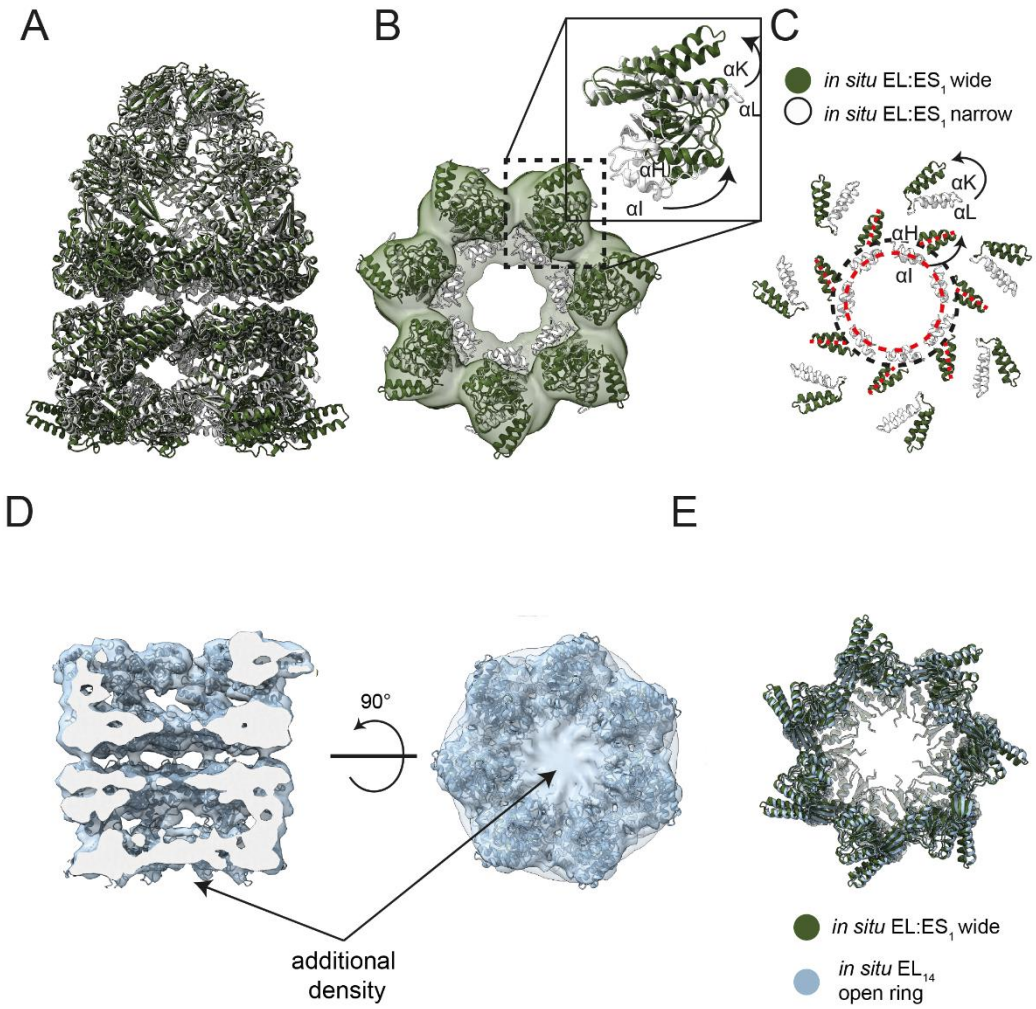

Figure 47: Comparison of the *in situ* EL:ES$_1$ complex structures and the EL$_{14}$ structure.

(A, B) The *in situ* structure of EL:ES$_1$ wide (dark green) is superimposed on the EL:ES$_1$ narrow conformation (white) using least square fitting of the equatorial domains conducted using ChimeraX. Structures are presented in ribbon representation as side (A) and top views (B). The cryo-ET density of the EL:ES$_1$ wide complex is represented in green in (B). An inset in (B) emphasizes the reorientation of the apical domains of the *trans*-ring. Shown are the SP-binding helices αI and αH along with the helical hairpin αL and αK. (C) Visualization of the ring opening expansion from ~50 Å to ~60 Å of the *trans*-ring of the *in situ* EL:ES$_1$ complex. Only the helices αI and αH as well as αK and αL are shown. The SP binding grooves are indicated by red dashed lines. (D) Structure of the EL$_{14}$ complex. Cross-section through the EL$_{14}$ complex shows additional density, not accounted for by the model, in the closed ring at the SP binding sites, as depicted in side and top view. No additional density at the defined contour level is visible in the opposite open ring. (E) Structural comparison. Overlay of the *trans*-ring of the *in situ* wide EL:ES$_1$ structure (dark green) and the open ring of the *in situ* EL$_{14}$ structure (light blue).

Conversely, the *in situ* structure of the EL:ES$_2$ complex presented no significant deviations from the published crystal structure of the non-cycling symmetrical complex

with ADP·BeF_x bound in both rings (PDB 4PKO(Fei, Ye et al. 2014)), showing an overall C-α root mean square deviation (r.m.s.d.) of 2.3 Å. Moreover, the *in situ* structure of GroEL alone, only found in the EL+ cells, was solved to a resolution of approximately 9.8 Å (Fig. 45 D and Fig. 47 C, D). It had a distinct ring conformation in contrast to the crystal structure of apo-GroEL (PDB 1SX3(Chaudhry, Horwich et al. 2004)) one ring resembles a conformation found in a structure of the slowly ATP-hydrolyzing GroEL mutant D398A (PDB 4AB3; Clare, 2012). In this structure both GroEL rings have nucleotide bound and in one ring the apical and intermediate domains are approximately 35º tilted, while the apical domains in the other ring show an additional 20º tilt (Clare, Vasishtan et al. 2012). This reorientation causes the positional difference of the apical domains between the rings (Fig. 47 D). The conformation of the more open ring mirrors that of the *trans*-ring of the wide EL:ES_1 complex discussed above (Fig. 47 E).

Interestingly, the ring in the EL_{14} complex with a continuous substrate binding cleft and a more closed conformation exhibits additional density at the apical domains, which is likely representative of symmetry-averaged, unfolded bound SP when GroES availability is limited (Fig. 47 D). Thus, the *in situ* EL_{14} complex exhibits intrinsic inter-ring asymmetry, probably because of the negative allosteric coupling between the rings, leading to preferential substrate binding to one ring.

4.2.4 Substrate folding *in situ*

To probe deeper into the substrate status within the EL:ES_1 complex and EL:ES_2 complex chambers *in situ*, we combined and analyzed all observed chambers (Fig. 48).

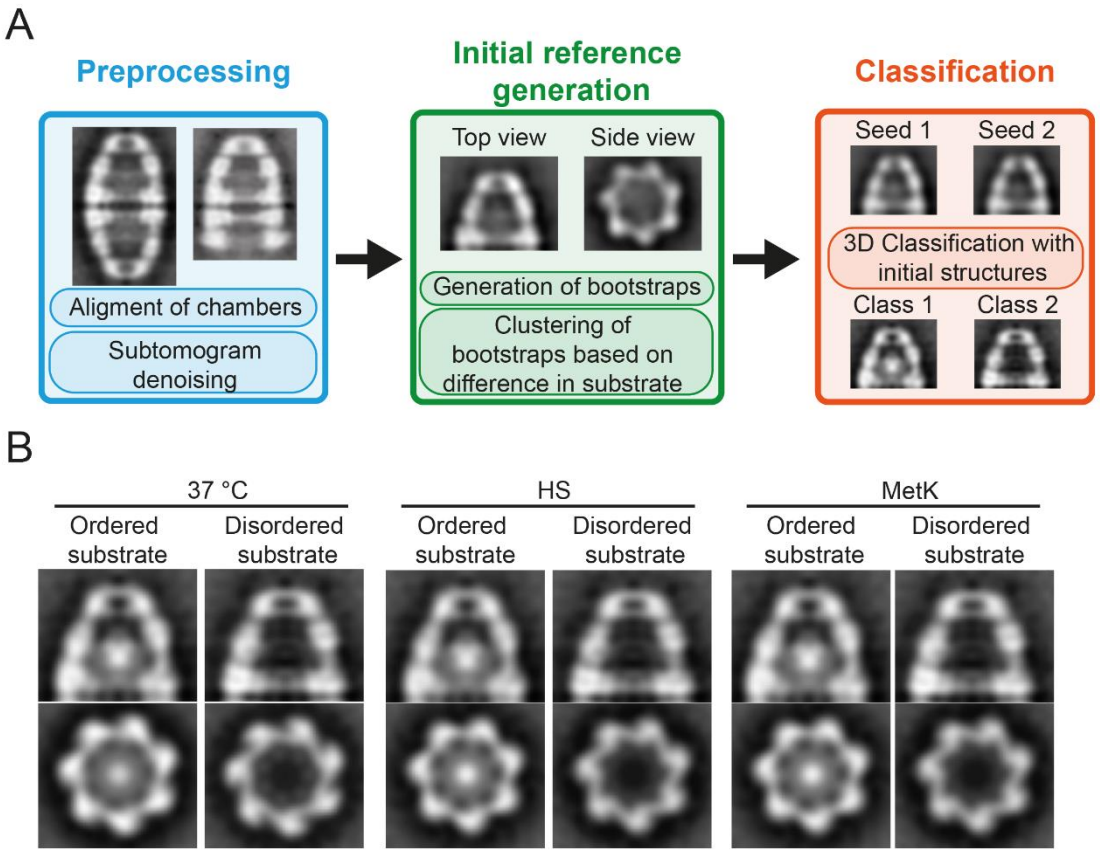

Figure 48: Averaging the substrate state in the *cis* chamber of GroEL

(A) Methodology for processing subtomograms to analyze encapsulated SP. To distinguish between different SP states in GroEL-ES chambers of EL:ES_1 and EL:ES_2 complexes, separate chambers were aligned. The resulting subtomograms were denoised, and preliminary structures for later 3D classification were generated via bootstrapping and k-means clustering. These averages served as starting points for 3D classification (for details, refer to Materials and Methods). (B) Slices through the *cis* chamber in different substrate states. Slice at the SP density level through subtomogram averages of GroEL:ES chambers from cells exposed to 37°C, HS, and MetK conditions. The process outlined in (A) yielded two unique classes of GroEL:ES chambers: one featuring a compact, localized density near the chamber base presumable representing ordered SP and another with a faint, delocalized density within the chamber. Upon separating the particles based on growth conditions (37 °C, HS, MetK), these two categories were identified across all three.

In all three experimental conditions, we identified two distinguishable classes: The first featured strong density in a location similar to MetK in our SPA structure and probably corresponding to the location of substrate in previously published EL:ES:Rubisco complex structures (Chen, Madan et al. 2013, Kim, Park et al. 2022). The second class exhibited only weak residual density within the chamber (Fig. 48 B).

Two extreme scenarios could explain these observations: One possibility is that a significant fraction of EL:ES chambers are empty under cellular conditions and the detected density represents both ordered and disordered proteins within the chamber. Alternatively, the observed classes might correspond to different folding states of the

client proteins themselves, with only highly ordered structures binding at a distinct location in the chamber manifesting as a density in EM.

To differentiate these possibilities we performed cryo-ET on isolated EL:ES:MetK complexes *in vitro* prepared using the same methodology as described above for SPA and in Material and Methods. The data was collected at the same magnification as for *in situ* cryo-ET and processed analogously. The classification yielded two classes of GroEL:ES chambers resembling those observed within cells *in vivo* (Fig. 49 A, B).

For comparison, we applied low-pass filtering to the SPA structures of ordered and unordered MetK within the GroEL-ES chambers described earlier and enforced 7-fold symmetry averaging onto these structures followed by down sampling to roughly the same resolution (Fig. 49 A, B). At present, it is not feasible to discriminate between empty EL:ES complexes and those with disordered substrates bound inside in *in situ* cryo-ET subtomograms, as these complexes appear nearly indistinguishable even under *in vitro* conditions. Previous MS data suggested a near 100% substrate occupancy of EL:ES (Kerner, Naylor et al. 2005), suggesting that a significant portion of complexes with weak density in the chamber may contain disordered SP.

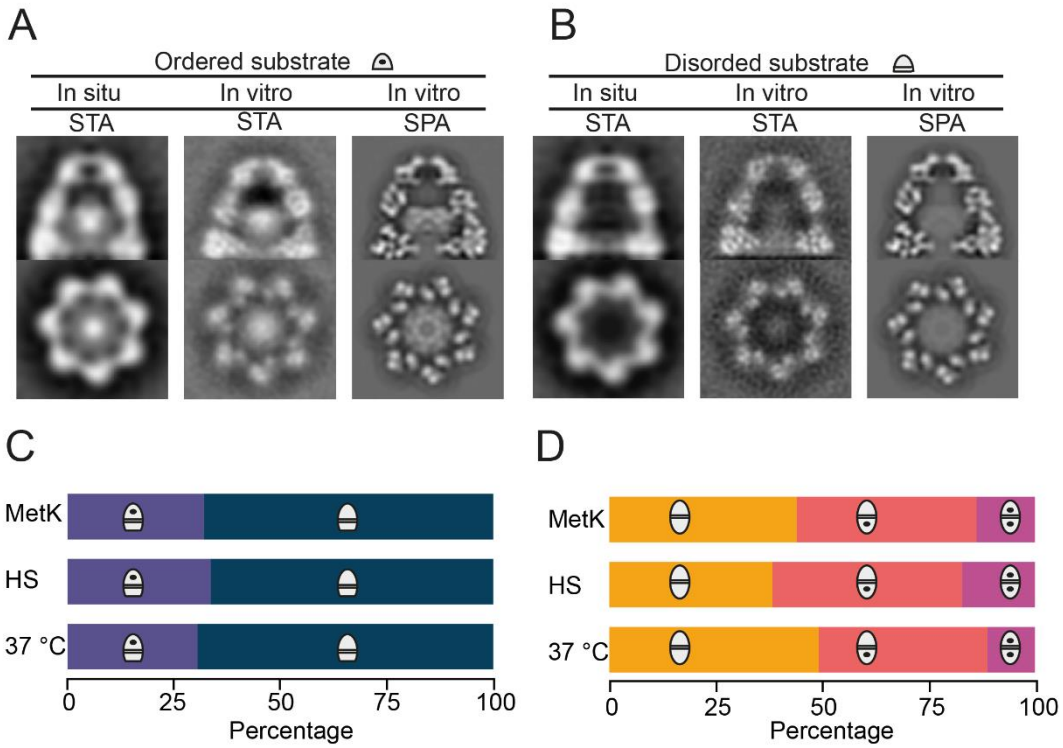

Figure 49: Substrate state of EL:ES *cis* chambers.

(A, B) Comparison of slices through *in situ, in vitro* and a down-filtered symmetrized SPA structure of EL:ES chambers from subtomogram averaging with ordered (left, A) and disordered (right, B) SP or MetK, respectively. Classification of the subtomogram averages

inside the chambers yielded two distinct classes for all conditions. One containing a strong density localized near the bottom of the chambers (A) and a second one only containing weak density (B). Shown are top and side views of the central slice each density. The gray values were normalized to a SD of 1 and a mean of 0 on the EL:ES chamber. (C, D) Quantification of EL:ES complexes of different stoichiometries. (C) Quantification of $EL:ES_1$ complexes with (purple) or without (dark blue) ordered SP in the GroEL:ES chamber for the experimental conditions indicated. (D) Quantification of $EL:ES_2$ complexes with two and one (dark orange and pink, respectively) or without (yellow) ordered SP in the EL:ES chamber. Pictograms indicate presence of ordered SP in one or two chambers.

The distribution of chamber classes with folded substrate proteins in $EL:ES_1$ and $EL:ES_2$ complexes for all three experimental conditions (Fig. 49 C, D), did not grossly change.

Sorting the $EL:ES_1$ and $EL:ES_2$ complexes in the *in situ* datasets according to the presence of encapsulated SP and/or bound SP and the *trans*-ring conformation of the $EL:ES_1$ complex allowed us to quantify the relative proportions of a total of seven different states of $EL:ES_1$ and $EL:ES_2$ complexes (Fig. 50). The relative population of these species was similar under cell growth at 37 °C (with or without MetK overexpression), with a subset of $EL:ES_2$ complexes containing structured SP in both chambers.

