Macromolecular machines in motion: Insights from long-timescale molecular dynamics and time-resolved cryo-EM

Lauren Alexandrescu



Scripps Research

Lincoln College
University of Oxford

and

The Scripps Research Institute

A thesis submitted for the degree of
Doctor of Philosophy
(Joint Oxford/TSRI)

October 2025

© 2025 Lauren Alexandrescu

Acknowledgements

I am immensely grateful for the help and support of my supervisors, Phil Biggin and Gabe Lander. To Phil, thank you for your patience and kindness during a particularly challenging period as I undertook research in a new field and country during the time of the pandemic. To Gabe, thank you for your above-and-beyond mentorship. I am truly appreciative of all of the opportunities you provided me with over the years and the genuine care you showed for my growth and success as a scientist. It has been a privilege to work with you, and I will always hold fond memories of my time in your lab.

I would also like to thank my undergraduate supervisor, Simon White, who opened my eyes to the world of research and greatly supported me in the earliest stages of my academic journey.

Thank you to all of the members of SBCB and the Biggin lab, past and present. I am particularly grateful to Professor (SA) Keith Cassidy and Dr. Michael Horrel who taught me foundational computational skills and provided support during the more quiet days in SBCB. I am also deeply thankful for the friendships of Will, Keith, Bjarne, Charlie, and Eddie, and for the late nights spent in Jericho, filled with laughter and music.

To the Lander lab, thank you to each of and every one of you for your friendship and guidance over the years. In particular, Alex, Anette, Chris, Jie, Kelly, Daniel, and Jeff welcomed me into the lab and taught me all things cryo-EM, Lon, and Pol theta. To Will L., thank you for your constant help and support on the microscopes and with troubleshooting the spray device. To Jan, thank you for all of the discussions that helped shape my technical understanding of EM and guided my experimental analyses. To Zeba, Prak, Cecilia Marie, and Maria thank you for lots of laughs and chats in the bays, especially in the final stages of my PhD. To Karen, thank you for steadfast willingness to help me with processing or just to offer advice and encouragement. To Noah and Mason, I am lucky to have worked with such incredible young scientists, and I drew much inspiration from your enthusiasm.

Throughout my PhD, I was fortunate to form a number of incredible friendships. Thank you to Alex, who always ensured that I felt welcomed in new spaces, celebrated at my biggest milestones, and supported during challenging times. To you and Nathan, I am grateful for the times you invited me to join in on your family's festivities, especially on holidays when I couldn't make it home. Thank you to Ringo, who is an absolute bundle of joy and the greatest pug I've ever known. To Salv, thank you for providing me with the music and playlists that defined the past five years of my life and for some of the most laughter-filled memories of this time. Lastly, thank you to Claire – you were the first true friend I made abroad, and I am

forever grateful for the memories we shared in Cowley and Leeds. Thank you for welcoming me into your communities, teaching me all things British, and for providing me with a sense of home away from home. I look forward to many future visits across the pond and long catch ups over glasses of wine.

A special thank you to the staff of Peleton, my favorite cafe in Oxford, who never failed to greet me with a smile and provided an oasis that centered me on the more difficult days of the PhD.

To my friends from home: Audrey, Camden, James, and Elliot. It has been a privilege to grow with you and I am thankful for our life-long friendships.

Thank you to my friends from college and forever roommates, Aisling and Fiona, who I've shared countless adventures with, at home and around the world. Thank you for always being there to support me when I'm away, and to greet me with open arms when I come home.

I owe immense gratitude to my parents for supporting me throughout my early life and beyond, and for providing me with all of the opportunities necessary to reach this stage. To my dad, thank your constant support of and interest in my scientific endeavors. To my mom, thank you for advocating for me throughout my early education and for always believing in my abilities.

Thank you to Eric for being my best friend for the past 24 years (the three years prior, I don't remember if this was the case). After leaving Connecticut, my "homebase" became wherever you were living at the time. I am grateful to have such a close friendship with you and to have been on so many adventures together.

Finally, to Will – thank you for being an everlasting source of light during this journey. Your groundedness and your gift for bringing out the best in others never ceases to amaze me. You have been the greatest support system during this journey and I am so thankful to have shared so many wonderful memories with you during this time.

Declaration of Generative AI Use

Generative artificial intelligence (AI) tools, including GPT4, GPT5, and Claude Sonnet 4 were used in the preparation of this thesis to improve writing clarity and style.

Abstract

Macromolecular machines undergo dynamic conformational changes across multiple timescales to execute their biological functions, yet conventional structural biology techniques typically capture only static snapshots. This thesis presents a multi-faceted approach to overcome these fundamental limitations by combining cutting-edge computational and experimental methodologies to investigate protein dynamics from microseconds to minutes. First, I present the most comprehensive molecular dynamics study of the human mitochondrial LONP1 protease to date, extending the timescales of large-complex molecular dynamics simulations and revealing transformative insights into the dynamic interplay between functional domains that governs enzymatic activity. Next, I expand experimental access to rapid conformational changes through the development of "Mix-it-up" (MIU), a cost-effective time-resolved cryo-EM platform that enables millisecond timescale studies of biochemical reactions. Finally, I employ time-resolved cryo-EM to elucidate the mechanistic details of early DNA binding by Polymerase θ, successfully capturing previously inaccessible early conformational ensembles that function in DNA repair. Together, these studies bridge the temporal gap between computational and experimental approaches, advancing our understanding of how macromolecular machines coordinate their functional cycles through exploration of diverse conformational landscapes. These mechanistic insights lay the groundwork for understanding how disruptions in protein dynamics drive disease and for guiding the design of targeted therapeutic interventions.

Contents

Acknowledgements **i**

Declaration of Generative AI Use **iii**

Abstract **v**

List of Figures **xi**

List of Tables **xiii**

1 Introduction **1**
1.1 Conformational landscapes in structural biology 1
1.1.1 Protein structure and hierarchy 1
1.1.2 The structure-function relationship 2
1.1.3 Timescales of protein motions 2
1.2 Biophysical techniques for assessing the structure-function relationship . . . 4
1.2.1 Overview of techniques 4
1.2.2 Light microscopy and super-resolution microscopy 4
1.2.3 Atomic Force Microscopy (AFM) 6
1.2.4 Small Angle X-ray Scattering 6
1.2.5 Nuclear magnetic resonance 7
1.2.6 X-ray crystallography 7
1.2.7 Molecular dynamics 9
1.2.8 Cryo-EM 9
1.3 Time-resolved cryo-EM methods for studying conformational landscapes of macromolecules 11
1.3.1 The challenge of time resolution in structural biology 11
1.3.2 Approaches for time-resolved cryo-EM sample preparation 11
1.3.3 Sample mixing strategies 14
1.3.4 Applications to biological systems 14
1.4 Objectives of this work 16
1.5 Significance of this work 18

2 Probing functional transitions of the human LONP1 protease through long-timescale simulations **19**

2.1 Chapter overview . . . 19
2.2 Introduction . . . 20
2.2.1 AAA+ proteins . . . 20
2.2.2 The mitochondrial protein quality control network . . . 20
2.2.3 Conserved mechanisms of AAA+ proteases . . . 23
2.2.4 The Lon protease . . . 26
2.3 Results . . . 29
2.3.1 Nucleotide flips within the LONP1 AAA+ domain reflect multiple stable binding poses . . . 29
2.3.2 Pore loop engagement with substrate is not strictly determined by nucleotide state . . . 33
2.3.3 Dynamics of substrate release . . . 37
2.3.4 Inactivation of LONP1 proteolytic sites occurs in the absence of substrate on the microsecond timescale . . . 37
2.3.5 Collective motions revealed by PCA provide a framework for probing nucleotide-dependent allosteric coupling . . . 41
2.4 Discussion . . . 43
2.5 Methods . . . 45
2.5.1 Long-timescale simulations of LONP1 . . . 45
2.5.1.1 Model . . . 45
2.5.1.2 Ligand parameterization . . . 47
2.5.1.3 Simulation setup . . . 47
2.5.1.4 Equilibration . . . 47
2.5.1.5 Production . . . 47
2.5.1.6 Analysis . . . 48
2.5.2 Substrate Release Simulations . . . 50
2.5.2.1 Model . . . 50
2.5.2.2 Simulation setup . . . 50
2.5.2.3 Equilibration . . . 50
2.5.2.4 Production . . . 51
2.5.2.5 Analysis . . . 51
2.5.3 Analysis Software . . . 51

3 Mix-it-up: Accessible time-resolved cryo-EM on the millisecond timescale 53
3.1 Chapter overview . . . 53
3.2 Introduction . . . 54
3.3 Results . . . 56
3.3.1 Single-sample blotting with MIU is effective for high-resolution structure determination . . . 56
3.3.2 High-resolution structures of mixed apoferritin and aldolase samples using MIU . . . 58
3.3.3 Application of MIU to study conformational changes . . . 61
3.3.4 pH-induced contraction of CCMV observed on the millisecond timescale 61
3.3.5 MIU captures ATP-dependent GroEL/ES complex formation . . . 65
3.4 Discussion . . . 67

3.5 Methods . . . 68
3.5.1 MIU Build . . . 68
3.5.2 Sample preparation . . . 69
3.5.2.1 Apoferritin . . . 69
3.5.2.2 Aldolase . . . 69
3.5.2.3 Cowpea Chlorotic Mottle Virus . . . 69
3.5.2.4 GroEL and GroES . . . 70
3.5.2.5 Gold fiducials . . . 70
3.5.3 Cryo-EM grid preparation . . . 70
3.5.3.1 Aldolase + Apoferritin. . . . 71
3.5.3.2 Contracted CCMV. . . . 71
3.5.3.3 Expanded CCMV. . . . 71
3.5.3.4 Expanded CCMV + low pH buffer / gold (700 ms). . . . 71
3.5.3.5 GroEL + GroES, no ATP. . . . 71
3.5.3.6 GroEL + GroES + ATP/gold (100 ms). . . . 71
3.5.3.7 GroEL + GroES + ATP/gold (300 ms). . . . 71
3.5.3.8 GroEL + GroES + ATP/gold (700 ms). . . . 72
3.5.4 Cryo-EM data collection . . . 72
3.5.5 Cryo-EM data analysis . . . 72
3.5.5.1 Aldolase + Apoferritin. . . . 72
3.5.5.2 Contracted CCMV. . . . 73
3.5.5.3 Expanded CCMV. . . . 74
3.5.5.4 Expanded CCMV + low pH buffer / gold (700 ms). . . . 74
3.5.5.5 GroEL + GroES, no ATP. . . . 75
3.5.5.6 GroEL + GroES + ATP/gold (100 ms) . . . 75
3.5.5.7 GroEL + GroES + ATP/gold (300 ms). . . . 76
3.5.5.8 GroEL + GroES + ATP/gold (700 ms) . . . 76

4 Time-resolved studies of polymerase theta-mediated DNA binding 79
4.1 Chapter overview . . . 79
4.2 Introduction . . . 80
4.2.1 DNA double-strand break repair . . . 80
4.2.1.1 Repair pathways . . . 80
4.2.1.2 Structural insights into DNA-bound polymerase theta . . . 82
4.3 Results . . . 84
4.3.1 Establishing a workflow for time-resolved studies of Polθ . . . 84
4.3.2 Early conformational ensembles of the Polθ helicase domain reflect tetrameric and DNA-bound dimeric states. . . . 86
4.4 Discussion . . . 88
4.5 Methods . . . 92
4.5.1 Sample preparation . . . 92
4.5.1.1 DNA-free PolθHel . . . 92
4.5.1.2 DNA oligomer design . . . 93
4.5.1.3 Formation of the PolθHel-DNA complex . . . 93
4.5.1.4 Negative Stain . . . 94

4.5.1.5 Cryo-EM grid preparation 94
4.5.2 Cryo-EM data collection 94
4.5.3 Cryo-EM data analysis 95
4.5.3.1 DNA-free PolθHel at 250 mM NaCl (control) 95
4.5.3.2 PolθHel at 100 mM NaCl ∼5 hr after DNA addition (control) 95
4.5.3.3 DNA-free PolθHel at 100 mM NaCl (control) 96
4.5.3.4 PolθHel at 100 mM NaCl 5 min after DNA addition 96
4.5.3.5 PolθHel at 100 mM NaCl 10 min after DNA addition 97

5 Concluding Remarks **99**

A Appendix: LONP1 **101**

B Appendix: MIU **109**

C Appendix: Polθ **123**

List of Figures

1.1 Timescale of protein motions. 3
1.2 Low-resolution Biophysical Techniques. 5
1.3 High-resolution Biophysical Techniques. 8
1.4 Molecular dynamics force field components. 10
1.5 Approaches for time-resolved cryo-EM sample preparation. 13
1.6 Strategies for sample mixing and application in time-resolved cryo-EM. . . . 15

2.1 AAA+ Proteases of the mitochondrial protein quality control network. . . . 21
2.2 The AAA+ module. 24
2.3 Model of substrate translocation. 25
2.4 Cryo-EM structures of inactive and active states of human LONP1. 27
2.5 Models of LONP1 generated for MD simulations. 28
2.6 Multiple poses of bound nucleotide. 31
2.7 Comparison of ATP poses from MD to experimental data from the PDB. . . 32
2.8 34
2.9 Change in ATPase intra-subdomain angle associated with nucleotide-induced conformational changes. 35
2.10 Pore loop engagement with substrate is not strictly nucleotide dependent. . . 36
2.11 Substrate release dynamics. 38
2.12 Inactivation of proteolytic sites upon substrate removal. 40
2.13 Per-chain principal component analysis reveals concerted rigid-body domain motions. 42

3.1 Mix-it-up design and method. 55
3.2 Modifications implemented in “Mix-it-up.” 57
3.3 Improved ice distribution of mixed samples using MIU. 59
3.4 EM grid architecture. 60
3.5 Aldolase and apoferritin particle distribution in mixed sample prepared with MIU. 62
3.6 pH-induced contraction of CCMV using MIU. 64
3.7 ATP-dependent complex formation of GroEL/ES observed following sample preparation with MIU. 66

4.1 DNA Repair Pathways. 81
4.2 Structural basis of Polθ-mediated microhomology recognition and annealing . 83

4.3 Negative stain PolθHel stability tests under varying salt conditions. 85
4.4 PolθHel particle distributions in cryo-EM. 87
4.5 Putative DNA-Bound Tetramer. 89
4.6 3D Classification of putative DNA-bound tetramer. 90

A.1 Complete analysis of time-dependent nucleotide binding pose clusters. 102
A.2 Complete analysis of pore loop to substrate engagement distances. 103
A.3 Complete analysis of substrate release distances. 104
A.4 Complete analysis of substrate release contacts. 104
A.5 Selection of residues for dihedral analysis to determine presence of the inactive state of the proteolytic site. 105
A.6 Complete analysis of proteolytic inactivation events. 106
A.7 Complete per-chain principal component analysis. 107

B.1 Processing pipeline of the mixed apoferritin and aldolase dataset prepared with MIU. 110
B.2 Control, contracted CCMV. 115
B.3 Processing pipeline of the control, contracted CCMV state dataset. 116
B.4 Processing pipeline of the control, expanded CCMV state dataset. 117
B.5 Processing pipeline of the pH-shift CCMV dataset prepared with MIU. . . . 118
B.6 Processing pipeline of the control dataset of GroEL in the presence of GroES and absence of ATP. 119
B.7 Processing pipeline of the 100 ms GroEL/ES + ATP dataset. 120
B.8 Processing pipeline of the 300 ms GroEL/ES + ATP dataset. 121
B.9 Processing pipeline of the 700 ms GroEL/ES + ATP dataset. 122

C.1 Full set of 2D class averages from negative stain PolθHel stability tests. . . . 123
C.2 Cryo-EM data processing pipeline of DNA-free PolθHel at 250 mM NaCl (control). 124
C.3 Cryo-EM data processing pipeline of PolθHel at 100 mM NaCl $\sim$5 hr after DNA addition (control). 125
C.4 Cryo-EM data processing pipeline of DNA-free PolθHel at 100 mM NaCl (control). 126
C.5 Cryo-EM data processing pipeline of PolθHel at 100 mM NaCl 5 min after DNA addition. 127
C.6 Cryo-EM data processing pipeline of PolθHel at 100 mM NaCl 10 min after DNA addition. 129

List of Tables

2.1 Summary of molecular dynamics simulations. 46

B.1 Cryo-EM parameters for apoferritin, aldolase, and CCMV samples 111
B.2 Cryo-EM parameters for GroEL samples (no ATP, 100ms) 112
B.3 Cryo-EM parameters for GroEL samples (300ms) 113
B.4 Cryo-EM parameters for GroEL samples (700ms) 114

C.1 Cryo-EM parameters for PolQθH 128

1 Introduction

1.1 Conformational landscapes in structural biology

1.1.1 Protein structure and hierarchy

Proteins are often depicted as static structures, though within their native environments they are highly dynamic and explore vast conformational landscapes. These landscapes reflect an ensemble of structural states dictated by the protein's amino acid sequence and the physical forces acting upon the system. While only twenty amino acids comprise the building blocks of proteins, their primary structure, defined by the sequence in which they are linked, dictates their spatial arrangement and determines the range of structural states a polypeptide can occupy. Ultimately, protein folds are driven by the balance of contributions from enthalpic interactions, such as van der Waals forces, hydrogen bonds, electrostatics, and disulfide bridges, and entropic forces, including hydrophobic packing and solvation. These forces sum and funnel the protein towards low free-energy states, though several local minima may be accessible and shape the observed "flexibility" of the protein [1].

The elements that define the architecture of a protein are organized hierarchically. Protein secondary structure consists of three main recurring components: α-helices, β-sheets, and loops. The relative position of these elements gives rise to higher-order motifs, such as helical bundles and β-barrels, which define the tertiary structure of a protein. During folding, hydrophobic residues will predominantly migrate towards a buried "core" of the protein, while charged residues remain exposed to the surrounding solvent. The resulting overall topology of a protein can influence its cellular localization and, ultimately, its function. For example, proteins that exist in cellular and organellar membranes typically contain transmembrane helices rich in hydrophobic residues that anchor them in lipid bilayers, whereas fibrous proteins such as collagen and keratin form elongated helical bundles that provide mechanical

strength and rigidity to connective tissue, muscle fibers, and hair.

1.1.2 The structure-function relationship

Beyond the impact of a protein's global fold on localization and scaffolding, local primary structural elements can encode specific biochemical activities. Conserved sequence motifs can serve as binding pockets, metal coordination sites, or catalytic residues that carry out enzymatic processes. A well-studied example is the Walker motifs found in many ATP-binding proteins [2]. The Walker A motif forms a glycine-rich loop and is responsible for nucleotide binding, cradling the phosphate groups through backbone amides and a conserved lysine. The Walker B motif, typically located downstream, contains conserved hydrophobic residues, followed by acidic side chains that coordinate a magnesium ion and facilitate hydrolysis. These elements can also influence larger conformational rearrangements to confer function. In the family of ATPases associated with diverse cellular activities (AAA+ family), the binding and hydrolysis of nucleotide at Walker motifs within the ATPase domain is thought to drive coordinated conformational changes that propagate through the complex, powering substrate remodeling to perform activities involved in membrane fusion, cytoskeleton remodeling, and mitochondrial protein quality control by AAA+ proteases such as the Lon protease [3, 4].

More broadly, few proteins function in a manner that is dictated by a single structural state, but rather by transitioning between multiple states within a defined conformational landscape to execute "tasks" within the cell. Thus, proteins are often likened to macromolecular "machines". The central goal of structural biology is to determine how static structures of these machines obtained from experimental or computational approaches are choreographed, much like assembling individual frames into a movie, in order to create a dynamic narrative that connects protein structure to function.

1.1.3 Timescales of protein motions

Conformational changes in proteins occur on a range of timescales as rapid responses to environmental stimuli, such as ligand binding, temperature or pH shifts, and interactions with other biomolecules (Fig 1.1) [5]. The fastest motions are atomic vibrations, occurring on the femtosecond timescale. Their oscillation frequency is determined by the mass of the atom, the strength of its chemical bonds, interactions with surrounding atoms, and the temperature of the system [6]. Bond torsions and angular motions typically occur on the picosecond timescale, often in response to the molecular environment. Steric repulsion and interactions such as hydrogen or ionic bonding can alter the strength and range of these motions. These ultrafast dynamics, occurring up to the picosecond timescale, are often inaccessible to direct observation by experimental methods, but can be explored computationally through molecular dynamics simulations [6].

In the picosecond to nanosecond range, side chain rotations are observed and can directly impact protein function by positioning residues for catalysis or coordinating ion transport through channels. Such rotations also facilitate preferred interactions depending on the charge or aromaticity of the residue. These rapid events precede slower processes such as the formation or rearrangement of secondary structure, which typically occurs on the nanosecond

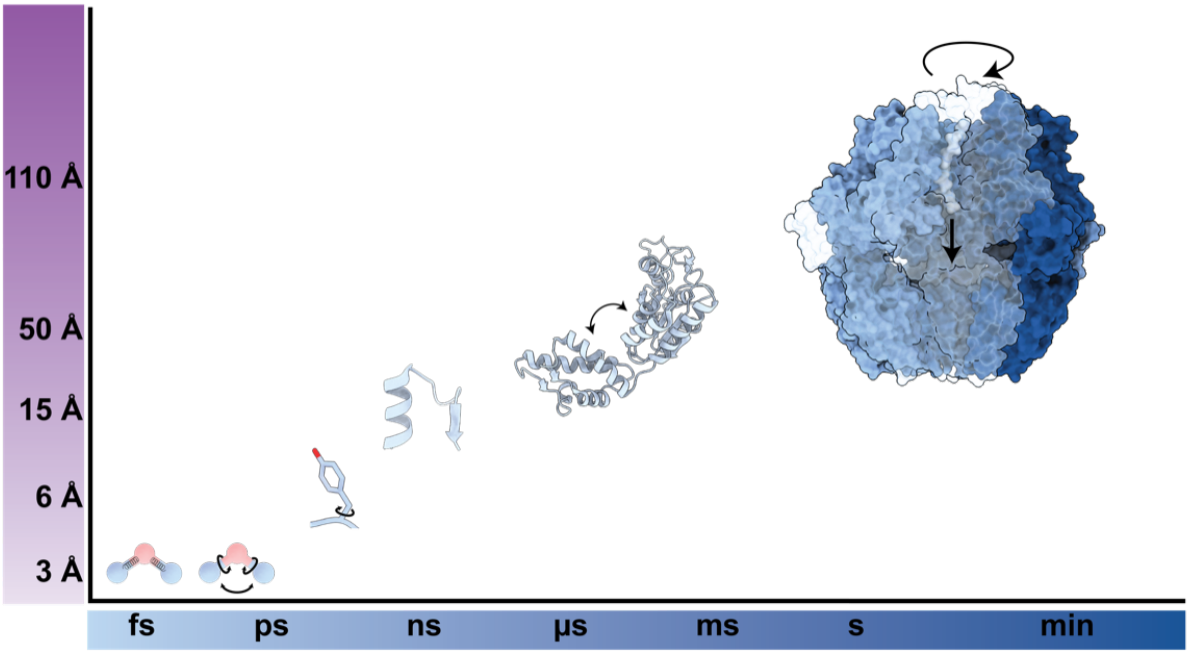


Figure 1.1: Timescale of protein motions. Protein dynamics occur at a range of spatial and temporal scales. Fast, small-scale motions include bond vibrations, angle bending (fs-ps). Torsional motion enables side chains to sample distinct rotamers on the picosecond-to-nanosecond timescale. Changes in secondary structure including loop motion and the formation of α-helices and β-sheets are longer timescale events, with dynamics observable ranging from nanoseconds to microseconds. Larger-scale motions, such as domain movements, range from microseconds to milliseconds, while global conformational changes, substrate remodeling, and the assembly of large-scale complexes can evolve over milliseconds to minutes.

to microsecond timescale [5]. Flexibility in secondary structural elements, induced by stimuli such as ligand binding or changes in salt concentration, can expose previously hidden binding pockets of the protein to solvent, enabling new functional interactions.

Beyond local conformational changes, proteins may undergo large-scale conformational transitions, enabling them to perform diverse functions. Domain movements often span microseconds to milliseconds, while significant structural reorganizations can occur on the order of milliseconds to minutes [5].

The sum of the dynamics spanning femtoseconds to minutes forms the foundation of a protein's functional landscape. While fast-timescale motions often appear random and uncoordinated, the pathways linking them to large-scale rearrangements are crucial for understanding biological mechanisms. For many macromolecular machines, this connection arises when energy derived from chemical catalysis, mediated by conformational changes at the amino acid level, is transformed into the mechanical force required to move entire protein domains, driving processes such as protein translocation, DNA unwinding, and intracellular transport [3, 7]. Understanding the conformational landscapes that govern protein function across these diverse timescales requires experimental approaches capable of capturing motions over the full spectrum of relevant temporal and spatial scales.

1.2 Biophysical techniques for assessing the structure-function relationship

1.2.1 Overview of techniques

Elucidating protein structure-function relationships requires a diverse toolkit of biophysical techniques. These methods can be categorized based on their resolution capabilities and the timescales they can access to assess the conformational landscapes of proteins. Low-resolution techniques (light microscopy, AFM, and SAXS) excel at capturing large-scale dynamics and conformational heterogeneity across ensembles, while higher-resolution methods (NMR, X-ray crystallography, and cryo-electron microscopy) provide atomic-level details of the structural states that underpin function (Fig 1.2,1.3). Computational approaches, such as molecular dynamics, complement experimental methods by accessing timescales and conformational states that are challenging to probe experimentally due to intrinsic limitations of measurement or sample preparation, while also driving the formation of hypotheses that guide experimental design (Fig 1.3). The choice of technique depends primarily on the biological question, the nature of the sample, and the timescales of interest. Ultimately, the integration of these methods can provide a comprehensive overview of macromolecular dynamics.

1.2.2 Light microscopy and super-resolution microscopy

Light microscopy is among the earliest tools used for direct biological observation, dating back to the 17^{th} century during which Antonie van Leeuwenhoek used a homemade microscope to observe bacteria for the first time. This method provides real-time visualization of biological processes, but has historically been limited to cellular and organelle-scale imaging due to the

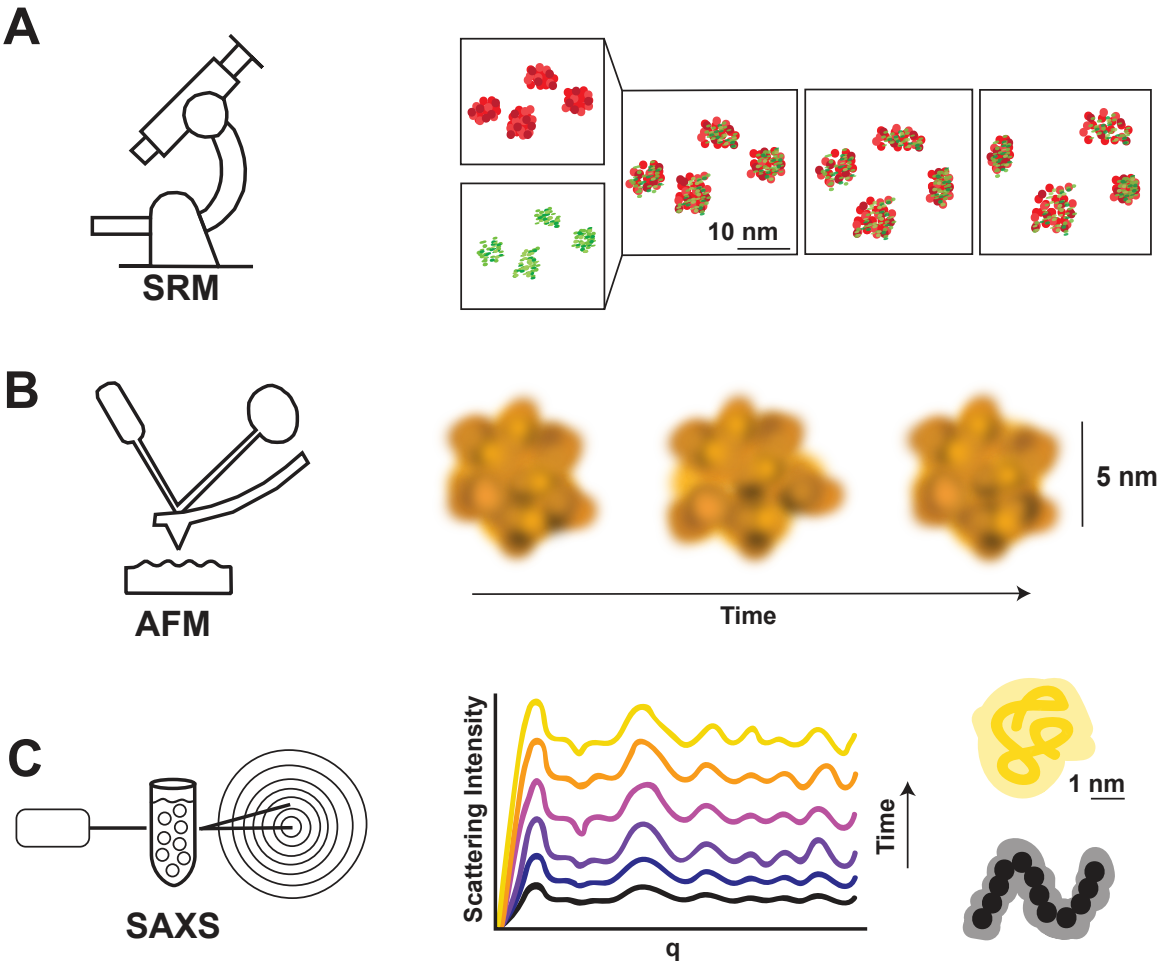


Figure 1.2: Low-resolution biophysical techniques. A) Super-resolution microscopy (SRM) enables visualization of molecular complexes beyond the diffraction limit of light, allowing spatial resolution on the nanometer scale and temporal resolution on the millisecond-second timescale. This method requires the use of fluorescent labels and is well-suited for studying cellular processes such as protein localization and organelle dynamics. **B)** Atomic force microscopy (AFM) provides topographical information of biomolecular structures and can capture low-resolution conformational changes in proteins over time. **C)** Small-angle X-ray scattering (SAXS) is a solution-based method that measures the scattering intensity of X-rays by the protein sample. These data can be used to generate a low-resolution, *ab initio* model of the protein ("SAXS envelope"). Data collected at multiple timepoints and can capture time-dependent changes in global conformation, folding pathways, and nanometer resolution structural changes.

diffraction limit of visible light (∼250 nm) [8]. While fluorophore labeling enables molecular specificity on the nanometer scale, it can occlude protein-protein binding sites or induce quenching of nearby fluorophores, thereby reducing image contrast and limiting effective resolution [9]. The advent of super-resolution microscopy (SRM) techniques (Fig 1.2A), including deterministic approaches such as stimulated emission depletion (STED) and structured illumination microscopy (SIM), as well as stochastic approaches such as photo-activated localization microscopy (PALM) and stochastic optical reconstruction microscopy (STORM), have pushed the resolution limit to ∼10–50 nm by manipulating fluorophore states, enabling visualization of subcellular architecture [10]. For time-resolved studies, conventional fluorescence microscopy achieves millisecond-to-second temporal resolution for large-scale processes like cell division, while super-resolution methods sacrifice temporal resolution (seconds-to-minutes) for improved spatial detail [11, 12]. Together, these techniques excel at studying protein localization and organelle dynamics in living systems, though detailed structural information remains beyond the spatial resolution of these methods.

1.2.3 Atomic Force Microscopy (AFM)

Atomic force microscopy (AFM) provides low-resolution information about the surface topography of biomolecules using a mechanical probe (Fig 1.2B). The probe is equipped with a sharp cantilever tip that scans sample surfaces. The cantilever deflection is monitored via laser reflection, enabling measurement of surface interactions and rigidity. The resulting data can be used to generate 3D topographical maps and inform on the mechanical properties involved in protein conformational changes and protein-protein interactions, providing insights into the energetic landscapes that govern macromolecular function. AFM offers advantages over techniques such as electron microscopy and X-ray crystallography, as it enables imaging of specimens in liquid [13]. Time-resolved AFM achieves millisecond temporal resolution, limited by feedback-loop delays in adjusting the cantilever height to maintain constant force. [14] Despite this constraint, time-resolved AFM has been applied to the study of dynamic biological processes, including actin-myosin interactions, DNA translocation, and amyloid assembly [15–18].

1.2.4 Small Angle X-ray Scattering

Small angle X-ray scattering (SAXS) is a low-resolution biophysical technique that can be used to assess the size and shape of macromolecules in solution (Fig 1.2C). The method measures the intensity of X-rays scattered by macromolecules in solution at small angles (0.1–5°) to yield structural information that can be used to generate a low resolution *ab initio* structure, commonly referred to as a "SAXS envelope" [19]. From these data, parameters such as radius of gyration, oligomeric state, and conformational heterogeneity can be inferred. Compared to crystallography, SAXS requires minimal sample preparation and enables rapid data collection [19]. Time-resolved SAXS achieves microsecond-to-millisecond temporal resolution using triggers such as laser pulses for temperature- and pH-jump experiments or the photorelease of caged substrates to initiate reactions [20, 21]. As a solution-state method, SAXS captures protein dynamics without crystallization constraints and can track rapid, non-equilibrium conformational changes in real-time. These capabilities have been applied to

studies of temperature-dependent insulin oligomerization, pH-dependent structural dynamics of poly-L-glutamic acid, and mechanistic insights into the Ca^{2+} ATPase transporter upon binding to ATP [21].

1.2.5 Nuclear magnetic resonance

Nuclear magnetic resonance was first demonstrated in the mid-20^{th} century and was soon after applied to the study of protein structure and dynamics (Fig 1.3A) [22, 23]. NMR spectroscopy provides atomic-resolution structural and dynamic information of proteins $< 50\,\text{kDa}$ in solution by detecting chemical shifts of NMR-active nuclei (^{1}H, ^{13}C, ^{15}N) in magnetic fields. NMR excels at studying conformational flexibility under physiological conditions and can characterize samples that remain challenging for electron microscopy and X-ray crystallography, such as intrinsically disordered proteins [24]. Despite its strengths, NMR is limited in its applications to larger systems, as increasing molecular size reduces tumbling rates and broadens resonance signals, complicating interpretation of spectra. Solid-state NMR, which does not rely on molecular tumbling, extends these applications to larger proteins, fibrils and membrane complexes [25]. Time-resolved NMR studies can probe dynamics on a wide range of temporal scales. Relaxation dispersion experiments capture microsecond-to-millisecond motions, revealing transient intermediate states and allosteric regulation, while real-time NMR monitors slower conformational changes (seconds-to-minutes) with atomic precision [26, 27]. Combined with its ability to measure ligand-binding kinetics and folding pathways, NMR offers an effective approach for dissecting the dynamic processes that confer macromolecular function.

1.2.6 X-ray crystallography

X-ray crystallography emerged as a promising tool for determining protein structure in the early 20^{th} century (Fig 1.3B). It has since become the gold standard for high-resolution protein structure determination, employing X-ray diffraction from crystalline protein arrays to generate electron density maps that reveal atomic positions [28]. However, the technique still faces several inherent challenges. The crystallization process can prove challenging for many proteins, and crystal packing forces can introduce conformational biases that can confound the interpretation of the resulting structure or give the impression of a less dynamic structure than that which exists in solution [29]. Despite these limitations, the method excels at revealing ligand binding pockets and catalytic sites with atomic precision. For time-resolved studies, X-ray crystallography accesses femtosecond-to-second timescales through X-ray free electron lasers (XFELs) and serial femtosecond crystallography (SFX), enabling high-throughput and room-temperature experiments, and delivering X-ray doses thousands of times higher than synchrotron sources [30]. Applications include studies of ligand binding as molecules diffuse into crystals or as photocaged substrates are released to initiate reactions, making the method particularly valuable for drug discovery and mechanistic studies [30].

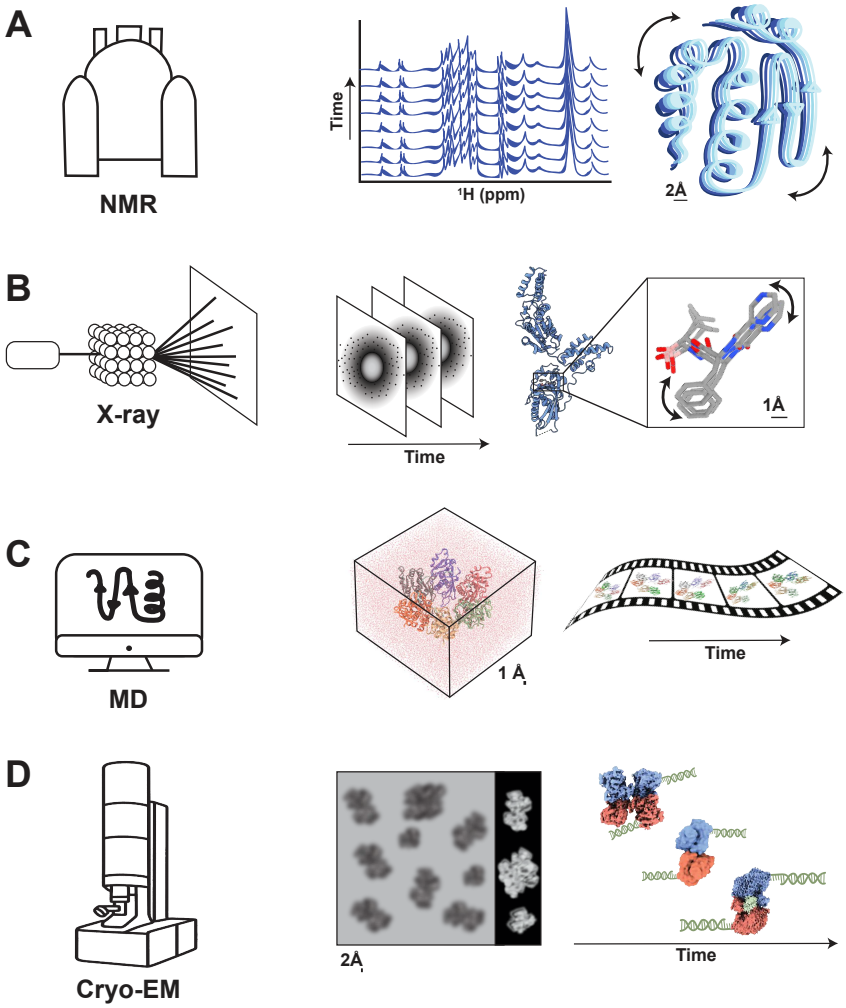


Figure 1.3: High-resolution biophysical techniques. A) Nuclear magnetic resonance (NMR) spectroscopy provides atomic-level resolution of proteins in solution by detecting chemical shifts of NMR-active nuclei. Time-resolved studies can be performed on the microsecond to minute timescale, and can be used to study folding pathways, allosteric regulation, and binding kinetics of proteins < 50 kDa. **B)** X-ray crystallography captures high-resolution structural information of proteins based on diffraction patterns obtained from crystalline samples. With recent technological advances, this method has become particularly powerful for high-throughput and time-resolved studies of ligand-bound targets at timescales as low as femtoseconds, greatly contributing to drug discovery efforts. **C)** Molecular dynamics (MD) simulations compute trajectories of molecular systems by numerically integrating Newton's equations of motion using forcefields that describe interatomic interactions. With femtosecond temporal resolutions, MD provides a "movie" of conformational dynamics from bond vibrations to large-scale structural rearrangements, offering atomic insight that often complements experimental time-resolved methods. **D)** Cryo-electron microscopy (cryo-EM) yields high-resolution structural information of proteins in a near-native environment through rapid vitrification of dynamic samples in solution. Vitrified samples are imaged with an electron beam, and the resulting particle images are averaged, aligned, and classified to generate 3D density maps. Time-resolved methods can be used to study protein interactions and conformational transitions at near-atomic resolution on microsecond-minute timescales.

1.2.7 Molecular dynamics

Molecular dynamics (MD) simulation is a physics-based computational technique that models the time-dependent evolution of a system of particles by numerically integrating Newton's equations of motion (Fig 1.3C, Fig 1.4) [6]. The interactions between atoms are described by a mathematical model known as a force field, which comprises a potential energy function and a set of parameters for each interaction type. These parameters are derived from experimental data or quantum mechanical calculations. The interactions modeled in a typical force field include both bonded (bond stretching, angle bending, dihedrals) and non-bonded (van der Waals and Coulumbic) interactions. An analytical equation for the force is determined by taking the derivative of the potential energy equation with respect to the distance (or angle) between atoms. After initializing the system with the initial coordinates and velocities, molecular dynamics calculations are performed iteratively: at each timestep, a force calculation is performed, followed by integration to obtain the updated coordinates and velocities. During force calculation, the force equation is evaluated for the set of atoms involved in each interaction. The resultant force vector on each atom is determined by summing all of the forces arising from its interactions. The force vector is then divided by the mass of the atom to obtain an acceleration vector, which is integrated first to determine the updated velocities, and then again to determine the new coordinates.

In practice, starting from experimentally determined or predicted structures, systems are solvated and neutralized *in silico*, after which an energy minimization step is performed to "relax" the system to an equilibrium state and prevent excessive forces due to clashing atomic positions or strained bond angles. The systems are then equilibrated under constant temperature and pressure, with the initial atomic velocities assigned randomly according to the Boltzmann distribution [31]. Finally, a production run is initiated, generating a "movie" of the dynamic evolution of the system over time. In contrast to most experimental methods, MD provides femtosecond temporal resolution with 2–4 fs timesteps, enabling visualization of dynamics ranging from bond vibrations to large-scale conformational changes. Though its computational demand typically limits the high end of temporal scales from nanoseconds to milliseconds for atomistic systems, longer timescales can be accessed through coarse-graining or enhanced sampling methods, such as metadynamics, replica exchange, and umbrella sampling, which bias the the free energy landscape of the system to promote sampling of high-energy states [32–34]. MD has been employed to model dynamics of a wide range of systems, ranging in size from a few amino acids to bacterial membranes and whole cells [35–37]. MD complements experimental time-resolved methods by providing atomic-level detail of dynamics at timescales often inaccessible experimentally, though validation with experimental data is required to assess the physiological relevance of the findings.

1.2.8 Cryo-EM

Single particle cryo-EM (SPA cryo-EM) emerged in the late 20$^{\text{th}}$ century, but was considered a low-resolution technique until the 2010s when the advent of direct electron detectors, along with improved image processing algorithms and advances in computational power, enabled the technique to reach sub-3 Å resolution (Fig 1.3D) [38]. In cryo-EM, a thin film of protein solution is plunge-frozen into liquid ethane, vitrifying the sample in a solution-like state

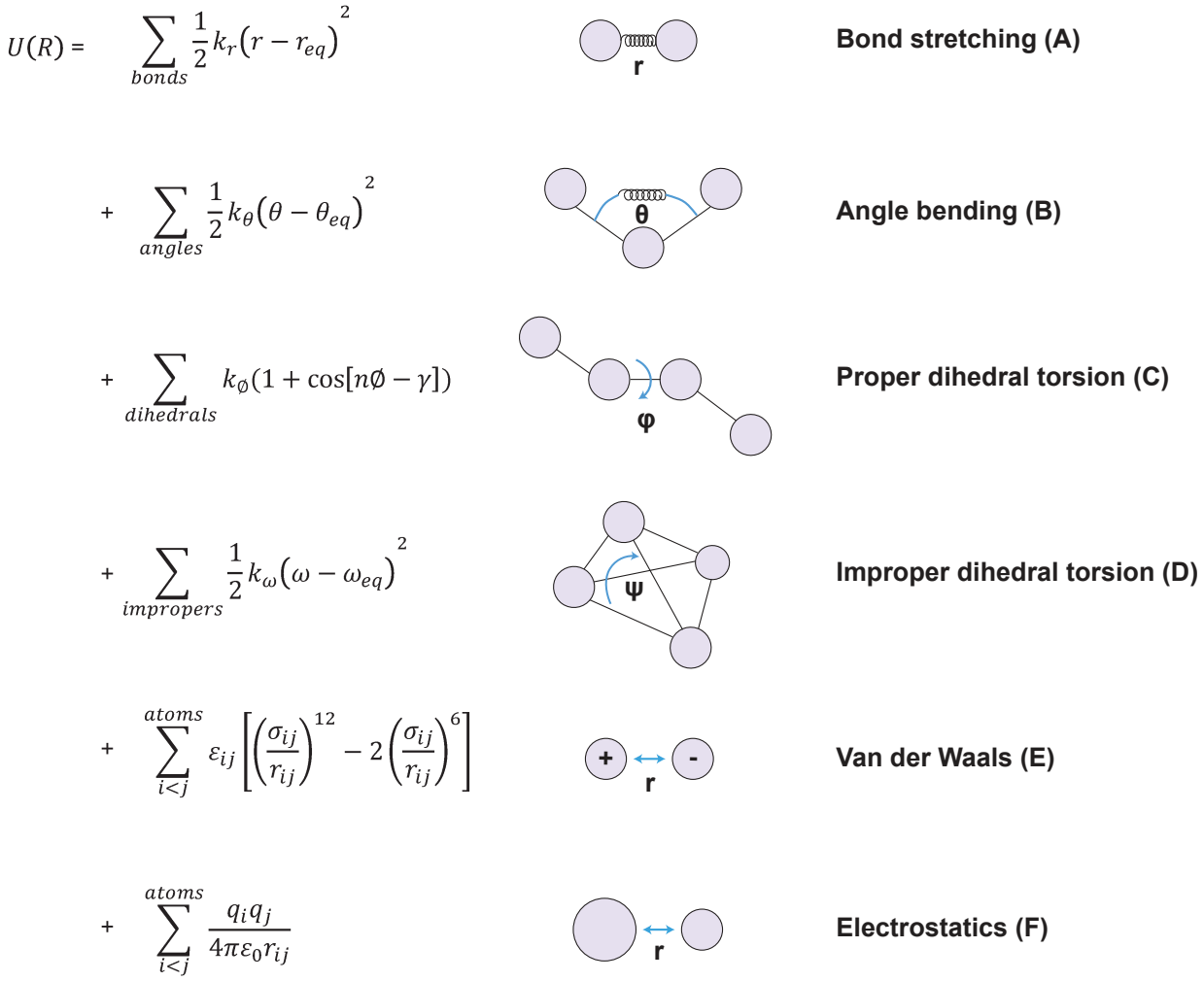


Figure 1.4: Potentials and interaction types described by force fields in molecular dynamics simulations. The potential energy equation comprises the potentials (left) for the bonded **(A-D)** and non-bonded **(E-F)** interactions (schematics shown, middle). Bond stretching **(A)**, angle bending, **(B)**, and improper dihedral torsions **(D)** are typically modeled as harmonic potentials. Proper dihedral torsions **(C)** are modeled using a periodic Fourier series to describe the multiple energy minima and maxima. Van der Waals interactions **(E)** are modeled by the Lennard-Jones potential, and electrostatic interactions **(F)** are modeled by the Coulumbic potential.

and preserving near-native conformations. A focused electron beam is then passed through the vitrified sample, and scattered electrons are recorded by direct electron detectors to generate micrographs. Protein particles are subsequently picked and aligned to generate 2D class averages, corresponding to different orientations of the protein. These are then computationally reconstructed into 3D density maps into which atomic models of the protein structure can be built. Cryo-EM excels at studying large complexes ($> 40\,\text{kDa}$) and membrane proteins with relatively straightforward sample preparation and near-atomic resolution, though smaller proteins remain challenging due to low signal-to-noise ratios [38]. In comparison to other high-resolution techniques, cryo-EM is better suited for accessing conformational flexibility than X-ray crystallography, and can be used to study larger systems than accessible by NMR. Applications to time-resolved studies are rapidly advancing as improvements in hardware, software, and sample preparation continue to push the spatial and temporal resolution limits of cryo-EM to the frontier of structural biology [30, 39]. These methods are further detailed in the following section.

1.3 Time-resolved cryo-EM methods for studying conformational landscapes of macromolecules

1.3.1 The challenge of time resolution in structural biology

While the biophysical techniques described above provide vast insights into protein structure and dynamics, capturing the transient intermediate states that underlie macromolecular function often requires specialized approaches. Despite advances in both experimental and computational methods, a persistent gap remains in our ability to observe these non-equilibrium states before they relax to equilibrium. Low-resolution experimental approaches yield valuable information on protein dynamics, while high-resolution biophysical and computational methods can provide detailed structural snapshots of key intermediates. However, most biophysical methods are limited to temporal resolutions of milliseconds to minutes, missing high-energy or short-lived states. Computational approaches complement these experiments by generating atomic-scale trajectories down to femtosecond timescales, though accessing slower dynamics on the order of microseconds and beyond is often infeasible for large systems due to the computational cost of force calculations. The development of time-resolved cryo-EM has been particularly transformative in bridging this gap, enabling visualization of near-native structural dynamics with microsecond to second resolution.

1.3.2 Approaches for time-resolved cryo-EM sample preparation

Preparation of mixed samples is a key component of time-resolved cryo-EM experiments (Fig 1.5). As the goal of these studies is to enable the investigation of the time-dependent stages of multi-component reactions, the time required for mixing and vitrification must be considered to determine the overall timescale of the experiment.

To prepare cryo-EM samples manually using blot-based approaches, a small volume of sample (3–$4\,\mu\text{L}$) is first applied manually, using a pipette, to a grid connected to tweezers. The

sample is then blotted for a brief period of time (1–5 s) with filter paper to produce a film that is sufficiently thin to form vitreous ice upon plunging into liquid ethane (Fig 1.5A) [38]. Conventionally, cellulose filter paper is used for blotting, but a recent study by Tan and Rubinstein demonstrated the use of glass fiber filter paper to enable faster sample wicking due to the increased surface area of the glass fibers as opposed to the cellulose fibers, which enables a greater absorbance capacity and more rapid sample thinning [40]. For blot-based approaches, sample plunging can be performed using a custom-built plunging device, in which the tweezers are attached to a gravity-based mechanism, or to an electromagnetic-based mechanism, such as a solenoid, that is triggered using applied voltage to enable faster plunging [41]. Plunging can also be automated using a commercial device such as the Vitrobot or Leica [42]. Together, these methods are relatively time-intensive compared to other approaches for time-resolved sample preparation, but do not require specialized equipment, facilitating their accessibility.

For faster time resolution, microfluidics enables precise control of sample mixing at relatively small volumes, and has been used in time-resolved experiments since the late 2000s (Fig 1.5E) [43]. Microfluidic chips consist of a network of channels to which multiple samples can be independently added before intersecting or merging to initiate the mixing process. Mixing efficiency depends on several factors, including the geometry of the channel intersection, the length and diameter of the mixing channel, the channel material, and the flow rate [44]. This method can also be limited by the requirement for relatively large dead volumes of sample [39]. Further, microfluidic chips are often made of PDMS or silica, making them expensive to customize and manufacture [44, 45].

An alternative approach for on-the-fly mixing is to use a laser pulse to release a photocaged substrate within a premixed sample (Fig 1.5A, 1.6C) [46–48]. This method aims to standardize the timing of the reaction initiation across the entirety of the sample by ensuring complete mixing of the components prior to triggering the reaction. To prevent crystalline ice formation, the laser wavelength must be carefully adjusted to release the caged substrate without heating the grid bars, as the gold or copper mesh can absorb laser energy and increase sample temperature [39]. A key limitation of this approach is its reliance on photocageable substrates, which are often challenging to synthesize. Timescales on the order of microseconds have been demonstrated when integrating this approach with specialized vitrification methods, though most studies have been limited to the millisecond timescale using more accessible, traditional plunge freezing methods [46–48].

Spray-based vitrification methods can be broadly classified into gas-assisted atomizers and voltage-assisted atomizers (Fig 1.5B). Gas-assisted atomizers use pressurized gas, such as nitrogen (N_2), to trigger aerosolization of a liquid droplet for application of sample onto a grid (Fig 1.5B). The aerosolized droplet sizes are often larger and more variable with this method compared to voltage-assisted atomizers, ranging from ∼30 to 100 μm in diameter [44]. In contrast, voltage-assisted atomizers apply a high voltage to a liquid sample to generate a fine aerosol. The resulting droplet sizes are typically more uniform and can be $< 1\mu$m in size (Fig 1.5B) [49]. Ultrasonic dispensers are also driven by applied voltage, but here the voltage is applied to a piezoelectric transducer (piezo), rather than the liquid itself [40, 50]. The piezo then produces mechanical (acoustic) vibrations that aerosolize a droplet placed on

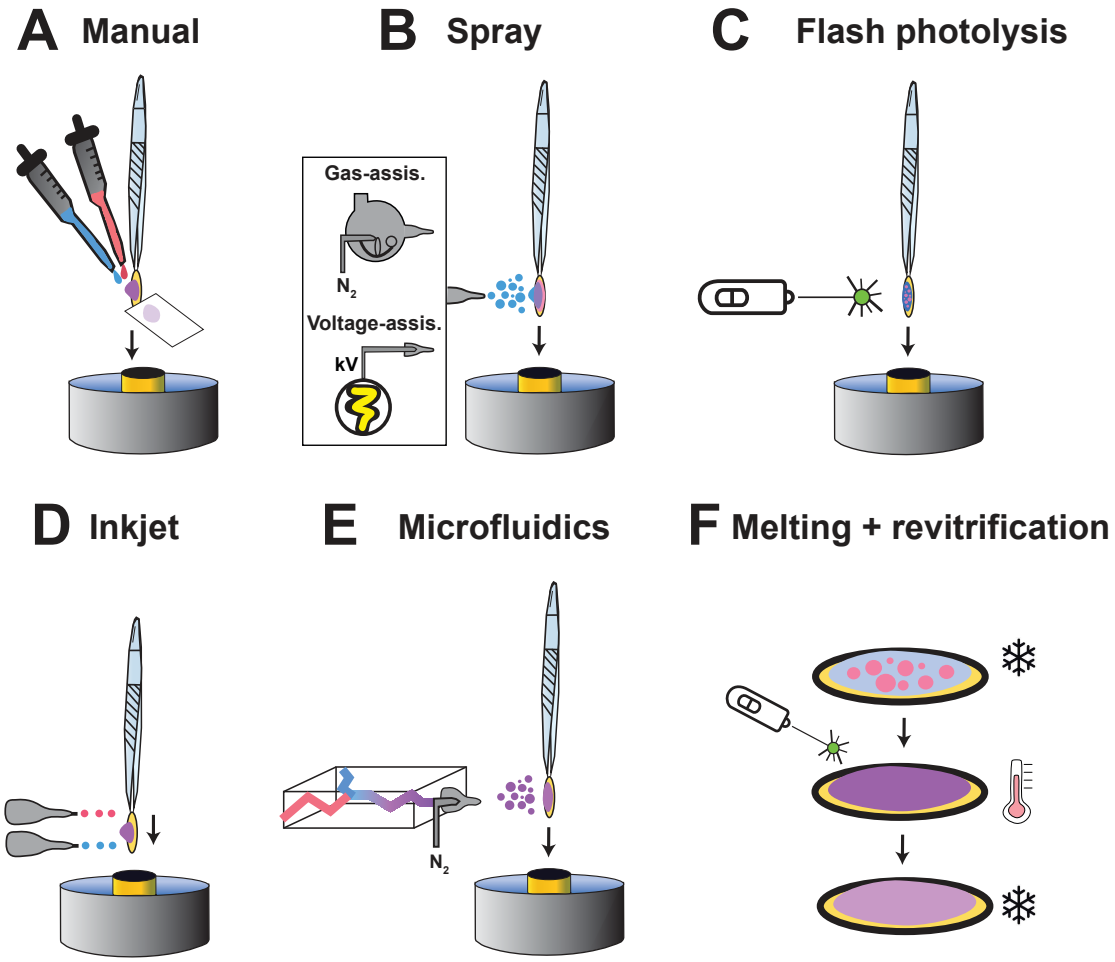


Figure 1.5: Approaches for time-resolved cryo-EM sample preparation. A) Manual application of two or more samples onto the grid is performed using pipetting techniques, and preparation time ranges from seconds to minutes. **B)** Spray-based strategies aerosolize sample to deliver fine droplets directly onto the grid on the millisecond timescale. The sample can either be delivered by a gas-assisted (inset, top) or voltage-assisted (inset, bottom) atomizer. For multi-sample systems, a primary sample is first manually applied to the grid, after which a secondary sample is sprayed to initiate on-grid mixing. **C)** Flash photolysis involves the use of caged substrates, which are mixed with a primary sample prior to application onto the grid. A laser pulse is then used to release the caged substrate, triggering the reaction. **D)** Inkjet-based methods enable precise control of droplet size, rapidly depositing picoliter volumes of sample onto the grid. Dual inkjets can be used to deposit multiple samples as the grid is plunging, resulting in on-grid mixing to enable the study of biochemical reactions. **E)** Microfluidics devices enable rapid mixing of reactants prior to aerosolization to deposit the sample onto the grid within milliseconds. **F)** Flash melting and revitrification involves vitrification of mixed samples, followed by rapid melting and the use of a secondary trigger to initiate the reaction, after which the sample is revitrified for observation by cryo-EM. This method achieves the fastest timescales of time-resolved sample preparation, capturing events that take place on the microsecond timescale.

its surface.

A more recent approach for sample application employs inkjet systems, which use similar hardware to conventional printers to enable application of nanoliter-sized droplets onto a grid (Fig 1.5D) [51, 52]. This method requires specialized grids (self-wicking grids), which contain copper nanowires along the grid bars to wick bulk liquid away and produce a thin layer of sample for vitrification. The removal of a blot-based approach for sample thinning enables higher-speed sample preparation with this method [39].

