# Using cryo-electron microscopy and molecular dynamics simulations to understand complex molecular systems

Zephyr Ford

St Anne's College
University of Oxford

*A thesis submitted for the degree of*
*Doctor of Philosophy*

Michaelmas 2025

## Abstract

Structural studies of biological complexes require an integration of techniques to understand both structure and function. While methods such as cryogenic electron microscopy (cryo-EM) can generate high-resolution static structures, the connection to function requires an approach which captures data over time. In this thesis I utilised molecular dynamics (MD) simulations and cryo-EM to study the structure and function of biological complexes.

MD simulations can probe the effects of mutations on the interactions between biological molecules. I investigated the consequences of changing the electrostatic surface of a potential protein nanopore on the translocation of double-stranded RNA through the pore lumen. I showed that the impact on translocation speed is due to direct interactions between the RNA and the protein, and to the indirect action of the change in ion flux.

The ribosome of the marine-dwelling archaeon *Nitrosopumilus maritimus* was studied using cryo-EM. I solved novel high-resolution structures of the large ribosomal subunit and built atomic models of the ribosomal RNA and proteins. I compared the presence of archaea-eukarya-specific proteins to the published structures of other archaeal ribosomes and found additional structural roles for some proteins compared to their homologues.

Many prokaryotic cells have proteinaceous lattices on their cell surface known as an S-layer. These S-layers are known to be very divergent in sequence and structure across domains down to species in the same genus. I used cryogenic electron tomography to solve medium-resolution structures of S-layers from four archaeal species, *N. maritimus*, *Saccharolobus solfataricus*, *Pyrobaculum oguniense*, and *Pyrobaculum arsenaticum*. I observed variation in S-layer architecture despite the broad presence of immunoglobulin-like folds. I used MD simulations to identify the divalent cation species likely to occupy the coordination sites I located in the *Haloferax volcanii* S-layer. This combined published cryo-EM data with MD data for an integrative approach to structural biology.

# Using cryo-electron microscopy and molecular dynamics simulations to understand complex molecular systems




Zephyr Ford

St Anne's College

University of Oxford

A thesis submitted for the degree of

*Doctor of Philosophy*

Michaelmas 2025

this thesis is dedicated to
the eighteen-year-old who
didn't believe they could do it

*we did it.*

# Acknowledgements

This thesis, while written all by me, has definitely been a team effort. Firstly I would like to thank my supervisors, Professor Syma Khalid and Dr. Tanmay Bharat, for their knowledge and guidance throughout my DPhil. The projects included in this thesis would not exist without them, and I have learned a lot about both the science and myself through their support over the years it took to get here.

I also have to show my appreciation for my advisors, official or otherwise: Dr. Lindsay Baker, Professor Phill Stansfeld, Dr. Anna Duncan, Dr. Kirsty McHugh, and Professor Andre Furger. There is no way I could have made it through this degree and the concurrent personal hurdles without their support, honesty, and kindness.

The biggest thanks go to all my friends who have helped drag me out of moments where I didn't have faith in myself. To Niko, who has seen me at my worst and continues to lift me back to my best, and has helped me become a mentally and physically stronger version of myself; and to Cyril, who has always been there at the moments I needed him and many others, and has shown me how passion and dedication can overcome huge hurdles: I could not have asked for a better trio to be finishing alongside, and long may the birdspotting continue. To Suzi, Larissa, Bora, Abbey, Bjarne, Danny, and all the other Wellcome Structural Biology students I've known along the way: thank you for always having an encouraging word and a willingness to chat over a cup of tea when things got rough. To Aym, my almost-twin and steadfast support: I know that any time we see each other, it's as if no time has passed at all and the love will always be there. To all of the Virtual Hangout, thank you for being there to let me vent and cheering my successes and organising social events where we can all keep in touch even as we spread further and further from Oxford; long may it continue. To my dearest "internet friends", most of whom I have now met in person: the mutual trust that we are not all secretly murderers has got us a long way so far, and I feel so lucky to have been living in a time where I could make such important connections with people that I wouldn't otherwise cross paths with, and I continue to appreciate you all every day. To all those friends who are past and present members of University College Chapel Choir, in particular Stuart and Giles: music has been such a life-saver over these past years, and it was always the greatest pleasure to sing with you and know that when I was coming back from whatever troubles, you'd still be there and delighted to see me. To my friends in SABC and Oxford Kites Wheelchair Basketball: learning new skills has been so much fun, and kept my spirits up in a way that only team sports with friends can.

Huge thanks to all the members of the Khalid group whom I have worked with, particularly Alek, Robbie, Mingyu, Iain, Reza, Astrid, Florencia, and Kahlan. The group has always had such a wonderful dynamic because of its members from undergraduates through to post-docs, and it is something really special that I know doesn't always happen in labs. I still can't always remember what exactly SBCB

stands for, but I would also like to thank everyone who has been part of SBCB throughout my time, who have made group lunches and pub trips and Christmas dinners feel relaxed, while making sure seminars were rigorous and probing questions were asked and answered. I would also like to thank the members of the Bharat group who have been so helpful over the years, in particular Andriko, Jan, and Abul, and hopefully we will cross paths again in the future.

I would like to thank all of the NHS staff who have looked after me during my DPhil, particularly my primary surgeon, Mr Bignell, who knows me "inside and out". I have been lucky to have such a wonderful team supporting me through difficult surgeries and recoveries, and I have always felt respected and well cared-for. Although it would have been nice to spend a total amount of time in the John Radcliffe that wasn't countable in months!

Finally, I would like to thank my family, in particular my parents, Dr. Mum and Dr. Dad. Having their support has been invaluable all throughout the DPhil process, from applying right down to proof-reading my thesis on a topic which is 0% related to their own areas of expertise. It has never felt like a burden to live up to their own successful PhDs, just a shared appreciation for pursuing academic knowledge to the best of our abilities. I can't wait to put my own thesis on the shelf next to theirs. Also a huge thanks to my sister, for putting up with this academic nonsense when she has so smartly escaped it! And biggest thanks to my fluffy orange boy, Cali, for always reminding me that the things I love the most are still worthy of love when he (or anything else) has put holes in them.

*be afraid and do it anyway.*

# Abstract

Structural studies of biological complexes require an integration of techniques to understand both structure and function. While methods such as cryogenic electron microscopy (cryo-EM) can generate high-resolution static structures, the connection to function requires an approach which captures data over time. In this thesis I utilised molecular dynamics (MD) simulations and cryo-EM to study the structure and function of biological complexes.

MD simulations can probe the effects of mutations on the interactions between biological molecules. I investigated the consequences of changing the electrostatic surface of a potential protein nanopore on the translocation of double-stranded RNA through the pore lumen. I showed that the impact on translocation speed is due to direct interactions between the RNA and the protein, and to the indirect action of the change in ion flux.

The ribosome of the marine-dwelling archaeon *Nitrosopumilus maritimus* was studied using cryo-EM. I solved novel high-resolution structures of the large ribosomal subunit and built atomic models of the ribosomal RNA and proteins. I compared the presence of archaea-eukarya-specific proteins to the published structures of other archaeal ribosomes and found additional structural roles for some proteins compared to their homologues.

Many prokaryotic cells have proteinaceous lattices on their cell surface known as an S-layer. These S-layers are known to be very divergent in sequence and structure across domains down to species in the same genus. I used cryogenic electron tomography to solve medium-resolution structures of S-layers from four archaeal species, *N. maritimus*, *Saccharolobus solfataricus*, *Pyrobaculum oguniense*, and *Pyrobaculum arsenaticum*. I observed variation in S-layer architecture despite the broad presence of immunoglobulin-like folds. I used MD simulations to identify the divalent cation species likely to occupy the coordination sites I located in the *Haloferax volcanii* S-layer. This combined published cryo-EM data with MD data for an integrative approach to structural biology.

# Certification

I hereby declare that this document has not already been submitted for any degree and is not being submitted as part of candidature for any other degree. I declare that the document has been written solely by me, and that all sources and materials presented but not authored by me have been acknowledged. I furthermore certify that the word count of this thesis does not exceed 50,000 words.

Signature of candidate:

*Z. C. Ford*

# Contents

**List of Figures** **xii**

**List of Tables** **xvi**

**List of Abbreviations** **xvii**

**List of Units** **xx**

**1 Introduction** **1**

1.1 Integrative structural biology . . . . . 1
1.2 Cryo-electron microscopy for solving biological structures . . . . . 4
1.2.1 Resolution revolution . . . . . 5
1.2.2 The rise of cryo-electron tomography . . . . . 6
1.3 Molecular dynamics for investigating biological structures . . . . . 6
1.4 Biomolecular Folding . . . . . 7
1.4.1 Importance of solvent . . . . . 9
1.4.2 Importance of binding partners . . . . . 11
1.4.3 Importance of other ligands . . . . . 12
1.5 Archaea . . . . . 12
1.6 Aims of this work . . . . . 14

**2 Theory and Methods** **15**

2.1 Molecular mechanics . . . . . 16
2.1.1 Classical description of biological systems . . . . . 16
2.1.2 Force fields . . . . . 16
2.1.3 Energy minimisation . . . . . 25
2.1.4 Simulations . . . . . 27
2.1.5 Applied electric fields . . . . . 36
2.1.6 Analysis . . . . . 36
2.2 Cryo-electron microscopy . . . . . 40
2.2.1 The electron microscope . . . . . 40
2.2.2 Grid preparation . . . . . 43
2.2.3 Data collection . . . . . 44
2.2.4 Data processing . . . . . 45
2.2.5 Model building . . . . . 48
2.2.6 Cryo-electron tomography . . . . . 49
2.3 Bioinformatics . . . . . 52
2.3.1 Sequence alignment . . . . . 52
2.3.2 Structure prediction . . . . . 53
2.3.3 Databases . . . . . 53

**3 Translocation of double-stranded RNA through a nanopore 55**
3.1 Introduction . . . 56
3.1.1 Nanopores . . . 56
3.1.2 FhuA . . . 61
3.1.3 Double-stranded RNA . . . 62
3.1.4 Aim of this chapter . . . 68
3.2 Materials and Methods . . . 70
3.2.1 FhuA model . . . 70
3.2.2 Calculating ion flux through non-occluded FhuA . . . 72
3.2.3 Double-stranded RNA model . . . 72
3.2.4 Simulation set-up . . . 73
3.2.5 Analysis . . . 76
3.3 Results . . . 78
3.3.1 FhuA has an appropriate pore for dsRNA translocation . . . 78
3.3.2 Conformational behaviour of dsRNA in MD simulations . . 80
3.3.3 Initial simulations of WT FhuA . . . 84
3.3.4 Ion conductance through WT FhuA is affected by mutations 87
3.3.5 dsRNA does not always escape the pore in simulations . . . 91
3.3.6 Interrogating the role of dsRNA structure in translocation . 92
3.3.7 Characterisation of protein-RNA electrostatic interactions . 96
3.3.8 Mutations impact protein-RNA electrostatic interactions . . 99
3.3.9 Mutations impact the speed of dsRNA translocation . . . . 104
3.3.10 Translocation in the opposite direction . . . 108
3.3.11 Towards nanopore sequencing . . . 110
3.4 Discussion . . . 113
3.4.1 Choice of force field . . . 113
3.4.2 Effect of using C-$\alpha$ restraints in simulations . . . 113
3.4.3 Comparison of results to previous experimental data . . . . 114
3.4.4 Alternative engineered FhuA nanopores . . . 114
3.4.5 Mechanism of effect of mutations on dsRNA translocation through FhuA . . . 115
3.5 Conclusions . . . 118
3.6 Future directions . . . 120

**4 Structural studies of the ribosome of *Nitrosopumilus maritimus* 121**
4.1 Introduction . . . 122
4.1.1 Ribosome structure and function . . . 123
4.1.2 History of structural studies of the ribosome . . . 126
4.1.3 Variation of ribosomes across the domains of life . . . 128

4.1.4 Ribosomes in *N. maritimus* cells . . . 131
4.1.5 Cryo-electron microscopy and model building . . . 132
4.1.6 Aim of this chapter . . . 134
4.2 Materials and Methods . . . 135
4.2.1 Production of *N. maritimus* ribosomes . . . 135
4.2.2 Cryo-EM . . . 137
4.2.3 Model building . . . 140
4.2.4 Other analyses . . . 141
4.3 Results . . . 142
4.3.1 50S map resolution and features . . . 142
4.3.2 70S map resolution and features . . . 142
4.3.3 Universal proteins in the LSU . . . 145
4.3.4 Archaea-eukarya proteins in the LSU . . . 159
4.3.5 rRNA model building in the LSU . . . 168
4.3.6 Unmodelled density in the LSU . . . 171
4.4 Discussion . . . 174
4.4.1 Ribosomal protein content . . . 174
4.4.2 Structural features . . . 176
4.4.3 Auto-building . . . 181
4.5 Conclusions . . . 182
4.6 Future work . . . 184

**5 Structural studies of archaeal S-layers 186**
5.1 Introduction . . . 187
5.1.1 S-layers . . . 187
5.1.2 Archaea . . . 190
5.1.3 Previous work referenced in this chapter . . . 194
5.1.4 Aims of this chapter . . . 196
5.2 Materials and Methods . . . 198
5.2.1 *N. maritimus* . . . 198
5.2.2 *S. solfataricus* . . . 200
5.2.3 *P. oguniense* and *P. arsenaticum* . . . 201
5.2.4 *H. volcanii* . . . 202
5.2.5 Other analysis . . . 204
5.3 Results . . . 205
5.3.1 Structural studies of the S-layer of *N. maritimus* . . . 205
5.3.2 Structural studies of the S-layer of *S. solfataricus* . . . 210
5.3.3 Structural studies of the S-layers of *P. oguniense* and *P. arsenaticum* . . . 217

5.3.4 MD simulations of the *H. volcanii* S-layer . . . . . 227
5.4 Discussion . . . . . 240
5.4.1 Structural comparisons . . . . . 240
5.4.2 Sequence comparisons . . . . . 244
5.4.3 Using MD to predict identity of ions in cryo-EM density . . 247
5.5 Conclusions . . . . . 249
5.6 Future work . . . . . 252

**6 Discussion and Outlook** **254**
6.1 Molecular dynamics . . . . . 254
6.2 Cryo-electron microscopy . . . . . 256
6.3 Bioinformatics tools . . . . . 257
6.4 Integrative structural biology . . . . . 259
6.5 Outlook . . . . . 259

**Appendices**

**A Translocation of double-stranded RNA through a nanopore** **262**
A.1 Using convex hulls to calculate cross-sectional area . . . . . 263
A.2 dsRNA at different concentrations of KCl . . . . . 265
A.3 dsRNA hydrogen bonding between base pairs in simulation of solution 266
A.4 Simulation of other fragments of ribosomal RNA . . . . . 267
A.5 Principal component analysis of dsRNA in water . . . . . 268
A.6 Membrane curvature . . . . . 269
A.7 Simulations of translocation without C-$\alpha$ restraints . . . . . 270
A.8 Ion conductance through FhuA . . . . . 271
A.9 Translocation at different concentrations of KCl . . . . . 273
A.10 dsRNA H-bonding between base pairs in simulation of translocation 274
A.11 Coarse-grained simulations . . . . . 275
A.12 Principal component analysis of dsRNA in FhuA . . . . . 277
A.13 dsRNA interactions with the membrane . . . . . 277
A.14 Ion "helter skelter" . . . . . 279
A.15 dsRNA velocity in the pore . . . . . 280

**B The ribosome of *Nitrosopumilus maritimus*** **281**
B.1 Ribosome particle picking . . . . . 282
B.2 Ribosome map and model data . . . . . 283
B.3 50S map FSC . . . . . 283
B.4 50S model-in-map FSC . . . . . 284
B.5 70S map FSC . . . . . 285

B.6 70S model-in-map FSC . . . . . 286
B.7 LSU Model validation . . . . . 287
B.8 Model validation of the SSU . . . . . 288

**C Structural studies of archaeal S-layers** **289**
C.1 Low density of *N. maritimus* cells on grids . . . . . 290
C.2 *Saccharolobus* SlaA sequence alignment . . . . . 290
C.3 *Pyrobaculum* S-layer sequence alignment . . . . . 290
C.4 Selective *Pyrobaculum* S-layer sequence alignment . . . . . 295
C.5 RMSD between *Pyrobaculum* S-layer domains . . . . . 296
C.6 TMHMM for *Pyrobaculum* S-layer sequences . . . . . 297
C.7 *Pyrobaculum* S-layers structure comparison . . . . . 298
C.8 *H. volcanii* potential cation binding sites . . . . . 299
C.9 *H. volcanii* VolMap density calculations . . . . . 302
C.10 Percentage *H. volcanii* binding site occupancies . . . . . 304
C.11 *H. volcanii* modelled binding site occupancies . . . . . 306

**Bibliography** **307**

# List of Figures


2.1 The bonded terms of force fields. . . . . 18
2.2 Plot of bond length against potential energy. . . . . 19
2.3 Plot of bond angle against potential energy. . . . . 20
2.4 Plot of plane angle against potential energy. . . . . 21
2.5 Plot of dihedral angle against potential energy. . . . . 22
2.6 Plot of the Lennard-Jones potential. . . . . 24
2.7 Schematic of energy minimisation. . . . . 26
2.8 A typical workflow for MD simulations. . . . . 29
2.9 The leapfrog integrator. . . . . 30
2.10 Visualisation of the periodic boundary condition. . . . . 32
2.11 Schematic of an electron microscope. . . . . 41

3.1 Example of a protein nanopore structure. . . . . 57
3.2 How nanopores work for DNA sequencing. . . . . 58
3.3 The crystal structure of FhuA. . . . . 62
3.4 Modifications to turn FhuA into a nanopore. . . . . 63
3.5 Non-canonical base pairs and common RNA folds *in vivo*. . . . . 65
3.6 Structure of large ribosomal subunit from *N. maritimus*. . . . . 67
3.7 Modified FhuA structure. . . . . 71
3.8 Sequence and structure of dsRNA used in simulations. . . . . 74
3.9 Example of dsRNA "threading". . . . . 75
3.10 Parameters for determining the presence of pi-stacking. . . . . 77
3.11 Calculations of $xy$ cross-sectional area of WT FhuA. . . . . 81
3.12 Graphs of conformational behaviour of dsRNA in simulations. . . . 83
3.13 Graphs of conformational behaviour of dsRNA in extended simulations. 85
3.14 PCA projection of dsRNA backbone structures in simulation. . . . . 86
3.15 RMSD of FhuA in simulation. . . . . 87
3.16 Speed of ion movement through FhuA under an applied electric field. 88
3.17 Instantaneous current, and current against applied electric field. . . 91
3.18 Non-productive dsRNA translocation. . . . . 93
3.19 Graphs of dsRNA structure during translocation. . . . . 95
3.20 PCA of dsRNA translocating through FhuA. . . . . 97
3.21 The electrostatic surface of WT FhuA. . . . . 98
3.22 Interactions of lysine and arginine residues with dsRNA. . . . . 100
3.23 Interactions of aspartic acid and glutamic acid residues with dsRNA. 101
3.24 Protein and dsRNA contacts in DA/EA pore. . . . . 103
3.25 Protein and dsRNA contacts in DA/EA/KA/RA pore. . . . . 105
3.26 Protein and dsRNA contacts in WT pore. . . . . 106
3.27 Positions of important dsRNA contact residues in FhuA. . . . . 107
3.28 dsRNA and coordinated potassium ions. . . . . 107

3.29 Initial threaded positions of dsRNA in simulations. . . . . 110
3.30 Graph of current against voltage for all simulations. . . . . 112
3.31 Proposed mechanism for how mutations affect dsRNA translocation. 116

4.1 Overview of prokaryotic ribosome composition. . . . . 124
4.2 Overview of ribosome function. . . . . 125
4.3 The division of conserved ribosomal proteins across domains of life. 130
4.4 Ammonia oxidation in ammonia-oxidising archaea. . . . . 131
4.5 Tomogram of a *N. maritimus* cell. . . . . 132
4.6 The RELION processing pipeline. . . . . 139
4.7 The 50S map. . . . . 143
4.8 The 70S map. . . . . 144
4.9 The model of the ribosome in the 70S map. . . . . 145
4.10 Ribosomal protein uL1. . . . . 147
4.11 Ribosomal protein uL2. . . . . 148
4.12 Ribosomal protein uL3. . . . . 149
4.13 Ribosomal protein uL4. . . . . 150
4.14 Ribosomal proteins uL5, uL18, and uL30. . . . . 151
4.15 The 5S ribonucleoprotein. . . . . 152
4.16 Ribosomal protein uL6. . . . . 152
4.17 A putative model of the ribosomal stalk. . . . . 153
4.18 Interaction between uL13 and 23S rRNA. . . . . 154
4.19 Ribosomal protein uL14. . . . . 155
4.20 Ribosomal protein uL15. . . . . 156
4.21 Ribosomal protein uL16. . . . . 157
4.22 Ribosomal proteins uL22, uL23, uL24, and uL29. . . . . 158
4.23 Proteins forming the polypeptide exit tunnel. . . . . 159
4.24 Placement of uL33 in the ribosome. . . . . 160
4.25 Ribosomal protein aL8. . . . . 162
4.26 Ribosomal proteins aL15, aL18, aL19, and aL21. . . . . 163
4.27 Ribosomal proteins aL24, aL37, aL40, and aL43. . . . . 165
4.28 Zinc coordination by ribosomal proteins. . . . . 166
4.29 Ribosomal protein aL30. . . . . 167
4.30 Secondary and tertiary structures of aL8 and aL30. . . . . 168
4.31 Ribosomal protein aL31. . . . . 169
4.32 Ribosomal protein aL32. . . . . 169
4.33 Ribosomal protein aL39. . . . . 170
4.34 Ribosomal RNA 5S. . . . . 170
4.35 Ribosomal RNA 23S. . . . . 171
4.36 Models in protein density. . . . . 172

4.37 Unmodelled RNA density. . . . . 173
4.38 aL proteins mapped on a phylogenetic tree. . . . . 177
4.39 50S model includes poly-cytosine. . . . . 178
4.40 Map of local resolution of the ribosome structure. . . . . 180

5.1 Overview of archaeal S-layer composition. . . . . 188
5.2 Phylogenetic tree of archaeal species studied. . . . . 191
5.3 *S. solfataricus* S-layer structure at low resolution. . . . . 193
5.4 The S-layer structure of *S.acidocaldarius*. . . . . 195
5.5 The S-layer structure of *H. volcanii*. . . . . 197
5.6 Tomogram slices of *N. maritimus* cells. . . . . 206
5.7 5.5 Å structure of *N. maritimus* S-layer. . . . . 207
5.8 Published structure of *N. maritimus* S-layer. . . . . 208
5.9 Model fitted to tomogram structure of *N. maritimus* S-layer. . . . . 209
5.10 Micrograph of an *S. solfataricus* cell. . . . . 210
5.11 Tomograms of vesicles from *S. solfataricus*. . . . . 211
5.12 Structure of the *S. solfataricus* S-layer. . . . . 213
5.13 Comparison of structures of the *S. solfataricus* S-layer. . . . . 214
5.14 Schematic of the *S. solfataricus* S-layer. . . . . 214
5.15 The *S. acidocaldarius* S-layer structure. . . . . 215
5.16 Comparison of the *S. solfataricus* and *S. acidocaldarius* sequences. . 216
5.17 Comparison of the *S. solfataricus* and *S. acidocaldarius* structures. . 218
5.18 Tomogram of a *P. oguniense* cell. . . . . 219
5.19 Structure of the *P. oguniense* S-layer. . . . . 220
5.20 Tomogram of a *P. arsenaticum* cell. . . . . 221
5.21 Structure of the *P. arsenaticum* S-layer. . . . . 223
5.22 AF2 models of *Pyrobaculum* S-layer monomers. . . . . 225
5.23 Simulation setup for *H. volcanii* S-layer. . . . . 227
5.24 Structural deformation of S-layer without divalent cations. . . . . 228
5.25 RMSF projected onto the S-layer structure. . . . . 229
5.26 Structural deformation of S-layer with only published divalent cations. 230
5.27 Electrostatic surface of the *H. volcanii* S-layer. . . . . 231
5.28 Ion density around the S-layer structure. . . . . 232
5.29 Divalent cation binding sites in the *H. volcanii* S-layer. . . . . 234
5.30 High occupancy binding sites. . . . . 235
5.31 Medium occupancy binding sites. . . . . 236
5.32 Low occupancy binding sites. . . . . 237
5.33 Occupancy of published binding sites. . . . . 238
5.34 Building ions into the electron density. . . . . 239
5.35 Rotational symmetry in p6 2D lattices. . . . . 242

5.36 *S. acidocaldarius* model coloured by homology to *S. solfataricus*. . . 246

A.1 Using convex hulls to calculate cross-sectional areas. . . . . 264
A.2 RMSD, SASA, and number of ions for each KCl concentration . . . 265
A.3 Percentage of frames where RNA is paired . . . . . 266
A.4 Simulations of RNA fragments in water . . . . . 267
A.5 Cumulative variance for PCA of dsRNA in water . . . . . 268
A.6 Membrane curvature induced by applied electric field . . . . . 269
A.7 Translocation of dsRNA through unrestrained FhuA . . . . . 270
A.8 Ion conductance and crossing times . . . . . 272
A.9 Translocation of dsRNA at varied KCl concentrations . . . . . 273
A.10 Percentage of frames where RNA in FhuA is paired . . . . . 274
A.11 Coarse-grained simulations of dsRNA translocation . . . . . 276
A.12 Cumulative variance for PCA of dsRNA in FhuA . . . . . 277
A.13 Interaction of dsRNA with the membrane . . . . . 278
A.14 Potassium ions helter skelter along dsRNA backbone . . . . . 279
A.15 Speed of dsRNA translocation in FhuA sections . . . . . 280

B.1 Picking ribosomes in RELION. . . . . 282
B.2 50S map FSC. . . . . 283
B.3 50S model-in-map FSC. . . . . 284
B.4 70S map FSC. . . . . 285
B.5 70S model-in-map FSC. . . . . 286

C.1 Arctica screening of *N. maritimus* cells. . . . . 290
C.2 BLAST alignment of SlaA sequences. . . . . 291
C.3 COBALT alignment of SlaA sequences. . . . . 292
C.4 BLAST alignment of S-layer sequences. . . . . 293
C.5 COBALT MSA alignment of *Pyrobaculum* S-layer sequences. . . . . 294
C.6 COBALT MSA alignment of selected *Pyrobaculum* S-layer sequences. 295
C.7 TMHMM analysis of *Pyrobaculum* S-layer sequences. . . . . 297
C.8 *Pyrobaculum* S-layers structure comparison. . . . . 298
C.9 Chain identification in the *H. volcanii* S-layer. . . . . 299
C.10 *H. volcanii* S-layer potential cation binding sites. . . . . 301
C.11 VolMap densities of ions. . . . . 303
C.12 Graph of *H. volcanii* S-layer binding site occupancies. . . . . 305
C.13 *H. volcanii* S-layer binding site occupancies. . . . . 306

# List of Tables


3.1 Modifications made to wild-type FhuA. . . . . 70
3.2 Mutations applied to each simulation system. . . . . 71
3.3 Simulation set-ups for simulations of FhuA without C-alpha restraints. 73
3.4 Simulation details for FhuA with dsRNA. . . . . 76
3.5 Time taken for dsRNA translocation through FhuA. . . . . 92
3.6 Time taken for dsRNA translocation through FhuA in reverse direction. 109

4.1 Building of universal r-proteins. . . . . 146
4.2 Proteins that are common to archaeal and eukaryotic ribosomes. . . 161
4.3 R-protein content overview. . . . . 175

5.1 *H. volcanii* S-layer simulation set-ups. . . . . 203
5.2 BLAST alignment of *Pyrobaculum* S-layers. . . . . 224
5.3 Assignments of ion identity. . . . . 233

A.1 Simulation set-ups for simulations of FhuA without dsRNA present 270
A.2 Time taken for dsRNA translocation through unrestrained FhuA . . 273

B.1 Model validation in PHENIX. . . . . 287
B.2 Model validation in PHENIX of the SSU. . . . . 288

C.1 RMSD calculated by domain, against H6Q8S4. . . . . 296
C.2 RMSD calculated by domain, against H6Q8U3. . . . . 296
C.3 RMSD calculated by domain, against A0A7L4P6S3. . . . . 296

# List of Abbreviations

**AF** . . . . . AlphaFold

**APBS** . . . . . Adaptive Poisson-Boltzmann Solver

**AOA** . . . . . ammonia-oxidising archaea

**BLAST** . . . . Basic Local Alignment Search Tool

**$CaCl_2$** . . . . . calcium chloride

**Cl** . . . . . chloride

**$CoCl_2$** . . . . . cobalt chloride

**cryo-EM** . . . cryo-electron microscopy

**cryo-ET** . . . . cryo-electron tomography

**CsgG** . . . . . curlin sigma S-dependent growth subunit G

**CTF** . . . . . contrast transfer function

**$CuCl_2$** . . . . . copper chloride

**DNA** . . . . . deoxyribonucleic acid

**ssDNA** single-stranded DNA

**DQE** . . . . . detective quantum efficiency

**$e^-$** . . . . . electron

**EDTA** . . . . . ethylenediaminetetraacetic acid

**EMDB** . . . . Electron Microscopy Data Bank

**FeNaEDTA** . . ferric sodium ethylenediaminetetraacetic acid

**FhuA** . . . . . ferric hydroxamate uptake component A

**FraC** . . . . . fragaceatoxin C

**FSC** . . . . . Fourier shell correlation

**HAMAP** . . . High-quality Automated and Manual Annotation of Proteins

**$H_3BO_3$** . . . . perboric acid

**HCl** . . . . . hydrochloric acid

**HEPES** . . . . 4-(2-hydroxyethyl)-1-piperasineethanesulfonic acid

**IDR** . . . . . intrinsically disordered region

**K** . . . . . potassium

**KBr** . . . . . potassium bromide

**KCl** . . . . . potassium chloride

**$KH_2PO_4$** . . . potassium hydrogen phosphate

**KOAc** . . . . . potassium acetate

**LJ** . . . . . Lennard-Jones

**LSU** . . . . . large subunit

**LUCA** . . . . . last universal common ancestor

**MD** . . . . . molecular dynamics

**MgCl** . . . . . magnesium chloride

**$MgSO_4$** . . . . magnesium sulfate

**$MnCl_2$** . . . . . manganese chloride

**MSA** . . . . . multiple sequence alignment

**MspA** . . . . . *Mycobacterium smegmatis* porin A

**NaCl** . . . . . sodium chloride

**$NaHCO_3$** . . . sodium hydrogen carbonate

**$Na_2MoO_4$** . . . sodium molybdate

**NCBI** . . . . . National Center for Biotechnology Information

**$NH_4Cl$** . . . . . ammonium chloride

**$NiCl_2$** . . . . . nickel chloride

**NPT** . . . . . number of atoms, pressure, temperature

**NVT** . . . . . number of atoms, volume, temperature

**OmpA** . . . . . outer membrane protein A

**PAE** . . . . . Predicted Alignment Error

**PBC** . . . . . periodic boundary condition

**PCA** . . . . . principal component analysis

**PDB** . . . . . Protein Data Bank

**pLDDT** . . . . predicted Local Distance Difference Test

**PME** . . . . . Particle Mesh Ewald

**POPC** . . . . . palmitoyl-oleoyl-phosphatidylcholine

**PTC** . . . . . peptidyl transferase centre

**RCSB** . . . . . Research Collaboratory for Structural Bioinformatics

**RMSD** . . . . . root mean square deviation

**RMSF** . . . . . root mean square fluctuation

**RNA** . . . . . ribonucleic acid

**dsRNA** double-stranded RNA

**mRNA** messenger RNA

**rRNA** ribosomal RNA

**tRNA** transfer RNA

**RNP** . . . . . ribonucleoprotein

**SASA** . . . . . surface-accessible surface area

**SCM** . . . . . synthetic Crenarchaeota medium

**SPA** . . . . . single-particle analysis

**SSU** . . . . . small subunit

**STA** . . . . . sub-tomogram averaging

**TMHMM** . . . transmembrane hidden Markov model

**VdW** . . . . . van der Waals

**v/v** . . . . . volume per volume

**v/w** . . . . . volume per weight

**WC** . . . . . Watson-Crick

**w/v** . . . . . weight per volume

**w/w** . . . . . weight per weight

**Zn** . . . . . zinc

**$ZnSO_4$** . . . . . zinc sulfate

# List of Units

**Å** . . . . . Ångstrom, $10^{-10}$ m

**bp** . . . . . base pair

**°C** . . . . . degree Celcius

**C** . . . . . Coulomb

**Da** . . . . . Dalton

**eV** . . . . . electron volt

***g*** . . . . . relative centrifugal force

**g** . . . . . gram

**Hz** . . . . . Hertz

**J** . . . . . joules

**K** . . . . . Kelvin

**L** . . . . . litre

**m** . . . . . metre

**M** . . . . . molar concentration

**min** . . . . . minute

**mol** . . . . . mole

**pH** . . . . . $-\log_{10}([H^+])$ in mol $L^{-1}$

**s** . . . . . second

***S*** . . . . . siemens

**S** . . . . . Svedberg

**V** . . . . . volt

**Prefixes** . . . .

**M** mega, $10^6$

**k** kilo, $10^3$

**m** milli, $10^{-3}$

**µ** micro, $10^{-6}$

**n** nano, $10^{-9}$

**p** pico, $10^{-12}$

**f** femto, $10^{-15}$

# 1 Introduction

## 1.1 Integrative structural biology

Integrative structural biology is an approach to solving the structures of biological molecules that incorporates data from a number of orthologous methods (Ward *et al.*, 2013). This approach acknowledges and compensates for the strengths and weaknesses of different methods, and aims to bridge the gap between structure and function in a reciprocal manner: data on the structure can inform the function, and data on the function can improve the structure (Rout and Sali, 2019; Srivastava *et al.*, 2020; Ziegler *et al.*, 2021; Schwalbe *et al.*, 2024). The emphasis on integrative structural biology has led to an increase in data availability efforts from researchers across the field, who aim to make it easier to incorporate existing results into new research (Ramalli *et al.*, 2022; Roy *et al.*, 2025; Vallat *et al.*, 2025). This builds on the pioneering work of the Protein Data Bank (PDB), which was initially launched in 1971 as the first open-access digital resource in the biological fields and allowed users to deposit and access 3D structures of macromolecules (Nature Structural Biology, 1998).

An early example of an integrative approach to structural biology is Molecular Dynamics Flexible Fitting (MDFF), which uses molecular dynamics (MD, see Section 1.3) to improve the fit of models into experimental structures (Trabuco *et al.*, 2008). The experimental structure is converted into an energy potential, with an energetic penalty for the model moving away from the structure density. This is combined with other bonded and non-bonded terms in the MD force field to ensure the model has a realistic conformation. MDFF is commonly used to fit models into lower-resolution structures, particularly where the target molecule is

in a different conformation in the structure compared to the model being fitted (Trabuco *et al.*, 2009, 2011).

Most commonly, high-resolution structures used in integrative structural biology come from X-ray crystallography and, increasingly, cryo-electron microscopy (cryo-EM, see Section 1.2) (Koukos and Bonvin, 2020). These static structures are then combined with other data to inform the function, dynamics, and biological context of the molecules. Cryo-electron tomography (cryo-ET) is often used to observe the molecules *in situ*, with sub-tomogram averaging (STA) allowing the solving of structures in a more physiologically relevant environment than a solvent developed to optimise a purified sample of the molecule for cryo-EM studies (Caspy *et al.*, 2025). These STA structures can then be fitted with models from high-resolution structures, using methods such as MDFF to improve the fitting, and then the resulting models can be compared to elucidate any conformational differences that may highlight a cellular function (Nogales and Mahamid, 2024). As cryo-EM structures can contain information about structural heterogeneity of the sample, tools such as MDFF can allow the exploration of these conformational states (McGreevy *et al.*, 2016). Such states may not be apparent in X-ray crystallography structures due to the impact of crystal packing (Luo *et al.*, 2015). Correlative light and electron microscopy (CLEM) is also used to study molecules in cells, with a light microscope being used to locate a fluorescence-tagged molecule that is then imaged at higher resolution using an electron microscope (de Boer *et al.*, 2015).