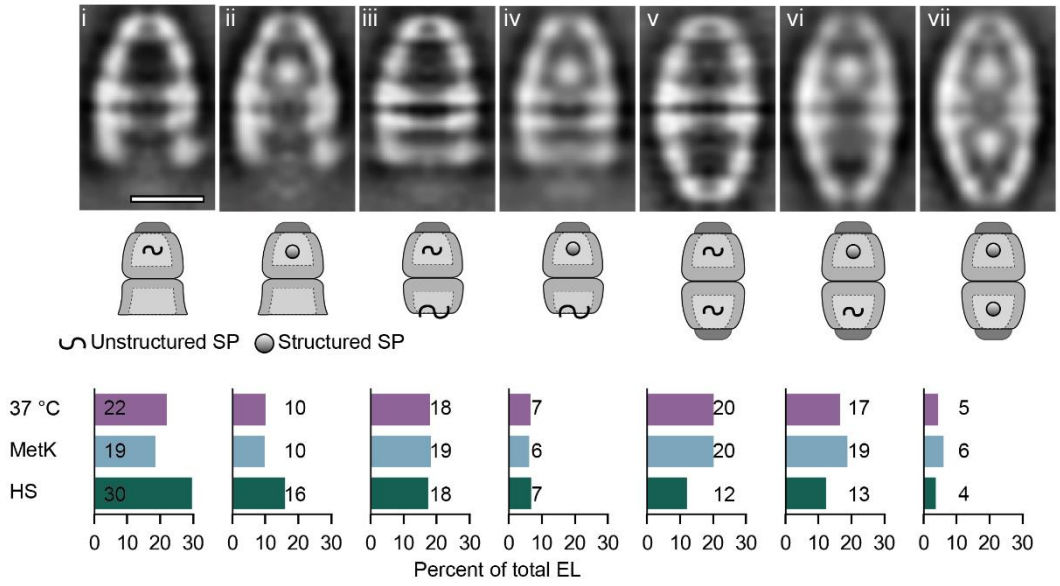

Figure 50: EL:ES substrate complex structures and quantification *in situ.*

Central slices (top) of all EL:ES species found *in situ* with different conformational states and SP occupancy, together with their relative fraction (bottom). Species (i) and (ii) are $EL:ES_1$ complexes with a *trans*-ring in the wide conformation and *cis*-rings with either disordered or no SP (i) or folded SP (ii). Species (iii) and (iv) are $EL:ES_1$ complexes with a *trans*-ring in narrow topology and *cis*-rings with either disordered or no SP (iii) or folded SP (iv). Species (v)-(vii) are $EL:ES_2$ complexes with either none (v), one (vi) or two (vii) folded SP.

Interestingly, upon HS, EL:ES_1 complexes with wide *trans*-ring conformation (no bound SP) were enriched (Fig. 50 i, ii) and EL:ES_2 complexes reduced (Fig. 50 v-vii), further suggesting that SP binding to the *trans*-ring facilitates EL:ES_2 formation.
Together these results define the chaperonin species that are populated *in vivo* and demonstrate that both EL:ES_1 and EL:ES_2 complexes are functionally active.

4.3 Analysis of polysomes in tomograms under selected conditions

In response to different growth conditions, cells are capable of forming multiple higher-order assemblies of macromolecules. These can either serve to up or down-regulate the activity of certain macromolecular machines. A prime example for this is the polysomes. Translating ribosomes form higher ordered assemblies called polysomes. Polysomes, or polyribosomes, are clusters of numerous ribosomes attached to a single messenger RNA (mRNA) molecule, collaboratively translating the genetic information within the mRNA into protein. It is theorized that several factors, including mRNA availability, cellular needs, and translation factors, regulate the formation and maintenance of polysomes. Furthermore, the activity of certain proteins, such as ribosome-associated chaperones, may also contribute to polysome assembly and maintenance. Yet, the exact processes underpinning polysome formation and maintenance remain unclear, necessitating further research for a comprehensive understanding.

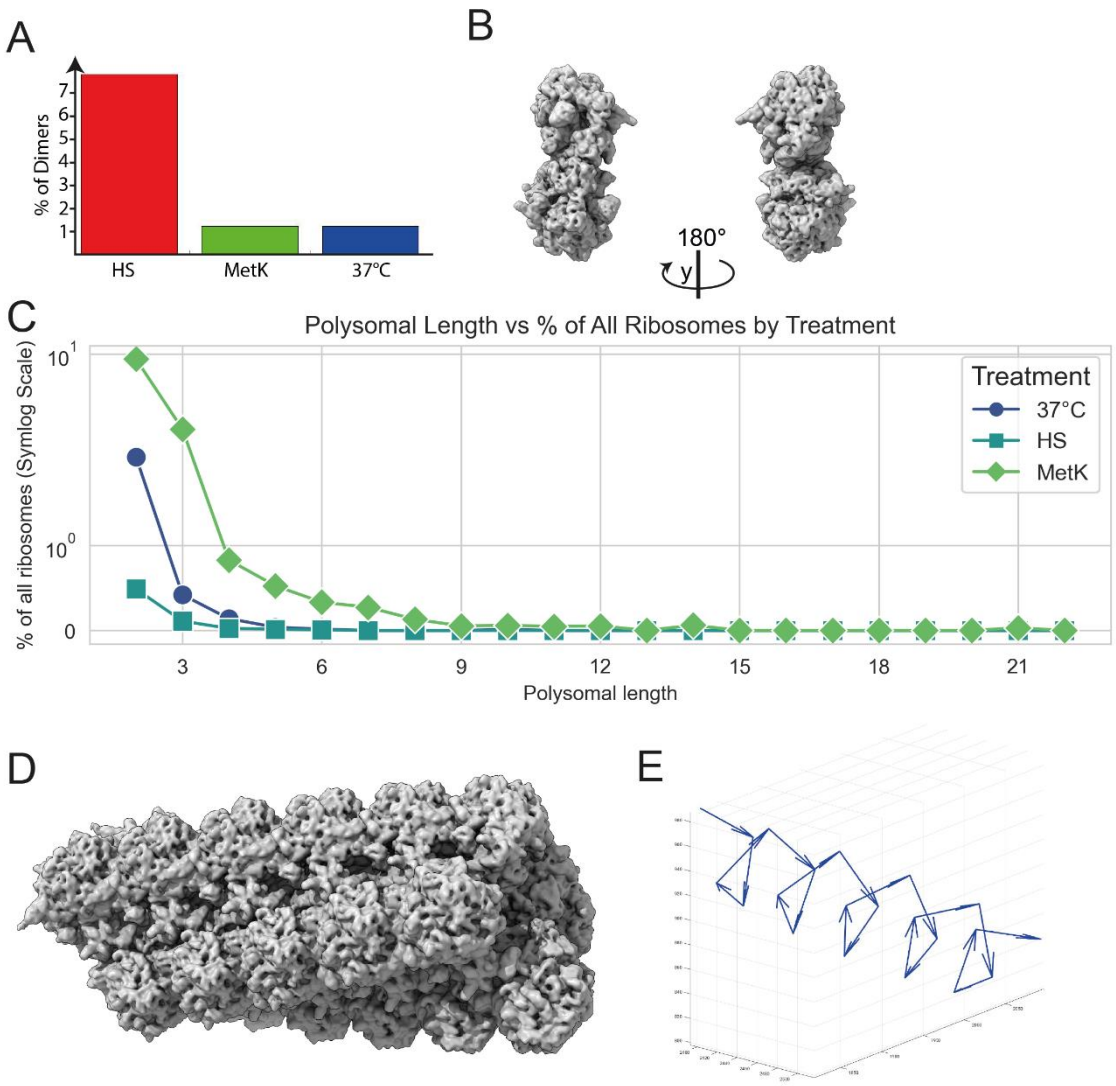

Figure 51: Quantification of different ribosomal assemblies and their spatial relationship with EL:ES.

(A) Quantification of ribosome dimers under different conditions. During heat stress a significant proportion of the cellular ribosomes form dimers with a fixed angle between them. (B) STA density of a ribosome dimer. Presumably all these ribosomes are translationally inactive hibernating ribosomes. (C) Quantification of the number and length of polysomes as a percentage of total cellular ribosomes. Heat stress significantly decreases the length as well as the number ribosomes in a polysomal orientation to each other, whereas the expression of MetK from a T7 promotor increases the length and abundance of polysomes. Data is presented on a symlog scale. (D, E) STA structure of a polysome of length 20. The ribosomal averages is shown at each position of the polysome found to be occupied at the orientation determined by STA. The 3D organization is shown in (E).

Contrarily, hibernating ribosomes, also known as 100S ribosomes, form in bacteria when two non-translating 70S ribosome copies associate into a translationally inactive dimer during periods of cellular stress. This prevalent bacterial adaptation likely serves a dual function in unfavorable conditions: conserving energy by downregulating

translation rates in the cell, and preserving these elaborate, energetically expensive cellular machines for swift reactivation when conditions revert back to normal. We utilized these two assemblies as local molecular markers for active translation (polysomes) or its absence (100S ribosomes). We identified all polysomes and hibernating 100s ribosomes using NEMO-TOC-based hierarchical transformation clustering (Jiang, Wagner et al. 2022). We then analyzed polysome length and abundance in the studied *E. coli* strains and growth conditions.

The abundance and length of polysomes reflected the anticipated pattern across the three conditions: During heat stress, when translation is suppressed, only a few short polysomes remain. They lengthen and their abundance increase during normal growth conditions and again they become significantly longer and more plentiful compared to normal growth at 37 °C when overexpressing MetK (Fig. 50 C-E). Length estimations of the average polysomes are limited by cryo-ET. Only a small portion of the cell remains after FIB milling.

The abundance of the hibernating 100S dimer displays the inverse pattern. Under heat stress, a substantial portion of all ribosomes (~ 8 %) adopts this state, while during normal growth or overexpression, only a few remain (~ 1 %) (Fig. 50 A, B). This substantiates the computational method's validity and confirms the cells expected cellular state.

4.4 Spatial relations between translation and protein folding by GroEL/ES

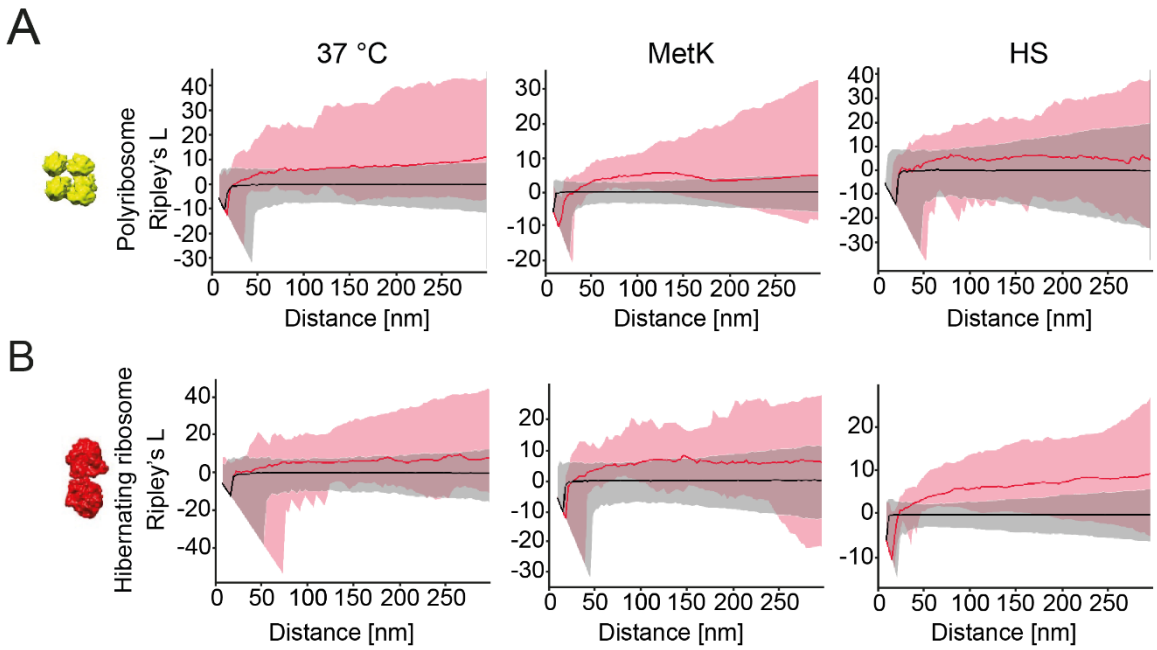

Figure 52: Spatial relationship of ribosome complexes and GroEL/ES under varying conditions.

(A, B) The red arrear shows the frequency of a relative distance between the two complexes found in the different tomograms whereas the grey shaded area corresponds to the simulated likelihood of a distance occurring under the assumption that all molecules are randomly distributed in the cellular volume. (A) During normal growth conditions we found that GroEL is slightly more likely to localize near a polysome than in the simulation. In cells overexpressing MetK there is a pronounced peak between ~50-150 nm from the polysome, indicating a preferential localization near the polysome upon overexpression of an obligate substrate. During heat stress all associations are abolished. (B) Testing for colocalization with hibernating 100S ribosomes on the other hand shows the inverse pattern. GroEL is preferentially localized closer to the 100S hibernating ribosomes than it should be when a randomly distributed, while there is no co-localization with the remaining hibernating ribosomes during either normal growth or MetK overexpression.

In additional to determining the *in situ* structures of macromolecules, cryo-ET can probed the spatial distribution of these molecules and their interactions subcellular organization and interactions. The EL:ES system is thought to be acting predominantly post translationally. We wanted to probe if there is a spatial correlation between EL:ES assisted folding and protein translation.

We observed no significant colocalization of polysomes with EL under normal cellular conditions and only a slight association during overexpression of the obligate substrate MetK (Fig. 52 A, B). This is consistent with previously published mass spectrometry data (Zhao, Castanie-Cornet et al. 2021), which indicated that native proteins are

initially managed by the trigger factor or the DnaK/J system, after which a subset of the client proteins is transferred to the EL:ES machinery for folding.
A thorough investigation on a bigger dataset would be needed to determine if the different substrate states and folding species of the EL:ES system (Fig. 50) show a more differentiated patter of co-localization.

5 Discussion

In this study we combined SPA and cryo-ET/STA to further understand the workings of chaperonins. Our results clarify several central aspects of the chaperonin mechanism:

1) The structural features of protein folding in the EL:ES chamber and the interactions of the substrate molecule with the EL:ES chamber in atomic detail *in vitro*.
2) The functional stoichiometry of GroEL and GroES, as well as the location of the folding reaction *in situ*.
3) The *in situ* structures of EL:ES_1 and EL:ES_2 complexes, as well as GroEL alone, all of which inform on distinct steps in the chaperonin reaction cycle.
4) Translational coupling with folding by the EL:ES machinery is weak to non-existent under normal conditions and only slightly more apparent when an obligate substate of the EL:ES machinery is overexpressed.

5.1 Selection of model system

Let us first discuss the decision on choosing *E. coli* as the model system. Over the past three decades *E. coli* has been the model organism of chois to study chaperonin function, as shown by the number of publications and structures deposited in PDB. This robust body of literature and previous experience simplifies experimental design and evaluation of results. Over these the amount of established biochemical tools for *E. coli* in the Hartl lab is unmatched, allowing for quick and easy experimentation. Even though previous work has shown that *C. crescentus* has good imaging properties in the context of single cell imaging, problems with ice quality and culture density made is less attractive for FIB milling, where the generation of a homogeneous carpet 2-3 layers thick is of advantage. The unique biology of *M. mazei*, characterized by the presence of both group I and group II chaperonin systems, initially appeared ideally suited for this project. The two different kinds of chaperonins could have served as an internal control, promising a robust characterization of the chaperonin system even at low resolution. However, despite the potential advantages, the complex imaging properties of *M. mazei*, compounded by the increased challenges in handling and genetic manipulation, ultimately rendered it unsuitable for our purposesWhile the exploration of further organisms would have been possible, screening is time

consuming in cryo-ET. During the time of this study, work on *Mycoplasma Pneumoniae* showed that it would have had distinct advantages in terms of imaging (Tegunov, Xue et al. 2021). It was the first organism where near atomic resolution tomography in the cell was demonstrated. A potential down side of this system would have been the comparative lack of previously published data on the chaperonin system that might have unique properties in this minimal bacterium. None the less its chaperonin system might warrant further investigation.