Rapid melting and revitrification is a recently developed method that enables the fastest timescales of investigating sample conformational variability accessible by cryo-EM (microseconds) (Fig 1.5F) [53]. This approach requires specialized hardware integrated with an electron microscope and involves rapidly melting a pre-vitrified sample using a high-power laser, increasing the kinetic energy of the system and allowing the sample to adopt more dynamic conformations. The sample is often exposed to a secondary trigger upon melting, such as a laser pulse to release a photo-caged substrate, in order to initiate a biochemical reaction or trigger alternate conformational states of the protein of interest [48]. The system is then rapidly cooled to revitrify the sample, trapping high-energy conformations that are inaccessible on the timescales required for other cryo-EM sample preparation methods. However, the melting step can cause sample evaporation, resulting in regions of crystalline ice. An approach involving sandwiching the sample between silicon dioxide layers was recently developed to mitigate these effects [54]. Despite its potential, this method has not yet been widely adopted due to the specialized equipment required.

1.3.3 Sample mixing strategies

The aforementioned approaches can be broadly grouped into two categories based on the order of mixing and application: 1. Mix-then-apply, and 2. On-grid mixing (Fig 1.6). The order of these events can dictate the time-resolution of the experiment as well as the mixing efficiency of the samples. In mix-then-apply approaches, samples are mixed prior to application (Fig 1.6A). This category encompasses manual mixing with blot-based application, as well as microfluidics combined with spray- or inkjet-based vitrification approaches.

Meanwhile, on-grid mixing can allow for shorter timescale studies of biochemical reactions due to the removal of a premixing step (Fig 1.6B,C) [55–57]. This method can be carried out in two ways. In the first, a primary sample is applied to the grid, followed by application of a secondary sample using a spray- or ink-jet-based method, triggering initiation of the reaction (Fig 1.6B) [58]. The second type of on-grid mixing involves addition of premixed samples onto a grid with a caged substrate that can be activated using a secondary trigger, such as light (Fig 1.6C) [46–48]. When photolysis is combined with a melting and revitrification strategy, microsecond timescales can be achieved [59].

1.3.4 Applications to biological systems

While time-resolved cryo-EM (tr-EM) has been under development for several decades, the applications of these methods to answering biological questions remain limited primarily

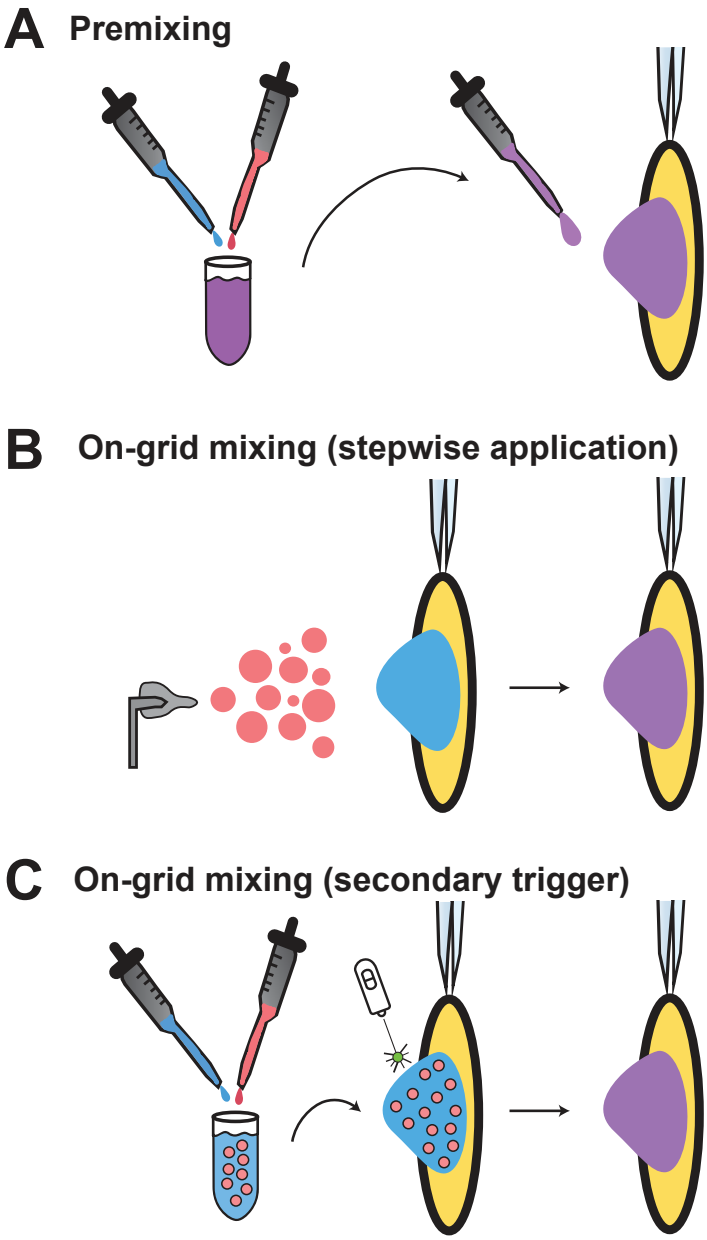


Figure 1.6: Strategies for sample mixing and application in time-resolved cryo-EM. A) Premixing. Reactants are combined in solution to initiate reactions prior to application onto the grid. This method enables time-resolutions on the millisecond timescale with use of microfludic devices, or on a range of seconds to minutes using manual approaches. **B) On-grid mixing with stepwise application.** Reactants are sequentially applied to the grid. A primary sample is first applied to the grid's surface, and a secondary sample is subsequently sprayed to initiate on-grid mixing with milliseconds. **C) On-grid mixing with a secondary trigger.** Reactants (typically, a protein and photocaged substrate) are applied together onto the grid, followed by a secondary trigger, such as a laser pulse, used to initiate the reaction immediately prior to freezing. This method typically enables time resolutions on the millisecond timescale, however, when combined with rapid melting and revitrification, microsecond timescales can be achieved.

due to cost, ease-of-use, and reproducibility of the equipment and methods. Though some successful examples have been reported, many of the studies have focused on well-behaving systems, especially those containing large macromolecular components. The most accessible method for tr-EM remains manual sample mixing and application, though this method is limited to slow timescale processes that occur on the range of seconds to minutes. Technologies that can be applied to faster timescale processes are still in their infancy, though recent development has seen a push towards more automated and reproducible systems that can access a wider range of biological samples.

Notable biological applications include studies by Maeots et al. [44], who designed custom-built microfluidics chips to investigate RecA-ssDNA filament growth on the millisecond timescale. Actomyosin dissociation has been extensively studied, from early work by Menetret et al. [46] using flash photolysis of non-hydrolyzable ATP analogs to capture transient dissociation states, to more recent approaches characterizing the dissociation by White, Walker, and Trinick [60] employing gas-assisted atomizers and Klebl et al. [55] using a blot-then-spray approach to achieve 10 ms and 7 ms temporal resolution of the dissociation event, respectively. Ribosomes have also served as model systems for tr-EM experiments. Fischer et al. [61] studied the process of translation, correlating tRNA movement through binding sites with large-scale 30S subunit rearrangements. More recently, Dandey et al. [56] applied the inkjet-based system, Spotiton [52], to time-resolved cryo-EM, demonstrating ribosome assembly using a novel, dual-inkjet method of sample application. Additionally, Bhattacharjee et al. [62] used SiO_2-coated microfluidics chips to study 70S ribosome interactions with a GTPase under heat shock conditions, revealing angular rotation of the 30S subunit preceding dissociation.

The fastest timescale dynamics of a biological system observed by time-resolved EM to date was demonstrated by Voss et al. [53] with 5 μs resolution of electron-beam induced GroEL disassembly, though under non-physiological conditions. Meanwhile, Torino et al. [45] studied biologically relevant GroEL/ES complex assembly using microfluidics and spray-based application with a time resolution of 13 ms.

More mechanistic biological applications of time-resolved EM include studies of ion channels and E3 ligases. Gangwar et al. [63] used a solenoid-driven plunging device modeled after a high-speed sample preparation device (Shake-it-off [50]) to study the kainate receptor gating mechanism on the second timescale, resolving transient intermediate conformations corresponding to open channel and non-conducting states. Studies of E3 ligase-mediated substrate polyubiquination conducted by Bodrug et al. [64] employed manual preparation (limiting timescales to seconds-minutes) and cryoDRGN, an AI-based method for solving heterogeneity in cryo-EM datasets, to reveal four major states of an E3 ligase heavily involved in cell cycle control.

1.4 Objectives of this work

Despite significant progress in time-resolved approaches, many molecular systems remain challenging to study due to the transient nature of their functionally important conformational states, as well as the cost barriers and technical complexity of the biophysical methods.

These large protein complexes perform essential, energy-dependent tasks within the cell by cycling through dynamic structural rearrangements that enable substrate translocation and remodeling, making them ideal subjects for structural biology studies that capture motion at the molecular level. To address the challenges of studying such dynamic systems, this work explores both computational and experimental approaches to investigate the behavior of macromolecular machines across multiple timescales. The thesis is organized around three objectives: 1) probing functional transitions of the human mitochondrial LONP1 protease with long-timescale molecular dynamics, 2) development of Mix-it-up for accessible, low-cost time-resolved cryo-EM, and 3) investigation of early-stage DNA-binding by the DNA repair enzyme Polymerase θ and complex formation. Together, these studies illuminate how structural dynamics underpin the function of molecular machines in their native cellular roles while advancing the methodological toolkit for studying dynamic biological processes.

The first chapter investigates the conformational dynamics underpinning function in the human mitochondrial LONP1 protease using microsecond-timescale molecular dynamics (MD) simulations. LONP1 plays a central role in mitochondrial protein quality control, but the dynamic processes governing the transition between its active and inactive functional states remain poorly understood. Through extensive MD simulations totaling over 100 μs, this work aims to: (1) characterize the conformational flexibility of nucleotide binding states within the AAA+ domain, (2) examine the relationship between nucleotide occupancy and substrate engagement by pore loop motifs, (3) investigate the spontaneous inactivation of proteolytic sites following substrate release, and (4) build a framework for identifying allosteric communication pathways between the ATPase and protease domains. The simulations presented in this work provide atomic-level detail into dynamic processes occurring on timescales invisible to conventional experimental methods and provide novel insights into the conformational landscape that describes LONP1 function.

The second chapter addresses the technological barriers limiting widespread adoption of time-resolved cryo-EM by developing "Mix-it-up" (MIU), a cost-effective tool for rapid sample mixing and vitrification. While existing time-resolved cryo-EM technologies can capture millisecond-timescale processes, they typically require expensive, specialized equipment that limits accessibility. MIU builds upon previous spray-based vitrification designs to enable on-grid sample mixing with vitrification delays as low as $\sim$120 ms, at a fraction of the cost of commercial alternatives. This work demonstrates MIU's capabilities through: (1) high-resolution structure determination of mixed protein samples, (2) capture of pH-induced viral capsid contraction, and (3) visualization of ATP-dependent chaperonin complex formation. The development establishes a practical platform for studying rapid biochemical processes and expands access to time-resolved structural biology.

The third objective employs time-resolved cryo-EM to investigate the early conformational dynamics of Polymerase θ upon DNA binding, addressing fundamental questions about the role of oligomeric states in DNA repair. Previous studies have shown that DNA binding triggers disassembly of the tetrameric form of the helicase domain of Polθ (PolθHel) into DNA-bound dimers, but the sequence of events and potential biological relevance of tetrameric intermediates remain unclear. Using time-resolved approaches, this work captures early conformational ensembles within minutes of DNA addition, revealing: (1) the presence of

both tetrameric and dimeric states during initial DNA binding and (2) putative DNA-bound tetrameric intermediates that may participate in early DNA recognition. These findings provide new details of how Polθ recognizes and processes DNA substrates in the context of DNA repair.

1.5 Significance of this work

The work presented here aims to bridge the time gap of observable protein motions in structural biology by combining computational and experimental strategies to reveal the dynamic behavior of macromolecular machines across multiple timescales. Through these objectives, this thesis establishes complementary case studies that highlight both methodological innovation and biological discovery. Collectively, these contributions advance our ability to interrogate transient structural states, provide broadly accessible tools for the structural biology community, and deepen our understanding of protein quality control and genome maintenance mechanisms.

2

Probing functional transitions of the human LONP1 protease through long-timescale simulations

2.1 Chapter overview

While conventional molecular dynamics studies of large macromolecular complexes are typically limited to short timescales due to computational expense, many biologically relevant conformational changes occur on microsecond-to-millisecond timescales that are often inaccessible to physical methods. This chapter applies long-timescale atomistic simulations to investigate the conformational landscape of the human mitochondrial LONP1 protease, a central regulator of mitochondrial proteostasis. With more than 100 μs of simulation time on a system exceeding 400 k atoms, this work captures conformational transitions associated with nucleotide binding, substrate engagement, and proteolytic site regulation at timescales invisible to physical methods. The findings illuminate dynamic processes that underlie LONP1's transition between active and inactive states and reveal long-range links between distant functional regions. Together, these findings provide a computational framework for understanding how large macromolecular machines coordinate their complex functional cycles through long-range communication networks.

Contributions This work was a collaborative effort between myself (Lauren Alexandrescu), William Rochira, Gabriel C. Lander, and Philip C. Biggin. LA, GCL, and PCB conceptualized and led the project. LA developed initial molecular dynamics models and performed all MD simulations. LA and WR performed the analysis and interpretation with guidance from GCL and PCB. LA prepared the figures and wrote the chapter.

2.2 Introduction

2.2.1 AAA+ proteins

AAA+ proteins (ATPases Associated with a variety of cellular Activities), encompass a family of protein machinery that share a highly conserved ATPase domain responsible for converting the chemical energy stored in ATP into the mechanical force required to perform a range of functions in the cell [3]. AAA+ proteins are found in all kingdoms of life, as well as viruses [65], and have evolved the capacity to interact with a wide range of proteins and nucleic acid substrates to carry out activities involved in cytoskeleton remodeling, membrane fusion, the heat shock stress response, and mitochondrial proteostasis [3, 66]. Within the mitochondria, AAA+ proteases undergo successive cycles of ATP hydrolysis to drive the translocation of substrates through a central pore, a mechanism that has now been recognized to be widely conserved across all protein translocases [3, 66, 67].

2.2.2 The mitochondrial protein quality control network

Mitochondria are the centers of a plethora of metabolic processes, including energy production, calcium signaling, and programmed cell death [68]. These organelles have an impressive capacity to respond to metabolic or environmental stimuli by rapidly changing their morphology through proteolytic remodeling of their proteome [67–69]. This coordinated proteolytic activity enables mitochondria to regulate important aspects of cellular biology, including oxidative phosphorylation, proteostasis, and apoptosis [68]. One of the primary determinants in regulating mitochondrial function is the mitochondrial protein quality control network, which consists of ATP-dependent chaperones and proteases that target misfolded or damaged proteins for repair or degradation (Fig 2.1) [67, 69]. This network is critical for mitigating mitochondrial stress, which is implicated in a plethora of human disorders, including cardiovascular disorders, cancer, and neurodegenerative disorders such as Alzheimer's [69–71]. To understand how dysregulation of the enzymes within this network can result in the progression of these disorders, a detailed understanding of the structural basis of the substrate remodeling and proteolytic mechanisms of ATP-dependent proteases, such as LONP1, must first be established.

The AAA+ proteases that make up the mitochondrial protein quality control network perform a variety of functions depending on their localization and substrate specificity [3, 67, 72, 73]. This introduction, however, will focus solely on members of the AAA+ family, though other proteases, such as OMA1 and PARL, additionally play key roles within the network [72]. The AAA+ proteases of the mitochondrial protein quality control network include i-AAA protease (YME1L), m-AAA protease (AFG3L2), ClpXP, and LonP1 (Fig 2.1).

The i-AAA protease is localized in the inner-mitochondrial membrane with its catalytic (protease) domain, YME1L, facing towards the intermembrane space [74]. YME1L belongs to the FtsH-like protease family [3]. It is a metalloprotease, containing a motif consisting of two histidines and a glutamate (HExxH) that coordinate a zinc ion required for proteolysis. YME1L oligomerizes into a hexameric barrel and is tethered to the intermembrane. The enzyme plays a central role in regulating mitochondrial morphology through its processing

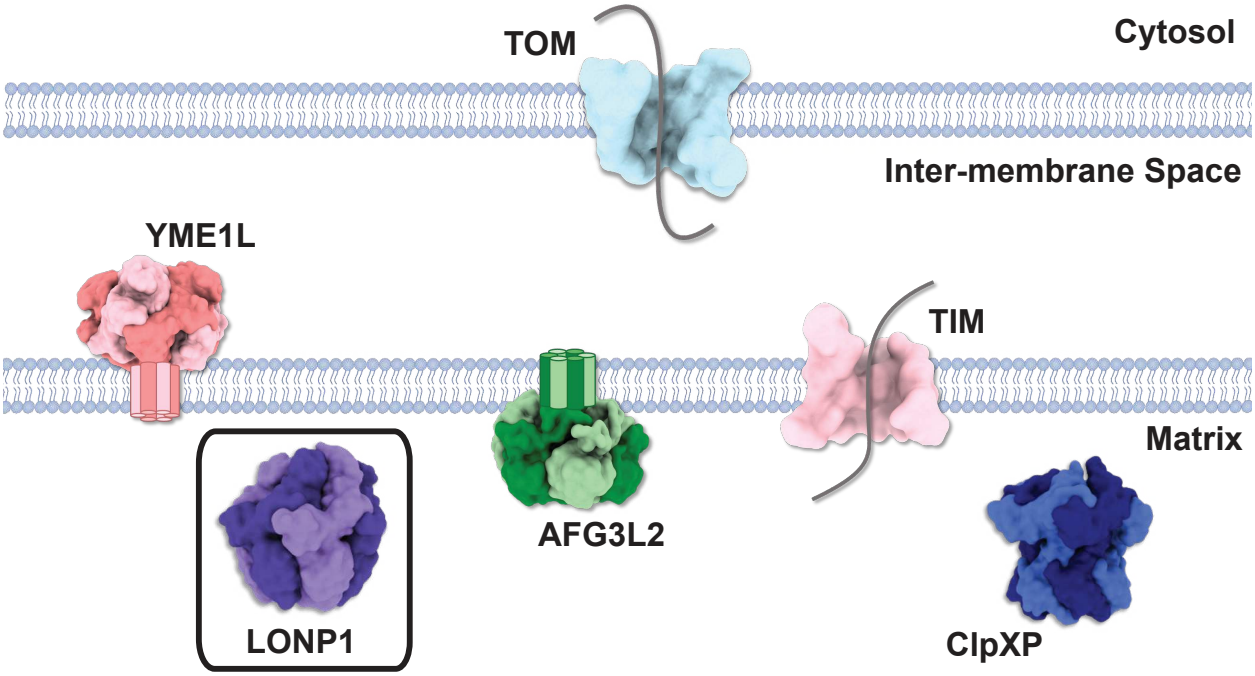


Figure 2.1: AAA+ Proteases of the mitochondrial protein quality control network. The mitochondrial protein quality control network consists of a diverse set of AAA+ proteases that coordinate to protect against mitochondrial stress. Common targets of the network include proteins misfolded upon import by the TOM (light blue) and TIM (light pink) complexes, as well as proteins that have become damaged due to reactive oxygen species produced by the electron transport chain and protein aggregates. Two AAA+ proteases are tethered to the inner mitochondrial membrane: the i-AAA protease, YME1L (coral), in which the proteolytic domain faces the intermembrane space, and the m-AAA protease, AFG3L2 (green), in which the catalytic domain is oriented towards the matrix of the mitochondria. Two soluble enzymes, ClpXP (blue) and LONP1 (purple), also play a key role in maintaining proteostasis within the mitochondria.

of proteins that mediate mitochondrial fission and fusion. One such protein, Optic Atrophy Protein 1 (OPA1), binds to YME1L as a long isoform (L-OPA1) and gets cleaved into shortened isoforms (S-OPA1) [74, 75]. The ratio of L-OPA1 to S-OPA1 in the intermembrane space controls the balance between fusion and fission processes. Furthermore, YME1L also regulates phospholipid composition of the mitochondrial membranes and apoptosis through the processing of PRELI Domain Containing 1, a lipid transport protein that regulates cardiolipin abundance [74]. Additional roles of YME1L include the regulation of mitochondrial protein import via proteolysis and putative functions as a chaperone and disaggregase [76].

The m-AAA protease is also embedded in the inner-mitochondrial membrane, however its protease domain, AFG3L2, faces the mitochondrial matrix. Similarly to YME1L, AFG3L2 is also a hexameric metalloprotease and member of the FtsH-like protease family, containing a HExxH zinc-binding motif that enables hydrolysis of the substrate's peptide bonds [3, 73]. Human m-AAA composition can consist either of homo-oligomers of AFG3L2, or hetero-oligomers of AFG3L2 and paraplegin, a metalloprotease encoded by the SPG7 gene [77]. The m-AAA protease also functions to degrade misfolded and oxidatively-damaged proteins, especially components of the electron transport chain and complexes that are involved in calcium transport and cristae organization. The enzyme also plays a role in mitochondrial ribosome assembly [78].

ClpXP is a soluble enzyme, located in the mitochondrial matrix. It is composed of two polypeptide chains: ClpX, an unfoldase, and ClpP, a peptidase [73, 79]. The independence of these chains enables modularity and regulatory flexibility of the enzyme. These two components form a stacked, tetradecameric barrel to carry out protein degradation activities [79]. ClpP is a serine protease and member of the Clp family [3]. Its catalytic motif consists of a serine-histidine-aspartate triad [79]. ClpXP primarily degrades misfolded and oxidatively-damaged proteins, including components of the electron transport chain. Additionally, its removal of proteins involved in mitochondrial ribosome assembly and RNA-binding, such as ERAL1 and DmLRPPRC1 [80, 81], contribute to the regulation of translation and protein synthesis. ClpXP also shares common substrates with LONP1, suggesting their cooperative activity in proteolysis within the mitochondrial matrix.

The human LONP1 protease is a central enzyme of the mitochondrial protein quality control network and is also located in the matrix of the mitochondria [82]. It is a hexameric, serine protease and member of the Lon family, but in contrast to ClpP, employs a serine-lysine catalytic dyad to perform proteolysis [82]. Further, the AAA+ and protease domains are encoded on a single polypeptide, enabling LONP1 to tightly couple substrate recognition, unfolding, and degradation within a single, self-contained unit [4, 83]. LONP1 targets oxidatively damaged or misfolded proteins, with the majority of its substrates comprising components of the electron transport chain that have been undergone damage as a result of the reactive oxygen species released into the mitochondrial matrix as a byproduct of oxidative phosphorylation [82, 84]. Degradation of these targets by LONP1 is an essential function of the mitochondrial protein quality control network and ultimately ensures organellar and cellular health. Numerous studies have shown that the dysregulation of LONP1 leads to the progression of age-related diseases such as Alzheimer's and cancer [85, 86].

2.2.3 Conserved mechanisms of AAA+ proteases

While these enzymes execute diverse roles, the principles governing their mechanics are highly conserved. Each consists of a AAA+ domain that binds nucleotide and converts the chemical energy from ATP hydrolysis into mechanical energy that can be used to unfold and translocate substrates through a central channel (Fig 2.2) [66].

Various models have been proposed to describe the mechanism of substrate translocation by AAA+ proteins (Fig 2.3A,B) [3, 87–89]. The hand-over-hand model states that the engagement of a particular subunit of the enzyme with the substrate is dictated by the nucleotide state of its AAA+ domain (Fig 2.3A) [3, 4, 90, 91]. In an actively translocating state, the AAA+ subunits are arranged in a spiral staircase around the substrate positioned in the central pore [4, 83, 89, 90]. ATP binding is thought to promote substrate engagement by positioning aromatic residues from a conserved "pore loop 1" motif (located between β2 and helix α2 of the AAA+ domain) to intercalate with the backbone of the substrate (Fig 2.3A) [3]. Upon hydrolysis and nucleotide release, the subunit disengages from the substrate and returns to the top of the register in the staircase, triggering a sequential downward shift of the remaining engaged subunits, which generates a pulling force ("power stroke") on the substrate. The disengaged subunit is conventionally referred to as the "seam subunit" (Fig 2.3A) [3]. In contrast to this model, the Brownian ratchet model posits that the thermal (Brownian) motion of the substrate enables it to diffuse within the central pore, while the pore loops act as spring-loaded pawls that stochastically grip and release the substrate (Fig 2.3B) [87, 88, 92, 93]. ATP binding and hydrolysis then reset the pawls further along the chain, preventing backsliding and producing a net downward translocation of the substrate [87, 88]. Some studies suggest that these two models are not strictly independent of one another, and rather a combination of the two is involved in substrate translocation [94].

The coordination of ATP hydrolysis and the resulting strength of the power stroke remains debated by the field. The main theories describing this event include the sequential, concerted, and probabilistic models of hydrolysis (Fig 2.3C-E) [95]. In sequential hydrolysis, ATP hydrolysis occurs in a strict order around the ring, with one subunit firing at a time, and triggering a shift in the subunit register with each cycle (Fig 2.3C) [3, 96]. This model is most consistent with cryo-EM data of AAA+ proteases describing the spiral conformation of subunits, and is thought to produce small, regular 2-residue steps of translocation [3, 4, 83, 89, 90]. A variant of this model proposes that three adjacent subunits fire together in each cycle, generating 6-residue steps, which correlate to experimental data [96]. The concerted model instead postulates that multiple or all six subunits hydrolyze ATP simultaneously, producing power strokes of as many as 30 residues in a single translocation step (Fig 2.3D) [95, 96]. Finally, the probabilistic model allows subunits to fire out of sequence, reducing the risk of stalling if an inactive subunit occupies a hydrolysis position (Fig 2.3E) [95, 96]. Several variations of this model exist, with one proposing a clockwise, two-residue-step mechanism and a second proposing an anti-clockwise long step model in which ATP hydrolysis at the top subunit drives anti-clockwise transition to the bottom, enabling translocation of six residues with bursts of up to 24 residues, consistent with single-molecule kinetic data [96, 97].

Following translocation of the substrate through the AAA+ domain, the unfolded polypeptide

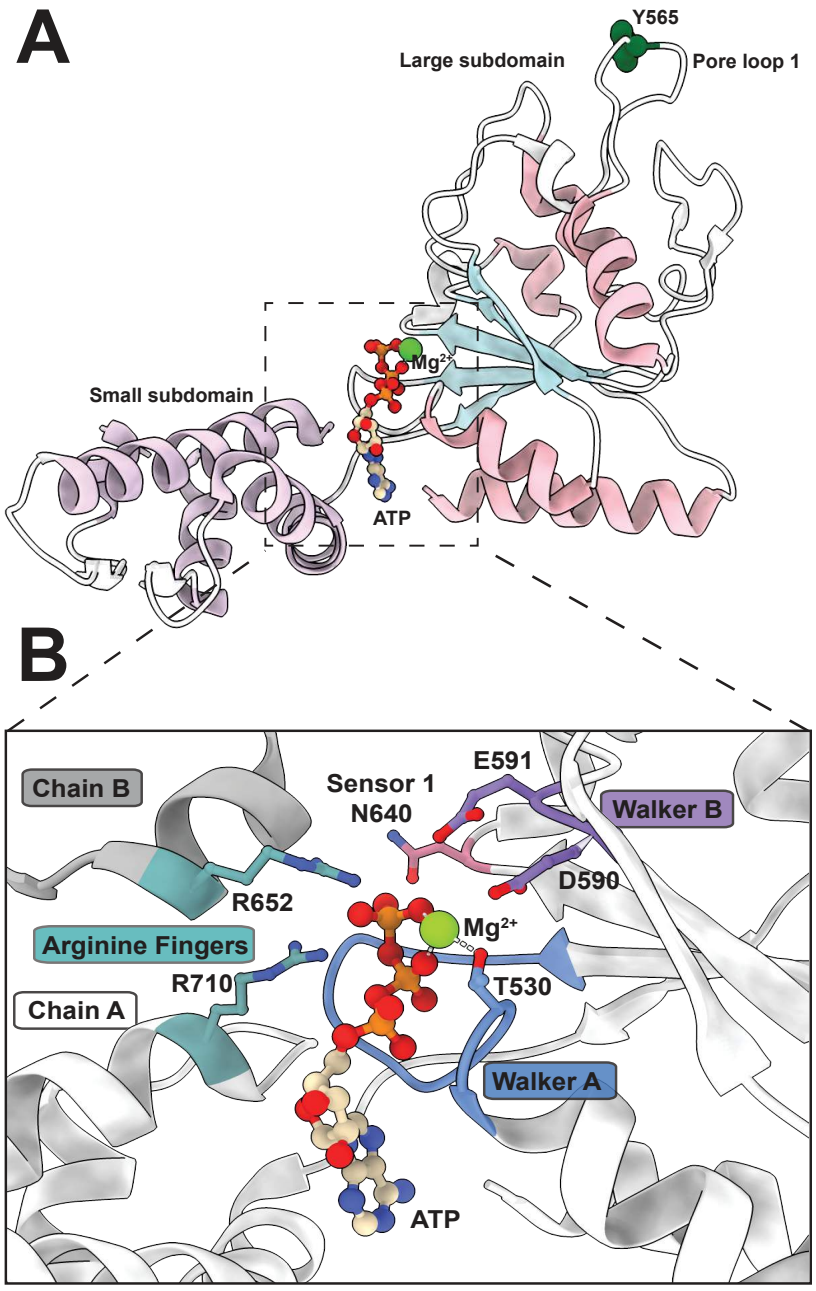


Figure 2.2: The AAA+ module. A-B) The AAA+ module. The AAA+ module (here shown of human LONP1, PDB:7KRZ) is highly conserved across the AAA+ family of proteins. It consists of the small (lavender) and large (pink and blue) subdomains, tethered by a loop motif **(A)**). The small subdomain is α-helical and contains an arginine finger (teal sticks shown in **(B)**) that senses the nucleotide binding state. The large subdomain consists of two α-helical motifs (pink) flanking a central, parallel β-sheet (blue). The aromatic residue (Y565, green) of the pore loop 1 motif is responsible for intercalating with the substrate backbone during translocation. Walker A and Walker B motifs (dark blue and purple, respectively) are responsible for binding ATP (wheat ball and stick) seen in **(B)**. A secondary arginine finger from the adjacent chain (gray cartoon), communicates the nucleotide state between subunits.

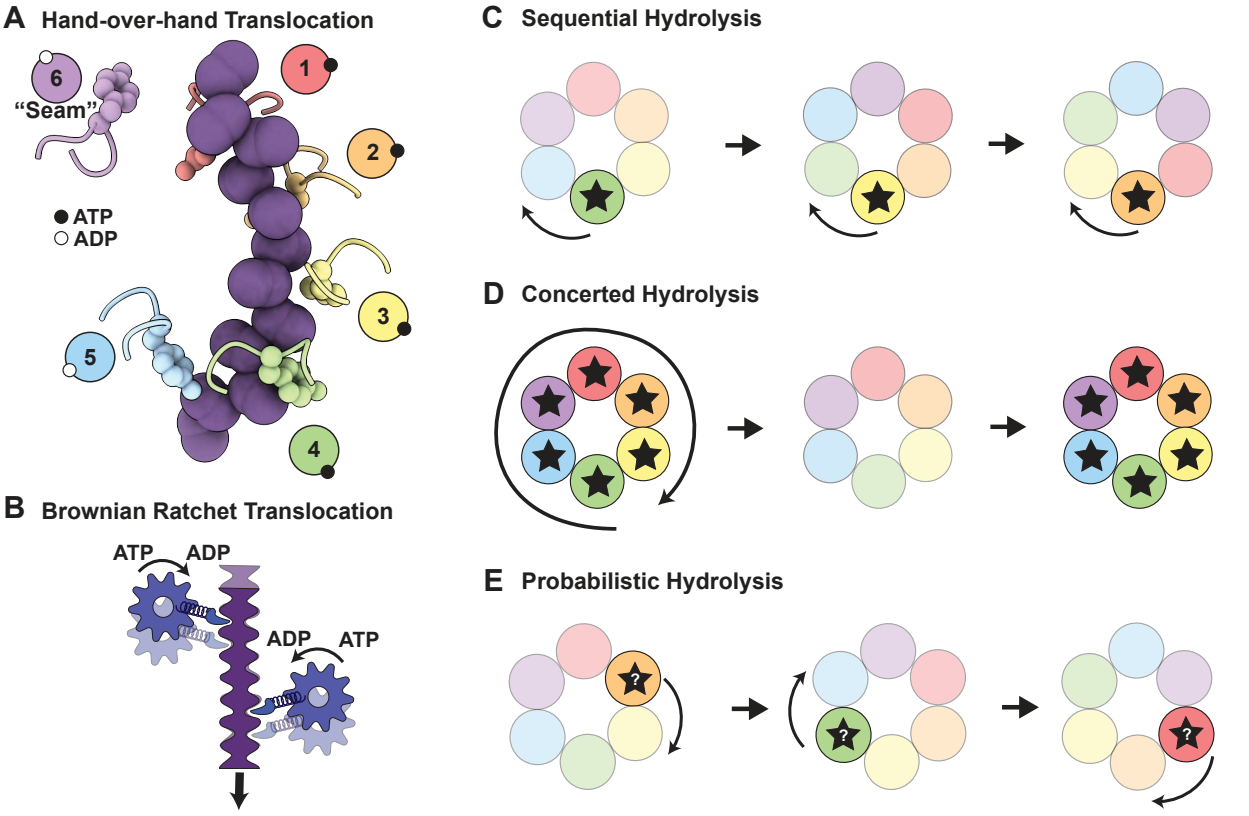


Figure 2.3: Models of substrate translocation and ATP hydrolysis in AAA+ proteases. A-B) Models of substrate translocation in AAA+ proteases. A) Hand-over-hand translocation. The hand-over-hand model of substrate translocation states that the nucleotide binding state of the subunits (rainbow circles) dictate pore loop (rainbow cartoon) engagement with the translocating substrate (plum spheres). Aromatic residues (rainbow spheres) of pore loop motifs on the ATPase domains intercalate with the substrate backbone as it is translocated through the central pore of the enzyme. The pore loops of subunits bound to ATP (black circles) actively engage with the subunit, with aromatic residues intercalating with the substrate backbone, gripping it like hands on a rope. Upon hydrolysis, a transition to an ADP-bound state (white circles) of the subunit occurs, resulting in disengagement of the corresponding pore loop and a shift in the subunit register, yielding a downward pull ("power stroke") on the substrate. **B) Brownian ratchet model of translocation.** The Brownian ratchet model states that substrate diffuses through the central pore of the enzyme due to thermal (Brownian motion), without the requirement of a power stroke. ATP exchange enables rounds of pore loop release and engagement, and the pore loops act as pawls to prevent backward motion of the substrate, yielding a net downward translocation. **B-D) Schematics models of ATP hydrolysis. C) Sequential hydrolysis.**) Hydrolysis (star) occurs sequentially, one-at-a-time, with each event causing a shift in the subunit register. **D) Concerted hydrolysis.**) Hydrolysis occurs in all subunits simultaneously, resulting in a larger power stroke and a full 360° rotation of the subunit positions within the register. **E) Probabilistic hydrolysis.**) Hydrolysis occurs randomly, with each event triggering a power stroke and register shift.

chain reaches the protease domain, where a short proteolytic motif, commonly consisting of either a catalytic serine or cysteine, performs peptide bond hydrolysis, leading to substrate degradation [98, 99].

2.2.4 The Lon protease

Domain organization
LONP1 is organized into three functional domains: an N-terminal domain responsible for substrate binding and recognition, an AAA+ domain that hydrolyzes ATP, and a protease domain consisting of a serine-lysine catalytic dyad that functions to degrade incoming polypeptide substrates (Fig 2.4A) [4, 82, 83]. The functional form of LONP1 is a hexameric barrel with a central pore through which substrate is translocated.

Distinct activity states of LONP1 revealed by cryo-EM
Structures of LONP1 resolved by cryo-EM reveal several distinct conformations of the enzyme, corresponding to different states of activity (Fig 2.4) [4, 83]. In the absence of substrate, the enzyme adopts an open lock washer form with a left-handed spiral staircase arrangement of the AAA+ and protease domains (Fig 2.4A). Upon addition of substrate, the AAA+ domains reconfigure to a closed conformation with a right-handed spiral, enclosing the unfolded polypeptide within the central pore (Fig 2.4C). These conformations reflect a substrate-free and substrate-bound state, respectively. Interestingly, proteolytic activation has only been observed in the presence of inhibitor or substrate bound within the protease domain. This state is characterized by a C_6-symmetric organization of the protease domains, as well as a rearrangement of secondary structure at the proteolytic site, resulting in the release of an auto-inhibitory 3_{10}-helix (Fig 2.4B) to form an elongated substrate-binding groove (Fig 2.4D) [4, 83].

Characterization of the LONP1 proteolytic site
The inactive conformation has been characterized by the presence of a 3_{10}-helix involving the catalytic serine (S855), as well as a hydrogen bond between an aspartate (D852) within the proteolytic loop and the catalytic lysine (K898) (Fig 2.4C) [4, 83, 100]. Both features work to inhibit proteolytic cleavage of substrate: the 3_{10}-helix blocks substrate binding, and the hydrogen bond prevents the catalytic lysine from functioning in proteolytic cleavage [4, 83, 101]. Upon activation, the hydrogen bond is broken and the 3_{10}-helix releases into an extended loop that acts as a binding groove for incoming peptide (Fig 2.4D) [4, 83, 100]. The active and inactive states are the only conformations observed in cryo-EM, indicating that any intermediate states are likely highly dynamic and short-lived. In order to understand how LONP1 regulates the transition between active and inactive states of the proteolytic sites, we endeavored to visualize these intermediate states with long-timescale MD.

This work
To capture a spectrum of conformational and nucleotide-dependent states that confer the function of LONP1, we performed unbiased, atomistic molecular dynamics simulations on four distinct models of the more than 400k atom system. These models comprised: (i) the active

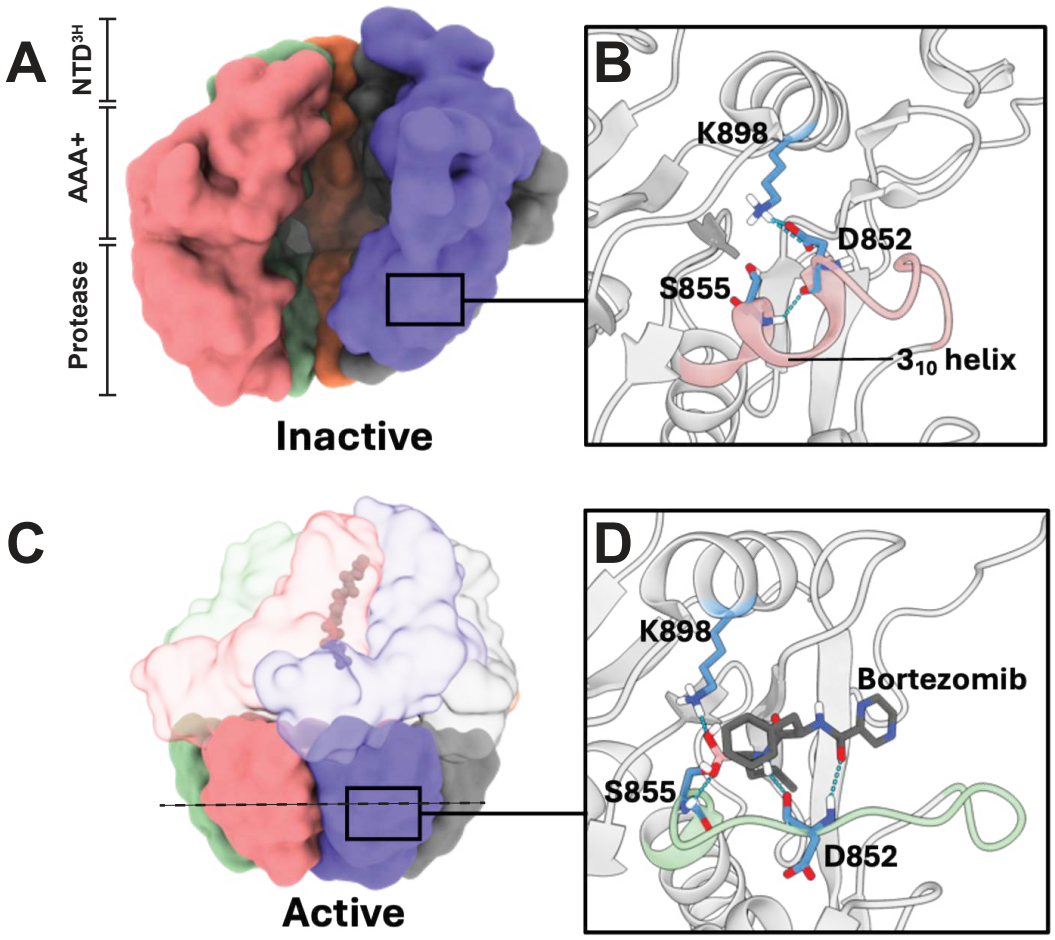


Figure 2.4: Cryo-EM structures of inactive and active states of human LONP1. A-B) The inactive state of human LONP1. (A) LONP1 consists of three functional domains encoded on a single polypeptide chain, including an N-terminal domain (of which only a tri-helical portion (NTD^{3H}) has been resolved to high resolution by cryo-EM), involved in substrate recognition, a AAA+ module responsible for the binding and hydrolysis of ATP to power substrate translocation, and a protease domain, where unfolded polypeptide substrates are directed for proteolysis. The functional form of LONP1 is a hexameric barrel. The macroscopic conformation of the inactive state is characterized by an open lockwasher form (PDB:7KSL, with one of the six subunits not shown), with the AAA+ and protease domains arranged in a left-handed spiral staircase. **(B)** At the proteolytic sites, the proteolytic loop (residues 846–857, coral) forms an auto-inhibitory 3_{10}-helix, containing the catalytic serine (S855), that sterically inhibits substrate binding. A hydrogen bond between the aspartic acid of the proteolytic loop (D852) and the catalytic serine (K898) chemically occludes the formation of the serine-lysine catalytic dyad. **C-D) The active state of human LONP1. (C)** In the presence of substrate, LONP1 forms a closed conformation (PDB:7KRZ), with the ATPase domains arranged in a right-handed spiral staircase around the translocating polypeptide chain. The protease domains adopt a planar, C_6-symmetric organization upon activation of the enzyme (dashed line). **(D)** At the proteolytic sites, the auto-inhibitory 3_{10}-helix relaxes to form an elongated protein-binding groove, here shown induced by a peptidomimetic inhibitor, bortezomib, which covalently binds to the catalytic serine (S855).

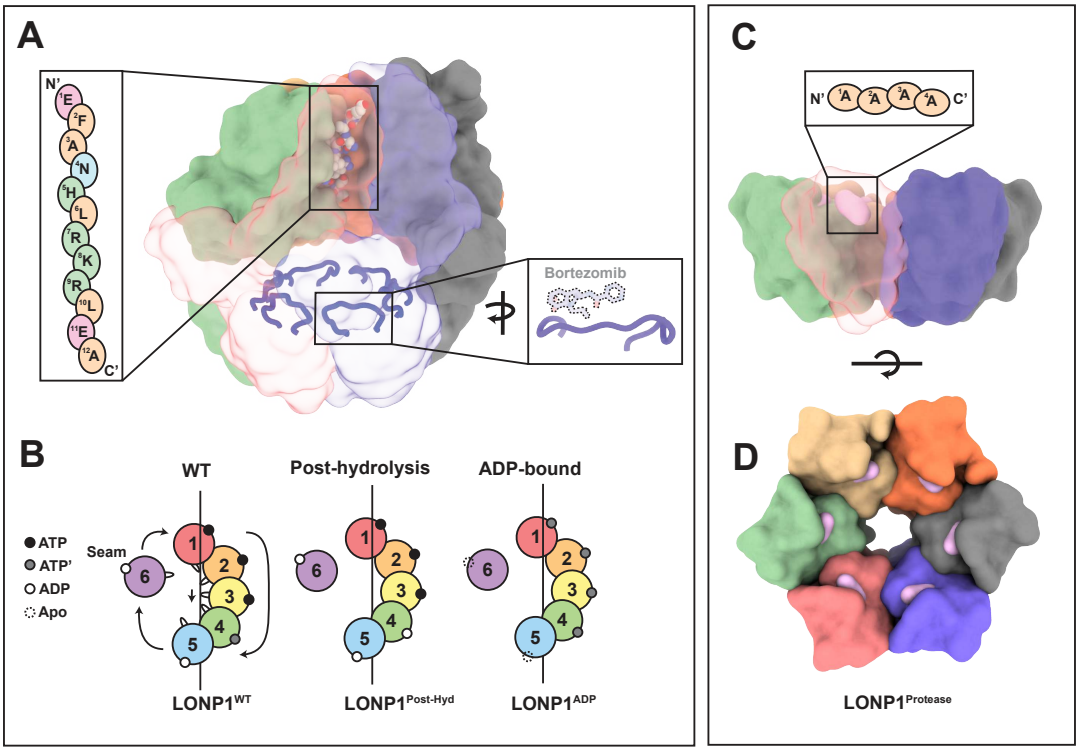


Figure 2.5: Models of LONP1 generated for MD simulations. A-B) Models used for microsecond-timescale simulations. A) The active conformation of LONP1 (PDB:7KRZ) was adapted for simulations. The deposited structure contained a poly-alanine 12-mer modeled into the substrate density observed within the central pore of the ATPase domain. In MD models, these residues were mutated to a sequence found in an endogenous substrate of LONP1, the steroidogenic acute regulatory protein (StAR). The sequence is depicted in the inset to the left of the enzyme, and colored by side chain classification (negatively-charged residues in pink, positively-charged residues in green, polar residues in blue, and hydrophobic residues in orange). The bortezomib molecule (dashed lines, transparent) is shown removed from the proteolytic sites (purple tubes) with a 180° rotation in the right inset of **(A)**. The resulting model was further delineated into the three models used for MD: 1) a wild-type model ($LONP1^{WT}$) that reflects the nucleotide states of the enzyme observed in cryo-EM with ATP (black and gray circles) bound to subunits in position one to four (red, orange, yellow, and green circles, respectively), and ADP (white circles) bound to subunits in position five and six (blue and purple circles, respectively), 2) a post-hydrolysis state of LONP1 ($LONP1^{Post}$), in which the ATP at the putative hydrolysis site (position four, green) is modeled as ADP to represent the nucleotide state following hydrolysis and release of the inorganic phosphate, and 4) an ADP-bound state of LONP1 ($LONP1^{ADP}$), in which the first four subunits are modeled bound to ADP and the nucleotides in positions five and six are removed entirely (dashed circles). **C-D) Model used for simulations of substrate release.** Side **(C)** and top **(D)** views of the protease-only model ($LONP1^{Protease}$) used for investigation of substrate release dynamics. $LONP1^{Protease}$ was prepared from a cryo-EM structure of human LONP1 in the active state with substrate density (pink surfaces) observed in each of the six proteolytic sites (PDB:7P0M) [100]. The density was modeled in the deposition as tetra-alanine (inset), and this sequence was also used for MD.

state model from cryo-EM, containing wild-type nucleotide states (PDB:7KRZ, hereafter referred to as "LONP1^{WT}"), (ii) a post-hydrolysis state where the ATP bound at subunit position four in the register was replaced with an ADP molecule (termed "LONP1^{Post}"), (iii) a fully ADP/apo state ("LONP1^{ADP}"), and (iv) a protease-only construct of the active state with a tetra-alanine substrate bound at the proteolytic sites (PDB:7P0M$^{\Delta(NTD^{3H}\text{-}AAA+)}$, hereafter referred to as "LONP1^{Protease}") (Fig 2.5). The full-length models derived from PDB:7KRZ (i-iii) were also modeled with a 12-residue polypeptide substrate bound within the central pore of the ATPase domain, for which density had been observed in the cryo-EM structure (Fig 2.5A). These simulations allowed us to investigate nucleotide dynamics within the ATPase domain, the impact of nucleotide state on AAA+ pore-loop engagement with translocating substrate, the dynamics of substrate release from proteolytic sites, and proteolytic inactivation events. The microsecond timescales achieved were sufficient to observe large-scale domain movements and conformational changes in secondary structure, enabling a better understanding of the conformational landscape underpinning LONP1 function.

2.3 Results

2.3.1 Nucleotide flips within the LONP1 AAA+ domain reflect multiple stable binding poses

Previous cryo-EM studies have proposed a strict relationship between the nucleotide state and enzymatic activity [3, 4, 89, 90]. To better understand this relationship, the conformational changes within the nucleotide binding pocket were monitored over the course of the microsecond timescale simulations. In contrast to the rigid poses depicted in cryo-EM structures of LONP1 (Fig 2.6A), the nucleotides proved to be highly dynamic in simulations. To classify the observed poses, affinity propagation clustering of the ATP and ADP coordinates was performed independently for each subunit of LONP1 (Fig 2.6B). The cluster granularity was tuned using a low preference parameter to promote fewer, more distinct clusters and increase the observable heterogeneity between poses. The coordinates of the nucleotides at each timestep of the simulation were then mapped to the nearest cluster match based on RMSD calculations in order to determine the time dependence of the poses (Fig 2.6C). This revealed multiple nucleotide binding poses that were stable on the microsecond timescale and that have not previously been observed by cryo-EM. More than one binding pose was observed in every subunit of the LONP1^{WT} model in at least two of the replicates, indicating the tendency for both ADP and ATP to adopt multiple conformations on the microsecond timescale. Further, we did not observe a significant change in nucleotide flexibility in the other simulated models (LONP1^{Post} and LONP1^{ADP}), indicating that the observed nucleotide dynamics are consistent across models at these timescales (Fig A.1). To further characterize the nucleotide binding poses, we compared the representative structures of each cluster identified. In several of the subunits, the bound nucleotide undergoes a "flip", characterized by a $\sim$180° rotation of the adenine ring that coincides with a $\sim$90° rotation of the ribose ring, while the phosphate tail remains relatively rigid (Fig 2.6B,D-E). The phosphate tails, however, demonstrate little flexibility as a result of rigid coordination by the Mg^{2+} ion and arginine fingers of the same chain (R710) and neighboring chain (R652), which are thought to

communicate nucleotide state between subunits (Fig 2.6D-E) [3]. Therefore, the movement of the ribose and adenine ring is responsible for the majority of the dynamics of the nucleotide observed in simulation.Full details of the time-dependent clusters observed across all models and replicates can be found in *Appendix A* (Fig A.1).

To determine whether these binding poses reflect physiologically relevant states, we performed a PDB search for experimental protein structures containing an ATP ligand. The ATP binding poses from structures of LONP1 and other AAA+ proteases such as ClpX and p97 did not show a high degree of conformational sampling, with the majority matching the first cluster, which reflects the initial binding pose based on our starting cryo-EM model (PDB:7KRZ). Given that the resolution of nucleotide binding pockets in cryo-EM structures is often too low to resolve atomic positions of nucleotide (Fig 2.6A), and that the conformations resolved by both cryo-EM and X-ray crystallography often reflect low-energy states, we also queried structures of ATP-binding proteins solved by NMR. In these, we expected to find a higher degree of conformational flexibility in the observed nucleotide binding poses that reflected conformations observed in solution [102]. Of the eight NMR structure ensembles containing an ATP ligand deposited in the PDB at the time of this study, two contained models with binding poses similar to the second and third cluster observed in simulation, corresponding to structures of the nucleotide-binding domain of rat α1 Na,K-ATPase (PDB:1MO8) and an ATP-binding DNA aptamer (PDB:9KTJ), respectively (Fig 2.7) [103, 104]. The all-residue heavy-atom RMSDs from the mean structures were 1.46 Å and 0.98 Å, respectively. Alignment of the simulated poses to the NMR poses yielded RMSDs of 3.0 Å and 2.8 Å for the second and third cluster, respectively. These findings suggest that the multiple nucleotide-binding conformations observed in our MD simulations reflect physiologically relevant states but often remain unresolved in static structural studies.

To understand the molecular basis for these nucleotide conformational transitions, we examined the local protein environment surrounding the binding sites where nucleotide flips were observed. This analysis revealed specific conformational dynamics within the binding pocket that facilitate and stabilize the different nucleotide poses. Specifically, four highly conserved aromatic residues that bridge the small and large subdomains of the ATPase module (H491, Y492, Y661, and Y673) help to coordinate the nucleotide binding pose (Fig 2.8A). In the initial ATP binding pose, three key hydrogen bonds stabilize the nucleotide: two form between the backbone atoms of Y492 and the adenine ring (one to the amino group and one to the imidazole nitrogen), while a third forms between the side chain hydroxyl of Y661 and the adenine amino group (Fig 2.8A). During the simulation, this hydrogen bonding network undergoes a dynamic rearrangement. After approximately 3.5 μs, the two hydrogen bonds involving the amino group are broken, while the bond to the imidazole nitrogen is retained. This selective loss of hydrogen bonds reduces the constraints on the adenine ring, allowing it to undergo a $\sim$180° flip into an alternative conformation that remains stable for approximately 6 μs (Fig 2.8B). These observations demonstrate that the aromatic residues within the binding pocket actively regulate nucleotide conformation through specific hydrogen bonding interactions.

The conformational flexibility of the aromatic residue was further assessed by measuring the range of side chain dihedral angles sampled during the simulation (Fig 2.8D). Interestingly,

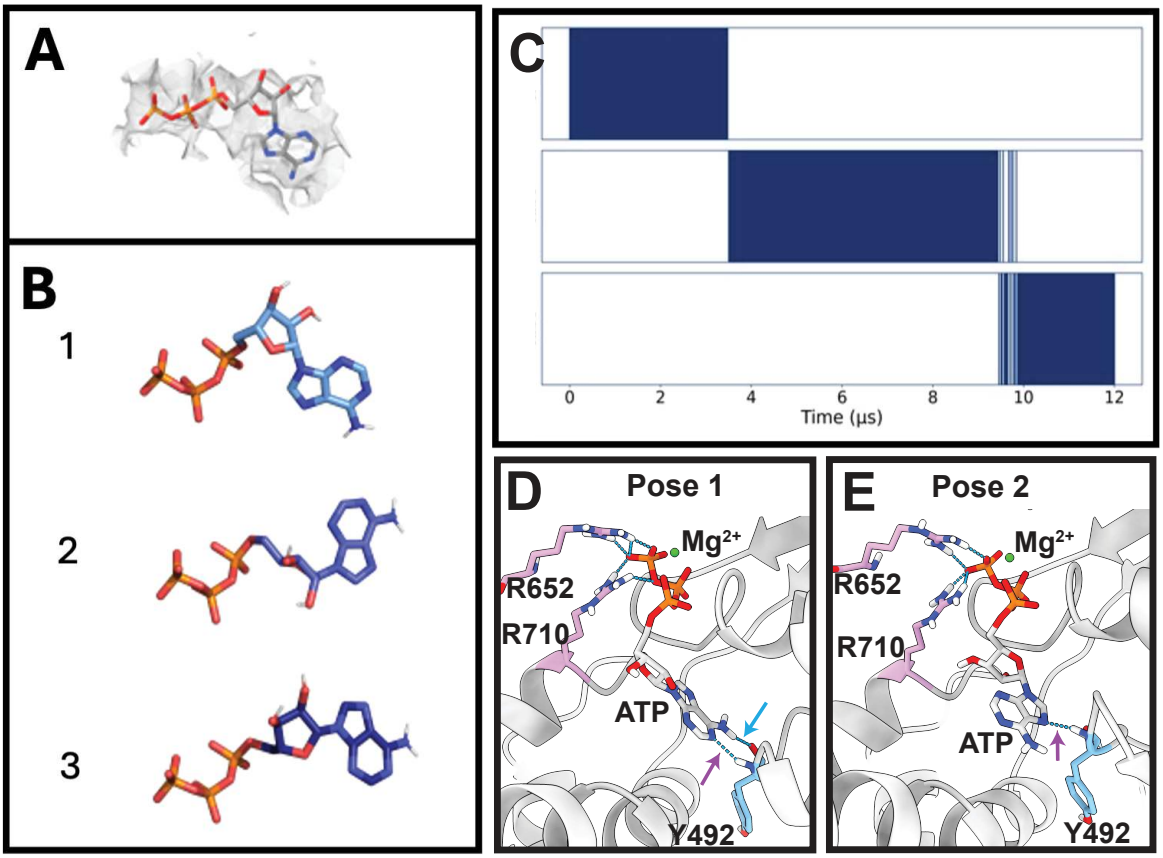


Figure 2.6: Multiple poses of bound nucleotide. A) Representative cryo-EM density (mesh) of nucleotide (ATP shown as gray sticks) from PDB:7KRZ. **B)** Representative structures obtained from clustering of nucleotide positions in a subunit of LONP1^{WT} during a microsecond-timescale simulation, revealing three distinct binding poses of ATP. **C)** RMSD of the ATP positions at each timepoint from the cluster centroid reflecting the time-dependent presence of clusters one (top), two (middle), and three (bottom) during the simulation. **D,E)** The nucleotide binding pocket conformation corresponding to cluster one (D) and cluster two (E). D) Pose one is stabilized by interactions between the backbone atoms of Y492 (blue sticks) and the adenine group of the ATP molecule (white sticks), forming hydrogen bonds (blue dashed lines) to the amino and imidazole groups (blue and purple arrows, respectively). Additionally, the phosphate tail is coordinated by a Mg^{2+} ion (green) and arginine residues of the same chain (R710, pink) and neighboring chain (R652, pink). E) Pose two coincides with the loss of one of these stabilizing hydrogen bonds, allowing the adenine ring to flip ∼180° into an upward conformation that remains stabilized by the hydrogen bond between Y492 and the imidazole group of ATP (purple arrow). Meanwhile, the phosphate tail remains rigidly coordinated by the arginine residues (pink) and Mg^{2+} ion (green).