Bioinformatics data can be used in a variety of ways to inform structural biology (Zheng *et al.*, 2017a). For example, sequence alignment can be used to identify homologous proteins to aid model building, or for comparison where sequences diverge at proposed functional sites (Gutmanas *et al.*, 2013). They can also be used to identify potential functions of a protein, inferred by features such as homology or by protein fold. There are a large number of databases of sequences, functions, and structures that can be utilised to support structural studies (Chen *et al.*, 2017).

There has been a significant increase in the use of protein and nucleic acid structure prediction methods in structural biology. While previous servers existed to predict secondary structure or common folds based on homology, these could only generate complete models of molecules when there was significant homology to solved structures (Sen *et al.*, 2006; Bertoline *et al.*, 2023). The development of AlphaFold and RoseTTAFold changed this, first focusing on single proteins using deep learning methods trained on the published structures in the PDB (Jumper *et al.*, 2021; Baek *et al.*, 2021). Both programs were then extended to include the ability to predict the structure of complexes including nucleic acids and other cofactors (Abramson *et al.*, 2024; Baek *et al.*, 2024). AlphaFold and RoseTTAFold can be used to provide models of complexes that do not yet have experimental structures, meaning these complexes can be studied using integrative structural methods before an experimental structure is solved. They also make the building of models into structures much faster, allowing researchers to spend more time analysing and understanding the structure rather than just building a model. The 2024 Nobel Prize in Chemistry was awarded for the development of these successful protein structure prediction methods (Hassabis, 2024).

Molecular simulations and nuclear magnetic resonance (NMR) are the common methods for obtaining dynamics of structures (Grandori, 2023). MD simulations allow the study of a structure over time, and experiments can be conducted *in silico* to observe interactions with other molecules or the effect of different environments (Vant *et al.*, 2022). There has been a recent move to make simulation trajectories available to researchers in a similar manner to the PDB and Electron Microscopy Data Bank (EMDB) structural databases, making results accessible for wider study (Caparotta and Perez, 2024). The field of cryo-EM has particularly benefitted from the utilisation of MD simulations, as the electron density can be used as a potential in the simulations as described above for MDFF (Bock *et al.*, 2024). Heterogeneity in cryo-EM structures can also be studied and modelled using MD, allowing the dynamics of the structure to be elucidated and visualised within the electron density (Toader *et al.*, 2023; Vuillemot *et al.*, 2023).

ISOLDE is an extension available for ChimeraX, and utilises short, contained MD simulations to relax a model into a cryo-EM structure and improve the quality of the model (Croll, 2018; Goddard *et al.*, 2018).

NMR is an experimental technique that has historically been used for structure determination, but the nature of the technique has made it more popular for understanding dynamics of proteins, particularly intrinsically disordered proteins and those forming heterogeneous aggregates that cannot be studied by averaging techniques (van der Wel, 2018; Zadorozhnyi *et al.*, 2024).

Integrative structural biology has been important in the improvement of molecular dynamics force fields, in particular those describing ribonucleic acid (RNA), as the increased availability of structural data on a range of RNA folds has allowed the development of more suitable parameters describing inter- and intra-molecular interactions (Li *et al.*, 2020; Liebl and Zacharias, 2023). This gives more data for computational biologists to use as benchmarks, which leads to more accurate force fields. The increase in the number of solved structures containing RNA (and DNA) has also led to better predictions of nucleic acid secondary and tertiary structure based on sequence. However, the accuracy of these predictions still lags behind those of protein structures (Budnik *et al.*, 2024). More structures will be required to train new, better predictive models.

## 1.2 Cryo-electron microscopy for solving biological structures

The first electron microscope was invented by Ruska (1934a,b), using magnets as objective lenses for the electron waves, and was quickly adopted across a range of scientific disciplines including biology (Ruska, 1934a,b; Marton, 1934; Krause, 1937). Electrons scattered by the sample in the microscope were captured on photographic film, and this produced the resulting image (Whiddington and Taylor, 1932). The development of the electron microscope was awarded the Nobel Prize in Physics in 1986 (Ruska, 1987).

The limitations of electron microscopy for biological samples were originally described by Marton (1934), with the consideration of high energy electrons destroying the samples. These limitations were then measured quantitatively by Glaeser (1971), who studied the damage rate at a range of electron doses. As described by Marton (1934), one method for reducing the rate of damage is to reduce the temperature of the sample, and the first cryogenic electron microscopy (cryo-EM) was performed in liquid ethane by Dubochet *et al.* (1988). However, the resolution obtainable by cryo-EM was still limited, leading the technique to be commonly referred to as "blobology" (Saibil, 2017). It was only possible to observe large molecules using the method, and atomic models (solved using X-ray crystallography) could only be roughly fitted to the resulting density maps (Rossmann, 2000). Samples needed to be thin, or sectioned using a diamond knife, to ensure that sufficient electrons were scattered and were captured on the film plate (Al-Amoudi *et al.*, 2004).

Henderson (1995) proposed that cryo-EM could yield resolutions high enough for atomic model building, foresight that would contribute to him being awarded the Nobel Prize in Chemistry in 2017 along with Dubochet and Frank (Henderson, 2017).

### 1.2.1 Resolution revolution

As technology, both hardware and software, improved for cryo-EM, the field underwent what was termed the "resolution revolution" by Kühlbrandt (2014). This was the combination of advancements that allowed cryo-EM to reach the predicted atomic resolution, via the process of single particle averaging (SPA). SPA was described by Frank (2006), and showed the possibility of high resolution structure determination of biological molecules. The sample needed to have random orientations to produce a sufficient distribution of 2D projections of the particle to cover all of the unique 3D space, accounting for molecule symmetry (Frank, 2002; Naydenova and Russo, 2017), and also needed to be large enough for the signal of scattered electrons to be observed above the noise (Russo *et al.*, 2022). Software improvements, such as enhanced SPA algorithms (Scheres, 2012), and hardware

improvements, such as the invention of direct electron detectors (McMullan *et al.*, 2009), ultimately led to the first atomic (< 1.5 Å) structures solved using cryo-EM (Nakane *et al.*, 2020; Yip *et al.*, 2020). Other developments such as focused-ion-beam milling have enabled the study of previously intractable samples (Villa *et al.*, 2013).

### 1.2.2 The rise of cryo-electron tomography

Cryo-electron tomography (cryo-ET) is the process of making 3D constructions from multiple images of the same sample, at different tilt angles (De Rosier and Klug, 1968; Hart, 1968). As it can be used to observe biological molecules *in situ*, within cells or organelles, cryo-ET has been a modestly popular technique for molecular biology throughout the past decades (Koster *et al.*, 1997; Baumeister *et al.*, 1999). However, recently the technique has become increasingly important in the arsenal of structural biologists since the work of Briggs (2013) showed the possibility of high-resolution structure determination using sub-tomogram averaging (STA). The potential of STA was first shown using stained samples of the 30S subunit of the *Escherichia coli* ribosome (Knauer *et al.*, 1983). The technological advancements that enabled the cryo-EM resolution revolution are now allowing a similar growth in cryo-ET (Hong *et al.*, 2023).

## 1.3 Molecular dynamics for investigating biological structures

Molecular dynamics (MD) simulations were first developed by Alder and Wainwright (1959), who modelled particles as hard spheres to study the dynamic properties of solids and liquids at phase transition points. This approach was first applied to proteins by McCammon *et al.* (1977) who studied the bovine pancreatic trypsin inhibitor over the course of an 8.8 ps simulation. Studies of nucleic acids followed with the DNA double helix and a tRNA molecule (Levitt, 1984; Harvey *et al.*, 1985). Simulations of biological systems were initially limited to globular molecules, but were already being utilised to refine X-ray crystallography structures and consider the dynamics of proteins (Karplus and Petsko, 1990). The first simulations of a

phospholipid bilayer were performed by Egberts *et al.* (1994), who achieved a liquid-crystalline phase with properties matching experimental values over the course of 80 ps simulations. The Nobel Prize in Chemistry was awarded to Levitt, Karplus, and Warshel in 2013 for their work in the computational multi-scale modelling of chemical systems (Levitt, 2013; Smith and Roux, 2013).

The field of MD simulations is constantly being advanced, with developments in methodology, software, improved use of new computer hardware, and analysis of data (Vlachakis *et al.*, 2014; Caparotta and Perez, 2024; Chipot, 2024; Cui *et al.*, 2025; Khrenova, 2025). This means that modern simulations of biological systems are vastly different to those on the scale of picoseconds that studied only one protein in a water box. MD is an increasingly utilised tool in structural biology studies, with many structure-based publications including molecular simulations to give insight into the dynamics of the solved structure such as the work of von Kügelgen *et al.* (2024) (Dror *et al.*, 2012; Hollingsworth and Dror, 2018). Additionally, MD simulations can give further mechanistic insights to previously published models (e.g. the work of Batebi *et al.* (2024) on a G-protein-coupled receptor), improving the understanding of the structure-function relationship. The scope of MD simulations has increased, with increases in computing power and the ability to parallelise simulation calculations leading to much longer timescales being accessible (Shaw *et al.*, 2008; Jung *et al.*, 2021). Also, much larger biological systems can now be studied computationally, with more atoms being simulated, as described by Bock *et al.* (2018) concerning simulations of the ribosome (Perilla *et al.*, 2015). Overall, this has meant that the complexity of systems that can be studied and are being studied has increased, with a corresponding improvement in the applicability of the data generated (Khalid *et al.*, 2023).

## 1.4 Biomolecular Folding

The interactions between biological macromolecules and their surroundings are key to the establishment and continuation of biological processes. They allow the transfer of information between macromolecules, the conversion of substrate into

product, and the correct assembly of cellular structures. Our understanding of how these processes work is therefore incomplete without an appreciation of the complex interaction networks between macromolecules. This is becoming more apparent as the field of biochemistry advances, and researchers are looking for more ways to investigate the interactions using methods that can probe these processes and the molecules involved in them.

Anfinsen's Nobel prize-winning work, based on Steiner's studies of Ribonuclease A, established the dogma that the native structure of a protein is determined by its primary sequence (Steiner, 1959; Anfinsen, 1973). As our understanding of proteins and protein complexes has evolved, we have grown to appreciate the role of the environment of the protein, as proteins do not fold in a vacuum. Cofactors, solvent, and the presence of other macromolecules can play a key role in determining the native structure of a protein (Dill, 1990; Frauenfelder and Wolynes, 1994; Brugè *et al.*, 1996; Martorana *et al.*, 1996, 1998; Dobson and Karplus, 1999). When interrogating protein structure, the complexity of the environment must be considered. However, our knowledge of these interacting partners is much more limited when compared to our knowledge of protein structures, particularly soluble protein structures. Thus, to fully understand the structure of a biological molecule, the role of interactions with other macromolecules must be considered. This allows a better understanding of the function of the molecule, as the structure-function paradigm links the function of a protein to its 3D structure (Redfern *et al.*, 2008; Fersht, 2008).

The folds of nucleic acid chains are also important for their function (Assmann *et al.*, 2023; Kwon, 2025). While DNA is typically double-stranded and most commonly adopts a B-helix structure in the cell (Watson and Crick, 1953; Neidle, 2010), RNA is much more broad in its conformational landscape due to the propensity for single-stranded RNA (Brion and Westhof, 1997; Tinoco and Bustamante, 1999; Westhof and Fritsch, 2000; Alberts *et al.*, 2002). This means there are a number of similarities to the folding of proteins (Thirumalai and Woodson, 1996; Thirumalai *et al.*, 1997; Thirumalai and Hyeon, 2005). The folds adopted

by RNA molecules are heavily influenced by their surroundings, and the change in fold induced by a change in environment can have important physiological consequences (Zemora and Waldsich, 2010).

### 1.4.1 Importance of solvent

The solvent in which the biological molecule exists will affect the folding of the molecule (Yu *et al.*, 2016; Arakawa, 2017). This includes water and positive and negative ions (such as in the cytoplasm), or the hydrocarbon tails of lipid molecules (such as in the membrane). Lipid tails are highly hydrophobic, meaning that optimal interactions with proteins will be hydrophobic. Therefore, hydrophobic residues will be facing towards the lipid tails, and hydrophilic residues will be buried away from the lipid environment (Corin and Bowie, 2022). This is an exception to the general rule where the formation of a molten globule from the hydrophobic protein residues is key to the final protein structure, because in membrane proteins these residues are not all protected from the environment (Dill, 1985, 1990; Ptitsyn, 1995). Additionally, in proteins where only part of the structure is contained within the membrane, different domains require different environments for correct folding and protein function. The polarity of the solvent is important for the structure of the protein and must be accounted for, particularly when purifying proteins for structural studies (Finney *et al.*, 1980; Bucciarelli *et al.*, 2020). Proteins in a solvent that is significantly non-native will adopt non-physiological folds, or potentially aggregate.

Cations are important to the structure of many biological molecules. RNA is a highly negatively-charged molecule, as every nucleic acid residue has one negative charge on the phosphate group in the backbone. This means that counter-ions are required for the RNA to be able to adopt structures where the backbone of one section is brought in close proximity to the backbone of another section. Physiologically, this role is usually filled by magnesium ions, $Mg^{2+}$ (Brion and Westhof, 1997; Grover, 2015; Guth-Metzler *et al.*, 2023). Without cations in the solution to act as counter-ions, RNA is unable to adopt a number of the complex folds that may be required for the function of the molecule. Ions such as calcium

$Ca^{2+}$, zinc $Zn^{2+}$, potassium $K^{+}$, and $Mg^{2+}$ are often coordinated by residues in protein structures and required for the maintenance of the protein structure (Cates *et al.*, 2002; Piovesan *et al.*, 2012; Rozov *et al.*, 2019; von Kügelgen *et al.*, 2021; Zhong and Barrientos, 2024). Many coordination sites for ions in proteins are specific, and require particular species for the correct fold to be adopted around the site. For example, while holmium ions, $Ho^{3+}$, can often bind in $Ca^{2+}$ binding sites in proteins, this has structural and physiological consequences (Weis *et al.*, 1991; Pidcock and Moore, 2001; von Kügelgen *et al.*, 2021).

"Extremophiles" is a non-taxonomic grouping of microorganisms that live in extreme conditions (Macelroy, 1974). These may be extreme temperatures, pH, salinity, or availability of energy sources. Often, extremophiles are species of archaea (see Section 1.5), and the biology underpinning the ability of these species to thrive in extreme environments is poorly understood. The proteins in these cells will have different properties due to the solvent in which they exist. Solvents in high temperature will have high energy, and this may be sufficient energy to disrupt the non-covalent bonds such as hydrogen bonds in the protein structure (Cooper, 2000). This means alternative methods are required to maintain protein structure. High salinity increases the ionic strength of the solvent, which is often exploited in X-ray crystallography to get proteins to "salt out" and form crystals (Möller *et al.*, 2012; Schröder, 2017). However, this does not happen with the proteins in extremophiles, which means that the folds of the proteins are adapted to the higher ionic strength, and may misfold in solvents with lower ionic strength (von Kügelgen *et al.*, 2021). Non-neutral pH solvents often cause protein misfolding for proteins from non-extremophiles (O'Brien *et al.*, 2012). This is particularly notable in enzymes, where small changes in the active site can denature the enzyme, and these changes can be induced by an altered concentration of protons in the solvent (Kishore *et al.*, 2012). Again, the enzymes and other proteins of extremophiles adapted to acidic or basic environments are functional in their native environments, but would not be functional in other environments such as the cytoplasm of a human cell.

Thus, the solvent heavily influences the fold of a protein, and when considering the structure of a protein, it must be studied in the native environment.

### 1.4.2 Importance of binding partners

Some biological complexes are only correctly folded when in the presence of all the binding partners. As these binding partners are not encoded within the protein sequence, this is outside of Anfinsen's dogma (Rogers *et al.*, 2013). Other proteins require chaperone proteins for correct folding, usually because the environment in which they are translated is not the same as the environment in which they function (Hendrick and Hartl, 1995). The chaperone proteins may engage with the target protein as soon as it begins to be translated by recruitment to the ribosome, or may be required to first unfold misfolded protein targets to allow the correct refolding (Beissinger and Buchner, 1998; Hartl *et al.*, 2011; Deuerling *et al.*, 2019; Buchner, 2019). These functions may require the chaperone to generate energy from ATP hydrolysis (Goloubinoff *et al.*, 2018). The ribosome itself can play a role in the folding of proteins, by interactions between the nascent chain and the polypeptide exit tunnel (Micic *et al.*, 2022).

Nucleic acids also require binding partners for correct or optimal folding. The ribosome is 60% RNA, and the remaining 40% is proteins (Moore, 1997). These proteins associate with the ribosomal RNA in a specific order that is important for the correct assembly of the ribosome (Gamalinda and Woolford Jr, 2015; Madru *et al.*, 2015; Razi and Ortega, 2017; Sheng *et al.*, 2023). Both protein and RNA are key to the inter-subunit interactions forming the full ribosome. DNA supercoiling is also reliant on the correct protein partners (Duprey and Groisman, 2021). Many factors control the dynamic conformational changes undertaken by DNA, and are required for the correct expression of genes and replication of the genome (Dorman and Dorman, 2016).

Intrinsically disordered proteins (IDP) or intrinsically disordered regions (IDR) of proteins are extremely flexible. There are a number of IDPs and IDRs that are known to adopt secondary structure and defined folds upon binding to partner

proteins, such as the transactivation domain of c-Myb, which folds upon binding to the CREB-binding protein (Gianni *et al.*, 2012; Mollica *et al.*, 2016; Bonetti *et al.*, 2018). The disorder in these proteins is also important to their function (Dyson and Wright, 2005), showing that an ordered structure is not required for a protein to have function, which has been termed the disorder-function paradigm in parallel to the previously established structure-function paradigm of proteins (van der Lee *et al.*, 2014; Pinet *et al.*, 2021; Trivedi and Nagarajaram, 2022).

### 1.4.3 Importance of other ligands

The quaternary structure of a protein refers to the inclusion of molecules that are not part of the primary sequence. This includes co-factors and prosthetic groups such as haem molecules in haemoglobin (Perutz, 1976). Molecules such as retinol and $NADP^+$ may bind to proteins (e.g. retinol to $\beta$-lactoglobulin, $NADP^+$ to dihydrofolate reductase) during the folding process and change the energy of the folding landscape in favour of the correct, functional fold (Kasper and Park, 2015; Minić *et al.*, 2020).

Ligands are also key to inducing conformational changes in proteins, which have consequences for their function (Li and Ji, 2019). This underpins a great number of cellular processes, and the presence or absence of ligand therefore affects the structure of the protein being studied. The structure of a protein without a physiological ligand may produce an inactive state, and subsequently have insufficient insight into the relationship between protein structure and function.

## 1.5 Archaea

Archaea are a diverse domain of prokaryotes, and species of archaea occupy every almost every possible ecological niche (Robinson, 2018). In order to adapt to these niches and the associated available nutrients, archaea have many unique proteins that are evolved to fold in the intracellular environment of the cells in these niches, which may have extreme conditions. Archaea are also the only domain with cell membranes formed of ether-linked lipids (Kates *et al.*, 1968; Albers and

Meyer, 2011), as opposed to ester-linked lipids, and this discovery was key to the identification of Archaea as a domain distinct from Bacteria (Woese and Fox, 1977). The evolutionary relationship between archaea, bacteria, and eukarya has been an active area of research since (Iwabe *et al.*, 1989; Woese *et al.*, 1990; Gribaldo and Brochier-Armanet, 2006; Doolittle, 2020).

The current model of the tree of life has two domains: Bacteria and Archaea. Eukaryotes branch off from within the domain of Archaea, with research suggesting that eukaryotes evolved from within the kingdom of Asgard archaea (Zaremba-Niedzwiedzka *et al.*, 2017; Zhang *et al.*, 2025). This disrupted the previous three-domain model of the tree of life, and brought renewed interest in the domain of Archaea as an interesting evolutionary way-point between the last universal common ancestor (LUCA) and eukaryotes (Williams *et al.*, 2020; Doolittle, 2020). Archaea were already a curiosity due to the wide range of complexity across the domain, the number of extremophiles in all environmental niches, and unique membrane structure (Corcelli *et al.*, 2012; Spang *et al.*, 2015; Williams *et al.*, 2020). However, the work linking Asgard archaea to eukaryotes sparked a new interest in their biology in all areas (Ithurbide *et al.*, 2025).

Like bacteria, archaea lack a nucleus and other organelles, and they share a number of other morphological features with bacteria. More complex archaea, generally those with larger genomes, express a number of proteins that are eukaryote-like, such as the cytoskeleton component actin (Stairs and Ettema, 2020; Williams *et al.*, 2020; Martinez-Gutierrez and Aylward, 2022). The evolutionary history of these proteins can be mapped out through identifying homologues in other archaea, however their origin has not yet been pinpointed on a phylogenetic tree (Stairs and Ettema, 2020; Roger *et al.*, 2021). Other proteins found in archaea are not found in eukaryotes. These include proteins that are important in the metabolism of complex energy sources, such as arsenate, or are adapted to the extreme environment in which a particular species of archaea inhabits (Huber *et al.*, 2000; Berezovsky and Shakhnovich, 2005). The latter brings up an interesting

question about how the structure enables the function in conditions that would denature human proteins (Rampelotto, 2013).

## 1.6 Aims of this work

The work presented in this thesis uses two methods, cryo-electron microscopy and molecular dynamics, to study the structure of biological complexes.

Molecular dynamics simulations are used in Chapter 3 to investigate dynamic interactions between a fragment of double-stranded RNA (dsRNA) and a protein nanopore based on FhuA. First, the suitability of FhuA for the translocation of dsRNA is examined. Then, mutations are made to change the electrostatic surface of the nanopore interior, and the effect on the movement of dsRNA is quantified. The role of the solvent in the translocation dynamics of the dsRNA is considered as an important factor.

The ribosome of the archaeon *Nitrosopumilus maritimus* is studied using cryo-electron microscopy in Chapter 4. *Nitrosopumilus maritimus* is a small cell with a small genome. The structural studies are used to inspect the protein composition of the large ribosomal subunit, including those which are shared between archaea and eukaryotes and those that are unique to archaea. This contributes to research on the evolution of the ribosome across the domains of life.

Finally, Chapter 5 shows research on the structural studies of a range of archaeal S-layers. Cryo-electron tomography is used to solve *in situ* structures of S-layers from four species: *Nitrosopumilus maritimus*, *Saccharolobus solfataricus*, *Pyrobaculum oguniense*, and *Pyrobaculum arsenaticum*. S-layers are known to have very divergent sequences and structures. The structures solved within Chapter 5 are compared to published structures and bioinformatics data. Additionally, the previously-published structure of the *Haloferax volcanii* S-layer is studied using molecular dynamics systems to investigate divalent cation binding sites.

# 2

# Theory and Methods

## Contents


**2.1 Molecular mechanics . . . . . 16**
2.1.1 Classical description of biological systems . . . . . 16
2.1.2 Force fields . . . . . 16
2.1.3 Energy minimisation . . . . . 25
2.1.4 Simulations . . . . . 27
2.1.5 Applied electric fields . . . . . 36
2.1.6 Analysis . . . . . 36
**2.2 Cryo-electron microscopy . . . . . 40**
2.2.1 The electron microscope . . . . . 40
2.2.2 Grid preparation . . . . . 43
2.2.3 Data collection . . . . . 44
2.2.4 Data processing . . . . . 45
2.2.5 Model building . . . . . 48
2.2.6 Cryo-electron tomography . . . . . 49
**2.3 Bioinformatics . . . . . 52**
2.3.1 Sequence alignment . . . . . 52
2.3.2 Structure prediction . . . . . 53
2.3.3 Databases . . . . . 53

The research presented in this thesis investigates the structure of molecules in biological systems using two approaches: imaging with electron waves and computational simulations with molecular mechanics. Bioinformatics techniques are used to complement the experimental structural studies.

## 2.1 Molecular mechanics

### 2.1.1 Classical description of biological systems

In classical molecular mechanics, molecules are modelled as "balls" (atoms, or groups of atoms) connected by "springs" (covalent bonds). The balls do not change identity but the springs are able to flex, bend, and stretch depending on the interactions between the balls, while maintaining connections between the balls. As the springs are not formed or removed, this means that covalent bonds are unable to be made or broken in molecular dynamics simulations. There is no explicit representation of electrons, hence the simulations are limited to classical mechanics.

### 2.1.2 Force fields

To work out how these model balls interact with each other, a force field is used to calculate the potential energy of each ball at its given position relative to other balls. The force field describes the properties of each ball based on experimental data or quantum mechanics calculations. It contains the mass, charge, and van der Waals (VdW) radii of each ball, as well as parameters for the bonds and interactions between different balls. These parameters can be used to calculate the potential energy terms for each interaction, and the force experienced by each ball. The work presented in this thesis uses the all-atom force field CHARMM36m, so every atom is described by an individual ball (MacKerell Jr, 2006; Best *et al.*, 2012; Vanommeslaeghe and MacKerell, 2015; Huang *et al.*, 2017).

The potential energy function, $U$, is a combination of the sum of all the bonded and non-bonded potential energy terms as shown in Equation 2.1.2.1.

$$U = U_{bonded} + U_{non-bonded} \tag{2.1.2.1}$$

$U_{bonded}$ includes all the terms from bonds, angles, improper dihedrals, and proper dihedrals as shown in Equation 2.1.2.2.

$$U_{bonded} = \sum U_{bonds} + \sum U_{angles} + \sum U_{dihedrals} + \sum U_{impropers} \tag{2.1.2.2}$$

$U_{non-bonded}$ includes the van der Waals and electrostatic interactions as shown in Equation 2.1.2.3.

$$U_{non-bonded} = \sum U_{VdWs} + \sum U_{Coulomb} \tag{2.1.2.3}$$

#### 2.1.2.1 Bonded terms

The bonded terms of the potential energy function describe the through-bond interactions between atoms, and the potential energies associated with those interactions. Three of the bonded terms (bonds, angles, and impropers) are modelled as harmonic oscillators using Hooke's Law (Hooke, 1678) and thus have reference value(s) $x_0$, displacement $x$, and spring (force) constants $k_x$ described in the force field. Figure 2.1 shows the relationship between the atoms and the type of displacement from equilibrium in each of the bonded terms. Atoms (single letters, e.g. $i$ or $j$) are connected by bonds (pairs of letters, e.g. $ij$ represents the bond between atoms $i$ and $j$), and bonds can have angles between them (three or four letters, e.g. $ijk$ represents the angle between bonds $ij$ and $jk$, and $ijkn$ represents the angle between bonds $ij$ and $kn$).

The first term accounts for the energy associated with bond stretching, or the change in bond length. This is modelled using a harmonic oscillator with reference length $r_{0,ij}$, displacement length $r_{ij}$, and force constant $k_{ij}$ between atoms $i$ and $j$. This is shown in Equation 2.1.2.4 and as a plot in Figure 2.2. The force constant for bond stretching is usually large in force fields compared to the energy of the system, which reflects the considerable energy that is required to cause a bond length to change.

$$U_{bonds} = \sum_{bonds} \frac{1}{2} k_{ij} (r_{ij} - r_{0,ij})^2 \tag{2.1.2.4}$$

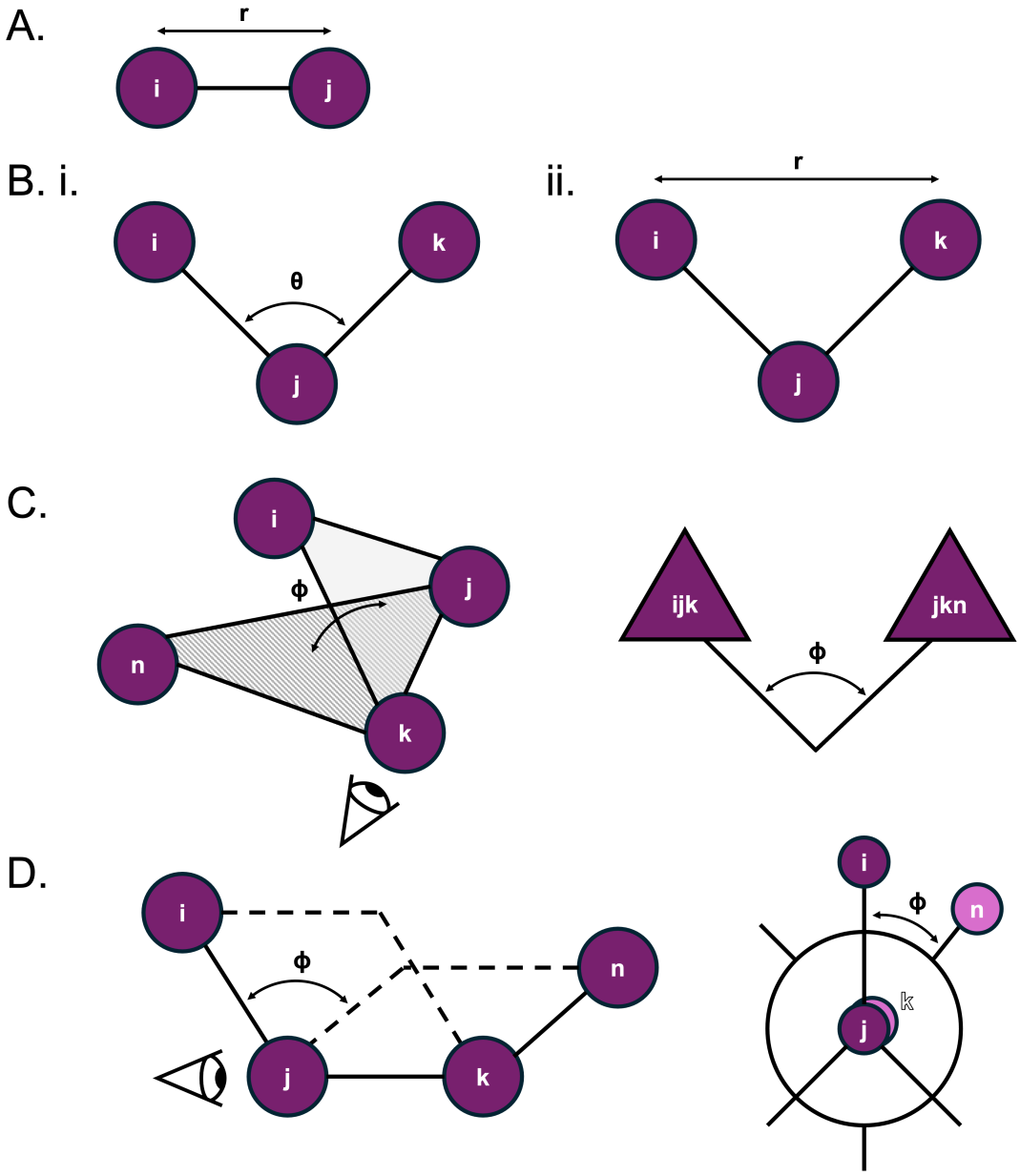



**Figure 2.1:** A. Bond stretching, with the change in distance $r$ between atoms $i$ and $j$. See Equation 2.1.2.4 and Figure 2.2.
B. i. Bond angle, with the change in angle $\theta$ between bonds $ij$ and $jk$. See Equation 2.1.2.5 and Figure 2.3. ii. The Urey-Bradley correction applies a restriction on the values of the distance $r$ between atoms $i$ and $k$. See Equation 2.1.2.6.
C. Bond rotation or torsion of improper dihedrals, with the change in angle in $\phi$ between the planes formed by atoms $ijk$ and $jkn$. See Equation 2.1.2.7 and Figure 2.4.
D. Bond rotation or torsion of proper dihedrals, with the change in angle $\phi$ between the bonds $ij$ and $kn$, when observing along bond $jk$. Also represented as a Newman projection (Newman, 1955). See Equation 2.1.2.8 and Figure 2.5.