5.2 Single particle analysis of GroEL:ES substrate complexes

In this work we for the first time resolved the molecular detail of a folded substrate inside the EL:ES chamber. Previous biochemical work demonstrated that enzymes could reached an active state inside the EL:ES chamber (Mayhew, Da Silva et al. 1996). This evidence was compounded by several structural studies, some of which were published concurrently with our work, that established a rough location as well as potential contract site of substrates with the EL:ES chamber (Chen, Madan et al. 2013, Kim, Park et al. 2022, Gardner, Darrow et al. 2023). Notably all these studies failed to resolve the structure of the encapsulated substrate protein and thereby left the exact nature of the interactions of the substrate protein with the EL:ES chamber walls in the dark. Previous studies showed a deformation of the EL:ES chamber when a substrate protein is encapsulated (Chen, Madan et al. 2013, Kim, Park et al. 2022). The exact nature of the displacement remained in question, as there was no way to assess the exact orientation of the substrate protein in the chamber and resolve the symmetry mismatch between the two. We unambiguously demonstrated that MetK, an obligate substrate protein of the EL:ES machinery, can fold to a near native state inside of the chamber. The displacements this folding induced in the C-α position of most of the GroEL residues was limited as shown in Fig. 32. This is in accordance with previously published structures showing that folding of a substrate inside the EL:ES chamber induces small divination from the *C7* symmetry of the EL heptamer (Chen, Madan et al. 2013, Kim, Park et al. 2022). We characterised potential contact sites by their distance to the folded MetK. Analysing these showed that in the folded form, MetK is mostly held in place by steric hinderance and not by high energy contacts, such as salt bridges or hydrophobic contacts. Notably the one loop/α-helix in MetK that is only partially resolved contacts the EL chamber and could be stabilised by it through a mixture of hydrophobic and electrostatic interactions. This loop forms the interface

between MetK monomers and would contact another two monomers in the native tetrameric form of MetK(Ishimoto, Fujiwara et al. 2014). An exact analysis of the interactions is unfortunately not possible do to the low local resolution in this region, most likely due to the high remaining flexibility. There are no clear contacts resolved between the folded MetK and the GGM repeats at the bottom of the EL chamber, even though previous work showed that these tails enhance protein folding (Tang, Chang et al. 2006, Chen, Madan et al. 2013, Weaver, Jiang et al. 2017). The nature of the contacts explains how folded MetK is readily discharged into the cytosol *in vivo*. The potential gain of entropy due to changes in water ordering could offset the decoupling of MetK with EL due to the low energy nature of the contacts.

5.3 *In situ* cryo-ET of GroEL/ES complexes and their reaction cycle

With this work we hope to resolve the debate about the functional *in vivo* reaction cycle for the EL:ES machinery. Since the 1990s there has been contradictory evidence about the role of the two potential complexes: $EL:ES_2$ and $EL:ES_1$. Earlier evidence *in vitro* suggested that, in an allosteric manner, ATP's interaction with the *trans*-ring of GroEL triggers the release of GroES from its *cis*-ring counterpart, consistent with previously established models of negative allosteric coupling between the chaperonin rings (Horovitz and Willison). Once GroES disengages, it subsequently attaches to the ring that had formerly acted as the *trans*-ring, yielding a reconfigured *cis*-GroEL-GroES complex (Saibil, Fenton et al. 2013, Hayer-Hartl, Bracher et al. 2016) (Fig. 9). Prior *in vitro* experiments largely align with the dominant role of such asymmetric structures (Engel, Hayer-Hartl et al. 1995, Hayer-Hartl, Martin et al. 1995, Haldar, Gupta et al. 2015, Yan, Shi et al. 2018). Yet, a body of evidence also exists, suggesting a sizable population of symmetrical $EL:ES_2$ under certain conditions (Azem, Kessel et al. 1994, Schmidt, Rutkat et al. 1994, Taguchi). This leads to hypotheses advocating for a symmetry shift under substrate protein (SP) presence (Ye and Lorimer 2013), resulting in a nearly complete shift towards the symmetric $EL:ES_2$ complex overcoming the negative cooperativity. Contradicting these hypotheses, our cryo-ET experiments indicate that both $EL:ES_1$ and $EL:ES_2$ structures coexist within the cellular environment, although $EL:ES_1$ complexes are more prevalent. Further experiments under conditions of thermal stress or with MetK, a GroEL-specific substrate, confirmed the presence of $EL:ES_2$ complexes. This points to a coexistence of both a symmetric

and asymmetric cycle under cellular conditions balancing the different functions of the EL:ES machinery (Horovitz, Reingewertz et al. 2022).

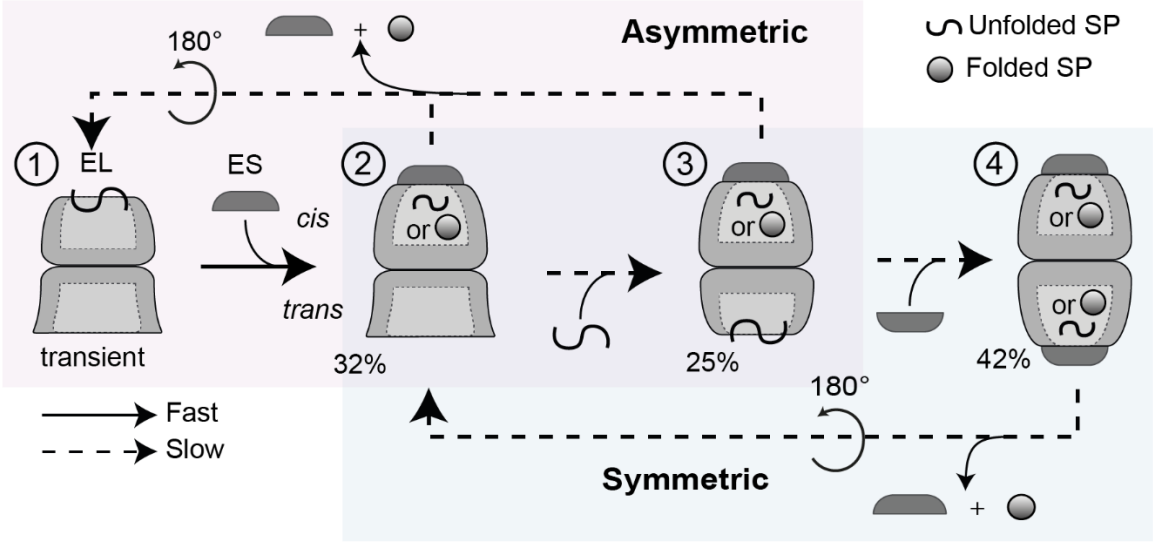

Figure 53: Mechanism for GroEL/ES assisted protein folding *in vivo*.

The interconnected reaction cycles involving asymmetric and symmetric chaperonin complexes are highlighted in light red and light blue, respectively. Folded SP is depicted as a sphere and unfolded SP as a wiggle. Presumably fast and slow steps are indicated by arrows with solid and dashed lines, respectively. The numbers below the pictograms indicate the fraction of the respective complex at 37 °C observed by cryo-ET. EL alone was only observable in EL+ cells and would have to be only transiently populated under physiological conditions to complete an asymmetric reaction cycle.

Two questions remained about the nature of the EL:ES cycle *in vivo*:

1) Is the symmetric EL:ES_2 complex necessary for protein folding to sustain normal cellular growth
2) Is the asymmetric EL:ES_1 complex sufficient for protein folding, or is substrate only folded in one of the chambers of the EL:ES_2 complex.

Let us address these questions in order. First, we showed by shifting the balance of EL_{14} and ES_7 from roughly 1:2 in wild type cells, to a condition with excess EL_{14} over ES_7 (EL+ cells) that these cells grow nearly as fast as 37 °C cells (doubling times 26 vs 25 min). Tomography of these cells showed that under these conditions the EL:ES_1 complex is the only folding active species, with a sizable portion of EL_{14} not bound by ES_7 at all. Further underscoring the negative allostery inherent in the EL:ES system is the fact that most of the identified “naked” EL_{14} molecules had a strong substrate density on one side of the double ring while the other side assumed a more open

substrate unbound conformation. These findings corroborate the view that the intrinsic affinity of GroEL for substrate and GroES varies naturally between the rings, which accords with an asymmetry in cycle occurring *in vivo*. The fact that EL:ES_2 was undetectable and that the cells exhibited near normal growth under laboratory conditions indicate that the EL:ES_1 complex is sufficient for protein folding *in vivo* (Hayer-Hartl, Bracher et al. 2016). Additionally, given the observed physiological 2:1 GroES-to-GroEL ratio(Fayet, Ziegelhoffer et al. 1989), the cellular environment ensures that every GroEL double ring can secure at least one GroES, thus sustaining efficient cell proliferation even at relatively low GroEL concentration. Yet, this behavior could be influenced by the crowded intracellular milieu(Ellis and Minton 2006).

Our investigation reveals that both asymmetric and symmetric chaperonin complexes are actively involved in protein folding through interconnected sub-reactions, thereby clarifying the longstanding uncertainty about the functional stoichiometry of GroEL and GroES. In fast-growing wild-type cells, instances of GroEL unaccompanied by GroES are undetectable using cryo-ET, suggesting that this complex might only be transiently populated (Fig 53; ①). In contrast, the EL:ES_1 complex is highly populated. Thus, the main acceptor for unfolded SP is the *trans*-ring of the asymmetric EL:ES_1 complex (Fig. 53; ②). In the asymmetric reaction, the GroEL rings are sequentially folding-active (Hayer-Hartl, Bracher et al. 2016). GroES dissociates from the *cis*-chamber upon ATP binding to the *trans*-ring (Fig. 53, ② or ③) (Horovitz, Reingewertz et al. 2022). This event leads to the release of the substrate protein from the *cis*-ring, aided by an expanded conformation of the apical domains. This leads to the forming a short lived EL:SP intermediate (Fig 53; ①) that is rapidly bound by ES for another cycle of SP folding. Alternatively, some EL:ES_1 complexes may not complete the asymmetric cycle and instead transition into the symmetric cycle, where both GroEL rings are actively involved in folding simultaneously. The binding of SP to the *trans*-ring of EL:ES_1 in a narrow conformation is the rate limiting step in forming EL:ES_2 complexes. This binding likely mitigates the negative inter-ring allostery and stabilizes the apical domains in a constricted state (Fig. 53, ③ to ④). In EL:ES_2 complexes, ATP hydrolysis and substrate protein folding proceed concurrently in both rings, although the timing may be asynchronous, leading to a staggered release of GroES and the SP (Fig. 53, ④ to ②). Regulation of the partitioning between asymmetric and symmetric chaperonin reactions remains speculative. In the typical asymmetric cycle, the GroEL rings are

interconnected through negative allostery. The binding of ATP to the *trans*-ring, irrespective of substrate protein binding, leads to the release of ADP and GroES from the *cis*-ring (Fig. 53, species ②/③ to ①) (Horovitz and Willison 2005). Experiments *in vitro* have demonstrated that ATP binding to the *trans*-ring diminishes the coupling between GroEL rings. Additionally, stabilizing the GroEL-GroEL interaction with inter-ring disulfide bonds has been shown to lead to the formation of $EL:ES_2$ complexes (Yan, Shi et al. 2018). This negative inter-ring allostery is also present *in vivo*, favoring the formation of $EL:ES_1$ complexes, as evidenced by our observations that $EL:ES_1$ (and unbound GroEL) complexes, but not $EL:ES_2$, are primarily involved in protein folding when there is an excess of GroEL over GroES. However, when GroES is in excess, $EL:ES_2$ complexes are also functional. Our data suggest that the binding of the substrate protein to the *trans*-ring of $EL:ES_1$, inducing a narrow ring conformation, is a critical factor for the formation of $EL:ES_2$ complexes. This binding possibly decreases negative inter-ring allostery in a subset of complexes, allowing ATP binding without prompting the disassociation of *cis*-GroES, and subsequently facilitating the attachment of a second GroES (Ye and Lorimer 2013). Nonetheless, the process of substrate protein binding in the *trans* ring followed by the binding of a second GroES occurs at a slower pace, as evident from the predominance of $EL:ES_1$ complexes with a wide *trans*-ring (lacking bound substrate protein) over those with a narrow *trans*-ring and bound substrate protein.

The time that the substrate protein spends in the encapsulated state for folding is controlled by the GroEL ATPase. Hence, in $EL:ES_2$ complexes, where both rings are active in folding, the substrate protein that is initially encapsulated progresses further in the folding trajectory than the subsequently encapsulated one. This aligns with our discovery that a limited portion of $EL:ES_2$ complexes contain structured substrate proteins in both chambers. In wild-type cells maintained at 37°C, approximately 53% of $EL:ES_2$ complexes have structured substrate protein in one ring, in contrast to about 30% in $EL:ES_1$ complexes. This suggests that the folding reaction starts in the *cis*-chamber of $EL:ES_1$ and continues in the $EL:ES_2$ complex. Consequently, the symmetric cycle, by providing additional time for folding, may be more beneficial for substrate proteins with slower folding kinetics.

The conflicting evidence present in the literature (Horwich and Fenton 2020) combined with our results shows why it is essential to understand proteins in their native in situ context.

5.4 Interactions between macromolecules

We further investigated if, under any of these conditions, we would observe a spatial coupling of protein folding by the EL:ES machinery to protein translation. As a first step towards this aim, we developed an unbiased approach to identify polysomes, an organizational unit of ribosomes closely linked to active translation(Mazor, Dong et al. 2018). For this we used a transformation clustering approach describing the 3D relationship between two macromolecules, in our case ribosomes. The analysis of the identified ribosomal assemblies showed several interesting results. Polysome length and abundance increase upon the overexpression of MetK, while they drastically decrease upon heat stress. We identified a second form of ribosome multimer that closely resembles the published SPA structure of the 100S hibernating ribosome (Beckert, Turk et al. 2018) (Fig. 52). This species is linked to a decrease in translation capacity and acts as a parking place for ribosomes, thereby buffering the need to degrade and re-synthesize this intricate machinery during changing conditions. These hibernating ribosomes showed an inverse change under the different conditions: A strong increase during heat stress and a near vanishing during MetK overexpression. This underscores the ability of cryo-ET to assess the cellular state of single cells by imaging and could be further used in the future in the context of complex samples, like tissues, to tease apart even local differences in the cellular milieu. This could for example be helpful in the context of neurodegenerative disease or other forms of localized stress within a cell.

As a second step we tested on a subset of data whether there is a correlation between the localization of polysomes and the EL:ES machinery. As an internal control we used the 100S ribosomes that should not co-localize with the EL:ES system. We observed that during normal growth there is only a mild co-localization of EL:ES and the polysome. This co-localization gets more pronounced when the obligate substrate MetK is overexpressed, resulting in a peak around 100 nm ± 50 nm distance from the polysome. This correlation is completely abolished during heat stress. These findings are consistent with previous observations that under normal conditions co-translational chaperones such as TF and the DnaJ/K system interact with newly synthesized proteins upstream of the EL:ES machinery (Zhao, Castanie-Cornet et al. 2021). Only if we tip the scales by overexpressing MetK do we observe a weak co-localization. Further experiments with cells where the DnaJ/K system or TF are depleted should show a shift of the EL:ES machinery towards co-translational folding. Interestingly we

saw a shift towards a colocalization of EL:ES and the hibernating 100s ribosomes during heat stress. This co-localization lacks any distinct peak in distance and seems more far-reaching in nature. This might indicate a restructuring of the bacterial cytosol during heat stress, potentially acting to prevent aggregation. This theory would need further careful investigation by complementary techniques such as super resolution light microscopy to characterize this shift in cellular organization and explore the potential of bacterial cytosolic higher ordered compartmentalization.

5.5 Outlook

Overall, this work shows the unique power of cellular cryo-ET to resolve macromolecular complexes in their *in situ* environment, without the need for labeling or chemical fixation, preserving the cell in its near native state. Cryo-ET offers very rich datasets, containing information about more than the initial specific biological question one wants to answer while acquiring a dataset. We demonstrated this using the ribosomal complexes as cellular markers confirming the cellular state of our cells during the experiment. This data was initially only intended for quality control purposes as shown in Fig. 41, but allowed us to probe a deeper layer of cellular organization. With the number of structures deposited in the EMDB resolved by cellular cryo-ET and STA exponentially increasing over the recent years, this type of fortuitous discovery will be key in driving our understanding of complex biology on a molecular level. This is especially true if we consider the push in the field towards more and more complex samples, providing even richer datasets closer to the native organization of cells in our own bodies. Cellular and tissue cryo-ET is uniquely positioned to provide the initial spark for interesting biological discovery. The currently low throughput and limited capability of *de novo* identification of proteins inside of tomographic reinstructions to only molecules that can be averaged to near molecular resolution will necessitate the integration with orthogonal techniques. Traditional cell biology, light microscopy, and different omics techniques will be increasingly needed to solved the mysteries of our own biology.

6 Material and Methods

6.1 Plasmids and strains

E. coli BL21(DE3) Gold cells from Stratagene were used in growth analysis, electron tomography, and protein expression studies. GroEL was expressed from a pBAD33 plasmid with the groEL gene regulated by an araBAD promoter in EL+ cells for tomography and biochemical tests, but not for purification (Kerner, Naylor et al. 2005). A pBAD33 plasmid, which contains both groEL and groES genes under the control of an araBAD promoter, was utilized for the overexpression of GroEL and GroES (Guzman, Belin et al. 1995). The expression of MetK was done using a pET22b plasmid as previously detailed(Kerner, Naylor et al. 2005).

6.2 Antibodies

The polyclonal antisera against GAPDH, MetK, GroEL and GroES were previously described (Kerner, Naylor et al. 2005) and all produced in house for previous studies. The antibody against lactalbumin was purchased from East Acres Biologicals.