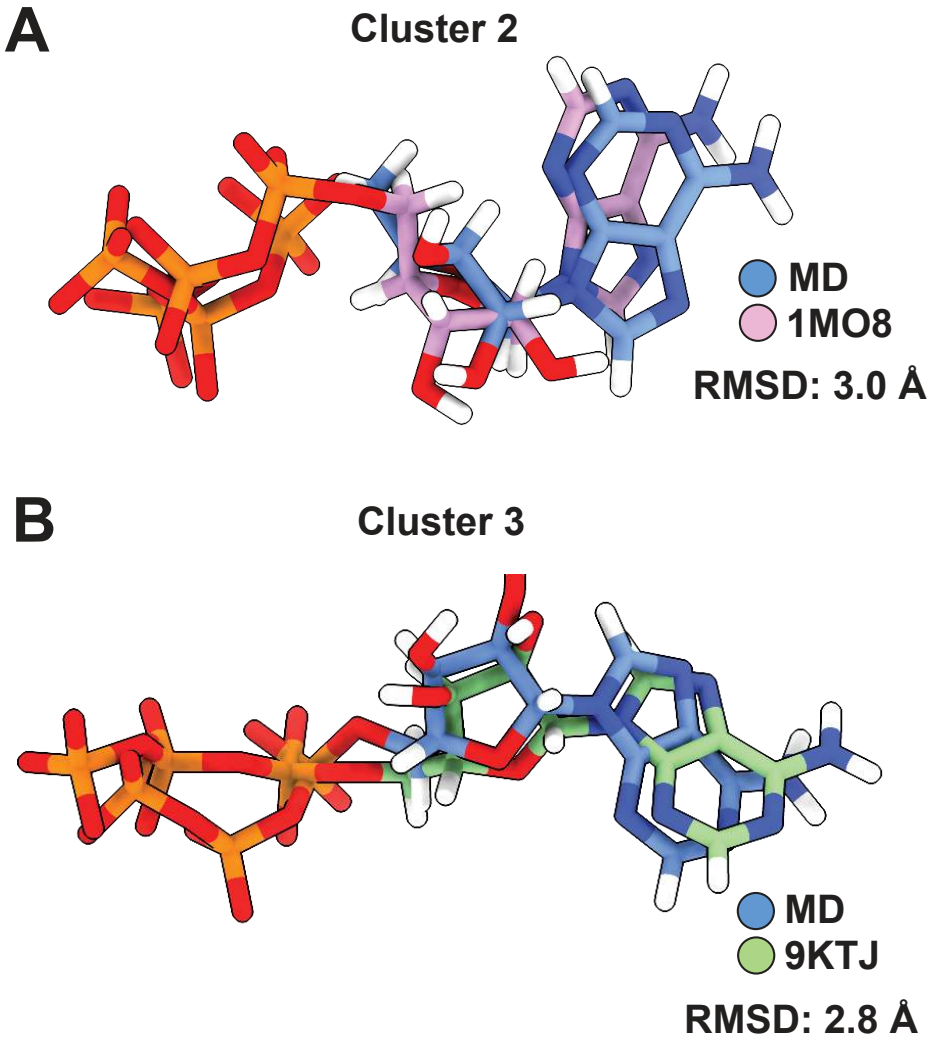


Figure 2.7: Comparison of ATP poses from MD to experimental data from the PDB. A PDB query was performed to identify nucleotide conformations similar to those observed in MD simulations. The nucleotide binding poses presented in Fig 2.6B corresponding to cluster 2 **(A)** and cluster 3 **(B)** are shown with their closest match from the PDB. Cluster 1 is not shown as this pose is present from the start of the trajectory, reflecting its similarity to the pose modeled in the cryo-EM structure. **A)** The pose represented by cluster 2 (blue sticks) matches most closely (RMSD of 3.0 Å) to an ATP binding pose (pink sticks) observed in PDB:1MO8, corresponding to an NMR ensemble of the ATP-binding domain of rat α1 Na,K-ATPase. **B)** Cluster 3 corresponds most closely (RMSD 2.8 Å) to an ATP conformation (green sticks) observed in a NMR ensemble of an ATP-binding DNA aptamer (PDB:9KTJ).

the dihedrals of residues H491, Y492, and Y673 in the subunit in which the nucleotide flip occurs sample a dihedral space distinct from those in the other subunits of the hexamer (Fig 2.8D). Notably, the $\chi 1$ angles of Y492 in this subunit exceed values of 240°, while in other chains, these values cluster around 180°. Y673 of this subunit also samples a wide range of dihedral space compared to others, forming three additional clusters that are distinct from the others. Interestingly, this residue exhibits a conformation that positions the tyrosine ring facing inward within the nucleotide binding pocket, and resulting in π-stacking interactions with the ATP molecule upon reaching the third binding pose. The local conformational flexibility within the nucleotide binding pocket, driven by these aromatic residues, leads to a $\sim$10° rotation of the large subdomain relative to the small subdomain over the course of the 12 μs simulation (Fig 2.9). These findings demonstrate that aromatic residues dynamically regulate nucleotide conformation while coupling local binding dynamics to larger-scale AAA+ domain movements.

2.3.2 Pore loop engagement with substrate is not strictly determined by nucleotide state

Nucleotide binding states have commonly been proposed to dictate substrate engagement and translocation within the ATPase domain of AAA+ proteases [3, 4, 89]. In this model, aromatic residues on the pore loops of ATP-bound subunits intercalate into the backbone of the substrate bound in the central pore of the ATPase domain, while the pore loops of ADP-bound or apo subunits remain unengaged [3, 4, 83]. Long-time scale simulations, however, conveyed a more dynamic nature of the pore loops than has been depicted in cryo-EM structures (Fig 2.10). Multiple events of ADP-bound subunit pore loop interactions with the substrate are present across all three of the simulated models. In LONP1$^{\mathrm{WT}}$, the aromatic residue of pore loop 1 (Y565) on the ADP-bound subunits is measured within 5–8 Å of the substrate during portions of the trajectory in at least two of the four replicates (Fig 2.10C). Full analyses across models and replicates are shown in (Fig A.2). Many of these contacts are stable for multiple microseconds of the simulation, though on the whole, the ADP-bound subunits exhibit increased dynamics, evident by the increased variability in the measured distances. Y565 of the ATP-bound subunits demonstrates more rigid contacts, persisting at distances of $\sim$5 Å from the substrate for the majority of the simulation. Of the ATP-bound subunits, the seam subunit (ATP1) is the most dynamic, though the pore loop residue remains within 8 Å of the substrate throughout the trajectories.

Interestingly, in subunits where ATP has been exchanged for ADP, including the fourth subunit of LONP1$^{\mathrm{Post}}$ and the second, third, and fourth position of LONP1$^{\mathrm{ADP}}$, the behavior of the pore loops remains largely unchanged (Fig A.2). This trend also holds for subunits where the nucleotide has been removed entirely, including in the seam subunit (position one) and the sixth subunit of LONP1$^{\mathrm{ADP}}$. Thus, nucleotide exchange did not have an observable impact on pore loop engagement at the simulated timescales and rather engagement appeared dependent on the position of the corresponding subunit's ATPase domain within the staircase register.

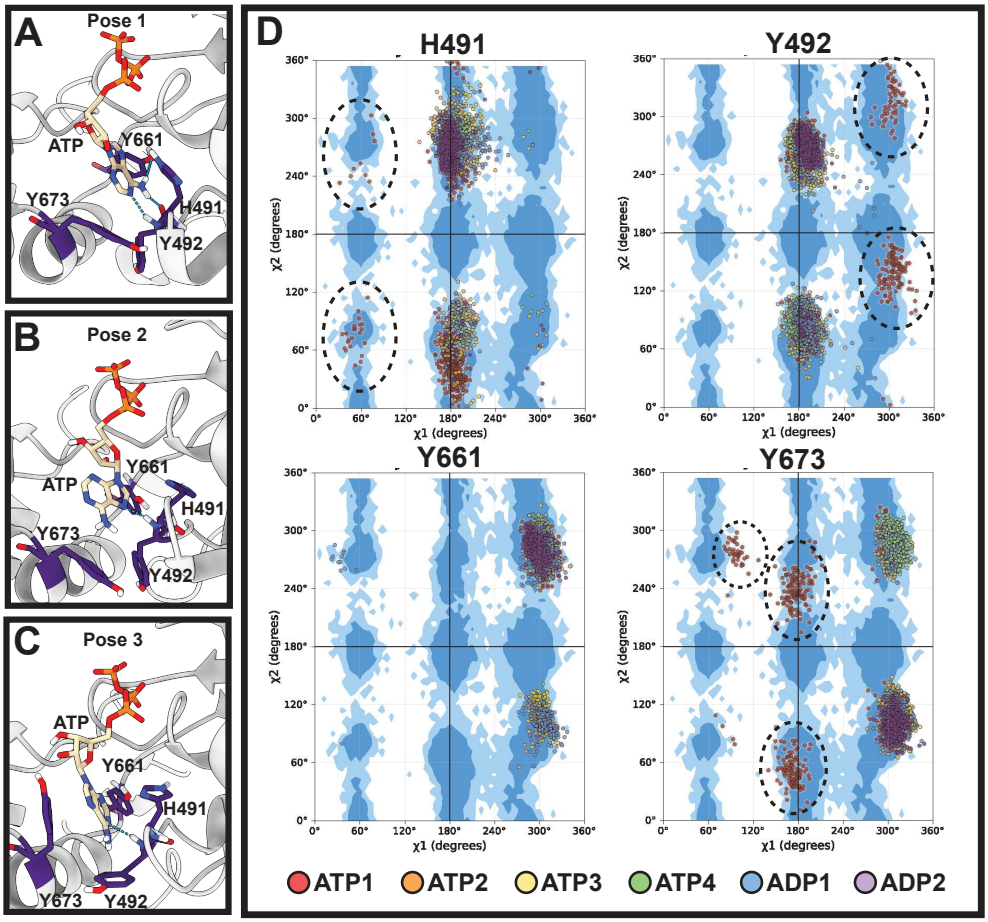


Figure 2.8: Aromatic residues in the nucleotide binding pocket modulate ATP binding poses. (A-C) Local dynamics associated with distinct nucleotide binding poses. A) Pose 1. The initial ATP binding pose (wheat sticks) is stabilized by three hydrogen bonds (blue dashed lines) between the backbone atoms of protein side chains of the large ATPase subdomain (Y492 and Y661, indigo sticks) and the adenine ring of ATP. The imidazole of H491 (indigo sticks) also appears to interact in a putative π-stacking interaction with the adenine ring. B) Pose 2. Upon the loss of hydrogen bonds to the amino group of the adenine ring, the nucleotide undergoes a $\sim$180° rotation to an upward conformation, stabilized by the hydrogen bond that persists between Y492 and the nitrogen of the imidazole group within the adenine ring. C) Pose 3. The adenine ring undergoes another rotation and engages in π-stacking interactions with Y673 (indigo sticks) of the small ATPase subdomain, following the transition of its phenol ring from an outward-facing conformation to a inward-facing conformation. D) The side chain dihedral angles (χ1 and χ2) of H491 (top left), Y492 (top right), Y661 (bottom left) and Y673 (bottom right) for each timestep in the 12 μs trajectory are shown overlaid on Janin plots. The dihedrals of the aromatic residues are plotted as circular markers with colors corresponding to their respective subunit in the hexamer (ATP1 subunit corresponding to the poses in (A-C) in red, ATP2 in orange, ATP3 in yellow, ATP4 in green, ADP1 in blue, and ADP2 in purple). Dashed ovals on the plots indicate clusters of the side chain dihedrals corresponding to the subunit shown in (A-C) that reflect a distinct side chain dihedral space not sampled by aromatic residues of other subunits.

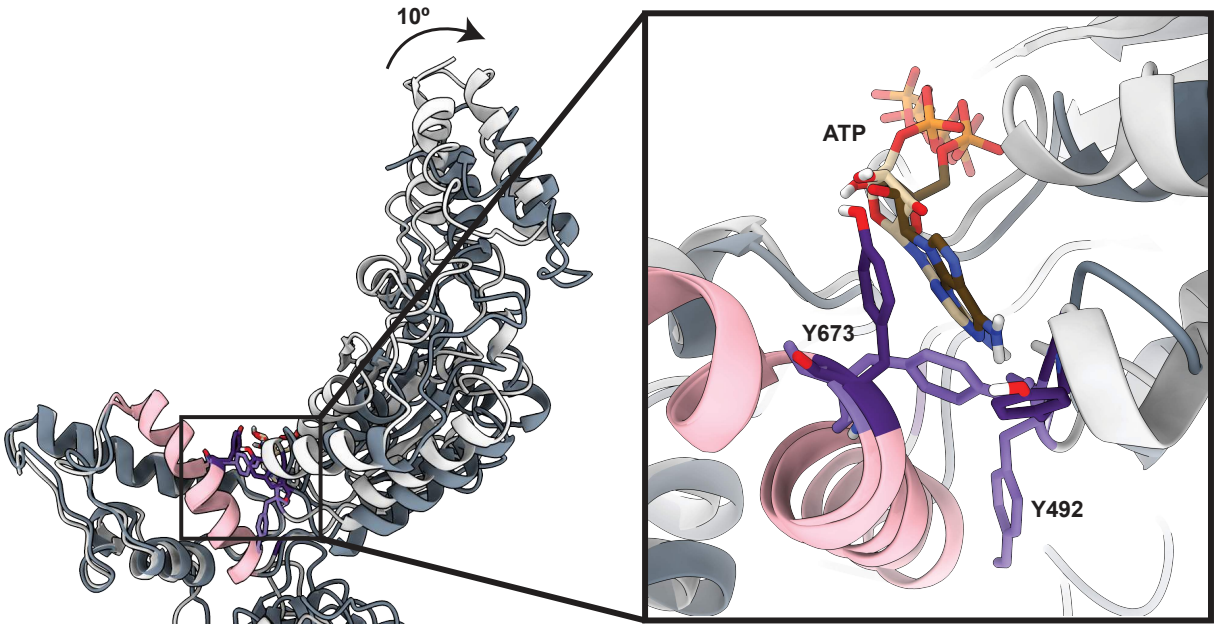


Figure 2.9: Change in intra-subdomain ATPase angle associated with nucleotide-induced conformational changes. Overlay of cartoon representations of the ATPase domain corresponding to the nucleotide analysis presented in Fig 2.6B, with the initial state (nucleotide pose 1, time = 0 μs) colored in white and the state observed after 10 μs in gray (nucleotide pose 3). Alignment of the ATPase domains was performed based on the position of the pink helix in order to determine the degree of rotation between the small and large subdomains. A $\sim$10° expansion of this angle is observed between these two states. The inset shows the corresponding nucleotide states (pose 1 in light tan and pose 3 in brown) and the conformational flexibility of the aromatic residues (Y673 and Y492 in light purple for the initial state and dark purple for the 10 μs state) within the nucleotide binding pockets. Both residues reposition their aromatic rings within the nucleotide binding pocket in the later stage of the simulation, contributing to the expansion of the binding pocket and the widening of the intra-subdomain angle.

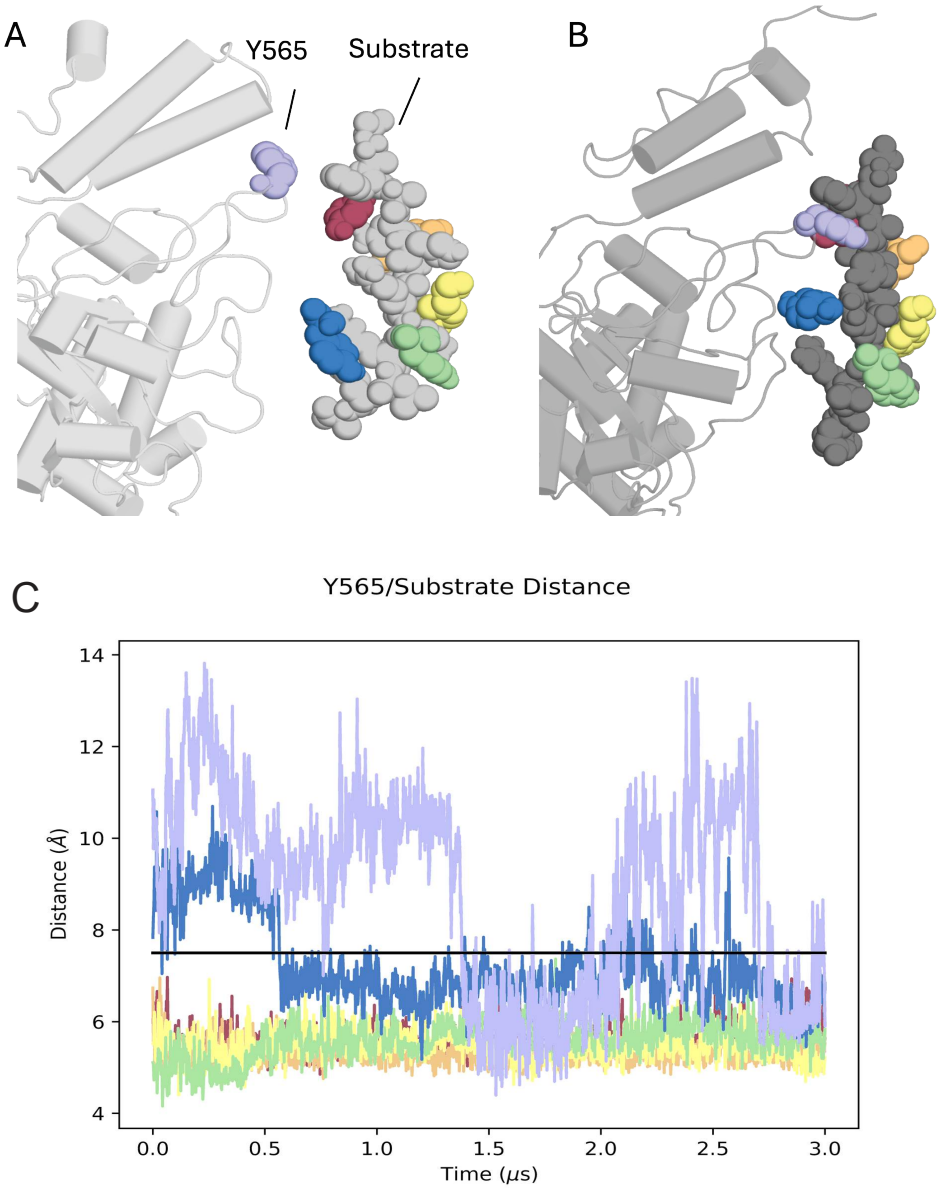


Figure 2.10: Pore loop engagement with substrate is not strictly nucleotide dependent. A,B) Structural representations of interactions between Y565 (colored spheres) of pore loop 1 and the substrate (gray spheres) with one ATPase domain shown (gray cartoon) for orientation. Y565 is colored according to the corresponding subunit (ATP1 in red, ATP2 in orange, ATP3 in yellow, ATP4 in green, ADP1 in blue, and ADP2 in purple). The initial poses obtained from the starting structure is shown in **(A)** and a snapshot from the first 3 μs of simulation is shown in **(B)** with Y565 of both ADP-bound subunits (blue and purple) coming within close contact of the translocating substrate. **C)** Distance between Y565 and substrate during the trajectory for each subunit, with the same coloring scheme as described above. A solid black line at 7.5 Å serves as a threshold for interaction distances.

2.3.3 Dynamics of substrate release

In order to understand the dynamics of the substrate once it reaches the protease domain, the isolated protease domains of LONP1 bound to substrate were modeled ($LONP1^{Protease}$) for simulation (Fig 2.11). A previous cryo-EM study [100] demonstrated the presence of substrate density within the proteolytic binding site of LONP1, without the use of inhibitors such as bortezomib to induce the activated state (PDB:7P0M). The group found that the signal for substrate was five times lower than than the surrounding protease, suggesting that only one of the six catalytic sites is occupied at a time in each enzyme. To determine the stability of the substrate within the binding groove, we conducted six 500 ns simulations of the six proteolytic subunits with tetra-alanine modeled in the substrate position (Fig 2.11A). While the substrate remained bound for the duration of the simulation in the majority of the subunits, a number of release events were also observed (Fig 2.11B). Across the six subunits and six replicates, twelve release events were identified (Fig A.3,A.4). These release events are characterized by the substrate dissociating from the proteolytic binding groove, with the distance between the substrate and the binding site increasing dramatically as the peptide exits the active site (Fig 2.11D). Concomitantly, the native contacts between the substrate and the proteolytic groove are progressively lost until complete dissociation occurs (Fig 2.11E).

The subunits with the highest propensity for substrate release were the central subunits at positions three, four, and six, exhibiting release events in half or more of the replicates (Fig 2.11F). The release events occurred fastest in subunit three, which had an average time-to-release of ∼200 ns. Zero release events occurred in subunit two and five, with the substrate bound to subunit five demonstrating the least dynamics and maintaining a distance of ∼5 Å for the majority of the simulation time. In proteolytic sites where substrate release occurred, a coordinated structural rearrangement of a β-hairpin positioned above the proteolytic groove undergoes a downward movement as the substrate exits the binding site. This conformational change brings a conserved tryptophan residue (W770) on the hairpin into direct contact with the binding groove, where it sterically occludes the active site following substrate departure (Fig 2.11B). Mutations of this residue have suggested it plays a role in peptidase activity and may be involved in substrate binding [105]. The repositioning of the hairpin creates a hydrophobic bridge that spans the substrate binding groove, blocking access to the polar patch where substrate binds (Fig 2.11C). This structural reorganization suggests a role of the hairpin and tryptophan residue in gating access to the substrate binding groove, potentially preventing nonproductive binding events or protecting the active site from solvent exposure until the next substrate arrives.

2.3.4 Inactivation of LONP1 proteolytic sites occurs in the absence of substrate on the microsecond timescale

The microsecond-timescale simulations using the 7KRZ-derived models ($LONP1^{WT}$, $LONP1^{Post}$, and $LONP1^{ADP}$) offered insight into the longer timescale dynamics of the proteolytic sites following substrate release (Fig 2.12). Through the removal of bortezomib from each of the six active proteolytic sites, these models reflect an enzymatic state immedi-

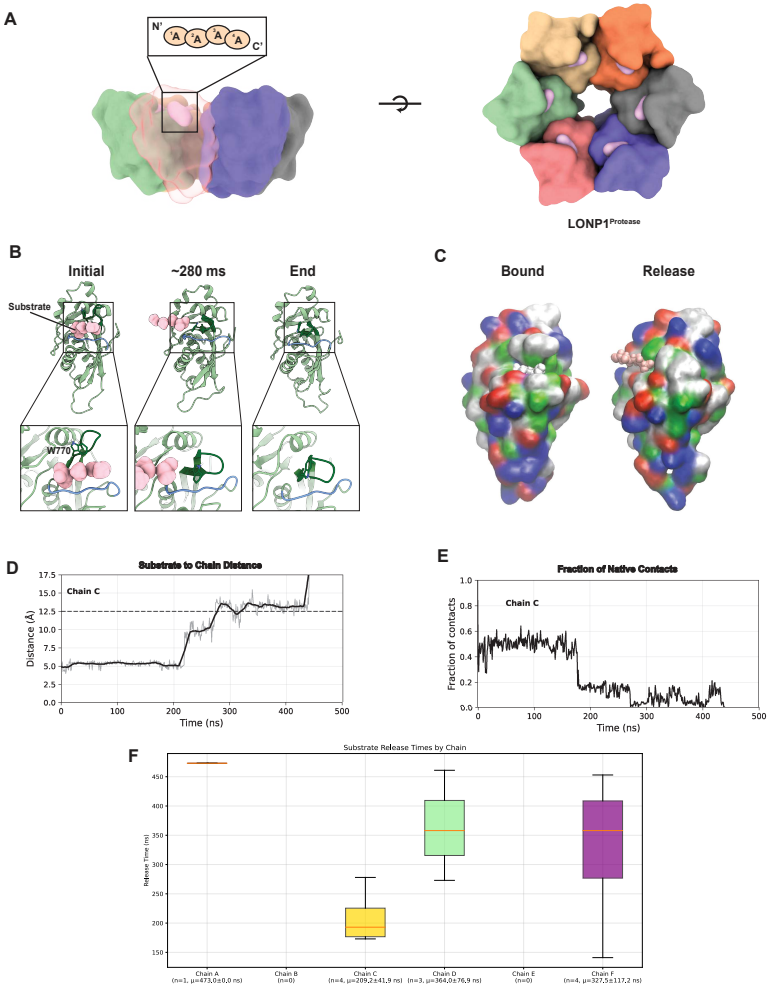


Figure 2.11: Substrate release dynamics. **A)** Side (left) and top (right) views of the ($LONP1^{Protease}$) model. Pink surfaces represent the tetra-alanine substrates bound to each of the six proteolytic sites, with the sequence depicted in the inset. **B)** Initial (left), intermediate (middle) and final (right) states of a subunit undergoing substrate (pink spheres) release. A key β-hairpin motif is colored in dark green, containing a tryptophan (W770, dark green sticks) that positions itself centrally above the proteolytic binding groove (light blue) in the later stages of substrate release. **C)** Electrostatic surface representations of a subunit with substrate bound (white spheres, left) and released (pink spheres, right). Substrate is positioned within a polar patch (green) in the bound state, that becomes occluded by the formation of a hydrophobic patch (white) following substrate release. **D,E)** Distance and fraction of native contacts (native contacts referring to those present in the initial trajectory frame) plots between the substrate and the proteolytic binding groove of the chain shown in **(B)**. Raw (light gray) and smoothed (black) data are shown. A dashed line indicates distances of 12.5 Å **(D)**, above which is considered a release event. The release event is also measured when the fraction of native contacts reaches zero **(E)**. **F)** Boxplot summary of the time to a release event in each subunit across all six 500 ns replicates. Chains B and E did not undergo any release events. The number (n) of release events as well as the mean time (μ) and standard deviation to a release event are shown below the chain labels on the x-axis. The mean times are also represented as a red line within each box.

ately following inhibitor or substrate release, and prior to the large conformational transition from the active, open form of LONP1, to the inactive, closed form (Fig 2.12A,B).

Intriguingly, multiple transitions of the proteolytic sites from an active, extended loop conformation to an inactive 3_{10}-helical conformation were observed (Fig 2.12C,D). This event also coincided with the formation of a hydrogen bond between an aspartate (D852) positioned on the proteolytic loop and a catalytic lysine (K898), indicative of the chemically inactive state of the proteolytic site. The inactivation of the proteolytic sites occurred rapidly in several subunits, with some sites inactivating within one microsecond.

Further, the inactivation of the proteolytic sites occurs during simulation despite the planar, C_6-symmetric organization of the protease domains, suggesting that inactivation events first occur locally and precede macroscopic domain level changes of the enzyme that transition it between an "on" and "off" form. The simulated systems also exhibit a mixture of inactive and active proteolytic sites within the hexameric ring (Fig 2.12E). These mixed states have not been observed in previous structures of LONP1, including those reconstructed without symmetry imposed. These structures present all six proteolytic sites in a homogenous activity state, likely due to the dynamic nature of their formation [4, 83, 100].

Upon visualization of the trajectories, several distinct paths to formation of the inactive state of the proteolytic sites were observed, with helix and hydrogen bond formation occurring in interchangeable sequences. To further investigate the order of these events, as well as to determine the stability and the propensity of these helices to form at a given proteolytic site, we first established a method for measuring the presence of the inactive proteolytic site conformation in simulation. The presence of the auto-inhibitory hydrogen bond between K898 and D852 was insufficient as a sole metric to assess the formation of the inactive state, as several instances were observed in the absence of the formation of a 3_{10}-helix. Likewise, we observed several instances where the 3_{10}-helix partially formed, without the formation of the hydrogen bond. Thus, three metrics were analyzed to identify the fully inactive proteolytic state during the simulation: 1) the presence of the auto-inhibitory hydrogen bond between D852 and K898, 2) the deviation of the dihedrals of the catalytic serine (S855) and 3) an adjacent proline (P854) from the dihedral values of an ideal 3_{10}-helix, as defined by the Ramachandran plot. While several other residues of the proteolytic site showed correlation between dihedral changes and helix formation, P854 and S855 were chosen based on empirical evidence demonstrating their predictive power of both the time of formation and stability of the assumed secondary structure (Fig A.5).

Based on meta-analysis, the formation of the inactive state at the proteolytic sites does not indicate a strong dependence on the subunit position within the hexamer (Fig 2.12F). The full dataset describing these events can be seen in *Appendix A* (Fig A.6). Our analysis identified both full inactivation events (exhibiting both steric and chemical inhibition) and partial inactivation events (steric-only or chemical-only), with the reported number of events and residence times encompassing all types. Inactivation of proteolytic sites is observed in both ATP and ADP-bound subunits, as well as in subunits in the core of the hexameric ring or closer to the seam subunit. The seam subunit had the highest tendency for inactivation, with seven events and an average residence time of 4.9 μs. While the proteolytic site at position three underwent the highest number of inactivation events (12), many of these were

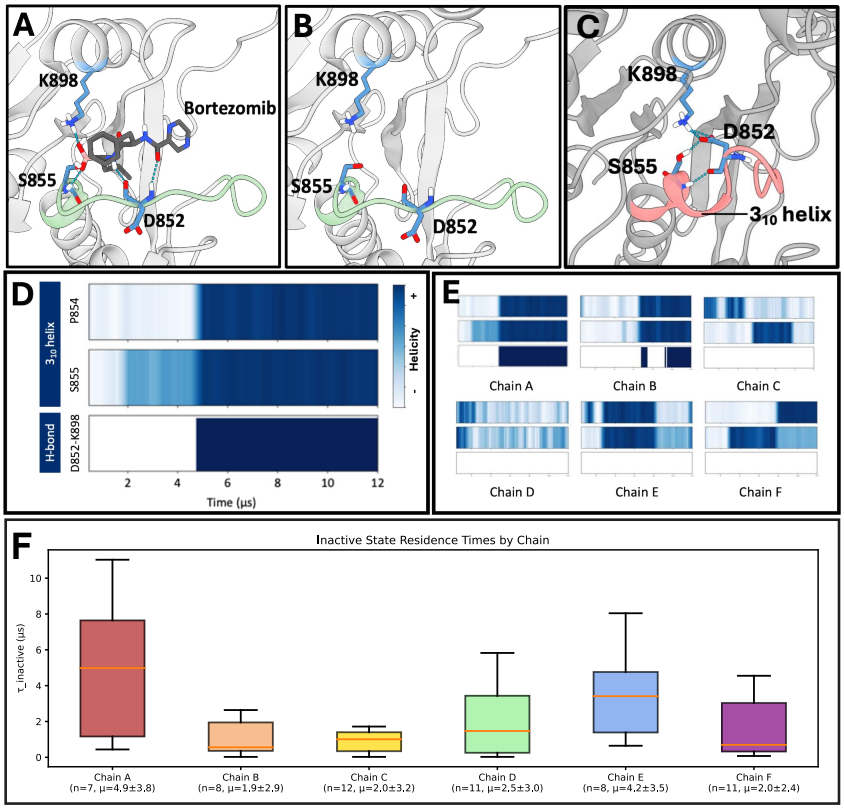


Figure 2.12: Inactivation of proteolytic sites upon substrate removal. A) The active conformation of the proteolytic site in the cryo-EM structure (PDB:7KRZ). The proteolytic loop (green cartoon) forms an elongated binding groove in the presence of a peptidomimetic inhibitor, bortezomib (gray sticks). Binding of bortezomib is coordinated through a covalent bond to the catalytic serine (S855, blue sticks), as well as hydrogen bonds (blue dashed lines) to the S855, the catalytic lysine (K898, blue sticks), and an central aspartate residue (D852, blue sticks) on the proteolytic loop. **B)** Representative proteolytic site in the initial starting state of the 7KRZ-derived models, ($LONP1^{WT}$,$LONP1^{Post}$, and $LONP1^{ADP}$), reflecting the removal of bortezomib from each of the six proteolytic sites. **C)** Final state (12 μ timepoint) of the proteolytic site represented analyzed in panel **(D)**. The formation of a 3_{10}-helix within the proteolytic loop (coral), as well as the formation of a hydrogen bond between D852 and the catalytic residues (K898 and S855) correspond to a fully inactive state, characterized by both steric occlusion of substrate binding and chemical inhibition of the catalytic dyad. **D)** A representative plot measuring the presence of the proteolytically inactive state. A 2D heatmap of the RMS of the dihedrals of P854 (top) and S855 (middle) from those of an ideal 3_{10}-helix are reflect the formation of the 3_{10}-helix (steric inactivation), with darker blues indicating increased helicity, and the presence of the chemically inhibiting hydrogen bond between D852 and K898 (bottom) reflects the presence of the chemically inactive state. The sites are considered fully inactive if all three metrics are "on" (dark blue). **E)** Representative per-subunit proteolytic inactivation analysis plots for a trajectory of 12 μ trajectory of $LONP1^{WT}$. **F)** Box plot of residence times of the inactive conformation of the proteolytic sites for each chain across all models and replicates, with mean residence times represented as a red line in each box. The number of inactivation events (n) and the mean residence times (μ in μs) along with the standard deviation are presented below each chain identifier on the x-axis.

categorized as partial inactivation events or occurred close to the end of the simulation, yielding an average residence time of 2.0 μs. Following formation of both the 3_{10}-helix and autoinhibitory D852-K898 hydrogen bond, the inactive conformation typically remain stable for the remainder of the simulation time, with few exceptions (Fig A.6). Despite the apparent stochastic nature and lack of dependence on subunit position and nucleotide state, our data reveal a tendency for inactivation events to cluster within groups of adjacent subunits (Fig A.6). This clustering pattern suggests that the inactivation of one proteolytic site may trigger a cascade of inactivation events in neighboring subunits, potentially serving as an early signal of substrate absence that precedes the large-scale conformational rearrangements that switch LONP1 to its inactive, open conformational state.

2.3.5 Collective motions revealed by PCA provide a framework for probing nucleotide-dependent allosteric coupling

To investigate dominant collective motions and reduce the dimensionality of the sampled conformational landscape, we performed Principal Component Analysis (PCA) on the backbone atom positions of each subunit (Fig. 2.13). The resulting distribution of structures in principal component space was partitioned into five groups using K-means clustering, which organizes similar conformations based on their proximity in the reduced space. For each group, the centroid was selected as the representative structure. The trajectory timepoints corresponding to these centroids were then identified and used to guide linear interpolation between successive representative structures, enabling visualization of the time-dependent transitions between key conformations. These PCA trajectories were further synchronized with the nucleotide cluster trajectories, allowing us to examine how the dominant motions of the subunits evolved in concert with the major nucleotide-binding poses observed over the course of the simulation.

The PCA analysis revealed several characteristic patterns of coordinated domain movements that provide insight into potential allosteric coupling mechanisms within individual subunits. In several instances, we observed a $\sim$5 Å downward displacement of the α0 helix in the AAA+ domain, followed by a $\sim$6 Å upward displacement of the protease domain, that coincided with a nucleotide flip (Fig A.6G). Additionally, the β-hairpin above the proteolytic site underwent a $\sim$4 Å upward displacement following inactivation of the proteolytic site, and shifted proximally to the linker connecting the two α-helical subdomains of the ATPase domain (Fig A.6G). The movement in this β-hairpin positions the closest residues (P524 of the linker and M773 of the β-hairpin) within $\sim$2.5 Å of one another, establishing a potential communication pathway between regions proximal to the nucleotide binding site and the proteolytic site. Although the order and magnitude of these events varied across subunits and replicates, PCA consistently highlighted recurrent motions of the ATPase and protease domains. Together, these analyses provide a framework for distinguishing direct causal relationships between nucleotide-binding states and the motions of the two functional domains.

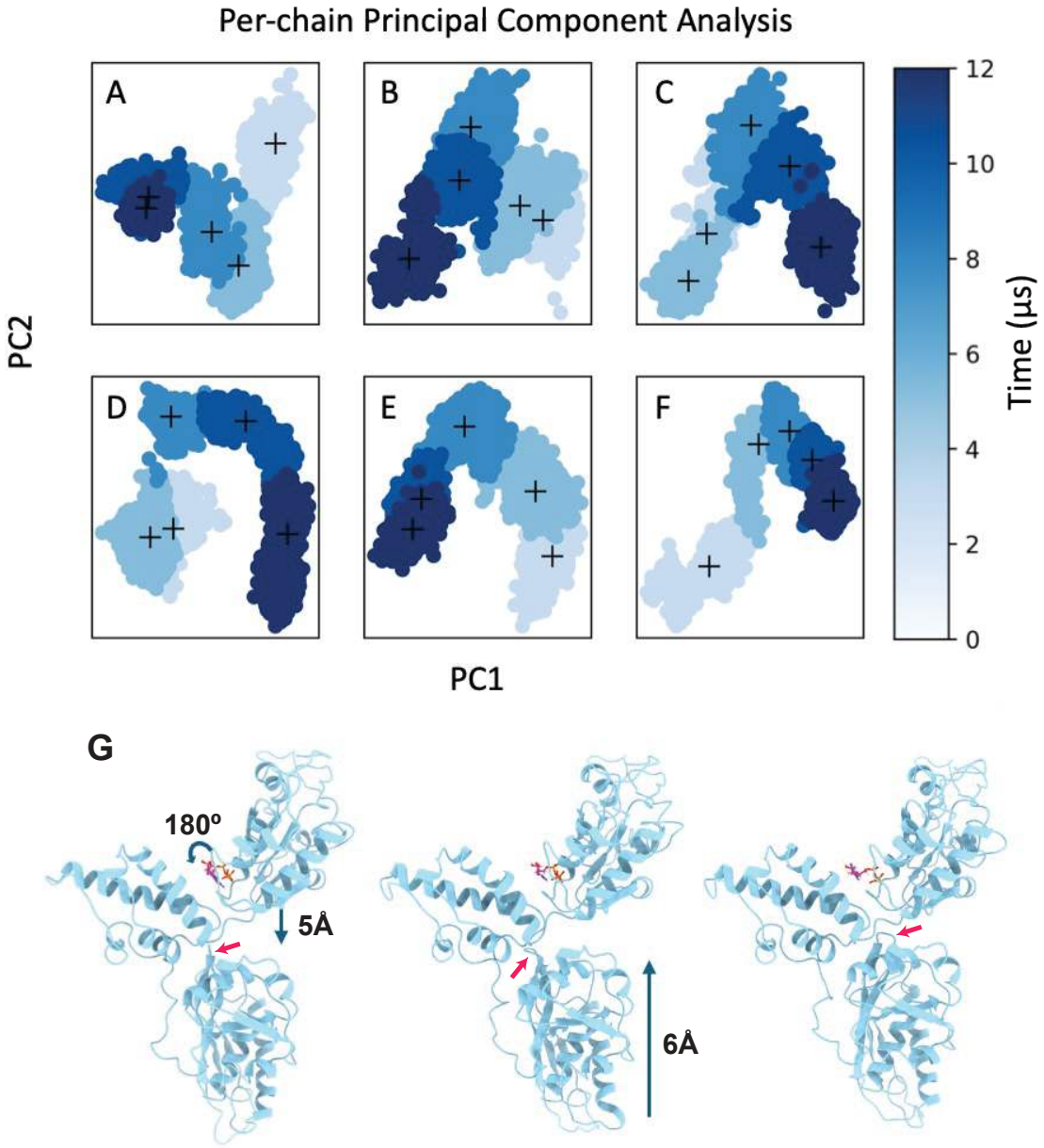


Figure 2.13: Per-chain principal component analysis reveals concerted rigid-body domain motions. A-F) Representative principal component analysis performed on the backbone atom positions of each subunit (chain A-F, respectively) at each timestep in the trajectory. The first and second principal components were plotted and coordinate positions were colored by timestep. Clustering was performed to classify the subunit coordinates into the top five conformations and the centroids of each cluster are marked with a black cross. **G)** Snapshots from a linear interpolation performed between the representative structure (centroid) of each cluster to determine the time-dependent nature of dominant subunit motions. A representative series of structural snapshots corresponding to chain A in **(A)** show large, rigid-body domain movements of the ATPase and protease domain coinciding with a nucleotide flip, positioning the two functional domains within close proximity of one another. A red arrow indicates the tryptophan β-hairpin.

2.4 Discussion

Our simulations provide a mechanistic framework for understanding how LONP1 transitions between states of activity. Details gained from MD help to inform on the temporal scale and coordination of key events involved in LONP1 function, across both the ATPase and protease functional domains.

Within the ATPase domain, we observed dynamic states of nucleotides corresponding to local conformational changes. Our simulations revealed multiple binding poses of both ATP and ADP, stable on the microsecond timescale. Several of these poses were characterized by a 180° rotation of the adenine ring, corresponding to a nucleotide "flip". In contrast to the flexibility observed in the adenine and ribose rings, phosphate tails remained relatively rigid, likely due to their coordination by the Mg^{2+} ions and arginine fingers. This rigid coordination may help to stabilize the nucleotides, priming them for hydrolysis, while the ribose and adenine rings are free to move and respond to or induce local conformational changes within the protein. Indeed, several conserved aromatic residues were observed interacting with the adenine ring of the nucleotides, sampling a variety of rotameric states in subunits where nucleotides displayed high flexibility. These changes also coincided with intra-subdomain motions of the ATPase domain, including a $\sim$10° rotation of the large subdomain relative to the small subdomain following changes in the nucleotide binding pose and conformations of the surrounding aromatic residues. Hinge motions within the ATPase domains have previously been postulated to play a role in shifting the register of subunits within the spiral staircase following hydrolysis in order to yield downward translocation of substrate [90, 106, 107]. Thus, these findings suggest that the nucleotide binding poses may play a role in local and domain-level conformational changes associated with the functional activity of the ATPase domain.

To further investigate the role of nucleotide state on substrate translocation, we monitored pore loop engagement with substrate across different nucleotide conditions. Surprisingly, we observed minimal changes to pore loop engagement when ATP was exchanged for ADP or removed entirely. Instead, engagement patterns showed a high degree of dependence on subunit position rather than nucleotide state, as previously proposed [3, 4, 90]. Pore loops at positions one to four remained relatively rigid in their substrate interactions, staying within $\sim$8 Å of the substrate, while pore loops at positions five and six (seam and neighboring subunit) showed high flexibility with distances ranging from $\sim$4.5 to 17 Å. Despite this flexibility, numerous close contacts were observed, indicating that "disengaged" pore loops may assist in guiding substrate through the central pore during translocation.

To assess conformational changes within the protease domain, we used a protease-only model of LONP1 bound to tetra-alanine ($LONP1^{Protease}$) and observed multiple substrate release events. While this model likely exhibits higher flexibility than the full enzyme and the tetra-alanine substrate may release more readily than bulkier substrates, it provides useful information on local conformational changes associated with release. Upon substrate release, we observed the downward motion of a β-hairpin above the proteolytic groove, which positioned a conserved tryptophan residue into the binding site and led to the formation of a hydrophobic patch that blocks groove access. This rearrangement may serve as an initial

gating mechanism prior to the formation of a more stable 3_{10}-helical conformation of the proteolytic loop when substrate abundance is low or after the enzyme has completed substrate processing.

Our long-timescale simulations revealed that proteolytic sites display considerable conformational flexibility on the microsecond timescale, sampling dynamic states before adopting a 3_{10}-helical conformation. Helix formation coincides with establishment of a hydrogen bond between D852 and the catalytic lysine K898, chemically inactivating the catalytic dyad. We observed partial inactivation events, characterized by helix formation without hydrogen bonding ("steric inactivation") or by hydrogen bonding without helix formation ("chemical inactivation"), as well as full inactivation events where both features were present. Chemical inactivation events were short-lived, while steric events persisted for microseconds and frequently preceded hydrogen bond formation. However, steric inactivation was more prone to reversal without hydrogen bonding, suggesting the bond stabilizes the helical conformation. Once both features were established, the inactive state persisted for the remainder of the trajectory, indicating this state represents a low-energy conformation with high propensity to form in the absence of substrate. Further, spontaneous transitions to the inactive state of the proteolytic site were observed, suggesting that this conformation is energetically favorable and can form in the absence of substrate. However, future work to identify a collective variable describing this transition will be necessary to characterize the free energy surface of the inactivation event.

Importantly, inactivation events were observed in the context of protease domains that maintained a C_6-symmetric, planar arrangement, characteristic of a macroscopic active state of LONP1 [4], indicating that global domain rearrangements are not strictly required for this transition to occur. This suggests that inactivation is a local event within each proteolytic site and precedes the large-scale rigid-body movements that distinguish the open and closed states of the enzyme and are observed on the second-minute timescales sampled with cryo-EM. The stochastic and asynchronous appearance of inactive sites across different subunits within the enzyme further highlights that proteolytic inactivation is a property of independent subunit dynamics, rather than a concerted, globally-coordinated event.

Nevertheless, our data indicate that intra-subunit communication may bias the occurrence of inactivation events. Inactivation events were often observed in clusters of adjacent subunits, suggesting a mechanism in which spontaneous inactivation events can propagate through the ring, progressively reducing the number of catalytically competent sites. Previous studies identified several residues that appear to stabilize the extended loop conformation, so the loss of these interactions may trigger a cascade of inactivation events [4, 83]. Such a mechanism would provide a means of tuning enzymatic activity in response to substrate availability, and allowing the enzyme to downregulate proteolysis after a substrate has been fully processed. Taken together, these data point to a model in which proteolytic inactivation of LONP1 is initiated by spontaneous loop rearrangements following substrate release and initial gating by the tryptophan β-hairpin, and is subsequently stabilized through the formation of a 3_{10}-helix and autoinhibitory hydrogen bond. In this framework, inactivation does not require large-scale conformational switching of the protease domains, but rather emerges from local fluctuations that feed into an allosteric regulatory network between subunits.

The coordinated domain movements revealed by PCA analysis provide additional evidence for potential allosteric coupling between the ATPase and protease domains. The observed correlation between nucleotide flips and rigid-body domain motions, particularly the positioning of the β-hairpin near the nucleotide-binding region, suggests candidate pathways for inter-domain communication that could link nucleotide state changes to proteolytic regulation.

In sum, these simulations reveal insights into dynamics of LONP1 that are currently invisible to physical methods due to their short temporal scales and transient nature. While the precise allosteric networks connecting the ATPase and protease domains require further investigation, our findings and PCA approach establishes a framework for systematically identifying coordinated motions. Statistical analysis of these observations across different subunit positions and nucleotide states will be essential to distinguish functionally relevant allosteric pathways from coincidental motions and fully elucidate the regulatory mechanisms governing LONP1 activity.

2.5 Methods

2.5.1 Long-timescale simulations of LONP1

2.5.1.1 Model

Our starting model was obtained using the cryo-EM structure corresponding to the active state of the hexameric human mitochondrial protease LONP1 (PDB:7KRZ) [4]. The system consisted of six protein chains (A-F) corresponding to the subunits of LONP1 (res 415–947) which include the NTD^{3H}, ATPase domain, and protease domain. A seventh protein chain corresponds to the 12-residue substrate modeled in the central pore of the ATPase domain. The side chains of the substrate could not be resolved by cryo-EM and were therefore modeled as a polyalanine peptide in the PDB entry. In our initial model for molecular dynamics simulations, we replaced the polyalanine 12-mer with the amino acid sequence of a known endogenous substrate, StAR. The sequence that we chose was based on work by Suzuki et al., in which mass-spectrometry was used to identify cleavage products of Lon-mediated degradation of StAR [84]. We modeled our substrate threaded through the central pore of the ATPase domain in the N' to C' direction. We additionally removed atoms corresponding to the bortezomib inhibitor bound at each of the six proteolytic sites in the 7KRZ PDB entry.

Addition of missing residues and model validation
Missing residues were added using Modeller v9.25 [108]. Twenty models were generated, and the ten with the best DOPE scores [109] were further evaluated using the SWISS-MODEL server [110]. The model with the best QMEAN [111] score was selected for loop refinement, in which loops containing the missing residues plus three adjacent residues on each side were refined. Following loop refinement, the model generation and evaluation procedure was repeated. The model with the highest QMEAN score was selected for MD.

Model modifications
Following the addition of the missing residues, three models of LONP1 were built for MD

Table 2.1: Summary of molecular dynamics simulations.

	$LONP1^{WT}$	$LONP1^{Post}$	$LONP1^{ADP}$	$LONP1^{Protease}$
System				
PDB Relative	7KRZ	7KRZ	7KRZ	7P0M
Functional Domains	All	All	All	Protease
Ligands	4 ATP, 2 ADP	3 ATP, 3 ADP	4 ADP	None
Substrate	ATPase-bound 12-mer (StAR)			Protease-bound tetra-alanine
System Size (atoms)	∼400k			∼180k
Replicates & Time (n × μs)	3×12, 1×6	3×12, 1×6	1×12, 2×6	6×0.5
Total Simulation Time (μs)	42	42	24	3
System Setup				
Force Field	Amber ff99SB-ILDN			
Water Model	TIP3P			
Ions / Salt (mM)	NaCl, 150			
Box Type / Size (Å)	Cubic, 120	Cubic, 120	Cubic, 120	Triclinic, 120
Equilibration				
Energy Minimization	10,000 SD + 10,000 CG; tol 100 kJ/mol/nm; 2 fs timestep			
NVT (ns)	1	1	1	0.5
NPT (ns)	10	10	10	5
Timestep (fs)	4	4	4	2
Restraints	Heavy atoms, all bond			Heavy atoms, H-bonds
Thermostat / T (K)	Velocity-rescaling, 300 K			
Barostat / P (bar)	Parrinello–Rahman, 1 bar			
HMR	yes	yes	yes	no
Production				
Timestep (fs)	4	4	4	2
Temperature (K)	300			
Thermostat	Nosé–Hoover	Nosé–Hoover	Nosé–Hoover	Velocity-rescaling
Pressure (bar)	1			
Barostat	MTK	MTK	MTK	Parrinello–Rahman
HMR	yes	yes	yes	no

simulation. LONP1^{WT} contained nucleotide states of the hexamer corresponding to those natively observed through cryo-EM (four ATP molecules bound to chains A-D and two ADP molecules bound to chains E-F). LONP1^{Post} included an exchange of the ATP at subunit position four in the wild-type model with ADP through the removal of the gamma phosphate and magnesium ion. This state, therefore, consists of three ATP molecules (chains A-C), and three ADP molecules (chains D-F). LONP1^{ADP} contains four ADP molecules (chains B-E) with chains A and F representing apo binding pockets.

2.5.1.2 Ligand parameterization

ADP and ATP coordinates were extracted from the PDB models and hydrogen atoms were added in PyMOL 4.6.0 [112] with the command `h_add`. The ligands were then parametrized with the Carlson polyphosphate parameters developed for use with the AMBER force field [113]. The tLEaP program of AmberTools 14 was used to generate a topology file from the input PDB coordinates and parameters. The topology files were then converted to the GROMACS topology file format with ACPYPE v0.1.1 [114]. Hydrogen mass repartitioning [115] was performed on the ligands with the hydrogen mass defined as 4.032 using the HMassRepartition class of ParmEd tools v3.2.0.

2.5.1.3 Simulation setup

The models were prepared for simulation using GROMACS 2020 [116] and parametrized with the Amber ff99SB-ILDN force field [117] and the Carlson parameters for ATP/ADP [113]. Hydrogen-mass repartitioning was applied to the full system in GROMACS, increasing the hydrogen mass by a factor of 4. The full system containing ∼400 k atoms was placed in cubic box of side length 120 Å, solvated in TIP3P water [118], and neutralized with 150 mM NaCl using genion in GROMACS. 10,000 steps of steepest descent minimization was performed, followed by 10,000 steps of conjugate gradient descent with one step of steepest descent minimization every 1,000 steps. For all energy minimization procedures, a force tolerance of < 100.0 kJ/mol/nm and a minimization step size of 2 fs were used.

2.5.1.4 Equilibration

Equilibration was performed in GROMACS. 1 ns of NVT equilibration was performed with the temperature maintained at 300 K using the velocity-rescaling thermostat [119]. The protein, ligands, and magnesium ions were contained in one temperature coupling group, while the water, sodium, and chlorine ions were contained in another. Following NVT equilibration, 10 ns of NPT was performed, maintaining the pressure at 1 bar using the Parrinello-Rahman barostat [120]. All bonds were constrained using the LINCS algorithm [121] during equilibration procedures, and position restraints on all heavy atoms were enforced. A timestep of 4 fs was used. The Verlet cut-off scheme was applied for non-bonded interactions.

2.5.1.5 Production

The equilibrated system was converted from GROMACS to Desmond format using a custom python script provided by the Pittsburgh Supercomputing Center (PSC) community.

Microsecond timescale simulations were carried out using the Anton 2 supercomputer [122] at the Pittsburgh Research Center (PSC). A 4 fs timestep was used, and the pressure and temperature were maintained at 1 bar with the Martyna, Tuckerman and Klein (MTK) barostat [123] and 300 K with the Nosé-Hoover thermostat [124], respectively. In total, more than 100 μs of the LONP1 models were simulated. Each simulation condition was run in quadruplicate (R1-R4).

2.5.1.6 Analysis

Trajectory analysis was performed on systems free of water, sodium ions, and chlorine ions. These atoms were stripped from the trajectory using a custom Tcl script provided by the PSC. The trajectories were also processed in GROMACS to remove translational and rotational motion.

All trajectory analysis was conducted using MDAnalysis v2.4.2 [125, 126] with custom analysis Python (v3.11) scripts. Simulation trajectories were loaded and processed with a timestep factor of 0.024 μs, corresponding to the simulation output frequency. Total simulation times for each replicate ranged from 6–12 μs.

Nucleotide clustering
Nucleotide binding poses were clustered for each chain of LONP1 using the Affinity Propagation clustering method of the MDAnalysis Encore clustering module over the 12 μs trajectory with a preference parameter of -40 (controlling cluster granularity) and a trajectory stepping factor of 10. The clustering was based on RMSD calculations between conformational frames to identify structurally similar binding poses. The analysis was conducted separately on each chain and nucleotide type (ATP/ADP), focusing on all nucleotide atoms (resname ATP ADP) within each binding site.

For each cluster, the ID, size (number of trajectory frames), and centroid frame (representative structure) were computed. Cluster occupancy was visualized as timelines, with binary bar plots indicating the presence or absence of each cluster conformation over the course of the simulation. To generate smooth transition trajectories, atomic coordinates were flattened into one-dimensional arrays, linearly interpolated between consecutive centroids, and reshaped into three-dimensional coordinates for visualization of conformational changes between binding poses.

Pore loop engagement analysis
Pore loop engagement analysis was performed using MDAnalysis. Distance analysis quantified the minimum distance between Cα atoms of the substrate modeled in the ATPase domain and the Cγ atom of residue 565 in each chain, using an 8 Å reference threshold to indicate close engagement.

Method of determining presence of inactivated proteolytic sites in simulation
The presence of inactivation events in the proteolytic sites (residues 854–857) was analyzed using dihedral angle analysis combined with hydrogen bond detection.

Dihedral angle analysis
Dihedral angle analysis was performed using MDAnalysis to calculate ϕ and ψ backbone dihedral angles for two residues, P854 and S855, within the proteolytic loop. These residues were selected after analyzing the RMS of the dihedral angles from ideal 3_{10}-helix values of several residues on the proteolytic binding groove (residue 853–857) in order to determine their correlation with the presence of the 3_{10}-helix (Fig). While several of the residues showed a significant reduction in RMS during the trajectory, P854 and S855 and had the strongest temporal correlation with the presence of the helix based on empirical evidence from observation of the trajectories by eye. In addition, the RMS of these two residues remains stable for the duration of the presence of the helix.

For each residue and each timepoint, the deviation of the dihedral angles of P854 and S855 from ideal 3_{10}-helix geometry as defined by the Ramachandran plot was calculated using reference values ($\phi = -71°$, $\psi = -18°$) from Barlow and Thornton [127]. The analysis employed a step factor of 10. The root-mean-square deviation (RMSD) of dihedral angles from ideal values was computed as:

$$\mathrm{RMSD}_{\mathrm{helix}} = \sqrt{(\phi - \phi_{\mathrm{ideal}})^2 + (\psi - \psi_{\mathrm{ideal}})^2}$$

The tolerance level was then defined at RMS values lower than 35° to reflect when P854 and S855 had attained a 3_{10}-helical conformation. The results were then translated to a binary measurement (0 for absence of the 3_{10}-helical dihedrals and 1 for presence). Smoothing was applied using a moving average with a window size of 20 frames.

Hydrogen bond analysis of proteolytic inactivation
Critical hydrogen bonding interactions were monitored between the catalytic lysine 898 (HZ* atoms) and the inhibitory aspartate 852 (OD* atoms) using the MDAnalysis HydrogenBondAnalysis module. Geometric criteria for hydrogen bond formation included donor-acceptor distance and angle cutoffs following standard conventions. The analysis was performed with the same step factor (10) as the dihedral analysis. Hydrogen bond counts were calculated as the total number of qualifying bonds at each timepoint. Smoothing was performed using a 20-frame moving average and the data were subsequently plotted as 2D heatmaps.