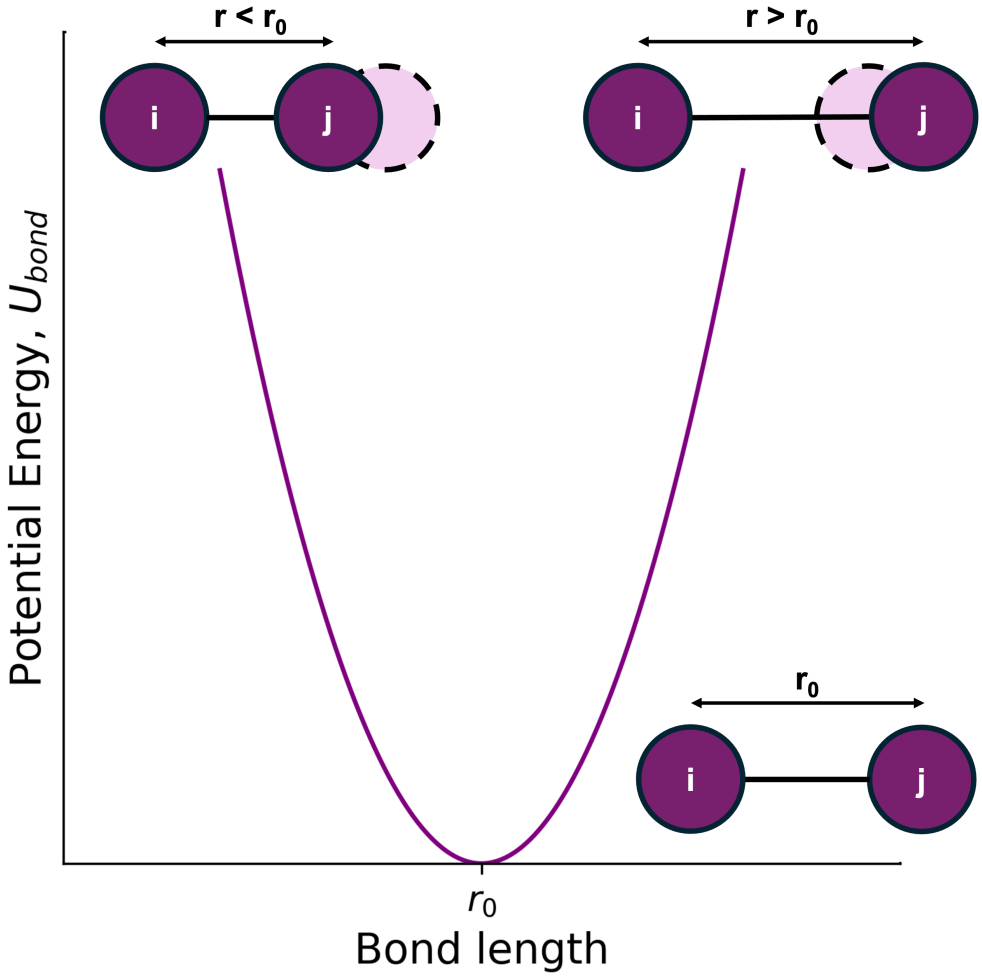



**Figure 2.2:** Plot of bond length against potential energy, as shown in Equation 2.1.2.4.

The angle between two bonds is also modelled using a harmonic oscillator. The reference angle $\theta_{0,ijk}$ describes the angle between bonds $ij$ and $jk$, which is displaced to $\theta_{ijk}$, and the oscillator has a force constant $k_{ijk}$. This is shown in Equation 2.1.2.5 and as a plot in Figure 2.3. The force constants for bond angle changing are also usually large in force fields compared to the energy of the system, as with those for bond stretching.

$$U_{angles} = \sum_{angles} \frac{1}{2} k_{ijk}(\theta_{ijk} - \theta_{0,ijk})^2 \tag{2.1.2.5}$$

The CHARMM36m force field applies a correction to this term, called the Urey-Bradley potential (Vanommeslaeghe and MacKerell, 2015). This applies an addi-

tional restraint on the distance between atoms $i$ and $k$, with a reference distance $r_{0,ik}$, displacement $r_{ik}$, and force potential $k_{UB,ijk}$, as shown in Figure 2.1.B.ii.

$$U_{Urey-Bradley} = \sum_{Urey-Bradley} \frac{1}{2} k_{UB,ijk}(r_{ik} - r_{0,ik})^2 \tag{2.1.2.6}$$

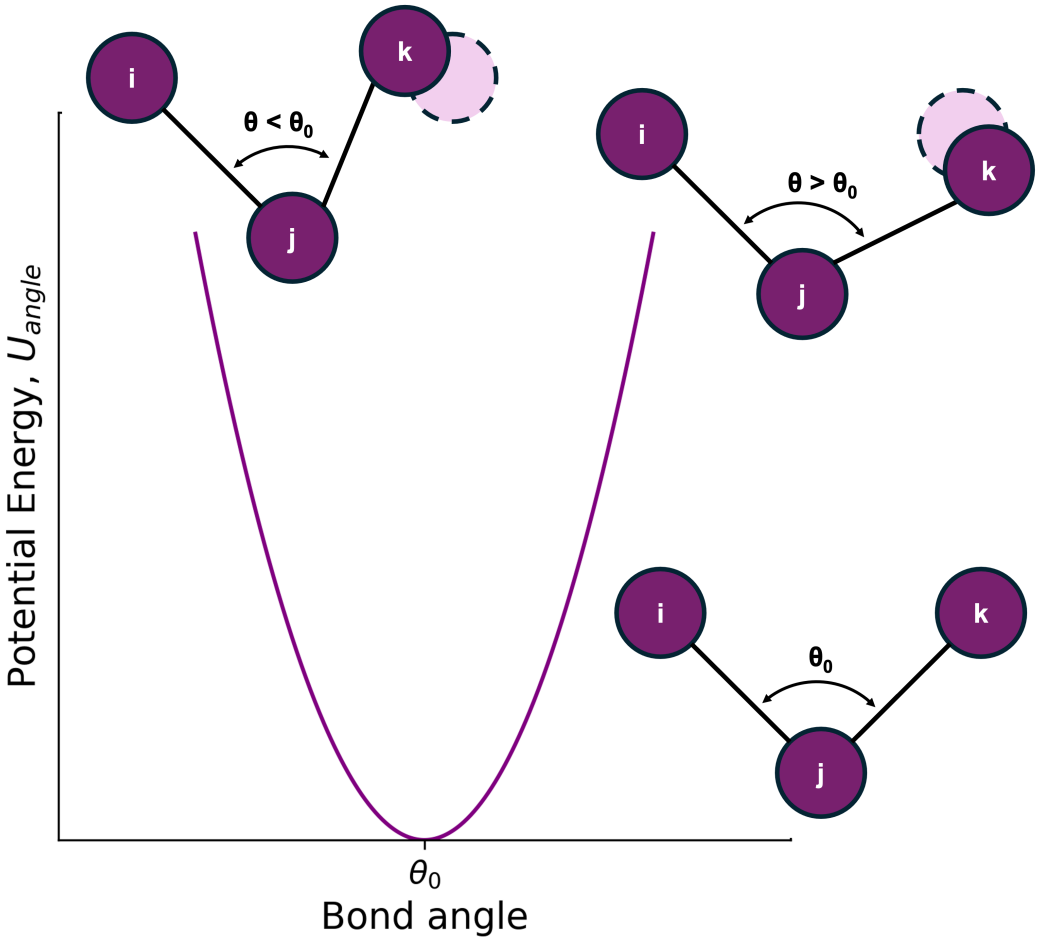



**Figure 2.3:** Plot of bond angle against potential energy, as shown in Equation 2.1.2.5.

Improper dihedrals are used to enforce the planarity of functional groups such as aromatic rings, with $\phi_{ijkn}$ being the torsion angle between two planes defined by atoms $ijk$ and $jkn$ (see Figure 2.1.C). The reference angle is $\phi_{0,ijkn}$, and the force constant $k_{ijkn}$ is high in force fields to maintain the planes of the groups of atoms, but not as large as those for bond lengths and angles. This is

shown in Equation 2.1.2.7 and as a plot in Figure 2.4.

$$U_{impropers} = \sum_{impropers} \frac{1}{2} k_{ijkn}(\phi_{ijkn} - \phi_{0,ijkn})^2 \tag{2.1.2.7}$$

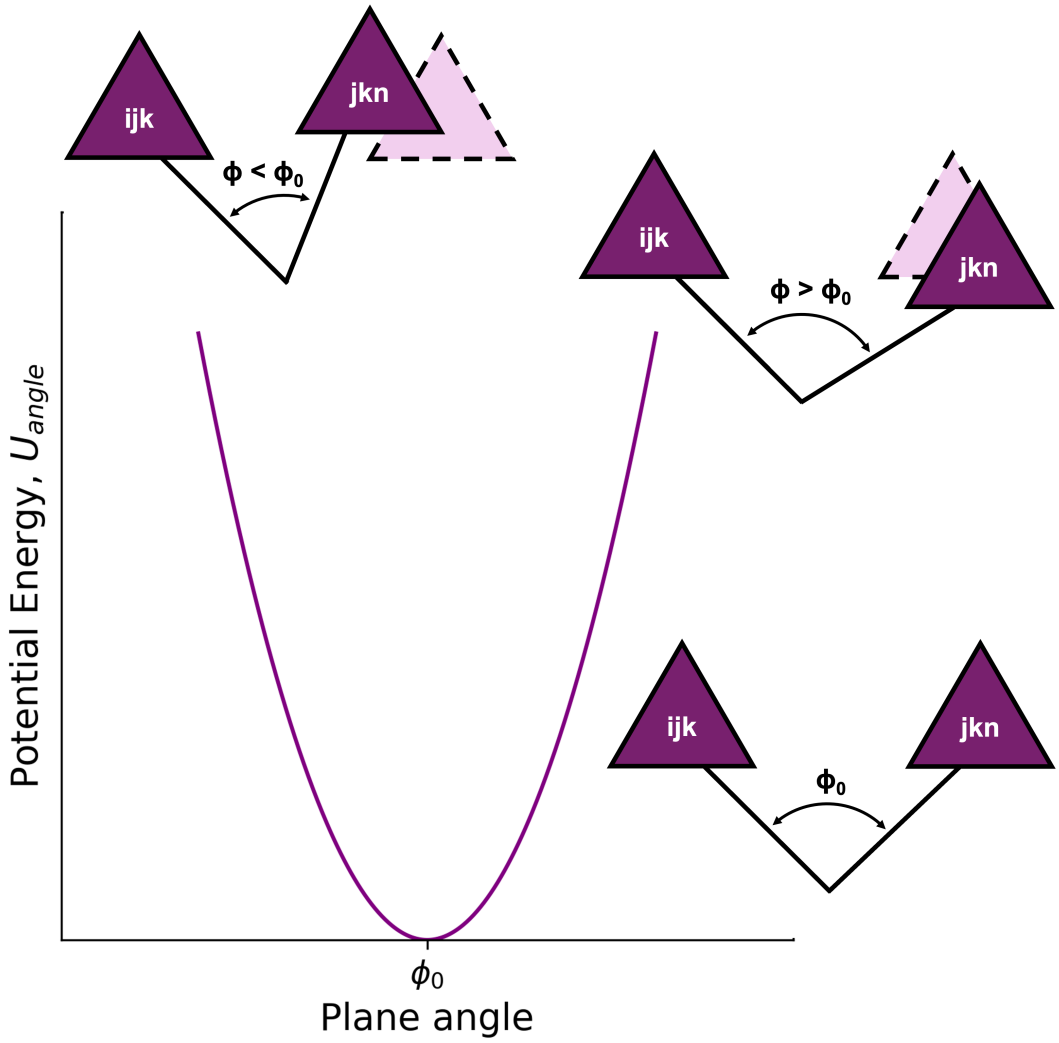



**Figure 2.4:** Plot of plane angle (improper dihedral angle) against potential energy, as shown in Equation 2.1.2.7.

Of the bonded energy terms, the one associated with the most physiological variation in molecule structure is the proper dihedral angle. This is expressed as a cosine series expansion to account for the multiple minima describing the potential energy of the dihedral angle. The torsion angle of the dihedral is $\phi_{ijkn}$, which has a periodicity $m$ giving the number of minima and a phase $\phi_0$. The function

has a force constant of $k_{ijkn}$. This is shown in Equation 2.1.2.8 and as a plot in Figure 2.5. The force constant for this term is very low compared to those for bond length and bond angle, as conformational changes of dihedral angle are less energetically costly and do occur commonly.

$$U_{dihedrals} = \sum_{dihedrals} k_{ijkn}[1 + \cos(m\phi_{ijkn} - \phi_0)] \tag{2.1.2.8}$$

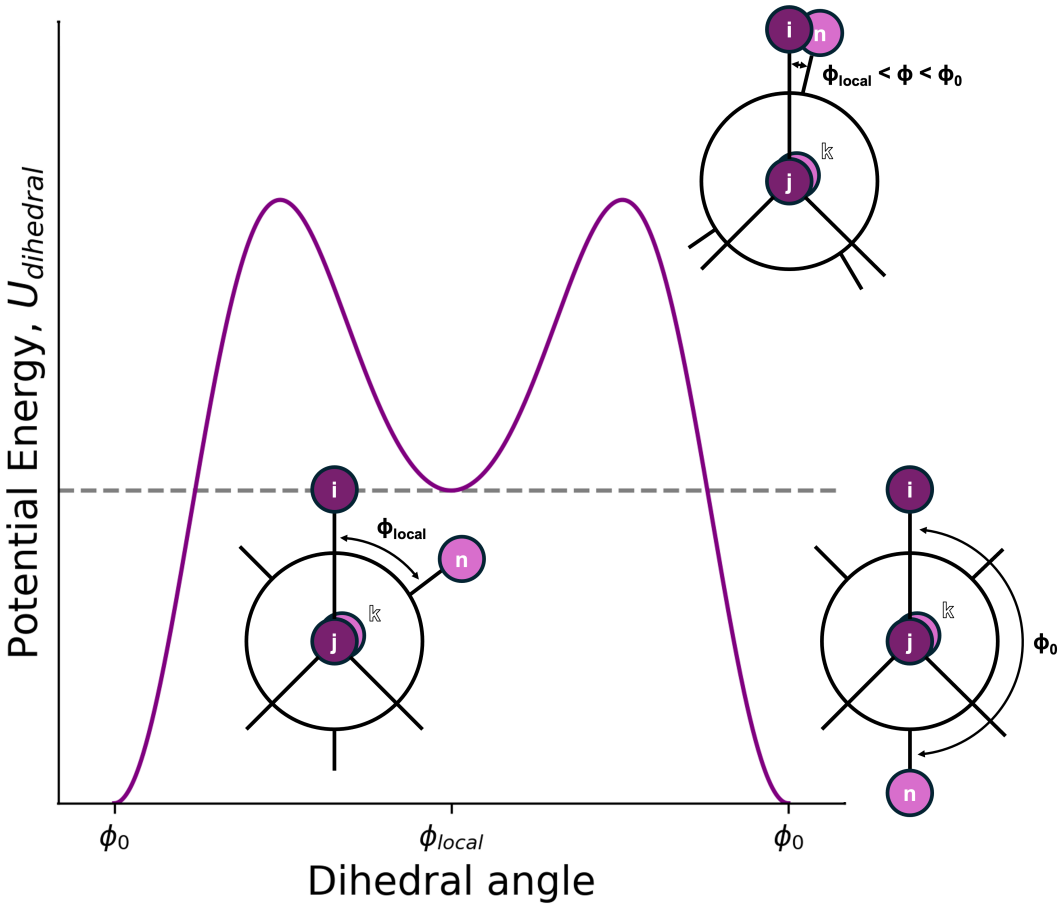



**Figure 2.5:** Plot of dihedral angle against potential energy, as shown in Equation 2.1.2.8. There is a local minimum representing a non-optimal angle, which is still lower in energy than having some overlap of atoms as represented in the associated Newman projections.

All of these terms are summed to calculate $U_{bonded}$ for a given system, for a specific set of coordinates.

#### 2.1.2.2 Non-bonded terms

Non-bonded terms describe the potential energies of all the possible intermolecular interactions between $N$ atoms, that is atoms connected via zero bonds, or four or more bonds. It combines the VdW and electrostatic terms. The potential energy, $U$, is for atom $i$ with respect to atom $j$, and the pair are separated by distance $r_{ij}$ in both terms.

The Lennard-Jones (LJ) potential is used to calculate VdW energies (Jones, 1924). The well-depth of the LJ potential is $\varepsilon_{ij}$ and sets the strength of the interaction, and $\sigma_{ij}$ is the distance between atoms $i$ and $j$ at which the potential energy is zero. The distance between the atoms, $r$, is involved in two terms in the LJ potential. $r^{-12}$ is the strong repulsive term when the distance between the two atoms is too small, and the electron orbitals of the two atoms overlap, known as Pauli repulsion (Pauli, 1925; Frenkel, 1930). $r^{-6}$ is a weak attractive term due to the attractive force between short-ranged instantaneous dipoles, known as London dispersion (London, 1930). This is shown in Equation 2.1.2.9, and as a plot in Figure 2.6.

$$U_{VdWs} = \sum_{non-bonded}^{N} 4\varepsilon_{ij} \left( \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right) \tag{2.1.2.9}$$

The LJ potential has a distance $r_{min,ij}$ where the potential energy is minimised, which can be calculated as shown in Equation 2.1.2.10, and, at this minimum distance, the potential energy $U_{min,ij}$ can also be calculated as shown in Equation 2.1.2.11.

$$r = r_{min,ij} = 2^{\frac{1}{6}} \sigma_{ij} \tag{2.1.2.10}$$

$$U_{min,ij} = -\varepsilon_{ij} \tag{2.1.2.11}$$

A distance cut-off is commonly used when calculating LJ potentials, as atoms far away from an atom $i$ will have a negligible contribution to the potential energy from the VdW energies, and the contribution of atoms further than the cut-off distance is assumed to be zero. This loss in accuracy is countered by the increase

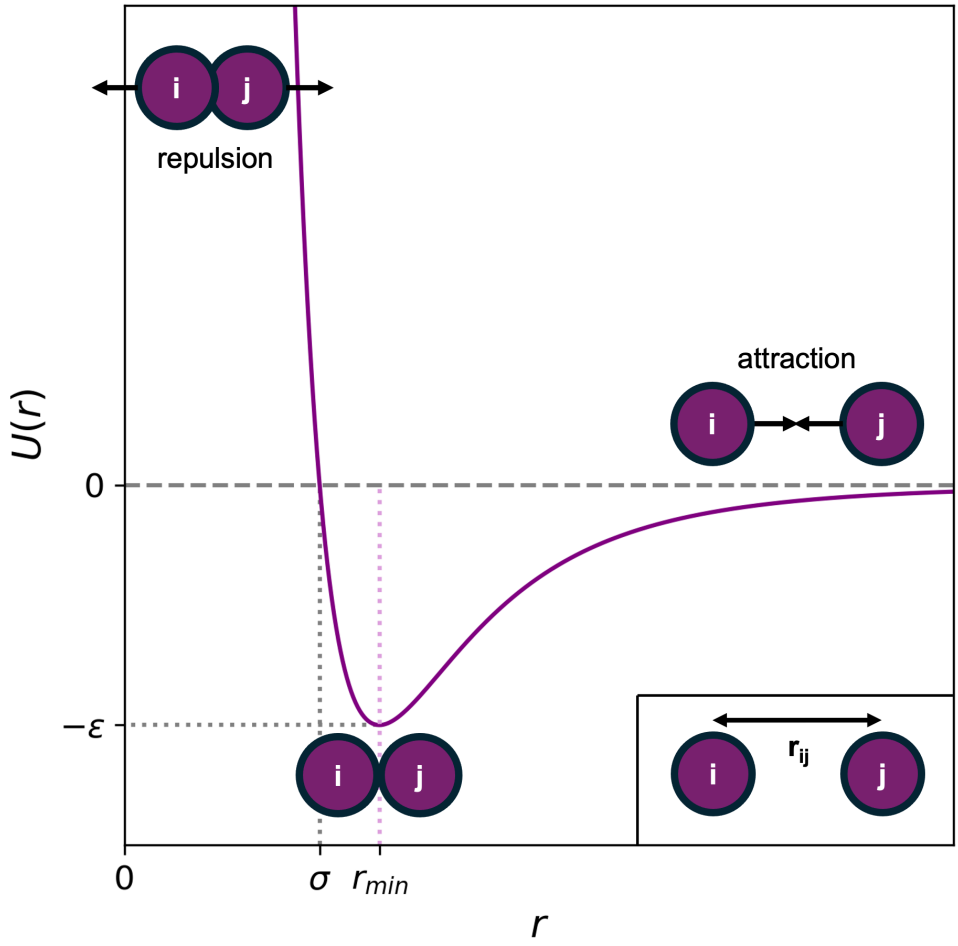



**Figure 2.6:** Plot of the Lennard-Jones potential $U(r)$ against distance $r$, as shown in Equation 2.1.2.9. $r_{min}$ can be calculated as shown in Equation 2.1.2.10, and $-\varepsilon$ is the potential energy of the interaction at $r_{min}$ as shown in Equation 2.1.2.11.

in computational efficiency as the number of calculations required is reduced. The cut-off is not defined in the force field.

The Coulomb potential is used to calculate potential energy from electrostatic interactions, and is based on Coulomb's law (de Coulomb, 1788). The work presented in this thesis uses the partial charge model, as it is more efficient than the alternative multipole moments model (Leach, 2001). In the partial charge model, the partial charges of the atoms involved are $q_i$ and $q_j$, separated by distance $r$, and $\epsilon_0$ is the permittivity of free space. The potential is shown in Equation 2.1.2.12.

$$U_{Coulomb} = \frac{1}{2} \sum_{non-bonded}^{N} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{2.1.2.12}$$

The $r_{ij}^{-1}$ component of the partial charges model means that this term is still important at long-range, compared to the LJ potential, which has $r^{-6}$ and $r^{-12}$ terms that make the term negligible above small distances. This means that the electrostatic interactions between all pairs of atoms in the system have to be considered as contributing to the potential energy, and the calculation of these interactions is often the slowest part of running molecular dynamics simulations. Therefore, a number of methods are used to calculate electrostatics more expeditiously, with minimal impact on accuracy. The simplest is a cut-off based on distance, as is used for LJ potentials. However, this has a much bigger impact on the accuracy of the final potential energy calculated, due to the long-range nature of electrostatic interactions. The reaction field method combines a distance cut-off, within which charges are treated explicitly, with a uniform dielectric constant for all atoms outside of the cut-off. The most sophisticated method is the Particle Mesh Ewald (PME) method, and it is also the most computationally intensive (Ewald, 1921; Darden *et al.*, 1993). In PME electrostatic calculations, charges are equally spaced on a grid to approximate the real positions of the charged particles. Short- and long-range interactions are calculated independently and then summed, with short-range interactions being calculated explicitly and long-range interactions being calculated using fast Fourier transforms. This allows the calculations to be performed more efficiently, with minimal loss of accuracy (Darden *et al.*, 1999). The method of calculating electrostatics is not defined in the force field, but is important for the calculations of potential energy using the information in the force field. The work presented in this thesis uses the PME consideration of electrostatics.

The VdW and Coulomb terms are summed, pair-wise, to calculate $U_{non-bonded}$ for a given system, for a specific set of coordinates.

### 2.1.3 Energy minimisation

The starting coordinates for a successful molecular dynamics simulation need to be at a minimum potential energy, which is unlikely to be the state immediately after system set-up. Thus, the potential energy of the system has to be minimised.

A common method for performing this is a first derivative method, the steepest descents algorithm. The coordinates of the atoms move in a direction that is parallel to the net force, which is downhill on the potential energy hypersurface. This movement is repeated for a set number of steps, or until a local potential energy minimum or another equilibration criterion is reached (see Figure 2.7). It is an algorithm that gives iterative improvement to the coordinates of the system with respect to the total potential energy. Steepest descents energy minimisation allows for steric clashes and other large problems within the system to be quickly resolved, but becomes less efficient as the potential energy approaches a minimum. There is also no guarantee that the minimum located will be the global energy minimum.

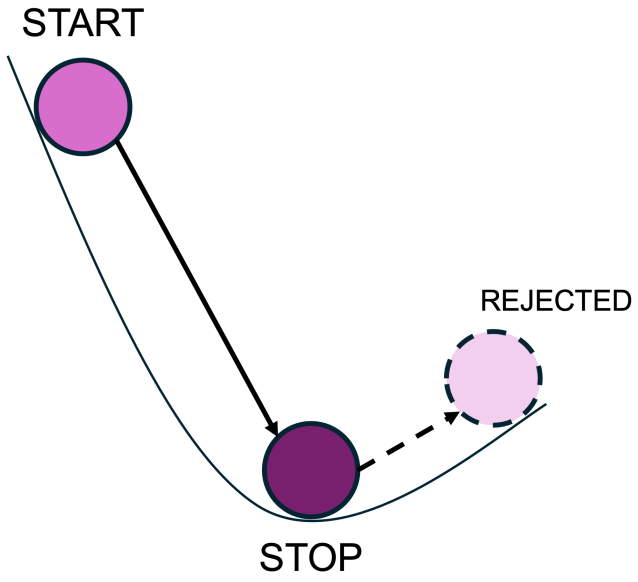



**Figure 2.7:** The atoms are represented as circles, on a 2D energy landscape (curved line). The steepest descents algorithm determines the first derivative of the force experienced by the atom at START. The coordinates of the atom are then moved a distance according to the step size, which places the atom at STOP. The potential energy of the atom at these coordinates is lower, so the position is accepted. Another round of steepest descents determines the first derivative of the force, which this time is a positive gradient and would move the atom to a position with higher potential energy, shown in a dotted circle. Thus, after evaluating the potential energy compared to the position at STOP, this set of coordinates is rejected.

Steepest descents minimisation is performed as shown in Equation 2.1.3.1.

First, the potential energy $U$ is calculated for the system at step $n$, which is then used to calculate the forces $\vec{F}_n$ (see Equation 2.1.4.1). These forces are then used to calculate a shift from the current coordinates of each atom at this step, $\vec{r}_n$, to the new coordinates $\vec{r}_{n+1}$. $\max(|\vec{F}_n|)$ is the largest scalar force acting on any atom in the system, and is used to normalise the coordinate shift for each atom. $h_n$ is the maximum allowed displacement, which acts as a "learning rate" and is modified for step $n+1$ depending on the result of the forces calculated for the new atom positions $\vec{r}_{n+1}$. $h_n$ is increased if the new forces are lower than for step $n$, and decreased if the new forces are higher. In the latter result, the coordinates $\vec{r}_{n+1}$ are rejected and recalculated using the smaller value of $h_n$. The steepest descents algorithm was sufficient for energy minimisation in the systems studied in this thesis.

$$\vec{r}_{n+1} = \vec{r}_n + \frac{\vec{F}_n}{\max(|\vec{F}_n|)} h_n \tag{2.1.3.1}$$

## 2.1.4 Simulations

### 2.1.4.1 Molecular dynamics

Newton's second law of motion states that force of an atom $\vec{F}$ is equal to the mass of the atom $m_i$ multiplied by the acceleration of the atom $\vec{a}_i$ (Newton, 1687). Acceleration can also be written as the second derivative of distance ($\vec{r}_i$ is the change in Cartesian coordinates $x_i$, $y_i$, $z_i$) with respect to time $t$. Force is also related to the potential energy $U$ by taking the negative first derivative of the potential energy. This directly relates the force experienced by the atom to the potential energy of the atom and the distance it has moved, as shown in Equation 2.1.4.1. From this, new coordinates can be calculated after the passing of a given timestep $\delta t$. This temporal dependence makes molecular dynamics a deterministic method.

$$\vec{F} = m_i \vec{a}_i = m_i \frac{d^2 \vec{r}_i(t)}{dt^2} = -\frac{dU}{d\vec{r}_i} \tag{2.1.4.1}$$

To perform molecular dynamics simulations, Newton's second law of motion is applied at each timestep, until the maximum simulation time is achieved as shown in Figure 2.8. Potential energy is calculated using the force field, and accounting for

the system coordinates $\vec{r}_i(t)$ and assigned (for the first step) or previously calculated velocities $\vec{v}_i(t)$. This is then used to calculate the forces on each atom, and Newton's second law is solved for all atoms using the timestep. Atom positions and velocities are then updated, and if the maximum time is not reached, then the cycle starts again by calculating the new potential energies.

In the work presented in this thesis, the GROMACS engine is used to perform the molecular dynamics simulations as detailed in the respective chapters (2022.1 in Chapter 3, 2020.3 in Chapter 5), using an integrator to calculate the movements of atoms in the simulation (Bekker *et al.*, 1993; Berendsen *et al.*, 1995; Van Der Spoel *et al.*, 2005; Abraham *et al.*, 2015, 2022).

#### 2.1.4.2 Integrators

The calculation of new positions and velocities for the atoms in a system is not a simple problem, as these two properties are coupled. This becomes a many-body problem that cannot be solved analytically, as the positions and velocities of all of the atoms are interdependent. To overcome this problem, Newton's second law of motion is integrated using a finite difference method. All algorithms using this method start with the assumption that the positions and dynamic properties of the atoms can be approximated as a Taylor series expansion (Taylor, 1715). A Taylor series expansion allows a complex function to be approximated by an infinite series of polynomials at a given point on the function. For $\vec{F}_i = m_i\vec{a}_i$, the expansion takes the form shown in Equation 2.1.4.2, Equation 2.1.4.3 and Equation 2.1.4.4, with $\vec{r}_i$ being the positions, $\vec{v}_i$ the velocity (first derivative of $\vec{r}_i$ with respect to time $t$), $\vec{a}_i$ the acceleration (second derivative), and $\vec{b}_i$ the third derivative and so on. The more polynomials that are included, the more accurate the approximation becomes.

$$\vec{r}_i(t+\delta t) = \vec{r}_i(t) + \delta t\vec{v}_i(t) + \frac{1}{2}\delta t^2\vec{a}_i(t) + \frac{1}{6}\delta t^3\vec{b}_i(t) + \cdots \tag{2.1.4.2}$$

$$\vec{v}_i(t+\delta t) = \vec{v}_i(t) + \delta t\vec{a}_i(t) + \frac{1}{2}\delta t^2\vec{b}_i(t) + \cdots \tag{2.1.4.3}$$

$$\vec{a}_i(t+\delta t) = \vec{a}_i(t) + \delta t\vec{b}_i + \cdots \tag{2.1.4.4}$$

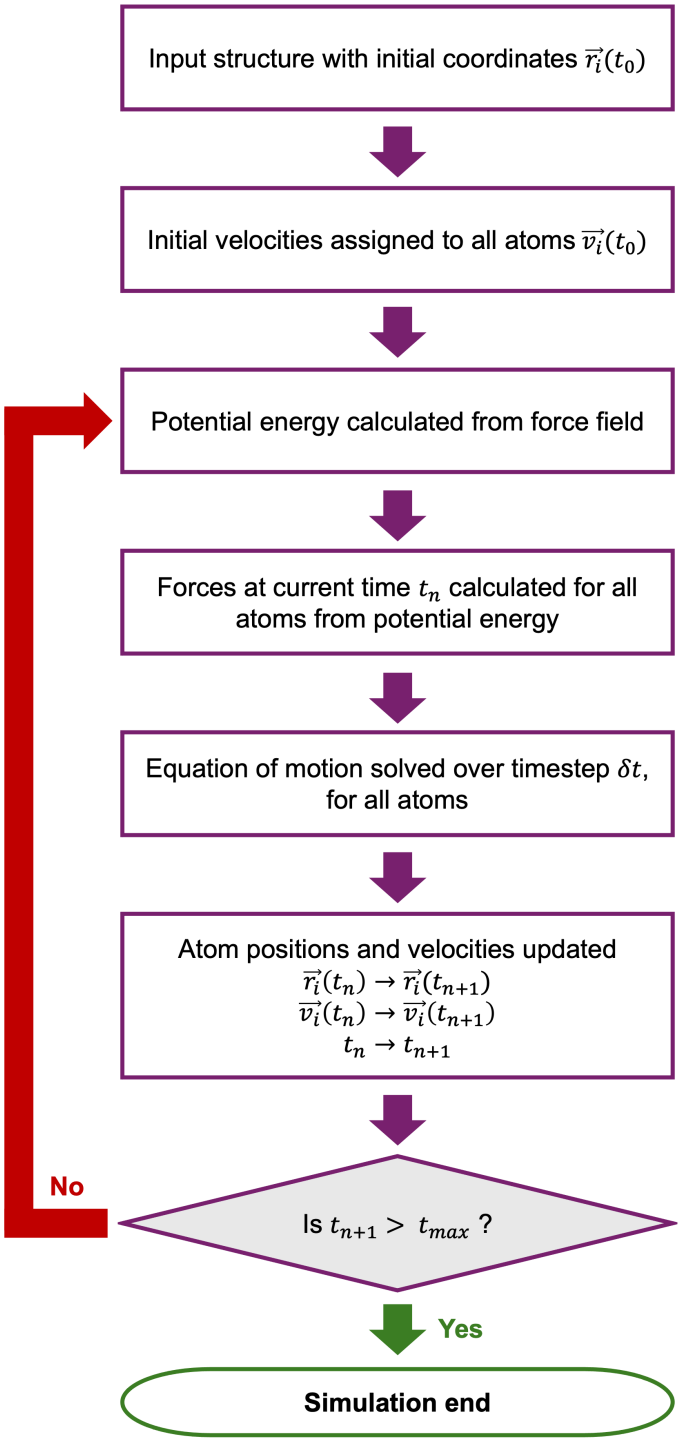



**Figure 2.8:** A typical workflow for equilibration and production MD simulations. Simulations only end when $t_{max}$ is reached, or an error in the system causes the software performing the simulations to crash.

The leapfrog algorithm is a variation on the Verlet algorithm (Verlet, 1967), and is used for integrating Newton's second law in molecular dynamics simulations uses the relationships shown in Equation 2.1.4.5 and Equation 2.1.4.6 (Hockney, 1970). First the velocities $\vec{v}_i(t + \frac{1}{2}\delta t)$ are calculated using $\vec{v}_i(t - \frac{1}{2}\delta t)$ and $\vec{a}_i$. The positions $\vec{r}_i(t + \delta t)$ can then be deduced from $\vec{v}_i(t + \frac{1}{2}\delta t)$ and $\vec{r}_i(t)$. This can be visualised as the calculations of velocities "leapfrogging" the positions from $t - \frac{1}{2}\delta t$ to $t + \frac{1}{2}\delta t$, and then the positions "leapfrogging" the velocities from $t$ to $t + \delta t$, as visualised in Figure 2.9. The leapfrog algorithm is utilised in all of the MD simulations in the work presented in this thesis.

$$\vec{r}_i(t + \delta t) = \vec{r}_i(t) + \delta t \vec{v}_i \left(t + \frac{1}{2}\delta t\right) \tag{2.1.4.5}$$

$$\vec{v}_i \left(t + \frac{1}{2}\delta t\right) = \vec{v}_i \left(t - \frac{1}{2}\delta t\right) + \delta t \vec{a}_i(t) \tag{2.1.4.6}$$

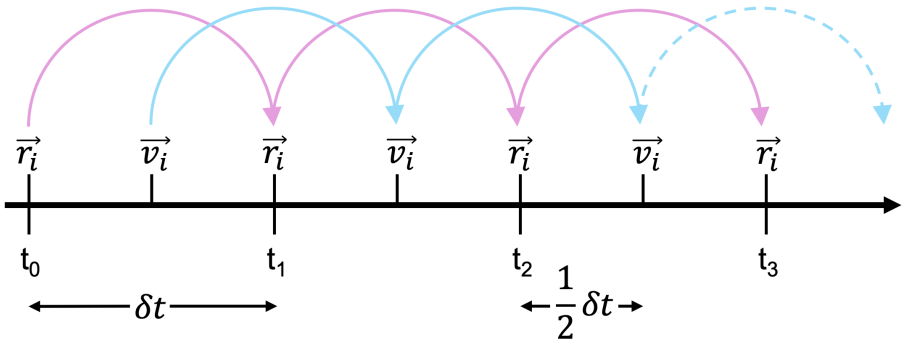



**Figure 2.9:** Diagram showing how the leapfrog algorithm works for integrating Newton's second law of motion explicitly involving $\vec{r}_i$ and $\vec{v}_i$. The positions $\vec{r}_i$ are calculated at each time $t$ in timesteps $\delta t$. The velocities $\vec{v}_i$ are calculated at times half a timestep from those for the positions $t + \frac{1}{2}\delta t$.

### 2.1.4.3 The periodic boundary condition

As the simulations are trying to replicate part of a larger environment, there cannot be hard boundaries present at the edges of a simulation box with a limited and finite number of particles. To overcome this, periodic boundary conditions are applied in

such a way that the particles experience forces as if they were in a bulk solution with no edges. The simulation box is effectively infinitely replicated in all directions to give a periodic array, as visualised in Figure 2.10. In three dimensions, five shapes are able to tesselate and fill all of the space: the cube, the hexagonal prism, the truncated octahedron, the rhombic dodecahedron, and the elongated dodecahedron. The work in Chapter 3 uses a cube and the work in Chapter 5 uses a hexagonal prism, as these shapes are the most appropriate for the systems being studied.

#### 2.1.4.4 Statistical mechanics

A molecular dynamics simulation is a microscopic state that aims to obey the macroscopic constraints of the biological system it is emulating. A thermodynamic ensemble is the set of all possible microscopic states that a system can have within these macroscopic limits. A foundational assumption of molecular dynamics is that the system is ergodic; averages of thermodynamic properties over a long trajectory will match the averages of a thermodynamic ensemble. This allows conclusions to be drawn from the limited number of microscopic states that are sampled by the simulations, with the assumption that they represent the average thermodynamic properties of the ensemble.

The standard thermodynamic ensembles used in molecular dynamics simulations of biological systems are the isochoric (NVT) and isobaric-isothermal (NPT) ensembles, so named because of the macroscopic properties that are maintained during the simulations. In NVT ensembles, the number of atoms (N), volume of the simulation (V), and temperature (T) are maintained. In NPT ensembles, the number of atoms (N), pressure (P), and temperature (T) are maintained. The latter is most commonly used for recreating realistic biological conditions, and was used for all production simulations in this thesis. The thermostats and barostats employed by these ensembles to maintain the conditions are discussed below.