6.3 *E. coli* growth

E. coli cells were cultured in LB medium. Depending on the plasmid in use, the medium was supplemented with either ampicillin (200 µg mL^{-1} for pET22b-MetK) or chloramphenicol (32 µg mL^{-1} for pBAD33 variants).

For the overexpression of GroEL, GroES, and MetK, *E. coli* Bl21 (DE3) cells, transformed with pBAD33-GroEL/ES and pET22b-MetK, were grown to early exponential phase at 37°C. GroEL/ES expression was initiated by adding arabinose to the LB medium until a concentration of 0.2% (w/v) was reached, and the culture was left to grow for 90 minutes. Following this, for harvesting, the cells were centrifuged at 8,000 g and 4°C for 10 minutes. The resulting pellet was resuspended to an OD_{600} between 0.1 and 0.2 in fresh LB medium, which contained both antibiotics. Expression of MetK was then induced by adding 1 mM IPTG followed by growth for 40 minutes.

In the case of GroEL expression (EL+), *E. coli* Bl21 (DE3) pBAD33-GroEL was induced with arabinose until a concentration of 0.1 % (w/v) was achieved, followed by 37 minutes of growth at 37 °C.

To subject *E. coli* Bl21 (DE3) cells to heat stress (HS), the cells, once they reached the early exponential growth phase at 37 °C, were transferred to a water bath set at 46 °C and shaken continuously for 2 hours.

6.4 *Caulobacter crescentus* growth

C. crescentus was cultured following the standard protocols provided by the DSMZ (DSMZ 2023) and handled analogously to *E. coli* for cryo-FIB milling and tomography.

6.5 Protein expression and purification

** This work was performed by the technical staff in the case of GroEL, GroES and as indicated in the citation for MetK.*

GroEL, GroES and MetK proteins were expressed and purified as described (Hayer-Hartl, Weber et al. 1996, Kerner, Naylor et al. 2005).

6.6 Measurement of protein concentration

Protein concentrations for purified samples were ascertained by observing the absorbance at 280 nm. Absorbance coefficients, derived from the protein sequence using the ProtParam program, were utilized for these measurements (Gasteiger, Hoogland et al. 2005). For cell lysates, protein concentrations were measured using the Pierce Coomassie Plus (Bradford) Assay Kit from Thermo Fisher Scientific, following the manufacturer's instructions.

6.7 Preparation of cell lysates

Cultures were prepared as previously outlined. Once grown, they were centrifuged, and the resulting cell pellet was immediately frozen using liquid nitrogen for subsequent steps. For spheroplast generation, the method detailed by Ewalt in 1997 was employed (Ewalt, Hendrick et al. 1997). The cells were initially resuspended in 100 mM Tris-HCl (pH 8.0) and then given two washes using 2 mL of this buffer. After washing, the pellet was dissolved in HMK buffer, which consists of 50 mM HEPES-KOH at pH 7.2, 20 mM Mg-acetate, and 50 mM K-acetate. This buffer was also supplemented with 20% sucrose (w/v) and 0.25 mg mL^{-1} lysozyme. The cell lysis mixture was incubated on ice for 7 minutes and then at 37°C for 10 minutes. The Complete EDTA-free protease inhibitor cocktail from Roche was introduced to the mix to avoid proteolysis. To achieve lysis of the spheroplasts, 0.1% Triton X-100 (v/v) was added, followed by sonication using a Branson Sonifier 250, set to a constant duty cycle and power level 1, for 3 minutes.

6.8 Mass spectrometry

** Mass spectrometry was performed in collaboration with Dr. Roman Köhler, Albert Ries and Alonso Izzat Carvajal Alvarez. Replicates of the work were independently carried out by Alonso Izzat Carvajal Alvarez. The data was analyzed by Roman Köhler or Alonso Izzat Carvajal Alvarez.*

Cell lysates were subjected to reduction by the addition of dithiothreitol (DTT) to a final concentration of 10 mM, followed by incubation at 56 °C for 45 minutes. The acylation of thiol groups was carried out by incorporating chloroacetamide to a final concentration of 55 mM and incubating for 45 minutes in a dark environment. An initial proteolytic digestion was performed using Lys-C (Wako) at a mass-to-mass ratio of 1:20 and incubation at 37 °C for 2 hours. This was followed by a secondary digestion stage with trypsin (Roche), also at a 1:20 mass-to-mass ratio, conducted overnight at 37 °C. The enzymatic reaction was terminated by adding trifluoroacetic acid (TFA) to a final concentration of 1%. Peptides were then desalted utilizing OMIX C18 (100 µL) tips (Agilent Technologies, Part No: A57003100), following the manufacturer's guidelines.

The desalted peptides were resuspended in 12 µL of 5% formic acid and subjected to sonication in a Bandelin Sonorex RK 100 ultrasonic bath. After centrifugation, the samples were transferred into autosampler vials (Waters) and analyzed using an Easy nLC-1200 nanoHPLC system (Thermo Fischer Scientific) coupled with a Q-Exactive Orbitrap HF mass spectrometer (Thermo Fischer Scientific). Peptides were separated on pulled spray-columns (ID 75 µm, 30 cm length, 8 µm tip opening, NewObjective), packed with 1.9 µm C18 particles (Reprosil-Pur C18-AQ, Dr Maisch GmbH). Depending on the sample, either a 196-minute or a 67-minute stepwise gradient was employed between buffer A (0.2% formic acid in water) and buffer B (0.2% formic acid in 80% acetonitrile). Column loading was automated via the nanoHPLC autosampler, operating at a pressure of 900 bar, with a set flow rate of 0.25 µL/min.

For mass spectrometric analysis, two distinct parameter sets were utilized. For comparisons involving 37°C, HS, and MetK conditions, the following settings were applied: MS resolution at 60,000 FWHM, MS-mass range from 300 to 1650 m/z, MS-AGC-setting at 3e6, and MS-MaxIT at 50 ms. All other samples were analyzed under a different set of parameters: MS resolution at 120,000 FWHM, same MS-mass range and MS-AGC-setting, but with MS-MaxIT adjusted to 100 ms.

Mass spectrometric data were processed using MaxQuant software with default settings. For protein identification, the *E. coli* K12 strain sequences from the UNIPROT database (version 2023-03-01) were used. MaxQuant's decoy database approach was employed to maintain the false discovery rates for proteins and peptides below 1%.

6.9 Quantification of MetK binding to GroEL

** This set of experiments was performed in collaboration with Alonso Izzat Carvajal Alvarez.*

To quantitatively assess the proportion of GroEL complexed with MetK in MetK overexpressing cells, immunoprecipitation followed by immunoblotting and liquid chromatography-tandem mass spectrometry (LC/MS-MS) for GroEL and MetK was conducted. The cellular preparation and lysis procedures adhered to the aforementioned protocols, with the modification of introducing apyrase to a final concentration of 25 U mL^{-1} into the lysis buffer, deplete the ATP pool within the lysate, effectively halting the GroEL reaction cycle as documented by (Kerner, Naylor et al. 2005). Post-lysis, the samples were subjected to centrifugation at 16,000 g at 4°C for 10 minutes for clarification.

Antibodies were immobilized on protein A Sepharose 4B beads (Thermo Fischer Scientific). Specifically, either 20 µL of a nonspecific α-lactalbumin antibody or a GroEL-specific antibody was coupled to 100 µL of the beads, following the manufacturer's instructions. The antibody-loaded beads were subsequently incubated with 180 µg of protein sample in 650 µL of HMK buffer for 1 hour. Post-incubation, the beads were washed twice with 600 µL of HMK buffer, followed by an additional two washes using HMK buffer containing 0.1% Triton X-100.

For immunoblotting assays, elution was conducted with 50 µL of 2x lithium dodecyl sulfate (Pierce) containing 5% (v/v) β-mercaptoethanol, as per manufacturer's guidelines. Alternatively, for LC/MS-MS analysis, elution and enzymatic digestion were performed utilizing the IST MS sample preparation kit (Preomics), adhering to the manufacturer's on-bead digestion protocol. Mass spectrometric analysis was executed as previously delineated.

6.10 *E. coli* growth curves

** The analysis of the growth curve data was performed by Alonso Izzat Carvajal Alvarez.*

Cell cultures were grown under the conditions previously outlined. At specified time intervals, aliquots were withdrawn for the assessment of optical density at 600 nm (OD_{600}) using an Eppendorf BioPhotometer Plus. To maintain uniform growth conditions, cultures that reached an OD_{600} exceeding 0.4 were diluted back to an OD_{600} of 0.1 using prewarmed LB medium supplemented with the requisite antibiotics and

arabinose. Subsequent OD_{600} readings were adjusted to account for the cumulative dilution factor. Data were subjected to computational processing and curve fitting utilizing the R statistical software package.

6.11 SDS-PAGE and Immunoblotting

** Some of the plots in this thesis were repeated by Alonso Izzat Carvajal Alvarez for improved image quality.*

Prior to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis, cellular samples were resuspended in HMK buffer supplemented with 2 mM DTT, 1 mM ethylenediaminetetraacetic acid (EDTA), and 5% glycerol. Sonication was subsequently conducted using a Bioruptor Plus (Diagenode) in accordance with the manufacturer's guidelines, followed by centrifugation at 16,000 g and 4°C for 20 minutes. The resulting protein samples were separated via electrophoresis using NuPAGE 10% Bis-Tris SDS gels (Invitrogen) and NuPAGE MES SDS running buffer (Invitrogen), at a constant voltage of 150 V.

The proteins were transferred from the polyacrylamide gels onto a polyvinylidene fluoride (PVDF) membranes (Roche) using a blotting buffer comprised of 25 mM Tris, 192 mM glycine, and 20% methanol. The transfer was conducted in a semidry plotter at a current of 150 mA. Following transfer, membranes were subjected to washes with TBST buffer (10 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.05% Tween-20) and subsequently blocked with 5% skimmed milk (Merck) in TBST buffer for 1 hour at ambient temperature.

Immunoblotting was performed by incubating the membranes with primary antibodies diluted in TBST buffer overnight at 4°C. Post-incubation, membranes were washed thrice with TBST, followed by a 1-hour room temperature incubation with horseradish peroxidase (HRP)-conjugated secondary antibodies (Sigma-Aldrich). The membranes were again washed three times with TBST. For chemiluminescent signal detection, Luminata Classico (Merck) was used as the HRP substrate and signals were captured using an Amersham ImageQuant 800 system (Cytiva).

For antibody stripping, membranes underwent two incubations in 200 mM sodium hydroxide (NaOH) for a duration of 5 minutes each.

6.12 Sample preparation and analysis of EL:ES:MetK complexes for single particle cryoEM

To synthesize substrate-bound EL:ES complexes, a 4 µM concentration of MetK was denatured alongside 1 µM of 14-mer GroEL in Buffer A, which contained 20 mM MOPS-NaOH (pH 7.4), 200 mM KCl, 10 mM MgCl2, 5 mM DTT, 30 mM NaF, and 5 mM BeSO4. This mixture was first incubated at 60°C for 15 minutes and then cooled down to 25 °C using an Eppendorf thermomixer over the course of one hour. The addition of 2 µM of 7-mer GroES and 1 mM ATP (pH 7) led to the formation of stable chaperonin complexes encapsulating MetK (Koike-Takeshita, Yoshida et al. 2008).

The resulting complexes were isolated by size exclusion chromatography on a Superdex 200 3.2/300 GL column. Fractions were analyzed by SDS-PAGE electrophoresis (NuPAGE™, Bis-Tris 4 to 12% gels) and by subsequent wester blotting. The fraction of MetK loaded of EL:ES complexes was further estimated by mass spectrometry using iBAQ values for quantification (Schwanhausser, Busse et al. 2011). For analysis by mass spectrometry, fractions F1 and F2 (Fig. 42) were analyzed separately but intensities pooled for the determination of iBAQ ratios.

For plunge freezing, the GroEL:GroES:MetK samples were concentrated tenfold through ultrafiltration using a 100 kDa Amicon centrifugal concentrator (Millipore) at room temperature. A control, consisting of GroEL and GroES without MetK, was treated in the same manner. Prior to freezing, a 1 µL of an n-octyl-β-D-glucopyranoside stock solution (87.5 mg/mL in Buffer A) was added per 50 µL of the sample.

For SPA and *in vitro* cryo-ET experiments, 4 µL of the concentrated sample was applied to a plasma-cleaned R 2/1 100 Holey carbon film Cu 200 mesh grid (Quantifoil). The grid was then blotted for 3.5 seconds at force 4 and flash-frozen in a liquid ethane-propane mixture cooled by liquid nitrogen using a Vitrobot Mark IV (Thermo Fisher Scientific), maintaining 100% humidity and a temperature of 4°C.

6.13 Data collection for single particle CryoEM

** SPA data collection and preprocessing was performed by Dr. Stefan Bohn*

Cryo-EM data for the EL:ES:MetK set was acquired using a FEI Titan Krios transmission electron microscope and the SerialEM software package (Mastronarde 2005). Images were recorded at a nominal magnification of 22,500x on a K3 direct electron detector (Gatan), accumulating a total electron dose of approximately 55

electrons per $Å^2$ over 30 frames and a calibrated physical pixel size of 1.09 Å. Micrographs were captured within a defocus range of -0.5 to -3.0 µm.

Real-time image processing and contrast transfer function (CTF) refinement of the cryo-EM micrographs were executed using the Focus software package (Biyani, Righetto et al. 2017). Only micrographs meeting specific criteria—including iciness below 1.05, drift between 0.4 Å and 70 Å, refined defocus between 0.5 µm and 5.5 µm, and estimated CTF resolution better than 6 Å—were retained. Frame alignment was done using MotionCor2 (Zheng, Palovcak et al. 2017) while CTF values for the aligned frames were determined via GCTF(Zhang 2016).

In a parallel control dataset, GroEL:ES complexes were obtained without MetK under similar conditions but at a nominal magnification of 29,000x, yielding a calibrated pixel size of 0.84 Å.

6.14 Image processing, classification and refinement for single particle analysis

From the EL:ES:MetK dataset, a total of 8,945 micrographs yielded 1,561,482 particles, which were picked using a crYOLO network trained specifically for this purpose (Wagner, Merino et al. 2019). These particles were extracted and 2D classified in RELION 3.1.3 (Zivanov, Nakane et al. 2018). The remaining particles were exported to CryoSPARC for additional 2D classification, ab initio model building, alignment, and 3D classification (Punjani, Rubinstein et al. 2017). $EL:ES_1$ and $EL:ES_2$ complexes were separated using 3D classification. This yielded $EL:ES_2$ (659,866 particles) and $EL:ES_1$ (294,250 particles) subsets that were exported back into RELION for additional alignment, CTF refinement and Bayesian Polishing. In the case of $EL:ES_2$ complexes, symmetry expansion around the *C2* axis was conducted, and the opposite half was subtracted through RELION's signal subtraction feature.

Further classification of the asymmetric $EL:ES_1$ complexes was performed using CryoDRGN (Zhong, Bepler et al. 2021), generating a clean subset of 242,276 particles. Signal subtraction in RELION was then used to remove the *trans* ring. The ES-bound single-ring particles, hereafter $EL_7:ES_7$ (1,562,002 particles in total), were aligned to a common reference in RELION and subsequently exported to CryoSPARC for further refinement without imposing symmetry. The resulting mask and reference were re-imported into RELION for an additional alignment step, targeting the asymmetric MetK substrate within the $EL_7:ES_7$ chamber. A second round of signal subtraction was performed to remove the $EL_7:ES_7$ chamber density, followed by 3D classification of

only the MetK density in RELION without angular search. Subsets displaying visible secondary structure elements were further analyzed and combined into a single frame of reference using custom Matlab 2015b scripts.

Local alignment of these folded MetK (fMetK) particles was conducted in CryoSPARC, followed by an additional 3D classification and a final round of local alignment, yielding 322,800 particles. This resulted in a MetK density with a global resolution of 3.7 Å.

To study MetK's interaction with the EL_7:ES_7 chamber, signal subtraction in RELION was reversed and the EL_7:ES_7:fMetK particles reconstructed. These were further refined and aligned in CryoSPARC, resulting in a global resolution of 3.04 Å for the EL_7:ES_7:fMetK complex.

The remaining particles without visible secondary structure were reconstructed separately after reversion of the signal subtraction. This resulted in a EL_7:ES_7:ufMetK complex structure at a nominal resolution of 2.94 Å.

As a control, GroEL:ES complexes without substrate proteins (SP) were processed similarly. Bayesian Polishing and CTF refinement stages in RELION was omitted. Signal subtraction was conducted in CryoSPARC, yielding a global resolution of 2.5 Å based on 293,974 particles and after the application of *C7* symmetry for the EL_7:ES_7 complex.

The resulting density maps were visualized and rendered using ChimeraX software package (Pettersen, Goddard et al. 2021).