Statistical analysis of proteolytic inactivation events Proteolytic inactivation events were analyzed using a multi-signal approach combining helix formation and hydrogen bond analysis. Inactivation states were defined using binary thresholding: helix formation (RMSD $< 35°$) and hydrogen bond presence (> 0.2 bonds/frame). The protease was considered inactive when ≥ 2 of the 3 monitored signals (hydrogen bond formation, P854 dihedrals, S855 dihedrals) were simultaneously active. State transitions were detected using first-difference analysis, and residence times (τ_{active}, τ_{inactive}) were calculated as intervals between transitions, including trajectory endpoints.

PCA Calculations
PCA was performed on the full 6–12 μs trajectory for each chain in all three of the models using MDAnalysis. PCA was conducted on backbone atoms (Cα, C, N, O) of residues 468–947 for each chain separately, maintaining focus on the catalytic domains while excluding the

flexible N-terminal region. Trajectory frames were aligned to the reference structure using mass-weighted least-squares fitting of backbone atoms. The covariance matrix was calculated from mean-centered atomic positions, diagonalized to obtain principal components, and the top five components accounting for 95% of the variance were retained for analysis.

Trajectory projection onto the principal component space was performed using a step factor of 10 for visualization and clustering analysis. The top two principal components were plotted for all chains and the K-means clustering algorithm [128] implemented in Scikit-learn (version 1.2.0) [129] was used to generate five clusters of distinct chain conformations. For each cluster, centroids were determined by identifying the frame with minimum RMS from the mathematical cluster center in principal component space, then ordered chronologically by their time coordinates. Time-resolved PCA trajectories were generated through smooth linear interpolation of atomic positions between consecutive cluster centroids in Cartesian coordinate space, with interpolated frames determined by the temporal spacing between centroids. Time coordinates were calculated as $t_{\mu s} = \frac{\texttt{frame_index} \times \texttt{trajectory.dt}}{10^6}$, where dt is the simulation timestep in picoseconds.

2.5.2 Substrate Release Simulations

2.5.2.1 Model

Substrate release dynamics were investigated using a model based on PDB:7P0M ($\text{LONP1}^{\text{Protease}}$) [100]. Residue addition, refinement, and model validation were performed as described in the long-timescale MD methods section. Unknown residues modeled in the substrate positions were replaced with alanine residues using command line file editing. The alanine residue at position 898 was mutated to lysine (K898) in PyMOL to restore the wild-type catalytic residue that had been mutated for cryo-EM studies. The ATPase domain and NTD^{3H} were removed from the PDB file through command line file editing, leaving only the six protease domains as the starting model for MD.

2.5.2.2 Simulation setup

The $\text{LONP1}^{\text{Protease}}$ system was prepared for simulation using GROMACS 2020. The system containing approximately 18k atoms was solvated in a 120 Å triclinic box using TIP3P water and neutralized with 150 mM NaCl using genion. Energy minimization was performed as described in the long-timescale MD methods section.

2.5.2.3 Equilibration

The system underwent 0.5 ns of NVT equilibration with the temperature maintained at 300 K using the velocity-rescaling thermostat was used. 5 ns of NPT equilibration was then performed, maintaining the pressure at 1.0 bar using the Parrinello-Rahman barostat. Position restraints on all heavy atoms were enforced for both NVT and NPT and the Verlet cut-off scheme was used for non-bonded interactions. Hydrogen bonds were also constrained using the LINCS algorithm during equilibration and a 2 fs timestep was used.

2.5.2.4 Production

Production simulations were performed using GROMACS 2020 with a 2 fs timestep and hydrogen bonds constrained using the LINCS algorithm. Six independent replicates of 500 ns each were performed, providing a total of 3 μs of simulation time. The temperature was maintained at 300 K and the pressure at 1.0 bar using the velocity-rescaling thermostat and Parrinello-Rahman barostat, respectively.

2.5.2.5 Analysis

Substrate release events of LONP1$^{\text{protease}}$ were identified using MDAnalysis by monitoring the distance between a central residue of the substrate (residue 3) and a central residue of the proteolytic binding groove (residue 852). The fraction of native contacts between the substrate (residues 1–4) and binding groove residues (845–855) was analyzed in parallel, with release confirmed when the fraction of native contacts reached zero. For statistical analysis, release events were defined as the first instance in which the distance exceeded 12.5 Å, after smoothing of distance trajectories using a 20-frame moving average.

2.5.3 Analysis Software

All analyses were implemented using custom Python (v3.11) scripts built on the MDAnalysis framework (version v2.4.2) [125, 126]. Additional computational libraries included: NumPy (version 1.24.1) for numerical operations, SciPy (version 1.10.0) [130] for statistical analysis and interpolation, Scikit-learn (version 1.2.0) [129] for clustering algorithms, Matplotlib (version 3.6.3) and Seaborn (version 0.12.2) for data visualization, and Pandas (version 1.5.2) for data manipulation and analysis.

3

Mix-it-up: Accessible time-resolved cryo-EM on the millisecond timescale

3.1 Chapter overview

While the preceding chapter explored protein dynamics through computational approaches using long-timescale molecular dynamics simulations, this chapter shifts focus to experimental methodologies for capturing transient conformational states in real-time. The computational insights from LONP1 revealed the importance of microsecond timescale motions in regulation and substrate processing, highlighting a critical gap in our ability to experimentally observe such rapid structural transitions. Traditional cryo-EM sample preparation methods operate on timescales of seconds to minutes, making them unsuitable for capturing the fleeting intermediate states that govern protein function. To address this limitation, we developed "Mix-it-up" (MIU), an accessible and cost-effective platform for time-resolved cryo-EM that enables visualization of biochemical processes on the millisecond timescale. This work aims to further expand the experimental window into dynamic events that coordinate complex molecular machinery in a broadly accessible manner.

Adapted from: Alexandrescu, L., Lessin, W., & Lander, G. C. (2025). 'Mix-it-up': Accessible time-resolved cryo-EM on the millisecond timescale. *IUCrJ*, 12(6):710–724. [131].

Contributions The work was a collaborative effort between myself (Lauren Alexandrescu), William Lessin, and Gabriel C. Lander. LA and GCL conceptualized and led the project. LA and WL assembled the device. LA adapted the device for time-resolved studies, prepared samples for cryo-EM, and performed all cryo-EM structure determination and interpretation with guidance from GCL. LA prepared the figures and LA wrote the manuscript with guidance from GCL.

3.2 Introduction

Biological processes within and outside of cells involve dynamic, multi-component pathways, wherein macromolecules interact with cofactors and substrates while undergoing rapid conformational changes. Structural biology techniques such as crystallography and single-particle cryo-electron microscopy (cryo-EM) have provided high-resolution structural insights into macromolecular complexes, enabling detailed investigation of molecular interactions and the mechanisms of biochemical reactions [132–136]. While crystallographic structures are conformationally constrained by the requirements of crystal packing [29], cryo-EM offers the capacity to visualize the conformational landscapes of macromolecules through rapid freezing and vitrification of aqueous samples [137, 138]. However, conventional blot-based methods commonly used for single particle cryo-EM sample preparation are carried out on timescales of seconds to minutes [41, 139]. These relatively prolonged timescales are more conducive to observing low-energy post-equilibrium states of biochemical reactions than the intermediate states associated with molecular recognition and binding or initial states of catalysis and processing. Reducing the timescales required to mix biological samples, substrates, and reagents can enable visualization of the initial interactions and molecular events that trigger important biological mechanisms [30, 39, 53, 61].

Time-resolved cryo-EM has emerged as an attractive area of development for the cryo-EM community, as it could enable visualization of key time-dependent states associated with biochemical reactions. In order to achieve this feat, however, the sample preparation steps must be performed on the timescale of these reactions, which typically occur on the order of microseconds to milliseconds [48, 56, 57, 61]. Further, the enzymes and substrates involved must be thoroughly mixed prior to vitrification [44, 55, 140]. Several devices are currently under development to perform rapid sample application using spray or inkjet technologies, but these devices require high-cost equipment and do not incorporate a method for achieving sample mixing, thus preventing their use in time-resolved cryo-EM studies [56]. To date, time-resolved cryo-EM studies have shown limited application for biological samples due to lack of widespread equipment availability, technology usability, or reproducibility.

Previously, a low-cost, high-speed, spray-based vitrification device, "Shake-it-off" [50], which was superseded by "Back it Up" (BIU) [40], demonstrated its ability to prepare samples for high-resolution cryo-EM structure determination on a timescale of tens of milliseconds. Compared to commercially available devices for high-speed sample preparation that can cost over $1 million and are unable to apply multiple samples to a grid, the BIU costs less than $1000 to build, and the design of the custom parts have been made open source. The BIU device incorporates a piezoelectric transducer (hereafter referred to simply as a "piezo") that enables aerosolization of a sample droplet when a high voltage is applied [40, 50]. BIU is controlled by a Raspberry Pi computer running a customized, open-source Python program that automates the spraying and plunging steps. This device was shown to be capable of high-resolution structure determination, evidenced by a $\sim$2 Å resolution map of apoferritin [40]. However, incorporation of mixed samples was not investigated in the study.

Our work builds on this framework by adapting the device for time-resolved cryo-EM sample preparation. We aimed to 1) establish a system for rapid, multi-component cryo-EM sample

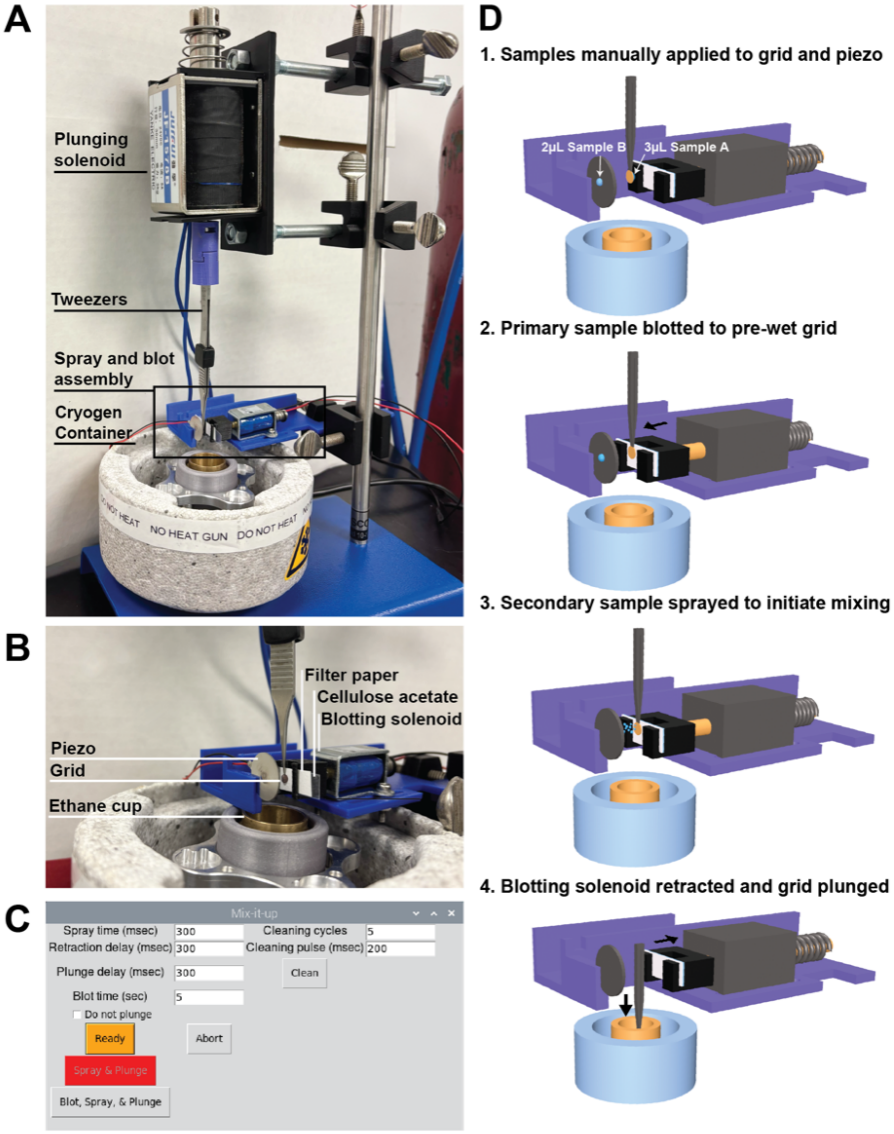


Figure 3.1: Mix-it-up design and method. A) The "Mix-it-up" (MIU) assembly is based on the previously described "Back-it-up" (BIU) design [40]. The grid and tweezers are connected to a plunging solenoid positioned in front of the spray and blot assembly, positioned above the cryogen container. **B)** The spray and blot assembly consists of a piezo electric transducer (piezo) that aerosolizes an applied droplet of sample, directing a spray towards a grid positioned in front of a strip of filter paper used for blotting. **C)** The MIU GUI includes updates to the BIU GUI, including a parameter for blotting time and a button that triggers the blot-spray-and-plunge sequence. **D)** The method used for high-speed sample mixing includes the following steps: 1) manual application of a primary sample to the grid and a secondary sample to the piezo, 2) blotting of the primary sample to pre-wet the grid for improved mixed sample distribution, 3) initiation of spray to apply the secondary sample and trigger on-grid mixing, and 4) retraction of the blotting solenoid and plunging into the cryogen bath.

preparation; 2) capture a conformational change induced by changes in environmental stimuli, and; 3) capture time-dependent intermediates along a protein-protein complex assembly pathway.

To achieve these aims, we first implemented a "blot-then-spray" methodology [58, 141] to enable high-speed, multi-sample mixing. In this approach, one sample is applied manually to the grid and the blotted automatically for a specified period of time, after which a secondary sample is sprayed to induce on-grid mixing, achieving time resolutions of ∼120–720 ms (Fig 3.1, Fig S1). We termed the third iteration of this spray technology "Mix-it-up" (MIU), following the basic design of the previous Shake-it-off and BIU implementations [40, 50]. As proof of principle, we first obtained high-resolution structures of apoferritin and aldolase prepared using MIU via the on-grid mixing approach. We then demonstrated MIU's ability to capture the pH-induced capsid contraction of the plant virus CCMV. Finally, we applied MIU to study the complex formation of the GroEL/ES chaperone system, capturing time-dependent intermediates along its assembly pathway. Together, these results establish MIU as a low-cost system for time-resolved cryo-EM sample preparation.

3.3 Results

3.3.1 Single-sample blotting with MIU is effective for high-resolution structure determination

BIU was designed for single-sample cryo-EM experiments, where sample is manually applied to a piezo, followed by automated aerosolization and application to the grid [40]. In the original BIU framework, through-grid wicking is used instead of conventional blotting to rapidly thin the sprayed sample prior to vitrification. After its publication, the BIU architecture was successfully adapted for conventional blotting for use in a light-coupled, time-resolved cryo-EM platform [47], and in time-resolved studies of kainate receptors [63]. However, in our studies using the BIU design, we found through-grid wicking to be ineffective at reproducibly thinning the sprayed sample across EM grids, leading to regions of thick ice surrounded by limited regions of ice that were amenable to imaging, but nonetheless too thick to yield optimal particle distributions for routine high-resolution particle reconstruction (Fig 3.3A).

To adapt the framework for mixed-sample studies, we implemented a blot-then-spray methodology combining conventional blotting and spraying to promote on-grid mixing while achieving optimal ice thickness (Fig 3.1, Fig 3.2). We hypothesized that by using conventional blotting to generate a thinned layer of the primary sample across the grid, a secondary sample could be sprayed onto the wet surface, where it could rapidly spread and diffuse into the primary sample. We surmised that the small size of the spray droplets would not substantially increase ice thickness, and the thinness of the primary sample (tens of nanometers) would accommodate rapid diffusion of the secondary sample, which would in turn promote interaction between the samples.

To more effectively thin the sample for optimal ice thickness across the EM grid, we modified the blotting assembly to recapitulate the force distribution that would be introduced by

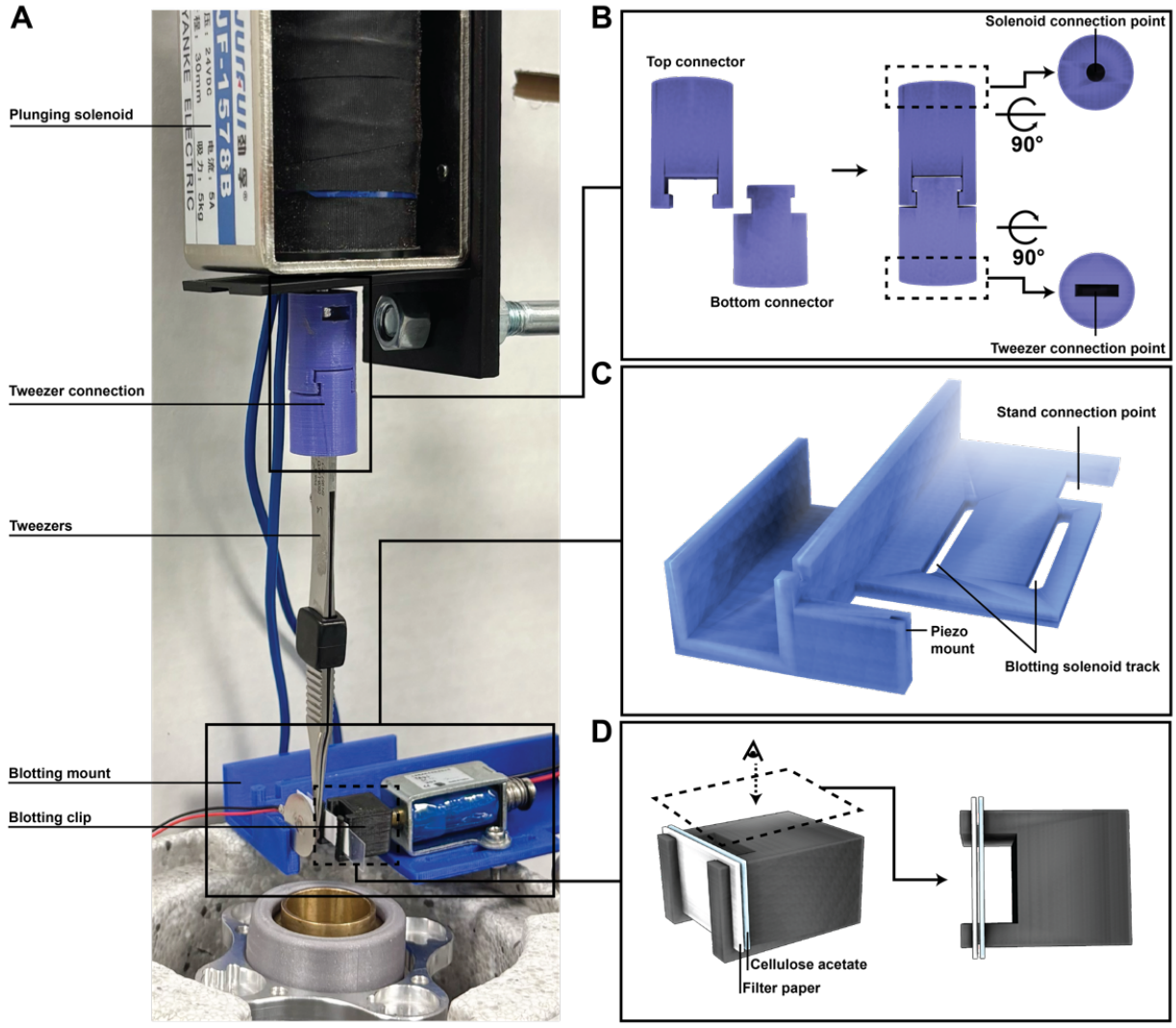


Figure 3.2: Modifications implemented in "Mix-it-up." A) Photograph of the Mix-it-up device denoting modifications introduced. B) Schematic of the updated tweezer connection design, which consists of a top component that connects to the plunging solenoid and a bottom component that connects to the tweezers. The components are shown disjoined (left) and joined (right). The top and bottom views show the connection points to the solenoid and tweezers, respectively. C) Schematic of the blotting mount. D) Side (left) and top (right) views of the blotting clip with filter paper positioned between the two pillars and a strip of cellulose acetate used as backing support.

manual blotting. In our studies, the circular design of the paper used in BIU resulted in inconsistent ice thickness across the grid due to an uneven distribution of force applied. To mitigate this effect, we designed a "clip" with two pillars on either end of a rectangular base that holds a strip of Whatman no. 1 filter paper securely in front of an equally sized strip of cellulose acetate (Fig 3.2D). This design aligns the point of contact of the filter paper and grid with the axis of movement of the blotting solenoid such that an even force is applied across the grid during the blotting process.

We first tested the capacity of the blot-then-spray method and updated framework to prepare cryo-EM grids that are amenable for single particle reconstructions. A homemade gold grid fenestrated with 2.2 μm holes spaced 2.2 μm apart was plasma cleaned and held by tweezers connected to the plunging solenoid. A volume of 3 μL of buffer was manually pipetted onto the side of the grid facing the blotting paper, and 2 μL of apoferritin was applied to the back of the piezo. Blotting was initiated with the "Ready" button on the graphical user interface (GUI), which triggers the forward movement of the blotting solenoid, positioning the filter paper in contact with the droplet of buffer. After 3 s, a 300 ms spray was triggered, immediately followed by simultaneous retraction of the blotting assembly and the grid was plunged into a liquid/solid ethane slurry. The time between the blotting paper retraction and the sample entering the ethane was measured to contribute an additional $\sim$23 ms.

The resulting atlas exhibited a gradient of ice thickness similar to that of grids prepared using a manual plunging device or automated blotting instruments such as the Thermo Fisher Vitrobot (Fig 3.3B). Greater sample thinning was observed within the squares compared to that of preparations with the original BIU framework, yielding improved particle distribution of the sprayed apoferritin (Fig 3.3B). This proof-of-principle demonstrated the ability of our MIU device to generate cryo-EM grids having a consistent distribution of thin, vitreous ice containing sprayed particles for high-resolution reconstruction.

3.3.2 High-resolution structures of mixed apoferritin and aldolase samples using MIU

Previous studies demonstrated the effectiveness of "on-grid mixing," a phenomenon in which samples are independently applied to a grid and mix upon contact within the foil holes [55–57]. We sought to demonstrate the ability of MIU to enable on-grid mixing by manually applying a primary sample to a grid, blotting, and subsequently spraying a secondary sample onto the grid (Fig 3.1D). Given that we observed sprayed apoferritin distributed across a cryo-EM grid that had been pre-wetted with buffer, we next aimed to characterize the sprayed apoferritin distribution across a grid that is pre-wetted with an aldolase sample.

To promote on-grid mixing, it is important to first consider the directionality of the grid during attachment to the plunging tweezers and sample application. The "front" of the grid, also referred to as the "sample side," corresponds to the side where the holey foil has been applied to the grid, where it is relatively flat across the grid bars (Fig 3.4A-B). The "back" side of the grid contains deep wells between the grid bars and the foil (Fig 3.4C), creating an area where sample can pool, making it difficult to generate a thin layer of sample even when blotted from this side, and increasing the likelihood of creating regions of thick ice that are not

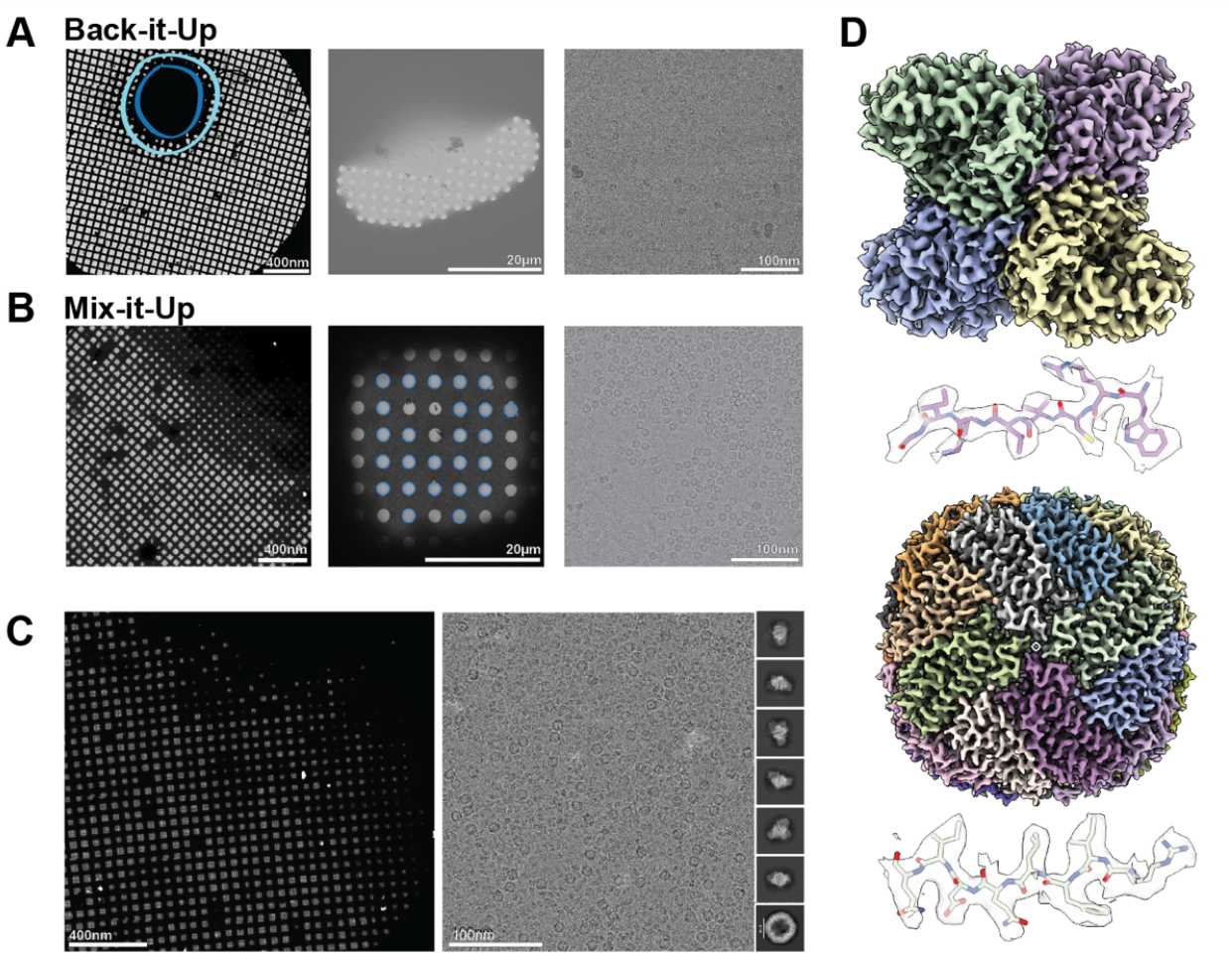


Figure 3.3: Improved ice distribution of mixed samples using MIU. A, B) Representative atlas (left), square (center), and high magnification images of apoferritin (right) obtained with BIU **(A)** and MIU **(B)**, demonstrating the ice and particle distributions achieved with each method. A dark blue circle indicates regions of thick ice, and a light blue circle indicates regions of usable squares for targeting on the leftmost panel of **(A)**. **C)** The atlas (left), representative micrograph (right), and representative 2D class averages obtained from the preparation of mixed apoferritin and aldolase samples using MIU. **D)** EM maps of aldolase (top) and apoferritin (bottom), both resolved to 2.9 Å , with representative peptide density shown beneath each map.

suitable for imaging. To prevent this, the sample side of the grid faces the blotting assembly, as the manually applied primary sample has a larger bulk volume compared to the sprayed droplets of the secondary sample. The spray is therefore targeted at the back side of the grid, where the droplets are expected to spread into sample-filled holes of the fenestrated film rather than pool. Due to the dual-sided sample application in our blot-then-spray method, both sides of the grid are plasma cleaned to ensure hydrophilicity and promote adherence and mixing of the sprayed droplets with the layer of manually applied sample.

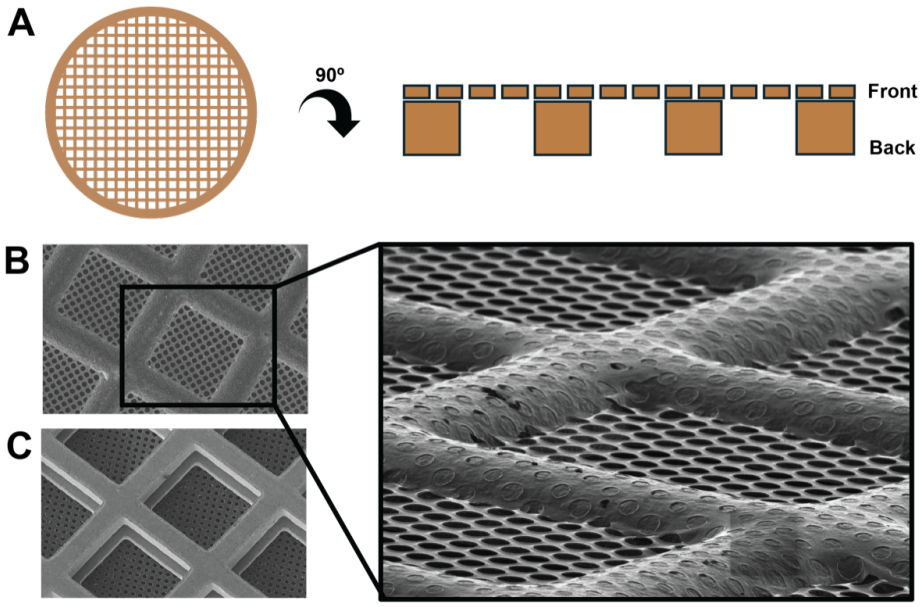


Figure 3.4: EM grid architecture. A) 2D schematic of an EM grid (left) and a representative schematic of the side view of several squares depicting the holey foil (front) and grid bars (back). B) Scanning electron microscopy (SEM) image of the front of a grid, showing a rounded surface of the grid bars overlaid with a holey foil (inset). C) SEM image of the back of a grid.

For our mixed apoferritin and aldolase studies, a 3 μL droplet of aldolase was first manually applied to the sample side of the grid, as described in the single-sample study. A 2 μL droplet of apoferritin was then applied to the back side of the piezo. Next, the blotting assembly was set to the forward position to initiate blotting of the aldolase sample. After 5 s of blotting, the spray was triggered, resulting in aerosolization of the apoferritin droplet. Apoferritin was sprayed onto the back side of the grid for 300 ms, during which the filter paper remained in contact with the grid. Following the spray period, the blotting assembly was retracted, and the grid was rapidly plunge frozen.

The resulting atlas demonstrates ice distribution similar to that obtained in the single-sample experiment, with several pockets of thick ice as well as areas of thinner ice radiating outward from these clusters (Fig 3.3C). Notably, high magnification imaging of this grid revealed an uneven amount of mixing across the grid. The acquired micrographs could be categorized into several phenotypic groups, containing: 1) high aldolase particle concentration with no apoferritin particles present (Fig 3.5A), 2) high aldolase particle concentration with apoferritin particles present at low concentration (Fig 3.5A), 3) optimally mixed aldolase and apoferritin

particles (Fig 3.5B), and 4) high apoferritin particle concentration with aldolase particles present at low concentration (Fig 3.5C).

The first case, in which only aldolase particles are observed in micrographs, is likely indicative of regions where aerosolized droplets of the sprayed sample did not contact the grid. Micrographs that exhibit a high concentration of aldolase but low apoferritin concentration could have resulted from 1) interaction of a small aerosolized droplet with the grid that did not spread across a large area, 2) the application of several individual droplets that did not pool following application to the grid leading to low, local concentrations of apoferritin, or 3) imaging at regions at the outer boundary of a sprayed droplet or pooled collection of droplets. Conversely, in micrographs that exhibit a high apoferritin concentration but low aldolase concentration, apoferritin appears to displace aldolase in order to engage in preferred inter-particle interactions, in some cases resulting in hexagonal particle packing. The micrographs with high apparent apoferritin concentrations may result from large droplets containing high amounts of apoferritin hitting the grid, or from many droplets hitting the grid with close proximity and pooling, leading to a localized increase in particle concentration.

While apoferritin and aldolase are readily distinguishable in our micrographs due to their large size and distinct structures, this ability to visually evaluate sample mixing will not extend to sprayed small molecules or buffer components that are often used to modulate biologically relevant reactions. Further, given the variation in the distribution and local concentrations of sprayed apoferritin throughout our dataset, we concluded that it was crucial to track the application of sprayed secondary samples that consist solely of buffer components, ligands, or small peptides, which cannot be resolved using single-particle cryo-EM. Thus, we incorporated 5nm gold fiducials into the sprayed sample of subsequent experiments in order to confirm areas of application on the grid. This approach has previously been used in foundational time-resolved EM studies to track the location of sprayed sample [60]. Nanogold fiducials are well-suited as high-contrast and readily identifiable markers that can be observed at high magnifications but may require computational deletion from the images for small or challenging particles due to artifacts that can occur during the analysis of images containing nanogold.

Processing of our mixed sample dataset yielded a $\sim$2.9 Å resolution map of apoferritin and a $\sim$2.9 Å resolution map of aldolase (Fig 3.3D, Fig B.1, Table B.1). These data demonstrate the capacity of MIU to enable high-resolution structure determination of multiple samples mixed on-grid using our blot-then-spray methodology.

3.3.3 Application of MIU to study conformational changes

3.3.4 pH-induced contraction of CCMV observed on the millisecond timescale

To assess whether MIU could capture conformational changes that take place on the millisecond timescale, we examined the well-established model system of the Cowpea chlorotic mottle virus (CCMV). CCMV is a plant virus of the Bromoviridae family that infects the cowpea species. The capsid has T=3 icosahedral symmetry, consisting of 180 identical copies of the

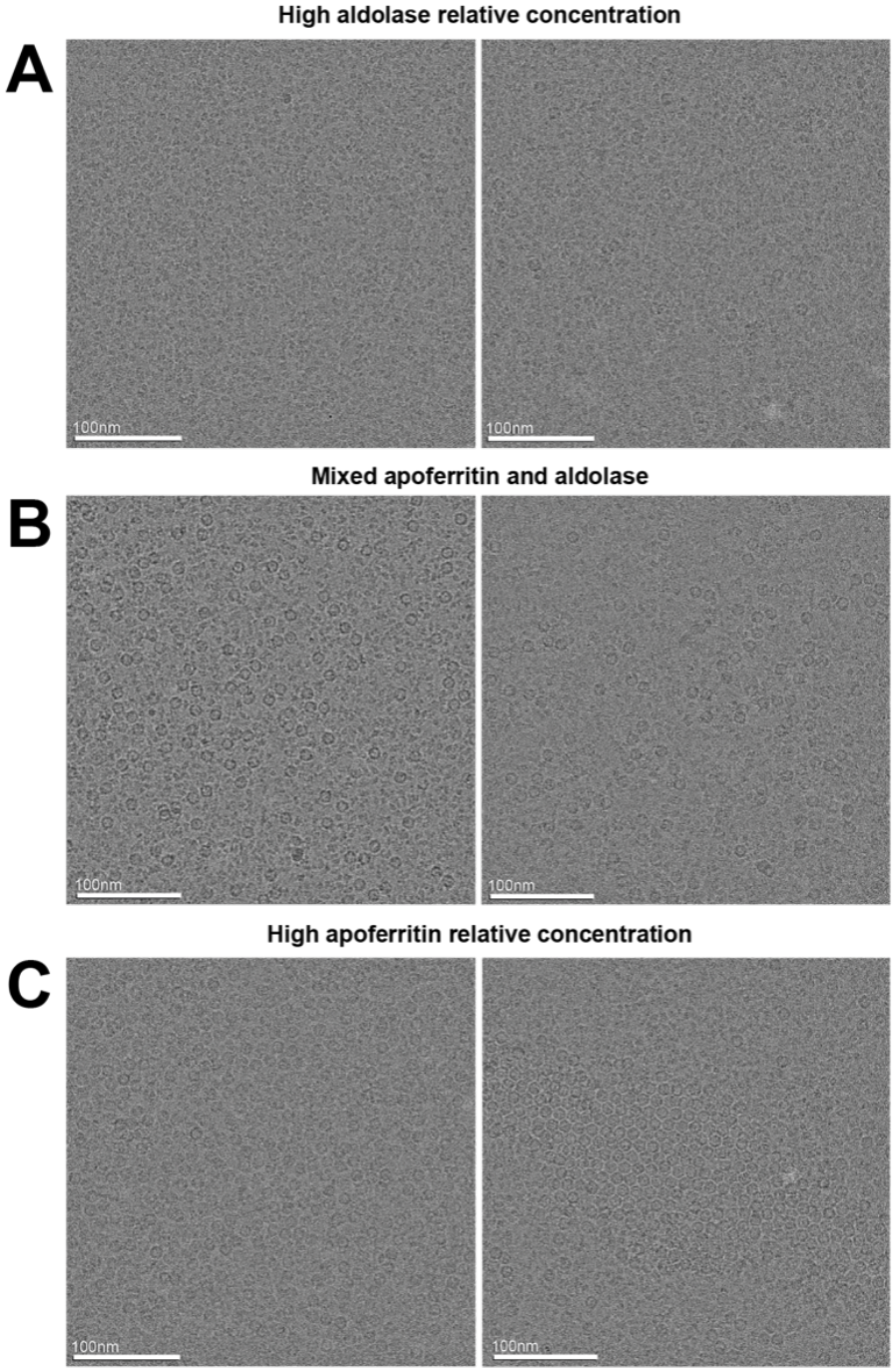


Figure 3.5: Aldolase and apoferritin particle distribution in mixed sample prepared with MIU. Representative micrographs from grids prepared with: A) high aldolase concentration, B) mixed apoferritin and aldolase concentrations, and C) high apoferritin concentrations.

coat protein [142]. During infection, CCMV capsids expand to release genetic material as they transition from the low-pH extracellular environment to higher pH conditions within the cellular cytosol. Both the contracted and expanded states of the capsid have been characterized by crystallography and cryo-EM [48, 142]. At a pH of 4.8, or in the absence of divalent cations such as Ca^{2+} or Mg^{2+}, the capsid favors a contracted form with a diameter of 29 nm [142]. At pH 7 or higher and under conditions of low ionic strength, the capsid expands to a diameter of 32 nm [48, 142]. Notably, this shift in conformation is reversible with purified capsid, so that a sample can be expanded or contracted in vitro by changing the pH of its buffer.

A recent study implemented specialized hardware for rapid melting and revitrification of cryo-EM samples to study the timescale of CCMV capsid contraction following a photoacid-induced pH shift in the surroundings [48]. With a temporal resolution of 30 μs, the CCMV capsids were found to undergo a $\sim$10 Å decrease in diameter, as well as a slight rearrangement in the capsid organization, corresponding to a partially contracted state.

In contrast, we focused on using MIU to trigger the full conformational change of CCMV capsids under low pH conditions to demonstrate MIU's capabilities as a low-cost method for studying large-scale conformational changes on the millisecond timescale. To induce a pH shift and trigger the contraction of expanded CCMV capsids, we sprayed low pH buffer (pH 3.6) onto a grid pre-blotted with expanded CCMV capsids. Nanogold fiducials were mixed with the low pH buffer to track the localization of the buffer on the grid and confirm areas of sample mixing.

Our expanded CCMV control was prepared by dialyzing CCMV capsids stored in low pH buffer (pH 4.6) into high pH buffer (pH 7.6) overnight. To confirm the conformational states of the CCMV control samples, the putative contracted (pH 4.6) and expanded (pH 7.6) forms were prepared separately by manually applying each sample to the grid and blotting for 5 s, followed by retraction of the blotting assembly and plunging of the grid. Our $\sim$3.2 Å reconstruction of the low pH CCMV (Fig B.2, Fig B.3, Table B.1) and our $\sim$6.9 Å reconstruction of the high pH CCMV (Fig B.4, Table B.1) are consistent with the contracted and expanded states previously described [48, 142], exhibiting diameters of 29 nm and 32 nm, respectively (Fig 3.6C).

To determine if capsid contraction could be triggered through sample preparation with MIU, we first applied expanded CCMV sample at pH 7.6 to the grid and blotted for 5 s. The filter paper was held in the blotting position as low pH buffer (pH 3.6) mixed with gold fiducials was sprayed for 300 or 700 ms onto the opposite side of the grid to promote on-grid mixing of the capsids and buffer. Immediately after spraying, the filter paper was retracted, and the grid was plunged into the cryogen.

CCMV capsids were observed in the presence of gold fiducials throughout the acquired micrographs, confirming mixing of the capsids and the low pH buffer. Most micrographs containing capsids also contained 1–3 nanogold fiducials, though the local concentration varied, with clusters of more than 15 beads sometimes present in the micrographs. Capsids were also observed infrequently in micrographs without fiducials, though the neighboring micrographs contained gold, indicating some variability in the local concentration of fiducials.

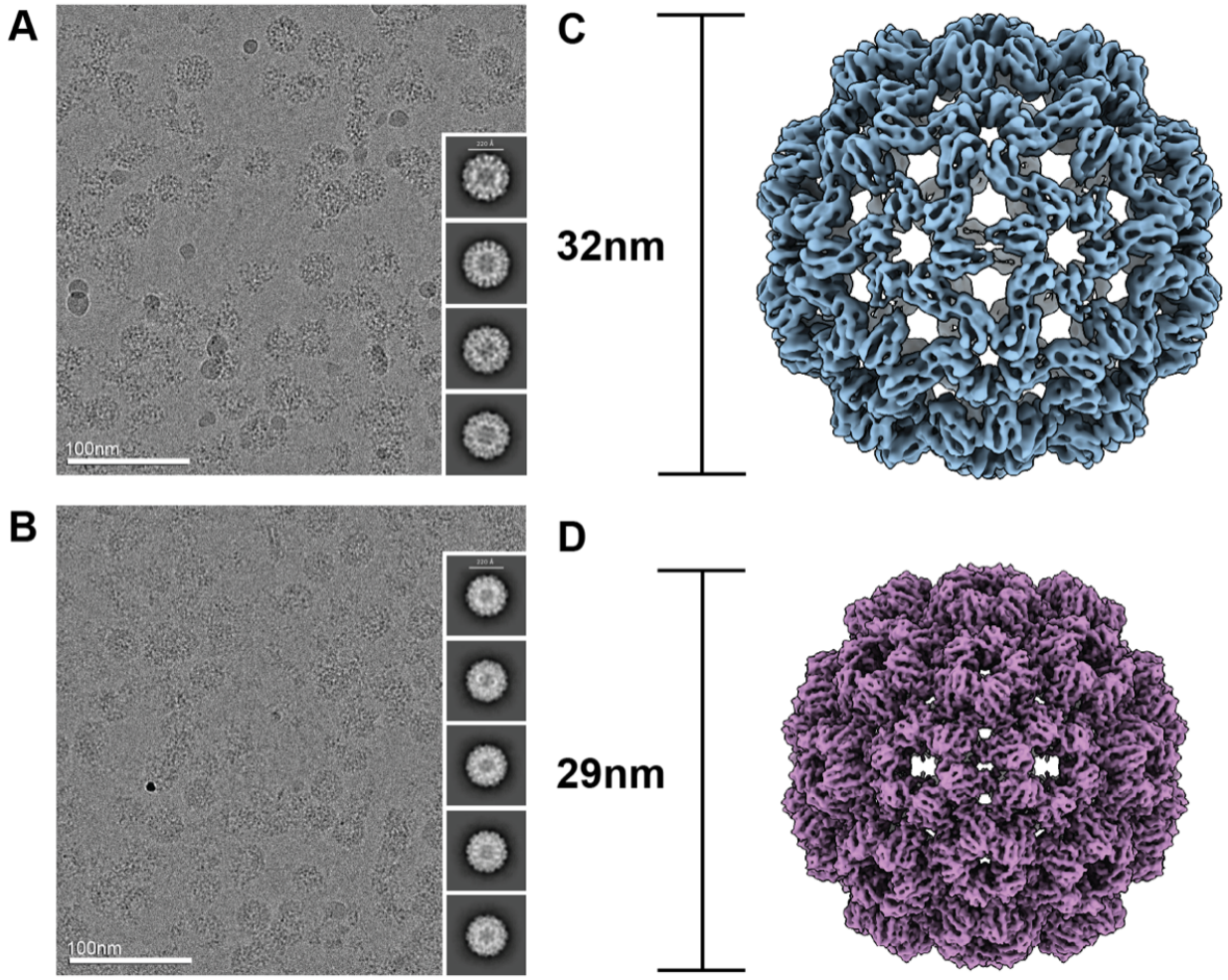


Figure 3.6: pH-induced contraction of CCMV using MIU. A, B) Representative micrographs and 2D class averages of the expanded state control **(A)** used as a starting point for the pH-shift experiment and the state resulting from contraction induced by application of low pH buffer with MIU **(B)**. **C, D)** Representative maps of the expanded state control with a diameter of 32 nm **(C)** and the pH-shift-induced contracted state at 29 nm **(D)**.

We observed a similar level of variation in the local distribution of fiducials during initial testing of a grid prepared with buffer as the primary sample and sprayed with a mixture of apoferritin and nanogold fiducials (data not shown). In this prior dataset, nanogold was observed in the majority of images containing apoferritin, with only a few instances of apoferritin appearing in the absence of nanogold; thus, we concluded that the fiducials serve as a good approximation of the location of sprayed sample on the grid.

In the CCMV dataset, many regions of denatured capsids were observed, which was consistent with the data presented in the melting/revitrification CCMV pH shift study [48], possibly indicating the structural instability of the capsid under such low pH conditions. Our ∼3.9 Å reconstruction corresponded to a fully contracted state, with a diameter of 29 nm (Fig 3.6D, Fig B.5, Table B.1), consistent with the expected conformational state of CCMV capsids under low pH conditions [48, 142].

3.3.5 MIU captures ATP-dependent GroEL/ES complex formation

Ligand-mediated protein conformational changes are attractive targets for time-resolved studies of biological systems due to rapid diffusion of small molecules. However, capturing these events remains technically challenging, with only a few successful examples demonstrated using highly specialized crystallographic techniques, such as serial femtosecond crystallography [143–146]. These methods often require extensive optimization and incur high costs for facility usage. Time-resolved cryo-EM offers a more versatile approach, enabling structural studies of larger and more complex proteins in a near-native state while accommodating a wider range of conformational dynamics. Further, the development of custom-built, cost-effective instrumentation for rapid sample preparation substantially reduces experiment-associated costs.

Several time-resolved cryo-EM studies have used ligands or small substrates to probe rapid conformational changes such as ATP-dependent dissociation of actomyosin [60], ss-DNA-RecA filament growth [44], and GTP-dependent dynamin conformational rearrangements [56]. We therefore investigated the capacity of rapid on-grid mixing using MIU to capture the assembly of the bacterial chaperonin GroEL/ES complex. The GroEL/ES machinery has served as a model system for understanding ATP-dependent protein folding and as a benchmark for studying protein complex assembly by cryo-EM [45, 147, 148]. GroEL forms a homo-tetradecameric assembly in the absence of GroES and ATP. Upon ATP binding, the apical heptamer undergoes a conformational change that facilitates GroES binding, thereby capping the complex and creating an enclosed cavity where unfolded or misfolded substrates can refold [149, 150]. GroEL/ES capping has previously been observed to occur on the millisecond to second timescale [45].

To establish the baseline organizational states of GroEL/ES in the absence of ATP, a mixture of GroEL/ES was prepared and incubated at room temperature for ten minutes prior to vitrification. The sample was manually applied to the grid, blotted, and vitrified, without addition of a sprayed sample. Uncapped GroEL, corresponding to a "barrel" conformation, was present uniformly in the data. Our ∼3.3 Å map shows a lack of nucleotide density in the nucleotide binding pocket (Fig 3.7A, Fig B.6, Table B.2). GroES was also present within the

sample and observed in 2D class averages, often forming a hexagonally-packed lattice with strong preferred orientation (Fig B.6). Extensive rounds of 2D classification were performed to confirm the absence of capped GroEL particles.

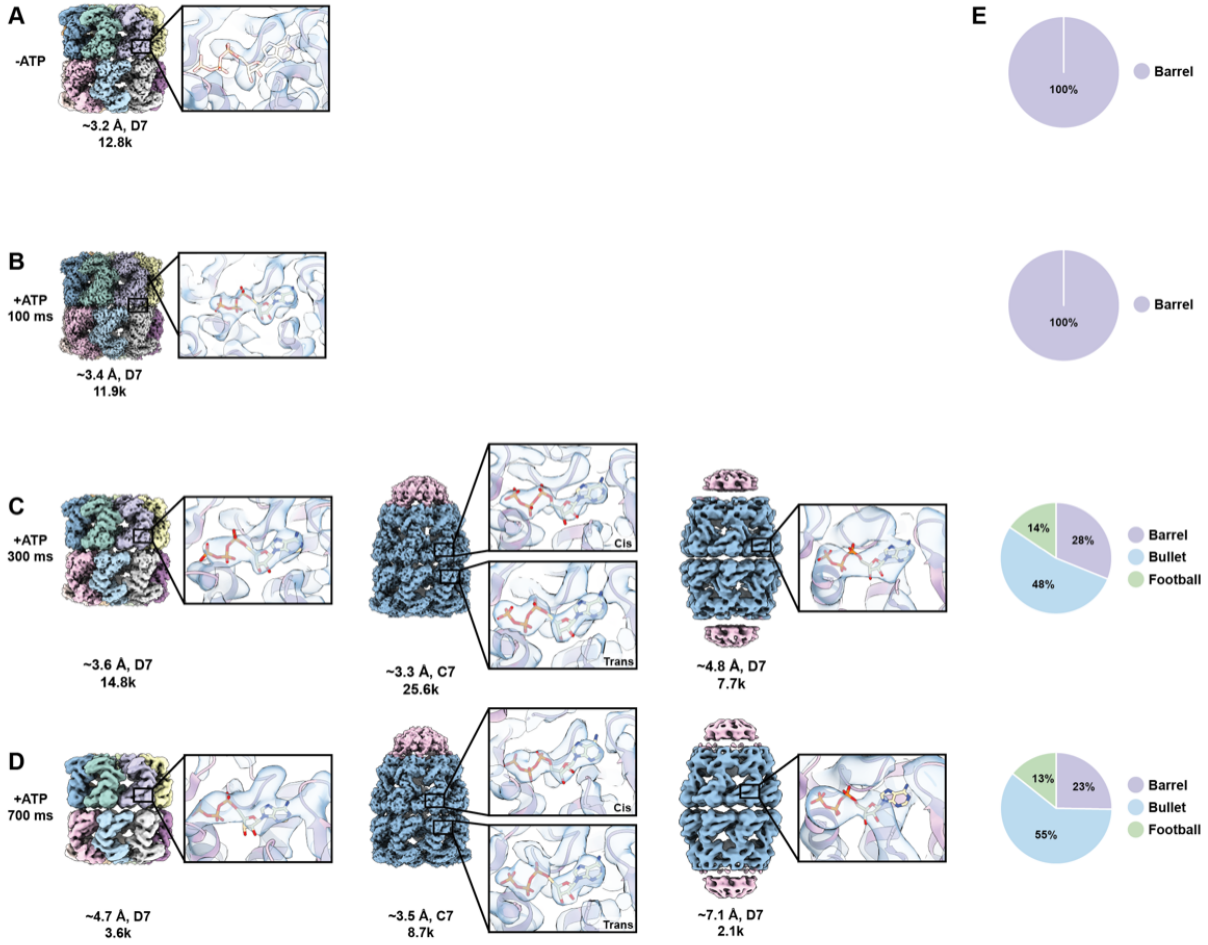


Figure 3.7: ATP-dependent complex formation of GroEL/ES observed following sample preparation with MIU. Density maps obtained from preparation of: **A)** the control GroEL in the presence of GroES without ATP; **B)** GroEL in the presence of GroES and sprayed ATP after 100 ms; **C)** GroEL + GroES with sprayed ATP after 300 ms; **D)** GroEL + GroES with sprayed ATP after 700 ms. The insets display representative subunit nucleotide density. **E)** Particle distributions for the control, 100 ms, 300 ms, and 700 ms from top to bottom, respectively.

We next assessed whether the blot-then-spray approach could drive GroEL/ES complex formation upon on-grid mixing. GroEL/ES samples were prepared as described for the ATP-free condition, with the addition of sprayed ATP supplemented with nanogold fiducials to identify regions of sprayed ATP localized on the grid. At the 100 ms timepoint, we were unable to detect GroEL/ES assembly, as all the GroEL particles were observed in the "barrel" conformation, despite performing several rounds of 2D classification in an attempt to identify a subpopulation of GroEL/ES particles (Fig 3.7B, Fig B.7, Table B.2). Despite remaining in an uncomplexed state, nucleotide density was observed in the nucleotide binding pocket corresponding to an ATP molecule, indicating that 100 ms was sufficient for nucleotide-

binding (Fig 3.7B). These data are consistent with fluorescence resonance energy transfer (FRET) data demonstrating ATP-binding to GroEL within 10 ms [151], as well as with a previous time-resolved cryo-EM study of GroEL/ES complex formation [45]. Although GroEL was observed complexed with GroES in the previous time-resolved study after 50 ms, the proportion of these complexes was low (∼4.2%). This discrepancy may be due to the difference in on-grid mixing behavior of particles in our study relative to the microfluidics system used in the former study.

At both the 300 ms and 700 ms timepoints, the majority of the GroEL particles were observed complexed with GroES, predominantly in the single-capped "bullet" conformation, with one GroES heptamer bound to one GroEL ring. Multiple rounds of 2D classification further revealed the presence of a double-capped GroEL/ES "football" complex in which GroES caps both ends of the GroEL barrel. 2D classification also revealed a small population of uncapped GroEL barrels. At the 300 ms timepoint, 48% of the extracted particles adopted the bullet conformation, 28% exhibited the barrel conformation, and 14% were in the football conformation (Fig 3.7C, Fig B.8, Table B.3). These proportions were similar in the 700 ms dataset, with 55%, 23%, and 13%, corresponding to the bullet, barrel, and football assemblies, respectively (Fig 3.7D, Fig B.9, Table B.4), suggesting that the GroEL/ES assembly reaction had largely reached equilibrium within the first 300 ms following ATP addition.

All three conformations exhibited nucleotide density in the nucleotide-binding pockets, including the uncapped GroEL barrels in both datasets (Fig 3.7C-D). The density was unambiguously resolved as ATP, except in the cases of the 700 ms barrel conformation and both the 300 ms and 700 ms football conformations, where low particle counts limited our attainable resolution, which prevented confident assignment of nucleotide state. These findings are consistent with pre-steady-state kinetic data showing that ATP hydrolysis occurs on the second timescale [152]. Interestingly, the barrel conformation of the 300 ms and 700 ms datasets closely resemble the GroEL conformation observed in the previous time-resolved study (PDB:8BL7), in which the apical domains of both heptameric rings are angled downward, reflecting an ADP-like state, despite exhibiting density for ATP in the nucleotide pockets. This may reflect a transient state between nucleotide binding and the rearrangement of the apical domains to accommodate GroES binding. The bullet conformation of these datasets exhibits a similar ADP-like conformation in the trans ring of GroEL and ATP-like conformation of the cis ring, despite density for ATP observed in both rings, reflecting a state that closely resembles another conformation observed in the previous time-resolved study (PDB:8BMO). Taken together, these data confirm the ability of MIU to capture ligand-induced conformational states in protein complexes through rapid on-grid mixing.

3.4 Discussion

In this study, we present Mix-it-up (MIU) - an updated and low-cost methodology for on-grid sample mixing, enabling the study of conformational changes in biological systems on the millisecond timescale. We first demonstrate that MIU improves the overall quality and consistency of ice on the grid suitable for imaging compared to previous iterations of the device (Shake-it-off, Back-it-up) [40, 50]. To illustrate the feasibility of on-grid mixing, we

used MIU to combine distinct protein samples - manually applied aldolase and sprayed apoferritin - serving as a proof of principle. We further show that this system supports the study of environmentally or ligand-induced conformational changes in protein samples with CCMV and the GroEL/ES chaperonin complex.

We also demonstrate the use of nanogold fiducials to localize spray distribution and localize regions of sample mixing on the grid. While we used fiducials in buffer or mixed with ligand solutions, this strategy could be extended to protein samples to track the spatial distribution of sprayed proteins that are too small or disordered to resolve by cryo-EM in an uncomplexed form.

Despite the accessibility and cost-effectiveness of the MIU setup, achieving full automation and greater reproducibility will require further engineering advances to improve precision and timing control. The current design also lacks the temporal resolution required to capture conformational changes on the single-millisecond timescale. Future modifications, such as enabling the spray to occur during grid plunging after pre-blotting of a primary sample, could improve the achievable time resolution. The addition of a nitrogen shield may also allow the grid to be positioned closer to the ethane bath, reducing plunge time. While we demonstrate MIU's ability to capture protein-protein complex assembly through ligand-mediated GroEL/ES assembly, further characterization of the wide-ranging types of protein-protein, protein-nucleic acid, or protein-ligand interactions accessible using MIU would help to expand the scope of this methodology.

Despite these limitations, MIU represents a substantial improvement over traditional blotting-based methods, which typically require several seconds for sample preparation. By enabling sample mixing and vitrification within a few hundred milliseconds, MIU opens new possibilities for examining dynamic biological systems that evolve on intermediate timescales. As such, MIU offers a robust and adaptable platform for investigating transient states in macromolecular assemblies by single-particle cryo-EM, and lays the groundwork for future time-resolved cryo-EM studies of rapid biological processes, such as catalysis, signaling, and transport.