#### 2.1.4.5 Temperature coupling

In a simulation, the effective temperature is represented by the total kinetic energy of the system. A higher total kinetic energy will lead to particles moving with

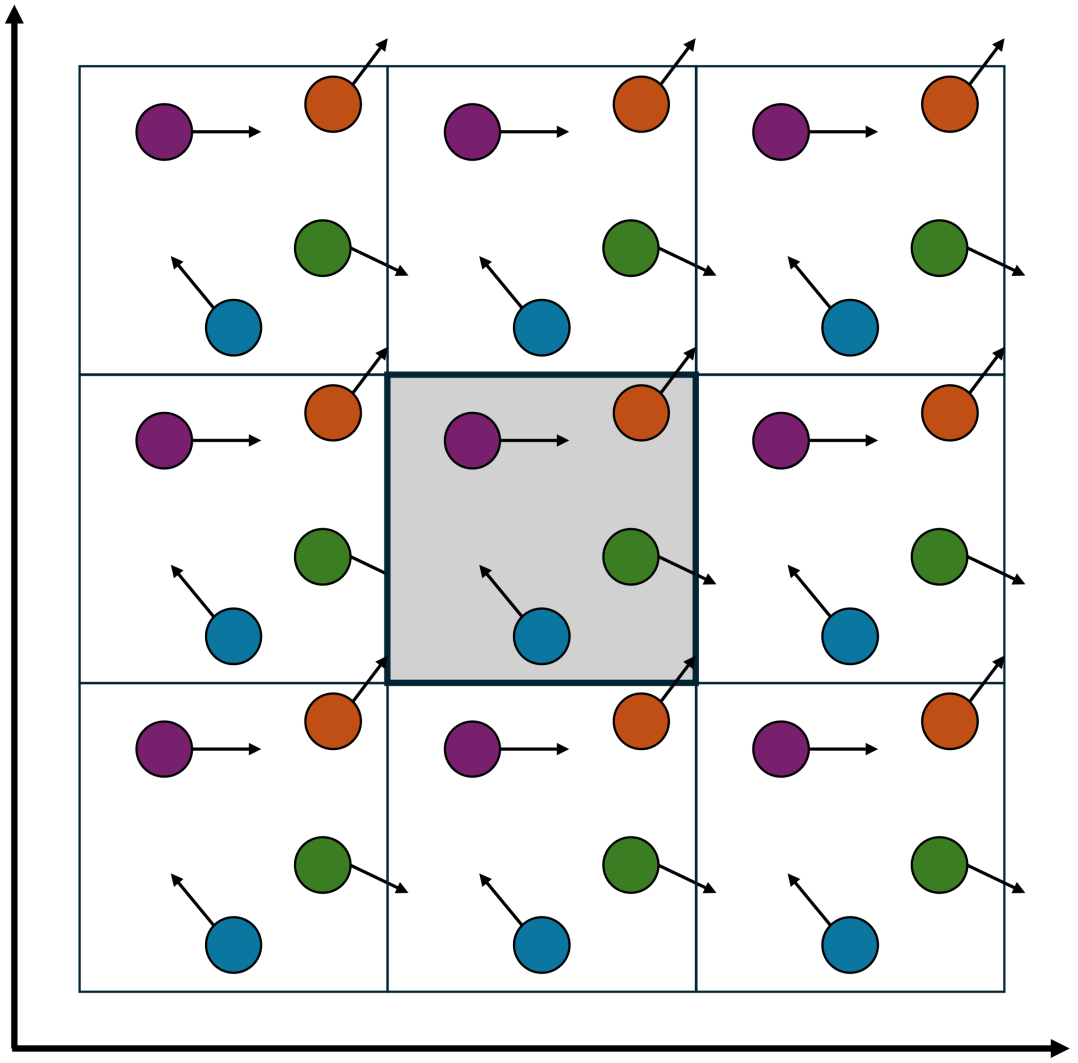


**Figure 2.10:** Visualisation of the periodic boundary condition in two dimensions, $x$ and $y$. The vertical and horizontal lines represent the boundary conditions, and the arrows indicate the direction in which a particle is moving. The central box, highlighted in grey, is the only one actually simulated; all others are projections showing that a particle leaving at the right edge of the box would immediately re-enter at the left edge, at the same height.

higher velocities on average. The distribution of velocities will follow the Maxwell-Boltzmann distribution, with a higher temperature giving a broader distribution of velocities (Maxwell, 1860; Boltzmann, 1868). For a given frame, the instantaneous temperature can be calculated from the velocities of the particles as shown in Equation 2.1.4.7. This is a rearrangement of the equipartition theorem (Waterston and Strutt, 1892), where $k_B$ is the Boltzmann constant and $N_f$ is the degrees of freedom of the atoms being used to calculate temperature.

$$T = \frac{2}{N_f k_B} \sum_{i=1}^{N} \frac{1}{2} m_i v_i^2 \tag{2.1.4.7}$$

The instantaneous temperature of a system will fluctuate greatly between frames, and thus it is more useful to evaluate an average temperature. This is done by time-averaging the instantaneous temperatures over a sufficiently long time period.

For the NPT and NVT ensembles, maintaining the temperature of the system is important and thus thermostats are used to constrain the temperature. There are a number of different thermostats available for molecular dynamics simulations. The production simulations presented in this thesis use the Velocity-rescale (V-rescale) thermostat (Bussi *et al.*, 2007). This is an extension of the Berendsen thermostat, shown in Equation 2.1.4.8 (Berendsen *et al.*, 1984). The Berendsen thermostat corrects deviations from the reference temperature $T_0$ by exponential decay. The rate of decay is determined by $\tau_T$, which is a coupling constant determining how quickly the decay happens.

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau_T} \tag{2.1.4.8}$$

The Berendsen thermostat is very effective in bringing a system to a target temperature, however this method also suppresses the natural kinetic energy fluctuations required for a correct thermodynamic ensemble. Therefore this thermostat is commonly used for equilibration simulations (as was the case for the research in this thesis), but not for production simulations. The V-rescale thermostat addresses the weaknesses of the Berendsen thermostat by including a stochastic term to ensure the system generates the correct kinetic energy distribution

$K_0$, as shown in Equation 2.1.4.9. $dK$ is the change in kinetic energy that happens over $dt$. $dW$ is a term representing the Wiener process, and forms part of the stochastic term that modulates the temperature fluctuations around the control by including the description of a random walk (Wiener, 1949).

$$dK = (K_0 - K)\frac{dt}{\tau_T} + 2\sqrt{\frac{KK_0}{N_f}}\frac{dW}{\sqrt{\tau_T}} \tag{2.1.4.9}$$

#### 2.1.4.6 Pressure coupling

The pressure of a system is related to the volume of the system. Equation 2.1.4.10 is the virial expression for pressure, combining an ideal gas component and a virial component that accounts for interactions (Clapeyron, 1834; Clausius, 1870). The ideal gas component $Nk_BT$ comes from the kinetic energy of the system and assumes the particles do not interact. $N$ is the number of particles, $k_B$ is the Boltzmann constant, and $T$ is the temperature of the system. The virial component is the time-averaged sum of the dot product $\vec{r}_{ij} \cdot \vec{F}_{ij}$, where $\vec{r}_{ij}$ is the vector between atoms $i$ and $j$, and $\vec{F}_{ij}$ is the force between atoms $i$ and $j$. This accounts for the interaction between particles, as they do not behave like an ideal gas.

$$P = \frac{Nk_BT}{V} + \frac{1}{3V}\left\langle \sum_{i<j} \vec{r}_{ij} \cdot \vec{F}_{ij} \right\rangle \tag{2.1.4.10}$$

Pressure needs to be controlled in the NPT thermodynamic ensemble. This is achieved in simulations using a barostat. In the research presented in this thesis, production simulations use the Parrinello-Rahman barostat (Parrinello and Rahman, 1981). This is a more rigorous extension of the Berendsen barostat, shown in Equation 2.1.4.11 (Berendsen *et al.*, 1984). In the same way that the Berendsen thermostat corrects deviations in temperature by exponential decay, the Berendsen barostat corrects deviations in pressure from the target tensor $\mathbf{P}_0$ by exponential decay, applied by scaling the vectors of the simulation box. The coupling constant is $\tau_{\mathbf{P}}$. The pressure tensors $\mathbf{P}_0$ and $\mathbf{P}$ describe the magnitude and direction of the pressure in the system.

$$\frac{d\mathbf{P}}{dt} = \frac{\mathbf{P}_0 - \mathbf{P}}{\tau_{\mathbf{P}}} \tag{2.1.4.11}$$

The Berendsen barostat does not produce a correct NPT thermodynamic ensemble, and is thus only used for equilibration simulations to achieve the target pressure tensor. The Parrinello-Rahman barostat includes terms that enable a change in simulation box size and shape by handling box dimension vectors as a matrix $\mathbf{b}'$. This allows pressure to be handled semi-isotropically, which is important for membrane systems such as those used in Chapter 3. $\mathbf{W}$ is a matrix determining the strength of coupling between the system and the barostat that accounts for the isothermal compressibility of the system. The barostat is expressed as a matrix equation of motion, as shown in Equation 2.1.4.12.

$$\frac{d^2\mathbf{b}}{dt^2} = \frac{V}{\mathbf{W}\mathbf{b}'}(\mathbf{P} - \mathbf{P}_0) \tag{2.1.4.12}$$

#### 2.1.4.7 System equilibration

After energy minimisation, the system needs to achieve thermodynamic equilibration. First, the NVT ensemble is used with the Berendsen thermostat to bring the system to a target temperature while maintaining the system volume. This allows the randomly-placed solvent molecules to rearrange to a more thermally stable configuration within the simulation, while the molecule(s) of interest have their positions restrained by applying a high force constant. The system pressure is monitored to assess the equilibration, and the temperature is monitored to check the thermostat has functioned correctly (Abraham *et al.*, 2022).

The NPT ensemble is then used with the Berendsen barostat, which allows the system volume to fluctuate and achieve the target pressure tensor. For membrane systems, which are semi-isotropic, further equilibration is needed with the Parrinello-Rahman barostat, as the Berendsen barostat cannot handle anisotropy. During equilibration with the NPT ensemble, position restraints on the target molecule(s) are reduced and then removed gradually, to allow the relaxation of the target within the equilibrated solvent environment. Equilibration is assessed by monitoring system density and ensuring no molecules are experiencing unrealistic forces (Abraham *et al.*, 2022).

Once the equilibration criteria have been met, production simulations can be run, generally using the NPT ensemble. The NPT ensemble was maintained for production simulations presented in this thesis using the V-rescale thermostat and the Parrinello-Rahman barostat.

The simulations described in this research required different equilibration protocols with the NPT ensemble. This is because the nanopore system involves a lipid membrane and therefore semi-isotropic pressure coupling. After an initial equilibration with the Berendsen barostat, three further equilibrations were performed with the Parrinello-Rahman barostat, starting with 1000 kJ $mol^{-1}$ $nm^{-2}$ position restraints on the protein and RNA, and 400 kJ $mol^{-1}$ $nm^{-2}$ position restraints on the lipids, and reducing these until a final NPT equilibration without position restraints.

### 2.1.5 Applied electric fields

In the GROMACS molecular dynamics engine, an external electric field can be applied according to Equation 2.1.5.1, where $E_0$ is field strength, angular frequency is $\omega = 2\pi v/\lambda$ ($v$ is the wave speed and $\lambda$ is wavelength, and $v/\lambda$ is equal to the frequency $f$), $t_0$ is time at the start of the peak in the field strength, and $\sigma$ is the width of the pulse. The variables are specified to GROMACS in the following order (and units): $E_0$ (V $nm^{-1}$), $\omega$ ($ps^{-1}$), $t_0$ (ps), $\sigma$ (ps).

$$E(t) = E_0 \exp\left[-\frac{(t-t_0)^2}{2\sigma_2}\right] \cos[\omega(t-t_0)] \qquad (2.1.5.1)$$

When an electric field is applied to a membrane system along the normal of the membrane, this gives the effect of opposite electrodes above and below the membrane, mimicking a membrane voltage (Roux, 2008). In Chapter 3, an electric field was utilised to move molecules based on the resulting charge gradient.

### 2.1.6 Analysis

A number of analyses were performed on the molecular dynamics simulations of this thesis, using the packages MDAnalysis (Michaud-Agrawal *et al.*, 2011; Gowers *et al.*, 2016), NumPy (van der Walt *et al.*, 2011), Pandas (McKinney, 2011), and SciPy

(Virtanen *et al.*, 2020) for Python, and the GROMACS tools gmx sasa, gmx covar, and gmx anaeig (Bekker *et al.*, 1993; Eisenhaber *et al.*, 1995; Van Der Spoel *et al.*, 2005; Abraham *et al.*, 2015). The theories behind the analyses are detailed below.

#### 2.1.6.1 Root mean square deviation

The root mean square deviation $\varrho(t)$ (RMSD) of a group of $N$ atoms, relative to a reference structure $\mathbf{x}_i^{\text{ref}}$ as a function of time is calculated as shown in Equation 2.1.6.1. RMSD is performed on structures that have been translationally and rotationally superpositioned to the reference structure. Calculations were performed in the MDAnalysis module for Python (Theobald, 2005; Liu *et al.*, 2010; Michaud-Agrawal *et al.*, 2011; Gowers *et al.*, 2016). RMSD was used to give a global assessment of structural change over time in simulations.

$$\varrho(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{x}_i(t) - \mathbf{x}_i^{\text{ref}}\right)^2} \tag{2.1.6.1}$$

#### 2.1.6.2 Root mean square fluctuation

The root mean square fluctuation (RMSF) of an atom $i$ is computed as the time average of positional change, as shown in Equation 2.1.6.2. RMSF is performed on structures that have been translationally and rotationally superpositioned to the reference structure. Calculations were performed in the MDAnalysis module for Python (Liu *et al.*, 2010; Michaud-Agrawal *et al.*, 2011; Gowers *et al.*, 2016). RMSF was used to identify residues with large motions during the simulations, suggesting regions of flexibility.

$$\varrho_i = \sqrt{\left\langle \left(\mathbf{x}_i - \langle \mathbf{x}_i \rangle\right)^2 \right\rangle} \tag{2.1.6.2}$$

#### 2.1.6.3 Solvent accessible surface area

The solvent accessible surface area (SASA) is the effective area of a molecule that can be contacted by solvent. This is calculated by probing the surface of a molecule with a sphere of representative size for the solvent being used. SASA was performed using gmx sasa in GROMACS (Eisenhaber *et al.*, 1995). SASA

was used to identify changes in molecule shape by following the area accessible to solvent over the course of simulations.

#### 2.1.6.4 Adaptive Poisson-Boltzmann Solver

The Adaptive Poisson-Boltzmann Solver (APBS) is used to calculate the electrostatic surface of biological molecules. It uses the Poisson-Boltzmann equation to provide a global solution for the electrostatic potential within a biomolecule in an implicit solvent (Fogolari *et al.*, 2002; Ren *et al.*, 2012; Jurrus *et al.*, 2018). APBS was used within PyMOL to visualise the electrostatic surface of molecules (PyMOL, 2015).

#### 2.1.6.5 Principal Component Analysis

Principal component analysis (PCA) is a statistical dimensionality reduction method used for reducing complex data to the essential features in a given number of components (Pearson, 1901; Hotelling, 1933; Greenacre *et al.*, 2022). These principal components explain the greatest variance in the data. PCA was performed using the GROMACS tools gmx covar and gmx anaeig, with gmx covar calculating a covariance matrix of fitted structures to obtain eigenvectors, which were then analysed with gmx anaeig to project the eigenvectors onto the structure, or the trajectory onto the eigenvectors. PCA was performed on reduced structures, where only a single bead represented each residue in the RNA (see Chapter 3, Figure 3.14 and Figure 3.20). The results were used to inspect the conformational space explored by the molecules during the simulations.

#### 2.1.6.6 Calculating current

The work of Göpfrich *et al.* (2016) was followed to calculate the instantaneous current $I$ (C $s^{-1}$) across the nanopore in Chapter 3, as shown in Equation 2.1.6.3. The sum over $i$ indicates a sum over all the ions, $\Delta t$ is the time interval between two consecutive frames of the MD simulations, $q_i$ is the charge of ion $i$, and $\zeta_i$ is defined in Equation 2.1.6.4. $z_i$ is the $z$ coordinate of ion $i$.

$$I(t + \frac{\Delta t}{2}) = \frac{1}{\Delta t\ \ell} \sum_{i}^{N} q_i(\zeta_i(t + \Delta t) - \zeta_i(t)) \tag{2.1.6.3}$$

$$\zeta_i(t) = \begin{cases} z_i(t), & |z_i(t)| \leq \ell/2 \\ -\ell/2, & z_i(t) < -\ell/2 \\ \ell/2, & z_i(t) > \ell/2 \end{cases} \tag{2.1.6.4}$$

Calculations were carried out within the $-\ell/2 \leq z_i \leq \ell/2$ region of the system. For this work, $\ell = 30$ Å and $q_i = 1.6 \times 10^{-19}$ C. The average current was calculated by summing up all of the instantaneous currents and dividing by the number of frames of the trajectory.

The conductance of the pore was calculated in siemens by using Ohm's Law, the relationship between conductance $G$, current $I$, voltage $V$, and resistance $R$, as shown in Equation 2.1.6.5 (Ohm, 1827).

$$G = \frac{I}{V} = \frac{1}{R} \tag{2.1.6.5}$$

## 2.2 Cryo-electron microscopy

### 2.2.1 The electron microscope

The architecture of the electron microscope is shown in Figure 2.11. Electrons can be used in a microscope because of wave-particle duality and the fact that electrons are charged. Wave-particle duality is a core principle of quantum physics, and the de Broglie relationship states that the wavelength $\lambda$ of a particle is dependent on its momentum $p$ (velocity multiplied by mass) as shown in Equation 2.2.1.1 (de Broglie, 1924; Davisson and Germer, 1927). Therefore electrons can act as waves, and these wave-like properties are exploited in the electron microscope (Ruska, 1934a,b). The de Broglie relationship can be combined with the relationship between energy $E$ and momentum to directly calculate wavelength from energy. At higher energies, and therefore greater velocities, the electrons have shorter wavelengths. This energy can be such that the electrons have an equivalent wavelength to X-rays, in the range of 1 Å and sufficient to distinguish between two adjacent atoms (Cheng, 2015). This gives a theoretical resolution sufficient for atomic model building, as wavelength is inversely proportional to resolution.

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{2mE}} \tag{2.2.1.1}$$

Energy can be expressed in terms of voltage $V$ and charge $Q$, using the relationship in Equation 2.2.1.2. In the research presented in this thesis, 300 kV is used to accelerate the electrons in the microscope, which is a commonly-used voltage in cryo-EM (Bakker *et al.*, 1996; Egerton, 2014).

$$V = \frac{E}{Q} \tag{2.2.1.2}$$

While there are no lenses suitable for X-rays (leading to the phase problem of X-ray crystallography, where only the diffraction pattern is collected), electrons are charged particles and therefore magnets can be used as lenses for the electron waves as shown in Figure 2.11 (Ruska, 1934a). Modern electron microscopes for structure determination have a sample chamber maintained in a vacuum at liquid nitrogen

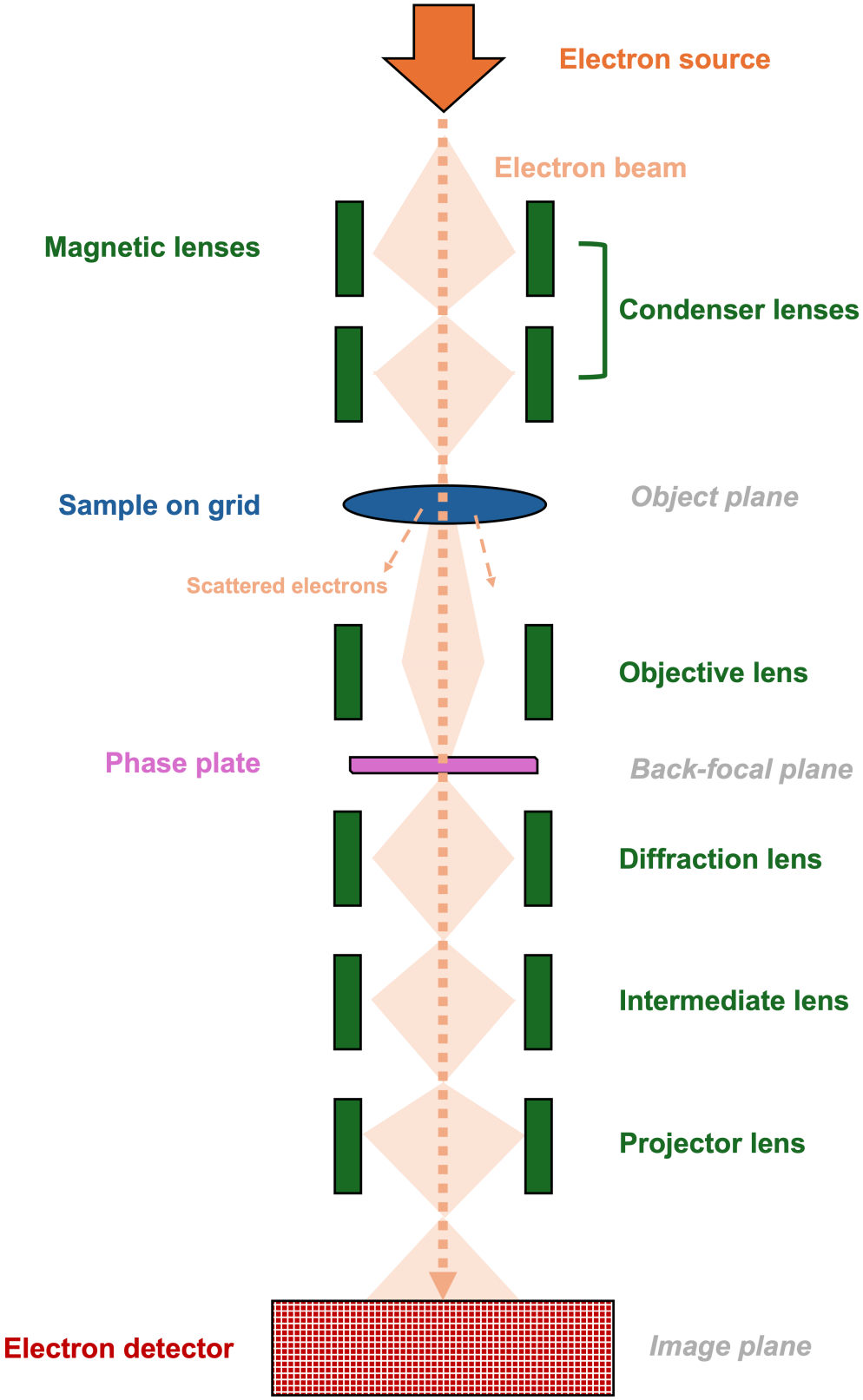



**Figure 2.11:** Schematic of an electron microscope. Electron beam is shown in orange, with the magnetic lenses (green) focusing the beam. The phase plate is at the back-focal plane, which is the plane where the diffraction pattern forms.

temperature to reduce the inelastic scattering of electrons and the rate of sample damage caused to biological samples by the application of high energy electrons (Knapek and Dubochet, 1980; Dubochet *et al.*, 1988; Dickerson *et al.*, 2022). This temperature is the reason the technique is named cryogenic-electron microscopy (cryo-EM). Many cryo-EM microscopes use a cold field emission gun as an electron source, which produces a localised beam of electrons with high temporal coherence and a small energy spread (Hamaguchi *et al.*, 2019; Konings and Bischoff, 2020). The low energy spread, combined with the use of an energy filter that removes inelastically scattered electrons, leads to an improved image contrast by reducing the effects of chromatic aberration (Nakane *et al.*, 2020; Chua *et al.*, 2022; Neselu *et al.*, 2023). The microscopes used in the research presented in this thesis were Titan Krios G3s (ThermoFisher Scientific), with an energy filter of 20 eV.

#### 2.2.1.1 Image formation

The interaction between the sample and the electrons forms the resulting image. The electrons are scattered by the atoms in the sample, either elastically (no loss of energy) or inelastically (electron loses energy as an emitted photon), and are focused onto the detector (Dickerson *et al.*, 2022). Inelastically scattered electrons are a source of noise in the final image. Elastically scattered electrons contribute to the contrast of the image on the detector, due to interference with unscattered electrons. The scattered and unscattered electrons combine constructively when they are in phase to increase the intensity of the signal, and combine destructively when they are out of phase. The microscope lens can be adjusted to change the focal point of the electrons, which changes the relative phases of the scattered and unscattered electrons. This leads to enhanced contrast as some signals are intensified and some are decreased, and is known as phase contrast (Toyoshima and Unwin, 1988). Phase contrast methods are necessary for cryo-EM of biological samples because the elements in biological molecules are poor scatterers due to their relatively low mass (Carroni and Saibil, 2016). Molecules containing nucleic acids

have increased scattering due to the presence of phosphorous atoms, which have double the mass of the common biological elements of carbon, nitrogen, and oxygen.

### 2.2.2 Grid preparation

Biological samples for cryo-EM are mounted on a grid made of materials such as copper or rhodium, and often coated with carbon (Russo and Passmore, 2014; Naydenova *et al.*, 2020). The aim is to have the sample in the holes between the mesh grid. However, the carbon grid becomes hydrophobic when exposed to the environment for extended periods of time, and this hydrophobicity would cause the sample to clump and distribute unevenly. Therefore, grids are glow discharged using ionised gas to make the grid more hydrophilic and allow the sample to spread evenly across the surface and the target particles to disperse randomly (Aebi and Pollard, 1987; Hirst *et al.*, 2024). After sample application to the grid, subsequent steps can be performed manually or by using a machine such as the Vitrobot Mark IV (ThermoFisher Scientific), as was used in the research in this thesis. The sample is blotted to get the layer as thin as possible ($< 100$ nm for single-particle studies, $< 500$ nm for tomography). This is necessary to reduce inelastic scattering, and because samples that are too thick cannot be imaged as there is insufficient contrast (Dickerson *et al.*, 2022). The grid is then cryo-vitrified to liquid nitrogen temperatures, by plunging into liquid ethane. Samples need to be plunge-vitrified to ensure that the freezing occurs rapidly, allowing for the formation of vitreous ice (Dubochet *et al.*, 1988). This ensures that there is minimal crystalline ice, which both damages the sample and causes strong diffraction of electrons. Liquid ethane must be used rather than direct plunging into liquid nitrogen because liquid nitrogen has a low heat capacity and vaporises upon contacting a material at temperatures higher than $-196$ °C, meaning that an insulating layer of nitrogen gas is formed and slows the cooling of the sample (the Leidenfrost effect (Leidenfrost, 1756)).

### 2.2.3 Data collection

Although sample damage is reduced at cryogenic temperatures, damage to the sample still accumulates and this limits the total dose of electrons per image (Glaeser, 1971; Dickerson *et al.*, 2025). Biological samples also do not scatter electrons strongly due to the low mass of the elements constituting biological molecules (Henderson, 1995). Therefore it is important that as many electrons as possible, that pass through the sample, are recorded. This increases the signal, and thus improves the signal-to-noise ratio. Direct electron detectors are used to record the data, and record every electron that lands on the detector leading to a very high detective quantum efficiency (DQE) (McMullan *et al.*, 2009). The time that the electron lands on the detector is also recorded, giving temporal resolution and allowing the image to be captured as a movie. This allows information from later frames to be given less importance during data processing, as the damage to the sample will have accumulated, in a process called dose-weighting (Wu *et al.*, 2016). Data can be collected at super-resolution, meaning that the landing of a single electron can be registered by multiple pixels that are all smaller than the size of an electron (Booth, 2012). Super-resolution counting of electrons has a low DQE, however the information can be used to achieve final resolutions beyond the physical Nyquist (Feathers *et al.*, 2021). Cryo-EM images of a sample are collected at a range of defoci, which gives a range of contrast in the images (Frank, 2006). Images with higher defoci will have higher contrast, which is important for particle identification above noise, at the expense of the highest resolution data. This is combined with images taken at lower defoci that contain this information at a lower contrast. Data collection in the research presented in this thesis used the EPU software (Thermo Fisher Scientific) or the SerialEM software (Mastronarde, 2005) to control the microscope.

### 2.2.4 Data processing

#### 2.2.4.1 Data corrections

Images taken with electron microscopes are not perfect. This is due to a number of factors including lens aberrations in the microscope, sample movement, and applied defoci. These factors can be corrected for if they are correctly modelled. The contrast transfer function (CTF) is a mathematical model of the Thon rings in the Fourier transform of a cryo-EM image (Erickson and Klug, 1971). Thon rings are alternating bright and dark rings that obscure true details in the collected image due to the distortion of contrast (Thon, 1966). CTFs can be accurately modelled using programs such as CTFFIND4, and then the Thon rings and modelled rings can be visually inspected (Rohou and Grigorieff, 2015). Additional corrections can be applied to images with a number of other programs. This includes spherical aberrations in the lens of the microscope, which distort the shape of the image and can be observed as deviations in the Thon rings from true circles (Zivanov *et al.*, 2020), and beam tilt, caused by the electron beam not running completely perpendicular to the detector and leading to a shift in the image (Zivanov *et al.*, 2018). A number of factors can contribute to the movement of the sample during data collection, including the drift of the sample stage and the movement of the sample within the ice due to the irradiation from the electron beam. Sample motion can be corrected for either by individual particle or by whole image, by aligning the raw images of the collected movie and averaging them to produce a single micrograph (Li *et al.*, 2013; Scheres, 2014; Zheng *et al.*, 2017b; Zivanov *et al.*, 2019). Sample deterioration due to electron dose can be accounted for using dose-weighting, which down-weights the data collected later in the image as the damage accumulates (Zheng *et al.*, 2017b). The consequence of this data processing by modelling of confounding factors increases the signal-to-noise ratio of the images collected (Baxter *et al.*, 2009).

#### 2.2.4.2 Single particle analysis

After such processing, the images can be combined to perform single particle analysis (SPA) of the sample. First, the particles must be picked, usually in two stages: a manual stage, and a computational phase. Software such as Topaz use convolutional neural networks to take the manual picks, and teach an algorithm to make picks (Bepler *et al.*, 2019). This allows the picking process to be much faster, and results in more particles being identified. As a cryo-EM sample optimised for SPA is a solution of the purified target particle, the particles are ideally randomly oriented on the cryo-EM grid, and thus each particle gives a random 2D projection in the resulting images (Tan *et al.*, 2017). However, there are samples which do not orientate randomly on a grid, and show preferred orientation thus limiting the projections. Orientation bias can be addressed by a variety of methods including changing the adsorption properties of the grid or adding detergents, so the sample is randomly distributed in the vitreous ice. With sufficient particles, this will result in 2D projections covering the whole sample. Once picked, these projections are classified in 2D so that any incorrect picks or poor quality images can be removed. The high quality projections can then be combined to make a 3D density by exploiting the fact that the Fourier transform of each 2D image is equivalent to a central slice of a 3D Fourier transform of the particle that generated that image (Sigworth, 2016). The 2D classifications are optimised to select the best 2D images, which are then combined to generate one or more 3D structures, depending on the homogeneity of the sample. Each 3D structure can then be refined using independently refined half-maps and assessment of angular assignments of each contributing image, until a high-resolution structure of the particle is obtained from the calculated electron density map. This electron density map can be sharpened to enhance contrast and amplify the high-resolution information which is typically down-weighted during 3D image reconstruction. B-factor sharpening specifically improves the visualisation of fine details by enhancing the contrast at the higher resolutions (Rosenthal and Henderson, 2003).

In the research presented in Chapter 4, SPA was performed in RELION, which implements the aforementioned corrections to collected images and the processing pipeline described above (Fernandez-Leiro and Scheres, 2017; Zivanov *et al.*, 2018, 2019). RELION uses a Bayesian approach to 3D refinement, improving on the Expectation-Maximisation-like approach previously used by a popular predecessor, FREALIGN, by reducing the likelihood of over-fitting (Li *et al.*, 2010).

#### 2.2.4.3 Calculating resolution

Calculating the resolution of cryo-EM structures is not as simple as it is for X-ray crystallography. A measure of resolution can be calculated by creating two independent half-maps of the structure through randomly assorting the contributing particles into two structures. These half-maps can then be compared by measuring the normalised cross-correlation coefficients that are calculated between the two 3D volumes over corresponding shells in Fourier space (Harauz and Van Heel, 1986). The coefficients give a measure of similarity between the corresponding Fourier shells for the two structures, and plotting these coefficients against the Fourier frequency gives the Fourier Shell Correlation (FSC) curve. An approximation of global resolution can be taken by reading the resolution (calculated as the reciprocal of Fourier frequency) at the point on the curve where the correlation coefficient is equal to 0.143, known as the gold-standard FSC (Henderson *et al.*, 2012; Scheres and Chen, 2012; Henderson, 2013; Lander, 2024). It is important to consider the shape of the FSC curve when assessing the accuracy of this resolution, with more accurate resolutions being associated with steep FSC curves. It is also important to visually inspect the structure, and compare the quality to other maps of the same resolution.

#### 2.2.4.4 Local resolution

Unlike in X-ray crystallography, resolution of density maps in cryo-EM is not uniform across the structure. There is variation in local resolution for a number of reasons, including flexibility in a region of the particle leading to poor alignment, and averaging heterogeneous conformations (Cardone *et al.*, 2013). It is possible to

calculate and visualise local resolution, which allows interpretation of the structure to be informed by these variations (Kucukelbir *et al.*, 2014).

### 2.2.5 Model building

Atomic models can be built into high-resolution structures of biological molecules. This is easiest when the sequence of amino or nucleic acids is known, and when homologous proteins can be used as templates. Models can be assessed by examining the model and comparing what has been built to known chemical rules about biological structures. This ensures realistic angle and bond parameters. Model building in Chapter 4 was performed using Coot (Emsley *et al.*, 2010). Refinement and validation were performed using PHENIX (Adams *et al.*, 2010).

#### 2.2.5.1 Refinement

Refinement refers to the assessment and improvement of a model by comparing it to known chemical and biological rules, which are modelled as restraints. In PHENIX, refinement is performed with RealSpaceRefine, and then validated by MolProbity, which assesses a number of measures (Williams *et al.*, 2018; Afonine *et al.*, 2018a). Ramachandran outliers and disallowed conformers are plotted and identified to check the $\phi$ and $\psi$ torsion angles of each amino acid residue compared to those allowed in real polypeptides (Ramachandran *et al.*, 1963; Ramachandran and Sasisekharan, 1968). Clashes are identified by examining where the VdW spheres of atoms would overlap, which is particularly important if the model has been built without hydrogen atoms. The C-Alpha Based Low-resolution Annotation Method (CaBLAM) evaluates polypeptide main chain geometry (Prisant *et al.*, 2020). C$\beta$ position outliers, cis or twisted proline residues, side chain rotamer outliers, and geometry restraints are also evaluated and refined within this process. It is worth noting that some outliers will be correct, such as rotamer outliers, Ramachandran outliers, and proline residues with cis peptide bonds. The number of outliers should be checked against the incidence in solved protein structures.

#### 2.2.5.2 Validation

Model validation additionally checks that the sequence of a model is correct by comparing to a given sequence. A map-to-model FSC correlation coefficient is also calculated, which looks at the correlation between the real density map and a density map calculated from the model (Afonine *et al.*, 2018a). This gives an assessment of the accuracy of the model compared to the structure, rather than checks based on chemistry and probability of a conformation being an outlier.

#### 2.2.5.3 Modelling of large complexes

The modelling and refinement of large complexes can be more difficult, as clashes have to be identified and fixed between all components of the complex. Refinement performed of individual components may produce errors as chains can be incorrectly moved to regions of density that belong to other molecules. Therefore, it is important to either refine the complex as a whole, or to carefully mask out regions of the map before refinement to ensure the molecule stays in the correct area of density. At lower resolutions, from around 3.5 Å, this can be added by the application of secondary structure restraints to maintain the model geometry which has been built.

### 2.2.6 Cryo-electron tomography

Cryo-ET is a structural method that obtains 3D volumes from samples by tilting the sample and taking images of the same area at multiple angles. These angles are then reconstructed into the 3D volume. As multiple images are being taken, and the dose-dependent damage of the sample still applies, lower doses of electrons can be applied during the collection of each image. This leads to a lower signal-to-noise ratio, and makes the reconstruction of tomograms and identification of structures above the noise more difficult. The Crowther criterion describes the minimum number of tilt images that are required to make a 3D reconstruction at a given resolution (Crowther *et al.*, 1970). The minimum size of objects that can be observed in cryo-ET images is much larger compared to cryo-EM. However,

there have been many recent advances in the field of cryo-ET that make it a valuable method to assess structures *in situ*.

#### 2.2.6.1 Tomogram processing

In addition to the image processing methods for electron microscopy data described in subsection 2.2.4, tomography data has specific processing requirements. To reconstruct the 3D volume, the 2D tilt images must be aligned. Commonly this is achieved by adding gold fiducials to the sample before imaging. These are small gold particles of a known size, which have a high contrast due to the large mass of the gold atoms. This allows the fiducials to be used as markers for alignment (Fung *et al.*, 1996; Brandt *et al.*, 2001). Cryo-ET data has a "missing wedge", so named because the tilt images do not cover 360° of the sample. This is due to imaging not being possible at high tilt angles, as the sample becomes too thick (Lucić *et al.*, 2005; Lučić *et al.*, 2013). Typically, the missing wedge is around 60° and must be accounted for in processing. Also, the images taken at higher tilt angles are commonly taken later in the imaging process, meaning the sample is more dose-damaged when these images are captured. These images are taken later as the increased sample thickness due to the tilting already decreases the attainable resolution, so the loss of high-resolution information is less impactful. Beam-induced sample motion has to be accounted for in cryo-ET just as in cryo-EM, however different solutions were required due to the position changing as well as the tilt angle changing, which makes the modelling of sample motion more difficult (Fernandez *et al.*, 2018, 2019). Also, 3D CTFs need to be calculated for the volumes to correct for deviations in contrast transfer (Turoňová *et al.*, 2017). For the work in Chapter 5, tomogram reconstruction was performed in IMOD (Kremer *et al.*, Feb; Mastronarde and Held, 2017).