6.15 Model building and refinement

**All models in this thesis were built by Dr. Andreas Bracher*

The process of model construction began with a rigid-body fitting of the subdomains of GroEL, GroES, and MetK, sourced from their respective crystal structures—PDB 1SX3 (Chaudhry, Horwich et al. 2004), PDB 5OPW(Yan, Shi et al. 2018), and PDB 7LOO(Gade, Tan et al. 2021)—into the cryo-EM density map. This initial step was followed by manual modifications using the Coot software (Emsley and Cowtan 2004). Further refinement of the models was performed in real space using the Phenix software suite (Liebschner, Afonine et al. 2019). Any residues exhibiting disordered side-chains were truncated at the C-beta position.

6.16 Plunge freezing for cryo-ET

The *E. coli* cells were cultured as described above for the different conditions. Cells were concentrated to OD_{600} of 7-10 by centrifugation at 8,000 g for 2 minutes at 4°C. R 2/1 100 Holey carbon film Cu 200 mesh grids (Quantifoil) that were plasma cleaned

for 30 sec in an air plasma. 4 µL of the cell suspension was applied onto the grid. The sample was blotted for 9 sec with force 10 and then plunge frozen in a mixture of liquid ethane and propane cooled by liquid nitrogen using a Vitrobot Mark IV (Thermo Fisher Scientific) at 70% humidity and 22°C. The grids were clipped into Autogrid support rings that had a special cut-out for ion beam accessibility at low angles (Thermo Fisher Scientific), and were stored in liquid nitrogen until they were utilized for FIB milling.

For plunge freezing of *C. crescentus* was cultured as described above and plunge frozen analogously to *E. coli*.

M. mazei was purchased as a culture from the German Collection of Microorganisms and Cell Cultures GmbH and not further grown in house (DSMZ 108209). Cultures were then briefly centrifuged to concentrate the cells to an OD_{600} nm of around ~1 and were immediately used for plunge-freezing. To prevent harmful exposure of the anaerobic bacteria to oxygen, the cells were handled and loaded into the micro-pipette in a nitrogen-rich atmosphere. To facilitate this, a makeshift cardboard enclosure with a small opening for manipulation was created and sealed with tape. A foam pad was positioned at the bottom of the enclosure, and all essential equipment was arranged on top of it. Concentrated *M. mazei* cultures were kept on a pre-warmed metal block to ensure consistent temperature during the plunge-freezing process. Liquid nitrogen was used to fill the enclosure and float the Styrofoam pad. The constant evaporation of the liquid nitrogen, along with the volume expansion that occurred during the liquid-to-gas transition, effectively displaced air, thereby eliminating oxygen from the environment. A candle flame was periodically used to verify the anaerobic conditions within the enclosure. The pipette tip was flushed with nitrogen before each use. Four microliters of *M. mazei* was directly applied to the EM grid and blotted for 10 seconds with a blot force of 10 onto R 2/1 carbon-foil 200-mesh copper EM grids (Quantifoil Micro Tools). This was then plunge-frozen in a liquid ethane–propane blend using a Vitrobot Mark 4 (Thermo Fisher Scientific) that was continuously flushed with nitrogen to minimize air exposure during plunging. Subsequently, the grids were secured in Autogrid support rings that had a special cut-out for ion beam accessibility at low angles (Thermo Fisher Scientific), and were stored in liquid nitrogen until they were utilized for focused ion beam (FIB) milling.

6.17 Focused Ion beam (FIB) Milling

The frozen cell grids were transferred into a dual beam cryo-focused ion beam (FIB)/scanning electron microscope (Thermo Fisher Scientific, either Scios, Quanta,

Aquilos or Aquilos 2). The cells were coated with a layer of inorganic platinum, if available in the used system, to help with sample conductivity. This was followed by the deposition of organometallic platinum using an *in situ* gas injection system (working distance, 10 mm; heating, 27°C; time, ~8 s). Removal of bulk material was carried out at a stage angle of 20-25° using gallium ions at 30 kV, 0.5 nA. Fine milling of lamellas preformed with successively lower currents between 0.3 nA and 30 pA as previously described in (Rigort, Bauerlein et al. 2012). Micro-expansion joints (relief cuts) were milled to prevent lamella from bending (Wolff, Limpens et al. 2019). Owing to the small size of individual bacteria, carpets consisting of around two to three layers of cells were milled at a low stage tilt angle (11–13°) to form short lamellae of ~100–200-nm thickness, for *E. coli* and *C. crescentus*. Milling for *M. mazei* was performed similarly but at a higher stage tilt angle of 15°.

The lamellas for the selective GroEL overexpression dataset were prepared using Serial FIB(Klumpe, Fung et al. 2021) and an additional layer of inorganic platinum was added after fine milling to avoid charging during image acquisitions (Khavnekar, Vrbovská et al. 2022).

6.18 *In situ* cryo-ET

The prepared lamellae were imaged using a Titan Krios TEM with a 300 kV field emission gun (Thermo Fisher Scientific). The microscope was equipped with a Quantum K2 energy filter and a K2 Summit direct detection camera from Gatan. Tomographic images were captured at a 42,000x magnification, corresponding to a pixel size of 3.52 $Å^2$. A defocus range of -5 to -3.0 µm was chosen. The images were captured in dose-fractionated super-resolution mode, at a total dose of approximately 120 $e^-/Å^2$ per tilt series, corresponding to around 3 $e^-/Å^2$ per image using SerialEM 3.9.0 (Mastronarde 2005).

For optimal image quality, a dose-symmetric tilt scheme was implemented, spanning a total range of ± 60° from an initial angle of 10°. This approach was designed to counterbalance any pre-existing tilt in the lamellae, which typically averaged around 11°. Image frames were aligned using MotionCor2 (version 1.4.0), and reconstruction was carried out in IMOD (version 4.10.49) through the TOMOgram MANager (TOMOMAN) wrapper scripts (Wan 2020). Image dose filtering and defocus estimation were performed using TomoMAN's implementation of the Grant and Grigorieff exposure filter (Grant and Grigorieff 2015) and CTFFIND4 (Rohou and Grigorieff 2015), respectively.

For the EL+ dataset, tomograms were acquired on a Krios G4 microscope equipped with a Selectris X energy filter and a Falcon 4 direct electron detector. Data collection was managed by the TEM Tomography 5 software (Thermo Fisher Scientific) and involved a dose-symmetric tilt scheme, with tilt ranges of ± 60° initiated at ± 10º. Autofocus adjustments were made for each tilt series, ranging from -2.5 µm to -5 µm. Data were captured in Falcon 4's EER mode, whit a calibrated physical pixel size of 3.02 $Å^2$ and a total dose of 3 $e^-/Å^2$ per tilt across 10 frames. A 10 eV slit was utilized throughout the data collection process.

Preprocessing was executed using TOMOMAN, with motion correction accomplished via Relion's version of MotionCor2. Defocus values were assessed using CTFFIND4, and the reconstruction was performed in IMOD. In this process, local deposits of the inorganic platinum applied during post-milling sputtering served as fiducial markers. CTF corrected versions of all tomograms were reconstructed at different binnings using NovaCTF (Turonova, Schur et al. 2017).

6.19 Subtomogram averaging

** Prof. Dr. William Wan wrote the stopgap software framework simultaneously with this work. I profited immensely from being an early alpha tester, help with averaging strategies and the chance to help add features to stopgap. Parts of the data for this thesis were used as test cases for development.*

For subtomogram averaging, datasets collected on the identical microscope (37°C, HS, MetK) were grouped and processed collectively. In contrast, the EL+ dataset underwent independent processing. A comprehensive overview of the workflow is provided in Fig. 31.

Initial templates were generated using previously published structures from the Protein Data Bank: 1AON for EL:ES_1, 4PKO for EL:ES2, and 5MDZ for 70S ribosomes. These structures were filtered to a 40 Å resolution using Chimera's (Pettersen, Goddard et al. 2004) molmap (Baker, Zhang et al. 2010) command.

To find the initial positions of a subset of EL:ES_1, EL:ES_2, and ribosomes, the noise-correlation template-matching technique within the STOPGAP software was used on tomographic volumes that had been binned four times to a pixel size of 14.08 Å (Wan 2020). This subset was then aligned and classified in STOPGAP to create a reference with a Fourier Shell Correlation (FSC) value nearing 1 at the 40 Å template matching resolution. A second round of template matching was conducted on the complete

dataset using these refined references. Efforts to identify EL_{14} structures using various templates generated from PDB structures were inconclusive and thus discontinued.

Each tomogram was visually inspected to determine the cross-correlation cut-off, comparing generated hits with the actual tomogram. To minimize false positives, a cytosolic mask for the bacterium was created in AMIRA (Thermo Fischer Scientific), which was then used to filter out external hits. This process yielded 176,408 initial subtomograms for EL:ES_1 and 125,860 for EL:ES_2, which were then separately classified and aligned in STOPGAP, resulting in 19,239 and 17,614 particles, respectively. As both references pick up a subset of the other particles, the particles were then combined and duplicates removed. The resulting combined dataset was split by reference free 3D classification in STOPGAP resulting in a set of 17,598 EL:ES_1 and 11,213 EL:ES_2 complexes that were then independently refined. This resulted in a resolution at the FSC cut-off of 0.143, after the application of symmetry, of 11.6 Å for the EL:ES_1 complex (*C*7) and 11.9 Å for the EL:ES_2 complex (*D*7).

Further processing of these datasets in a RELION / WARP / M pipeline as outlined in (Hoffmann, Kreysing et al. 2022) yielded no further gains in resolution.

Classification techniques employed simulated annealing stochastic hill climbing multi-reference alignment, as outlined in prior research (Erdmann, Hou et al. 2021). All classifications were carried out repeatedly with different, random initial starting sets of 250-500 subtomograms to generate the initial references. Only particles that ended up in the same class for all independent rounds of classifications were retained (Erdmann, Hou et al. 2021).

The EL+ dataset followed a similar processing pipeline, with low-pass filtering applied to the final structures of the previous datasets as initial references for template matching. Template matching was repeated once with structures generated by averaging a subset of this dataset. To improve the resolution for model building, the dataset was exported to WARP (Tegunov and Cramer 2019) and angles and positions refined using RELION 3.0 (Zivanov, Nakane et al. 2018). EL 14-mer particles were co-refined for geometric distortions with ribosomes. The resulting EL 14-mer particles were exported for further alignment and classification in RELION. Classification was performed with a regularization parameter T of 4 and 6 classes for 25 iterations without angular search. This resulted in a more homogeneous subset of 12,421 particles. These particles were again co-refined in M for geometric distortions and per particle

defocus for CTF estimation. This resulted in a final structure with nominal resolution of 9.8 Å at the 0.143 FSC cut off.

Due to their high molecular weight and density, ribosome template matching achieves a higher precision and recall. During initial rounds of classification in STOPGAP, no false positive particles were detected. Therefore, all ribosomal hits from template matching were aligned first in STOPGAP at progressively lower binnings (bin4, bin2, bin1). The resulting particles were then exported to WARP using TOMOMAN. Subtomograms were reconstructed for RELION 3.0 using WARP with a nominal pixel size of 3.52 $Å^2$/pix. An iterative approach with subtomogram alignment in RELION and tilt-series refinement in M (Tegunov, Xue et al. 2021) was performed until no further improvement in gold standard Fourier Shell Correlation (FSC) was obtained. This resulted in a final structure of the ribosome at a resolution of 8.6 Å for the combined 37°C, HS and MetK OE datasets and 6.3 Å for the EL+ dataset that was processed separately.

6.20 Classification of the GroEL:ES substrate state *in situ*

** This work was a close collaboration with Florian Beck how had the idea of bootstrapping. Thank you for that Florian!*

To elucidate substrate protein densities within the GroEL-ES chamber, we employed a multi-stage analysis strategy. Initially, we executed a symmetry expansion around the *C2* axis of the $EL:ES_2$ complexes and combined the resulting chambers with the $EL:ES_1$ chambers, generating a new set of aligned $EL_7:ES_7$ chambers. All subtomograms were subsequently denoised using TOPAZ's 3D pretrained function(Bepler, Morin et al. 2019).

Our early classification attempts via STOPGAP's multi-reference-based alignment led primarily to separation by missing wedge. To circumvent this, we created 5,000 random averaged bootstraps, each comprising 250 random subtomograms. These bootstraps were then subjected to k-means clustering to yield two distinct classes. The averaged bootstraps from these classes served as initial structures for multi-reference alignment in STOPGAP as described in the previous paragraph, where we implemented stochastic hill climbing with a simulated annealing temperature factor of 10. This was followed by 40 iterations of alignment, focusing on the chamber's interior by masking. The classification with the two staring classes was repeated five times. Only particles that consistently fell into the same classes across all repeats were retained for a final round of subtomogram averaging.

This resulted in two distinct classes: one exhibiting weak, diffuse density within the chamber and another showing strong density near the chamber's base. Attempts to further subdivide these classes only led to separations based on missing wedge. Given the inability to resolve the C7 symmetry mismatch between the substrate and its enclosing chamber, we generated final averages under imposed C7 symmetry to improve the signal-to-noise ratio. The class with strong density included 12,255 subtomograms, while the weak-density class comprised 24,435 subtomograms.

6.21 Cryo-ET analysis of *in vitro* reconstituted GroEL-ES complexes

To generate a comparison structure for the *in situ* EL_7:ES_7 complexes found, tomographic analysis of EL:ES complexes containing a defined substrate protein in a folded state and in a known topology we imaged at the same nominal magnification (42,000 x) and pixel size (3.52 $Å^2$ per pixel). The grids were prepared in the same batch as the ones used for SPA ensuring that folded MetK would be presented in the chambers in the imaged sample.

The defocus range was kept identical, ranging from -5 to -3.0 µm and imaging was performed same TEM that was used for the 37°C, HS and MetK datasets (Titan Krios with a field emission gun 300 kV (Thermo Fisher Scientific) equipped with an energy filter (Quantum K2, Gatan), a direct detection camera (K2 Summit, Gatan).

Tomograms were recorded in dose fractionated super resolution mode, with a total dose of ~120 e-/Å2 per tilt series using SerialEM 3.9.0 (RRID: SCR_017293, https://bio3d.colorado.edu/SerialEM/)(Mastronarde 2005). A dose-symmetric tilt scheme was used with an increment of 3° in a total range of ± 60° from a starting angle of 0°(Hagen, Wan et al. 2017). Frames were aligned using MotionCor2 (v.1.4.0, https://emcore.ucsf.edu/ucsf-software)(Zheng, Palovcak et al. 2017). The reconstruction was performed in IMOD using patch tracking (v.4.10.49, RRID:SCR_003297, https://bio3d.colorado.edu/imod/) by using the TOMOMAN wrapper scripts, analogously to what was described in the chapter *In situ cryo-ET* (Wan 2020). The tilt series were dose filtered using TOMOMANs implementation of Grant & Grigorieff exposure filter (Grant and Grigorieff 2015). The CTF was determined using CTFFIND4 (Rohou and Grigorieff 2015) and CTF corrected tomograms reconstructed using NovaCTF at a nomial binning of 4, 2 and 1 corresponding to a pixel size of 14.08 , 7.04 and 3.52 $Å^2$ (Turonova, Schur et al. 2017).

Initial spatial coordinates and orientations of EL:ES_1 and EL:ES_2 particles were determined by noise-correlation template-matching techniques, as implemented within the STOPGAP software package on tomograms with a nominal pixel size of 14.08 $Å^2$. For this purpose, *in situ* structures of EL:ES_1 and EL:ES_2 served as template references, thereby enhancing the specificity and reliability of the particle recognition process (Wan 2020).

This resulted in 39,518 initial hits for the EL:ES_2 template and 46,093 hits for the EL:ES_1 template, with both sets showing a significant overlap. These were then further aligned and classified separately in STOPGAP combined to remove duplicate particles

and separated again by classification as described above, yielding 5,832 and 13,688 particles, respectively.

Classification of the substrate inside the chamber was performed as described above in section “Classification of the GroEL:ES substrate state *in situ*”.

The resulting classes were then again split into $EL:ES_1$ and $EL:ES_2$ complexes corresponding to their substrate state and exported to WARP using TOMOMANs export function (Tegunov and Cramer 2019).