3.5 Methods

3.5.1 MIU Build

MIU was initially assembled following instructions for the Shake-it-up and BIU devices designed by the Rubinstein group, as previously described [40, 50]. Several modifications were made to the device, which included a redesign of the 3D-printed tweezer-to-solenoid connection, the blotting clip, and the blotting mount, as shown in (Fig 3.2). The STL files of these redesigned components are available at https://github.com/laurenalex77/Mix-it-up. The software repository was downloaded from https://github.com/johnrubinstein/BIUcontrol.git and modified as follows: A method to enable the blot-spray-and-plunge sequence was added to the BIUapplyandplunge.py script and BIUgui.py, along with a parameter to specify the blot time within the GUI. A 60 second delay was also added following the plunge instruction in BIUapplyandplunge.py to enable adequate time for disconnection of the tweezers prior to transferring the grid to liquid nitrogen from the ethane bath.

3.5.2 Sample preparation

3.5.2.1 Apoferritin

Heavy-chain mouse apoferritin was purified from frozen cell pellets prepared as previously described [153]. A pET24a vector was provided by M. Kikkawa (University of Tokyo) and was transformed into Rosetta *E. coli* cells. Cells were grown at 37° C in LB media and protein expression was induced with 1 mM IPTG at OD600nm = 0.5. Protein was expressed for 3 hours at 37° C, after which cells were harvested via centrifugation and stored at −80° C. Cell pellets were thawed and resuspended in lysis buffer containing 30 mM HEPES pH 7.5, 300 mM NaCl, 1 mM $MgSO_4$, 1x Roche inhibitor, 30 mg lysozyme and lysed by sonication for a total of 10 min at 30% amplification with a pulse of 3 s on / 8 s off. The lysate was clarified via centrifugation at 20,000 RCF, 4°C for 20 min. Heat precipitation was performed to remove host cell proteins by incubating the clarified lysate in a 70°C water bath for 10 min and performing centrifugation at 20,000 RCF, 4°C for 15 min. Ammonium sulfate precipitation was performed by adding 60% (w/v) ammonium sulfate to the supernatant, stirring on ice for 10 min, and centrifuging at 14,000 RCF, 4°C for 20 min. The pellet containing apoferritin was resuspended in PBS and dialyzed in a 10 kDa MW cut-off dialysis cassette overnight in dialysis buffer (30 mM HEPES pH 7.5, 20 mM NaCl, 1 mM DTT) to remove ammonium sulfate. Ion exchange was performed the next day using two connected 5 mL Hi Trap Q HP columns. The column was washed with low salt buffer (30 mM HEPES pH 7.5, 20 mM NaCl, 1 mM DTT) and eluted from a gradient of 0–100% high salt buffer (30 mM HEPES pH 7.5, 500 mM NaCl, 1 mM DTT). Fractions containing apoferritin were pooled and concentrated using a 100 kDa MW concentrator. Apoferritin was further purified by size exclusion chromatography (SEC) using a Superdex 200 10/300 GL (Cytiva) column and SEC buffer (30 mM HEPES pH 7.5, 150 mM NaCl, 1 mM DTT). Fractions containing apoferritin were pooled, concentrated to 14 mg/mL, and flash-frozen in 30 mM HEPES pH 7.5 + 150 mM NaCl.

3.5.2.2 Aldolase

Lyophilized rabbit muscle aldolase (Sigma-Aldrich) was solubilized in 20 mM HEPES pH 7.5, 50 mM NaCl to a final concentration of 1.4 mg/mL.

3.5.2.3 Cowpea Chlorotic Mottle Virus

CCMV was obtained as a gift from the Nicole Steinmetz lab at the University of California, San Diego. Preparation of CCMV was performed as previously described [154]. CCMV was harvested 14 days after inoculation of *Vigna unguiculata*, California black-eyed pea No. 5 plants. CCMV was purified following homogenization of leaves in buffer consisting of 0.2 M NaOAc and 1 mM EDTA at pH 4.8. The homogenate was filtered through cheesecloth, centrifuged, and the supernatant was precipitated overnight with 0.02 M NaCl and 8% (w/v) PEG. Following centrifugation, the pellet was resuspended in the contracted state buffer consisting of 0.1 M NaOAc and 1 mM EDTA at pH 4.8. The sample was subjected to further centrifugation and the supernatant was applied to a 20% (w/v) sucrose cushion and centrifuged for an additional 2 hr at 4°C, 148,000 RCF. The pellet was resuspended in the contracted state buffer and stored as an intact virion at −20°C until further use.

3.5.2.4 GroEL and GroES

E. coli GroEL and Gro ES were received as a gift from the Steven Johnson lab at the University of Indiana. Protein expression and purification of GroEL and GroES were performed as previously described [155]. T7 Express *E. coli* (camR) containing a Trc-promoted GroES/GroEL plasmid (ampR) were cultured in LB media and induced with 1 mM IPTG at $OD_{600} \approx 0.8$ for 4 h at 37°C. Cells were lysed by microfluidization after resuspension of cell pellets in 50 mM Tris pH 7.4, 50 mM KCl, 1 mM DTT, and 1 mM PMSF. The lysed cells were clarified by centrifugation (22,000 RCF for 45 min), and the lysate was purified via cation exchange using FFQ (Cytiva) resin in buffer containing 50 mM Tris pH 7.4, 1 mM DTT, with a 0 to 1 M NaCl gradient. GroEL-containing fractions were precipitated with 1.2 M ammonium sulfate at 4°C and further purified via hydrophobic interaction chromatography using source 15ISO resin (Cytiva) in buffer containing 50 mM Tris pH 7.4, 50 mM KCl, 150 mM NaCl, and 1 mM DTT. GroEL fractions were then dialyzed into storage buffer containing 50 mM Tris pH 7.4, 150 mM NaCl, and 1 mM DTT.

E. coli GroES was similarly purified with the following exceptions: after cation exchange, Gro-ES containing fractions were adjusted to pH 4.6 with 50 mM sodium acetate buffer. The sample then further purified by additional cation exchange using a FFSP (Cytiva) column in buffer containing 50 mM Tris pH 4.6, 1 mM DTT, with a 0 to 1 M NaCl gradient. GroES-containing fractions were size-exchanged on a Hiload 26/600 Superdex 75 (GE) column and eluted in 50 mM Tris pH 7.4, 300 mM NaCl, and 1 mM DTT. Fractions were concentrated GroES were concentrated and stored.

GroEL and GroES stocks used for cryo-EM sample preparation were diluted as needed in buffer containing 50 mM Tris, pH 7.4, 50 mM KCl, 1 mM $MgCl_2$.

3.5.2.5 Gold fiducials

Ted Pella 5nm gold colloid was prepared according to Iancu et al., 2006 [156] (section "First Phase", Tube B). 5x the desired sample volume of fiducials was centrifuged at 18,000 RCF for 25 min at 4°C. The pellet containing the fiducials was resuspended in sample, resulting in a 5x final concentration of fiducials.

3.5.3 Cryo-EM grid preparation

The general procedure for grid preparation is as follows: A homemade, holey gold 2.2/2.2, 400 EM grid prepared as previously described [157] or an UltrAuFoil R2/2, 200 was plasma cleaned for 7 s at 25 Watts (75% argon, 25% oxygen) using a Solarus plasma cleaner (Gatan, Inc.). For single-sample experiments, only the sample side of the grid was plasma cleaned. Both sides were cleaned for spray experiments. The grid was then mounted to the tweezers above the cryogen bath with the smooth side facing towards the blotting paper and the back side (copper) positioned in front of the piezo. A 0.5 x 2 cm rectangle of Whatman No.1 filter paper was positioned in the blotting clip in front of a similarly sized strip of cellulose acetate. 3 μL of sample A was manually applied to the smooth side of the grid. 2 μL of sample B was applied to the back side of the piezo (opposite the side facing the grid). The spray time,

retraction delay, and plunge delay were set to 100, 300, or 700 ms, and the blot time was set to 5 s on the MIU GUI. The "Blot, Spray, and Plunge" button was pressed to trigger 5 s of blotting, followed by the specified spray time, filter paper retraction, and plunging into liquid ethane.

3.5.3.1 Aldolase + Apoferritin.

Sample A, 7 mg/mL of apoferritin in 30 mM HEPES pH 7.5 + 150 mM NaCl, was manually applied to a homemade grid. Sample B, 1.4 mg/mL of aldolase in 20 mM HEPES pH 7.5, + 50 mM NaCl, was applied to the piezo.

3.5.3.2 Contracted CCMV.

Sample A, 10 mg/mL of contracted CCMV in 0.1 M NaOAc, 1 mM EDTA, pH 4.8, was manually applied to a homemade grid.

3.5.3.3 Expanded CCMV.

Sample A, 20 mg/mL of expanded CCMV in 10mM Tris (base), 50mM NaCl, 1mM EDTA, pH 7.6, was manually applied to a homemade grid.

3.5.3.4 Expanded CCMV + low pH buffer / gold (700 ms).

Sample A, 20 mg/mL expanded CCMV in 10mM Tris (base), 50mM NaCl, 1mM EDTA, pH 7.6, was manually applied to an UltrAuFoil 2/2, 200 grid. Sample B, low pH buffer consisting of 0.1 M NaOAc, 1 mM EDTA at pH 3.6, was applied to the piezo.

3.5.3.5 GroEL + GroES, no ATP.

Sample A, 8.6 μM of GroEL + 17.1 μM of GroES, was incubated at room temperature for 10–15 min and applied manually to the grid. Spray time, plunge delay, and retraction delay were all set to 700 ms.

3.5.3.6 GroEL + GroES + ATP/gold (100 ms).

Sample A, 8.6 μM GroEL + 17.1 μM GroES, was incubated at room temperature for 10–15 min prior to freezing and applied manually to a homemade grid. Sample B, 83.3 mM ATP, was applied to the piezo. Spray time, plunge delay, and retraction delay were all set to 100 ms.

3.5.3.7 GroEL + GroES + ATP/gold (300 ms).

Sample A, 8.6 μM of GroEL + 17.1 μM of GroES, was the same mixture used for the no-ATP-control and was applied manually to an UltrAuFoil 2/2 grid following incubation at RT. Sample B, 83.3 mM ATP, was applied to the piezo. Spray time, plunge delay, and retraction delay were all set to 300 ms.

3.5.3.8 GroEL + GroES + ATP/gold (700 ms).

Sample A, 8.6 μM of GroEL + 17.1 μM of GroES, was the same mixture used for the no-ATP-control and was applied manually to an UltrAuFoil 2/2 grid following incubation at RT. Sample B, 83.3 mM ATP, was applied to the piezo. Spray time, plunge delay, and retraction delay were all set to 700 ms.

The plunge delay for MIU experiments is approximately 23.1 ms. This average was calculated based on three separate videos from experiments where the spray time was set to 100 ms, 300 ms, and 700 ms. The number of frames between the initiation of spray and the submergence of the grid in liquid ethane were counted. There were 24 frames in the 100 ms experiment video (200.58 fps), 64 frames in the 300 ms experiment video (200.43 fps), and 147 frames in the 700 ms experiment video (201.21 fps). Subtracting the time of spray, the resulting plunge delays were calculated to be 19.6 ms, 19.3 ms, and 30.4 ms, yielding an average of 23.1 ms.

3.5.4 Cryo-EM data collection

Cryo-EM data was collected on a Talos Arctica equipped with a 200 kV field emission gun and a Falcon 4i direct electron detector. Micrographs were collected at 150,000 x nominal magnification (yielding a pixel size of 0.94 Å/pixel at the detector) using the EPU data collection software. The specified defocus range was -1.4, -1.2, -1 μm. The mixed apoferritin/aldolase dataset was collected at an exposure rate of 10.53 e^-/pixel/s with an exposure time of 4.22 s. The contracted CCMV control dataset was collected at an exposure rate of 10.75 e^-/pixel/s with an exposure time of 4.13 s. The expanded CCMV control dataset was collected at an exposure rate of 9.93 e^-/pixel/s with an exposure time of 4.48 s. The CCMV pH shift dataset was collected at an exposure rate of 10.9 e^-/pixel/s with an exposure time of 4.08 s. The GroEL/ES, no ATP and +ATP 700 ms datasets were collected at an exposure rate of 10.9 e^-/pixel/s with an exposure time of 4.08 s. The GroEL/ES + ATP 300 ms dataset was collected at an exposure rate of 9.66 e^-/pixel/s with an exposure time of 4.61 s. The GroEL/ES + ATP 100 ms dataset was collected at an exposure rate of 10.94 e^-/pixel/s with an exposure time of 4.08 s.

3.5.5 Cryo-EM data analysis

3.5.5.1 Aldolase + Apoferritin.

The image processing pipeline is shown in (Fig B.1). 2,476 movies were collected. Motion correction and CTF correction were performed in cryoSPARC Live v4.6.2 [158]. After applying a resolution cutoff of 6 Å based on CTF estimation, 1,673 micrographs remained for further processing. Apoferritin particles were picked with cryoSPARC live Blob Picker with a particle diameter of 100–110 Å, an NC score > 0.3 and a power score range of -17 to 1,329. 1,071,352 particles were extracted with a box size of 200 px, binned to 100 px. 2D classification was performed with fifty classes. Subset selection was performed with stringent cleanup (the single class having the most well-defined secondary structure was selected) and minimal cleanup (only one class that did not have features consistent with apoferritin was excluded) in parallel, with 81,562 particles selected in the stringent instance and 1,049,548 selected in

the minimal instance. Ab initio reconstruction with one class, octahedral symmetry, and the initial resolution set to 6 Å was performed on the stringent cleanup particles, followed by non-uniform refinement with octahedral symmetry, the initial low-pass resolution set to 8 Å, and per-particle defocus and per-group CTF parameter optimization turned on. Ab initio reconstruction with three classes and octahedral symmetry, and the initial resolution set to 8 Å was performed on the minimal cleanup particles. The reconstruction from non-uniform refinement of the stringent clean up particles was input into a heterogenous refinement job, along with two classes from the minimal clean up ab-initio to serve as junk classes for sorting. All extracted particles were input, and the refinement was run with octahedral symmetry and hard classification enforced. 465 k particles were sorted into the "good" class as the output of the heterogenous refinement job. These particles were re-extracted without binning and further refined using non-uniform refinement with octahedral symmetry, the initial low-pass resolution set to 8 Å, and per-particle defocus and per-group CTF parameter optimization turned on, yielding a ∼2.9 Å map, based on a Fourier Shell Correlation (FSC) cutoff of 0.143 (Table B.1).

Aldolase particles were picked with cryoSPARC Blob Picker with a particle diameter of 100–110 Å, an NC score > 0.12 and a power score range of 409 to 658. 820,058 particles were extracted with a box size of 200 px, binned to 100 px, and 2D classification was performed with fifty classes. Subset selection was performed to yield 389,144 particles. Ab initio reconstruction with two classes, C1 symmetry, and the initial resolution set to 8 Å was performed. The "good" reconstruction was then further refined with non-uniform refinement using D2 symmetry, the initial low-pass resolution set to 8 Å, and per-particle defocus and per-group CTF parameter optimization turned on. This reconstruction was input into a heterogenous refinement job, along with duplicates of the "junk" class yielded during Ab initio reconstruction. All extracted particles were input and the refinement was run with D2 symmetry. 428 k particles were sorted into the "good" class as the output of the heterogenous refinement job. These particles were re-extracted without binning and further refined using non-uniform refinement with D2 symmetry, the initial low-pass resolution set to 8 Å, and hard classification enforced. Further exposure curation was performed to apply a defocus cutoff of 1.3 μm, a CTF fit resolution cutoff of 3.5 Å, a relative ice thickness cutoff of 1.1, and a defocus tilt angle cutoff of 10 degrees. 622 exposures and 185,420 particles remained after exposure curation. A reconstruction with the resulting particles was further refined with non-uniform refinement with D2 symmetry, the initial low-pass resolution set to 6 Å, and per-particle defocus and per-group CTF parameter optimization turned on, yielding a ∼2.9 Å map, based on a Fourier Shell Correlation (FSC) cutoff of 0.143 (Table B.1).

3.5.5.2 Contracted CCMV.

The image processing pipeline is shown in (Fig B.3). 292 micrographs were collected. Motion correction and CTF correction were performed in cryoSPARC live v4.6.2. CCMV particles were initially picked with cryoSPARC live blob picker using a diameter of 100–500 Å. After initial 2D classification in cryoSPARC live, the best two 2Ds were used as templates for live template picker with the particle diameter set to 280 Å, the minimum separation distance set to 0.2 diameters, the NC score set to > 0.3, and the power score range of -86 to 2,775. 7,994

particles were extracted with a box size of 500 px, binned to 100 px, and 2D classification was performed with fifty classes and a circular mask of 280 Å. Ab initio reconstruction with two classes, C1 symmetry, and the initial resolution set to 20 Å was performed. The "good" class containing 5,397 particles was further refined using non-uniform refinement with icosahedral symmetry, the initial low-pass resolution set to 20 Å, and per-particle defocus and per-group CTF parameter optimization turned on. These particles were re-extracted without binning and further refined using non-uniform refinement with icosahedral symmetry, the initial low-pass resolution set to 20 Å, and per-particle defocus and per-group CTF parameter optimization turned on, yielding a ∼3.2 Å map, based on a Fourier Shell Correlation (FSC) cutoff of 0.143 (Table B.1).

3.5.5.3 Expanded CCMV.

The image processing pipeline is shown in (Fig B.4). 595 micrographs were collected. A resolution CTF cutoff of 6 Å was applied. 342 micrographs remained after preprocessing. Motion correction and CTF correction were performed in cryoSPARC live v4.6.2. CCMV particles were initially picked with cryoSPARC live blob picker using a diameter of 200–500 Å. After initial 2D classification in cryoSPARC live, the best four 2Ds were used as templates for live template picker with the particle diameter set to 310 Å, the minimum separation distance set to 0.2 diameters, the NC score set to > 0.257, and the power score range of 1,030 to 1,839. 62,017 particles were extracted with a box size of 560 px, binned to 100 px, and 2D classification was performed with fifty classes and a circular mask of 370 Å. Ab initio reconstruction on a selected subset of 2269 particles with 2 classes, C1 symmetry, and the initial resolution set to 20 Å was performed. The "good" class containing 1,885 particles was further refined using non-uniform refinement with icosahedral symmetry. These particles were re-extracted without binning and further refined using non-uniform refinement with icosahedral symmetry to yield a ∼6.9 Å map, based on a Fourier Shell Correlation (FSC) cutoff of 0.143 (Table B.1).

3.5.5.4 Expanded CCMV + low pH buffer / gold (700 ms).

The image processing pipeline is shown in (Fig B.5). 6,007 micrographs were collected. A resolution CTF cutoff of 12 Å was applied. 3,050 micrographs remained after preprocessing. Motion correction and CTF correction were performed in cryoSPARC live v4.6.2. CCMV particles were picked with cryoSPARC live blob picker using a diameter of 275–325 Å, an NC score > 0.156 and a power score range of -942 to 2,434. 68,810 particles were extracted with a box size of 560 px, binned to 100 px, and 2D classification was performed with fifty classes and a circular mask of 390 Å. Ab initio reconstruction on a selected subset of 6,952 particles with 1 class, C1 symmetry, and the initial resolution set to 20 Å was performed. The reconstruction was further refined using non-uniform refinement with icosahedral symmetry and the initial low-pass resolution set to 15 Å. These particles were re-extracted without binning and further refined using non-uniform refinement with icosahedral symmetry, the initial low-pass resolution set to 6 Å, and per-particle defocus and per-group CTF parameter optimization turned on, yielding a ∼3.9 Å map, based on a Fourier Shell Correlation (FSC) cutoff of 0.143 (Table B.1).

3.5.5.5 GroEL + GroES, no ATP.

The image processing pipeline is shown in (Fig B.6). 601 micrographs were collected. A resolution CTF cutoff of 7.5 Å was applied. 376 micrographs remained after preprocessing. Motion correction and CTF correction were performed in cryoSPARC live v4.6.2. Particles were picked in cryoSPARC with Topaz using the ResNet16 pretrained dataset, an estimated particle diameter of 240 Å, and a particle threshold of -5. The picks were selected based on an NC score > -15.56 and a power score range of -30 to 143. 60,116 particles were extracted with a box size of 420 px, binned to 100 px, and initial 2D classification was performed with fifty classes. Extensive 2D classification was performed iteratively in parallel to confirm the absence of capped GroEL. Classes resembling GroEL from the initial 2D classification were selected (10,969 particles) and an ab initio reconstruction with 2 classes and C1 symmetry, and the initial resolution set to 15 Å was performed. Both classes were used as inputs for a heterogenous refinement job with 17,477 particles from a less stringent 2D class subset selection, the initial low-pass resolution set to 12 Å, and hard classification enforced. The 11,859 particles sorted into the "good" class were then re-extracted without binning and the reconstruction was further refined using non-uniform refinement with D7 symmetry, the initial low-pass resolution set to 6 Å, and per-particle defocus and per-group CTF parameter optimization turned on, yielding a ∼3.2 Å map, based on a Fourier Shell Correlation (FSC) cutoff of 0.143 (Table B.2).

3.5.5.6 GroEL + GroES + ATP/gold (100 ms)

The image processing pipeline is shown in (Fig B.7). 2,645 micrographs were collected. A CTF cutoff of 6 Å, relative ice thickness cutoff of 1.1, total motion cutoff of 50 pixels, and defocus range cutoff of 110 Å was applied. 1,434 micrographs remained after preprocessing. Motion correction and CTF correction were performed in cryoSPARC live v4.6.2. Particles were picked in cryoSPARC with Topaz using the ResNet16 pretrained dataset, an estimated particle diameter of 240 Å, and a particle threshold of -5. The picks were selected based on an NC score > 0 and a power score of 25–120. 119,235 particles were extracted with a box size of 420 px, binned to 100 px, and initial 2D classification was performed with fifty classes. Classes resembling GroEL were selected for a total of 11,602 particles. Extensive 2D classification was performed iteratively to confirm the absence of capped GroEL. Ab initio reconstruction with C1 symmetry was performed with 1 and 2 classes and the initial resolution set to 15 Å in parallel. The reconstruction from the 1-class job, which corresponds to the single-capped GroEL in the "bullet" conformation, was further refined using non-uniform refinement with C7 symmetry and the initial low-pass resolution set to 20 Å. The reconstructions were further refined in independent non-uniform refinement jobs with D7 symmetry and the initial low-pass resolution set to 15 Å. The refined "best" class from the 2-class ab initio and refined class from the 1-class ab initio were used as inputs for heterogenous refinement (C1 symmetry, enforced hard classification, initial resolution 12 Å) containing 17,663 particles resulting from a less stringent 2D classification subset selection. After sorting, the 11.9k particles from the best reconstruction were re-extracted without binning and further refined using non-uniform refinement with optimized per-particle defocus and per-group CTF parameters, D7 symmetry and an initial low-pass resolutions of 6 Å. The resulting reconstruction was resolved at ∼3.4 Å,

based on a Fourier Shell Correlation (FSC) cutoff of 0.143 (Table B.2).

3.5.5.7 GroEL + GroES + ATP/gold (300 ms).

The image processing pipeline is shown in (Fig B.8). 5,301 micrographs were collected. A resolution CTF cutoff of 7.5 Å was applied. 3,665 micrographs remained after preprocessing. Motion correction and CTF correction were performed in cryoSPARC live v4.6.2. Particles were picked in cryoSPARC with Topaz using the ResNet16 pretrained dataset, an estimated particle diameter of 240 Å, and a particle threshold of -5. The picks were selected based on an NC score > 27.16 and a power score range of 30 to 149. 255,398 particles were extracted with a box size of 420 px, binned to 100 px, and initial 2D classification was performed with fifty classes. Classes resembling GroEL were selected for a total of 53,790 particles. Ab initio reconstruction with C1 symmetry was performed with 1 and 2 classes and the initial resolution set to 15 Å in parallel. The reconstruction from the 1-class job, which corresponds to the single-capped GroEL in the "bullet" conformation, was further refined using non-uniform refinement with C7 symmetry and the initial low-pass resolution set to 15 Å. Further rounds of 2D classification revealed a class corresponding to uncapped GroEL in the "barrel" conformation containing 2,230 particles, as well as a class corresponding to a double-capped GroEL in the "football" conformation with 761 particles. Both particle subsets were reconstructed independently using Ab initio reconstruction with one class, C1 symmetry, and the initial resolution set to 20 Å. The reconstructions were further refined in independent non-uniform refinement jobs with D7 symmetry and the initial low-pass resolution set to 15 Å. The refined reconstructions of the bullet, barrel, and football conformations, along with a "junk" class from the initial 2-class Ab initio job were used as inputs for heterogenous refinement (C7 symmetry, enforced hard classification, initial resolution 12 Å) containing the 53,790 particles resulting from the initial 2D classification subset selection. After sorting, 14.8k particles were assigned to the barrel conformation, 7.7k to the football, and 25.6k to the bullet. Particles from each conformation were re-extracted without binning and the reconstructions were further refined using non-uniform refinement with optimized per-particle defocus and per-group CTF parameters, using D7, D7, and C7 symmetry and initial low-pass resolutions of 6 Å, 8 Å, and 6 Å for the barrel, football, and bullet conformations, respectively. The resulting reconstructions were resolved at ∼3.6 Å for the barrel conformation, ∼4.8 Å for the football conformation, and ∼3.3 Å for the bullet conformation, based on a Fourier Shell Correlation (FSC) cutoff of 0.143 (Table B.3).

3.5.5.8 GroEL + GroES + ATP/gold (700 ms)

The image processing pipeline is shown in (Fig B.9). 6,063 micrographs were collected. A resolution CTF cutoff of 7.5 Å was applied. 2,535 micrographs remained after preprocessing. Motion correction and CTF correction were performed in cryoSPARC live v4.6.2. Particles were picked in cryoSPARC with Topaz using the ResNet16 pretrained dataset, an estimated particle diameter of 240 Å, and a particle threshold of -5. The picks were selected based on an NC score > 25.96 and a power score range of 47 to 139. 118,543 particles were extracted with a box size of 420 px, binned to 100 px, and initial 2D classification was performed with fifty classes. Classes resembling GroEL were selected for a total of 15,977 particles. The

data were processed as described in the 300 ms dataset above with the following exceptions: ab initio reconstructions were performed with the initial resolution set to 20 Å, except for the bullet ab initio (15 Å). After sorting with heterogenous refinement, 3.6k particles were assigned to the barrel conformation, 2.1k to the football, and 8.8k to the bullet. Particles from each conformation were re-extracted without binning and further refined as described in the 300 ms dataset. The resulting reconstructions were resolved at ∼4.7 Å for the barrel conformation, ∼7.1 Å for the football conformation, and ∼3.5 Å for the bullet conformation, based on a Fourier Shell Correlation (FSC) cutoff of 0.143 (Table B.4).

4

Time-resolved studies of polymerase theta-mediated DNA binding

4.1 Chapter overview

Extending the temporal scope of this thesis beyond the microsecond MD simulations of LONP1 and the millisecond time-resolved cryo-EM methodology development of Mix-it-up, this chapter investigates conformational heterogeneity on the minute timescale by probing early interactions of the DNA repair enzyme, polymerase theta (PolθHel), with DNA using time-resolved cryo-EM. The helicase domain of Polθ (PolθHel) plays a key role in recognizing and rectifying DNA double strand breaks, yet the initial stages of its DNA binding remain poorly understood. As a molecular machine central to error-prone DNA repair pathways, Polθ's upregulation in cancer cells deficient in canonical high-fidelity repair pathways makes it an attractive target for structural studies. However, mechanistic insight into theta-mediated repair has been largely constrained to post-equilibrium states, leaving unresolved questions about how its oligomeric state regulates DNA interactions. Previous cryo-EM studies suggested that DNA binding to PolθHel triggers a conformational change that disrupts the tetrameric assembly to yield functional dimers, but they were unable to clarify the role of tetramers in DNA binding. Although the temporal window explored here is longer than in previous chapters, it proves well-suited to capture early binding events and conformational ensembles that were previously inaccessible due to lengthy (hour-scale) preparation procedures used in prior structural studies of the PolθHel-DNA complex. The time-resolved data presented here raise new insights into the potential involvement of the tetrameric PolθHel species in initial DNA recognition, building a more complete picture of the conformational landscape of Polθ and its engagement with DNA.

Contributions This work was a collaborative effort between myself (Lauren Alexandrescu), Noah Mathew, Gabriel C. Lander. LA and GCL conceptualized and led the project. LA and NM performed sample purification and negative stain data collection and analysis. LA and NM performed cryo-EM sample preparation. LA performed all of the cryo-EM data collection. LA and NM performed cryo-EM structure determination. LA analyzed and interpreted the results with guidance from GCL. LA prepared the figures and wrote the chapter.

4.2 Introduction

4.2.1 DNA double-strand break repair

4.2.1.1 Repair pathways

Up to 50 DNA double-strand breaks (DSBs) occur in each cell of the human body per day [159]. These breaks can be incurred by UV radiation damage, reactive oxygen species or free radicals, and dysfunction in DNA replication and transcription pathways, among other sources [160, 161]. Ultimately, DSBs lead to genome instability and require repair [160, 162]. Errors or mutations made during DNA repair can be propagated through cells during division and proliferation, which can ultimately lead to cancer [160, 163]. Multiple pathways exist to repair DNA DSBs (Fig 4.1), with the most dominant being non-homologous end-joining (NHEJ) and homologous recombination (HR) [161, 162]. These high-fidelity pathways are responsible for repairing the vast majority of DSBs, with pathway choice depending on the stage of the cell cycle and the accessibility and structure of the double-strand ends [161]. More error-prone pathways, such as microhomology-mediated end-joining (MMEJ), also exist as backup mechanisms when NHEJ or HR pathways malfunction [163, 164].

NHEJ repairs DNA DSBs in the absence of a homologous template (Fig 4.1) [161]. Despite exhibiting a higher error rate than homologous recombination (HR), NHEJ is the dominant DSB repair pathway in cells, due to its accessibility in both non-dividing and proliferating cells [161, 165]. The pathway is initiated through the binding of the Ku complex (Ku70-Ku80) to blunt, double-stranded ends of DNA, or those with short single-stranded overhangs, at the break point of the DNA [161, 165]. Recruitment of NHEJ cofactors by Ku, including DNA-PKcs, enables DNA alignment and, in some cases, end processing by the Artemis nuclease or DNA polymerases to trim single-strand overhangs of DNA that may be present in order to "clean up" the ends prior to ligation by the DNA ligase IV complex (LIG4, XRCC4, and XLF) [161, 165–167]. Short stretches of complementary DNA, known as microhomologies, can also be used to guide ligation when exposed on single-stranded overhangs, though NHEJ has a limited capacity for base pairing (up to 4 base pairs) [168]. While NHEJ does not require microhomology for repair, its presence can improve the precision of the pathway [161]. Mismatches within the microhomology, however, can lead to lower fidelity repairs and can result in insertions or deletions that may give rise to genomic mutations or chromosomal translocations, though these consequences are more pronounced in lower fidelity repair pathways, such as MMEJ [161, 163].

HR is the second most common DNA DSB repair pathway and is considered the highest fidelity mechanism, due to its use of a homologous DNA template to mediate repair (Fig 4.1)

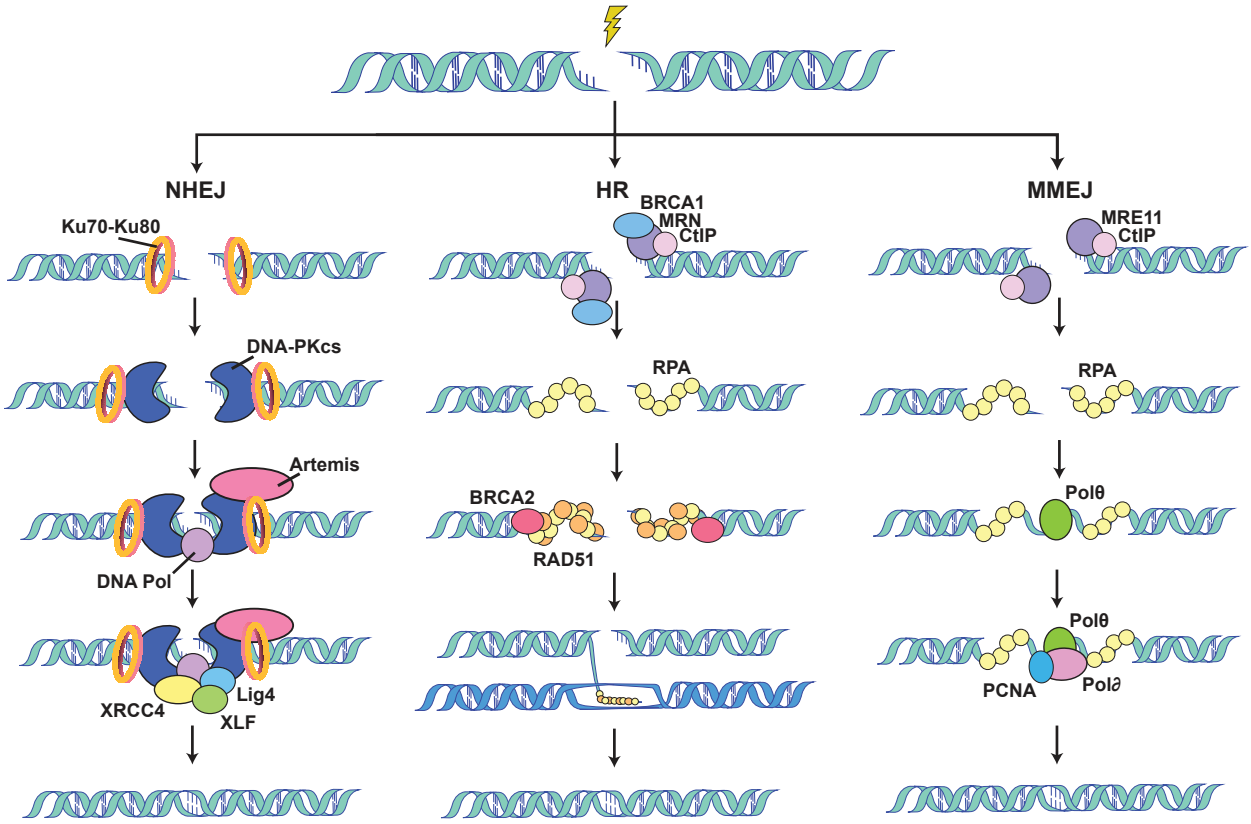


Figure 4.1: Major repair pathways of DNA double-strand breaks (DSBs) [161]. A) Non-homologous end-joining (NHEJ). NHEJ is the dominant DSB repair pathway, accessible throughout the cell cycle. In the first stage of repair, the Ku complex (Ku70-Ku80, orange rings) binds to blunt or short overhang DNA ends. Ku then recruits DNA-PKcs (dark blue), which stabilize the DNA ends and align them for repair. In some cases, the Artemis nuclease (pink oval), activated by the phosphorylation of DNA-PKcs, and DNA polymerase (lavender) are recruited for end processing. In the final step, DNA ends are ligated by a complex including Ligase IV (blue), XLF (green), and XRCC4 (yellow), repairing the DNA double strand break. **B) Homologous recombination (HR).** HR represents the highest fidelity repair mechanism, restricted to S and G_2 cell cycle phases due to its reliance on sister chromatid availability. Repair is initiated when CtBP-interacting protein (CtIP, pink), in cooperation with BRCA1 (blue), and the MRE11-RAD50-NBS1 (MRN, purple) complex binding to DNA ends, where MRE11 creates 5' strand nicks and performs end-resection to generate 3' overhangs up to 300 base pairs in length. Replication protein A (RPA, yellow) then coats the exposed single-stranded DNA to prevent degradation. These proteins are subsequently displaced by Rad51 (orange), mediated by BRCA2 (red) to form nucleoprotein filaments capable of strand invasion into homologous sister chromatid templates (blue DNA) for high-fidelity repair synthesis. **C) Microhomology-mediated end-joining (MMEJ).** MMEJ serves as a low-fidelity backup pathway upregulated when NHEJ or HR are deficient. Following DSB formation, CtIP (pink) and MRE11 (lavender) resect 5' DNA ends to expose microhomology sequences (typically 2–6 base pairs). RPA (yellow) then binds to the single-stranded DNA to protect it from degradation. Polymerase theta (Polθ, green) displaces RPA and recognizes these microhomologous regions, aligning and annealing the complementary 3' strand ends, followed by performing initial gap-filling DNA synthesis (< 15 nt). Polδ (pink) then performs the remaining strand synthesis, in complex with the proliferating cell nuclear antigen (PCNA, blue) processivity clamp, repairing the DSB.

[161]. Unlike NHEJ, which can ligate DNA ends directly, HR is reliant on exonuclease activity to carry out end-resection of the 5' ends of the double-stranded DNA, exposing single-stranded 3' ends that can invade the DNA duplex of a sister chromatid to guide high-fidelity repair using a homologous template [169]. The pathway is initiated by CtBP-interacting protein (CtIP) and the MRE11-RAD50-NBS1 (MRN) complex binding to DNA ends [170], with BRCA1 facilitating CtIP recruitment and promoting end-resection [171]. MRE11 creates 5' strand nicks and performs end-resection to generate 3' overhangs up to 300 base pairs in length [161]. Replication protein A (RPA) then coats the exposed single-stranded DNA to prevent degradation and secondary structure formation [161]. BRCA2 mediates the displacement of RPA by Rad51, which forms nucleoprotein filaments capable of strand invasion into the sister chromatid for template-guided repair synthesis [161, 171]. Despite its high fidelity, HR is restricted to S and G_2 cell cycle phases due to its reliance on sister chromatid availability [161].

MMEJ is a low-fidelity DNA DSB repair pathway that is upregulated in cells deficient in NHEJ or HR [161, 163, 172]. The pathway relies heavily on sequences of microhomology (typically 2–6 base pairs in length) to mediate repair [173]. Following a DSB, the 5' ends of the DNA are resected by the exonuclease activity of the MRN complex mediated by CtIP, generating two 3' single-stranded DNA overhangs and exposing regions of microhomology [174]. These microhomologies are recognized by polymerase theta (Polθ), which displaces the RPA that coats the ssDNA following end-resection [175]. While there is some evidence to suggest that the process of RPA displacement by Polθ is mediated by ATP binding and hydrolysis within the helicase domain (PolθHel), FRET data demonstrates that the process of microhomology recognition by PolθHel is ATP-independent [176, 177]. After RPA displacement, Polθ aligns the microhomology at the ends of the two 3' strands to anneal them [163, 178, 179]. Subsequently, the polymerase domain of Polθ performs gap-filling to synthesize the new 5' ends, completing the repair of the DSB [180]. Due to its reliance on microhomology sequences for DNA repair, MMEJ is prone to mismatches and errors that can be propagated to proliferating cells, leading to genomic instability [163]. Indeed, Polθ is upregulated in several cancers, including breast, ovarian, and lung cancer, and knockdown of Polθ results in substantial loss of tumor cells in those that are HR-deficient or contain BRCA1/2 mutations [163, 172]. As a result, Polθ has become an attractive target for anti-cancer therapeutics.

4.2.1.2 Structural insights into DNA-bound polymerase theta

Until recently, the role of Polθ in microhomology recognition remained unclear due to the absence of DNA-bound structures [163]. Recent cryo-EM studies determined the structure of PolθHel bound to DNA, revealing its mechanism for positioning and annealing microhomologies on single-stranded DNA (ssDNA) during DSB repair [178, 179]. Three conformational states were determined of the PolθHel-DNA complex, corresponding to microhomology "searching", "aligning", and "annealing" states.

In the absence of DNA, PolθHel exists in both dimeric and tetrameric forms, with the tetramer assembling as a "dimer of dimers" [178, 179, 181, 182]. DNA binding appears to drive tetrameric disassembly through a large conformational rearrangement involving the

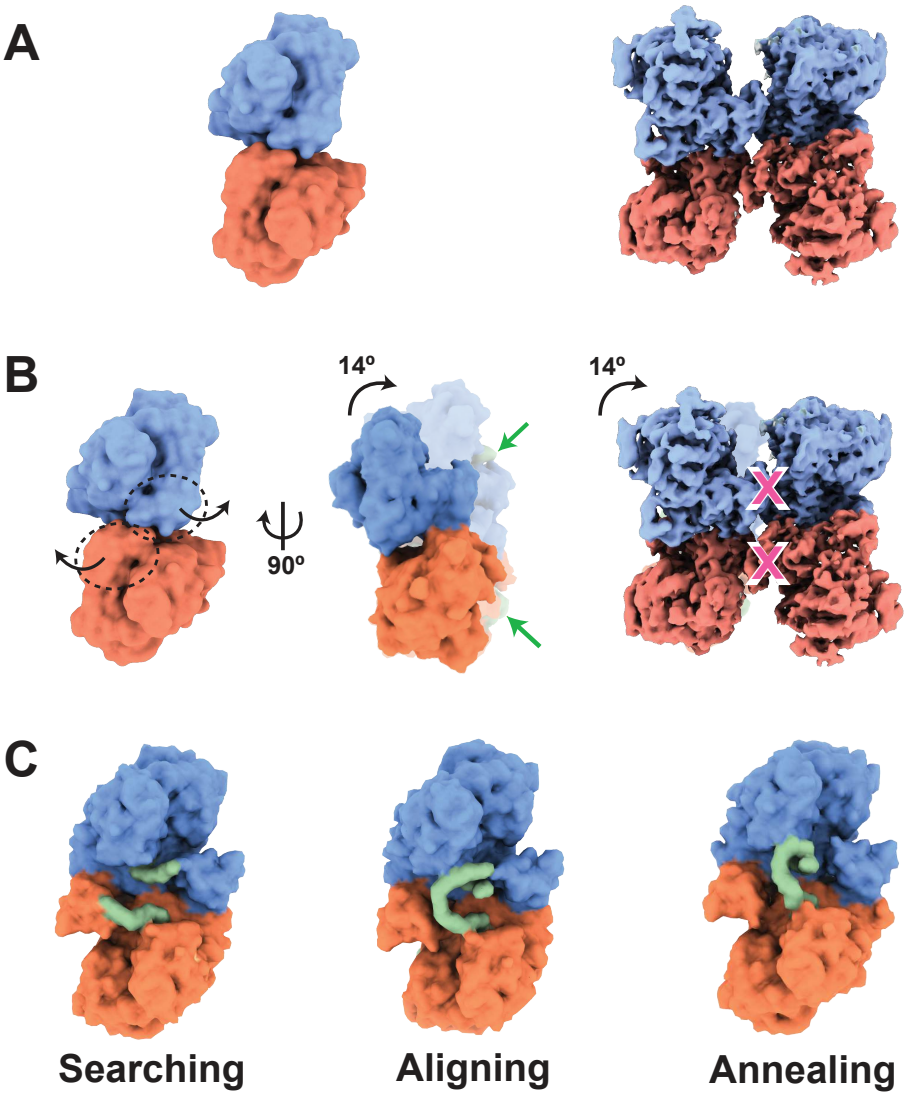


Figure 4.2: Structural basis of Polθ-mediated microhomology recognition and annealing [178, 179] (A) Apo structures of the helicase domain of Polθ (PolθHel) show both dimeric (left) and tetrameric (right) assemblies, with the tetramer adopting a "dimer of dimers" organization. **(B)** Cryo-EM reconstructions of PolθHel bound to DNA reveal that DNA binding induces a large conformational rearrangement of the C-terminal D5 domains (residues 790–894, dashed circles) (left), creating space for 3' single-stranded DNA ends at the protomer interface shown in **(C)**. The ~14° rotation of intraprotomer angle and the rearrangement of the D5 domains upon DNA binding (transparent dimer, middle) sterically block the dimer–dimer interface (pink X's) of the tetramer (right), consistent with DNA binding promoting tetramer disassembly. Entry points of ssDNA ends into the PolθHel DNA tunnel are indicated by green arrows. **(C)** Three distinct DNA-bound conformations were observed by Zerio et al. [178] upon binding of PolθHel to DNA (green): 1) A "searching" state (left), in which ssDNA ends are sequestered by the helicase domain, 2) an "aligning" state, where 3' overhangs are positioned in close proximity, and 3) an "annealing" state, in which dsDNA density is resolved at the protomer interface, consistent with successful microhomology pairing and strand annealing.

C-terminal D5 domains (residues 790–894) and expansion of the intra-protomer angle by ∼14° [178]. This rearrangement creates steric clashes that block the dimer-dimer interface, helping to explain the preference for the dimeric state in the presence of DNA [178, 179]. In the DNA-bound structures, density corresponding to 3' ssDNA ends was found at the protomer interface and could be traced through the DNA tunnels of both protomers, with double-stranded DNA (dsDNA) density observed at one tunnel entry point [178, 179]. The ssDNA ends in complex with PolθHel adopt three distinct conformational states. First, a "microhomology searching" state sequesters ssDNA ends and positions them in close proximity to the repositioned D5 domains, as the enzyme searches for microhomologous sequences on the 3' ends [178]. In the next stage of repair, a "microhomology aligning" state brings the 3' ends together for alignment. Finally, a "microhomology annealing" state, containing density corresponding to double-stranded DNA at the protomer interface, indicates successful strand annealing. Despite the large population of tetrameric PolθHel species in the absence of DNA, tetramers have not previously been observed bound to DNA, suggesting that the dimeric DNA-bound complexes reflect a post-equilibrium conformational ensemble. Thus, it is unclear whether DNA binding proceeds or follows tetrameric disassembly, and the early interactions of DNA and PolθHel required to form the DNA-bound complex remain unknown.

4.3 Results

4.3.1 Establishing a workflow for time-resolved studies of Polθ

DNA-binding proteins are typically stored under high-salt concentrations, but DNA binding is significantly weakened under these conditions due to competition at the DNA-binding pockets [181]. To address this challenge, previous workflows purified and stored PolθHel in high salt conditions (250–500 mM NaCl) and prepared DNA-bound complexes by adding DNA oligonucleotides at high salt, then gradually reducing the salt concentration to 100 mM NaCl over a long period (hour-scale) to promote DNA binding while maintaining protein stability (see *Methods*) [178, 179]. This multi-hour procedure, however, precluded time-resolved studies of transient intermediate states. Therefore, we sought to determine the minimum salt concentration at which PolθHel remained stable in solution while retaining the ability to bind DNA rapidly upon addition. This would enable direct observation of early DNA-binding states without the confounding effects of lengthy complex formation procedures.

To this end, we performed a series of negative stain experiments to test the stability of PolθHel at a range of salt concentrations, including 100 mM, 150 mM, 200 mM, and 250 mM NaCl (Fig 4.3). As expected, the high salt (250 mM) condition showed intact particles, primarily representing the dimeric state (∼86% of particles) with few classes corresponding to tetramers (∼14% of particles) (Fig 4.3A). At 200 mM NaCl, the protein structure remained preserved, and we were able to easily identify dimeric classes, as well as tetrameric classes (Fig 4.3B). At 150 mM NaCl, PolθHel was still structurally intact, with no signs of aggregation on the grid. Interestingly, the 50/50 balance of tetramers to dimers in the absence of DNA that had previously been reported was not observed [178]. Instead, the majority of the classes reflected the tetrameric state (∼76% of particles), with few dimeric classes (∼24% of particles)

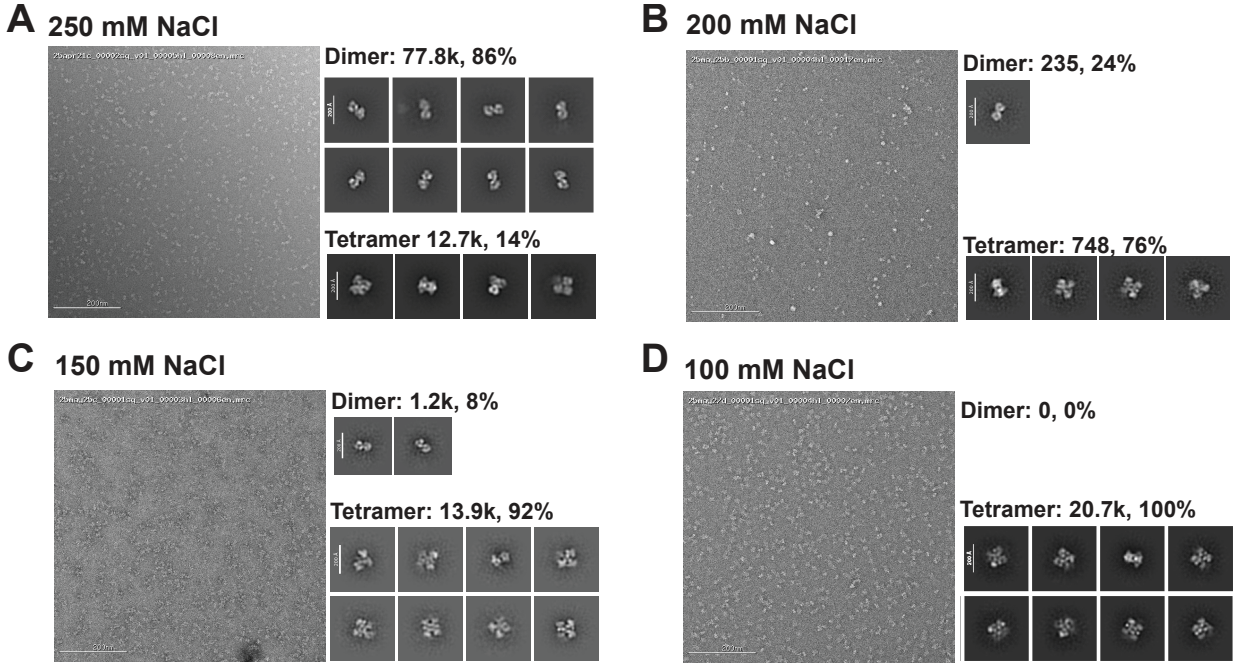


Figure 4.3: Negative stain PolθHel stability tests under varying salt conditions. Representative micrograph and 2D class averages of PolθHel at 250 mM **(A)**, 200 mM **(B)**, 150 mM **(C)**, and 100 mM NaCl **(D)**. **A)** 250 mM conditions show primarily dimeric classes (∼86% of the particles) with ∼14% of the particles reflecting the tetrameric state. **B)** 200 mM NaCl conditions resulted in ∼24% dimeric particles and ∼76% tetramers. **C)** 150 mM conditions resulted in ∼8% dimeric particles and 92% tetramers. **D)** Under 100 mM NaCl conditions, no dimeric particles were observed, with all PolθHel 2D class averages reflecting the tetrameric species. Full 2D class average data for all salt conditions is shown in *Appendix C* (Fig C.1).

observed (Fig 4.3C). This trend continued at 150 mM salt conditions, shifting the balance of oligomeric states further in favor of the tetramer, with ∼8% of particles corresponding to the dimeric state and ∼92% of particles adopting the tetrameric state. To our surprise, PolθHel was also stable at our lowest salt condition (100 mM NaCl), and we did not observe aggregation of particles on the grid (Fig 4.3D). Further, all of the classes represented the tetrameric state (100% of particles), with no dimers observed. Together, these data indicated that the protein could be prepared at low-salt conditions prior to DNA addition, laying the groundwork for studies of early interactions between PolθHel and DNA.

We next confirmed that the low-salt (100 mM) PolθHel samples maintained structural integrity in cryo-EM. Consistent with our negative stain results, we only observed the tetrameric state under these conditions (Fig 4.4B). This is in stark contrast to the control dataset that we collected of PolθHel under high-salt conditions (250 mM NaCl) (Fig 4.4A), which showed roughly 50% apo dimers and 50% tetramers, as was also reported by Zerio et al. [178]. Our structures from both the 100 mM NaCl and 250 mM NaCl datasets are in good agreement (Fig 4.4A,B), demonstrating that though the salt concentration appears to alter the distribution of oligomeric states, the states themselves do not exhibit distinct conformations.

4.3.2 Early conformational ensembles of the Polθ helicase domain reflect tetrameric and DNA-bound dimeric states.

We next investigated early-stage structural rearrangements in PolθHel-DNA binding to assess the biological relevance of the tetrameric state. We prepared the PolθHel-DNA complex and froze samples at five and ten minutes after DNA addition, hypothesizing that these early conformational ensembles would differ from the equilibrium complex frozen several hours after DNA addition. Indeed, five minutes following DNA addition, we observed a significant population of tetramer particles (∼48%) alongside the expected DNA-bound dimeric states that had previously been observed in our control dataset and were reported by Zerio et al. and Ito et al. (Fig 4.4D,C) [178, 179]. Ten minutes after DNA addition, the tetrameric population falls to (∼5%), indicating that the system had reached a near-equilibrium state (Fig 4.4E). The DNA-bound dimers are readily observable in 2D class averages, as the particles appear more rectangular than the "hourglass"-shaped particles observed in the absence of DNA (Fig C.3, C.5, C.6). Additionally, weak density corresponding to the flexible D5 domains that mediate DNA binding can be observed flanking the equator of the dimer. While previously only the dimeric form had been observed upon DNA addition [178], or alongside a small population of tetramers (∼6%) [179], our five-minute timepoint reveals an approximately 50/50 balance of tetramers and DNA-bound dimers, suggesting that the tetrameric state represents a significant early intermediate in the DNA-binding process that is rapidly lost during conventional complex formation procedures.

In our control tetramer structure (DNA-free), we observed additional density flanking either side of the DNA entry tunnel in one of the subunits, as well as on one side of the tunnel in the complementary subunit of the same dimer within the tetramer (Fig 4.5). This density was not present in the tetramer structure deposited by Zerio et al. (EMDB: 44534), nor in the structure of the tetramer obtained at 100 mM NaCl in the absence of DNA (Fig 4.5B,

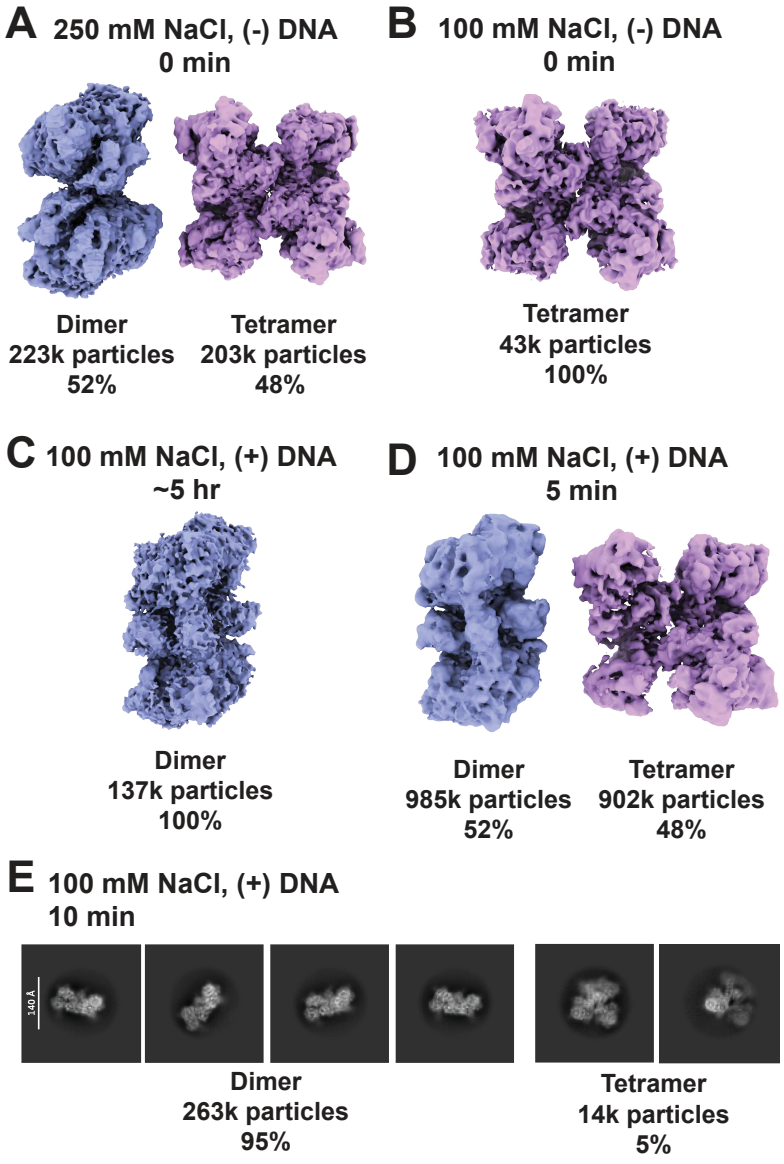


Figure 4.4: PolθHel particle distributions in cryo-EM. A-B) Particle distributions of PolθHel observed in the absence of DNA under varying salt conditions. **A)** At high-salt conditions (250 mM NaCl), PolθHel adopted both dimeric (blue) and tetrameric (lavender) classes, with ~52% of particles corresponding to the former and ~48% of particles corresponding to the latter. **B)** Under low salt conditions (100 mM NaCl), PolθHel exclusively adopted the tetrameric state (100% of particles). **C-E)** Particle distributions of PolθHel observed in the presence of DNA under low-salt conditions at varying timepoints. **C)** Using conventional complex formation workflows (~5 hr), PolθHel particles reflected 100% DNA-bound dimers. **D)** 5 min following DNA addition, ~52% of particles adopted a DNA-bound dimeric state, while ~48% reflected a tetrameric state. **E)** Under low salt conditions, 10 min following DNA addition, 95% of particles adopted a DNA-bound dimeric state, while only ~5% reflected a tetrameric state. Representative 2D class averages are shown here, and the full dataset can be found in *Appendix C* (Fig C.6).

4.5C) [178].