#### 2.2.6.2 Sub-tomogram averaging

The equivalent process to SPA for cryo-ET is sub-tomogram averaging (STA). This is the process of extracting sub-volumes from the tomograms and aligning them to produce an average structure of higher resolution (Briggs, 2013). Alignment

is improved iteratively, with 3D masks being used to refine the volumes being aligned and the rotation and translation of alignment being reduced with each iteration to allow more fine-grain sampling of alignments. There has been significant work to improve the process of sub-tomogram averaging, both in aligning the sub-volumes and optimising the data collection (Bharat *et al.*, 2015; Böhning and Bharat, 2021). Data acquisition of tilt-series has been automated on electron microscopes (Mastronarde, 2005), and accelerated by the development of parallel collection modes (Bouvette *et al.*, 2021). The tilt-scheme has been optimised to decrease the impact of sample damage (Hagen *et al.*, 2017). In data processing, the 3D CTF correction for sub-volumes has been improved and refined using a Bayesian approach (Bharat and Scheres, 2016; Zivanov *et al.*, 2022). Sub-tomogram averaging in Chapter 5 was performed using the AV3 and TOM toolboxes in MATLAB (Nickell *et al.*, 2005; Förster and Hegerl, 2007; Förster *et al.*, 2008).

## 2.3 Bioinformatics

A number of bioinformatics tools offer useful complementary information to structural studies of biological molecules and were utilised in the research presented in Chapter 4 and Chapter 5 of this thesis.

### 2.3.1 Sequence alignment

#### 2.3.1.1 Basic Local Alignment Search Tool

The Basic Local Alignment Search Tool (BLAST) is a tool, currently hosted by the National Center for Biotechnology Information (NCBI), used to identify homology between protein or nucleic acid sequences (Altschul *et al.*, 1990, 1997; Johnson *et al.*, 2008). Using the Accepted Point Matrix (PAM) proposed by Dayhoff *et al.* (1979), it calculates a maximum segment pair measure by scoring the similarity or lack thereof between residues in the target and search sequences. BLAST can be used to scan complete databases or specific subsections of databases, or even provided sequences. The segments with high scores are then relayed to the user, along with the score and the annotated alignment. The parameters given to BLAST can be optimised based on the use case, including the method of scoring, the depth of search, and the length of segment necessary to identify homology.

#### 2.3.1.2 COnstraint-Based ALignment Tool

The NCBI tools server uses the constraint-based alignment tool (COBALT) to perform multiple sequence alignment (MSA) of user-submitted amino acid sequences (Papadopoulos and Agarwala, 2007). COBALT derives pairwise constraints from databases of conserved domains (Marchler-Bauer *et al.*, 2005), protein motifs (Hulo *et al.*, 2006), and sequence similarity (Altschul *et al.*, 1990), and uses these to perform a progressive multiple alignment of the submitted sequences. Similarities across alignments are scored, and then the algorithm iteratively realigns, rescores, and refines the alignment. COBALT can only perform alignments on sequences with sufficient homology as identified using BLAST, or the alignment fails.

### 2.3.2 Structure prediction

#### 2.3.2.1 Transmembrane Hidden Markov Model

Transmembrane Hidden Markov Model (TMHMM) is a sequence-based structure prediction tool, which calculates the likelihood of a protein sequence containing a transmembrane helix segment (Sonnhammer *et al.*, 1998). It utilises the positive-inside rule (von Heijne, 1986), and established probabilities that sequences form $\alpha$-helices and interact with the membrane, and uses a hidden Markov model to predict if the observed sequence features combine to become a transmembrane helix (Baum and Petrie, 1966). The presence or absence of a transmembrane helix in the structure is the hidden state being assessed by the model. TMHMM was trained using previously established databases of transmembrane proteins where transmembrane helices were confirmed (Jones *et al.*, 1994; Wallin and von Heijne, 1998). The output of TMHMM is a graph of the likelihood that residues within a sequence are part of a transmembrane helix, with a confidence score.

### 2.3.3 Databases

A number of databases were used during the course of the research presented in this thesis. These databases are all open-source, with the goal of making the process of research and knowledge more accessible. The Universal Protein Resource (UniProt) consortium was first launched by Bairoch *et al.* (2005) (The UniProt Consortium, 2025). It is regularly updated with new annotations of protein sequences, proteomes, and published structures. Annotations of protein sequences are taken from the HAMAP database (High-quality Automated and Manual Annotation of microbial Proteomes), which has been extended from the original limitation to microbial proteomes (Lima *et al.*, 2009; Pedruzzi *et al.*, 2015). Within UniProt, sequence alignment can be run using BLAST or ClustalOmega (Altschul *et al.*, 1990; Sievers *et al.*, 2011). Each protein sequence in UniProt has a unique ID, which starts with a code identifying the species from which it originates. UniProt was used to obtain the sequences of ribosomal proteins in Chapter 4 and S-layer proteins in Chapter 5. GenBank, hosted by NCBI, is a database of genome sequences and was used to obtain

the sequences of ribosomal RNA in Chapter 4 (Benson *et al.*, 2008). The Research Collaboratory for Structural Bioinformatics (RCSB) hosts the Protein Data Bank (PDB) (Berman *et al.*, 2000). This is a repository of validated protein and nucleic acids structures, and the models are available to download along with the statistics comparing the model to other published models. All of the published models utilised in this thesis were taken from the PDB, and the accompanying statistical data was used to confirm the accuracy of the model. Where the structure was solved using cryo-EM, the electron density was downloaded from the Electron Microscopy Data Bank (EMDB) (Tagari *et al.*, 2002). The PDB and EMDB are linked, to make it simple to find associated structures and models. They are also linked directly to UniProt. Where experimental protein structures do not exist, UniProt also provides predicted models of proteins sourced from the AlphaFold database (Varadi *et al.*, 2022). This is a database of protein structure models predicted using AlphaFold2 (Jumper *et al.*, 2021), and contains models for all annotated proteins in UniProt. The Encyclopedia of Domains takes the AlphaFold database and uses the CATH database to classify domains in the protein models (Lau *et al.*, 2024). CATH classifies domains into hierarchical levels: Class (C), Architecture (A), Topology (T) and Homologous superfamilies (H)(Orengo *et al.*, 1999; Sillitoe *et al.*, 2021). This denotes the level of structural similarity between domains, as each homologous superfamily has a four-part code that relates to the levels.

# 3

# Translocation of double-stranded RNA through a nanopore

## Contents


**3.1 Introduction . . . . . 56**
3.1.1 Nanopores . . . . . 56
3.1.2 FhuA . . . . . 61
3.1.3 Double-stranded RNA . . . . . 62
3.1.4 Aim of this chapter . . . . . 68
**3.2 Materials and Methods . . . . . 70**
3.2.1 FhuA model . . . . . 70
3.2.2 Calculating ion flux through non-occluded FhuA . . . . 72
3.2.3 Double-stranded RNA model . . . . . 72
3.2.4 Simulation set-up . . . . . 73
3.2.5 Analysis . . . . . 76
**3.3 Results . . . . . 78**
3.3.1 FhuA has an appropriate pore for dsRNA translocation 78
3.3.2 Conformational behaviour of dsRNA in MD simulations 80
3.3.3 Initial simulations of WT FhuA . . . . . 84
3.3.4 Ion conductance through WT FhuA is affected by mutations 87
3.3.5 dsRNA does not always escape the pore in simulations . 91
3.3.6 Interrogating the role of dsRNA structure in translocation 92
3.3.7 Characterisation of protein-RNA electrostatic interactions 96
3.3.8 Mutations impact protein-RNA electrostatic interactions 99
3.3.9 Mutations impact the speed of dsRNA translocation . . 104
3.3.10 Translocation in the opposite direction . . . . . 108
3.3.11 Towards nanopore sequencing . . . . . 110
**3.4 Discussion . . . . . 113**
3.4.1 Choice of force field . . . . . 113
3.4.2 Effect of using C-$\alpha$ restraints in simulations . . . . . 113
3.4.3 Comparison of results to previous experimental data . . 114
3.4.4 Alternative engineered FhuA nanopores . . . . . 114
3.4.5 Mechanism of effect of mutations on dsRNA translocation through FhuA . . . . . 115
**3.5 Conclusions . . . . . 118**
**3.6 Future directions . . . . . 120**

# Statement of Authorship

All the work in this chapter was conducted by the doctoral candidate.

# 3.1 Introduction

The dynamics of substrate translocation through nanopores are governed by intermolecular interactions. In this chapter, I focus on the electrostatic interactions between a nanopore, FhuA, and double-stranded RNA.

## 3.1.1 Nanopores

A nanopore can be any nanometre-scale hole through an otherwise non-porous membrane and is the only passageway between two separated compartments. Nanopores allow the controlled movement of a substrate between the two compartments. As shown in Figure 3.1, movement of the substrate between the two compartments is dependent on the interactions between the substrate and the pore (Berezhkovskii and Bezrukov, 2023). Protein pores in biological membranes exemplify this, covering a spectrum of diameters appropriate for metabolites that are key for cell function and thus biologically-relevant substrates (Zhang *et al.*, 2022). Commonly these proteins are formed from $\beta$-barrels if the membrane is not energy-transducing (Wiswedel *et al.*, 2024), giving them a stable structure and a constitutively open pore. This opening can be occluded by extra- or intracellular "plugs", which may respond to signals and open or close the pore according to the needs of the cell (Mayse and Movileanu, 2023). Additional extra- and intracellular domains may have other functions for the cell but are extraneous to their function as a nanopore. These proteins exist in membranes and therefore have hydrophobic exteriors and hydrophilic cavities that allow correct, stable folding in the lipid environment (Kleinschmidt, 2015).

### 3.1.1.1 Nanopores in biotechnology

Due to their structural stability and biologically relevant diameter, nanopores have been harnessed for a number of biotechnological applications (Kasianowicz *et al.*,

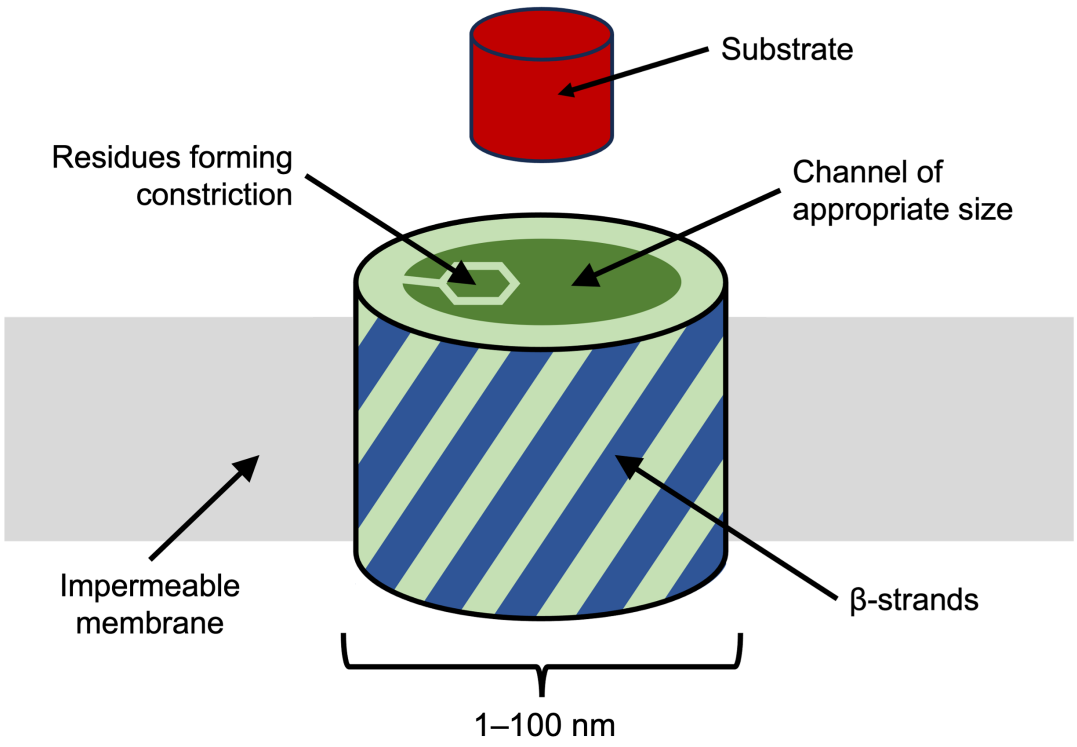



**Figure 3.1:** Example of protein nanopore structure, showing the pore formed of $\beta$-strands inserted into an impermeable membrane. The substrate can only move to the other side of the membrane via the nanopore, where it encounters a constriction point. This constriction makes the movement controlled and predictable.

1996; Maglia *et al.*, 2010; Deamer *et al.*, 2016). The most relevant application to this study is DNA sequencing. DNA is negatively charged due to the anionic nature of each phosphate group, and thus it can be moved through the pore by applying an electric field to the system, where the DNA will move towards the positively charged electrode. The strength of the field will affect the speed at which the DNA will move, with an increase in voltage leading to an increase in current, caused by the faster movement of ions towards the electrodes (Noakes *et al.*, 2019). Larger pores also lead to an increased baseline current due to the increased number of ions that can pass through. The optimal speed of DNA movement through a nanopore for sequencing is 450 bases per second, and most commonly adaptations to the pore need to be made to slow down translocation and make it more controlled and predictable (Wang *et al.*, 2021b; MacKenzie and Argyropoulos, 2023).

Several DNA-sensing nanopores have been developed by different teams across the globe, of which Oxford Nanopore Technologies (ONT) is a well-known example (Bayley, 2015). ONT utilise the bacterial outer membrane transport lipoprotein curlin sigma S-dependent growth subunit G (CsgG, from *Escherichia coli*) in a membrane, across which a −180 mV voltage is applied (Brown and Clarke, 2016). Double-stranded DNA is unwound by an attached DNA helicase, which ratchets a resulting single-stranded DNA (ssDNA) through the pore. DNA bases occlude the pore to different degrees, creating distinct and detectable changes in the electric current carried by the charged ions moving through the pore (Tang *et al.*, 2022). These changes can then be computationally identified to give the sequence of the DNA that has been passed through the pore, as shown in Figure 3.2.

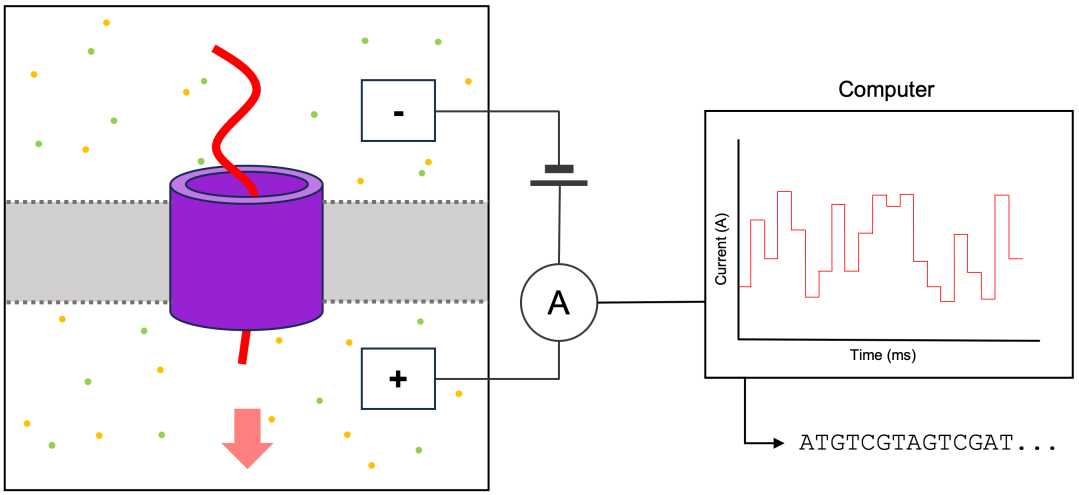



**Figure 3.2:** How nanopores work for DNA sequencing. Ions, shown in green and yellow, flow through the pore and generate a current. The change in current, caused by the single strand of DNA passing through the nanopore, is detected by a computer. It is then interpreted into a sequence, as the fluctuation in current is dependent on the base in the constriction point.

Generally, a number of mutations are made to optimise proteins for use as nanopores. For example, several aspartate residues were mutated to asparagines in the constriction of *Mycobacterium smegmatis* porin A (MspA), which allowed the translocation of single-stranded DNA (Butler *et al.*, 2008). Further mutations

involved replacing negatively charged residues with positively charged residues, underlying the importance of the electrostatic surface of the pore when considering the design principles of nanopores.

The internal diameter of the pore is important for identifying the potential use of the nanopore and is dependent on the structure of the pore. For $\beta$-barrels, proteins with more $\beta$-strands have larger pore diameters (Nanda, 2019). A well-studied nanopore, $\alpha$-hemolysin, has 14 strands making up the $\beta$-barrel. This gives an internal diameter of 14 Å at the narrowest constriction compared to a diameter of 10 Å for single-stranded DNA (Song *et al.*, 1996; Heng *et al.*, 2006), meaning it is a suitable choice for ssDNA translocation (Ashkenasy *et al.*, 2005; Mathé *et al.*, 2005; Muzard *et al.*, 2010; Zhou *et al.*, 2020). CsgG has an internal diameter of 12 Å, and a barrel structure formed from nine copies of a four-strand $\beta$-sheet, although the constriction suitable for DNA probing is formed by the periplasmic eyelet loops (Cao *et al.*, 2014). Another pore used for DNA sequencing is MspA, which also has a constriction of 12 Å (Laszlo *et al.*, 2016).

Nanopores are increasingly being investigated as potential biosensors for other biological molecules, such as detecting post-translational modifications on proteins and looking at RNA molecules (Ying *et al.*, 2022; Shaw *et al.*, 2024). These either utilise the intrinsic charge on the investigated molecules or current caused by the applied charge across the membrane. Nanopores will have to be redesigned and optimised for detecting these biological molecules and their specific chemical properties (Singh, 2023; Wu *et al.*, 2024).

#### 3.1.1.2 Molecular dynamics simulations of nanopores

MD simulations have been used previously to study nanopores, including the translocation of nucleic acids through protein pores (Bond *et al.*, 2011; Choudhary *et al.*, 2022). The field has advanced since the early models using coarse-grained force fields and simple representations of nucleic acids, largely due to the increase in computing resources allowing access to longer timescales at greater molecular resolution (Hollingsworth and Dror, 2018). This has allowed *in silico* investigation at

atomic resolution of a number of currently-used and potential nanopores, developing our understanding and improving their function (Bond *et al.*, 2011; Willems *et al.*, 2020; Rattu *et al.*, 2021a,b).

Specifically, MD simulations have been used to determine design principles for nanopores with varying functions, such as investigating the effect of the position of rings of large hydrophobic residues (Rattu *et al.*, 2021a). It has been shown that MD is a suitable method to approach such questions as it gives atomic-resolution insights into the interactions that complement and are in agreement with experimental data (Zhang and Ding, 2025). MD simulations also allow high-throughput screening of mutations, as mutations can be generated *in silico* and screened for the effect on translocation. This method of investigation gives a better understanding of the molecular basis behind successful design principles.

Simulations of single pores have several differences compared to experimental set-ups. *In vitro*, translocation typically occurs on a millisecond timescale (time between two consecutive bases entering the constriction of the nanopore) with a typical temporal resolution for data collection of 10 $\mu$s (Wang *et al.*, 2021b). The voltage applied in these set-ups is calibrated so that multiple current readings are taken for each base to allow for more accurate sequence determination.

Currently, timescales accessible by most all-atom MD simulations are much lower than tens of microseconds due to the computational cost. This means simulations cannot bridge the gap between two consecutive data points from *in vitro* experiments, and translocation events would not be observed if the voltage used was the same as in experiments. Simulations would therefore be limited in the information they could provide on the molecular detail of translocation. To overcome this limitation, the speed of translocation needs to be increased. Higher electric fields are applied across the membrane to push the substrate in a single direction through the pore at a faster rate (Rattu *et al.*, 2021b). While the voltages used are much higher than those used in experiments, the principles of translocation that are identified in MD agree qualitatively with experimental studies (Bond *et al.*, 2011). This is considered an acceptable compromise (Zhang and Ding, 2025).

### 3.1.2 FhuA

FhuA (ferric hydroxamate uptake component A) is a 22-strand $\beta$-barrel protein located in the outer membrane of *Escherichia coli* and many other gammaproteobacteria. It is part of the import mechanism of ferrichrome as well as being exploited by colicin M and bacteriophage such as T5 for injection of their DNA into the bacterial cell (Bonhivers *et al.*, 1998; Locher *et al.*, 1998; Zeth *et al.*, 2008). Locher *et al.* (1998) solved the structure by X-ray crystallography (PDB 1BY3) in both the liganded and unliganded states, showing the changes that occur when ferrichrome binds and interacts with the energy-transducing TonB system in the bacterial inner membrane. The structure of FhuA is shown in Figure 3.3. FhuA has an N-terminal plug domain, which acts to occlude the pore and prevent the free diffusion of molecules through it, and this plug is affected by ligand binding via the energisation from TonB (Chérier *et al.*, 2021). FhuA also has several large extracellular loops, which are important for interacting with phages and reducing access to the pore (van den Berg *et al.*, 2022). Beyond the context of the cell, these domains are not needed for the protein to function as a nanopore and they need to be removed so they cannot intermittently occlude the pore and prevent predictable behaviour (Endriß and Braun, 2004). Removal of these domains must be done carefully to ensure the pore is assembled correctly when expressed.

Niedzwiecki *et al.* (2012) took FhuA as a potential scaffold for a novel nanopore and successfully engineered and reconstituted it as a channel capable of acting as a nanopore, with its conductance being modified by the addition of organic glycols that could occlude the pore. They created a pore that was very stable over a broad range of experimental conditions including pH and ionic concentration, and had a conductance of 4 n$S$, by removing four extracellular loops and the plug domain as seen in Figure 3.4. Using polymer-exclusion experiments, they compared the size and geometry of the engineered FhuA (termed FhuA $\Delta$C/$\Delta$4L) to the native FhuA. Previous work by Mohammad *et al.* (2011) had found that FhuA $\Delta$C/$\Delta$4L inserted into the membrane in conformations that resulted in a wide range of conductances. Therefore Niedzwiecki *et al.* (2012) selected pores with a narrow

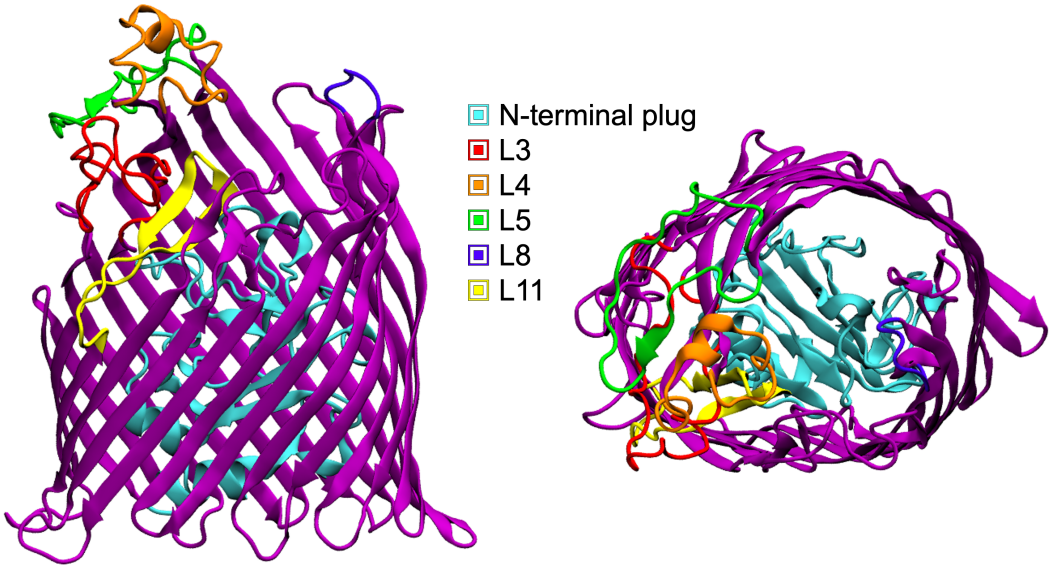



**Figure 3.3:** Diagram showing the crystal structure of FhuA, from 1BY3 (Locher *et al.*, 1998). The N-terminal plug is shown in blue. Loops important in ferrichrome binding or phage binding are highlighted. L3, L4, L5, and L11 are important in ferrichrome binding; L8 is important for binding bacteriophage T5; L4, L5, and L8 are important for binding bacteriophage T1; L3, L4, L5, L8, and L11 are important for binding bacteriophage phi80. Binding data from Lewis *et al.* (2023).

range of conductances for their analysis of single-particle data. The interaction of water-soluble, flexible poly(ethylene glycol)s demonstrated that the inner dimensions and the geometry of the FhuA ΔC/Δ4L pore agreed with those predicted from the crystal structure of the native protein. This suggested that the removal of the plug and loops had not had a large impact on the structure, and FhuA ΔC/Δ4L was relatively stiff. This structure was further characterised to show it was stable under the application of ± 0.1 V transmembrane potentials. Due to the large internal diameter identified by Niedzwiecki *et al.* (2012), I theorised that FhuA ΔC/Δ4L may be suitable for the passage of double-stranded RNA (dsRNA).

### 3.1.3 Double-stranded RNA

RNA has a ribose 2'-OH group that is not present in DNA. The presence of this group allows for the adoption of complex 3D structures by RNA that are not possible by DNA molecules as it provides an additional H-bond donor and acceptor (Leontis

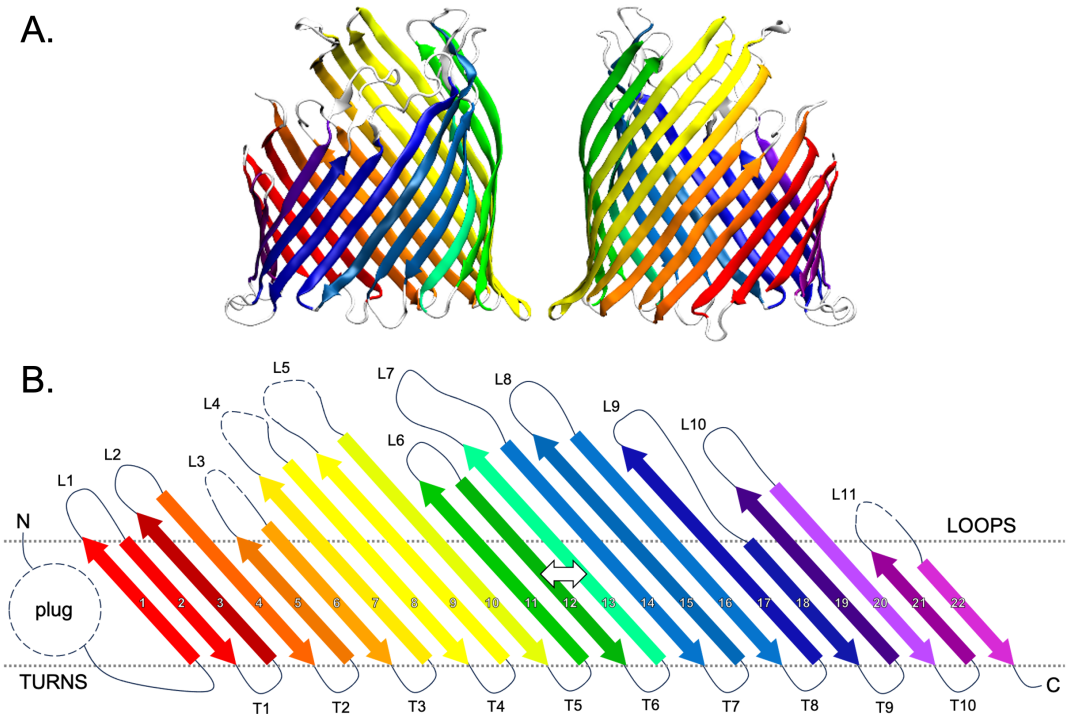



**Figure 3.4:** A. Crystal structure of FhuA from 1BY3, with removed plug and loops in white, and strands coloured from red (N-terminus) to pink (C-terminus).
B. Diagrammatic structure of FhuA showing where the plug and loops were removed (dotted lines), and the relative lengths of each $\beta$-strand. Strands labelled 1–22 and coloured as in A. Double-headed arrow in white indicates a potential opening of the $\beta$-barrel to the outer leaflet of the membrane.

*et al.*, 2002b). Additionally, RNA molecules utilise uracil bases instead of thymine. While uracil has less efficient base-stacking compared to thymine, due to the lack of methyl group on the C5 position of the base, the hydrogen bonds between uracil and adenine are stronger than those between thymine and adenine (Swart *et al.*, 2004; Vakonakis and LiWang, 2004; Pérez *et al.*, 2005; Svozil *et al.*, 2010).

Most of the RNA in a cell is ribosomal RNA (rRNA), with less than 5% of the total content being unfolded messenger RNA. rRNAs adopt complex, specific folds that are key to their function (Tinoco and Bustamante, 1999; Zemora and Waldsich, 2010). Base-pairing enables the folding of RNA molecules and helps give the RNA structural stability, including some protection from native exonucleases (Hayne *et al.*, 2023). The fold is also important for the RNA to be identified as "self" by the innate immune system of the cell and not targeted for specific

degradation (Hull and Bevilacqua, 2016; Uehata and Takeuchi, 2020; Nguyen *et al.*, 2023). Currently these folds are not currently well-predicted, and therefore more experimental data on RNA folds (both structural and bioinformatic) will improve our understanding of RNA folding and the potential for accurate predictions (Schroeder, 2018; Nithin *et al.*, 2024; Kwon, 2025).

RNA can base-pair in the same way as DNA using the Watson-Crick (WC) edges of the bases: guanine forming three hydrogen bonds with cytosine and adenine forming two hydrogen bonds with uracil (instead of thymine in DNA). These WC base pairs are the "canonical" interactions, along with the "wobble" pair of guanine and uracil involving the WC edges and the formation of two hydrogen bonds (Mokdad *et al.*, 2006; Stombaugh *et al.*, 2009). Additionally, RNA can readily form "non-canonical" base pairs as shown in Figure 3.5 (Westhof and Fritsch, 2000; Leontis and Westhof, 2001; Lemieux and Major, 2002; Stombaugh *et al.*, 2009). Other non-canonical base pairs utilise the other edges of the base, such as guanine-adenine and adenine-adenine pairs that form between the Hoogsteen edge of one base and the sugar edge of the other (Leontis *et al.*, 2002a; Hoogsteen, 1963). Calculating the strength of these non-canonical interactions is non-trivial, but research has shown that the average energy of hydrogen-bonding for the non-canonical interactions is less than the energy of a WC guanine-cytosine base pair (Halder *et al.*, 2019).

Due to the promiscuity of RNA base-pairing, RNA can form secondary structures beyond the archetypal double helices of WC duplexes. These other structures are known as "loops", and can be further divided into "hairpins", "bulges", and "junctions" as shown in Figure 3.5 (Tinoco and Bustamante, 1999; Hermann and Patel, 2000). These are usually energetically stable because of the non-canonical interactions, and do not require additional factors such as the magnesium ions often required for tertiary fold stability.

#### 3.1.3.1 3D structure of rRNA

The ribosome is a well-studied structure containing a lot of dsRNA (Noller, 2005). The structure of the large subunit of the ribosome from *Nitrosopumilus maritimus*

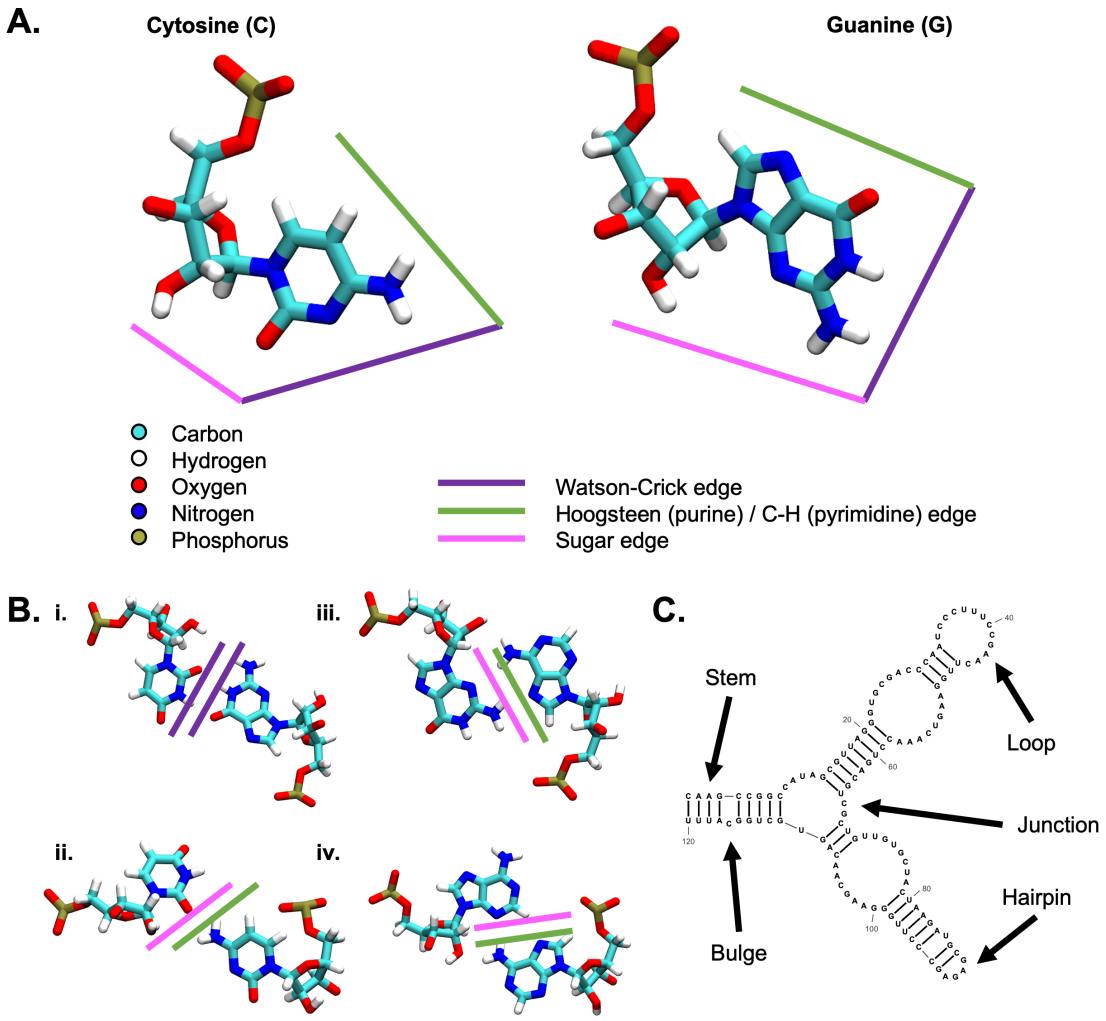



**Figure 3.5:** Non-canonical base pairs and common RNA folds *in vivo*.
A. Annotated purine (cytosine) and pyrimidine (guanine) interacting edges. Adapted from Leontis and Westhof (2001).
B. RNA base pairs. i. is a G-U wobble pair between WC edges; ii., iii., and iv. are U-C, G-A, and A-A respectively, between the Hoogsteen and sugar edges.
C. 2D stem-and-loop representation of 5S rRNA from *Nitrosopumilus maritimus*, highlighting the structural features (see Chapter 4, subsection 4.3.5 and Figure 4.34, for 3D structure).

is shown in Figure 3.6 to illustrate the high proportion of rRNA (for more detail, see Chapter 4). The formation of the peptide-peptide bond during protein synthesis is catalysed by the RNA component of the ribosome and not the proteins. Therefore the secondary and tertiary structure of the RNA is paramount to the ribosome performing its functions correctly, by exposing the catalytic nitrogen on an adenine residue (Nissen *et al.*, 2000; Świderek *et al.*, 2015).