An additional round of alignments was performed in RELION 3.0 for all different classes and complexes with prior set for all angles to avoid divergence. Local search was performed with a sigma of 0.5 and a search angle of 0.9°. The resulting particles were separately refined in M correcting for geometrical distortions. Particles were again exported from M (Tegunov, Xue et al. 2021) at a nominal pixel size of 3.52 $Å^2$ and signal subtraction preformed in RELION of the *trans* ring for $EL:ES_1$ and the opposing chambers for $EL:ES_2$. Based on their previous classification results in STOPGAP, the refined signal-subtracted single-chamber complexes were combined in two groups based on the substrate density that could be identified. This resulted in 7,087 $EL_7:ES_7$ chambers containing an ordered SP and 14,371 $EL_7:ES_7$ chambers that either contained a disordered SP or that were empty. The resulting chambers were again locally refined in RELION using priors and a sigma on all angles. This resulted in a global resolution of 9.4 $Å$ for the GroEL-ES chambers containing ordered SP and a global resolution of 8.8 $Å$ for the remaining chambers. Even at these resolutions attempts to break the symmetry mismatch between the $EL_7:ES_7$ chamber and the SP remained fruitless.

6.22 Clustering of ribosomes

* *The algorithm described in this section was originally designed by Florian Beck. The work in this section was performed in close collaboration with Florian Beck how also designed some of the panels presented in Fig 48.*

Polysome were detected by the MATLAB implementation of NEMO-TOC (Jiang, Wagner et al. 2022). For Clustering a threshold of 25 pseudo degrees was used. The method consists of 3 major steps: 1) Transformations (rotations, translations) between ribosome pairs were calculated. 2) Distance matrices for rotations and translations were computed and combined. 3) Transformations were clustered based on the combined distance matrix.

6.23 Co-localization analysis of ribosomal super complexes and the EL:ES machinery

** This work was performed in close collaboration with Prof. Dr. Antonio Martínez-Sánchez*

After all molecular species were identified we used the PyOrg package (Martinez-Sanchez, Baumeister et al. 2022) to simulate a random distribution of the identified molecules within the cellular cytosol as previously described in (Engel, Schaffer et al. 2015). Mesh structures of the molecules in question were prepared using VTK (Schroeder, Avila et al. 2000).

7 Table of figures

8 Abbreviations

2D	Tow dimensional
37 °C cells	Exponentially growing *E. coli* BL21(DE3) cells
3D	Three dimensional
Å	Angstrom
AAA-proteins	ATPases associated with diverse cellular activitie
ADP	Adenosine diphosphate
ADP·BeF_x	Adenosine diphosphate beryllium fluorid, an ADP-Pi mimic
ALS	Amyotrophic lateral sclerosis
ap	Apical domain
ART	Algebraic reconstruction techniques
Asp	Aspartic acid
ATP	Adenosine triphosphate
AWI	Air-water-interface
C. crescentus	*Caulobacter crescentus*
CCD	Charge-coupled device
CEMOVIS	Cryo-EM of vitreous sections
CMOS	Complementary metal-oxide-semiconductor
CNN	Convolutional neural network
Cryo-EM	Cryo electron microscopy
cryo-ET	cryo electron tomography
CTD	C-terminal domain
CTF	contrast transfer function
DED	Direct electron detector
DNA	Desoxyribonucleic acid
DQE	Detector quantum efficiency
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
DTT	Dithiothreitol

E. coli	*Escherichia Coli*
EDTA	Ethylenediaminetetraacetic
EELS	Electron energy loss spectroscopy
EL:ES_1	'bullet' shaped complexes consisting of 14 GroEL monomers and 7 GroES monomes bound to one side of the double ring
EL:ES_2	'football' shaped complexes consisting of 14 GroEL monomers and 14 GroES monomes bound to both side of the double ring
EL+ cells	Cellular system where GroEL was overexpressed, while ES levels remained endogenous
EM	Electron microscopy
eq	equatorial domain
ER	Endoplasmic reticulum
eV	Electron volt
FEG	Field emission gun
FIB	Focused Ion beam
FSC	Fourier Shell Correlation
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
Glu	Glutamic acid
GroEL	Chaperonine
GroES	Co chaperonine to GroEL
HEPES-KOH	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid potassium hydroxide buffer
HMK	Hepes-Magnesium-Potassium buffer
HRP	Horseradish peroxidase
HS	Heat stress

HS cells	*E. coli* BL21(DE3) cells at 46 °C for 2 hours
Hsp10	Heat shock protein of an approximate size of 10 kDa
Hsp40	Heat shock protein of an approximate size of 40 kDa
Hsp60	Heat shock protein of an approximate size of 60 kDa
Hsp70	Heat shock protein of an approximate size of 70 kDa
Hsp90	Heat shock protein 90
HtpG	High Temperature Protein G
iBAQ	Intensity-based absolute quantification
int	Intermediate domain
IP	Immunoprecipitated
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kDa	Kilo Dalton
km	Kilo meter
kV	Kilo volt
LA	lactalbumin
LB	Luria-Bertani medium
LC/MS-MS	Liquid chromatography-tandem mass spectrometry
LED	Light emitting diode
M. mazei	*Methanosarcina mazei*
MD	Middle domain
MetK	S-Adenosylmethionine synthetase
MetK cells	Cells that mildly overexpress GroEL and GroES as well as strongly overexpress the obligate substrate protein MetK
mg	Milli gram
ml	Milli Liter
mM	Milli molar

mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
NAC	Nascent chain–associated complex
NBD	Nucleotide-binding domain
NEF	Nucleotide exchange factor
nm	Nano meter
nm	Nano meter
NMR	Nuclear magnetic resonance
NTD	N-terminal domain
OD_{600}	Optical density at a wave length of 600 nm
PDB	Protein Data Bank
Pfd	Prefoldin
Phe	Phenylalanine
PhLP	Phosducin-like protein
Pi	Inorganic phosphate
PVDF	Polyvinylidene fluoride
RAC	Ribosome-associated complex
RuBisCo	Ribulose-1,5-bisphosphate carboxylase
SBD	Substrate-binding domain
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SE	Secondary electrons
sec	Seconds
SEM	Scanning Electron Microscopy
sHsp	Small heat shock protein
SIRT	Simultaneous iterative reconstruction techniques
SNR	Signal-to-noise ratio
SP	Substrate protein
SPA	Single particle analysis
STA	Subtomogram averaging
STED	Stimulated emission depletion

STEM	Scanning transmission electron microscopy
TBST	Tris-buffered saline with Tween 20
TEM	Transmission electron microscope
TF	Trigger factor
TRiC	TCP-1 ring complex
Tyr	Tryptophan
UI	User interface
w/v	Weight per volume
WBP	Weighted back-projection
µm	Micro meter

9 Bibliography

Al-Amoudi, A., J. J. Chang, A. Leforestier, A. McDowall, L. M. Salamin, L. P. Norlén, K. Richter, N. S. Blanc, D. Studer and J. Dubochet (2004). "Cryo-electron microscopy of vitreous sections." The EMBO journal **23**(18): 3583-3588.

Albert, S., M. Schaffer, F. Beck, S. Mosalaganti, S. Asano, H. F. Thomas, J. M. Plitzko, M. Beck, W. Baumeister and B. D. Engel (2017). "Proteasomes tether to two distinct sites at the nuclear pore complex." Proceedings of the National Academy of Sciences **114**(52): 13726-13731.

Albert, S., W. Wietrzynski, C.-W. Lee, M. Schaffer, F. Beck, J. M. Schuller, P. A. Salomé, J. M. Plitzko, W. Baumeister and B. D. Engel (2020). "Direct visualization of degradation microcompartments at the ER membrane." Proceedings of the National Academy of Sciences **117**(2): 1069-1080.

Anfinsen, C. (1972). "Studies of the principles that govern the folding of protein chains (Nobel Lecture)." Norstedt & Sons, Stockholm.

Anfinsen, C. B. (1973). "Principles that govern the folding of protein chains." Science **181**(4096): 223-230.

Athanasiou, D., M. Aguila, J. Bellingham, W. Li, C. McCulley, P. J. Reeves and M. E. Cheetham (2018). "The molecular and cellular basis of rhodopsin retinitis pigmentosa reveals potential strategies for therapy." Progress in retinal and eye research **62**: 1-23.

Azem, A., M. Kessel and P. Goloubinoff (1994). "Characterization of a functional GroEL14(GroES7)2 chaperonin hetero-oligomer." Science **265**(5172): 653-656.

Baker, M. L., J. Zhang, S. J. Ludtke and W. Chiu (2010). "Cryo-EM of macromolecular assemblies at near-atomic resolution." Nat Protoc **5**(10): 1697-1708.

Balchin, D., M. Hayer-Hartl and F. U. Hartl (2016). "In vivo aspects of protein folding and quality control." Science **353**(6294): aac4354.

Balchin, D., M. Hayer-Hartl and F. U. Hartl (2020). "Recent advances in understanding catalysis of protein folding by molecular chaperones." FEBS Lett **594**(17): 2770-2781.

Bäuerlein, F. J. and W. Baumeister (2021). "Towards visual proteomics at high resolution." Journal of molecular biology **433**(20): 167187.

Bäuerlein, F. J. B. (2018). In Situ Architecture and Cellular Interactions of PolyQ Inclusions, Technische Universität München.

Bäuerlein, F. J. B., M. Renner, D. E. Chami, S. E. Lehnart, J. C. Pastor-Pareja and R. Fernández-Busnadiego (2021). Cryo-electron tomography of large biological specimens vitrified by plunge freezing, Cold Spring Harbor Laboratory.

Beckert, B., M. Turk, A. Czech, O. Berninghausen, R. Beckmann, Z. Ignatova, J. M. Plitzko and D. N. Wilson (2018). "Structure of a hibernating 100S ribosome reveals an inactive conformation of the ribosomal protein S1." Nature Microbiology **3**(10): 1115-1121.

Bepler, T., A. Morin, M. Rapp, J. Brasch, L. Shapiro, A. J. Noble and B. Berger (2019). "Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs." Nat Methods **16**(11): 1153-1160.

Biyani, N., R. D. Righetto, R. McLeod, D. Caujolle-Bert, D. Castano-Diez, K. N. Goldie and H. Stahlberg (2017). "Focus: The interface between data collection and data processing in cryo-EM." Journal of Structural Biology **198**(2): 124-133.

Bond, C., A. N. Santiago-Ruiz, Q. Tang and M. Lakadamyali (2022). "Technological advances in super-resolution microscopy to study cellular processes." Molecular Cell **82**(2): 315-332.

Brandt, F., S. A. Etchells, J. O. Ortiz, A. H. Elcock, F. U. Hartl and W. Baumeister (2009). "The native 3D organization of bacterial polysomes." Cell **136**(2): 261-271.

Buchholz, T.-O., M. Jordan, G. Pigino and F. Jug (2019). Cryo-care: content-aware image restoration for cryo-transmission electron microscopy data. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), IEEE.
Calloni, G., T. Chen, S. M. Schermann, H.-c. Chang, P. Genevaux, F. Agostini, G. G. Tartaglia, M. Hayer-Hartl and F. U. Hartl (2012). "DnaK functions as a central hub in the E. coli chaperone network." Cell reports **1**(3): 251-264.
Castaño-Díez, D., M. Kudryashev, M. Arheit and H. Stahlberg (2012). "Dynamo: a flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments." Journal of structural biology **178**(2): 139-151.
Castaño-Díez, D., M. Kudryashev and H. Stahlberg (2017). "Dynamo Catalogue: Geometrical tools and data management for particle picking in subtomogram averaging of cryo-electron tomograms." Journal of structural biology **197**(2): 135-144.
Chaillet, M. L., G. van der Schot, I. Gubins, S. Roet, R. C. Veltkamp and F. Förster (2023). "Extensive Angular Sampling Enables the Sensitive Localization of Macromolecules in Electron Tomograms." International Journal of Molecular Sciences **24**(17): 13375.
Chaudhry, C., A. L. Horwich, A. T. Brunger and P. D. Adams (2004). "Exploring the structural dynamics of the E.coli chaperonin GroEL using translation-libration-screw crystallographic refinement of intermediate states." J Mol Biol **342**(1): 229-245.
Chaudhuri, T. K., V. K. Verma and A. Maheshwari (2009). "GroEL assisted folding of large polypeptide substrates in Escherichia coli: Present scenario and assignments for the future." Prog Biophys Mol Biol **99**(1): 42-50.
Chen, D. H., D. Madan, J. Weaver, Z. Lin, G. F. Schroder, W. Chiu and H. S. Rye (2013). "Visualizing GroEL/ES in the act of encapsulating a folding protein." Cell **153**(6): 1354-1365.
Cheng, A., C. Negro, J. F. Bruhn, W. J. Rice, S. Dallakyan, E. T. Eng, D. G. Waterman, C. S. Potter and B. Carragher (2021). "Leginon: New features and applications." Protein Science **30**(1): 136-150.
Choi, S. I., S. Kwon, A. Son, H. Jeong, K. H. Kim and B. L. Seong (2013). "Protein folding in vivo revisited." Curr Protein Pept Sci **14**(8): 721-733.
Clare, D. K., D. Vasishtan, S. Stagg, J. Quispe, G. W. Farr, M. Topf, A. L. Horwich and H. R. Saibil (2012). "ATP-triggered conformational changes delineate substrate-binding and -folding mechanics of the GroEL chaperonin." Cell **149**(1): 113-123.
Costa, T. R., A. Ignatiou and E. V. Orlova (2017). "Structural analysis of protein complexes by cryo electron microscopy." Bacterial Protein Secretion Systems: Methods and Protocols: 377-413.
Cruz-Leon, S., T. Majtner, P. C. Hoffmann, J. P. Kreysing, M. W. Tuijtel, S. L. Schaefer, K. Geissler, M. Beck, B. Turonova and G. Hummer (2023). "High-confidence 3D template matching for cryo-electron tomography." bioRxiv: 2023.2009. 2005.556310.
Dandey, V. P., H. Wei, Z. Zhang, Y. Z. Tan, P. Acharya, E. T. Eng, W. J. Rice, P. A. Kahn, C. S. Potter and B. Carragher (2018). "Spotiton: New features and applications." Journal of Structural Biology **202**(2): 161-169.
Danev, R., M. Belousoff, Y.-L. Liang, X. Zhang, F. Eisenstein, D. Wootten and P. M. Sexton (2021). "Routine sub-2.5 Å cryo-EM structure determination of GPCRs." Nature Communications **12**(1): 4333.
de Teresa-Trueba, I., S. K. Goetz, A. Mattausch, F. Stojanovska, C. E. Zimmerli, M. Toro-Nahuelpan, D. W. Cheng, F. Tollervey, C. Pape and M. Beck (2023). "Convolutional networks for supervised mining of molecular patterns within cellular context." Nature Methods **20**(2): 284-294.

Deuerling, E., M. Gamerdinger and S. G. Kreft (2019). "Chaperone interactions at the ribosome." Cold Spring Harbor perspectives in biology **11**(11): a033977.
Dietrich, H. M., R. D. Righetto, A. Kumar, W. Wietrzynski, R. Trischler, S. K. Schuller, J. Wagner, F. M. Schwarz, B. D. Engel, V. Müller and J. M. Schuller (2022). "Membrane-anchored HDCR nanowires drive hydrogen-powered CO2 fixation." Nature **607**(7920): 823-830.
Dobson, C. M., A. Sali and M. Karplus (1998). "Protein Folding: A Perspective from Theory and Experiment." Angew Chem Int Ed Engl **37**(7): 868-893.
DSMZ. (2023). "Caulobacter vibrioides growth and purchasing information." from https://www.dsmz.de/collection/catalogue/details/culture/DSM-4727.
Dubochet, J. (2012). "Cryo-EM—the first thirty years." Journal of microscopy **245**(3): 221-224.
Dubochet, J., M. Adrian, J.-J. Chang, J.-C. Homo, J. Lepault, A. W. McDowall and P. Schultz (1988). "Cryo-electron microscopy of vitrified specimens." Quarterly reviews of biophysics **21**(2): 129-228.
Dung, N. H. T., G. Perone, R. Vazzana, F. K. Don, M. Silva, S. Sorrentino, P. Swuec, F. Leroux, N. Kalebic, F. Coscia and P. S. Erdmann (2023). Serialized On-grid Lift-In Sectioning for Tomography (SOLIST), Cold Spring Harbor Laboratory.
Eisenstein, F., H. Yanagisawa, H. Kashihara, M. Kikkawa, S. Tsukita and R. Danev (2023). "Parallel cryo electron tomography on in situ lamellae." Nature Methods **20**(1): 131-138.
Electron, D. (2023). "CCD CAMERA V. DIRECT DETECTION." Retrieved 20.09, 2023, from https://directelectron.com/de-series-cameras/.
Ellis, R. J. and A. P. Minton (2006). "Protein aggregation in crowded environments."
Ellis, R. J. and A. P. Minton (2006). "Protein aggregation in crowded environments." Biol Chem **387**(5): 485-497.
Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2126-2132.
Engel, A., M. K. Hayer-Hartl, K. N. Goldie, G. Pfeifer, R. Hegerl, S. Muller, A. C. da Silva, W. Baumeister and F. U. Hartl (1995). "Functional significance of symmetrical versus asymmetrical GroEL-GroES chaperonin complexes." Science **269**(5225): 832-836.
Engel, B. D., M. Schaffer, L. Kuhn Cuellar, E. Villa, J. M. Plitzko and W. Baumeister (2015). "Native architecture of the Chlamydomonas chloroplast revealed by in situ cryo-electron tomography." elife **4**: e04889.
Erdmann, P. S., Z. Hou, S. Klumpe, S. Khavnekar, F. Beck, F. Wilfling, J. M. Plitzko and W. Baumeister (2021). "In situ cryo-electron tomography reveals gradient organization of ribosome biogenesis in intact nucleoli." Nat Commun **12**(1): 5364.
Ewalt, K. L., J. P. Hendrick, W. A. Houry and F. U. Hartl (1997). "In vivo observation of polypeptide flux through the bacterial chaperonin system." Cell **90**(3): 491-500.
Fayet, O., T. Ziegelhoffer and C. Georgopoulos (1989). "The groES and groEL heat shock gene products of Escherichia coli are essential for bacterial growth at all temperatures." J Bacteriol **171**(3): 1379-1385.
Fei, X., X. Ye, N. A. LaRonde and G. H. Lorimer (2014). "Formation and structures of GroEL:GroES2 chaperonin footballs, the protein-folding functional form." Proc Natl Acad Sci U S A **111**(35): 12775-12780.
Fujiwara, K., Y. Ishihama, K. Nakahigashi, T. Soga and H. Taguchi (2010). "A systematic survey of in vivo obligate chaperonin-dependent substrates." EMBO J **29**(9): 1552-1564.
Gade, M., L. L. Tan, A. M. Damry, M. Sandhu, J. S. Brock, A. Delaney, A. Villar-Briones, C. J. Jackson and P. Laurino (2021). "Substrate Dynamics Contribute to