To determine whether the density aligns with the expected position of DNA at this site, we docked the protomer of the DNA-bound dimeric model (PDB:9BHA) containing the dsDNA hairpin into the protomer of our tetrameric structure containing putative DNA density (Fig 4.5D). The alignment positioned the modeled dsDNA hairpin at the site of the extra density, though the low resolution of this region limits our interpretation (Fig 4.5D,E). The modeled DNA oligo differs from the design of that used in our experiments, which may lead to differences in the positioning of the double-stranded region relative to the protomer (see *Methods*). While the oligo used in the structural determination of the DNA-bound dimer from Zerio et al. [178] contained several mismatched base pairings leading to a kink in the double-stranded region, as well as a 5' overhang, the oligo used in this study lacks the additional secondary structure in the double-stranded portion and does not contain a 5' overhang. These discrepancies between our studies may account for the differences in the binding state of the DNA at the entrance to the DNA tunnel of PolθHel. However, despite the strong density observed at the tunnel entrance in our structure, no density is observed within the DNA tunnel (Fig 4.5F), indicating that single-stranded DNA has not entered the PolθHel protomer in this state.

To further classify potential subspecies, we performed 3D classification of the tetrameric particles from the five-minute dataset without alignment. Interestingly, several states of the putative DNA density were observed (Fig 4.6). Eight of the classes (80.4% of the particles) contain density flanking either end of the DNA tunnel, as observed in the consensus structure (Fig 4.6B, 4.5). One class (15.4% of the particles) contains density on one side of the tunnel, and only 4.2% of the particles are observed without any additional density. The heterogeneity observed in these early time-point structures supports our hypothesis that the tetrameric state serves as a dynamic platform for initial DNA capture and positioning.

4.4 Discussion

In this study, early-stage interactions between PolθHel and DNA are explored using time-resolved cryo-EM. Our findings reveal novel insights into the mechanism of PolθHel-mediated DNA repair. We first confirm the stability of PolθHel under low salt conditions (under 150 mM NaCl) in the absence of DNA, providing a system primed for DNA binding and establishing an experimental workflow for time-resolved studies. Interestingly, we observe a correlation between salt concentration and the oligomeric state of PolθHel. At high salt concentrations (250 mM), PolθHel predominantly adopts a dimeric form, while decreasing salt concentrations shift the population towards the tetrameric state, with 100 mM NaCl yielding 100% tetrameric structures in cryo-EM. At 250 mM NaCl, cryo-EM reveals a more balanced distribution of dimers and tetramers ($\sim$52% and $\sim$48%, respectively) compared to negative stain data ($\sim$86% and $\sim$14%, respectively). The shifted distribution in negative stain may be due to inherent differences in sample preparation procedures, with cryo-EM reflecting a population more representative of the native, solution state. Together, these data may explain the discrepancies in oligomeric states of previously solved structural ensembles and suggest that changes in salt concentration may impact the functional activity of PolθHel

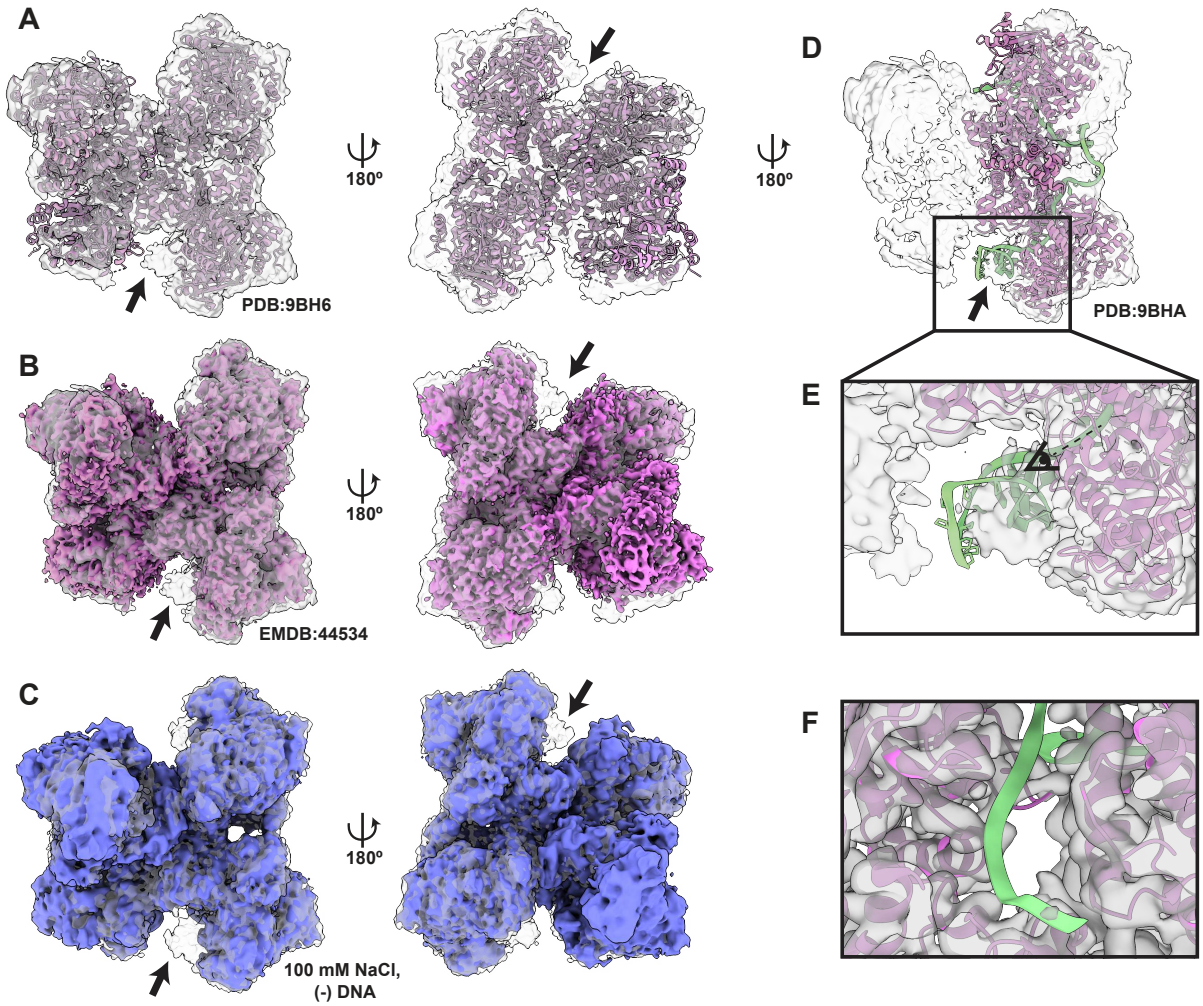


Figure 4.5: Putative DNA-Bound Tetramer. A-C) Alignment of the tetrameric model (PDB:9BH6, pink) **(A)** and structure (EMDB:44534, pink) **(B)** of PolθHel in the absence of DNA from Zerio et al [178], as well as our low-salt structure of PolθHel in the absence of DNA (indigo) **(C)** docked into our low-salt structure obtained five minutes after DNA addition (transparent). Black arrows indicate regions of extra density unique to the tetrameric structure obtained following DNA addition. Two views are shown with 180° rotation respective to one another. **D)** Docking of the model of the DNA-bound dimer (PDB:9BHA, pink) from Zerio et al. [178] into the structure obtained under 100 mM salt conditions 5 min after DNA addition (transparent). Alignment was focused on the protomer containing a dsDNA hairpin (green) and the protomer of the tetramer structure containing additional density. **E)** The inset shows the alignment of the DNA hairpin to the extra region of density shown in **(D)**. The viewpoint indicated is shown in **(F)** which highlights the lack of density in the DNA tunnel.

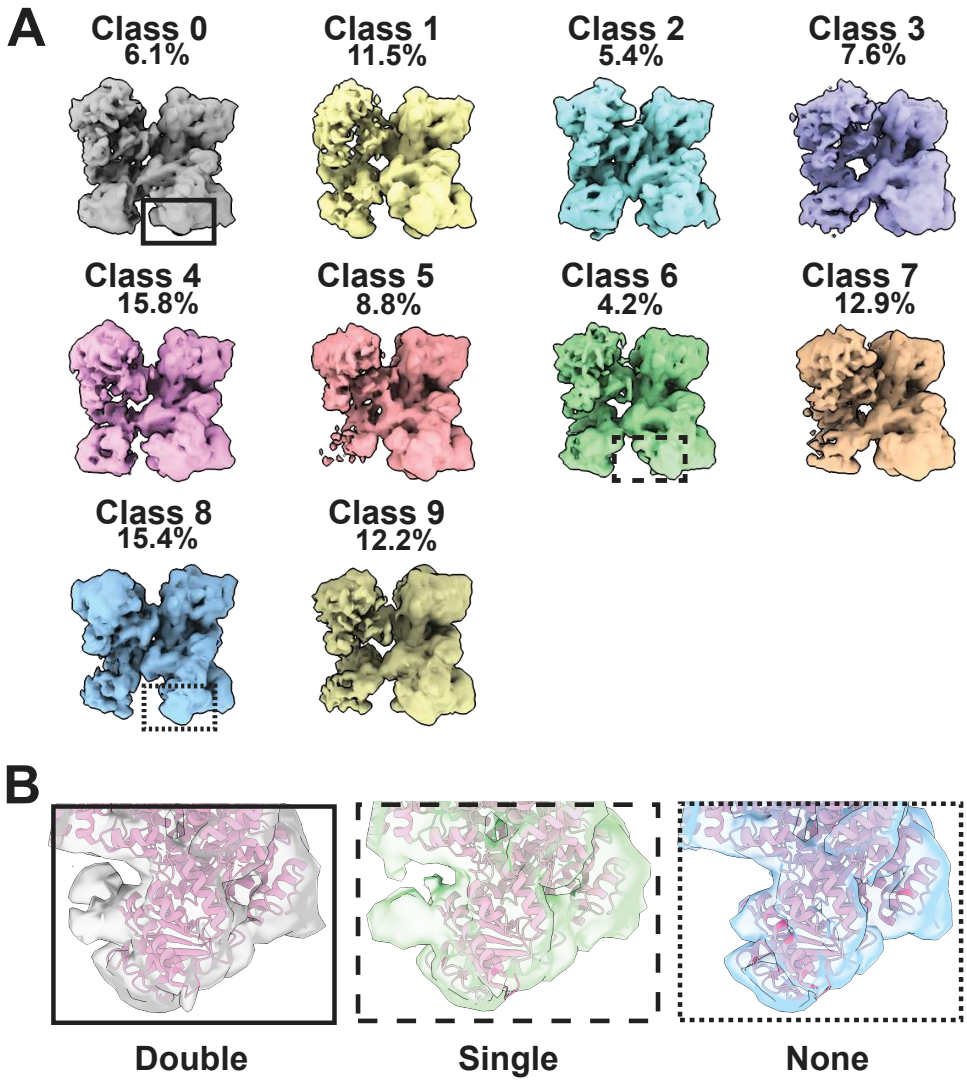


Figure 4.6: 3D Classification of putative DNA-bound tetramer. A) 3D classification of the DNA-bound tetramer five minutes following DNA addition performed without alignment. A solid, dashed, and dotted rectangle indicates the region shown in **(B)** for class 0, 6, and 8, respectively. **B)** Structures of class 0, 6, and 8 highlighting the bottom right protomer and the presence and absence of extra density near the entrance to the DNA tunnel with a tetramer model docked in (PDB:9BHA). Class 0–5 and class 7–9 (80.4% of the particles) showed two extra regions of density at this area, with class 0 used as a representative structure (left). Class 6 (4.2% of particles, middle) showed extra density flanking only one end of the DNA tunnel. Class 8 (15.4% of particles, right) showed no extra DNA density surrounding the DNA tunnel.

[178, 179, 181, 182]. Low salt concentrations could promote a tetrameric state to sequester and stabilize PolθHel monomers when there is a low abundance of 3' single-stranded DNA ends with exposed microhomology that require repair, potentially suppressing MMEJ in favor of higher-fidelity DNA repair pathways when the cell is not overwhelmed by DNA DSBs.

Next, we used time-resolved cryo-EM to identify early conformational ensembles of PolθHel in the presence of DNA. After five minutes, we observe a significant population of tetramers alongside the expected DNA-bound dimers. Previous studies observed negligible (6%) or no tetramers upon DNA addition, likely due to lengthy complex formation procedures (hours) that shift the conformational ensemble to a post-equilibrium state [178, 179]. After ten minutes, the tetrameric population significantly decreases to 5%, suggesting that the tetramer may represent an early binding intermediate prior to disassembly into catalytically competent dimers that align and anneal DNA for repair.

Further, we observed additional density at the entrance to the DNA tunnel in two protomers of the tetrameric state, while the remaining two protomers were resolved at lower resolution, suggesting increased flexibility that may correspond to destabilization prior to tetrameric disassembly. Since the DNA-binding site is positioned within $\sim$1 nm of the oligomerization interface, it is plausible that the electrostatic changes induced upon binding of the negatively charged phosphate backbone of DNA influence interactions at the dimer-dimer interface, promoting disassembly. Computational analyses such as Poisson-Boltzmann electrostatic difference mapping of PolθHel in the presence and absence of DNA, as well as molecular dynamics simulations assessing the stability of the tetramer upon DNA binding could help to elucidate how DNA binding modulates the oligomeric state of PolθHel.

Moreover, we did not observe additional density within the DNA tunnel, reflecting a state prior to single-stranded DNA entry. Upon classification of the putative DNA-bound tetramers, several populations of the extra density were observed, including those with density flanking either side of the DNA tunnel, one side, or an absence of the density. These states may reflect the flexibility of the DNA as it is being positioned prior to entering the PolθHel DNA tunnel.

In sum, these data reflect novel insights into the conformational landscape of PolθHel at early stages following DNA addition. The observation of a significant population of the tetrameric state in the presence of DNA, in addition to the identification of putative DNA density at the entrance to the PolθHel DNA tunnel, reflects a role of the tetrameric state in DNA-binding by PolθHel. Future work to determine a high-resolution structure of the DNA-tetramer binding site and further investigate the heterogeneity exhibited by the tetramer is required to fully characterize these interactions. Additional structural studies and functional assays (EMSAs, FRET) performed under varying salt concentrations in the absence and presence of DNA would also help to further characterize the effect of salt concentration on PolθHel oligomeric state. There is an additional possibility that states may be impacted by oligomer design, which vary from study to study. The secondary structure of the double-stranded kink used to stall DNA-PolθHel interactions, as well as the presence of a 5' overhang likely also impact early binding states. A comprehensive study of DNA oligo structure and its impact on PolθHel binding, expanding on the work of Zerio et al. [178], may help to contextualize the differences in oligomeric states observed across studies [179, 181, 182]. Finally, early timepoints in time-resolved structural studies may reveal varying degrees of tetramer flexibility

and dimer destabilization, helping to trace the pathway of disassembly. Together with the results presented in this work, these future studies will serve to provide further insight into early PolθHel-DNA-binding interactions and solidify our understanding of the mechanism of microhomology-mediated end-joining.

4.5 Methods

4.5.1 Sample preparation

4.5.1.1 DNA-free PolθHel

Samples used for the cryo-EM dataset of DNA-free PolθHel at 250 mM NaCl were prepared in-house as described in Zerio et al. [178]. 10xHis-SUMO-Avi-tagged human PolθHel was expressed in Rosetta DE3 competent cells. The cells were grown in 8 L of 2xYT media (Sigma-Aldrich) at 37°C in a shaking incubator until $OD_{600} = 1$, at which point the cells were moved to a shaking incubator at 16°C for 1 hour. Protein production was then induced with 0.2 mM IPTG and continued overnight at 16°C. The cells were harvested via centrifugation at 7,000 RPM, 4°C for 10 min using a JLA 8.1 rotor. The supernatent was decanted and the cell pellets were stored at −20°C until further use.

At the start of the protein purification, the cell pellets were thawed and resuspended in 38 mL of lysis buffer (40 mM HEPES, pH 7.5, 0.5 M NaCl, 5% glycerol, 0.5 mM TCEP, and 1 mM PMSF). The resuspended pellets were pooled and sonicated for 7 min with a pulse of 3s on / 8s off and an amplification of 75%. Cellular debris was removed via centrifugation at 15,000 RPM, 4°C for 40 min using a JA 30.50 rotor. The tagged protein was then purified using nickel affinity chromatography. For this step, the clarified lysate was batch bound for 1 hr at 4°C to Ni-NTA resin that had been washed and pre-equilibrated with lysis buffer. The mixture was then applied to a gravity column and the flow through was discarded. The resin was washed with 25 mL of lysis buffer (10x the 2 mL bed volume), followed by 3 low-salt washes (lysis buffer + 25 mM imidazole), and three elutions of the tagged protein with high-salt buffer (lysis buffer + 250 mM imidazole). The elution fractions were pooled, and BME and EDTA were added for final concentrations of 3 mM and 0.5 mM, respectively. A His-tagged tobacco etch virus (TEV) protease was added to digest the tagged-proteins during overnight dialysis at 4°C in dialysis buffer containing 40 mM HEPES, pH 7.5, 0.5 M NaCl, 5% glycerol, 0.5 mM TCEP, 3 mM BME, and 0.5 mM EDTA. The next day, the dialysis product was collected and batch bound for 1 hr at 4°C to Ni-NTA resin that had been pre-equilibrated with dialysis buffer. The sample was then applied to a gravity column and the flow through, containing untagged PolθHel, was collected. The salt concentration of the sample was lowered to 0.1 M NaCl with the addition of no-salt lysis buffer. Next, the sample was loaded onto a 5-mL HiTrap Heparin column (Cytiva) pre-equilibrated with Heparin buffer A (50 mM HEPES, pH 7.5, 100 mM NaCl, 5% glycerol, and 0.5 mM TCEP) to perform heparin affinity chromatography. The column was washed with 5x column volumes of Buffer A at rate of ∼1 mL/min. A 0–50% gradient of Heparin Buffer B (50 mM HEPES, pH 7.5, 1 M NaCl, 5% glycerol, and 0.5 mM TCEP) was used to elute PolθHel, followed by ∼20 mL of a Buffer B wash. Fractions containing protein were pooled and concentrated to ∼500 μL using

a 50 kDA MW-cut-off concentrator. The protein was further purified using size-exclusion chromatography (SEC) with a Superdex 200 10/300 GL (Cytiva) column. The SEC buffer consisted of 10 mM HEPES, pH 7.5, 250 mM NaCl, and 0.5 mM TCEP. Fractions containing protein were pooled, concentrated, aliquoted, and flash-frozen prior to storage in SEC buffer at −80°C.

Samples used for the cryo-EM dataset of DNA-free PolθHel at 100 mM NaCl were obtained from MOMA Therapeutics and were purified as described in Zerio et al. [178].

4.5.1.2 DNA oligomer design

The design of the DNA oligomer is similar to that used in Zerio et al. with several modifications [178]. The 5' overhang was removed, as well as the secondary structure and mismatched base pairings in the double-stranded hairpin. Eight bases of ssDNA separate the microhomology (CCC GGG) from the dsDNA region. The full sequence is as follows: 5'-ACT GTG ACG CTT AGC GTC ACA GTT TTT TTT CCC GGG-3'.

4.5.1.3 Formation of the PolθHel-DNA complex

The PolθHel-DNA complex used in the cryo-EM dataset of the control PolθHel-DNA (100 mM NaCl, 5 hr) was prepared as described in Zerio et al. from the in-house purified DNA-free PolθHel and the DNA oligo described above [178]. The oligo was annealed at 94°C for 2 min and then allowing to gradually cool on the benchtop for 75–90 min. Apo PolθHel was thawed on ice and spun down at 15,000 RPM, 4°C for 10 min to remove aggregates. Oligos were added in 2x excess with respect to the monomeric protein concentration and incubated on ice for 30 min. The salt concentration was then lowered over the course of an hour from 250 mM NaCl to 100 mM NaCl through the addition of no-salt storage buffer (10 mM Hepes, pH 7.5 and 0.5 mM TCEP). After reaching the desired salt concentration, the mixture was incubated on ice for an additional 30 min. The sample was spun down at 15,000 RPM, 4°C for 10 min to remove aggregates prior to SEC. SEC was performed using a Superdex 200 10/300 GL (Cytiva) column pre-equilibrated with low-salt SEC buffer (10 mM Hepes, pH 7.5, 0.5 mM TCEP, and 100 mM NaCl). Fractions containing protein were analyzed using native PAGE to confirm the presence of DNA-bound PolθHel. Glycerol was added to the samples for a final concentration of 5%. The fractions were then loaded onto a 4.5% native PAGE gel that was prerun in a 4°C cold room for 60 min at 75 V in 0.2x TBE buffer, pH 7.6. The gel was run with samples at 70 V for 90 min in the cold room. The gel was then rinsed with water and stained with 1x Diamond nucleic acid dye (Promega) in 0.2x TBE buffer, pH 7.6 on a rocker for 30 min. The gel was then stained and destained with Coomassie protein stain and imaged on a lightbox. Fractions containing DNA-bound protein were pooled, concentrated with a 50kDA MW-cut-off concentrator, and used to prepare EM grids.

The PolθHel-DNA complex used in the cryo-EM datasets of the five- and ten-minute post DNA addition experiments, were prepared using apo PolθHel protein obtained from MOMA Therapeutics. Aliquots were thawed on ice and the salt was lowered from 250 mM NaCl to 100 mM NaCl with the addition of no-salt storage buffer (10 mM Hepes, pH 7.5 and 0.5 mM TCEP). 2x DNA oligos were added and grids were frozen after 5 or 10 min of incubation.

4.5.1.4 Negative Stain

Negative stain grids were prepared by adding a layer of nitrocellulose acetate onto copper/rhodium Maxtaform 400 mesh EM grids and coating using a carbon evaporator to create a carbon support layer. The grids were glow discharged under vacuum immediately prior to use in a Pelco easiGlow 91000 Glow Discharge Cleaning System (Ted Pella, Inc.) for 30 s at 15 mA. Apo PolθHel samples (MOMA Therapeutics) were diluted to a concentration of 0.175 mg/mL in storage buffer (10 mM Hepes, pH 7.5, 0.5 mM TCEP) with salt added for a final concentration of 100 mM, 150 mM, 200 mM, or 250 mM NaCl. 3 μL of sample was then applied to a grid. After a 10 s incubation, the sample was blotted with Whatman no. 1 filter paper. A 6 μL droplet of 2% uranyl formate stain was applied to the grid for 30 s, and then blotted. Another 6 μL droplet of stain was applied and blotted immediately. Finally, the grid was circled 10x in a 6 μL stain droplet place on a strip of parafilm, after which it was blotted and placed in a fume hood to dry for 2 min.

4.5.1.5 Cryo-EM grid preparation

High-salt DNA-free PolθHel and PolθHel-DNA (5 hr) were left undiluted at concentrations of 0.7 mg/mL and 0.46 mg/mL, respectively. Low-salt DNA-free PolθHel and PolθHel-DNA (5 min and 10 min) were prepared at 1.25 mg/mL and DNA was added in 2x excess.

All grids were prepared in a 4°C cold room at 93–94% humidity. Samples were kept on ice for the duration of the grid preparation. UltrAuFoil 300 mesh R 1.2/1.3 holey grids were plasma cleaned under vacuum using a Solarus plasma cleaner (Gatan, Inc.) set to 25 Watts (75% argon, 25% oxygen) for 7 s. 3 μL of sample was applied to the grid and blotted using Whatman no. 1 filter paper for 2 s after the solvent front of the droplet on the paper had stopped expanding. The grid was then plunged into liquid ethane using a manual plunge-freezer.

4.5.2 Cryo-EM data collection

Cryo-EM data was acquired using a Talos Arctica with a 200 kV field emission gun and a Falcon 4i direct electron detector. The EPU data collection software was used to collect micrographs at a nominal magnification of 190,000 x, corresponding to a pixel size of 0.74 Å/pixel at the detector. The defocus range was set to -1.4, -1.2, -1 μm. The DNA-free, 250 mM control dataset was collected using an exposure rate of 7.42 e^-/pixel/s and an exposure time of 3.67 s. Two grids of the same sample were used for this dataset. The DNA-bound, 250 mM control dataset was collected using an exposure rate of 7.41 e^-/pixel/s and an exposure time of 3.67 s. The DNA-free, 100 mM control dataset was collected using an exposure rate of 6.35 e^-/pixel/s and an exposure time of 4.28 s. For the five-minute post-DNA addition, 100 mM experiment, two datasets of the same grid were collected on separate dates. The first was collected using an exposure rate of 6.35 e^-/pixel/s and an exposure time of 4.28 s. The second was collected using an exposure rate of 6.34 e^-/pixel/s and an exposure time of 4.31 s. The ten-minute post-DNA addition dataset was collected using an exposure rate of 6.34 e^-/pixel/s and an exposure time of 4.31 s.

4.5.3 Cryo-EM data analysis

4.5.3.1 DNA-free PolθHel at 250 mM NaCl (control)

The image processing pipeline is shown in (Fig C.2). 5,270 micrographs were collected. Micrographs were curated with a resolution CTF cutoff of 6.0 Å, leaving 3,485 micrographs remaining for processing. Motion correction and CTF correction were performed in cryoSPARC live v4.6.2 [158]. Particles were picked in cryoSPARC live using Blob Picker, with a particle diameter of 80 Å to 150 Å. The particles were filtered with an NC score of 0.247 and a power score range of 446 to 967. 693,505 particles were extracted with a box size of 440 px, binned to 100 px. Initial 2D classification was performed with fifty classes and a mask diameter of 200 Å. Minimal cleanup was performed by selecting classes resembling PolθHel dimers and tetramers (5 classes excluded) for a total of 627,556 particles. Stringent cleanup of dimer classes was performed in parallel, resulting in 7 classes selected for a total of 136,703 particles. *Ab initio* reconstructions with C1 symmetry were performed with 3 and 4 classes (in parallel) on the particle set from minimal 2D cleanup and with 1 class for the particle set from the stringent cleanup. All *ab initio* jobs were set to use an initial resolution of 20 Å. The reconstruction from the 3-class job corresponding to a tetrameric state was further refined using non-uniform refinement with C2 symmetry and the initial low-pass resolution set to 10 Å. The reconstruction from the 1-class job corresponding to a dimer state was also further refined using non-uniform refinement with C2 symmetry and the initial low-pass resolution set to 10 Å. The refined reconstructions of the tetramer and dimer conformations, along with a "junk" class from the initial 4-class *ab initio* job were used as inputs for heterogenous refinement (C1 symmetry, enforced hard classification, initial resolution 10 Å) containing the 668,558 particles resulting from the initial 2D classification. After sorting, 206k particles were assigned to the tetrameric state, 237k to the "junk" class, and 226k to the dimeric state. Particles from each conformation were re-extracted without binning and the reconstructions were further refined using non-uniform refinement with optimized per-particle defocus and per-group CTF parameters, an initial low-pass resolutions of 6 Å, and D2 and C2 symmetry imposed for the tetramer and dimer, respectively. The resulting tetramer reconstruction was resolved at ∼3.3 Å, and the dimer reconstruction was resolved to ∼3.4 Å, based on a Fourier Shell Correlation (FSC) cutoff of 0.143 (Table C.1).

4.5.3.2 PolθHel at 100 mM NaCl ∼5 hr after DNA addition (control)

The image processing pipeline is shown in (Fig C.3). 6,220 micrographs were collected. Micrographs were curated with a resolution CTF cutoff of 6.0 Å, leaving 2,686 micrographs remaining for processing. Motion correction and CTF correction were performed in cryoSPARC live v4.6.2. Particles were picked in cryoSPARC live using Blob Picker, with a particle diameter of 80 Å to 150 Å. The particles were filtered with an NC score of 0.460 and a power score range of 371 to 750. 221,139 particles were extracted with a box size of 440 px, binned to 100 px. Initial 2D classification was performed with fifty classes and a mask diameter of 200 Å. 2D selection of PolθHel classes resulted in a total of 129,806 particles. Two *ab initio* reconstructions were performed in parallel, with C1 symmetry and an initial resolution of 35 Å, and with C2 symmetry and an initial resolution of 25 Å. The C1 reconstruction was further refined using non-uniform refinement with C1 symmetry. The reconstruction from

the 1-class job corresponding to a dimer state was also further refined using non-uniform refinement with C2 symmetry and the initial low-pass resolution set to 10 Å. The refined reconstruction was put in duplicate into a heterogenous refinement job, along with the C2 *ab initio* used as a "junk" class. Heterogenous refinement was run with C1 symmetry. After sorting, 45k particles were assigned to one of the dimer duplicates, 39k particles to the second, and 76k to the "junk" class. Particles from the best dimer class were further refined using non-uniform refinement. The particles from this refinement, along with the particles from the initial non-uniform refinement job were combined and duplicates were removed before performing a third non-uniform refinement with an initial low-pass resolution of 10 Å. Finally, the particles were re-extracted without binning, and a non-uniform refinement with C2 symmetry, optimized per-particle defocus and per-group CTF parameters, and an initial low-pass resolutions of 10 Å was performed. The resulting dimer reconstruction was resolved at ∼3.3 Å, based on a Fourier Shell Correlation (FSC) cutoff of 0.143 (Table C.1).

4.5.3.3 DNA-free PolθHel at 100 mM NaCl (control)

The image processing pipeline is shown in (Fig C.4). 2,774 micrographs were collected. Micrographs were curated with a resolution CTF cutoff of 6.0 Å, leaving 1,244 micrographs remaining for processing. Motion correction and CTF correction were performed in cryoSPARC live v4.6.2. Particles were picked in cryoSPARC live using Blob Picker, with a particle diameter of 80 Å to 150 Å. The particles were filtered with an NC score of 0.319 and a power score range of 356 to 622. 174,581 particles were extracted with a box size of 440 px, binned to 100 px. Initial 2D classification was performed with fifty classes and a mask diameter of 200 Å. Classes resembling PolθHel tetramers were selected for a total of 66,965 particles. 2D selection of classes loosely resembling PolθHel dimers were selected in parallel, resulting in a total of 17,373 particles. Two *Ab initio* reconstruction jobs were performed in parallel. The first job contained the tetramer particles, and was run with 3 classes, C1 symmetry and an initial resolution of 15 Å. The second job contained the dimer-like particles and was run with 2 classes, C1 symmetry, and an initial resolution of 15 Å. This job did not yield any dimer-like 3D structures. The best tetramer reconstruction was then further refined using non-uniform refinement with C1 symmetry and the initial low-pass resolution set to 15 Å. A heterogenous refinement job with the refined tetramer and two junk classes, along with the full set of particles from the initial 2D classification, was run with C1 symmetry, the initial low-pass resolution set to 15 Å, and force hard classification enabled. After sorting, 43k particles were assigned to the tetramer class. The tetramer reconstruction was further refined using non-uniform refinement with D2 symmetry imposed, optimized per-particle defocus and per-group CTF parameters, and an initial low-pass resolutions of 6 Å. The resulting tetramer reconstruction was resolved at ∼3.9 Å, based on a Fourier Shell Correlation (FSC) cutoff of 0.143 (Table C.1).

4.5.3.4 PolθHel at 100 mM NaCl 5 min after DNA addition

The image processing pipeline is shown in (Fig C.4). 21,425 micrographs were collected. Micrographs were curated with a resolution CTF cutoff of 6.0 Å, leaving 21,414 micrographs remaining for processing. Motion correction and CTF correction were performed in cryoSPARC

live v4.6.2. Particles were picked in cryoSPARC using Blob Picker, with a particle diameter of 90 Å to 175 Å, yielding ∼3.2 mil particles. The particles were filtered with an NC score of 0.360 and a power score range of 401 to 526. ∼2.5 mil particles were extracted with a box size of 440 px, binned to 100 px. Initial 2D classification was performed with fifty classes and a mask diameter of 200 Å. Classes resembling PolθHel tetramers and dimers were selected for a total of ∼1.4 mil particles. *Ab initio* reconstruction was performed with 4 classes and an initial resolution of 20 Å. The best tetramer and dimer classes were further refined using non-uniform refinement with the initial low-pass resolution set to 15 Å. 2D classification was then run in parallel on the particle set from each reconstruction to yield templates for template picking. 45,972 tetramer particles corresponding to four 2D classes and 73,197 dimer particles corresponding to four 2D classes were used as templates for template picking. Template picking was run with an estimated particle diameter of 175 Å and a minimum separation distance of 0.6 diameters, yielding 8.9 mil particles. The particles were filtered with an NC score of 0.330 and a power score range of 541 to 1191. ∼3.5 million particles were extracted and 2D classified with 100 classes, a mask diameter of 180 Å, a batchsize of 500 particles per class, 50 online-EM iterations, 5 final full iterations, an initial classification uncertainty factor of 4, and hard classification enabled for the final iteration. 2D selection was then performed to yield ∼1.8 mil particles. These were input into an *Ab initio* reconstruction job with 3 classes and an initial low-pass resolution of 20 Å. The tetramer class was further refined using non-uniform refinement with an initial low-pass resolution of 10 Å. The refined tetramer, along with the dimer class from non-uniform refinement following blob picking, two junk classes, and the full ∼3.5 mil particles from 2D classification following template picking were input into a heterogenous refinement job with an initial low-pass resolution of 10 Å and hard classification enforced. After sorting, 902k particles were assigned to the tetramer class, 985k to the dimer class, and the remaining to the two junk classes. The tetramer reconstruction was then further refined using non-uniform refinement with an initial low-pass resolution of 15 Å. Tetramer particles were re-extracted without binning and refined using non-uniform refinement with optimized per-particle defocus and per-group CTF parameters, and an initial low-pass resolutions of 8 Å. The resulting tetramer reconstruction was resolved at ∼3.4 Å, based on a Fourier Shell Correlation (FSC) cutoff of 0.143 (Table C.1).

4.5.3.5 PolθHel at 100 mM NaCl 10 min after DNA addition

The image processing pipeline is shown in (Fig C.6). 6,313 micrographs were collected. Micrographs were curated with a resolution CTF cutoff of 6.0 Å, leaving 5,431 micrographs remaining for processing. Motion correction and CTF correction were performed in cryoSPARC live v4.6.2. Particles were picked in cryoSPARC live using Blob Picker, with a particle diameter of 100 Å to 150 Å. The particles were filtered with an NC score of 0.271 and a power score range of 474 to 905, yielding 496,830 particles. 496,482 particles were extracted with a box size of 440 px, binned to 100 px. 2D classification was performed with fifty classes and a mask diameter of 200 Å (Table C.1).

5

Concluding Remarks

This thesis presents a comprehensive investigation of macromolecular dynamics across multiple temporal scales, from microseconds to minutes, using complementary computational and experimental approaches. Through the integration of long-timescale molecular dynamics simulations and the development and application of time-resolved cryo-EM methodologies, we have explored conformational landscapes of a diverse range of macromolecular machines, including the LONP1 mitochondrial protease and the DNA-repair enzyme Polθ. Together, these studies demonstrate the power of multi-scale approaches to uncover dynamic processes that govern macromolecular function.

The microsecond-timescale molecular dynamics simulations of LONP1 provide quantitative insight into conformational landscapes on temporal scales inaccessible to experimental approaches. These simulations reveal the dynamic nature of ATP and ADP binding, identifying conserved aromatic residues that stabilize nucleotide binding poses through hydrogen bonding and stacking interactions. Nucleotide conformational fluctuations propagate through the binding pocket, while the aromatic residues mediate hinge-like motions between the subdomains of the ATPase module in response. The temporal dynamics of pore loop motion during simulation support a link between subunit position and substrate engagement, consistent with the hand-over-hand translocation model of AAA+ proteases. Substrate dissociation from the protease domains occurs on sub-microsecond timescales in simulation, and the conserved Trp770 β-hairpin motif undergoes rapid closure to block proteolytic groove access upon substrate loss, indicating a gating-mechanism of the proteolytic binding sites mediated by this structural element. We additionally establish the timescale for proteolytic site inactivation (microseconds) and demonstrate the stochastic nature of its transition, revealing that local conformational rearrangements within the proteolytic sites precede the larger inter-domain movements responsible for shifting LONP1 into a catalytically inactive state. Further computational studies are required to establish the energy landscape describing the transition

between activity states of the proteolytic sites. While a direct allosteric communication pathway linking ATPase and protease activities remains unresolved, correlated motions between functional domains are quantifiable on microsecond timescales. Future computational studies should employ mutual information analysis to identify causal relationships between residue fluctuations to establish a direct communication network between functional domains.

The development of Mix-it-up (MIU) establishes a cost-effective platform for millisecond-resolution time-resolved cryo-EM studies. We demonstrate high-resolution structures of rapidly mixed samples within (as low as ∼120 ms), validating the method's capability to capture rapid conformational transitions, including CCMV capsid contraction and GroEL/ES chaperonin complex assembly. The use of gold nano-fiducials demonstrates an extensible method for tracking spray localization for components to small to observe by electron microscopy, that can be adopted by other time-resolved tools to aid user-driven screening and data collection. Future applications of these fiducials could also allow quantitative assessment of mixing efficiency and droplet dynamics within the MIU framework, as well as optimization of other spray-based systems. While millisecond temporal resolution represents a substantial advancement over conventional cryo-EM sample preparation, further design optimizations could enhance temporal resolution, such as spray-while-plunge approaches or ethane jet vitrification to eliminate the rate-limiting plunge step. Further, the applicability of MIU to more complex biological systems remains to be tested, with promising targets detailed in this thesis, such as transient substrate recognition events in LONP1 and millisecond timescale DNA-binding interactions with PolθHel.

Our investigations of early PolθHel conformational ensembles in the presence of DNA reveal an unexpected balance of oligomeric states. We reveal a salt-dependence of the oligomeric states, rationalizing the observed discrepancies in previous structural studies. We also identify additional density outside the DNA entry tunnel of the tetrameric state of PolθHel, consistent with putative DNA-binding sites. Several states of the density are observed, suggesting flexibility of the DNA as it begins to position its 3' single stranded DNA ends for microhomology search by PolθHel. Further high-resolution refinement of the density is required to characterize the DNA interactions with the tetrameric state. These findings suggest a prominent role of the tetrameric species in early DNA recognition, opening questions into its direct function in microhomology-mediated end joining.

In sum, this work highlights the need for multi-disciplinary studies of conformational landscapes within macromolecular complexes in order to assess dynamics on the range of timescales associated with function.

A

Appendix: LONP1

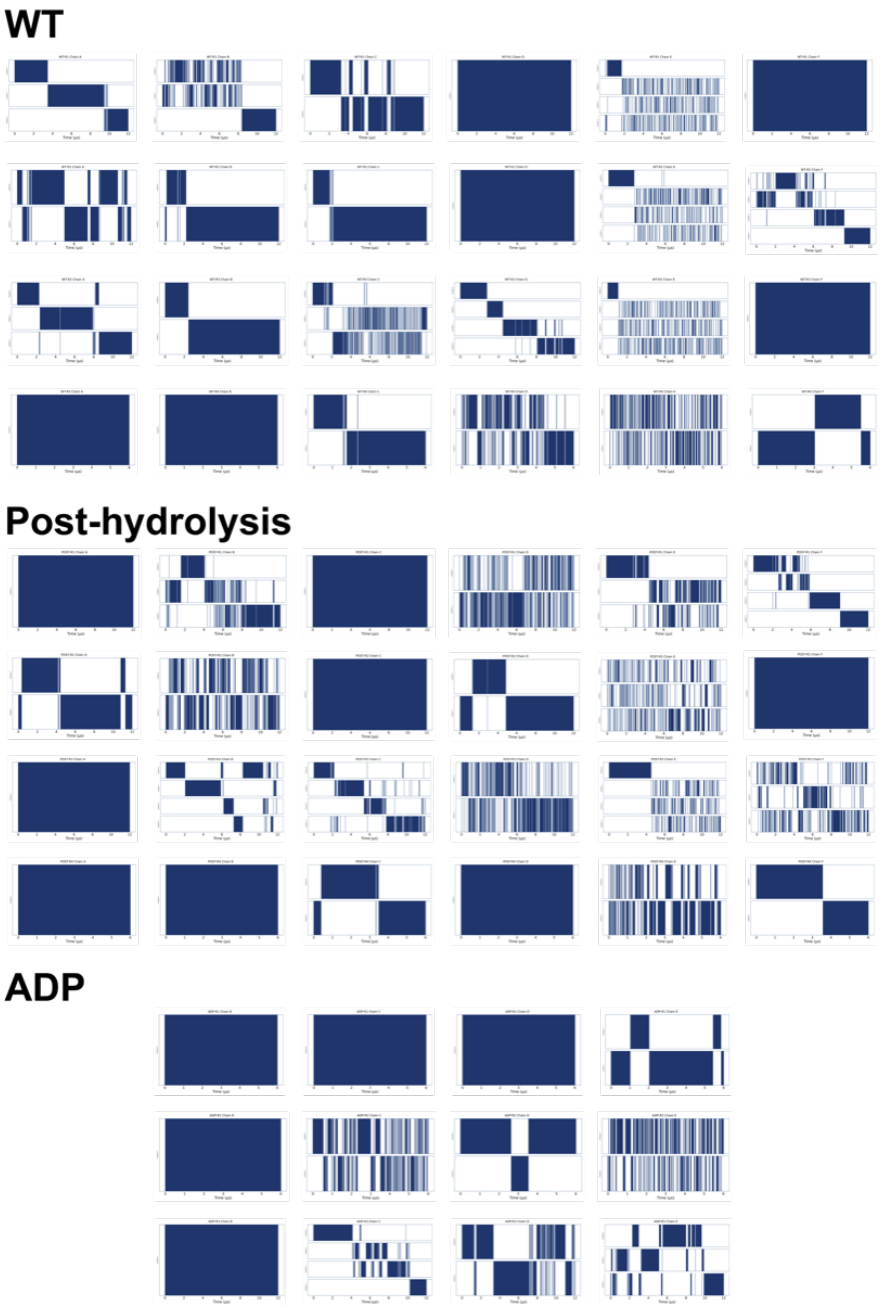


Figure A.1: Complete analysis of time-dependent nucleotide binding pose clusters. Analysis of the time-dependence of nucleotide binding clusters for LONP1^{WT} (top), LONP1^{Post} (middle), and LONP1^{ADP}. Each column of plots reflects one chain of the hexamer in order from chain A-F, left to right (chain A and F of LONP1^{ADP} are not show due to the absence of nucleotide at these positions). Each row represents a distinct simulation replicate. The rows within each plot represent a distinct cluster of the nucleotide conformations, with solid colored plots reflecting the presence of a single cluster for the entirety of the simulation.

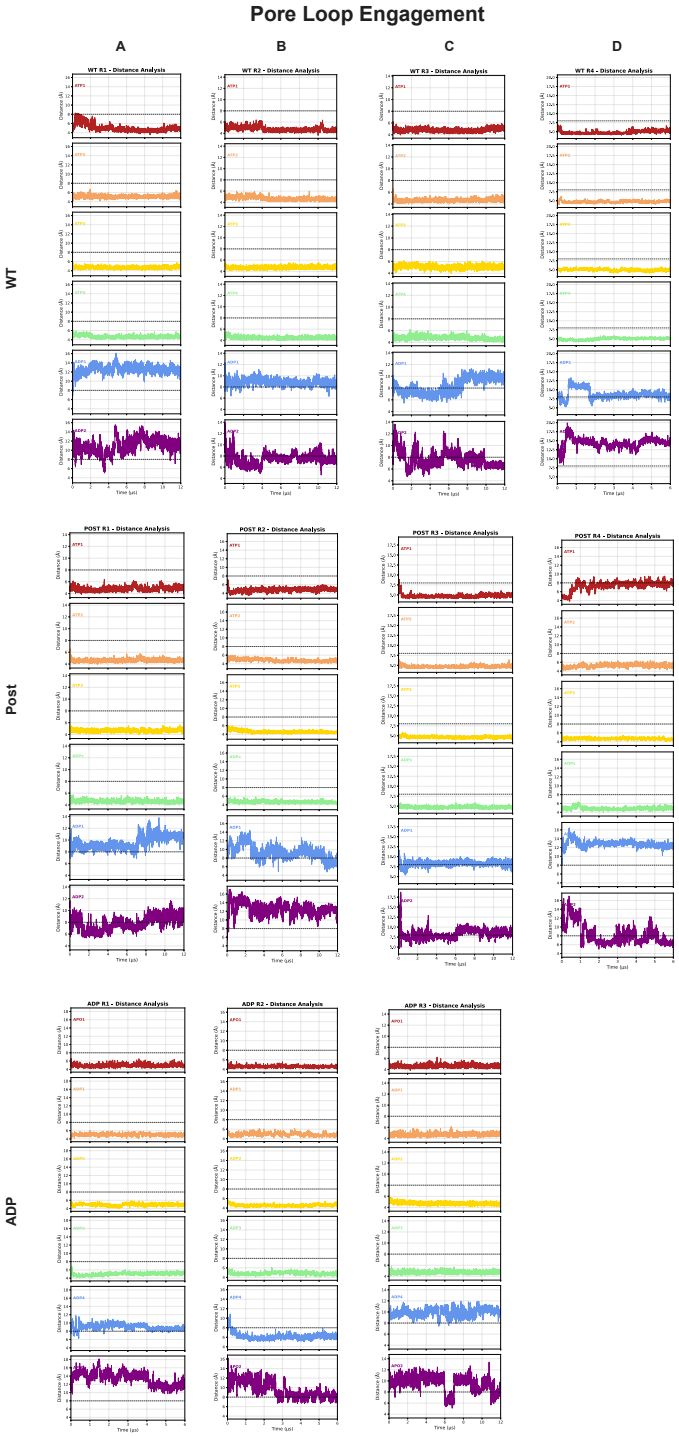


Figure A.2: Complete analysis of pore loop to substrate engagement distances. Analysis of the distance between Y565 of the pore loop 1 motif and the substrate engaged in the central pore of the ATPase domain for LONP1^{WT} (top), LONP1^{Post} (middle), and LONP1^{ADP}. Each column of plots reflects a distinct replicate. The rows correspond to each chain of the hexamer. A dotted line reflects an 8 Å threshold for interaction distances.

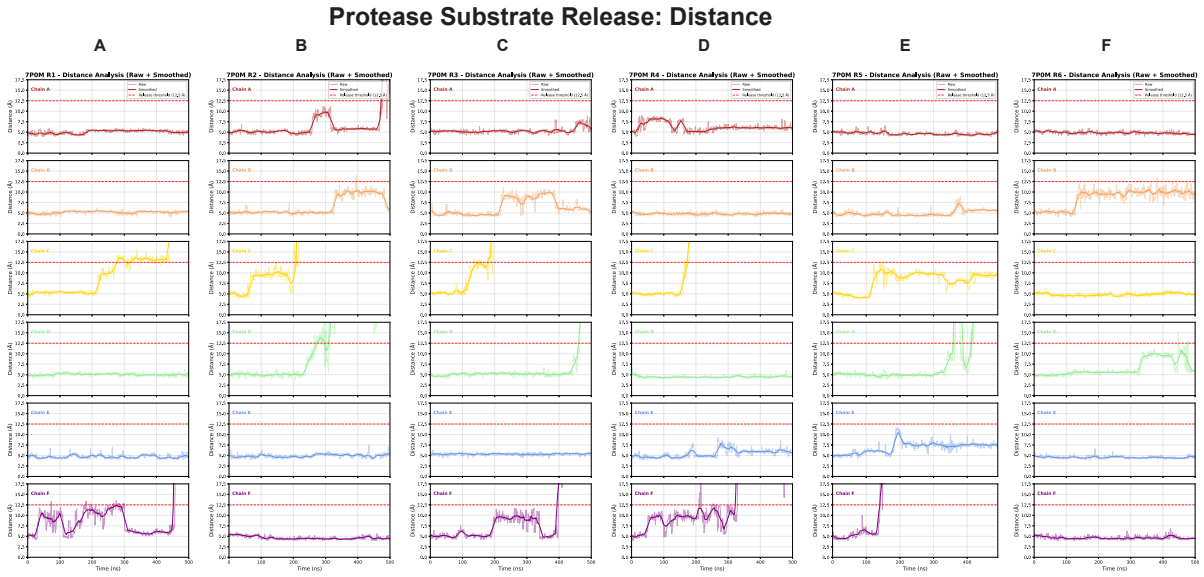


Figure A.3: Complete analysis of substrate release distances. Analysis of the distance between the tetra-alanine substrate and the proteolytic loop for LONP1$^{\text{Protease}}$. Each column of plots reflects a distinct replicate. The rows correspond to each chain of the hexamer. The smoothed data (solid lines) are plotted atop the raw data (transparent). A dotted line reflects an 12.5 Å threshold for substrate release.

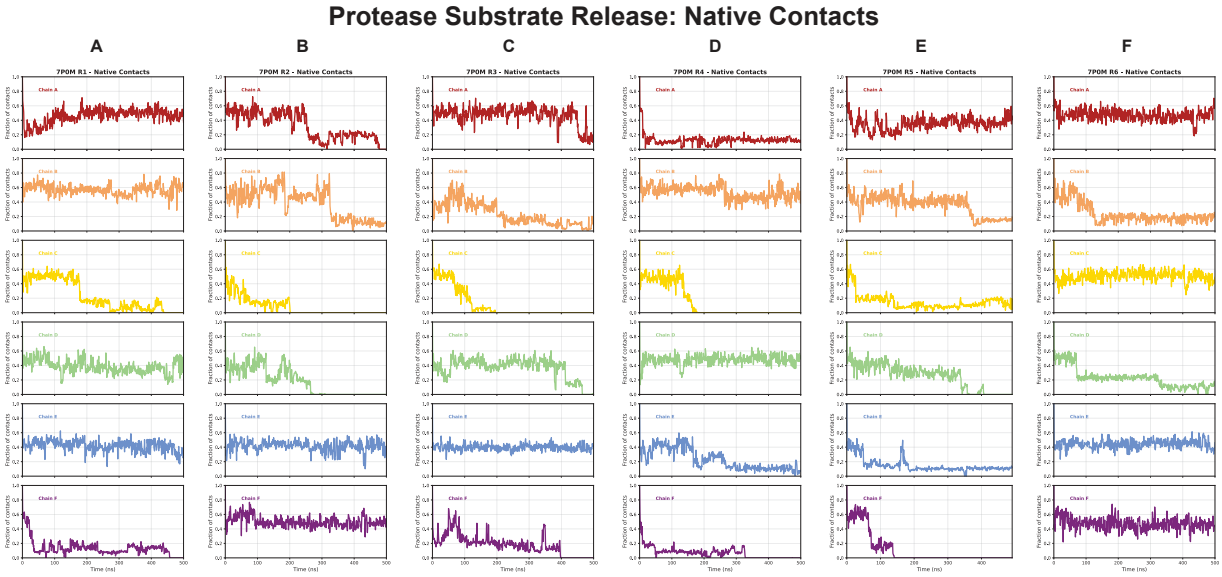


Figure A.4: Complete analysis of substrate release contacts. Analysis of the fraction of native contacts between the tetra-alanine substrate and the proteolytic loop for LONP1$^{\text{Protease}}$. Native contacts refer to contacts present in the initial frame of the trajectory. Each column of plots reflects a distinct replicate. The rows correspond to each chain of the hexamer. A release event is identified when the fraction of native contacts reaches zero.

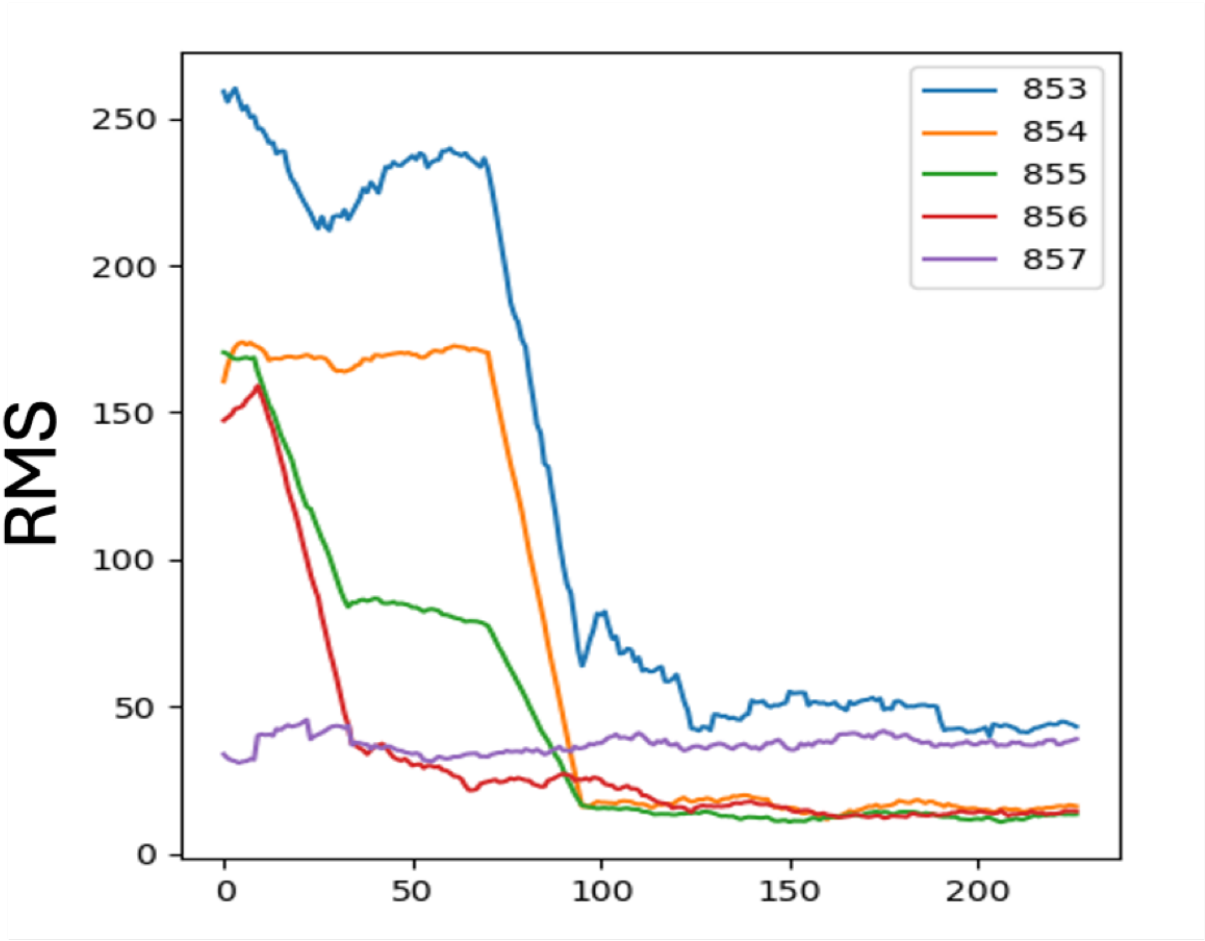


Figure A.5: Selection of residues for dihedral analysis to determine presence of the inactive state of the proteolytic site. Representative dihedral analysis for several residues (853–857) that form the 3_{10}-helix in the inactive proteolytic site. The RMS of the dihedrals of each residue from those of an ideal 3_{10}-helix are plotted during the course of the trajectory. The threshold chosen to identify the helix corresponds to an RMS of 35°. The dihedrals of P854 (orange) and S855 (green) were empirically found to demonstrate the highest temporal correlation and most robust response to formation events of the helix, as observed upon visualization of the trajectories.

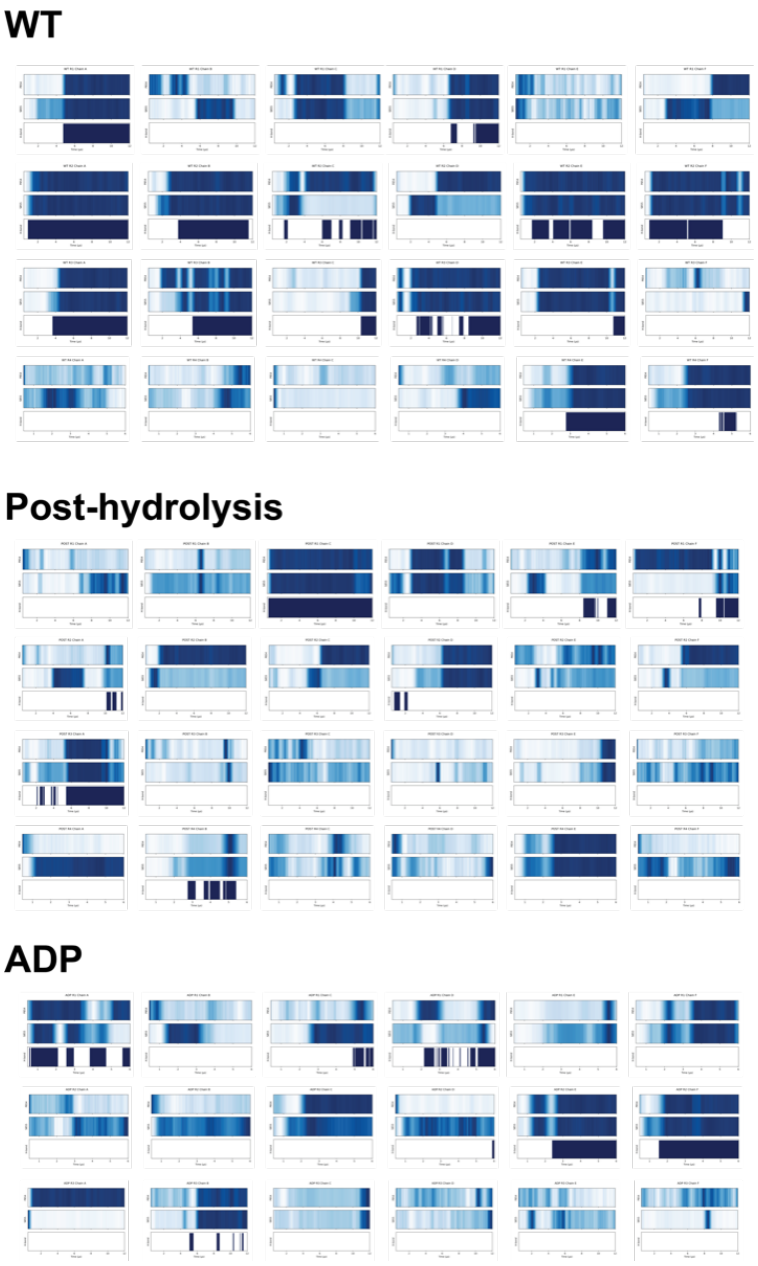


Figure A.6: Complete analysis of proteolytic inactivation events. Analysis of the time-dependence of proteolytic inactivation events in LONP1^{WT} (top), LONP1^{Post} (middle), and LONP1^{ADP}. Each column of plots reflects one chain of the hexamer in order from chain A-F, left to right. Each row represents a replicate trajectory. Within each plot, the top two rows correspond to the RMS of the dihedrals of P854 and S855, respectively, from those of an ideal 3_{10}-helix, with darker blues corresponding to a lower RMS and higher helicity. Together, these two metrics confer the presence of the sterically inactive state of LONP1 proteolytic sites. The third row indicates the presence of a hydrogen bond (dark blue) between D852 and the catalytic lysine (K898), indicating the presence of the chemically inactive state. When all three metrics are "on" (dark blue), the proteolytic sites are considered fully inactive.