The correct location of RNA sequences is a challenging aspect of building models into structures of ribosomes and other riboproteins (Bonilla and Jang, 2024). Without extremely high-resolution data (around 2 Å), it can be difficult to unambiguously distinguish between each pyrimidine base and each purine base. Additionally, while in DNA structures the assignment of one strand of a double-stranded structure can predict the sequence of the other strand with high confidence, RNA has a high percentage (currently estimated to be around 40%) of non-WC base pairs (Leontis and Westhof, 2001; Lemieux and Major, 2002; Olson *et al.*, 2009). The variety of other occurring base pairs make the unambiguous assignment more time-consuming since there is little complementary data to assist. Therefore, methods that can increase the amount of data we have on potential conformations of dsRNA will improve RNA modelling in the future, which in turn will help structural and functional understanding (Šponer *et al.*, 2010; Li *et al.*, 2024; Wang *et al.*, 2023a).

Due to the highly-negatively charged backbone of RNA, cations are required to counter the anionic phosphate moieties. In cells, this is usually a role played by magnesium ions, $Mg^{2+}$ (Guth-Metzler *et al.*, 2023). However, RNA structures in solution can be successfully coordinated by other cations such as potassium, $K^{+}$ (Grover, 2015).

#### 3.1.3.2 Molecular dynamics simulations of RNA

RNA is one of the least well-studied of all the biological macromolecules with regard to MD simulations (Kührová *et al.*, 2016; Mlýnský and Bussi, 2018; Šponer *et al.*, 2018). Due to a relative lack of structural and biochemical data on the behaviour of RNA, the current force fields describing RNA are not well-developed

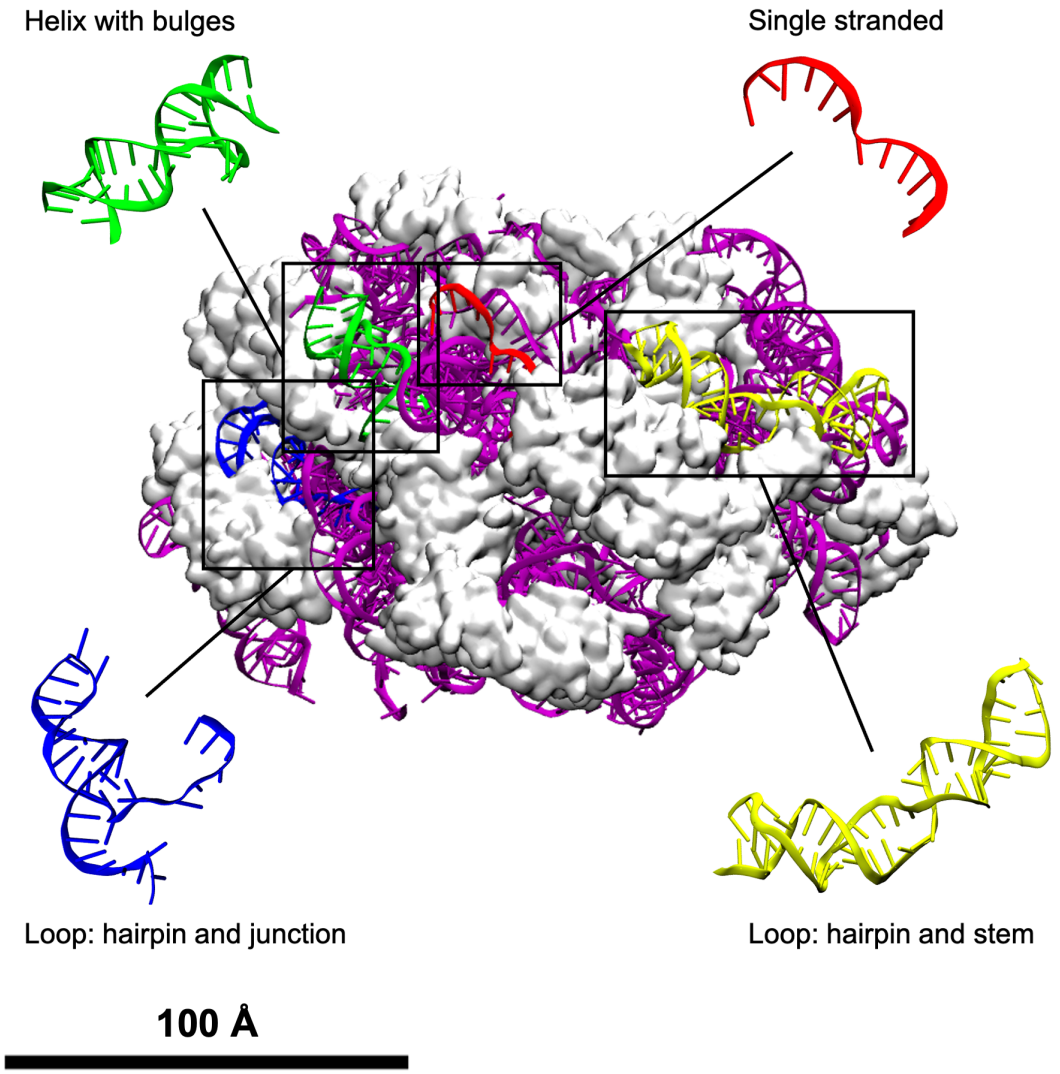



**Figure 3.6:** Structure of large ribosomal subunit from *N. maritimus*. rRNA shown in purple, with fragments of different secondary structures and their original location in the 50S subunit. The fragment in green, helix with bulges, is used in this chapter. Riboproteins shown in white. Further detail in Chapter 4, Section 4.3.

in comparison to DNA, and particularly in comparison to lipids and proteins (Choi *et al.*, 2024). In recent years, quantum mechanics simulations have improved the performance of RNA force fields, however the behaviour of some specific RNA structures such as G-quadruplexes remain challenging to model (Languin-Cattoën and Bussi, 2026). Force fields for RNA have often been developed for specific-use cases, or for RNA-only simulations (Vangaveti *et al.*, 2017; Chen *et al.*, 2022). Therefore, simulations of RNA must be analysed with this in mind. Currently,

the most popular, universally applicable RNA force fields for all-atom simulations are CHARMM36 and AMBER OL3, with the latter most commonly used (Chen *et al.*, 2022; Huang *et al.*, 2017). Both were assessed to be of similar quality in recent reviews, with differing strengths and weaknesses based on the application (Liebl and Zacharias, 2023; Knappeová *et al.*, 2024). Therefore, CHARMM36 was used for direct comparison to previous nanopore simulations with DNA utilising the same methodology (Rattu *et al.*, 2021a).

RNA conformational changes range from base repuckering, which occurs on a nanosecond timescale, to conformational switching on a timescale of seconds (Al-Hashimi and Walter, 2008; Rinnenthal *et al.*, 2011). MD simulations using all-atom force fields can access the timescale of base repuckering, up to some base pair opening when simulations are extended to microseconds. Obtaining simulations on longer timescales with all-atom MD requires the use of dedicated hardware such as Anton (Shaw *et al.*, 2008), massive parallel approaches such as folding@home (Shirts and Pande, 2000; Pande *et al.*, 2003), or enhanced sampling methods such as metadynamics and non-equilibrium MD simulations (Abrams and Bussi, 2014; Dellago and Hummer, 2014; Bernardi *et al.*, 2015). Such approaches for obtaining longer timescales have not been used in the work presented here.

### 3.1.4 Aim of this chapter

Currently, the lack of structural information on RNA is limiting the power of machine learning methods to predict RNA 3D structures, due to a lack of training data. The promiscuity of base pairing in RNA and the number of possible structures means that it is difficult to work out where fragments may associate with each other, and there is no equivalent to AlphaFold or RoseTTAFold that successfully produces RNA 3D structures. If a nanopore that translocated dsRNA could read out the order of base pairs in a segment of dsRNA, along with the length of unpaired segments and other features, this would provide a large amount of contextual data that would then allow the construction of a 3D model prediction for that specific RNA.

This chapter investigates the use of FhuA ΔC/Δ4L as a nanopore for the electric field-induced translocation of dsRNA, using a fragment from the author's unpublished ribosome structure (detailed in Chapter 4). There has previously been no published studies on simulations of double-stranded nucleotides through protein nanopores, and this work is also unique for taking the approach of investigating a pore specifically for double-stranded RNA rather than single-stranded RNA or DNA. The interactions between charged protein residues that line the pore and the RNA are investigated by mutating acidic, basic, or acidic and basic residues to alanines and measuring potential changes in RNA translocation. As RNA has one negative charge for every phosphate group and thus every nucleotide, I hypothesise that mutating the charged residues lining the pore interior will affect the translocation dynamics of the RNA (just as it affects the movement of charged ions), because the distribution of electrostatic attractions and repulsions will be impacted. This will help elucidate the mechanism of dsRNA movement through FhuA ΔC/Δ4L and the effect of mutations on residence time in the nanopore. Moreover, it will contribute to the body of work on the underlying design principles when using nanopores to detect biological macromolecules, with the overall aim of having controlled, predictable translocation of the macromolecule through the pore.

## 3.2 Materials and Methods

### 3.2.1 FhuA model

The model of the engineered FhuA nanopore was constructed in PyMOL (PyMOL, 2015) by taking a 2.7 Å X-ray crystallography model structure from the PDB (1BY3 (Locher *et al.*, 1998)) and removing residues to recreate the experimentally successful variant of Niedzwiecki *et al.* (2012), FhuA ΔC/Δ4L. Henceforth, this variant will be referred to as WT FhuA, with the wild-type amino acid sequence but including the designed truncations. This WT FhuA model was also later mutated *in silico* to change all the pore-lining aspartic acid, glutamic acid, lysine, and arginine residues to alanine, using the PyMOL mutagenesis wizard. A total of 22 negatively charged residues were chosen for mutation, and 19 positively charged residues. The models of FhuA ΔC/Δ4L were inserted into a palmitoyl-oleoyl-phosphatidylcholine (POPC) bilayer using CHARMMGUI (Jo *et al.*, 2007, 2008; Brooks *et al.*, 2009; Lee *et al.*, 2016).

Table 3.1 and Table 3.2 detail the mutations made to the FhuA ΔC/Δ4L pore, with the resulting pores henceforth referred to as WT, DA/EA, KA/EA, and DA/EA/KA/RA. Figure 3.7 shows the locations of these mutations on the pore structure.

**Table 3.1:** Modifications made to FhuA . "Residues removed" applies to all modified forms of FhuA. A total of 22 negatively charged residues were chosen for mutation, and 19 positively charged residues.

| | |
|---|---|
| **Residues removed** | 1–160, 243–273, 318–339, 394–419, 682–704 |
| **ASP to ALA** | 179, 287, 348, 358, 371, 379, 454, 469, 473, 619, 659, 665 |
| **GLU to ALA** | 163, 210, 238, 279, 289, 350, 501, 522, 571, 573 |
| **LYS to ALA** | 167, 206, 364, 526, 545, 575 |
| **ARG to ALA** | 193, 199, 212, 222, 277, 297, 301, 452, 471, 479, 637, 663, 713 |

#### 3.2.1.1 Calculating $xy$ cross-sectional area

The $xy$ cross-sectional area of the WT FhuA pore along the $z$-axis was calculated using HOLE (Smart *et al.*, 1996) and the extension PoreAnalyser (Seiferth and

**Table 3.2:** Mutations from Table 3.1 applied to each system as named in the first column.

| | ASP to ALA | GLU to ALA | LYS to ALA | ARG to ALA |
|---|---|---|---|---|
| **WT** | No | No | No | No |
| **DA/EA** | Yes | Yes | No | No |
| **KA/RA** | No | No | Yes | Yes |
| **DA/EA/KA/RA** | Yes | Yes | Yes | Yes |

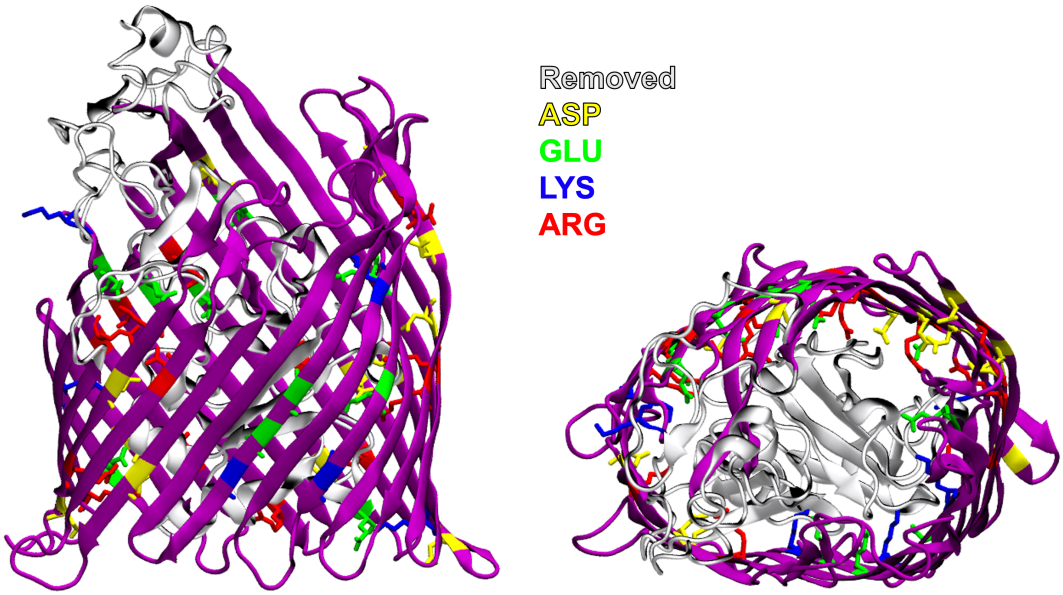



**Figure 3.7:** FhuA with removed residues (in white) taken from the design of Niedzwiecki *et al.* (2012), with residues mutated to alanine within this research highlighted by colour.

Biggin, 2024). VMD (Humphrey *et al.*, 1996) was used to obtain the large and small radii needed to calculate an elliptical area of a base pair of dsRNA, accounting for the van der Waals radius specified in the CHARMM36m force field. The largest $\sigma$ value was used to calculate the radius from the Lennard-Jones minimum distance of the atoms in nucleic acids (0.405 for CN7) to ensure the largest atoms were counted (Equation 2.1.2.10). This gives a Lennard-Jones minimum of 4.54 Å, and a radius for each atom of 2.27 Å.

#### 3.2.1.2 Applied restraints

Positional restraints were applied to FhuA for the duration of all the simulations used for analysis unless specified, because of the flexibility of the pore (see subsection 3.3.3,

also observed in Rodríguez-Ropero and Fioroni (2012); Li *et al.* (2021)). A positional restraint of 1000 kJ $mol^{-1}$ $nm^{-2}$ was applied to each C-$\alpha$ atom in the protein, in a conformation where the remaining extracellular loops of the pore were more open, to allow the dsRNA sufficient space for translocation. This conformation was chosen after simulating FhuA in a membrane (both with and without an applied electric field) and visualising the flexibility of the extended $\beta$-strands in the outer portion of the protein. For simulations where the RNA was being restrained artificially, position restraints of 1000 kJ $mol^{-1}$ $nm^{-2}$ were applied to every atom in residue 572 and 622 of the dsRNA molecule.

### 3.2.2 Calculating ion flux through non-occluded FhuA

Flux of ions through the WT FhuA or one of the mutated pores was investigated by varying the applied electric field strength between 0 and $-0.1$ V $nm^{-1}$ in a 15 nm x 15 nm x 16 nm box as shown in Table 3.3. An electric field is defined as the direction in which positive ions flow; as the potassium ions flow towards the bottom of the box, where $z = 0$, the electric field is negative.

Electric field strengths of $-0.03$ and $-0.05$ V $nm^{-1}$ were chosen for further experiments as the pore and membrane did not suffer any instances of deformation from the subsequent charge gradient and allowed RNA translocation on a computationally viable timescale. This gives a total applied voltage of $-0.57$ or $-0.95$ V respectively, compared to $-0.18$ V used by Oxford Nanopore Technologies and other biological nanopores for nucleotide sequencing (Noakes *et al.*, 2019).

### 3.2.3 Double-stranded RNA model

A dsRNA fragment was excised from the author's unpublished cryo-EM ribosome structure, as shown in Figure 3.6 (see Chapter 4). The sequence, structure, and dot-bracket notation are shown in Figure 3.8 (Hofacker *et al.*, 1994; Ramlan and Zauner, 2008). The structure was energy minimised and equilibrated in a $100^3$ $Å^3$ water box with 0.15 M KCl or neutralising $K^+$. $K^+$ was chosen preferentially over $Mg^{2+}$ as force field parameters for $K^+$ are significantly more accurate, and

**Table 3.3:** Simulation set-ups for simulations of FhuA without dsRNA present.

| **Pore** | **Restraints** | **Replicas** | **Electric field ($V\ nm^{-1}$)** | **Box $z$-dimension (nm)** | **Electric field (total, V)** |
|---|---|---|---|---|---|
| WT | none | 3 | 0 | 16 | 0 |
| WT | none | 3 | −0.01 | 16 | −0.16 |
| WT | none | 3 | −0.025 | 16 | −0.4 |
| WT | none | 3 | −0.05 | 16 | −0.8 |
| WT | none | 3 | −0.075 | 16 | −1.2 |
| WT | none | 3 | −0.1 | 16 | −1.6 |
| WT | C-$\alpha$s | 3 | 0 | 16 | 0 |
| WT | C-$\alpha$s | 3 | −0.01 | 16 | −0.16 |
| WT | C-$\alpha$s | 3 | −0.025 | 16 | −0.4 |
| WT | C-$\alpha$s | 3 | −0.05 | 16 | −0.8 |
| WT | C-$\alpha$s | 3 | −0.075 | 16 | −1.2 |
| WT | C-$\alpha$s | 3 | −0.1 | 16 | −1.6 |
| DA/EA | none | 3 | 0 | 16 | 0 |
| DA/EA | none | 3 | −0.01 | 16 | −0.16 |
| DA/EA | none | 3 | −0.025 | 16 | −0.4 |
| DA/EA | none | 3 | −0.05 | 16 | −0.8 |
| DA/EA | none | 3 | −0.075 | 16 | −1.2 |
| DA/EA | none | 3 | −0.1 | 16 | −1.6 |
| DA/EA | C-$\alpha$s | 3 | 0 | 16 | 0 |
| DA/EA | C-$\alpha$s | 3 | −0.01 | 16 | −0.16 |
| DA/EA | C-$\alpha$s | 3 | −0.025 | 16 | −0.4 |
| DA/EA | C-$\alpha$s | 3 | −0.05 | 16 | −0.8 |
| DA/EA | C-$\alpha$s | 3 | −0.075 | 16 | −1.2 |
| DA/EA | C-$\alpha$s | 3 | −0.1 | 16 | −1.6 |
| KA/RA | C-$\alpha$s | 3 | −0.025 | 17.5 | −0.44 |
| KA/RA | C-$\alpha$s | 3 | −0.075 | 17.5 | −1.32 |
| DA/EA/KA/RA | C-$\alpha$s | 3 | −0.025 | 17.5 | −0.44 |
| DA/EA/KA/RA | C-$\alpha$s | 3 | −0.075 | 17.5 | −1.32 |

$K^+$ has been used experimentally for nanopores (Grotz *et al.*, 2021; MacKenzie and Argyropoulos, 2023).

### 3.2.4 Simulation set-up

The dsRNA was placed manually with 2–3 bases entering the pore, twice with the RNA entering from the intracellular side of the pore (at the bottom of the simulation box) and once with the dsRNA entering from the extracellular side. A potassium chloride (KCl) concentration of 0.15 M was chosen for all the simulations as this is physiologically appropriate (Stautz *et al.*, 2021), and within the range

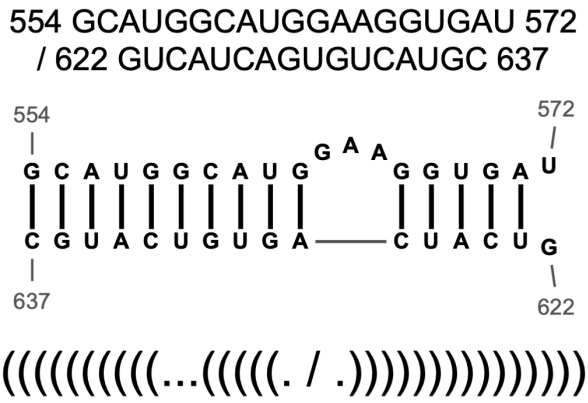



**Figure 3.8:** Sequence, stem-and-loop structure, and dot-bracket notation of dsRNA fragment used in simulations, taken from the ribosome of *N. maritimus*.

of successful experimental concentrations and previous simulations (Bond *et al.*, 2011; Basdevant *et al.*, 2019), and the TIP3P water model was used (Berendsen *et al.*, 1981; Price and Brooks, 2004).

Simulations were performed using the GROMACS 2022.1 engine (Bekker *et al.*, 1993; Berendsen *et al.*, 1995; Lindahl *et al.*, 2001; Van Der Spoel *et al.*, 2005; Hess *et al.*, 2008; Pronk *et al.*, 2013; Abraham *et al.*, 2015; Szilárd *et al.*, 2015). The all-atom force field CHARMM36m was chosen because of the agreement of simulation data with experimental results, and the wide body of previous work to which the results can be compared (MacKerell Jr, 2006; Best *et al.*, 2012; Huang *et al.*, 2017). The systems were equilibrated in stages as described in Chapter 2, subsubsection 2.1.4.7, to ensure temperature, pressure, and overall energy were stabilised. Six different RNA starting conformations were then generated for each of the three set-ups. The protein and lipids were restrained, and the RNA and solvent were allowed to move freely. Simulations were performed at 330 K for 10 ns, and the first, last, and four intermediary frames were taken as the starting points for further simulations.

The slowest part of RNA translocation in simulation was identified as the initial movement of the RNA through the pore. Preliminary simulations were carried out to rapidly "thread" the RNA in the pore, so that the first residues

of the RNA were level with the headgroups of the top leaflet of the bilayer at the start of all analysed production runs (see Figure 3.9). This was achieved by applying an electric field with a strength of $-0.075$ V nm$^{-1}$ across the simulation box (in the $z$-direction) for the WT FhuA, and $-0.05$ V nm$^{-1}$ for the DA/EA and DA/EA/KA/RA mutants. An appropriate frame from each resultant simulation was chosen by visual inspection to be the seed for the production simulation. In each case, the desired amount of "threading" was achieved within 50 ns. Translocation did not occur in any of the simulations with the KA/RA mutant.

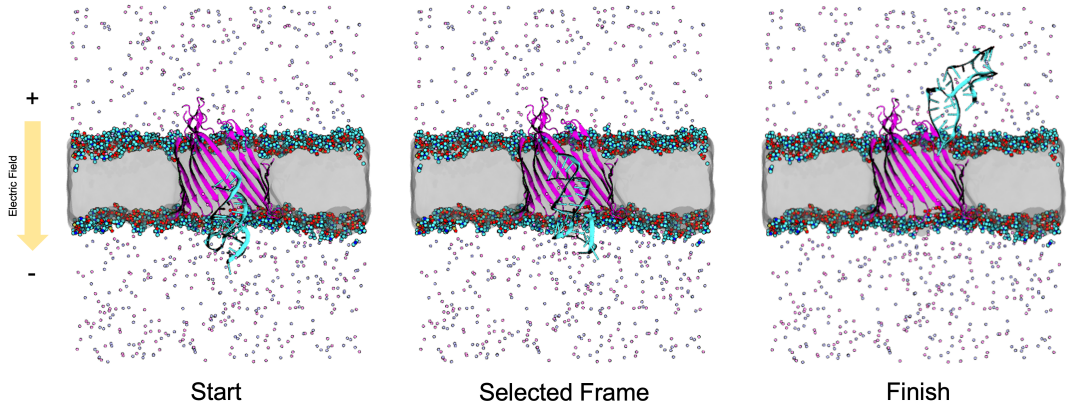



**Figure 3.9:** Example of how a frame was chosen for starting production simulations, from a simulation for "threading" the dsRNA through the pore by applying a higher electric field strength.

Production simulations were 50 ns with an integration step size of 2 fs. Temperature was maintained at 310 K by the V-rescale thermostat (Bussi *et al.*, 2007), and semi-isotropic pressure maintained at 1 bar by the Parrinello-Rahman barostat (Parrinello and Rahman, 1981). Non-bonded interactions were cut-off at 1.2 nm. Electrostatics were treated using the particle mesh Ewald method (Darden *et al.*, 1993). Hydrogen bonds and constraints were managed by the LINCS algorithm (Hess *et al.*, 1998). A uniform, static electric field of $-0.05$ V nm$^{-1}$ or $-0.03$ V nm$^{-1}$ was applied in the $z$-direction for the duration of all production simulations, coming to an overall electric field strength of $\sim -1$ V or $\sim -0.6$ V. A summary of production simulations is provided in Table 3.4. Position-restrained simulations were run for 500 ns.

**Table 3.4:** Simulation set-ups for simulations of FhuA with dsRNA, detailing the variation of electric field. KA/RA pores did not successfully thread with dsRNA, so no production runs were performed.

| **Pore** | **Replicas** | **Box $z$-dimension (nm)** | **Threading electric field (V nm$^{-1}$)** | **Production electric field (V nm$^{-1}$)** | **Production electric field (total, V)** |
|---|---|---|---|---|---|
| WT | 12 | 19 | −0.075 | −0.05 | −0.95 |
| WT | 12 | 19 | −0.075 | −0.03 | −0.57 |
| WT | 6 | 19 | −0.075 | 0.05 | 0.95 |
| WT | 6 | 19 | −0.075 | 0.03 | 0.57 |
| DA/EA | 12 | 19 | −0.05 | −0.05 | −0.95 |
| DA/EA | 12 | 19 | −0.05 | −0.03 | −0.57 |
| DA/EA | 6 | 19 | 0.05 | 0.05 | 0.95 |
| DA/EA | 6 | 19 | 0.05 | 0.03 | 0.57 |
| KA/RA | 6 | 20 | −0.05 | n/a | n/a |
| KA/RA | 6 | 20 | −0.05 | n/a | n/a |
| DA/EA/KA/RA | 6 | 20 | −0.05 | -0.05 | −1.00 |
| DA/EA/KA/RA | 6 | 20 | −0.05 | -0.03 | −0.60 |

### 3.2.5 Analysis

Analysis was performed using GROMACS tools and the MDAnalysis module for Python (Theobald, 2005; Liu *et al.*, 2010; Michaud-Agrawal *et al.*, 2011; Gowers *et al.*, 2016). Other Python packages used were NumPy, Pandas, MatPlotLib, and SciPy (Hunter, 2007; McKinney, 2011; van der Walt *et al.*, 2011; Virtanen *et al.*, 2020). The tools utilised from the GROMACS suite were gmx sasa, gmx covar, and gmx anaeig (Eisenhaber *et al.*, 1995). Visualisation was performed in VMD (Humphrey *et al.*, 1996). Electrostatic surface calculations were performed with the ABPS package for PyMOL (PyMOL, 2015; Jurrus *et al.*, 2018). Python scripts were developed by the author specifically for this project. Custom contact analysis scripts used a cutoff of 5 Å. When determining the importance of contacts, residues were identified that were in contact with the RNA for more than 20% of the simulation where RNA was threaded in the pore. Pi-stacking was identified using the values in Figure 3.10 (Hunter and Sanders, 1990). Calculations of average ion currents were performed by combining instantaneous currents from Equation 2.1.6.3, following the protocol of Göpfrich *et al.* (2016).

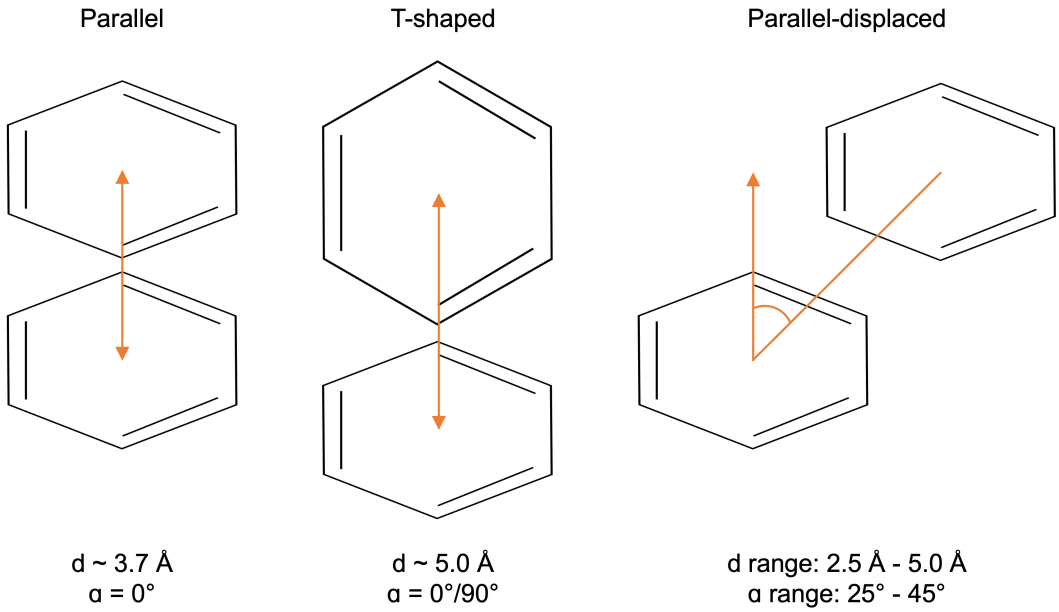



**Figure 3.10:** Parameters used for determining the presence of pi-stacking, from Hunter and Sanders (1990).

## 3.3 Results

### 3.3.1 FhuA has an appropriate pore for dsRNA translocation

Preliminary analyses of the $xy$ cross-sectional area of both the dsRNA fragment and the internal pore of FhuA ΔC/Δ4L (WT FhuA) were performed before running simulations. These were necessary to check that FhuA was an appropriate choice for translocating dsRNA.

#### 3.3.1.1 Cross-sectional area of dsRNA

Previous work suggested that dsRNA has a diameter of 23 Å and thus an $xy$ cross-sectional area of 415 Å$^2$, assuming that the RNA is a cylinder (Hur, 2019; Laurent and Carlson, 2023). Using the VdW radii from the CHARMM36 force field, the model dsRNA was shown to have an average diameter of 22 Å, giving an $xy$ cross-sectional area of 400 Å$^2$. However, in the pore the dsRNA is not necessarily in a cylindrical conformation, and it proved more useful to consider the area of an RNA base pair individually. For this section of RNA, this gave an average $xy$ cross-sectional area of 210 Å$^2$ and a maximum of 360 Å$^2$ where there are non-canonical interactions.

#### 3.3.1.2 Cross-sectional area of WT FhuA pore

Niedzwiecki *et al.* (2012) estimated the internal diameter of the WT FhuA pore to be an average of 24 Å, with a minimum constriction of 10 Å at the extracellular entrance, by using impermeable dextran polymers to restrict nanopore access. These calculations agree with the predictions derived by Locher *et al.* (1998), from their crystal structure (1BY3), which suggested a conical pore going from a 12 Å opening on the extracellular side and a 28 Å opening on the intracellular side. Their results gave an approximate average $xy$ cross-sectional area of 500 Å$^2$ with a maximum of 600 and minimum of 300 Å$^2$, assuming the pore is circular. As the pore is not circular due to the large number of $\beta$-strands that form FhuA (allowing for a greater deformation from circular), these values are not likely to be accurate.

Calculations of pore $xy$ cross-sectional area were performed independently for the work presented here. These were done initially using the program HOLE, which functions by determining the largest possible sphere that fits in the calculated pore pathway (Smart *et al.*, 1996). In the case of WT FhuA, which is a non-perfect cylinder, this underestimated the $xy$ cross-sectional area of the inside of the pore. HOLE predicted a constriction with a radius of 5.6 Å, giving an $xy$ cross-sectional area of 100 Å$^2$. This is smaller than that required even for a single base of dsRNA and would suggest that WT FhuA is a non-viable pore choice for dsRNA translocation. However, by eye, it was judged that WT FhuA should be large enough for dsRNA to fit through. Therefore, PoreAnalyser was used, which does not assume the pore is a perfect cylinder (Seiferth and Biggin, 2024).

PoreAnalyser calculates the largest ellipsoids that can fit along the pore pathway and produces a more accurate calculation of the $xy$ cross-sectional area of the pore. It is worth noting that it is still likely to be an underestimate, but it is a much closer representation of what is accessible to the dsRNA. This is because dsRNA does not have as much ability to access small crevices in the pore compared to water and ions. First, PoreAnalyser was used to analyse the crystal structure, which again showed a constriction smaller than the $xy$ cross-sectional area of dsRNA at around 150 Å$^2$. As this constriction was around the position in the pore where the extracellular loops collapse inwards, I hypothesised that this constriction would not exist in simulation when the pore was relaxed. Therefore, a short simulation was performed to generate a more open structure of the extracellular loops. HOLE and PoreAnalyser analyses were then repeated for comparison. This showed that even at the narrowest constriction WT FhuA is a viable pore choice for dsRNA translocation (see Figure 3.11). This result demonstrates the importance of considering a dynamic structure that is not affected by crystal packing.

The above analysis was also performed on the DA/EA mutant pore to check that mutating some of the residues to alanines did not have a significant impact on the internal diameter; the results are shown in Figure 3.11. This confirmed that WT FhuA was a good choice of nanopore for dsRNA, as the opening was

wide enough for dsRNA to pass through but also narrow enough that ion flux would be constricted when dsRNA was in the pore. This was intended to enable a measurable difference in the current across the pore.

Theoretically, dsDNA would also be able to translocate through WT FhuA as it has a diameter of 20 Å, and therefore an $xy$ cross-sectional area of 315 Å$^2$. Since dsDNA is 25% smaller in area, it would be likely to have a lower effect on restricting ion flow through the pore.

Additionally, a convex hull was calculated for an alternative method of assessing WT FhuA and dsRNA $xy$ cross-sectional area (see Appendix Section A.1).

### 3.3.2 Conformational behaviour of dsRNA in MD simulations

Equilibrium simulations were performed to assess the behaviour of the dsRNA fragment in water using the CHARMM36 all-atom force field. The simulations (analysis shown in Figure 3.12) demonstrated that the secondary structures remained stable under physiological (0.15 M KCl) and low salt (neutralising $K^+$ only) conditions. Data for simulations with other KCl concentrations is in Appendix Section A.2. For example, plots of RMSD showed that the fragment behaved similarly in both conditions (physiological: 3.51 ± 0.47 Å; neutralising: 3.78 ± 0.07 Å). RMSF showed that unpaired bases had much more flexibility, as well as the ends of the fragments. Plots of end-to-end distance of the fragment showed that the extension of each fragment remained within 5% of the mean length throughout the simulation. SASA showed that structure compressed slightly but not significantly over the course of the simulations in both physiological (68.1 ± 1.1 Å$^2$ to 66.0 ± 1.3 Å$^2$) and low (67.9 ± 1.0 Å$^2$ to 66.5 ± 1.3 Å$^2$) salt, comparing the first and last 10 ns. As expected, more $K^+$ ions are associated with the backbone in the physiological salt conditions (10.8 ± 2.4 compared to 8.3 ± 2.1), as far fewer were available at neutralising conditions. Analysis of base-pairing showed that > 90% of the hydrogen bonds between the base pairs were maintained for the course of the simulation (see Appendix Section A.3, Figure A.3).