Enzymatic Specificity in Human and Bacterial Methionine Adenosyltransferases." JACS Au **1**(12): 2349-2360.
Gardner, S., M. C. Darrow, N. Lukoyanova, K. Thalassinos and H. R. Saibil (2023). "Structural basis of substrate progression through the bacterial chaperonin cycle." Proceedings of the National Academy of Sciences **120**(50): e2308933120.
Gasteiger, E., C. Hoogland, A. Gattiker, S. e. Duvaud, M. R. Wilkins, R. D. Appel and A. Bairoch (2005). Protein Identification and Analysis Tools on the ExPASy Server. The Proteomics Protocols Handbook. J. M. Walker. Totowa, NJ, Humana Press**:** 571-607.
Gemmer, M., M. L. Chaillet, J. van Loenhout, R. Cuevas Arenas, D. Vismpas, M. Gröllers-Mulderij, F. A. Koh, P. Albanese, R. A. Scheltema and S. C. Howes (2023). "Visualization of translation and protein biogenesis at the ER membrane." Nature **614**(7946): 160-167.
Georgopoulos, C. (2006). "Toothpicks, serendipity and the emergence of the Escherichia coli DnaK (Hsp70) and GroEL (Hsp60) chaperone machines." Genetics **174**(4): 1699-1707.
Gestaut, D., A. Limatola, L. Joachimiak and J. Frydman (2019). "The ATP-powered gymnastics of TRiC/CCT: an asymmetric protein folding machine with a symmetric origin story." Curr Opin Struct Biol **55**: 50-58.
Gestaut, D., Y. Zhao, J. Park, B. Ma, A. Leitner, M. Collier, G. Pintilie, S. H. Roh, W. Chiu and J. Frydman (2022). "Structural visualization of the tubulin folding pathway directed by human chaperonin TRiC/CCT." Cell **185**(25): 4770-4787 e4720.
Glaeser, R. M. (2018). "Proteins, interfaces, and cryo-EM grids." Current opinion in colloid & interface science **34**: 1-8.
Gloge, F., A. H. Becker, G. Kramer and B. Bukau (2014). "Co-translational mechanisms of protein maturation." Current opinion in structural biology **24**: 24-33.
Grant, T. and N. Grigorieff (2015). "Measuring the optimal exposure for single particle cryo-EM using a 2.6 A reconstruction of rotavirus VP6." Elife **4**: e06980.
Gruber, R. and A. Horovitz (2016). "Allosteric mechanisms in chaperonin machines." Chemical Reviews **116**(11): 6588-6606.
Grunewald, K., O. Medalia, A. Gross, A. C. Steven and W. Baumeister (2003). "Prospects of electron cryotomography to visualize macromolecular complexes inside cellular compartments: implications of crowding." Biophys Chem **100**(1-3): 577-591.
Guo, Q., C. Lehmer, A. Martinez-Sanchez, T. Rudack, F. Beck, H. Hartmann, M. Perez-Berlanga, F. Frottin, M. S. Hipp, F. U. Hartl, D. Edbauer, W. Baumeister and R. Fernandez-Busnadiego (2018). "In Situ Structure of Neuronal C9orf72 Poly-GA Aggregates Reveals Proteasome Recruitment." Cell **172**(4): 696-705 e612.
Guzman, L. M., D. Belin, M. J. Carson and J. Beckwith (1995). "Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter." J Bacteriol **177**(14): 4121-4130.
Hagen, W. J. H., W. Wan and J. A. G. Briggs (2017). "Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging." J Struct Biol **197**(2): 191-198.
Haldar, S., A. J. Gupta, X. Yan, G. Milicic, F. U. Hartl and M. Hayer-Hartl (2015). "Chaperonin-Assisted Protein Folding: Relative Population of Asymmetric and Symmetric GroEL:GroES Complexes." J Mol Biol **427**(12): 2244-2255.
Han, W., M. Jin, C. Liu, Q. Zhao, S. Wang, Y. Wang, Y. Yin, C. Peng, Y. Wang and Y. Cong (2023). "Structural basis of plp2-mediated cytoskeletal protein folding by TRiC/CCT." Science Advances **9**(11): eade1207.
Hartl, F. U. (1996). "Molecular chaperones in cellular protein folding." Nature **381**(6583): 571-579.

Hartl, F. U. (2017). "Protein Misfolding Diseases." Annual Review of Biochemistry **86**(1): 21-26.
Hartl, F. U., A. Bracher and M. Hayer-Hartl (2011). "Molecular chaperones in protein folding and proteostasis." Nature **475**(7356): 324-332.
Hayer-Hartl, M., A. Bracher and F. U. Hartl (2016). "The GroEL-GroES Chaperonin Machine: A Nano-Cage for Protein Folding." Trends Biochem Sci **41**(1): 62-76.
Hayer-Hartl, M. K., J. Martin and F. U. Hartl (1995). "Asymmetrical interaction of GroEL and GroES in the ATPase cycle of assisted protein folding." Science **269**(5225): 836-841.
Hayer-Hartl, M. K., F. Weber and F. U. Hartl (1996). "Mechanism of chaperonin action: GroES binding and release can drive GroEL-mediated protein folding in the absence of ATP hydrolysis." EMBO J **15**(22): 6111-6121.
Henderson, R. (1992). "Image contrast in high-resolution electron microscopy of biological macromolecules: TMV in ice." Ultramicroscopy **46**(1-4): 1-18.
Henkelman, R. and F. Ottensmeyer (1974). "An energy filter for biological electron microscopy." Journal of Microscopy **102**(1): 79-94.
Himes, B. A. and P. Zhang (2018). "emClarity: software for high-resolution cryo-electron tomography and subtomogram averaging." Nature methods **15**(11): 955-961.
Hoffmann, P. C., J. P. Kreysing, I. Khusainov, M. W. Tuijtel, S. Welsch and M. Beck (2022). "Structures of the eukaryotic ribosome and its translational states in situ." Nat Commun **13**(1): 7435.
Horovitz, A., T. H. Reingewertz, J. Cuellar and J. M. Valpuesta (2022). "Chaperonin Mechanisms: Multiple and (Mis)Understood?" Annu Rev Biophys **51**(1): 115-133.
Horovitz, A. and K. R. Willison (2005). "Allosteric regulation of chaperonins." Curr Opin Struct Biol **15**(6): 646-651.
Horwich, A. L. and W. A. Fenton (2020). "Chaperonin-assisted protein folding: a chronologue." Q Rev Biophys **53**: e4.
Instruments, E. (2023). "Resolution Criteria." from https://www.edinst.com/de/news/the-rayleigh-criterion-for-microscope-resolution/.
Ishii, N. (2017). GroEL and the GroEL-GroES Complex. Macromolecular Protein Complexes: Structure and Function. J. R. Harris and J. Marles-Wright. Cham, Springer International Publishing**:** 483-504.
Ishimoto, T., K. Fujiwara, T. Niwa and H. Taguchi (2014). "Conversion of a chaperonin GroEL-independent protein into an obligate substrate." J Biol Chem **289**(46): 32073-32080.
Jiang, W., J. Wagner, W. Du, J. Plitzko, W. Baumeister, F. Beck and Q. Guo (2022). "A transformation clustering algorithm and its application in polyribosomes structural profiling." Nucleic Acids Research **50**(16): 9001-9011.
Jumper, J., R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli and D. Hassabis (2021). "Highly accurate protein structure prediction with AlphaFold." Nature **596**(7873): 583-589.
Kaiser, C. M., H.-C. Chang, V. R. Agashe, S. K. Lakshmipathy, S. A. Etchells, M. Hayer-Hartl, F. U. Hartl and J. M. Barral (2006). "Real-time observation of trigger factor function on translating ribosomes." Nature **444**(7118): 455-460.
Kelley, K., A. M. Raczkowski, O. Klykov, P. Jaroenlak, D. Bobe, M. Kopylov, E. T. Eng, G. Bhabha, C. S. Potter, B. Carragher and A. J. Noble (2022). "Waffle Method: A

general and flexible approach for improving throughput in FIB-milling." Nature Communications **13**(1): 1857.
Kelly, J. J., D. Tranter, E. Pardon, G. Chi, H. Kramer, L. Happonen, K. M. Knee, J. M. Janz, J. Steyaert and C. Bulawa (2022). "Snapshots of actin and tubulin folding inside the TRiC chaperonin." Nature structural & molecular biology **29**(5): 420-429.
Kerner, M. J., D. J. Naylor, Y. Ishihama, T. Maier, H. C. Chang, A. P. Stines, C. Georgopoulos, D. Frishman, M. Hayer-Hartl, M. Mann and F. U. Hartl (2005). "Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli." Cell **122**(2): 209-220.
Khavnekar, S., V. Vrbovská, M. Zaoralová, R. Kelley, F. Beck, S. Klumpe, A. Kotecha, J. Plitzko and P. S. Erdmann (2022). "Optimizing cryo-FIB lamellas for sub-5Å in situ structural biology." BioRxiv: 2022.2006. 2016.496417.
Kim, H., J. Park, S. Lim, S. H. Jun, M. Jung and S. H. Roh (2022). "Cryo-EM structures of GroEL:ES(2) with RuBisCO visualize molecular contacts of encapsulated substrates in a double-cage chaperonin." iScience **25**(1): 103704.
Kim, Y. E., M. S. Hipp, A. Bracher, M. Hayer-Hartl and F. U. Hartl (2013). "Molecular chaperone functions in protein folding and proteostasis." Annu Rev Biochem **82**: 323-355.
Klebl, D. P., M. S. C. Gravett, D. Kontziampasis, D. J. Wright, R. S. Bon, D. C. F. Monteiro, M. Trebbin, F. Sobott, H. D. White, M. C. Darrow, R. F. Thompson and S. P. Muench (2020). "Need for Speed: Examining Protein Behavior during CryoEM Grid Preparation at Different Timescales." Structure **28**(11): 1238-1248 e1234.
Klumpe, S., H. K. Fung, S. K. Goetz, I. Zagoriy, B. Hampoelz, X. Zhang, P. S. Erdmann, J. Baumbach, C. W. Muller, M. Beck, J. M. Plitzko and J. Mahamid (2021). "A modular platform for automated cryo-FIB workflows." Elife **10**.
Klumpe, S., H. K. H. Fung, S. K. Goetz, I. Zagoriy, B. Hampoelz, X. Zhang, P. S. Erdmann, J. Baumbach, C. W. Müller, M. Beck, J. M. Plitzko and J. Mahamid (2021). "A modular platform for automated cryo-FIB workflows." eLife **10**: e70506.
Kohl, H. and L. Reimer (2008). "Transmission Electron Microscopy." Springer Series in Optical Sciences **36**.
Koike-Takeshita, A., M. Yoshida and H. Taguchi (2008). "Revisiting the GroEL-GroES reaction cycle via the symmetric intermediate implied by novel aspects of the GroEL(D398A) mutant." J Biol Chem **283**(35): 23774-23781.
Kudryavtseva, S. S., E. B. Pichkur, I. A. Yaroshevich, A. A. Mamchur, I. S. Panina, A. V. Moiseenko, O. S. Sokolova, V. I. Muronetz and T. B. Stanishneva-Konovalova (2021). "Novel cryo-EM structure of an ADP-bound GroEL-GroES complex." Sci Rep **11**(1): 18241.
Kühlbrandt, W. (2014). "The resolution revolution." Science **343**(6178): 1443-1444.
Labbadia, J. and R. I. Morimoto (2015). "The biology of proteostasis in aging and disease." Annual review of biochemistry **84**: 435-464.
Langmore, J. P. and M. F. Smith (1992). "Quantitative energy-filtered electron microscopy of biological molecules in ice." Ultramicroscopy **46**(1-4): 349-373.
Liebschner, D., P. V. Afonine, M. L. Baker, G. Bunkoczi, V. B. Chen, T. I. Croll, B. Hintze, L. W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams and P. D. Adams (2019). "Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix." Acta Crystallogr D Struct Biol **75**(Pt 10): 861-877.
Lin, Z., D. Madan and H. S. Rye (2008). "GroEL stimulates protein folding through forced unfolding." Nature Structural & Molecular Biology **15**(3): 303-311.

Liu, Y.-T., H. Zhang, H. Wang, C.-L. Tao, G.-Q. Bi and Z. H. Zhou (2022). "Isotropic reconstruction for electron tomography with deep learning." Nature communications **13**(1): 6482.
Lovell, S. C., I. W. Davis, W. B. Arendall III, P. I. De Bakker, J. M. Word, M. G. Prisant, J. S. Richardson and D. C. Richardson (2003). "Structure validation by Cα geometry: ϕ, ψ and Cβ deviation." Proteins: Structure, Function, and Bioinformatics **50**(3): 437-450.
Lucas, B. A. and N. Grigorieff (2023). "Quantification of gallium cryo-FIB milling damage in biological lamellae." Proceedings of the National Academy of Sciences **120**(23): e2301852120.
Lucas, B. A., B. A. Himes, L. Xue, T. Grant, J. Mahamid and N. Grigorieff (2021). "Locating macromolecular assemblies in cells by 2D template matching with cisTEM." Elife **10**: e68946.
Mahamid, J., S. Pfeffer, M. Schaffer, E. Villa, R. Danev, L. Kuhn Cuellar, F. Förster, A. A. Hyman, J. M. Plitzko and W. Baumeister (2016). "Visualizing the molecular sociology at the HeLa cell nuclear periphery." Science **351**(6276): 969-972.
Martinez-Sanchez, A., W. Baumeister and V. Lučić (2022). "Statistical spatial analysis for cryo-electron tomography." Computer methods and programs in biomedicine **218**: 106693.
Martynowycz, M. W., M. T. Clabbers, J. Unge, J. Hattne and T. Gonen (2021). "Benchmarking the ideal sample thickness in cryo-EM." Proceedings of the National Academy of Sciences **118**(49): e2108884118.
Masahide, K. (2022). "Number of realesed structures in PDB per year." from https://twitter.com/mkikkawa/status/1518391846644645888.
Mastronarde, D. N. (1997). "Dual-axis tomography: an approach with alignment methods that preserve resolution." Journal of structural biology **120**(3): 343-352.
Mastronarde, D. N. (2005). "Automated electron microscope tomography using robust prediction of specimen movements." J Struct Biol **152**(1): 36-51.
Mastronarde, D. N. and S. R. Held (2017). "Automated tilt series alignment and tomographic reconstruction in IMOD." Journal of structural biology **197**(2): 102-113.
Mayhew, M., A. C. R. Da Silva, J. Martin, H. Erdjument-Bromage, P. Tempst and F. U. Hartl (1996). "Protein folding in the central cavity of the GroEL–GroES chaperonin complex." Nature **379**(6564): 420-426.
Mazor, K. M., L. Dong, Y. Mao, R. V. Swanda, S.-B. Qian and M. H. Stipanuk (2018). "Effects of single amino acid deficiency on mRNA translation are markedly different for methionine versus leucine." Scientific Reports **8**(1): 8076.
Moebel, E., A. Martinez-Sanchez, L. Lamm, R. D. Righetto, W. Wietrzynski, S. Albert, D. Larivière, E. Fourmentin, S. Pfeffer and J. Ortiz (2021). "Deep learning improves macromolecule identification in 3D cellular cryo-electron tomograms." Nature methods **18**(11): 1386-1394.
Mosalaganti, S., A. Obarska-Kosinska, M. Siggel, R. Taniguchi, B. Turoňová, C. E. Zimmerli, K. Buczak, F. H. Schmidt, E. Margiotta, M.-T. Mackmull, W. J. H. Hagen, G. Hummer, J. Kosinski and M. Beck (2022). "AI-based structure prediction empowers integrative structural analysis of human nuclear pores." Science **376**(6598): eabm9506.
Nelson, D. L. and M. M. Cox (2008). "Absolute ultimate guide for lehninger principles of biochemistry." UK: Palgrave Macmillan.
Niwa, T., K. Fujiwara and H. Taguchi (2016). "Identification of novel in vivo obligate GroEL/ES substrates based on data from a cell-free proteomics approach." FEBS Lett **590**(2): 251-257.