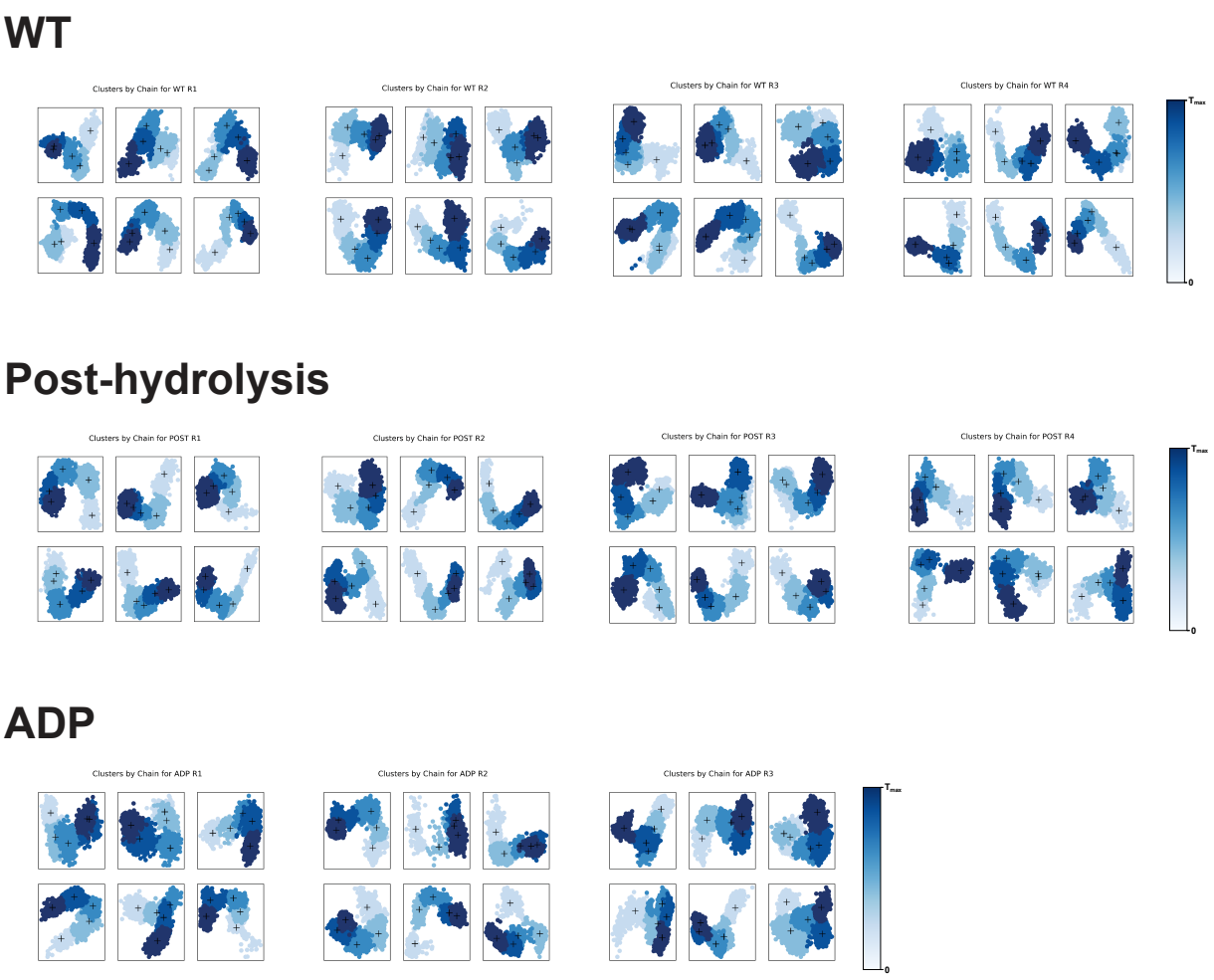


Figure A.7: Complete per-chain principal component analysis. Principal component analysis on the backbone atoms of each subunit for LONP1$^{\mathrm{WT}}$ (top), LONP1$^{\mathrm{Post}}$ (middle), and LONP1$^{\mathrm{ADP}}$. Each column represents a distinct simulation replicate. Within each plot, the six panels correspond to chain A-F (left to right and top to bottom, respectively.) For each chain, principal component 1 is plotted against principal component 2, and the backbone atom positions for each timestep in the trajectory are plotted onto the PC space. The points are colored according to time, with lighter blue colors reflecting initial timepoints and darker blues corresponding to end timepoints. The centroids of the five groups obtained from clustering are marked with black crosses.

B

Appendix: MIU

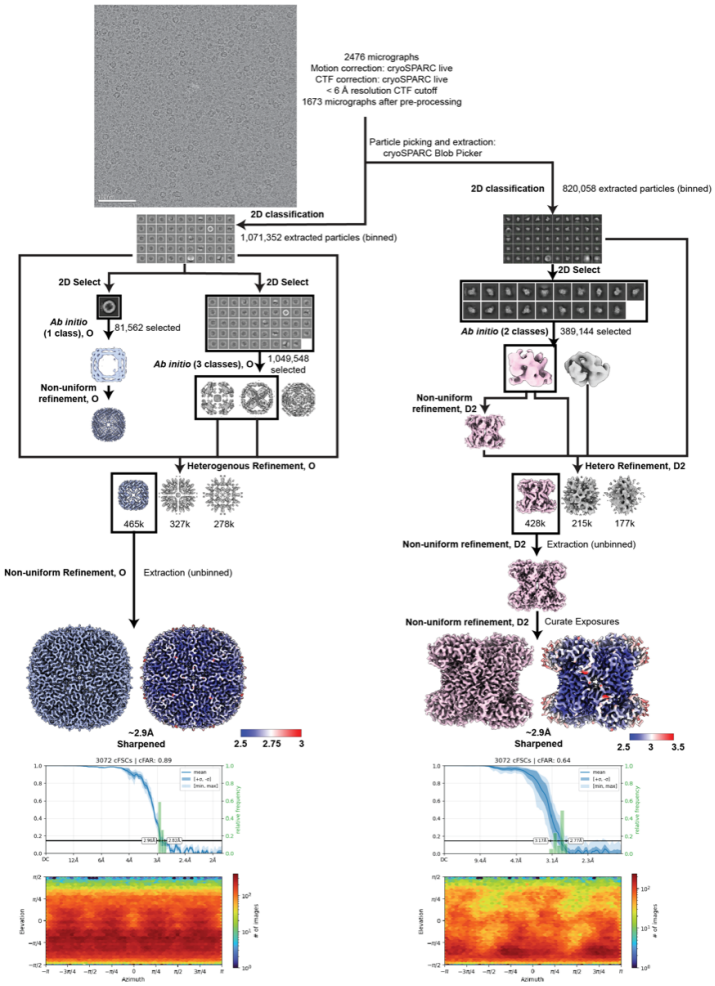


Figure B.1: Processing pipeline of the mixed apoferritin and aldolase dataset prepared with MIU. Representative processing pipeline for mixed sample of apoferritin and aldolase. Motion and CTF correction were performed in cryoSPARC Live, as well as picking and extraction of apoferritin particles. Micrographs were exported to cryoSPARC for aldolase particle picking and extraction. Following 2D classification of apoferritin particles, two subsets were generated using stringent and minimal cleanup strategies and ab initio reconstructions were generated with octahedral symmetry imposed. The highest resolution ab initio class was used for initial 3D refinement with imposed symmetry. All extracted particles were then combined with several "good" and "junk" initial volumes in a heterogeneous refinement job with symmetry imposed to separate high-quality particles. Aldolase particles were processed in a similar manner with the following exceptions: a single subset of particles was generated of the best class averages following 2D classification. Following ab initio reconstruction, initial refinement, and heterogenous refinement, the particles were further refined, and exposures were curated to further filter out poor quality particles. For both apoferritin and aldolase, final refinement of the selected particles yielded a map at ∼2.9 Å resolution. Final maps are shown next to a local resolution map. The conical FSC plot overlaid with a resolution histogram and corresponding angular distribution plot are shown below the final maps.

Table B.1: Cryo-EM data collection and image analysis parameters for apoferritin, aldolase, and CCMV samples prepared with MIU

Parameter	**Apoferritin**	**Aldolase**	**CCMV contracted**	**CCMV expanded**	**CCMV pH shift**
EMDB ID	EMDB:70747	EMDB:70748	EMDB:70749	EMDB:70750	EMDB:70751
Data collection					
Microscope	Talos Arctica		Talos Arctica	Talos Arctica	Talos Arctica
Detector	TFS Falcon 4i		TFS Falcon 4i	TFS Falcon 4i	TFS Falcon 4i
Voltage (kV)	200		200	200	200
Magnification	150,000/148,936		150,000/148,936	150,000/148,936	150,000/148,936
Data acq. software	EPU v3.9.1.8206		EPU v3.9.1.8206	EPU v3.9.1.8206	EPU v3.9.1.8206
Exposure navigation	Image shift to 8μm		Image shift to 8μm	Image shift to 8μm	Image shift to 8μm
Total exposure (e^-/Å^2)	50		50	50	50
Exposure rate (e^-/pixel/sec)	10.53		10.75	9.93	10.90
Exposure time (s)	4.2		4.1	4.5	4.1
Number of frames	1,296		1,008	1,377	1,251
Pixel size (Å) per pixel	0.94		0.94	0.94	0.94
Defocus range (μm)	-1 to -1.4		-1 to -1.4	-1 to -1.4	-1 to -1.4
Micrographs collected (no.)	2,476		292	595	6,007
Data analysis					
Processing package	cryoSPARC		cryoSPARC	cryoSPARC	cryoSPARC
Total extracted picks (no.)	1,701,352	820,058	7,994	62,017	68,810
Particles used for 3D (no.)	465,201	185,420	5,397	1,885	6,952
Final particles (no.)	451,448	185,420	5,369	1,665	6,889
Symmetry	Oct	D2	Icos	Icos	Icos
Global resolution (Å)					
FSC = 0.143 (no mask / masked) (Å)	3.0 / 2.9	3.2 / 2.9	3.8 / 3.2	8.4 / 6.9	4.6 / 3.9
Local resolution range (Å)	2.5–9.7	2.5–38.1	2.0–49.5	1.9–17.6	2.0–61.5
Map sharpening B-factor (Å^2)	-150.6	-121.2	-79.8	-383.1	-82.8
cFSCs cFAR	0.89	0.64	0.97	0.85	0.98

Table B.2: Cryo-EM data collection and image analysis parameters for GroEL and GroES samples (no ATP, 100ms) prepared with MIU

Parameter	GroEL barrel, no ATP	GroEL barrel, 100ms
EMDB ID	EMDB:70752	EMDB:70759
Data collection		
Microscope	Talos Arctica	Talos Arctica
Detector	TFS Falcon 4i	TFS Falcon 4i
Voltage (kV)	200	200
Magnification	150,000/148,936	150,000/148,936
Data acq. software	EPU v3.9.1.8206	EPU v3.9.1.8206
Exposure navigation	Image shift to 8μm	Image shift to 8μm
Total exposure ($e^-/Å^2$)	50	50
Exposure rate (e^-/pixel/sec)	10.90	10.94
Exposure time (s)	4.1	4.1
Number of frames	1,251	1,251
Pixel size (Å) per pixel	0.94	0.94
Defocus range (μm)	-1 to -1.4	-1 to -1.4
Micrographs collected (no.)	601	2,645
Data analysis		
Processing package	cryoSPARC	cryoSPARC
Total extracted picks (no.)	60,116	119,235
Particles used for 3D (no.)	11,859	11,931
Final particles (no.)	11,777	11,864
Symmetry	D7	D7
Global resolution (Å)		
FSC = 0.143 (no mask / masked)	3.8 / 3.2	3.9 / 3.4
Local resolution range (Å)	2.0–31.8	3.1–41.5
Map sharpening B-factor ($Å^2$)	-75.4	-75.2
cFSCs cFAR	0.80	0.65

Table B.3: Cryo-EM data collection and image analysis parameters for GroEL and GroES samples (300ms) prepared with MIU

Parameter	**GroEL barrel, 300ms**	**GroEL bullet, 300ms**	**GroEL football, 300ms**
EMDB ID	EMDB:70753	EMDB:70754	EMDB:70755
Data collection			
Microscope	Talos Arctica		
Detector	TFS Falcon 4i		
Voltage (kV)	200		
Magnification	150,000/148,936		
Data acq. software	EPU v3.9.1.8206		
Exposure navigation	Image shift to 8μm		
Total exposure (e^-/Å^2)	50		
Exposure rate (e^-/pixel/sec)	9.66		
Exposure time (s)	4.6		
Number of frames	1,395		
Pixel size (Å) per pixel	0.94		
Defocus range (μm)	-1 to -1.4		
Micrographs collected (no.)	5,301		
Data analysis			
Processing package	cryoSPARC		
Total extracted picks (no.)	255,389		
Particles used for 3D (no.)	14,813	25,588	7,696
Final particles (no.)	14,730	25,430	7,653
Symmetry	D7	C7	D7
Global resolution (Å)			
FSC = 0.143 (no mask / masked) (Å)	4.2 / 3.6	3.9 / 3.3	6.8 / 4.8
Local resolution range (Å)	4.3–38.2	1.9–17.4	5.0–51.4
Map sharpening B-factor (Å^2)	-84.7	-74.3	-180.9
cFSCs cFAR	0.69	0.79	0.31

Table B.4: Cryo-EM data collection and image analysis parameters for GroEL and GroES samples (700ms) prepared with MIU

Parameter	**GroEL barrel, 700ms**	**GroEL bullet, 700ms**	**GroEL football, 700ms**
EMDB ID	EMDB:70756	EMDB:70757	EMDB:70758
Data collection			
Microscope	Talos Arctica		
Detector	TFS Falcon 4i		
Voltage (kV)	200		
Magnification	150,000/148,936		
Data acq. software	EPU v3.9.1.8206		
Exposure navigation	Image shift to 8μm		
Total exposure (e^-/Å^2)	50		
Exposure rate (e^-/pixel/sec)	10.90		
Exposure time (s)	4.1		
Number of frames	1,251		
Pixel size (Å) per pixel	0.94		
Defocus range (μm)	-1 to -1.4		
Micrographs collected (no.)	6,063		
Data analysis			
Processing package	cryoSPARC		
Total extracted picks (no.)	118,543		
Particles used for 3D (no.)	3,605	8,753	2,121
Final particles (no.)	3,588	8,711	2,111
Symmetry	D7	C7	D7
Global resolution (Å)			
FSC = 0.143 (no mask / masked) (Å)	7.0 / 4.7	4.4 / 3.5	8.1 / 7.1
Local resolution range (Å)	2.0–70.5	2.0–55.3	1.9–17.4
Map sharpening B-factor (Å^2)	-118.4	-60.5	-685.9
cFSCs cFAR	0.31	0.79	0.42

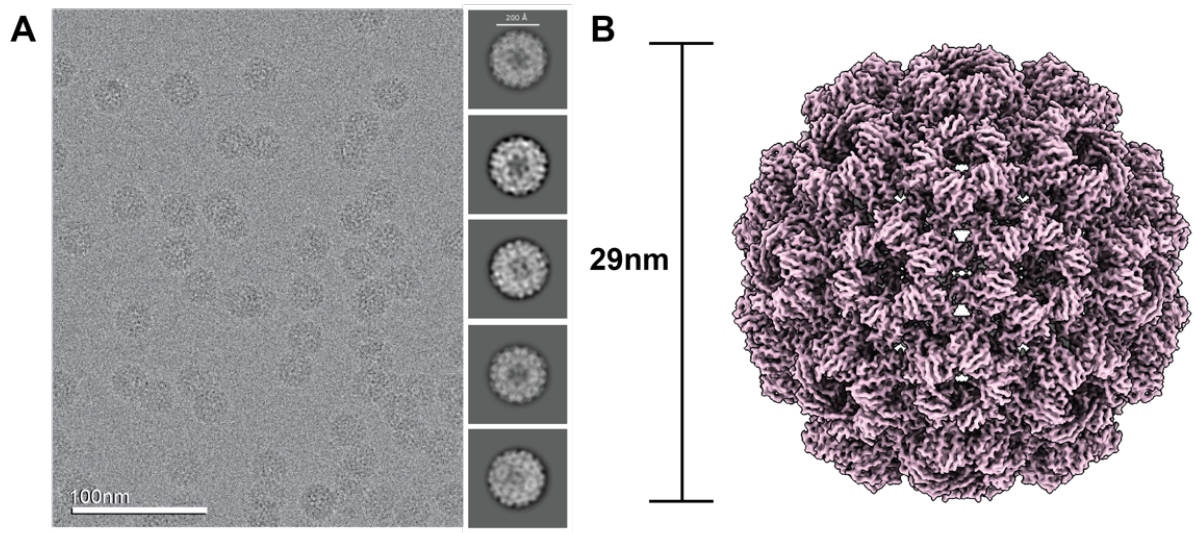


Figure B.2: Control, contracted CCMV. A) Representative micrograph and 2D class averages of CCMV sample in pH 4.6 buffer. B) EM density map of the contracted state exhibiting a 29 nm diameter.

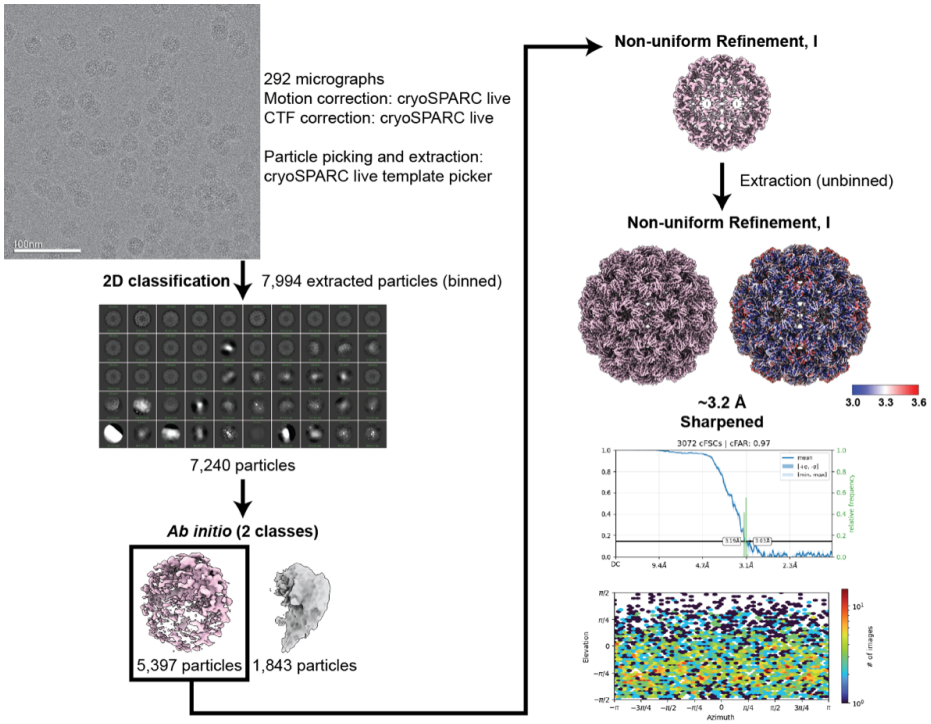


Figure B.3: Processing pipeline of the control, contracted CCMV state dataset. Representative processing pipeline for CCMV dataset. Motion and CTF correction, initial particle picking, and initial 2D classification were performed in cryoSPARC Live. Template-based picking was used to refine particle selection, followed by additional 2D classification and ab initio reconstruction. Particles from the best class were used for 3D refinement with imposed icosahedral symmetry. After re-extraction, particles were further refined to yield a final map at ∼3.2 Å resolution. The final map is shown next to a local resolution map. A conical FSC plot overlaid with a histogram of resolutions and an angular distribution plot are shown below the final maps.

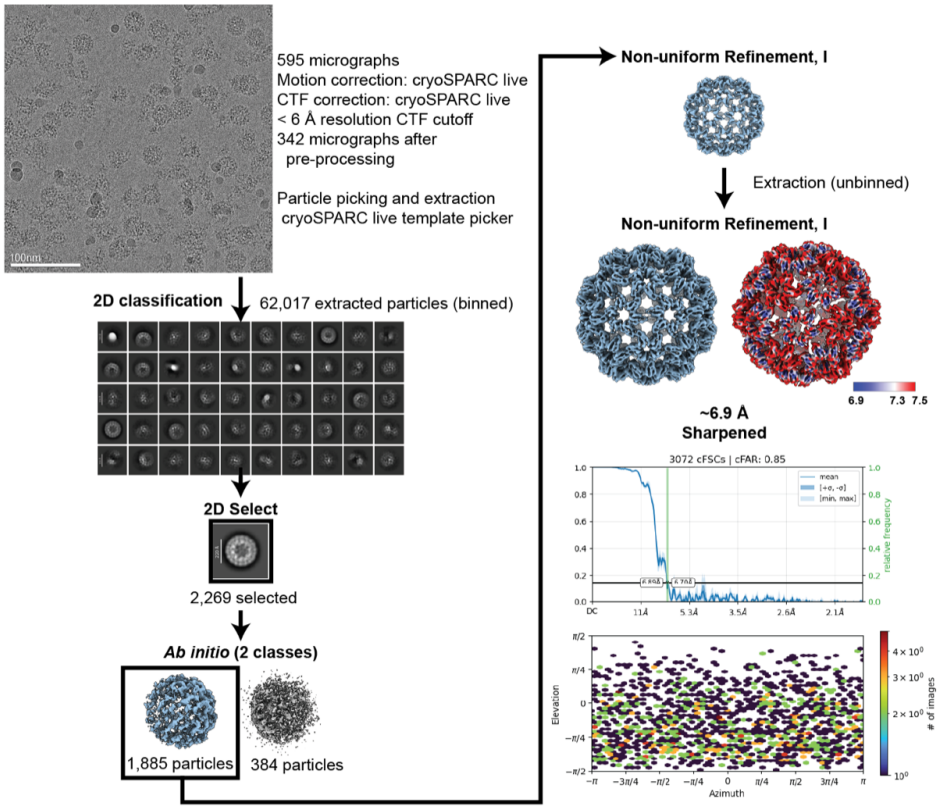


Figure B.4: Processing pipeline of the control, expanded CCMV state dataset. Motion and CTF correction were performed in cryoSPARC Live, with a resolution cutoff applied during preprocessing. Initial blob-based particle picking was followed by template-based picking and 2D classification. A selected subset of particles was used for ab initio reconstruction and 3D refinement was performed with imposed icosahedral symmetry. After re-extraction, particles were further refined to yield a final map at ∼6.9 Å resolution. The final map is shown next to a local resolution map. The conical FSC plot overlaid with a resolution histogram and corresponding angular distribution plot are shown below the final map.

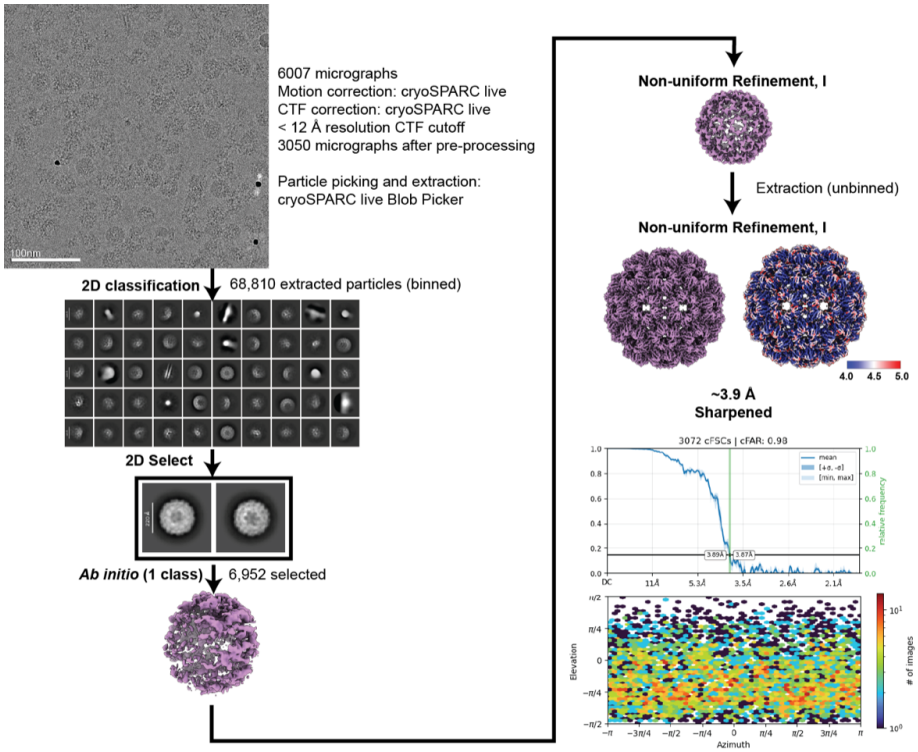


Figure B.5: Processing pipeline of the pH-shift CCMV dataset prepared with MIU. Representative processing pipeline for the pH-shift CCMV dataset. Motion and CTF correction were performed in cryoSPARC Live. Particles were picked using blob-based picking, followed by 2D classification and selection of a particle subset for ab initio reconstruction. The resulting volume was refined with icosahedral symmetry imposed. After re-extraction, particles were further refined to yield a final map at ∼3.9 Å resolution. The final map is shown next to a local resolution map. The conical FSC plot overlaid with a resolution histogram and corresponding angular distribution plot are shown below the final map.

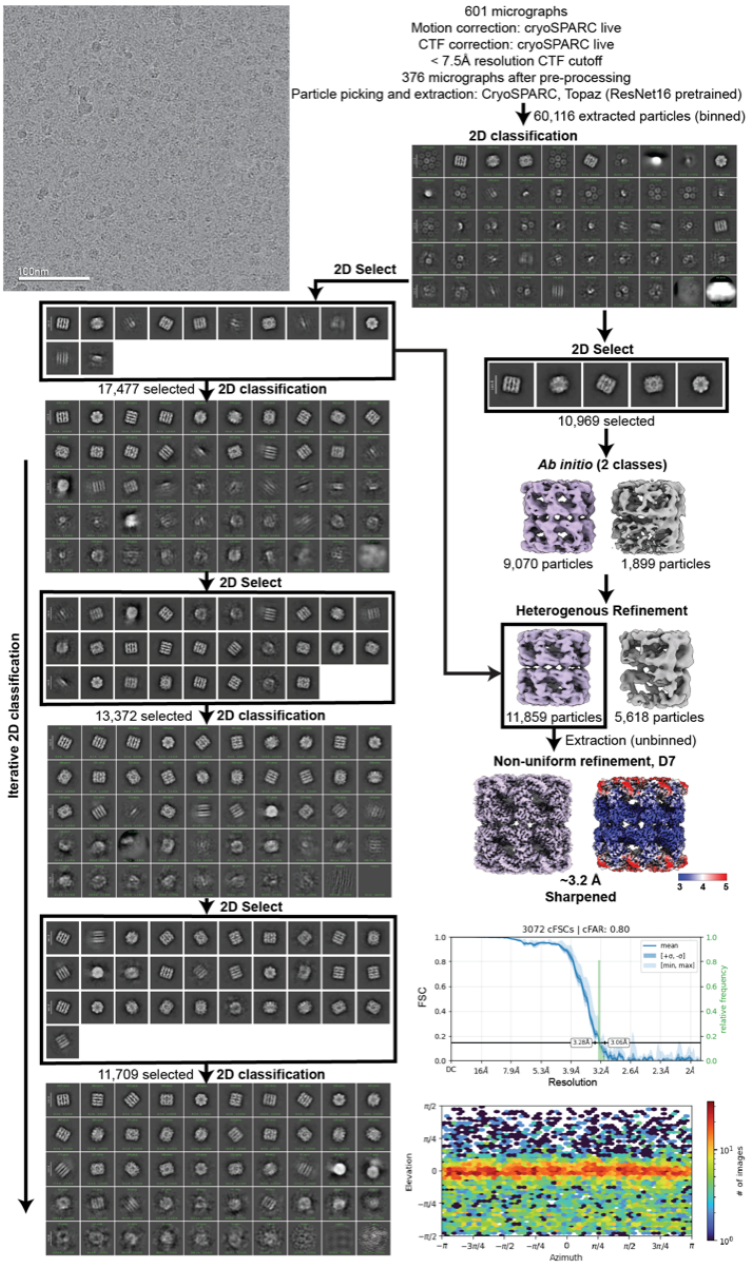


Figure B.6: Processing pipeline of the control dataset of GroEL in the presence of GroES and absence of ATP. Representative processing pipeline for the GroEL + GroES control dataset prepared in the absence of ATP. Motion and CTF correction were performed in cryoSPARC Live. Micrographs were exported to cryoSPARC, and particles were picked using Topaz with a pretrained model, followed by initial 2D classification. Iterative 2D classification was also performed to confirm the absence of GroEL/ES complexes. A subset of particles resembling GroEL was selected for ab initio reconstruction, and the resulting volumes were further refined in a heterogeneous refinement job using a broader particle set. High-quality particles were then re-extracted and refined with imposed D7 symmetry to yield a final map at ∼3.2 Å resolution. The conical FSC plot overlaid with a resolution histogram and corresponding angular distribution plot are shown below the final map.

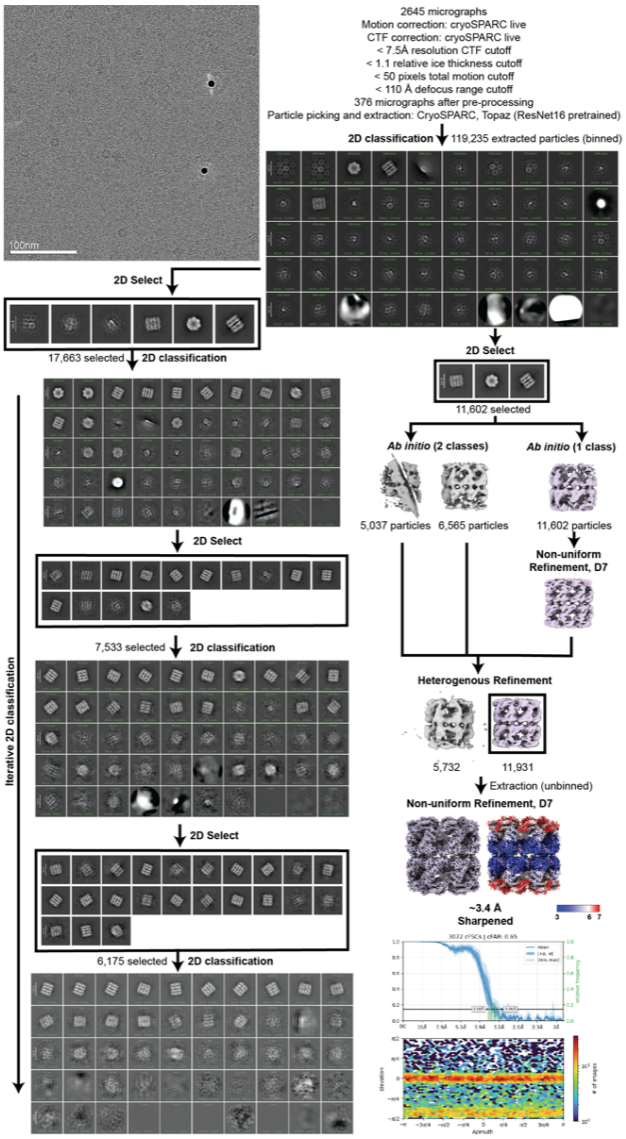


Figure B.7: Processing pipeline of the 100 ms GroEL/ES + ATP dataset. Representative processing pipeline for the 100 ms GroEL + GroES + ATP dataset. Motion and CTF correction were performed in cryoSPARC Live. Micrographs were exported to cryoSPARC, and particles were picked using Topaz with a pretrained model, followed by 2D classification and selection of GroEL particles, resembling the "barrel" (uncapped) conformational state. Iterative 2D classification was performed in parallel to confirm the absence of additional conformational states. Ab initio reconstructions were generated and the "best" class was further refined with symmetry imposed. Heterogeneous refinement was then used to sort high-quality particles from a broader subset selection. The best particles were re-extracted and refined with D7 symmetry to yield a final map at ∼3.4 Å resolution. The final map is shown next to a local resolution map. The conical FSC plot overlaid with a resolution histogram and corresponding angular distribution plot are shown below the final maps.

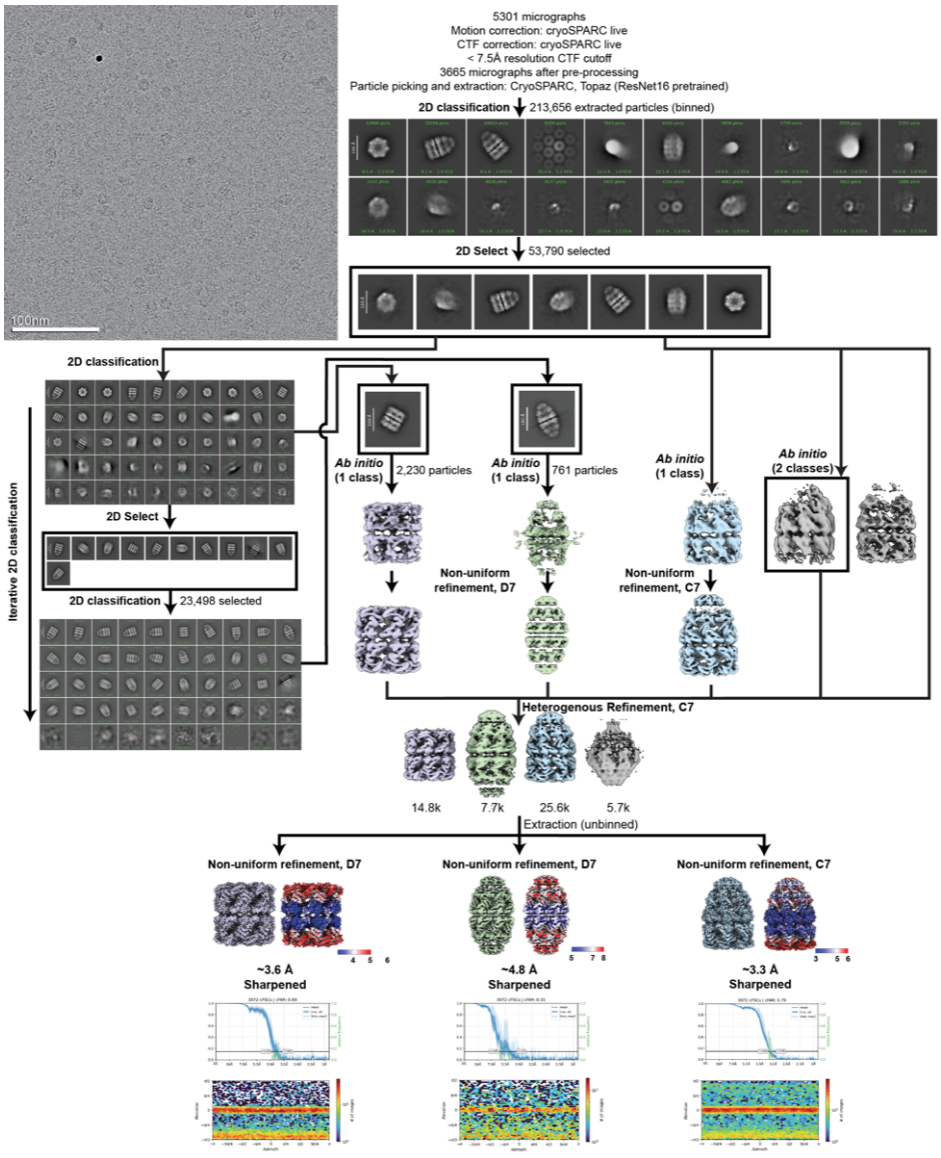


Figure B.8: Processing pipeline of the 300 ms GroEL/ES + ATP dataset. Representative processing pipeline for the 300 ms GroEL+ GroES + ATP dataset. Motion and CTF correction were performed in cryoSPARC Live. Micrographs were exported to cryoSPARC, and particles were picked using Topaz with a pretrained model, followed by initial 2D classification and selection of GroEL/ES complexes, all resembling the "bullet" (single-capped) conformation. Further rounds of 2D classification revealed two additional conformational states - the barrel (uncapped), and football (double-capped). Ab initio reconstructions were performed independently for each of the three conformational states. These reconstructions, along with a junk volume, were used in heterogeneous refinement with C7 symmetry imposed to further refine particle sets. Sorted particles were re-extracted and refined with symmetry to yield final maps at ∼3.3 Å (bullet), ∼3.6 Å (barrel), and ∼4.8 Å (football) resolution. Final maps are shown next to local resolution maps. The conical FSC plot overlaid with a resolution histogram and corresponding angular distribution plot are shown below the final maps.

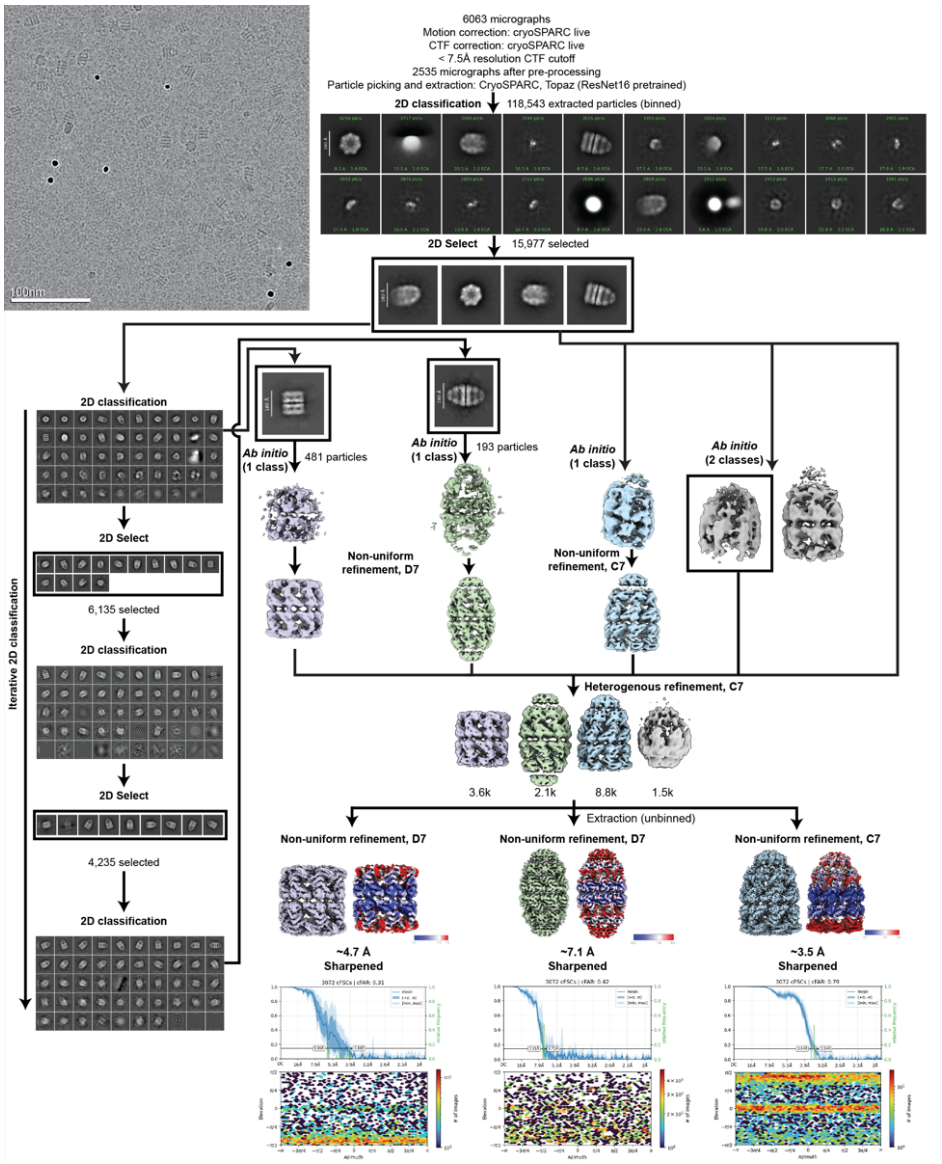


Figure B.9: Processing pipeline of the 700 ms GroEL/ES + ATP dataset. Representative processing pipeline for the 700 ms GroEL + GroES + ATP dataset. Motion and CTF correction were performed in cryoSPARC Live. Micrographs were exported to cryoSPARC, and particles were picked using Topaz with a pretrained model, followed by initial 2D classification and selection of GroEL/ES complexes. Further 2D classification revealed three distinct conformational states: bullet (single-capped), barrel (uncapped), and football (double-capped). Ab initio reconstructions were performed independently for each conformation, and the resulting volumes were used in a heterogeneous refinement job with C7 symmetry imposed to sort particles. Sorted particles were re-extracted and refined with appropriate symmetry to yield final maps at ∼3.5 Å (bullet), ∼4.7 Å (barrel), and ∼7.1 Å (football) resolution. Final maps are shown next to local resolution maps. The conical FSC plot overlaid with a resolution histogram and corresponding angular distribution plot are shown below the final maps.

C

Appendix: Polθ

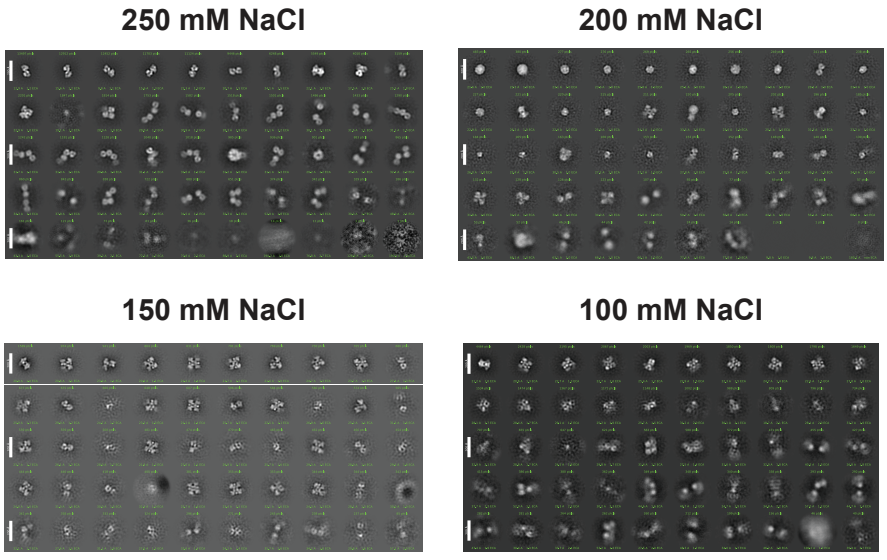


Figure C.1: Full set of 2D class averages from negative stain PolθHel stability tests. Complete 2D class average data from negative stain data of PolθHel under 250 mM, 200 mM, 150 mM, and 100 mM NaCl conditions in the absence of DNA.

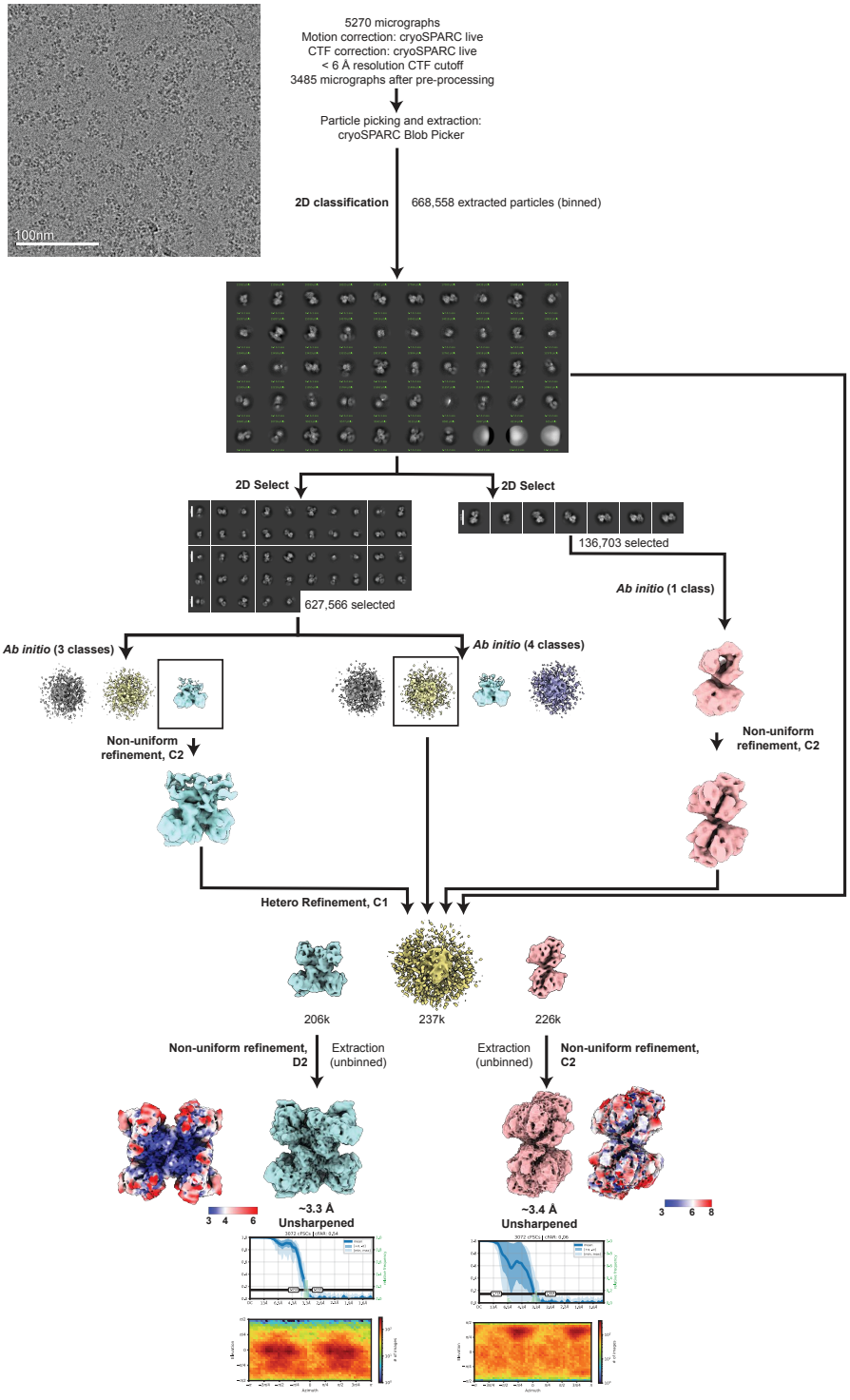


Figure C.2: Cryo-EM data processing pipeline of DNA-free PolθHel at 250 mM NaCl (control).

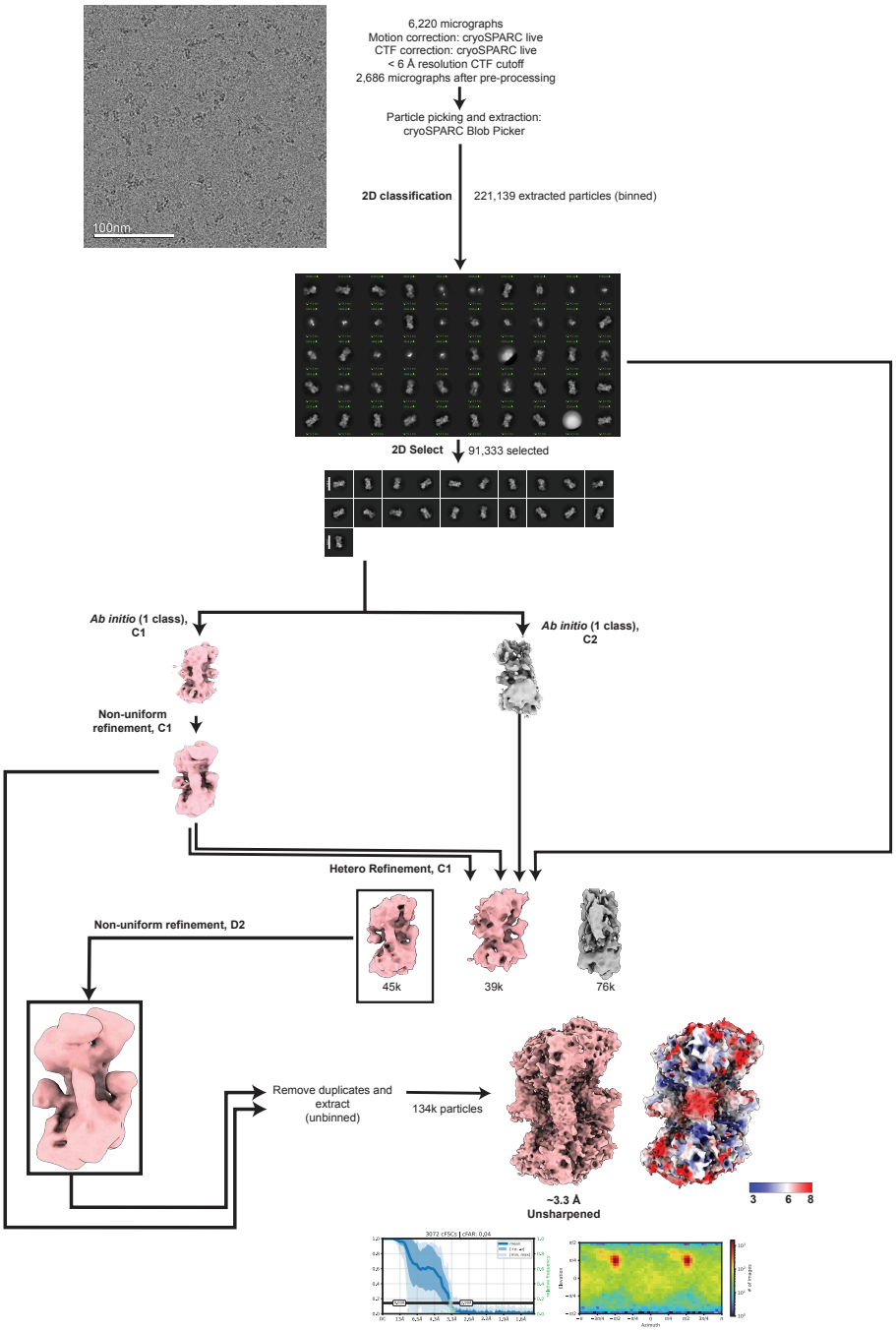


Figure C.3: Cryo-EM data processing pipeline of PolθHel at 100 mM NaCl ~5 hr after DNA addition (control).

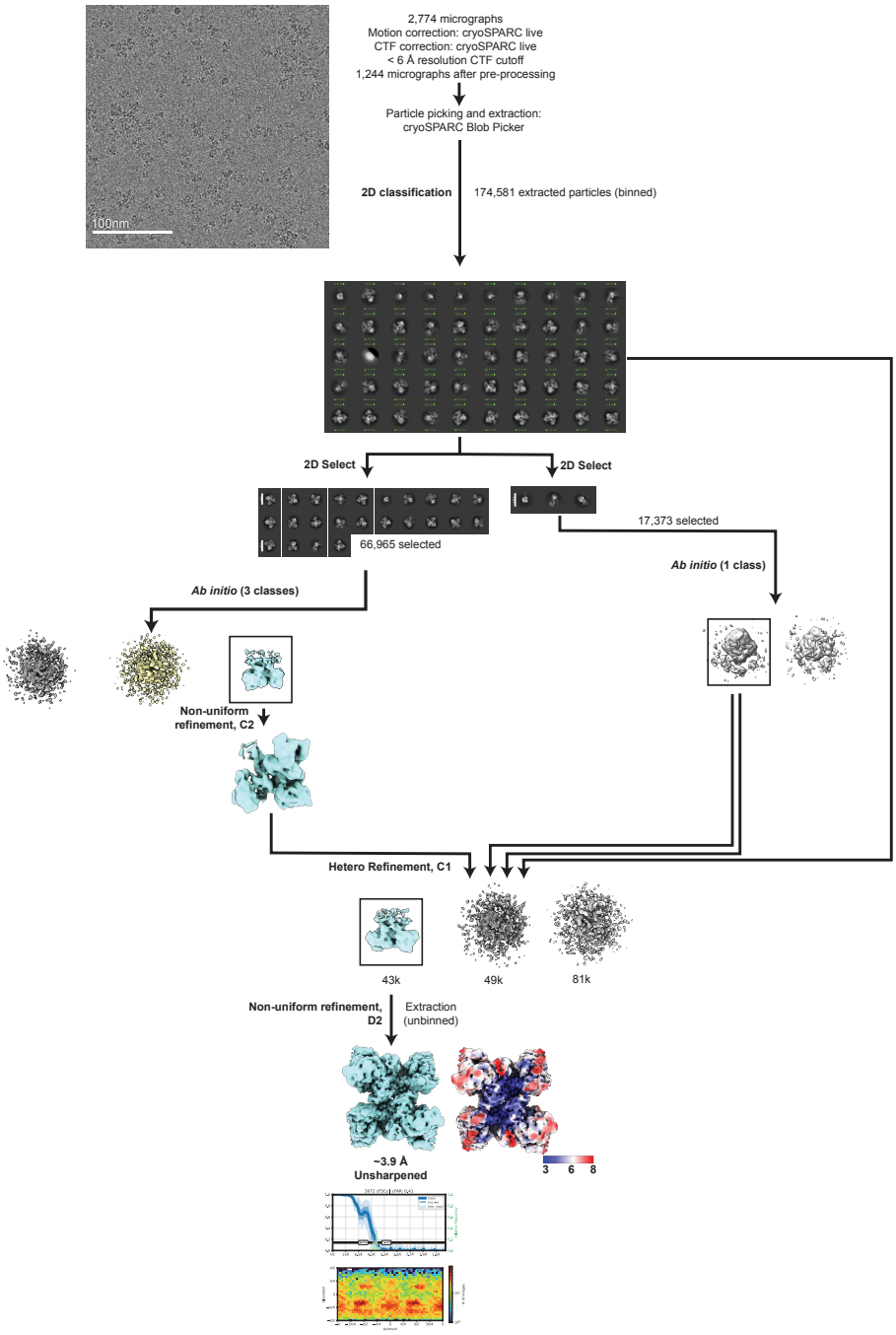


Figure C.4: Cryo-EM data processing pipeline of DNA-free PolθHel at 100 mM NaCl (control).