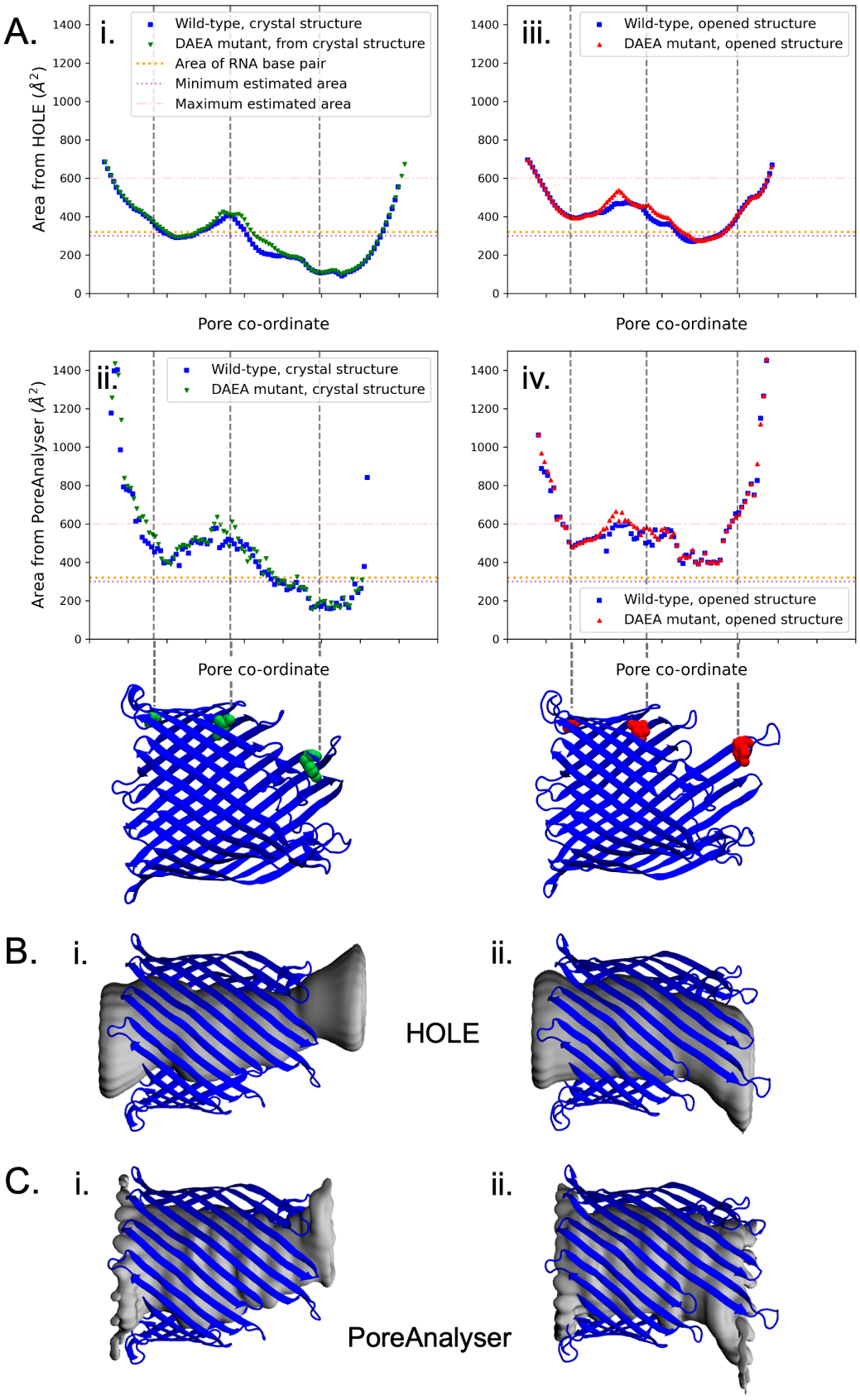



**Figure 3.11:** *(Caption on next page.)*

**Figure 3.11:** A. Cross-sectional area, in $xy$, of FhuA $\Delta$C/$\Delta$4L pore as estimated by HOLE (A. i. WT; A. ii. DA/EA mutant) and PoreAnalyser (A. iii. WT; A. iv. DA/EA mutant). Reference residues for pore co-ordinate are shown below the graphs, on both the crystal structure (green residues) and the opened structure (red residues) of the pore. B. Pathways of the $xy$ cross-sectional area as calculated by HOLE in WT (B. i.) and DA/EA mutant (B. ii.) FhuA pore. Grey surface shows the calculated path from HOLE. C. Pathways of the $xy$ cross-sectional area as calculated by PoreAnalyser in WT (C. i.) and DA/EA mutant (C. ii.) FhuA pore. Grey surface shows the calculated path from PoreAnalyser.

Equilibrium simulations of the other dsRNA fragments shown in Figure 3.6 were performed, with data shown in Appendix Figure A.4.

Overall, this data showed that the dsRNA was stable in MD simulations performed with GROMACS and the CHARMM36 force field over 50 ns, which is the length of later simulations of dsRNA translocation. These results show a plateau in metrics such as RMSD, and agree with previous assessments that the CHARMM36 RNA force field is stable on the scale of tens of nanoseconds (Šponer *et al.*, 2010). Generally, *in vitro* experimental data on RNA dynamics is limited due to the difficulty of obtaining dynamic data, so atomistic MD simulations are an important source of information to exceed the size limitations of NMR and the difficulties associated with producing and purifying RNA fragments (Al-Hashimi and Walter, 2008; Rinnenthal *et al.*, 2011; Deka, 2023).

One simulation of dsRNA in 0.15 M KCl was extended to 1 µs. The behaviour of the dsRNA was consistent with the findings for the 50 ns simulations until 100–150 ns, after which point there were more distinct changes to the conformations as shown in the plots of RMSD and SASA in Figure 3.13. This is reflected in the change in base pairings that are maintained over the course of the simulation, with more hydrogen bonds breaking and most not reforming after 100 ns. Previous assessments of the CHARMM36 force field for RNA have shown that dsRNA is stable at least on the scale of tens of microseconds, so these are likely to be plausible conformational changes (Liebl and Zacharias, 2023).

PCA was performed on the conformations of the dsRNA backbone for the short and extended simulations. The first two eigenvectors covered more than

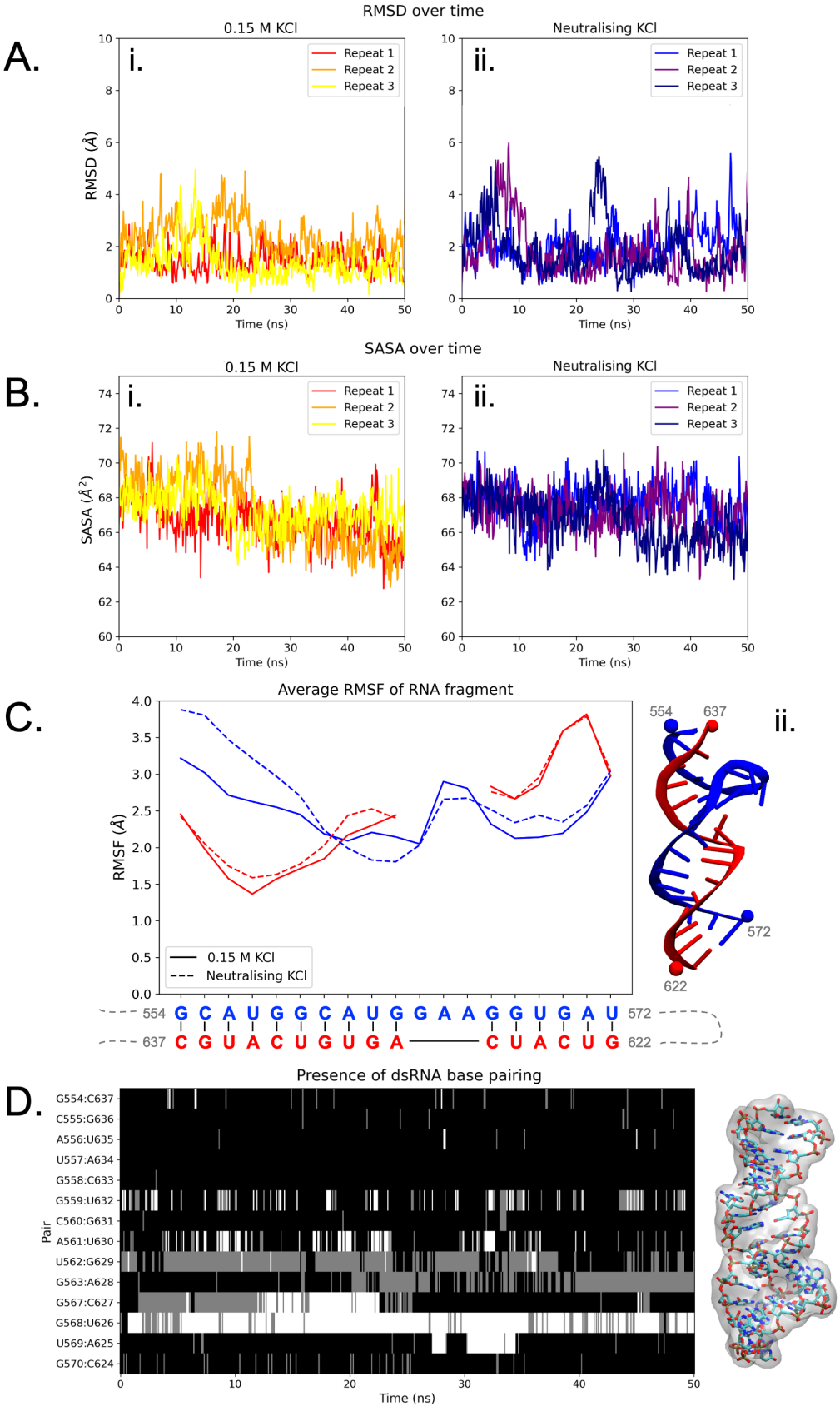



**Figure 3.12:** *(Caption on next page.)*

**Figure 3.12:** A. RMSD of dsRNA over time in 0.15 M KCl (i.) and neutralising $K^+$ (ii.). There is no significant difference between the two conditions.
B. SASA of dsRNA over time in 0.15 M KCl (i.) and neutralising $K^+$ (ii.). There is no significant difference between the two conditions.
C. i. RMSF of dsRNA fragment in 0.15 M KCl, with the sequence and base pairing shown along the $x$-axis. C. ii. Structure of the fragment coloured as in i. in cartoon representation, and in licorice representation showing all non-hydrogen atoms and the surface area.
D. Base pairing of dsRNA fragment in 0.15 M KCl over time. Black shows all hydrogen bonds maintained, grey is a least one is present, white is no hydrogen bonds remain.

60% of the variance and the first ten eigenvectors covered more than 90% of the variance (see Appendix Section A.5, Figure A.5), and the trajectories were projected across the first three principal components as shown in Figure 3.14. This projection shows that the salt concentration did not impact the conformational space explored, and that the extended simulation allowed the dsRNA to explore a greater conformational space.

### 3.3.3 Initial simulations of WT FhuA

Equilibrium MD simulations of WT FhuA in a POPC membrane revealed that the pore is highly flexible including in the $\beta$-sheets. This is likely due to its size and caused a lot of variation in the shape of the pore as shown in Figure 3.15. Previous studies have shown that nanopore shape has an impact on the flux of ions and the translocation of nucleotides through the pore (Omori *et al.*, 2021; Wang *et al.*, 2021a; Seiferth and Biggin, 2024). Upon observing this shape variation in all repeats, it was decided that to study the effect of residue mutation as a design principle, this property should be isolated from the stochastic flexing of the protein. Thus, position restraints were applied to the C-$\alpha$ atoms of FhuA, which allowed the side chains to move but maintained the pore shape throughout the simulation. The use of restraints will have to be considered when assessing the results of any experiments of WT FhuA as a nanopore. This is because *in vivo* the protein would have the ability to be relatively flexible, and additional modifications may have to be made to mitigate this (Niedzwiecki *et al.*, 2012). The application of an electric field in the simulations induces some membrane curvature due to the

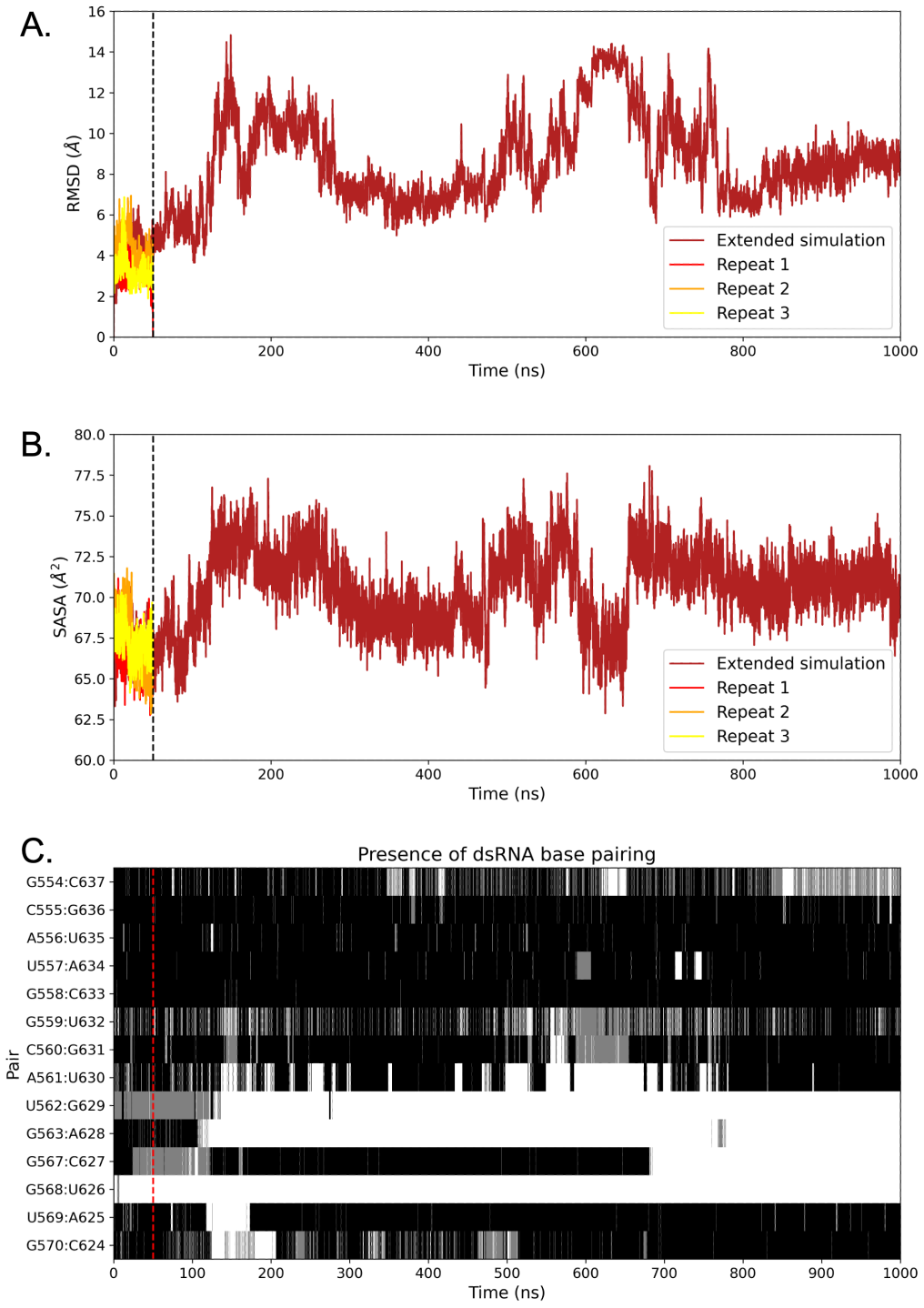



**Figure 3.13:** A. RMSD of dsRNA over time in 0.15 M KCl, with the 50 ns simulations shown to the left of the dashed black vertical line. RMSD increases significantly after approximately 150 ns.
B. SASA of dsRNA over time in 0.15 M KCl, with the 50 ns simulations shown to the left of the dashed black vertical line. SASA increases after approximately 100 ns.
C. Base pairing of dsRNA fragment in 0.15 M KCl over time. Black shows all hydrogen bonds maintained, grey is a least one is present, white is no hydrogen bonds remain.

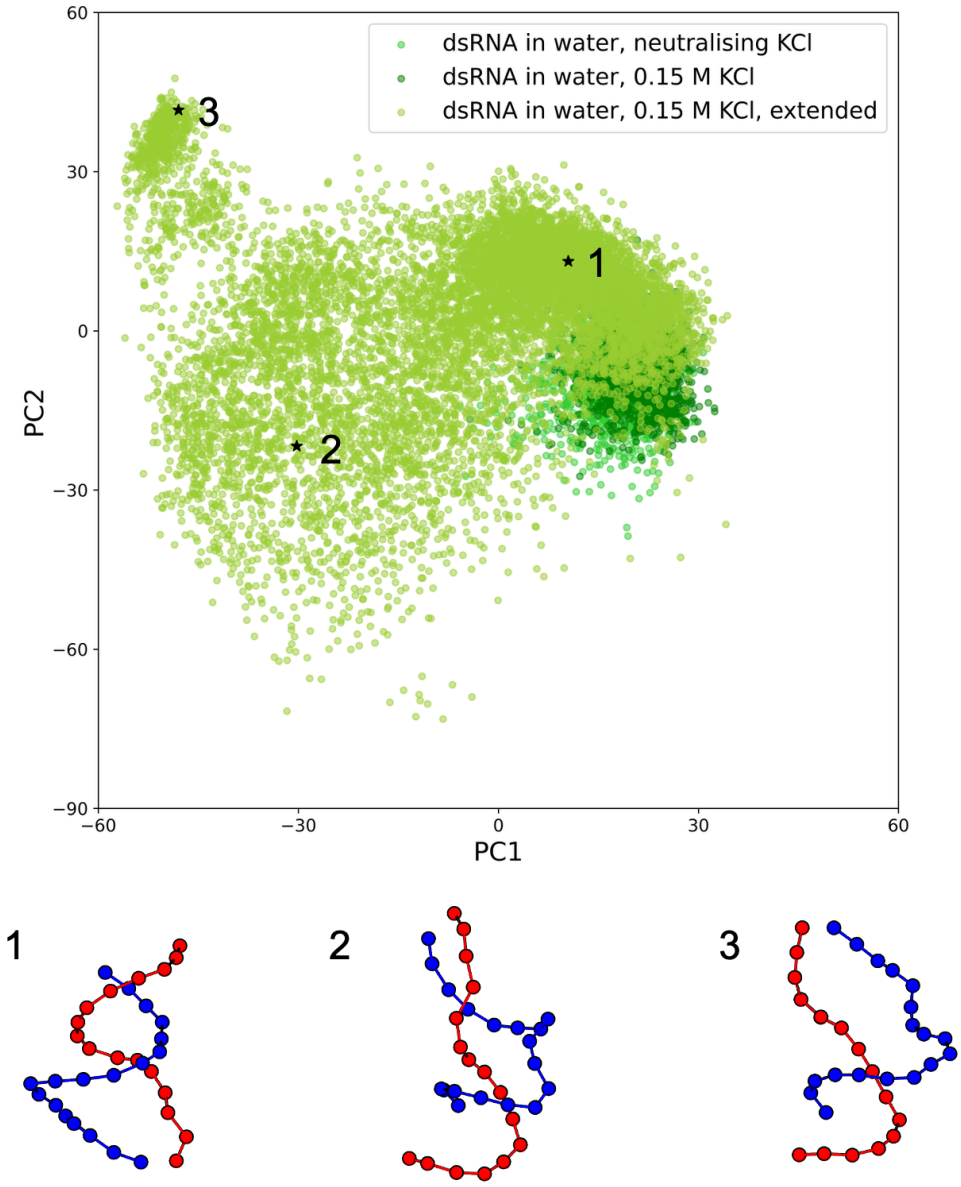



**Figure 3.14:** Projection of dsRNA backbone structures from simulations in neutralising $K^+$ and 0.15 M KCl into space calculated from concatenating all mentioned simulations, showing two largest principle components (which cover $\geq$60% of the variance). Extended simulation was run for 500 ns after the initial 50 ns. Structures 1, 2, and 3 are shown to represent the conformations of the three portions of the projection.

overall charge of the protein being non-zero (−8 for WT FhuA, +16 for DA/EA FhuA, −27 for KA/RA FhuA, −5 for DA/EA/KA/RA FhuA). The results of these are detailed in Appendix Section A.6, Figure A.6.

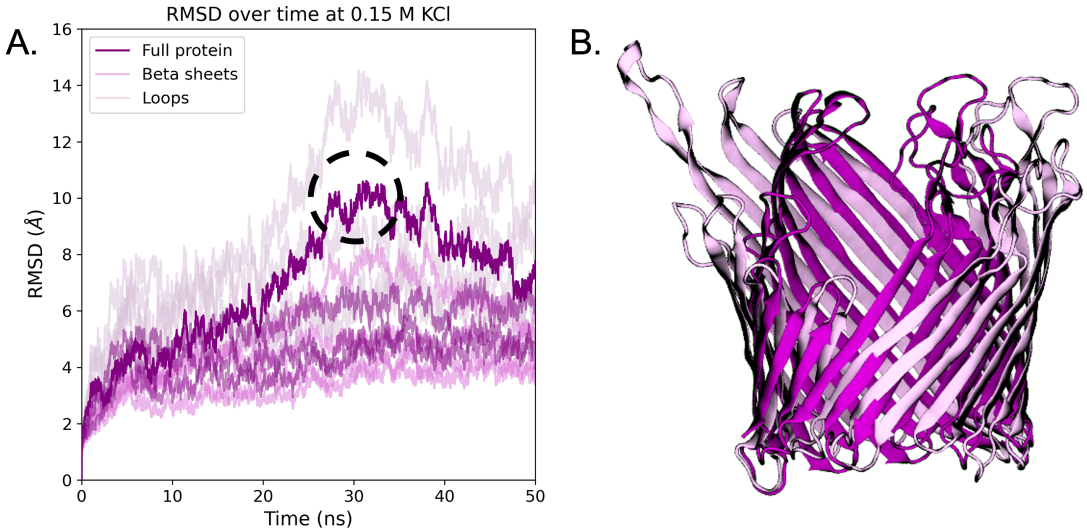



**Figure 3.15:** A. Plot of RMSD of WT FhuA over time. Darkest lines show RMSD for four replicas. Data is also shown divided for residues in $\beta$-sheets and residues in flexible loops.
B. Dark purple structure is crystal structure of FhuA; lilac structure is the structure at 30 ns of simulation of the replica circled in A.

### 3.3.4 Ion conductance through WT FhuA is affected by mutations

Ion flux was calculated through the engineered FhuA at a range of electric fields, from 0 to −0.1 V nm$^{-1}$, without RNA in the pore. Simulations were also performed without position restraints (see Appendix Section A.7) showing that the trends of behaviours are comparable with and without C-$\alpha$ restraints. The median number of frames for ion crossing (0.01 ns per frame), conductance (ions ns$^{-1}$), and ion speed in $z$ (nm ns$^{-1}$) were calculated and plotted for both potassium and chloride ions. Data for ion speed is shown in Figure 3.16; data for number of frames for crossing and conductance are shown in Appendix Section A.8.

The interpretation of MD simulations of ions moving through nanopores can be limited due to the fixed nature of force fields, which may not account for

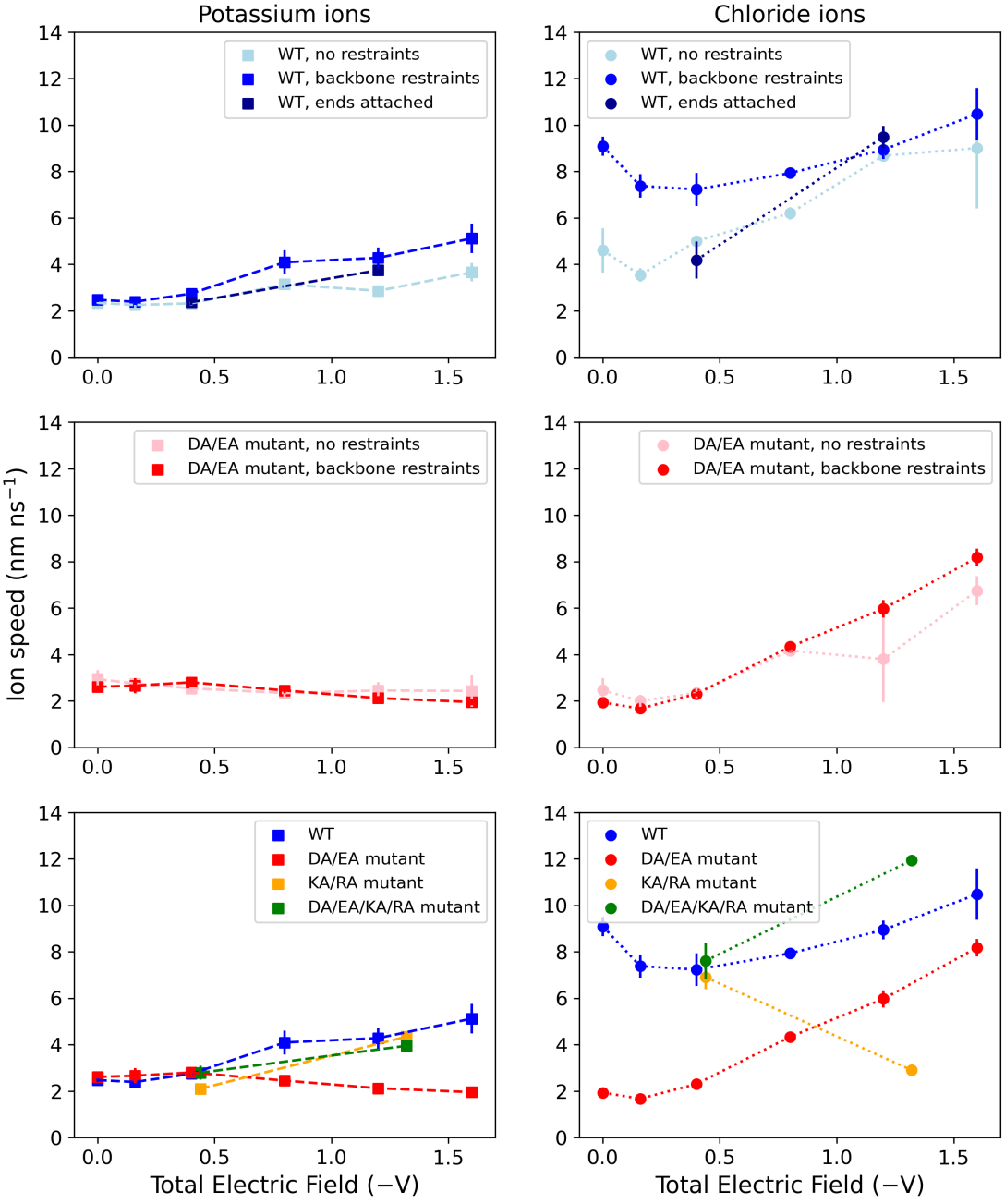



**Figure 3.16:** Plots of ion speed in the $z$ dimension against applied electric field for potassium (left column) and chloride (right column) ions. Standard deviation shown by vertical lines from each point. The correlation between ion speed in $z$ and total electric field is inverse for the DA/EA mutant compared to the others when considering potassium ions; the correlation is inverse for the KA/RA mutant compared to the others when considering chloride ions.

the polarizability of the system components without significant computational cost (Modi *et al.*, 2012). Studies of the dependence of KCl conductivity on concentration, using OmpF, showed that there is agreement with experimental data between 0 and 1 M (Pezeshki *et al.*, 2009). OmpF is a 16-strand $\beta$-barrel with an internal constriction and is thus not completely comparable to FhuA. However, since a concentration of 0.15 M KCl is used for the simulations of FhuA in this work, the experimental results lend confidence to the simulation data when interpreting the limitations of the results. Additionally, the work of Pezeshki *et al.* (2009) suggests at least 10 ns of simulation is required for reasonable statistics, a standard met by this work. Studies of OmpF also showed the influence of point mutations on the transport of ions.

#### 3.3.4.1 Impact of DA/EA mutations

Comparing WT and the DA/EA mutant, the trends for median frame count were different for potassium and chloride ions. As the electric field strength was increased, the length of time for a potassium ion to cross the pore decreased in the WT pore after an initial peak yet increased in the DA/EA mutant. This result was reversed for chloride ions; in the WT pore there was a slow increase in the amount of time, whereas in the DA/EA mutant there was an exponential decrease in the amount of time taken for ions to cross with the increase in electric field strength. Therefore, this suggests that electrostatic interactions are important in the length of time it takes for ions to pass completely through the pore, which is reflected in the calculations of ion speed. Potassium speed decreased with increased electric field strength in the DA/EA mutant.

#### 3.3.4.2 Impact of KA/RA mutations

The trends for KA/RA mutants were the opposite to the results for DA/EA mutants. Chiefly, the speed of chloride ions decreased as electric field increased, inversely to the trend for the WT pore. This suggests that the important factor in speed of ion translocation is oppositely charged side chains in the pore, and the absence of the electrostatic attraction hinders the movement of those ion species. More

simulations are needed to test the limits of this paradoxical relationship between electric field strength and ion speed in the DA/EA and KA/RA mutants, for potassium and chloride ions respectively.

#### 3.3.4.3 Impact of DA/EA/KA/RA mutations

By combining both the removal of negatively and positively charged side chains, the behaviour of ion movement returned to following the trends for the WT pore. Slight deviations from the behaviour in WT pores were likely due to the removal of an unequal number of negatively and positively charged side chains, leading to a different overall electrostatic profile of the pore.

#### 3.3.4.4 Calculating ionic current through the pore

The data in subsection 3.3.4 was used to calculate the current across the membrane in each system, following the method from Göpfrich *et al.* (2016) (see Chapter 2, subsubsection 2.1.6.6, Equation 2.1.6.3). The results, shown in Figure 3.17, showed that current scaled with applied voltage in a linear fashion. Studies of KCl transport through OmpC showed that the relationship between temperature and conductance are modelled more accurately in MD simulations with concentrations lower than 0.3 M KCl and temperatures between 15 and 35 °C (288–308 K) (Biró *et al.*, 2010). Additionally, the qualitative change in gradient of the graphs of conductance through the mutants are in agreement with earlier studies of OmpF (Pezeshki *et al.*, 2009). The experimental and simulation data for KCl transport through OmpF and OmpC give conductivity values between 10 and 100 m$S$ cm$^{-1}$ for concentrations under 1 M KCl, and between 1 and 1.5 n$S$ for temperatures between 288 and 308 K. The results for the simulations in subsection 3.3.4 can be used to calculate a conductivity of 48.4 m$S$ cm$^{-1}$, which is 2.88 n$S$ (see Chapter 2, subsubsection 2.1.6.6, Equation 2.1.6.5). These values are in the same order of magnitude as previous experimental data from Biró *et al.* (2010), with the difference in conductivity between OmpC and FhuA as a likely contributor to the difference since FhuA has a larger internal pore. The simulation results also agree with

experimental work using OmpG, which showed unitary conductance values around 1.2 n$S$, which again is in the same order of magnitude (Chen *et al.*, 2008).

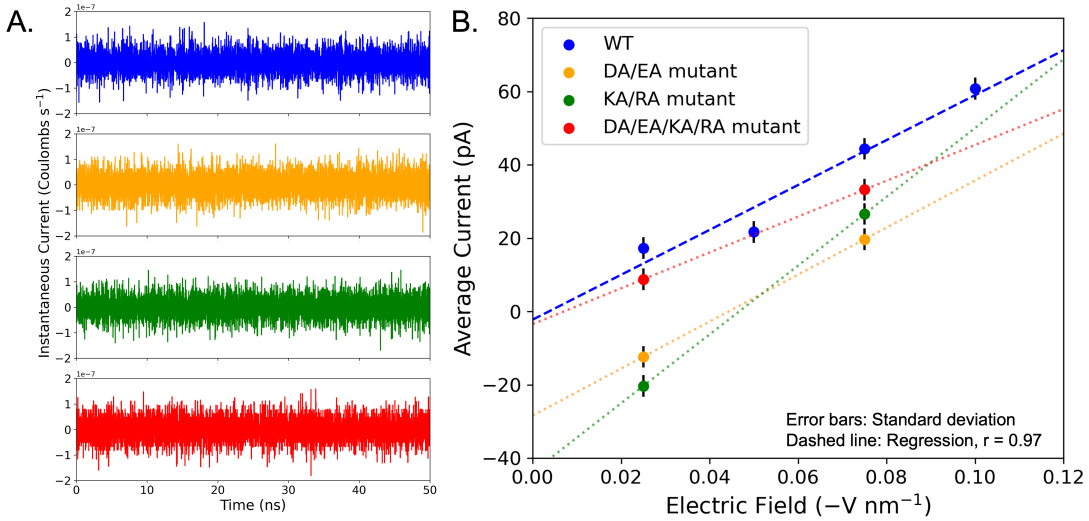



**Figure 3.17:** A. Example plots of instantaneous current, in $10^{-7}$ Coulombs s$^{-1}$, for different FhuA pores.
B. Calculated average current plotted against applied electric field. Regression given only for the data with the WT pore.

### 3.3.5 dsRNA does not always escape the pore in simulations

Simulations were performed using $-0.05$ V nm$^{-1}$ and $-0.03$ V nm$^{-1}$ electric fields. For translocation of the dsRNA to occur within the accessible timescale of 50 ns, $-0.05$ V nm$^{-1}$ was often required but was sometimes insufficient to overcome the factors holding the RNA in the pore. Table 3.5 shows the overview of translocations, together with the time taken for the dsRNA centre of mass to move above the upper lipid leaflet. Some simulations were also performed with different concentrations of KCl, with the data in Appendix Section A.9.

This data showed that dsRNA translocation happened much faster in DA/EA mutants and slightly faster in DA/EA/KA/RA mutants. However, in KA/RA mutants, the ability for translocation was abolished.

Three simulations of the WT pore that were performed with a $-0.03$ V nm$^{-1}$ electric field were extended to 500 ns. In two cases, the RNA still did not escape

**Table 3.5:** Time taken, in nanoseconds, for the dsRNA to fully translocate through the pore if applicable.

| **Pore** | **Replicas** | **Electric Field (V nm$^{-1}$)** | **Number of translocations** | **Crossing times (ns) and *average*** |
|---|---|---|---|---|
| WT | 12 | −0.05 | 6 | 42.06, 19.09, 11.92, 39.05, 37.84, 30.91 *30.15 ± 11.07* |
| WT | 12 | −0.03 | 1 | 26.24 *n/a* |
| DA/EA | 12 | −0.05 | 12 | 3.78, 24.52, 17.46, 7.58, 11.14, 8.62, 7.00, 6.60, 29.65, 18.65, 17.00, 4.25 *13.02 ± 8.00* |
| DA/EA | 12 | −0.03 | 4 | 38.19, 38.73, 45.34, 26.30 *37.14 ± 6.86* |
| KA/RA | 6* | −0.05 | 0 | n/a *n/a* |
| KA/RA | 6* | −0.03 | 0 | n/a *n/a* |
| DA/EA/KA/RA | 6 | −0.05 | 5 | 29.42, 3.49, 19.93, 12.35, 24.14 *17.87 ± 9.10* |
| DA/EA/KA/RA | 6 | −0.03 | 2 | 28.81, 30.01 *29.41 ± 0.60* |

the pore and did not make significant further progress. However, in one set-up, the RNA did escape the pore at 103 ns. This showed that translocation at a lower applied electric field is possible over a longer timescale, which is more representative of what happens *in vitro*. The extended simulations also suggest that the initial threading is important: if the RNA is allowed to find a crevice in the pore that is non-productive for translocation, then it cannot escape and will not successfully translocate on a sub-microsecond timescale (see Figure 3.18).