Noble, A. J., H. Wei, V. P. Dandey, Z. Zhang, Y. Z. Tan, C. S. Potter and B. Carragher (2018). "Reducing effects of particle adsorption to the air–water interface in cryo-EM." Nature methods **15**(10): 793-795.
O'Reilly, F. J., L. Xue, A. Graziadei, L. Sinn, S. Lenz, D. Tegunov, C. Blötz, N. Singh, W. J. Hagen and P. Cramer (2020). "In-cell architecture of an actively transcribing-translating expressome." Science **369**(6503): 554-557.
Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin (2004). "UCSF Chimera--a visualization system for exploratory research and analysis." J Comput Chem **25**(13): 1605-1612.
Pettersen, E. F., T. D. Goddard, C. C. Huang, E. C. Meng, G. S. Couch, T. I. Croll, J. H. Morris and T. E. Ferrin (2021). "UCSF ChimeraX: Structure visualization for researchers, educators, and developers." Protein Sci **30**(1): 70-82.
Plitzko, J. and D. Bollschweiler (2024). Fundamentals of instrumentation and electron optics for cryo-electron tomography. Focus on structural biology. Briegel A. and F. F., Springer.
Prats, C., T. E. Graham and J. Shearer (2018). "The dynamic life of the glycogen granule." J Biol Chem **293**(19): 7089-7098.
Preissler, S. and E. Deuerling (2012). "Ribosome-associated chaperones as key players in proteostasis." Trends in biochemical sciences **37**(7): 274-283.
Punjani, A. (2021). "Real-time cryo-EM structure determination." Microscopy and Microanalysis **27**(S1): 1156-1157.
Punjani, A. and D. Fleet (2022). "3D Flexible Refinement: Structure and Motion of Flexible Proteins from Cryo-EM." Microscopy and Microanalysis **28**(S1): 1218-1218.
Punjani, A. and D. J. Fleet (2021). "3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM." Journal of structural biology **213**(2): 107702.
Punjani, A., J. L. Rubinstein, D. J. Fleet and M. A. Brubaker (2017). "cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination." Nat Methods **14**(3): 290-296.
Punjani, A., J. L. Rubinstein, D. J. Fleet and M. A. Brubaker (2017). "cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination." Nature Methods **14**(3): 290-296.
Ramachandran, G. N., C. Ramakrishnan and V. Sasisekharan (1963). "Stereochemistry of polypeptide chain configurations." Journal of Molecular Biology **7**(1): 95-99.
Reboul, C. F., M. Eager, D. Elmlund and H. Elmlund (2018). "Single-particle cryo-EM—Improved ab initio 3D reconstruction with SIMPLE/PRIME." Protein Science **27**(1): 51-61.
Rice, G., T. Wagner, M. Stabrin, O. Sitsel, D. Prumbaum and S. Raunser (2023). "TomoTwin: generalized 3D localization of macromolecules in cryo-electron tomograms with structural data mining." Nature Methods: 1-10.
Richardson., J. a. D. (2023). from https://en.wikipedia.org/wiki/Ramachandran_plot#/media/File:Ramachandran_plot_general_100K.jpg.
Rigort, A., F. J. Bauerlein, E. Villa, M. Eibauer, T. Laugks, W. Baumeister and J. M. Plitzko (2012). "Focused ion beam micromachining of eukaryotic cells for cryoelectron tomography." Proc Natl Acad Sci U S A **109**(12): 4449-4454.
Rigort, A., E. Villa, F. J. B. Bäuerlein, B. D. Engel and J. M. Plitzko (2012). Chapter 14 - Integrative Approaches for Cellular Cryo-electron Tomography: Correlative Imaging and Focused Ion Beam Micromachining. Methods in Cell Biology. T. Müller-Reichert and P. Verkade, Academic Press. **111:** 259-281.

Rohou, A. and N. Grigorieff (2015). "CTFFIND4: Fast and accurate defocus estimation from electron micrographs." J Struct Biol **192**(2): 216-221.
Rosenzweig, R., N. B. Nillegoda, M. P. Mayer and B. Bukau (2019). "The Hsp70 chaperone network." Nature reviews molecular cell biology **20**(11): 665-680.
Ross, J. F. and M. Orlowski (1982). "Growth-rate-dependent adjustment of ribosome function in chemostat-grown cells of the fungus Mucor racemosus." J Bacteriol **149**(2): 650-653.
Rye, H. S., A. M. Roseman, S. Chen, K. Furtak, W. A. Fenton, H. R. Saibil and A. L. Horwich (1999). "GroEL-GroES cycling: ATP and nonnative polypeptide direct alternation of folding-active rings." Cell **97**(3): 325-338.
Saibil, H. R., W. A. Fenton, D. K. Clare and A. L. Horwich (2013). "Structure and allostery of the chaperonin GroEL." J Mol Biol **425**(9): 1476-1487.
Sameshima, T., T. Ueno, R. Iizuka, N. Ishii, N. Terada, K. Okabe and T. Funatsu (2008). "Football- and bullet-shaped GroEL-GroES complexes coexist during the reaction cycle." J Biol Chem **283**(35): 23765-23773.
Scheres, S. H. (2016). "Processing of structurally heterogeneous cryo-EM data in RELION." Methods in enzymology **579**: 125-157.
Schiøtz, O. H., C. J. O. Kaiser, S. Klumpe, D. R. Morado, M. Poege, J. Schneider, F. Beck, C. Thompson and M. J. Plitzko (2023). Serial Lift-Out – Sampling the Molecular Anatomy of Whole Organisms, Cold Spring Harbor Laboratory.
Schmidt, M., K. Rutkat, R. Rachel, G. Pfeifer, R. Jaenicke, P. Viitanen, G. Lorimer and J. Buchner (1994). "Symmetric complexes of GroE chaperonins as part of the functional cycle." Science **265**(5172): 656-659.
Schroeder, W. J., L. S. Avila and W. Hoffman (2000). "Visualizing with VTK: a tutorial." IEEE Computer graphics and applications **20**(5): 20-27.
Schwanhausser, B., D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen and M. Selbach (2011). "Global quantification of mammalian gene expression control." Nature **473**(7347): 337-342.
Sharma, S., K. Chakraborty, B. K. Muller, N. Astola, Y. C. Tang, D. C. Lamb, M. Hayer-Hartl and F. U. Hartl (2008). "Monitoring protein conformation along the pathway of chaperonin-assisted folding." Cell **133**(1): 142-153.
Stagg, S. M., A. J. Noble, M. Spilman and M. S. Chapman (2014). "ResLog plots as an empirical metric of the quality of cryo-EM reconstructions." Journal of structural biology **185**(3): 418-426.
Suloway, C., J. Pulokas, D. Fellmann, A. Cheng, F. Guerra, J. Quispe, S. Stagg, C. S. Potter and B. Carragher (2005). "Automated molecular microscopy: the new Leginon system." Journal of structural biology **151**(1): 41-60.
Taguchi, H. (2015). "Reaction Cycle of Chaperonin GroEL via Symmetric "Football" Intermediate." J Mol Biol **427**(18): 2912-2918.
Tang, Y. C., H. C. Chang, A. Roeben, D. Wischnewski, N. Wischnewski, M. J. Kerner, F. U. Hartl and M. Hayer-Hartl (2006). "Structural features of the GroEL-GroES nano-cage required for rapid folding of encapsulated protein." Cell **125**(5): 903-914.
Tegunov, D. and P. Cramer (2019). "Real-time cryo-electron microscopy data preprocessing with Warp." Nat Methods **16**(11): 1146-1152.
Tegunov, D., L. Xue, C. Dienemann, P. Cramer and J. Mahamid (2021). "Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 A in cells." Nat Methods **18**(2): 186-193.
Toro-Nahuelpan, M., I. Zagoriy, F. Senger, L. Blanchoin, M. Théry and J. Mahamid (2020). "Tailoring cryo-electron microscopy grids by photo-micropatterning for in-cell structural studies." Nature Methods **17**(1): 50-54.

Tuijtel, M. W., J. P. Kreysing, S. Welsch, G. Hummer, M. Beck and B. Turoňová (2023). "Thinner is not always better: Optimising cryo lamellae for subtomogram averaging." bioRxiv: 2023.2007. 2031.551274.
Turonova, B., F. K. M. Schur, W. Wan and J. A. G. Briggs (2017). "Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4A." J Struct Biol **199**(3): 187-195.
van den Hoek, H. G. (2022). In Situ structural studies of the ciliary pore, Technische Universität München.
Vila-Sanjurjo, A., W. K. Ridgeway, V. Seymaner, W. Zhang, S. Santoso, K. Yu and J. H. Cate (2003). "X-ray crystal structures of the WT and a hyper-accurate ribosome from Escherichia coli." Proc Natl Acad Sci U S A **100**(15): 8682-8687.
WA, R. (1959). "Principles of optics."."
Wagner, T., F. Merino, M. Stabrin, T. Moriya, C. Antoni, A. Apelbaum, P. Hagel, O. Sitsel, T. Raisch, D. Prumbaum, D. Quentin, D. Roderer, S. Tacke, B. Siebolds, E. Schubert, T. R. Shaikh, P. Lill, C. Gatsogiannis and S. Raunser (2019). "SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM." Commun Biol **2**(1): 218.
Wan, W. (2020). "williamnwan/STOPGAP: STOPGAP 0.7. 1." Zenodo **1**.
Wan, W. (2020). "williamnwan/TOMOMAN: TOMOMAN 08042020." Zenodo.
Wan, W. and J. A. Briggs (2016). "Cryo-electron tomography and subtomogram averaging." Methods in enzymology **579**: 329-367.
Wan, W. and J. A. G. Briggs (2016). Chapter Thirteen - Cryo-Electron Tomography and Subtomogram Averaging. Methods in Enzymology. R. A. Crowther, Academic Press. **579:** 329-367.
Wang, J. and L. Chen (2003). "Domain motions in GroEL upon binding of an oligopeptide." J Mol Biol **334**(3): 489-499.
Waudby, C. A., C. M. Dobson and J. Christodoulou (2019). "Nature and Regulation of Protein Folding on the Ribosome." Trends Biochem Sci **44**(11): 914-926.
Weaver, J., M. Jiang, A. Roth, J. Puchalla, J. Zhang and H. S. Rye (2017). "GroEL actively stimulates folding of the endogenous substrate protein PepQ." Nat Commun **8**: 15934.
Weis, F. and W. J. Hagen (2020). "Combining high throughput and high quality for cryo-electron microscopy data collection." Acta Crystallographica Section D: Structural Biology **76**(8): 724-728.
Weissenberger, G., F. Nijpels, R. Henderikx, B. Beulen, R. Ravelli and P. Peters (2020). "VitroJet: Bridging the gap in sample prep." Microscopy and Microanalysis **26**(S2): 328-328.
Wickner, S., T. L. Nguyen and O. Genest (2021). "The Bacterial Hsp90 Chaperone: Cellular Functions and Mechanism of Action." Annu Rev Microbiol **75**(1): 719-739.
Williams, D. B., C. B. Carter, D. B. Williams and C. B. Carter (2009). "Electron sources." Transmission electron microscopy: A textbook for materials science: 73-89.
Wolff, G., R. W. Limpens, S. Zheng, E. J. Snijder, D. A. Agard, A. J. Koster and M. Bárcena (2019). "Mind the gap: Micro-expansion joints drastically decrease the bending of FIB-milled cryo-lamellae." Journal of structural biology **208**(3): 107389.
Xu, Z., A. L. Horwich and P. B. Sigler (1997). "The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex." Nature **388**(6644): 741-750.
Xue, L., S. Lenz, M. Zimmermann-Kogadeeva, D. Tegunov, P. Cramer, P. Bork, J. Rappsilber and J. Mahamid (2022). "Visualizing translation dynamics at atomic detail inside a bacterial cell." Nature **610**(7930): 205-211.

Xue, L., C. M. Spahn, M. Schacherl and J. Mahamid (2023). "Structural insights into context-dependent inhibitory mechanisms of chloramphenicol in cells." BioRXiv: 2023.2006.2007.544107.
Xun, L., D. R. Boone and R. A. Mah (1988). "Control of the Life Cycle of Methanosarcina mazei S-6 by Manipulation of Growth Conditions." Appl Environ Microbiol **54**(8): 2064-2068.
Yan, X., Q. Shi, A. Bracher, G. Milicic, A. K. Singh, F. U. Hartl and M. Hayer-Hartl (2018). "GroEL Ring Separation and Exchange in the Chaperonin Reaction." Cell **172**(3): 605-617 e611.
Yang, D., X. Ye and G. H. Lorimer (2013). "Symmetric GroEL:GroES2 complexes are the protein-folding functional form of the chaperonin nanomachine." Proc Natl Acad Sci U S A **110**(46): E4298-4305.
Ye, X. and G. H. Lorimer (2013). "Substrate protein switches GroE chaperonins from asymmetric to symmetric cycling by catalyzing nucleotide exchange." Proc Natl Acad Sci U S A **110**(46): E4289-4297.
You, X., X. Zhang, J. Cheng, Y. Xiao, J. Ma, S. Sun, X. Zhang, H.-W. Wang and S.-F. Sui (2023). "In situ structure of the red algal phycobilisome–PSII–PSI–LHC megacomplex." Nature **616**(7955): 199-206.
Yu, C.-H., Y. Dang, Z. Zhou, C. Wu, F. Zhao, M. S. Sachs and Y. Liu (2015). "Codon usage influences the local rate of translation elongation to regulate co-translational protein folding." Molecular cell **59**(5): 744-754.
Zhang, K. (2016). "Gctf: Real-time CTF determination and correction." J Struct Biol **193**(1): 1-12.
Zhao, L., M. P. Castanie-Cornet, S. Kumar, P. Genevaux, M. Hayer-Hartl and F. U. Hartl (2021). "Bacterial RF3 senses chaperone function in co-translational folding." Mol Cell **81**(14): 2914-2928 e2917.
Zheng, S., G. Wolff, G. Greenan, Z. Chen, F. G. Faas, M. Bárcena, A. J. Koster, Y. Cheng and D. A. Agard (2022). "AreTomo: An integrated software package for automated marker-free, motion-corrected cryo-electron tomographic alignment and reconstruction." Journal of Structural Biology: X **6**: 100068.
Zheng, S. Q., E. Palovcak, J.-P. Armache, K. A. Verba, Y. Cheng and D. A. Agard (2017). "MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy." Nature methods **14**(4): 331-332.
Zheng, S. Q., E. Palovcak, J. P. Armache, K. A. Verba, Y. Cheng and D. A. Agard (2017). "MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy." Nat Methods **14**(4): 331-332.
Zhong, E. D., T. Bepler, B. Berger and J. H. Davis (2021). "CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks." Nat Methods **18**(2): 176-185.
Zimmerli, C. E., M. Allegretti, V. Rantos, S. K. Goetz, A. Obarska-Kosinska, I. Zagoriy, A. Halavatyi, G. Hummer, J. Mahamid and J. Kosinski (2021). "Nuclear pores dilate and constrict in cellulo." Science **374**(6573): eabd9776.
Zivanov, J., T. Nakane, B. O. Forsberg, D. Kimanius, W. J. Hagen, E. Lindahl and S. H. Scheres (2018). "New tools for automated high-resolution cryo-EM structure determination in RELION-3." elife **7**: e42166.
Zivanov, J., T. Nakane and S. H. Scheres (2019). "A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis." IUCrJ **6**(1): 5-17.
Zivanov, J., T. Nakane and S. H. Scheres (2020). "Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1." IUCrJ **7**(2): 253-267.

Zivanov, J., J. Oton, Z. Ke, A. von Kugelgen, E. Pyle, K. Qu, D. Morado, D. Castano-Diez, G. Zanetti, T. A. M. Bharat, J. A. G. Briggs and S. H. W. Scheres (2022). "A Bayesian approach to single-particle electron cryo-tomography in RELION-4.0." Elife **11**: e83724.