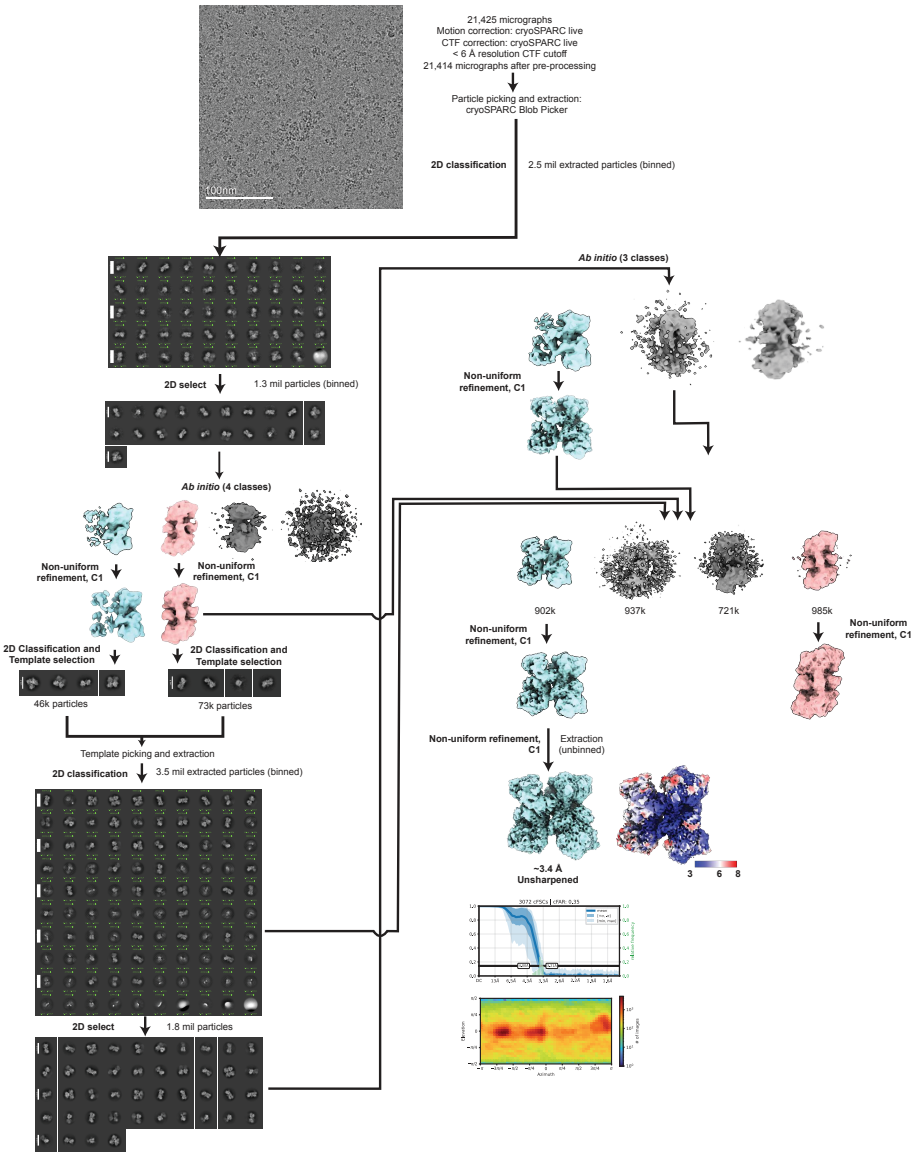


Figure C.5: Cryo-EM data processing pipeline of PolθHel at 100 mM NaCl 5 min after DNA addition.

Table C.1: Cryo-EM data collection and image analysis parameters for PolθHel

Parameter	**250 mM, (-) DNA**	**100 mM, (+) DNA, 5 hr**	**100 mM, (-) DNA**	**100 mM, (+) DNA, 5 min**	**100 mM, (+) DNA, 10 min**
Data collection					
Microscope	Talos Arctica	Talos Arctica	Talos Arctica	Talos Arctica	Talos Arctica
Detector	TFS Falcon 4i	TFS Falcon 4i	TFS Falcon 4i	TFS Falcon 4i	TFS Falcon 4i
Voltage (kV)	200	200	200	200	200
Magnification	150,000/189,189	150,000/189,189	150,000/189,189	150,000/189,189	150,000/189,189
Data acq. software	EPU v3.9.1.8206	EPU v3.9.1.8206	EPU v3.9.1.8206	EPU v3.9.1.8206	EPU v3.9.1.8206
Exposure navigation	Image shift to 8μm	Image shift to 8μm	Image shift to 8μm	Image shift to 8μm	Image shift to 8μm
Total exposure ($e^-/Å^2$)	50	50	50	50	50
Exposure rate (e^-/pixel/sec)	7.42	7.41	6.35	6.35	6.34
Exposure time (s)	3.7	3.7	4.28	4.28	4.31
Number of frames	1,296	1,296	1,008	1,377	1,251
Pixel size (Å) per pixel	0.74	0.74	0.74	0.74	0.74
Defocus range (μm)	-1 to -1.4	-1 to -1.4	-1 to -1.4	-1 to -1.4	-1 to -1.4
Micrographs collected (no.)	5,270	6,220	2,774	21,425	6,313
Data analysis					
Processing package	cryoSPARC	cryoSPARC	cryoSPARC	cryoSPARC	cryoSPARC
Total extracted picks (no.)	668,558	221,139	174,581	2,652,826	496,482
Tetramers					
Particles used for 3D (no.)	205,655	N/A	174,581	3,544,297	N/A
Final particles (no.)	203,264	N/A	42,961	902,281	N/A
Symmetry	D2	N/A	D2	C1	N/A
Global resolution (Å)					
FSC = 0.143 (no mask / masked) (Å)	3.7 / 3.3	N/A	4.4 / 3.9	3.7 / 3.4	N/A
Local resolution range (Å)	2.8–10.8	N/A	3.4–32.4	1.6–47.7	N/A
Map sharpening B-factor ($Å^2$)	-123.4	N/A	-144.0	-134.4	N/A
cFSCs cFAR	0.54	N/A	0.41	0.35	N/A
Dimers					
Particles used for 3D (no.)	226,064	221,139	N/A	3,544,297	N/A
Final particles (no.)	223,464	134,360	N/A	984,520	N/A
Symmetry	C2	C2	N/A	C1	N/A
Global resolution (Å)					
FSC = 0.143 (no mask / masked) (Å)	3.8 / 3.4	3.8 / 3.3	N/A	N/A	N/A
Local resolution range (Å)	1.6–50.4	2.8–49.7	N/A	N/A	N/A
Map sharpening B-factor ($Å^2$)	-114.0	-87.5	N/A	N/A	N/A
cFSCs cFAR	0.06	0.04	N/A	N/A	N/A

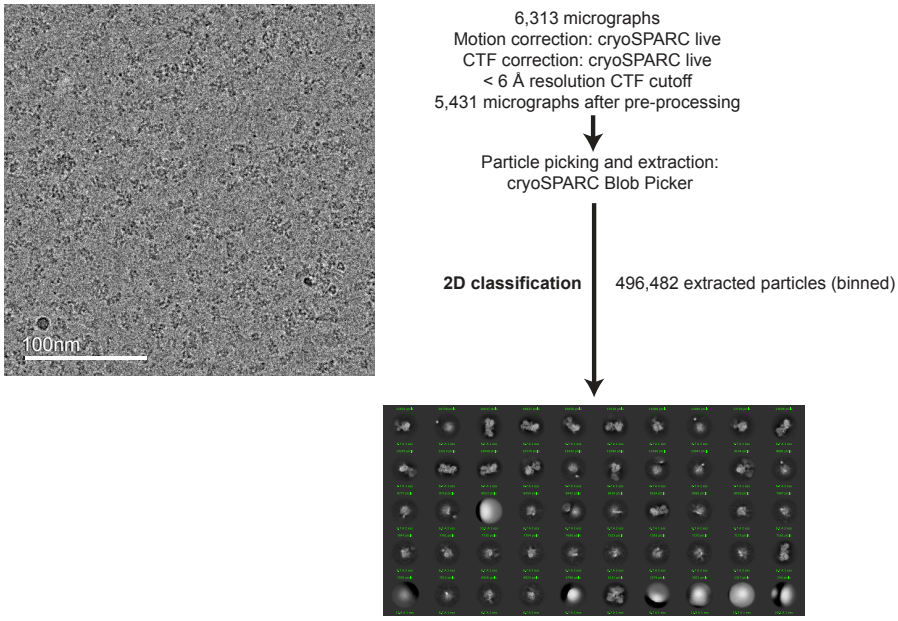


Figure C.6: Cryo-EM data processing pipeline of PolθHel at 100 mM NaCl 10 min after DNA addition.

Bibliography

[1] Christopher M. Dobson. Protein folding and misfolding. *Nature*, 426(6968):884–890, December 2003.

[2] J.E. Walker, M. Saraste, M.J. Runswick, and N.J. Gay. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *The EMBO Journal*, 1(8):945–951, August 1982.

[3] Cristina Puchades, Colby R. Sandate, and Gabriel C. Lander. The molecular principles governing the activity and functional diversity of AAA+ proteins. *Nature Reviews Molecular Cell Biology*, 21(1):43–58, January 2020.

[4] Mia Shin, Edmond R. Watson, Albert S. Song, Jeffrey T. Mindrebo, Scott J. Novick, Patrick R. Griffin, R. Luke Wiseman, and Gabriel C. Lander. Structures of the human LONP1 protease reveal regulatory steps involved in protease activation. *Nature Communications*, 12(1):3239, May 2021.

[5] Katherine Henzler-Wildman and Dorothee Kern. Dynamic personalities of proteins. *Nature*, 450(7172):964–972, December 2007.

[6] Andrew Leach. *Molecular Modelling: Principles and Applications*. Pearson Education, second edition, January 2001.

[7] Amy J. Fernandez and James M. Berger. Mechanisms of hexameric helicases. *Critical Reviews in Biochemistry and Molecular Biology*, 56(6):621–639, August 2021.

[8] James Pawley. *Handbook of Biological Confocal Microscopy*. Springer Science & Business Media, June 2006.

[9] Erin M. Conroy, Jia Jun Li, Hyungki Kim, and W. Russ Algar. Self-Quenching, Dimerization, and Homo-FRET in Hetero-FRET Assemblies with Quantum Dot Donors and Multiple Dye Acceptors. *The Journal of Physical Chemistry C*, 120(31):17817–17828, August 2016.

[10] Lothar Schermelleh, Alexia Ferrand, Thomas Huser, Christian Eggeling, Markus Sauer, Oliver Biehlmaier, and Gregor P. C. Drummen. Super-resolution microscopy demystified. *Nature Cell Biology*, 21(1):72–84, January 2019.

[11] Rupsa Datta, Tiffany M. Heaster, Joe T. Sharick, Amani A. Gillette, and Melissa C. Skala. Fluorescence lifetime imaging microscopy: Fundamentals and advances in

instrumentation, analysis, and applications. *Journal of Biomedical Optics*, 25(7):071203, May 2020.

[12] Guillaume Jacquemet, Alexandre F. Carisey, Hellyeh Hamidi, Ricardo Henriques, and Christophe Leterrier. The cell biologist's guide to super-resolution microscopy. *Journal of Cell Science*, 133(11):jcs240713, June 2020.

[13] Yves F. Dufrêne, Toshio Ando, Ricardo Garcia, David Alsteens, David Martinez-Martin, Andreas Engel, Christoph Gerber, and Daniel J. Müller. Imaging modes of atomic force microscopy for application in molecular and cell biology. *Nature Nanotechnology*, 12(4):295–307, April 2017.

[14] Toshio Ando. High-speed atomic force microscopy. *Current Opinion in Chemical Biology*, 51:105–112, August 2019.

[15] Takayuki Uchihashi and Simon Scheuring. Applications of high-speed atomic force microscopy to real-time visualization of dynamic biomolecular processes. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1862(2):229–240, February 2018.

[16] Kien Xuan Ngo, Noriyuki Kodera, Eisaku Katayama, Toshio Ando, and Taro QP Uyeda. Cofilin-induced unidirectional cooperative conformational changes in actin filaments revealed by high-speed atomic force microscopy. *eLife*, 4:e04806, February 2015.

[17] Neal Crampton, Masatoshi Yokokawa, David T. F. Dryden, J. Michael Edwardson, Desirazu N. Rao, Kunio Takeyasu, Shige H. Yoshimura, and Robert M. Henderson. Fast-scan atomic force microscopy reveals that the type III restriction enzyme EcoP15I is capable of DNA translocation and looping. *Proceedings of the National Academy of Sciences of the United States of America*, 104(31):12755–12760, July 2007.

[18] Pierre-Emmanuel Milhiet, Daisuke Yamamoto, Olivia Berthoumieu, Patrice Dosset, Christian Le Grimellec, Jean-Michel Verdier, Stéphane Marchal, and Toshio Ando. Deciphering the Structure, Growth and Assembly of Amyloid-Like Fibrils Using High-Speed Atomic Force Microscopy. *PLOS ONE*, 5(10):e13240, October 2010.

[19] Greg L. Hura, Angeli L. Menon, Michal Hammel, Robert P. Rambo, Farris L. Poole II, Susan E. Tsutakawa, Francis E. Jenney Jr, Scott Classen, Kenneth A. Frankel, Robert C. Hopkins, Sung-jae Yang, Joseph W. Scott, Bret D. Dillard, Michael W. W. Adams, and John A. Tainer. Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nature Methods*, 6(8):606–612, August 2009.

[20] Nigel M Kirby and Nathan P Cowieson. Time-resolved studies of dynamic biomolecules using small angle X-ray scattering. *Current Opinion in Structural Biology*, 28:41–46, October 2014.

[21] Hyun Sun Cho, Friedrich Schotte, Valentyn Stadnytskyi, and Philip Anfinrud. Time-resolved X-ray Scattering Studies of Proteins. *Current opinion in structural biology*, 70:99–107, October 2021.

[22] Arthur Kowalsky. Nuclear Magnetic Resonance Studies of Proteins. *Journal of Biological Chemistry*, 237(6):1807–1819, June 1962.

[23] L. D. Hall. Nuclear Magnetic Resonance. In Melville L. Wolfrom, editor, *Advances in Carbohydrate Chemistry*, volume 19, pages 51–93. Academic Press, January 1964.

[24] Shiou-Ru Tzeng, Yin-Chu Tseng, Chien-Chu Lin, Chia-Ying Hsu, Shing-Jong Huang, Yi-Ting Kuo, and Chung-I Chang. Molecular insights into substrate recognition and discrimination by the N-terminal domain of Lon AAA+ protease. *eLife*, 10:e64056, April 2021.

[25] Bernd Reif, Sharon E. Ashbrook, Lyndon Emsley, and Mei Hong. Solid-state NMR spectroscopy. *Nature Reviews Methods Primers*, 1(1):2, January 2021.

[26] Matthew J Smith, Christopher B Marshall, Francois-Xavier Theillet, Andres Binolfi, Philipp Selenko, and Mitsuhiko Ikura. Real-time NMR monitoring of biological activities in complex physiological environments. *Current Opinion in Structural Biology*, 32:39–47, June 2015.

[27] Johann J. Eicher, Jacky L. Snoep, and Johann M. Rohwer. Determining Enzyme Kinetics for Systems Biology with Nuclear Magnetic Resonance Spectroscopy. *Metabolites*, 2(4):818–843, December 2012.

[28] Lawrence Bragg. X-Ray Crystallography. *Scientific American*, 219(1):58–74, 1968.

[29] Eran Eyal, Sergey Gerzon, Vladimir Potapov, Marvin Edelman, and Vladimir Sobolev. The Limit of Accuracy of Protein Modeling: Influence of Crystal Packing on Protein Structure. *Journal of Molecular Biology*, 351(2):431–442, August 2005.

[30] Amir Banari, Amit K. Samanta, Anna Munke, Tim Laugks, Saša Bajt, Kay Grünewald, Thomas C. Marlovits, Jochen Küpper, Filipe R. N. C. Maia, Henry N. Chapman, Dominik Oberthür, and Carolin Seuring. Advancing time-resolved structural biology: Latest strategies in cryo-EM and X-ray crystallography. *Nature Methods*, 22(7):1420–1435, July 2025.

[31] Carlo Cercignani. The boltzmann equation. In *The Boltzmann Equation and Its Applications*, pages 40–103. Springer, 1988.

[32] Rafael C. Bernardi, Marcelo C. R. Melo, and Klaus Schulten. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1850(5):872–877, May 2015.

[33] Alessandro Barducci, Massimiliano Bonomi, and Michele Parrinello. Metadynamics. *WIREs Computational Molecular Science*, 1(5):826–843, 2011.

[34] Johannes Kästner. Umbrella sampling. *WIREs Computational Molecular Science*, 1(6):932–942, 2011.

[35] Fabrizio Marinelli, Fabio Pietrucci, Alessandro Laio, and Stefano Piana. A Kinetic Model of Trp-Cage Folding from Multiple Biased Molecular Dynamics Simulations. *PLOS Computational Biology*, 5(8):e1000452, August 2009.

[36] Thomas J. Piggot, Daniel A. Holdbrook, and Syma Khalid. Electroporation of the E. coli and S. Aureus Membranes: Molecular Dynamics Simulations of Complex Bacterial

Membranes. *The Journal of Physical Chemistry B*, 115(45):13381–13388, November 2011.

[37] Jan A. Stevens, Fabian Grünewald, P. A. Marco van Tilburg, Melanie König, Benjamin R. Gilbert, Troy A. Brier, Zane R. Thornburg, Zaida Luthey-Schulten, and Siewert J. Marrink. Molecular dynamics simulation of an entire cell. *Frontiers in Chemistry*, 11:1106495, January 2023.

[38] Robert M. Glaeser, Eva Nogales, and Wah Chiu. *Single-Particle Cryo-EM of Biological Macromolecules*. IOP Publishing, May 2021.

[39] Sascha Josef Amann, Demian Keihsler, Tatyana Bodrug, Nicholas G. Brown, and David Haselbach. Frozen in time: Analyzing molecular dynamics with time-resolved cryo-EM. *Structure*, 31(1):4–19, January 2023.

[40] Yong Zi Tan and John Rubinstein. Through-grid wicking enables high-speed cryoEM specimen preparation. *Microscopy and Microanalysis*, 27(S1):526–528, August 2021.

[41] Megan J. Dobro, Linda A. Melanson, Grant J. Jensen, and Alasdair W. McDowall. Chapter Three - Plunge Freezing for Electron Cryomicroscopy. In Grant J. Jensen, editor, *Methods in Enzymology*, volume 481 of *Cryo-EM Part A Sample Preparation and Data Collection*, pages 63–82. Academic Press, January 2010.

[42] Cristina V. Iancu, William F. Tivol, Jordan B. Schooler, D. Prabha Dias, Gregory P. Henderson, Gavin E. Murphy, Elizabeth R. Wright, Zhuo Li, Zhiheng Yu, Ariane Briegel, Lu Gan, Yongning He, and Grant J. Jensen. Electron cryotomography sample preparation using the Vitrobot. *Nature Protocols*, 1(6):2813–2819, December 2006.

[43] Zonghuan Lu, Tanvir R. Shaikh, David Barnard, Xing Meng, Hisham Mohamed, Aymen Yassin, Carmen A. Mannella, Rajendra K. Agrawal, Toh-Ming Lu, and Terence Wagenknecht. Monolithic microfluidic mixing–spraying devices for time-resolved cryo-electron microscopy. *Journal of Structural Biology*, 168(3):388–395, December 2009.

[44] Märt-Erik Mäeots, Byungjin Lee, Andrea Nans, Seung-Geun Jeong, Mohammad M. N. Esfahani, Shan Ding, Daniel J. Smith, Chang-Soo Lee, Sung Sik Lee, Matthias Peter, and Radoslav I. Enchev. Modular microfluidics enables kinetic insight from time-resolved cryo-EM. *Nature Communications*, 11(1):3465, July 2020.

[45] Stefania Torino, Mugdha Dhurandhar, Annelore Stroobants, Raf Claessens, and Rouslan G. Efremov. Time-resolved cryo-EM using a combination of droplet microfluidics with on-demand jetting. *Nature Methods*, 20(9):1400–1408, September 2023.

[46] J.-F. Ménétret, W. Hofmann, R.R. Schröder, G. Rapp, and R.S. Goody. Time-resolved cryo-electron microscopic study of the dissociation of actomyosin induced by photolysis of photolabile nucleotides. *Journal of Molecular Biology*, 219(2):139–144, May 1991.

[47] Alejandra Montaño Romero, Calli Bonin, and Edward C. Twomey. C-SPAM: An open-source time-resolved specimen vitrification device with light-activated molecules. *IUCrJ*, 11(1):16–22, January 2024.

[48] Oliver F. Harder, Sarah V. Barrass, Marcel Drabbels, and Ulrich J. Lorenz. Fast viral dynamics revealed by microsecond time-resolved cryo-EM. *Nature Communications*, 14(1):5649, September 2023.

[49] H. D. White, K. Thirumurugan, M. L. Walker, and J. Trinick. A second generation apparatus for time-resolved electron cryo-microscopy using stepper motors and electrospray. *Journal of Structural Biology*, 144(1):246–252, October 2003.

[50] J. L. Rubinstein, H. Guo, Z. A. Ripstein, A. Haydaroglu, A. Au, C. M. Yip, J. M. Di Trani, S. Benlekbir, and T. Kwok. Shake-it-off: A simple ultrasonic cryo-EM specimen-preparation device. *Acta Crystallographica Section D: Structural Biology*, 75(12):1063–1070, December 2019.

[51] Michele C. Darrow, John P. Moore, R. John Walker, Klaus Doering, and Russell S. King. Chameleon: Next Generation Sample Preparation for CryoEM based on Spotiton. *Microscopy and Microanalysis*, 25(S2):994–995, August 2019.

[52] Tilak Jain, Patrick Sheehan, John Crum, Bridget Carragher, and Clinton S. Potter. Spotiton: A prototype for an integrated inkjet dispense and vitrification system for cryo-TEM. *Journal of Structural Biology*, 179(1):68–75, July 2012.

[53] Jonathan M. Voss, Oliver F. Harder, Pavel K. Olshin, Marcel Drabbels, and Ulrich J. Lorenz. Rapid melting and revitrification as an approach to microsecond time-resolved cryo-electron microscopy. *Chemical Physics Letters*, 778:138812, September 2021.

[54] Wyatt A. Curtis, Jakub Hruby, Constan R. Kruger, Sarah V. Barrass, Marcel Drabbels, and Ulrich J. Lorenz. Ultrathin Liquid Cells for Microsecond Time-Resolved Cryo-EM, May 2025.

[55] David P. Klebl, Howard D. White, Frank Sobott, and Stephen P. Muench. On-grid and in-flow mixing for time-resolved cryo-EM. *Acta Crystallographica. Section D, Structural Biology*, 77(Pt 10):1233–1240, September 2021.

[56] Venkata P. Dandey, William C. Budell, Hui Wei, Daija Bobe, Kashyap Maruthi, Mykhailo Kopylov, Edward T. Eng, Peter A. Kahn, Jenny E. Hinshaw, Nidhi Kundu, Crina M. Nimigean, Chen Fan, Nattakan Sukomon, Seth A. Darst, Ruth M. Saecker, James Chen, Brandon Malone, Clinton S. Potter, and Bridget Carragher. Time-resolved cryo-EM using Spotiton. *Nature Methods*, 17(9):897–900, September 2020.

[57] Ruth M. Saecker, Andreas U. Mueller, Brandon Malone, James Chen, William C. Budell, Venkata P. Dandey, Kashyap Maruthi, Joshua H. Mendez, Nina Molina, Edward T. Eng, Laura Y. Yen, Clinton S. Potter, Bridget Carragher, and Seth A. Darst. Early intermediates in bacterial RNA polymerase promoter melting visualized by time-resolved cryo-electron microscopy. *Nature Structural & Molecular Biology*, 31(11):1778–1788, November 2024.

[58] D. Kontziampasis, D. P. Klebl, M. G. Iadanza, C. A. Scarff, F. Kopf, F. Sobott, D. C. F. Monteiro, M. Trebbin, S. P. Muench, and H. D. White. A cryo-EM grid preparation device for time-resolved structural studies. *IUCrJ*, 6(6):1024–1031, November 2019.

[59] Monique S. Straub, Oliver F. Harder, Nathan J. Mowry, Sarah V. Barrass, Jakub Hruby, Marcel Drabbels, and Ulrich J. Lorenz. Laser flash melting cryo-EM samples to overcome preferred orientation. *Nature Methods*, 22(9):1880–1886, September 2025.

[60] H. D White, M. L Walker, and J Trinick. A Computer-Controlled Spraying-Freezing Apparatus for Millisecond Time-Resolution Electron Cryomicroscopy. *Journal of Structural Biology*, 121(3):306–313, January 1998.

[61] Niels Fischer, Andrey L. Konevega, Wolfgang Wintermeyer, Marina V. Rodnina, and Holger Stark. Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature*, 466(7304):329–333, July 2010.

[62] Sayan Bhattacharjee, Xiangsong Feng, Suvrajit Maji, Prikshat Dadhwal, Zhening Zhang, Zuben P. Brown, and Joachim Frank. Time resolution in cryo-EM using a PDMS-based microfluidic chip assembly and its application to the study of HflX-mediated ribosome recycling. *Cell*, 187(3):782–796.e23, February 2024.

[63] Shanti Pal Gangwar, Maria V. Yelshanskaya, Kirill D. Nadezhdin, Laura Y. Yen, Thomas P. Newton, Muhammed Aktolun, Maria G. Kurnikova, and Alexander I. Sobolevsky. Kainate receptor channel opening and gating mechanism. *Nature*, 630(8017):762–768, June 2024.

[64] Tatyana Bodrug, Kaeli A. Welsh, Derek L. Bolhuis, Ethan Paulakonis, Raquel C. Martinez-Chacin, Bei Liu, Nicholas Pinkin, Thomas Bonacci, Liying Cui, Pengning Xu, Olivia Roscow, Sascha Josef Amann, Irina Grishkovskaya, Michael J. Emanuele, Joseph S. Harrison, Joshua P. Steimel, Klaus M. Hahn, Wei Zhang, Ellen D. Zhong, David Haselbach, and Nicholas G. Brown. Time-resolved cryo-EM (TR-EM) analysis of substrate polyubiquitination by the RING E3 anaphase-promoting complex/cyclosome (APC/C). *Nature Structural & Molecular Biology*, 30(11):1663–1674, November 2023.

[65] Thiago V. Seraphim and Walid A. Houry. AAA+ proteins. *Current Biology*, 30(6):R251–R257, March 2020.

[66] Yousuf A. Khan, K. Ian White, and Axel T. Brunger. The AAA+ superfamily: A review of the structural and mechanistic principles of these molecular machines. *Critical Reviews in Biochemistry and Molecular Biology*, 57(2):156–187, March 2022.

[67] Michael J. Baker, Takashi Tatsuta, and Thomas Langer. Quality Control of Mitochondrial Proteostasis. *Cold Spring Harbor Perspectives in Biology*, 3(7):a007559, July 2011.

[68] Timothy Wai and Thomas Langer. Mitochondrial Dynamics and Metabolic Regulation. *Trends in Endocrinology & Metabolism*, 27(2):105–117, February 2016.

[69] Yew Mun Lee, Dongxue Hu, and Yih-Cherng Liou. Importance of Mitochondrial Quality Control in Parkinson's Disease: The Potential Interplay of Mitochondrial Unfolded Protein Response and Mitophagy. In Canhua Huang and Yuanyuan Zhang, editors, *Oxidative Stress: Human Diseases and Medicine*, pages 103–131. Springer, Singapore, 2021.

[70] Ying Guo, Teng Guan, Kashfia Shafiq, Qiang Yu, Xin Jiao, Donghui Na, Meiyu Li, Guohui Zhang, and Jiming Kong. Mitochondrial dysfunction in aging. *Ageing Research Reviews*, 88:101955, July 2023.

[71] Xi He, Jiankang Liu, and Wei-Jin Zang. Mitochondrial homeostasis and redox status in cardiovascular diseases: Protective role of the vagal system. *Free Radical Biology and Medicine*, 178:369–379, January 2022.

[72] Roman M. Levytskyy, Iryna Bohovych, and Oleh Khalimonchuk. Metalloproteases of the Inner Mitochondrial Membrane. *Biochemistry*, 56(36):4737–4746, September 2017.

[73] Yuichi Matsushima. Chapter Eight - Mitochondrial AAA+ proteases. In Laurie S. Kaguni and Fuyuhiko Tamanoi, editors, *The Enzymes*, volume 54 of *History of , Current Topics and Future Perspectives*, pages 205–220. Academic Press, January 2023.

[74] Yohsuke Ohba, Thomas MacVicar, and Thomas Langer. Regulation of mitochondrial plasticity by the i-AAA protease YME1L. *Biological Chemistry*, 401(6-7):877–890, May 2020.

[75] Naotada Ishihara, Yuu Fujita, Toshihiko Oka, and Katsuyoshi Mihara. Regulation of mitochondrial morphology through proteolytic cleavage of OPA1. *The EMBO Journal*, 25(13):2966–2977, July 2006.

[76] Klaus Leonhard, Alexandra Stiegler, Walter Neupert, and Thomas Langer. Chaperone-like activity of the AAA domain of the yeast Yme1 AAA protease. *Nature*, 398(6725):348–351, March 1999.

[77] Sandro Banfi, Maria Teresa Bassi, Grazia Andolfi, Anna Marchitiello, Stefania Zanotta, Andrea Ballabio, Giorgio Casari, and Brunella Franco. Identification and Characterization of AFG3L2, a Novel Paraplegin-Related Gene. *Genomics*, 59(1):51–58, July 1999.

[78] Mark Nolden, Sarah Ehses, Mirko Koppen, Andrea Bernacchia, Elena I. Rugarli, and Thomas Langer. The m-AAA Protease Defective in Hereditary Spastic Paraplegia Controls Ribosome Assembly in Mitochondria. *Cell*, 123(2):277–289, October 2005.

[79] Robert T. Sauer, Xue Fei, Tristan A. Bell, and Tania A. Baker. Structure and function of ClpXP, a AAA+ proteolytic machine powered by probabilistic ATP hydrolysis. *Critical Reviews in Biochemistry and Molecular Biology*, 57(2):188–204, March 2022.

[80] Karolina Szczepanowska, Priyanka Maiti, Alexandra Kukat, Eduard Hofsetz, Hendrik Nolte, Katharina Senft, Christina Becker, Benedetta Ruzzenente, Hue-Tran Hornig-Do, Rolf Wibom, Rudolf J Wiesner, Marcus Krüger, and Aleksandra Trifunovic. CLPP coordinates mitoribosomal assembly through the regulation of ERAL1 levels. *The EMBO Journal*, 35(23):2566–2583, December 2016.

[81] Yuichi Matsushima, Yuta Hirofuji, Masamune Aihara, Song Yue, Takeshi Uchiumi, Laurie S. Kaguni, and Dongchon Kang. Drosophila protease ClpXP specifically degrades DmLRPPRC1 controlling mitochondrial mRNA and translation. *Scientific Reports*, 7(1):8315, August 2017.

[82] Sundararajan Venkatesh, Jae Lee, Kamalendra Singh, Irene Lee, and Carolyn K. Suzuki. Multitasking in the mitochondrion by the ATP-dependent Lon protease. *Biochimica Et Biophysica Acta*, 1823(1):56–66, January 2012.

[83] Mia Shin, Cristina Puchades, Ananya Asmita, Neha Puri, Eric Adjei, R. Luke Wiseman, A. Wali Karzai, and Gabriel C. Lander. Structural basis for distinct operational modes and protease activation in AAA+ protease Lon. *Science Advances*, 6(21):eaba8404, May 2020.

[84] Gabriela Ondrovičová, Tong Liu, Kamalendra Singh, Bin Tian, Hong Li, Oleksandr Gakh, Dusan Perečko, Jiří Janata, Zvi Granot, Joseph Orly, Eva Kutejová, and Carolyn K. Suzuki. Cleavage Site Selection within a Folded Substrate by the ATP-dependent Lon Protease*. *Journal of Biological Chemistry*, 280(26):25103–25110, July 2005.

[85] Daniela A. Bota and Kelvin J. A. Davies. Mitochondrial Lon protease in human disease and aging: Including an etiologic classification of Lon-related diseases and disorders. *Free Radical Biology and Medicine*, 100:188–198, November 2016.

[86] Marcello Pinti, Lara Gibellini, Yongzhang Liu, Shan Xu, Bin Lu, and Andrea Cossarizza. Mitochondrial Lon protease at the crossroads of oxidative stress, ageing and cancer. *Cellular and Molecular Life Sciences*, 72(24):4807–4824, December 2015.

[87] Wonmuk Hwang and Martin Karplus. Structural basis for power stroke vs. Brownian ratchet mechanisms of motor proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 116(40):19777–19785, October 2019.

[88] Hisham Mazal, Marija Iljina, Inbal Riven, and Gilad Haran. Ultrafast pore-loop dynamics in a AAA+ machine point to a Brownian-ratchet mechanism for protein translocation. *Science Advances*, 7(36):eabg4674, September 2021.

[89] Andres H. de la Peña, Ellen A. Goodall, Stephanie N. Gates, Gabriel C. Lander, and Andreas Martin. Substrate-engaged 26S proteasome structures reveal mechanisms for ATP-hydrolysis–driven translocation. *Science*, 362(6418):eaav0725, November 2018.

[90] Cristina Puchades, Anthony J. Rampello, Mia Shin, Christopher J. Giuliano, R. Luke Wiseman, Steven E. Glynn, and Gabriel C. Lander. Structure of the mitochondrial inner membrane AAA+ protease YME1 gives insight into substrate processing. *Science*, 358(6363):eaao0464, November 2017.

[91] Stephanie N. Gates and Andreas Martin. Stairway to translocation: AAA+ motor structures reveal the mechanisms of ATP-dependent substrate translocation. *Protein Science*, 29(2):407–419, 2020.

[92] Remi Casier, Dorit Levy, Inbal Riven, Yoav Barak, and Gilad Haran. Millisecond bidirectional protein translocation by a AAA+ disaggregation machine, May 2025.

[93] Inayathulla Mohammed, Kai A. Schmitz, Niko Schenck, Dimitrios Balasopoulos, Annika Topitsch, Timm Maier, and Jan Pieter Abrahams. Catalytic cycling of human mitochondrial Lon protease. *Structure*, 30(9):1254–1268.e7, September 2022.

[94] Jeffrey T. Mindrebo and Gabriel C. Lander. Structural and mechanistic studies on human LONP1 redefine the hand-over-hand translocation mechanism. *bioRxiv*, 2024.

[95] Andreas Martin, Tania A. Baker, and Robert T. Sauer. Rebuilt AAA + motors reveal operating principles for ATP-fuelled machines. *Nature*, 437(7062):1115–1120, October 2005.

[96] Xue Fei, Tristan A Bell, Simon Jenni, Benjamin M Stinson, Tania A Baker, Stephen C Harrison, and Robert T Sauer. Structures of the ATP-fueled ClpXP proteolytic machine bound to protein substrate. *eLife*, 9:e52774, February 2020.

[97] Xiang Ye, Leland Mayne, and S. Walter Englander. A conserved strategy for structure change and energy transduction in Hsp104 and other AAA+ protein motors. *Journal of Biological Chemistry*, 297(3):101066, September 2021.

[98] Carlos López-Otín and Judith S. Bond. Proteases: Multifunctional Enzymes in Life and Disease *. *Journal of Biological Chemistry*, 283(45):30433–30437, November 2008.

[99] Özlem Doğan Ekici, Mark Paetzel, and Ross E. Dalbey. Unconventional serine proteases: Variations on the catalytic Ser/His/Asp triad configuration. *Protein Science*, 17(12):2023–2037, December 2008.

[100] Genís Valentín Gesé, Saba Shahzad, Carlos Pardo-Hernández, Anna Wramstedt, Maria Falkenberg, and B. Martin Hällberg. A dual allosteric pathway drives human mitochondrial Lon. *bioRxiv*, August 2021.

[101] Javier García-Nafría, Gabriela Ondrovičová, Elena Blagova, Vladimir M Levdikov, Jacob A Bauer, Carolyn K Suzuki, Eva Kutejová, Anthony J Wilkinson, and Keith S Wilson. Structure of the catalytic domain of the human mitochondrial Lon protease: Proposed relation of oligomer formation and activity. *Protein Science*, 19(5):987–999, May 2010.

[102] Eri Kobayashi, Kei Yura, and Yoshinori Nagai. Distinct Conformation of ATP Molecule in Solution and on Protein. *Biophysics*, 9:1–12, January 2013.

[103] Mark Hilge, Gregg Siegal, Geerten W. Vuister, Peter Güntert, Sergio M. Gloor, and Jan Pieter Abrahams. ATP-induced conformational changes of the nucleotide-binding domain of Na,K-ATPase. *Nature Structural & Molecular Biology*, 10(6):468–474, June 2003.

[104] Yan Jiang, Yuchao Zhang, Liqi Wan, Cheng Cui, Pei Guo, Da Han, and Weihong Tan. Structural basis and affinity improvement for an ATP-binding DNA aptamer. *Proceedings of the National Academy of Sciences of the United States of America*, 122(33):e2506491122, August 2025.

[105] Ľuboš Ambro, Vladimír Pevala, Gabriela Ondrovičová, Jana Bellová, Nina Kunová, Eva Kutejová, and Jacob Bauer. Mutations to a glycine loop in the catalytic site of human Lon changes its protease, peptidase and ATPase activities. *The FEBS Journal*, 281(7):1784–1797, 2014.

[106] Shanshan Li, Kan-Yen Hsieh, Shih-Chieh Su, Grigore D. Pintilie, Kaiming Zhang, and Chung-I. Chang. Molecular basis for ATPase-powered substrate translocation by the Lon AAA+ protease. *Journal of Biological Chemistry*, 297(4), October 2021.

[107] Christoph Bieniossek, Barbara Niederhauser, and Ulrich M. Baumann. The crystal structure of apo-FtsH reveals domain movements necessary for substrate unfolding and translocation. *Proceedings of the National Academy of Sciences of the United States of America*, 106(51):21579–21584, December 2009.

[108] Narayanan Eswar, David Eramian, Ben Webb, Min-Yi Shen, and Andrej Sali. Protein Structure Modeling with MODELLER. In Bostjan Kobe, Mitchell Guss, and Thomas Huber, editors, *Structural Proteomics: High-Throughput Methods*, Methods in Molecular Biology™, pages 145–159. Humana Press, Totowa, NJ, 2008.

[109] Min-yi Shen and Andrej Sali. Statistical potential for assessment and prediction of protein structures. *Protein Science*, 15(11):2507–2524, 2006.

[110] Andrew Waterhouse, Martino Bertoni, Stefan Bienert, Gabriel Studer, Gerardo Tauriello, Rafal Gumienny, Florian T. Heer, Tjaart A. P. de Beer, Christine Rempfer, and Lorenza Bordoli. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*, 46(W1):W296–W303, 2018.

[111] Pascal Benkert, Silvio C. E. Tosatto, and Dietmar Schomburg. QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Structure, Function, and Bioinformatics*, 71(1):261–277, 2008.

[112] Warren L. DeLano. Pymol: An open-source molecular graphics tool. *CCP4 Newsl. protein crystallogr*, 40(1):82–92, 2002.

[113] Kristin L. Meagher, Luke T. Redman, and Heather A. Carlson. Development of polyphosphate parameters for use with the AMBER force field. *Journal of Computational Chemistry*, 24(9):1016–1025, 2003.

[114] Alan W. Sousa da Silva and Wim F. Vranken. ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Research Notes*, 5(1):367, July 2012.

[115] Chad W. Hopkins, Scott Le Grand, Ross C. Walker, and Adrian E. Roitberg. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *Journal of Chemical Theory and Computation*, 11(4):1864–1874, April 2015.

[116] Henk Bekker, Herman JC Berendsen, EJ Dijkstra, S Achterop, R Vondrumen, David Vanderspoel, A Sijbers, H Keegstra, and MKR Renardus. GROMACS - parallel computer for molecular-dynamics simulations. In RA DeGroot and J Nadrchal, editors, *4th International Conference on Computational Physics (PC 92)*, pages 252–256, SINGAPORE, 1993. World Scientific Publishing.

[117] Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L. Klepeis, Ron O. Dror, and David E. Shaw. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function, and Bioinformatics*, 78(8):1950–1958, 2010.

[118] Daniel J. Price and Charles L. Brooks, III. A modified TIP3P water potential for simulation with Ewald summation. *The Journal of Chemical Physics*, 121(20):10096–10103, November 2004.

[119] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1):014101, January 2007.

[120] Michele Parrinello and Aneesur Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics*, 52(12):7182–7190, 1981.

[121] Berk Hess, Henk Bekker, Herman J. C. Berendsen, and Johannes G. E. M. Fraaije. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, 1997.

[122] David E. Shaw, J.P. Grossman, Joseph A. Bank, Brannon Batson, J. Adam Butts, Jack C. Chao, Martin M. Deneroff, Ron O. Dror, Amos Even, Christopher H. Fenton, Anthony Forte, Joseph Gagliardo, Gennette Gill, Brian Greskamp, C. Richard Ho, Douglas J. Ierardi, Lev Iserovich, Jeffrey S. Kuskin, Richard H. Larson, Timothy Layman, Li-Siang Lee, Adam K. Lerer, Chester Li, Daniel Killebrew, Kenneth M. Mackenzie, Shark Yeuk-Hai Mok, Mark A. Moraes, Rolf Mueller, Lawrence J. Nociolo, Jon L. Peticolas, Terry Quan, Daniel Ramot, John K. Salmon, Daniele P. Scarpazza, U. Ben Schafer, Naseer Siddique, Christopher W. Snyder, Jochen Spengler, Ping Tak Peter Tang, Michael Theobald, Horia Toma, Brian Towles, Benjamin Vitale, Stanley C. Wang, and Cliff Young. Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer. In *SC '14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 41–53, November 2014.

[123] Glenn J. Martyna, Michael L. Klein, and Mark Tuckerman. Nosé–Hoover chains: The canonical ensemble via continuous dynamics. *The Journal of Chemical Physics*, 97(4):2635–2643, 1992.

[124] Denis J. Evans and B. L. Holian. The nose–hoover thermostat. *J. Chem. Phys*, 83:4069–4074, 1985.

[125] Richard Gowers, Max Linke, Jonathan Barnoud, Tyler Reddy, Manuel Melo, Sean Seyler, Jan Domański, David Dotson, Sébastien Buchoux, Ian Kenney, and Oliver Beckstein. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. In *Python in Science Conference*, pages 98–105, Austin, Texas, 2016.

[126] Fiona B. Naughton, Irfan Alibay, Jonathan Barnoud, Estefania Barreto-Ojeda, Oliver Beckstein, Cédric Bouysset, Orion Cohen, Richard J. Gowers, Hugo MacDermott-Opeskin, and Micaela Matta. MDAnalysis 2.0 and beyond: Fast and interoperable, community driven simulation analysis. *Biophysical Journal*, 121(3):272a–273a, 2022.

[127] D. J. Barlow and J. M. Thornton. Helix geometry in proteins. *Journal of Molecular Biology*, 201(3):601–619, June 1988.

[128] Aristidis Likas, Nikos Vlassis, and Jakob J. Verbeek. The global k-means clustering algorithm. *Pattern Recognition*, 36(2):451–461, 2003.

[129] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, and Vincent Dubourg. Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, 12:2825–2830, 2011.

[130] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, and Jonathan Bright. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, 2020.

[131] L. Alexandrescu, W. Lessin, and G. C. Lander. 'Mix-it-up': Accessible time-resolved cryo-EM on the millisecond timescale. *IUCrJ*, 12(6):710–724, November 2025.

[132] Zoe L Watson, Fred R Ward, Raphaël Méheust, Omer Ad, Alanna Schepartz, Jillian F Banfield, and Jamie HD Cate. Structure of the bacterial ribosome at 2 Å resolution. *eLife*, 9:e60482, September 2020.

[133] Ka Man Yip, Niels Fischer, Elham Paknia, Ashwin Chari, and Holger Stark. Atomic-resolution protein structure determination by cryo-EM. *Nature*, 587(7832):157–161, November 2020.

[134] Takanori Nakane, Abhay Kotecha, Andrija Sente, Greg McMullan, Simonas Masiulis, Patricia M. G. E. Brown, Ioana T. Grigoras, Lina Malinauskaite, Tomas Malinauskas, Jonas Miehling, Tomasz Uchański, Lingbo Yu, Dimple Karia, Evgeniya V. Pechnikova, Erwin de Jong, Jeroen Keizer, Maarten Bischoff, Jamie McCormack, Peter Tiemeijer, Steven W. Hardwick, Dimitri Y. Chirgadze, Garib Murshudov, A. Radu Aricescu, and Sjors H. W. Scheres. Single-particle cryo-EM at atomic resolution. *Nature*, 587(7832):152–156, November 2020.

[135] Søren G. F. Rasmussen, Brian T. DeVree, Yaozhong Zou, Andrew C. Kruse, Ka Young Chung, Tong Sun Kobilka, Foon Sun Thian, Pil Seok Chae, Els Pardon, Diane Calinski, Jesper M. Mathiesen, Syed T. A. Shah, Joseph A. Lyons, Martin Caffrey, Samuel H. Gellman, Jan Steyaert, Georgios Skiniotis, William I. Weis, Roger K. Sunahara, and Brian K. Kobilka. Crystal structure of the $B2$ adrenergic receptor–Gs protein complex. *Nature*, 477(7366):549–555, September 2011.

[136] Jan Pieter Abrahams, Andrew G. W. Leslie, René Lutter, and John E. Walker. Structure at 2.8 Â resolution of F1-ATPase from bovine heart mitochondria. *Nature*, 370(6491):621–628, August 1994.

[137] J. Dubochet, J. Lepault, R. Freeman, J. A. Berriman, and J.-C. Homo. Electron microscopy of frozen water and aqueous solutions. *Journal of Microscopy*, 128(3):219–237, 1982.

[138] M. Adrian, J. Dubochet, J. Lepault, and A. W. McDowall. Cryo-electron microscopy of viruses. *Nature*, 308(5954):32–36, 1984-03-01/0007.

[139] Robert A. Grassucci, Derek J. Taylor, and Joachim Frank. Preparation of macromolecular complexes for cryo-electron microscopy. *Nature Protocols*, 2(12):3239–3246, December 2007.

[140] Xiangsong Feng and Joachim Frank. A PDMS-based Microfluidic Chip Assembly for Time-Resolved Cryo-EM (TRCEM) Sample Preparation. *Bio-Protocol*, 15(4):e5193, February 2025.

[141] Nigel Unwin. Acetylcholine receptor channel imaged in the open state. *Nature*, 373(6509):37–43, January 1995.

[142] Jeffrey A Speir, Sanjeev Munshi, Guoji Wang, Timothy S Baker, and John E Johnson. Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo-electron microscopy. *Structure*, 3(1):63–78, January 1995.

[143] Friedrich Schotte, Manho Lim, Timothy A. Jackson, Aleksandr V. Smirnov, Jayashree Soman, John S. Olson, George N. Phillips, Michael Wulff, and Philip A. Anfinrud. Watching a Protein as it Functions with 150-ps Time-Resolved X-ray Crystallography. *Science*, 300(5627):1944–1947, June 2003.

[144] Maximilian Wranik, Tobias Weinert, Chavdar Slavov, Tiziana Masini, Antonia Furrer, Natacha Gaillard, Dario Gioia, Marco Ferrarotti, Daniel James, Hannah Glover, Melissa Carrillo, Demet Kekilli, Robin Stipp, Petr Skopintsev, Steffen Brünle, Tobias Mühlethaler, John Beale, Dardan Gashi, Karol Nass, Dmitry Ozerov, Philip J. M. Johnson, Claudio Cirelli, Camila Bacellar, Markus Braun, Meitian Wang, Florian Dworkowski, Chris Milne, Andrea Cavalli, Josef Wachtveitl, Michel O. Steinmetz, and Jörg Standfuss. Watching the release of a photopharmacological drug from tubulin using time-resolved serial crystallography. *Nature Communications*, 14(1):903, February 2023.

[145] J. R. Stagno, Y. Liu, Y. R. Bhandari, C. E. Conrad, S. Panja, M. Swain, L. Fan, G. Nelson, C. Li, D. R. Wendel, T. A. White, J. D. Coe, M. O. Wiedorn, J. Knoska, D. Oberthuer, R. A. Tuckey, P. Yu, M. Dyba, S. G. Tarasov, U. Weierstall, T. D. Grant, C. D. Schwieters, J. Zhang, A. R. Ferré-D'Amaré, P. Fromme, D. E. Draper, M. Liang, M. S. Hunter, S. Boutet, K. Tan, X. Zuo, X. Ji, A. Barty, N. A. Zatsepin, H. N. Chapman, J. C. H. Spence, S. A. Woodson, and Y.-X. Wang. Structures of riboswitch RNA reaction states by mix-and-inject XFEL serial crystallography. *Nature*, 541(7636):242–246, January 2017.

[146] Gisela Brändén and Richard Neutze. Advances and challenges in time-resolved macromolecular crystallography. *Science*, 373(6558):eaba0954, August 2021.

[147] Hyunmin Kim, Junsun Park, Seyeon Lim, Sung-Hoon Jun, Mingyu Jung, and Soung-Hun Roh. Cryo-EM structures of GroEL:ES2 with RuBisCO visualize molecular contacts of encapsulated substrates in a double-cage chaperonin. *iScience*, 25(1):103704, January 2022.

[148] Sofia S. Kudryavtseva, Evgeny B. Pichkur, Igor A. Yaroshevich, Aleksandra A. Mamchur,

Irina S. Panina, Andrei V. Moiseenko, Olga S. Sokolova, Vladimir I. Muronetz, and Tatiana B. Stanishneva-Konovalova. Novel cryo-EM structure of an ADP-bound GroEL–GroES complex. *Scientific Reports*, 11(1):18241, September 2021.

[149] Alan M. Roseman, Shaoxia Chen, Helen White, Kerstin Braig, and Helen R. Saibil. The Chaperonin ATPase Cycle: Mechanism of Allosteric Switching and Movements of Substrate-Binding Domains in GroEL. *Cell*, 87(2):241–251, October 1996.

[150] W. A. Fenton and A. L. Horwich. GroEL-mediated protein folding. *Protein Science*, 6(4):743–760, April 1997.

[151] Navneet K. Tyagi, Wayne A. Fenton, and Arthur L. Horwich. GroEL/GroES cycling: ATP binds to an open ring before substrate protein favoring protein binding and production of the native state. *Proceedings of the National Academy of Sciences of the United States of America*, 106(48):20264–20269, December 2009.

[152] Damian Madan, Zong Lin, and Hays S. Rye. Triggering Protein Folding within the GroEL-GroES Complex. *The Journal of Biological Chemistry*, 283(46):32003–32013, November 2008.

[153] Mengyu Wu, Gabriel C. Lander, and Mark A. Herzik. Sub-2 Angstrom resolution structure determination using single-particle cryo-EM at 200 keV. *Journal of Structural Biology: X*, 4:100020, February 2020.

[154] Akhtar Ali and Marilyn J. Roossinck. Rapid and efficient purification of Cowpea chlorotic mottle virus by sucrose cushion ultracentrifugation. *Journal of Virological Methods*, 141(1):84–86, April 2007.

[155] Jack Godek, Jared Sivinski, Edmond R. Watson, Felicidad Lebario, Wenli Xu, Mckayla Stevens, Christopher J. Zerio, Andrew J. Ambrose, Xiaoyi Zhu, Carlee A. Trindl, Donna D. Zhang, Steven M. Johnson, Gabriel C. Lander, and Eli Chapman. Bis-sulfonamido-2-phenylbenzoxazoles Validate the GroES/EL Chaperone System as a Viable Antibiotic Target. *Journal of the American Chemical Society*, 146(30):20845–20856, July 2024.

[156] Cristina V. Iancu, William F. Tivol, Jordan B. Schooler, D. Prabha Dias, Gregory P. Henderson, Gavin E. Murphy, Elizabeth R. Wright, Zhuo Li, Zhiheng Yu, Ariane Briegel, Lu Gan, Yongning He, and Grant J. Jensen. Electron cryotomography sample preparation using the Vitrobot. *Nature Protocols*, 1(6):2813–2819, December 2006.

[157] Chelsea R. Marr, Samir Benlekbir, and John L. Rubinstein. Fabrication of carbon films with ∼500nm holes for cryo-EM with a direct detector device. *Journal of Structural Biology*, 185(1):42–47, January 2014.

[158] Ali Punjani, John L. Rubinstein, David J. Fleet, and Marcus A. Brubaker. cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*, 14(3):290–296, 2017.

[159] Michael M. Vilenchik and Alfred G. Knudson. Endogenous DNA double-strand breaks:

Production, fidelity of repair, and induction of cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 100(22):12871–12876, 2003.

[160] Kum Kum Khanna and Stephen P. Jackson. DNA double-strand breaks: Signaling, repair and the cancer connection. *Nature Genetics*, 27(3):247–254, March 2001.

[161] Ralph Scully, Arvind Panday, Rajula Elango, and Nicholas A. Willis. DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nature Reviews Molecular Cell Biology*, 20(11):698–714, November 2019.

[162] Raphael Ceccaldi, Beatrice Rondinelli, and Alan D. D'Andrea. Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends in Cell Biology*, 26(1):52–64, January 2016.

[163] Dale A. Ramsden, Juan Carvajal-Garcia, and Gaorav P. Gupta. Mechanism, cellular functions and cancer roles of polymerase-theta-mediated DNA end joining. *Nature Reviews Molecular Cell Biology*, 23(2):125–140, February 2022.

[164] Agnel Sfeir and Lorraine S. Symington. Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? *Trends in Biochemical Sciences*, 40(11):701–714, November 2015.

[165] Eric Weterings and David J. Chen. The endless tale of non-homologous end-joining. *Cell Research*, 18(1):114–124, January 2008.

[166] Markus Löbrich and Penny Jeggo. A Process of Resection-Dependent Nonhomologous End Joining Involving the Goddess Artemis. *Trends in Biochemical Sciences*, 42(9):690–701, September 2017.

[167] Michael P. Conlin, Dylan A. Reid, George W. Small, Howard H. Chang, Go Watanabe, Michael R. Lieber, Dale A. Ramsden, and Eli Rothenberg. DNA Ligase IV Guides End-Processing Choice during Nonhomologous End Joining. *Cell Reports*, 20(12):2810–2819, September 2017.

[168] Nicholas R. Pannunzio, Go Watanabe, and Michael R. Lieber. Nonhomologous DNA end-joining for repair of DNA double-strand breaks. *Journal of Biological Chemistry*, 293(27):10512–10523, July 2018.

[169] G R Smith. Homologous recombination in procaryotes. *Microbiological Reviews*, 52(1):1–28, March 1988.

[170] Shunichi Takeda, Kyoko Nakamura, Yoshihito Taniguchi, and Tanya T. Paull. Ctp1/CtIP and the MRN Complex Collaborate in the Initial Steps of Homologous Recombination. *Molecular Cell*, 28(3):351–352, November 2007.

[171] Rohit Prakash, Yu Zhang, Weiran Feng, and Maria Jasin. Homologous recombination and human health: The roles of BRCA1, BRCA2, and associated proteins. *Cold Spring Harbor Perspectives in Biology*, 7(4):a016600, 2015.

[172] Raphael Ceccaldi, Jessica C. Liu, Ravindra Amunugama, Ildiko Hajdu, Benjamin Primack, Mark I. R. Petalcorin, Kevin W. O'Connor, Panagiotis A. Konstantinopou-

los, Stephen J. Elledge, Simon J. Boulton, Timur Yusufzai, and Alan D. D'Andrea. Homologous-recombination-deficient tumours are dependent on Polθ-mediated repair. *Nature*, 518(7538):258–262, February 2015.

[173] Peng He and Wei Yang. Template and primer requirements for DNA Pol θ-mediated end joining. *Proceedings of the National Academy of Sciences of the United States of America*, 115(30):7747–7752, July 2018.

[174] Hailong Wang and Xingzhi Xu. Microhomology-mediated end joining: New players join the team. *Cell & Bioscience*, 7(1):6, January 2017.

[175] Pedro A. Mateos-Gomez, Tatiana Kent, Sarah K. Deng, Shane McDevitt, Ekaterina Kashkina, Trung M. Hoang, Richard T. Pomerantz, and Agnel Sfeir. The helicase domain of Polθ counteracts RPA to promote alt-NHEJ. *Nature Structural & Molecular Biology*, 24(12):1116–1123, 2017.

[176] Jeffrey M Schaub, Michael M Soniat, and Ilya J Finkelstein. Polymerase theta-helicase promotes end joining by stripping single-stranded DNA-binding proteins and bridging DNA ends. *Nucleic Acids Research*, 50(7):3911–3921, April 2022.

[177] Carel Fijen, Lea Drogalis Beckham, Dante Terino, Yuzhen Li, Dale A. Ramsden, Richard D. Wood, Sylvie Doublié, and Eli Rothenberg. Sequential requirements for distinct Polθ domains during theta-mediated end joining. *Molecular Cell*, 84(8):1460–1474.e6, April 2024.

[178] Christopher J. Zerio, Yonghong Bai, Brian A. Sosa-Alvarado, Timothy Guzi, and Gabriel C. Lander. Human polymerase θ helicase positions DNA microhomologies for double-strand break repair. *Nature Structural & Molecular Biology*, 32:1061–1068, June 2025.

[179] Fumiaki Ito, Ziyuan Li, Leonid Minakhin, Htet A. Khant, Richard T. Pomerantz, and Xiaojiang S. Chen. Structural basis for Polθ-helicase DNA binding and microhomology-mediated end-joining. *Nature Communications*, 16(1):3725, April 2025.

[180] Susanna Stroik, Juan Carvajal-Garcia, Dipika Gupta, Alyssa Edwards, Adam Luthman, David W. Wyatt, Rachel L. Dannenberg, Wanjuan Feng, Thomas A. Kunkel, Gaorav P. Gupta, Mark Hedglin, Richard Wood, Sylvie Doublié, Eli Rothenberg, and Dale A. Ramsden. Stepwise requirements for polymerases δ and θ in theta-mediated end joining. *Nature*, 623(7988):836–841, November 2023.

[181] Joseph A. Newman, Christopher D. O. Cooper, Hazel Aitkenhead, and Opher Gileadi. Structure of the Helicase Domain of DNA Polymerase Theta Reveals a Possible Role in the Microhomology-Mediated End-Joining Pathway. *Structure*, 23(12):2319–2330, December 2015.

[182] Hanbo Guo, YanXia Wang, Jun Mao, Huimin Zhao, Yuntong He, Yuandong Hu, Jing Li, Yujie Liu, Zheng Guan, Allen Guo, Xiaodan Ni, Fengying Zhang, and Jie Heng. Cryo-EM structure of DNA polymerase θ helicase domain in complex with inhibitor novobiocin, January 2023.