## 3.3.6 Interrogating the role of dsRNA structure in translocation

### 3.3.6.1 Hydrogen bonding between dsRNA base pairs during translocation

The persistence of hydrogen bonds between the dsRNA base pairs was calculated, and compared to the data from subsection 3.3.2 (see Appendix Section A.10,

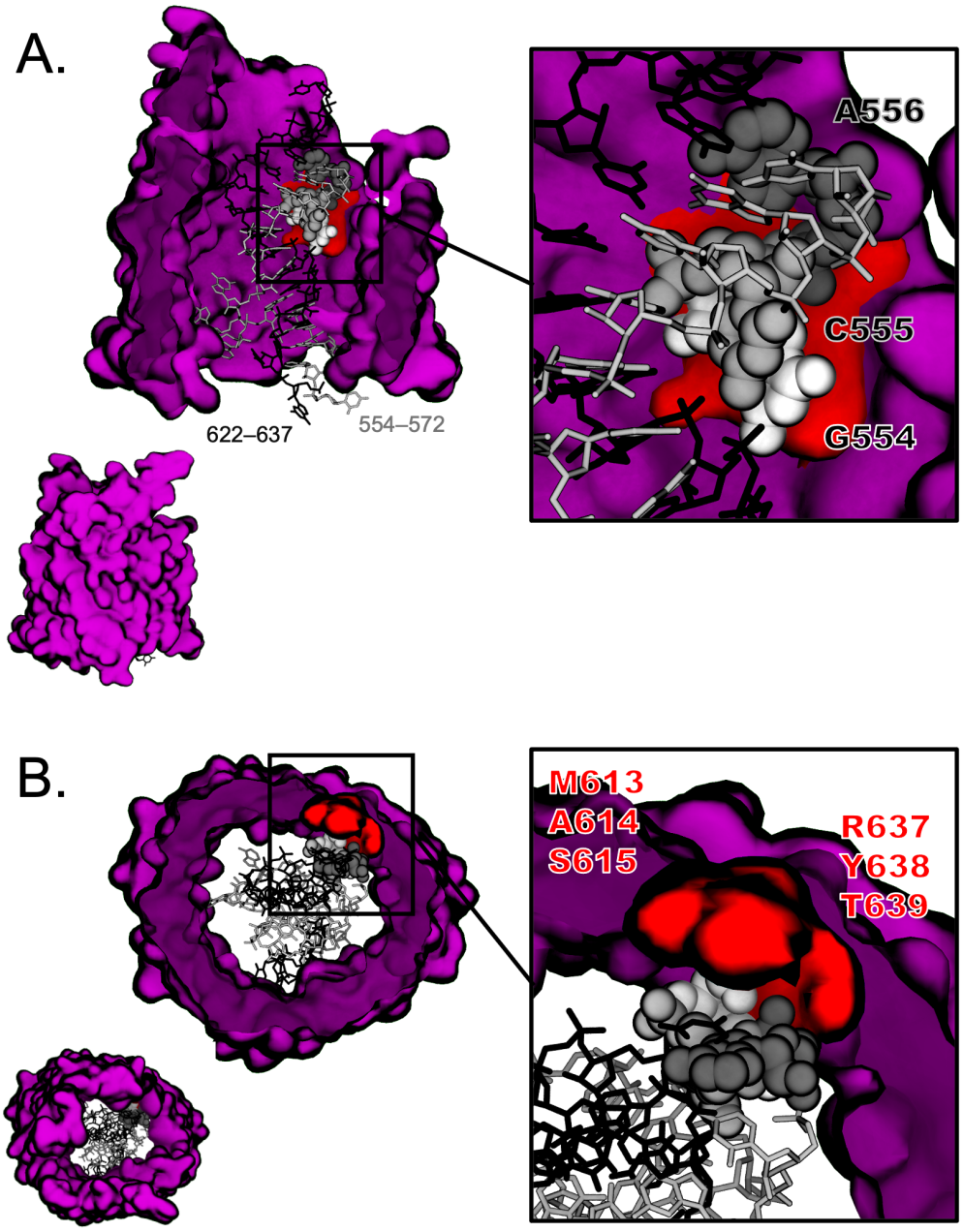



**Figure 3.18:** A. Side-on, cut-through view of dsRNA in non-productive translocation position, during an extended simulation of dsRNA in a WT pore with a $-0.03$ V nm$^{-1}$ electric field. Most of the dsRNA shown in licorice representation, with three residues in van der Waals representation. Residue shown in white is the 5' end of the strand. Patch of pore (shown in QuickSurf representation) in red is the residues within 5 Å of the three dsRNA residues shown in VdW representation.
B. Top down view of the same frame shown in A.

Figure A.10). This data showed that translocation led to the breaking of hydrogen bonds holding the secondary structure together. The biggest decreases (> 50%) were observed for the wobble pairs G-U, and the base pair that translocated first.

#### 3.3.6.2 Influence of dsRNA starting position on translocation speed

During the simulations, all the charged residues of the pore were in contact with the dsRNA at some point. The most common interactions with the dsRNA may have been influenced by the starting position of the nucleotides, though several methods were employed to try and lessen this bias (see subsection 3.2.4). Allowing the spontaneous insertion of dsRNA over multiple repeats was likely to be unproductive and a waste of computational resources, limiting the possible approaches and introducing a potential source of bias. Coarse-grained simulations were attempted (see Appendix Section A.11) however these did not provide sufficient additional data to justify the loss of molecular detail.

#### 3.3.6.3 Role of dsRNA structure in translocation speed

Interestingly, one set of six simulations of dsRNA translocating through the WT pore resulted in a much greater distortion to the dsRNA structure. In several of these simulations, there was a near-complete breakdown of the base pairing and a buckling of the helix, with the bases being pushed out in one direction (Figure 3.19). This also led to a greater increase in RMSD over the course of the simulations. In these examples, base pairs that usually had fewer strong interactions remained paired. It is unclear why this happened in some simulations and not others.

#### 3.3.6.4 Principal component analysis of dsRNA conformations

To compare the conformational space explored by the dsRNA within the pore to the data from subsection 3.3.2, PCA was performed on the dsRNA backbone structures from the concatenated trajectories of all the simulations. However, the first two eigenvectors covered less than 30% of the variance and the first ten eigenvectors covered only 60% of the variance, limiting the interpretation of this analysis (see Appendix Section A.12, Figure A.12). This suggests there are not any dominating

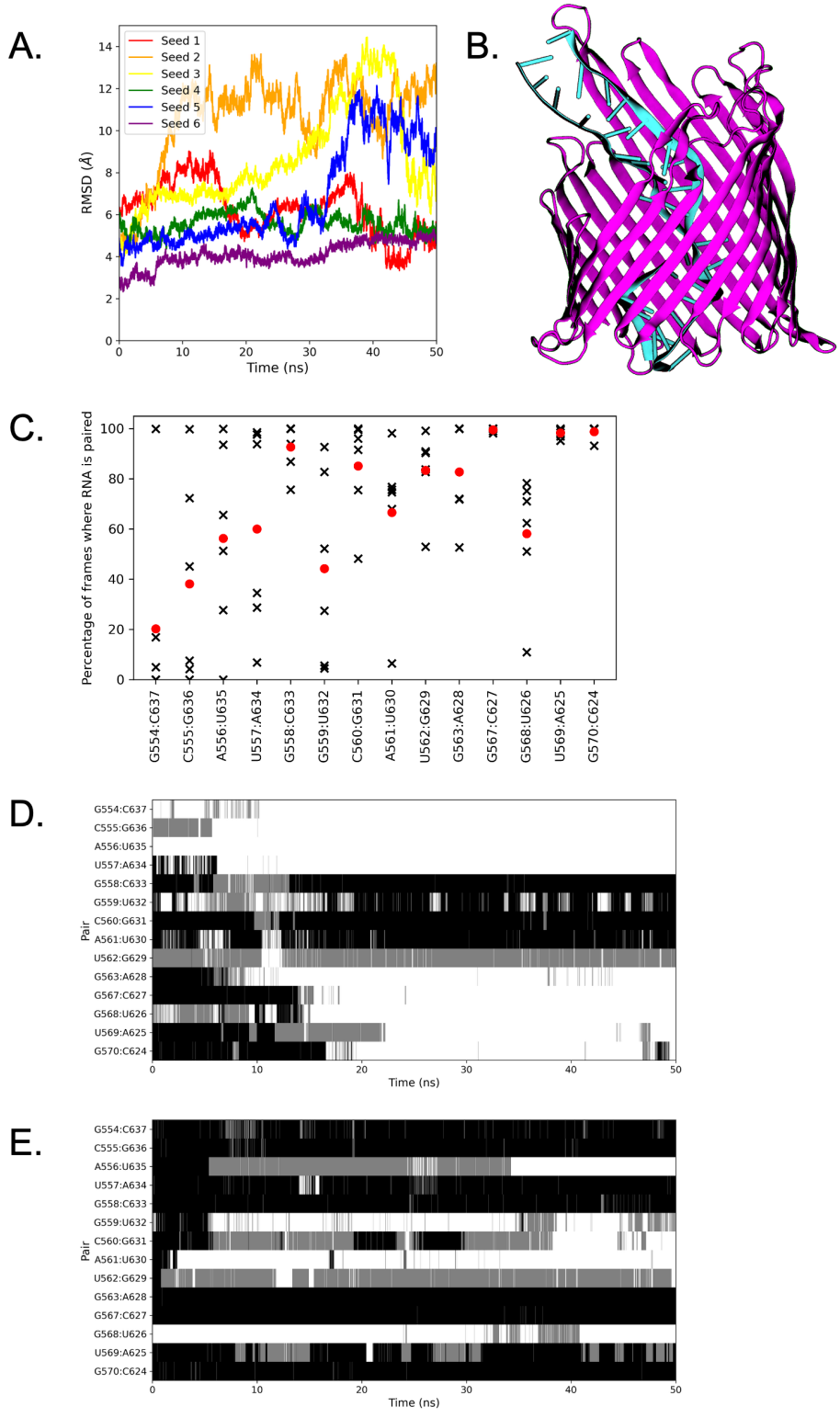



**Figure 3.19:** *(Caption on next page.)*

**Figure 3.19:** A. RMSD of dsRNA against time compared to equilibrated dsRNA structure from Figure 3.12, with seed 2 shown in orange and seed 6 in purple.
B. Structure showing deformation of dsRNA in seed 6 with high RMSD value.
C. Percentage of frames where dsRNA is base-paired as at the beginning of the simulation. Xs indicate individual simulations, circles indicate average percentage.
D. Base pairing of dsRNA fragment over time for simulation seed 2.
E. Base pairing of dsRNA fragment over time for simulation seed 6.

conformational changes as the dsRNA passes through the pore. The backbone structures were projected onto the two largest components. Figure 3.20 shows that the conformational space explored by dsRNA in water had a large degree of overlap with the dsRNA in simulations with the pore. Generally, the data suggests that conformations of dsRNA are independent of the applied electric field in this range. Simulations of dsRNA in the DA/EA pore showed the dsRNA exploring a more distinct conformational space than the other simulations, but the lack of variance explained by the eigenvectors makes detailed interpretation of this result unreliable.

### 3.3.7 Characterisation of protein-RNA electrostatic interactions

The simulations of dsRNA translocating through WT FhuA were analysed for common interactions based on electrostatics and charged amino acid side chains. Each nucleotide in RNA has a charge of $-1$ from the phosphate group, and thus interactions with the positively charged residues arginine and lysine are likely to be important while the RNA moves through the pore. Additionally, dsRNA interactions with the acidic residues glutamate and aspartate could either lead to repulsion or be mediated by positively charged ions that are commonly coordinated by the dsRNA phosphate groups. The electrostatic interior of the pore was calculated (Figure 3.21) and shows both negatively and positively charged patches. Zhou *et al.* (2020) showed that the electrostatic surface of the pore had a particular impact at the constriction in MspA, CsgG, and $\alpha$-hemolysin, though this is likely to not be so pronounced for FhuA due to the wider pore $xy$ cross-sectional area.

The residency of charged protein residues by dsRNA over the duration of the simulations was calculated, to identify which residues are important for interacting

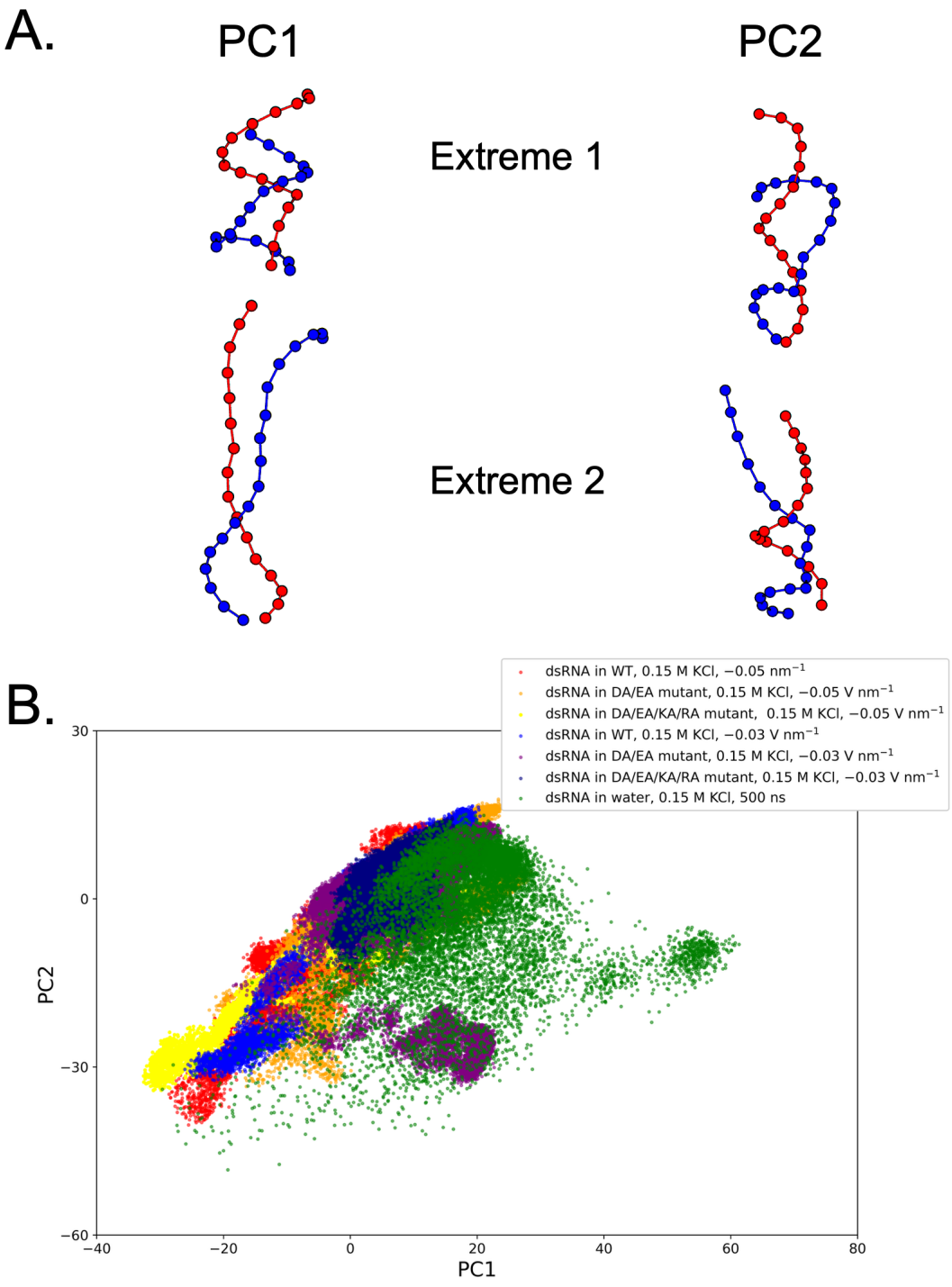



**Figure 3.20:** A. Projections of the principal component analysis onto the dsRNA backbone, showing the extreme conformations of the first two principal components. B. Projection of dsRNA backbone structures from all simulations where dsRNA is in the pore and simulations of dsRNA in water, into space calculated from concatenating all the dsRNA backbone structures of the mentioned simulations, showing the first two principle components (which cover $\leq 30\%$ of the variance).

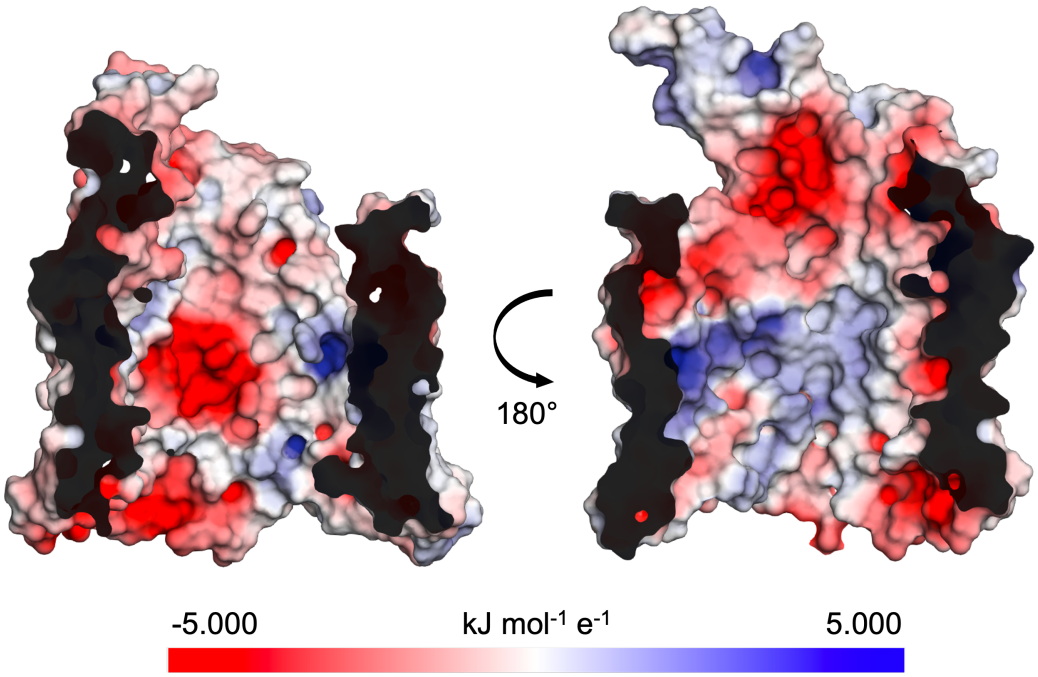



**Figure 3.21:** Visualisation of electrostatic surface of the pore of WT FhuA. Positive electrostatic potential shown in blue representing positively charged residues, negative electrostatic potential shown in red representing negatively charged residues.

with the RNA phosphate groups. These interactions are non-specific, that is they are not altered by the identity of the nucleotide. An interaction was defined as a distance of less than 5 Å between a nucleic acid phosphate and an amino acid functional group, or between one of the former and a potassium ion.

The lipid used in simulations, POPC, has a charged headgroup. In some simulations, the dsRNA formed electrostatic interactions with the headgroups. See Appendix Section A.13 for more details.

#### 3.3.7.1 Interactions with lysine and arginine residues

Residency time and residency percentage were calculated over the simulations for each of the pore-lining arginines and lysines with the phosphate groups of the dsRNA, allowing for key residues to be identified (example of analysis in Figure 3.22). Residues with over 20% residency (taken as a percentage of the time the dsRNA was inside the pore) were designated “important”. These were LYS526, LYS575,

ARG193, ARG199, ARG212, ARG222, ARG277, ARG297, ARG637, and ARG713 (10 residues of 19 total pore-lining basic residues).

#### 3.3.7.2 Interactions with aspartic acid and glutamic acid residues

Interactions between aspartic and glutamic acids and the nucleic backbone (both negatively charged species) are mediated by potassium ions. The most important interactions were ASP179, ASP287, ASP358, GLU163, GLU210, GLU279, GLU522, GLU571, and GLU573 (9 residues of 22 total pore-lining acidic residues). Example analysis is shown in Figure 3.23.

Potassium ions were observed interacting transiently with the dsRNA backbone phosphate groups. This is expected as potassium ions play an important role in the stability of RNAs *in vivo* (Grover, 2015; Rozov *et al.*, 2019). When the dsRNA was present in the pore but not translocating, the potassium ions were observed to move through the pore via interactions with adjacent phosphate groups, moving along one of the strands of the helix. This "helter skelter"-like movement of potassium ions was also observed when the dsRNA was actively translocating, although it was impacted by the conformational changes of the dsRNA as it became less helical under the effect of higher applied electric fields (see Appendix Section A.14, Figure A.14).

### 3.3.8 Mutations impact protein-RNA electrostatic interactions

#### 3.3.8.1 Interactions with the DA/EA mutant

Once the opportunities for interactions with aspartic and glutamic acid residues were removed, it was expected that other electrostatic interactions may become more important. Analysis showed that the interactions LYS526, LYS575, ARG193, ARG199, ARG212, ARG277, ARG637, and ARG713 were conserved, with ARG222 no longer appearing important. Additional interactions appeared with LYS167, LYS545, and ARG663.

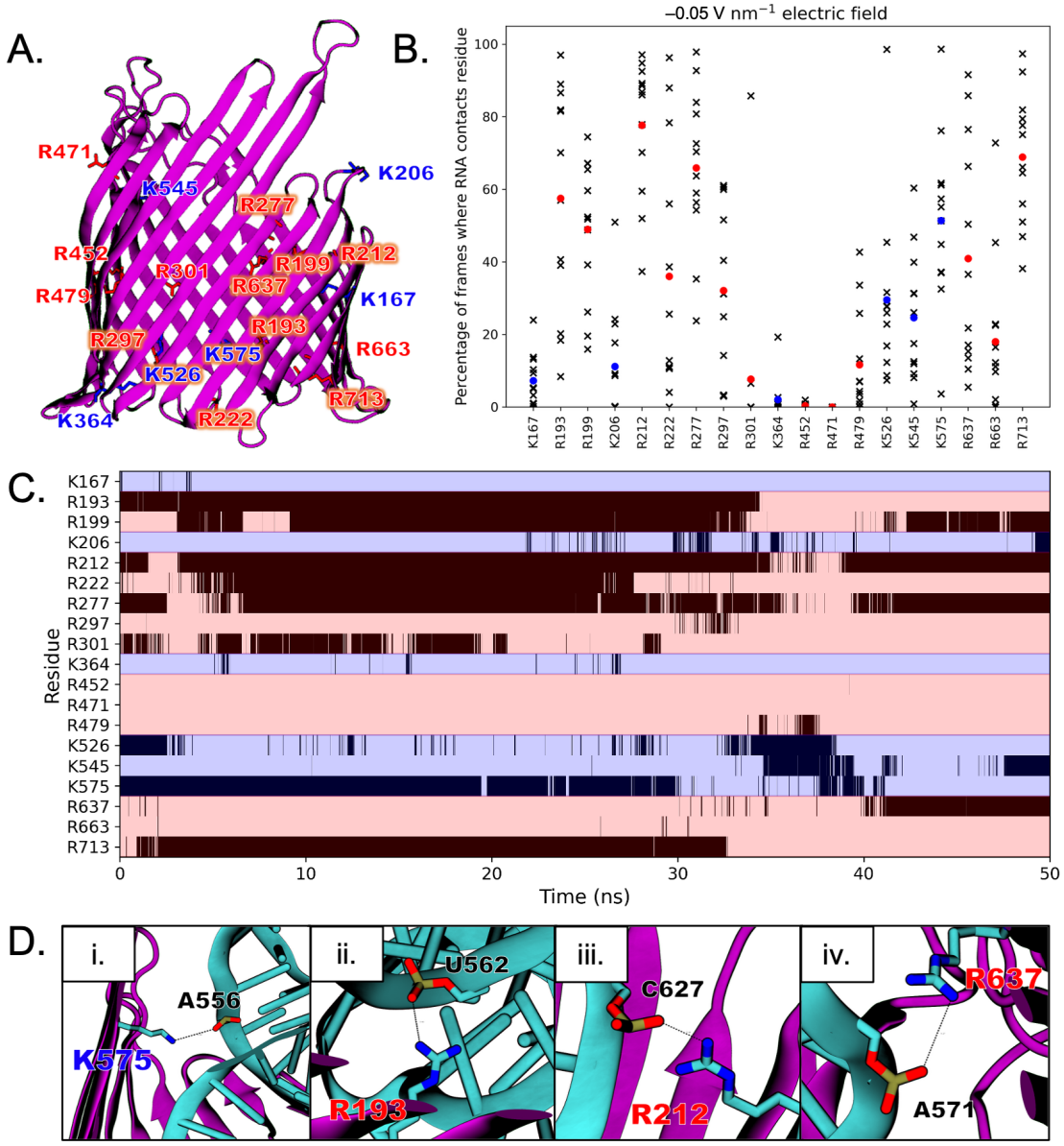



**Figure 3.22:** A. Positions of each lysine and arginine residue, with those identified as important highlighted in orange.
B. Percentage of frames where there is a contact, with residues having contact with dsRNA in $> 20\%$ of frames, in both $-0.05$ V nm$^{-1}$ and $-0.03$ V nm$^{-1}$ electric fields, being identified as important. Xs indicate individual simulations, circles indicate average percentage (red: arginine; blue: lysine).
C. Interaction of each protein residue with dsRNA backbone over time, with a black line indicating an inter-molecular distance of less than 5 Å.
D. Interactions between RNA and protein. Residue side chain and RNA phosphate group shown in licorice style, without hydrogen atoms. i. LYS575; ii. ARG193; iii. ARG212; iv. ARG637.

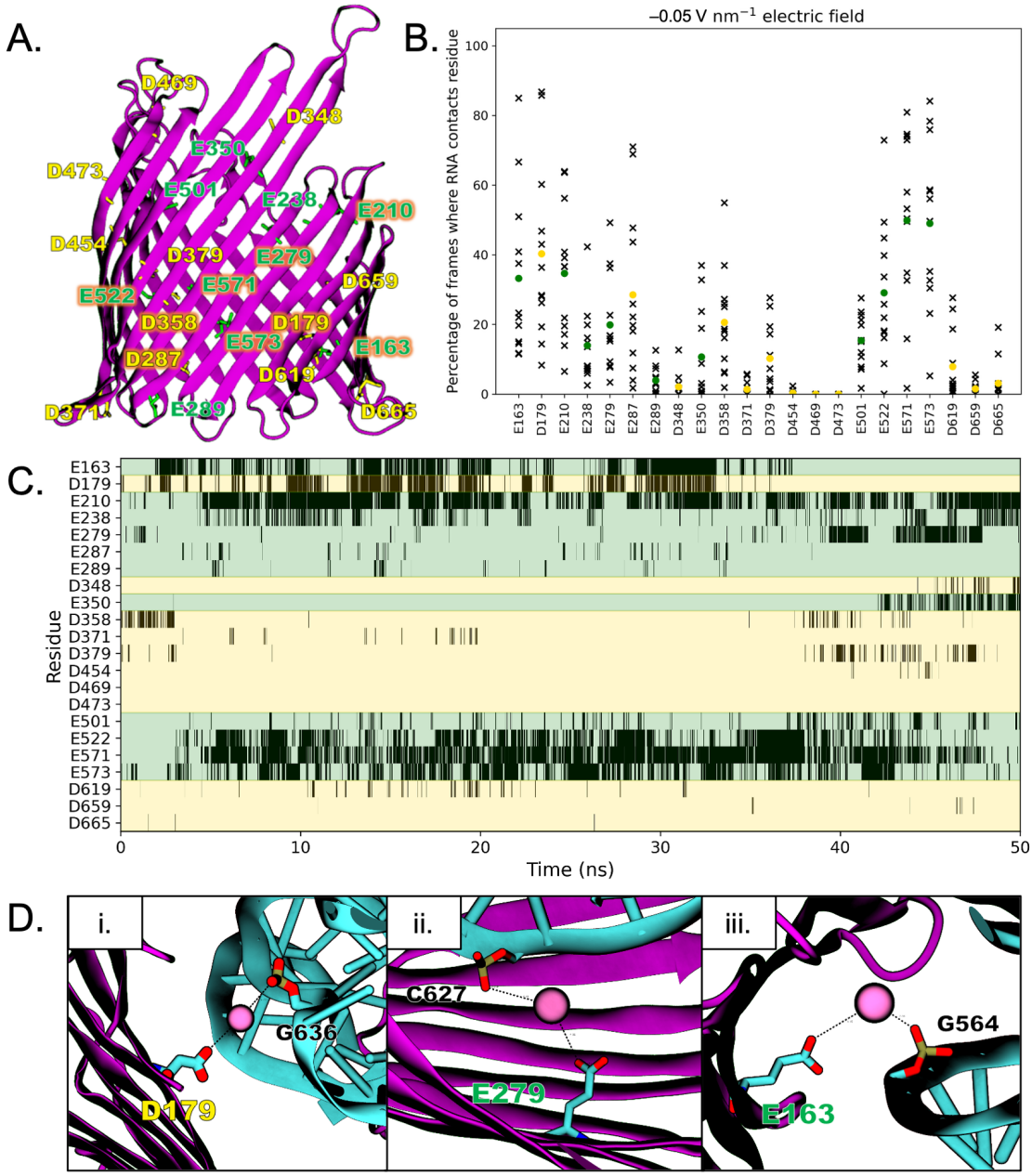



**Figure 3.23:** A. Positions of each aspartic acid and glutamic acid residue, with those identified as important highlighted in orange.
B. Percentage of frames where there is a contact, with residues having contact with dsRNA in $> 20\%$ of frames, in both $-0.05$ V nm$^{-1}$ and $-0.03$ V nm$^{-1}$ electric fields, being identified as important. Xs indicate individual simulations, circles indicate average percentage (green: aspartic acid; yellow: glutamic acid).
C. Interaction of each protein residue with the dsRNA backbone (via a potassium ion) over time, with a black line indicating an inter-molecular distance of less than 5 Å.
D. Interactions between RNA and protein. Residue side chain and RNA phosphate group shown in licorice style, without hydrogen atoms; potassium ions shown as pink spheres. i. ASP179; ii. GLU279; iii. GLU163.

The possibility of polar interactions was also considered, involving residues with partial charges. This showed interactions with ASN305, ASN675, GLN165, GLN175, GLN236, GLN505, and GLN608, and pi-stacking interactions with PHE500, PHE557, PHE650, and TYR655 (Figure 3.24).

#### 3.3.8.2 Interactions with the DA/EA/KA/RA mutant

This mutant also showed polar interactions that may be compensating for the lack of electrostatic interactions (Figure 3.25). Some of these were conserved from the DA/EA mutant (ASN305, GLN165, GLN236, PHE557, PHE650); some were additional (ASN583, ASN681, GLN705, TYR234, TYR275, TYR435).

#### 3.3.8.3 Polar interactions in the WT pore

With the identification of important polar interactions in the modified pores, analysis was performed to see which also appear in the WT pore. A small number were evaluated to be important. These were GLN608, GLN705, PHE650, and TYR655 (Figure 3.26).

#### 3.3.8.4 Relative positions of amino acid contacts

Some of the amino acids involved in interactions with the dsRNA were found in clusters, with the cluster involving ARG212, ARG199, ARG277, GLU210, and GLU279 being the most prevalent. This suggests that dsRNA translocates by initially interacting with the cluster of ARG713, GLU163, and ASP179 or the cluster with ARG297 and ASP287, and then moving up towards the ARG212 cluster. In the DA/EA/KA/RA mutant, with none of these charged residues present, there are nearby residues in clusters GLN236 and TYR234, and GLN705 and ASN681. The interactions seem to be prominently on one face of the pore, leading to the “low wall” of the extracellular side of FhuA (Figure 3.27).

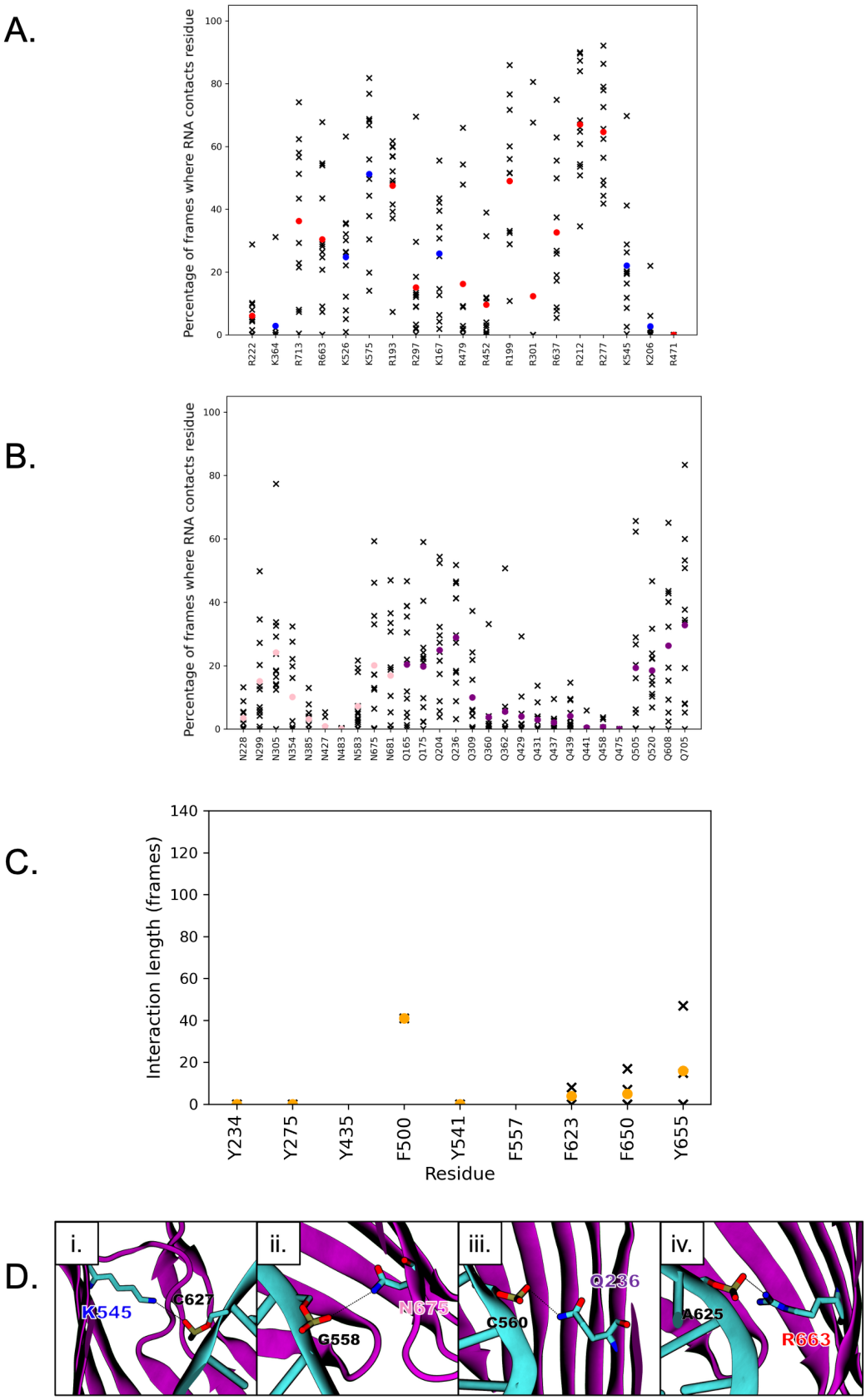



**Figure 3.24:** *(Caption on next page.)*

**Figure 3.24:** A. Percentage of frames where RNA contacts basic residues in DA/EA pore. Xs indicate individual simulations, circles indicate average percentage (red: arginine; blue: lysine).
B. Percentage of frames where RNA contacts polar residues in DA/EA pore. Xs indicate individual simulations, circles indicate average percentage (pink: asparagine; purple: glutamine).
C. Length of pi-stacking interactions between RNA and aromatic residues in DA/EA pore. Xs indicate individual simulations, circles indicate average interaction length.
D. Interactions between RNA and DA/EA pore. Residue side chain and RNA phosphate group shown in licorice style, without hydrogen atoms. i. LYS545; ii. ASN675; iii. GLN236; iv. ARG663.

### 3.3.9 Mutations impact the speed of dsRNA translocation

The pore-lining mutations of FhuA had the most pronounced impact on the number of ions that were associated with the dsRNA for the duration of the dsRNA translocating through the pore (see Figure 3.28 for visualisation of association). For the WT pore, there was an average of 14 potassium ions ($-0.03$ V nm$^{-1}$: $14.4 \pm 2.6$; $-0.05$ V nm$^{-1}$: $14.0 \pm 2.6$) associated with the phosphate backbone. This reduced to only an average of 8 potassium ions ($-0.03$ V nm$^{-1}$: $8.4 \pm 2.0$; $-0.05$ V nm$^{-1}$: $8.4 \pm 1.9$) for the DA/EA mutant, suggesting that the aspartic and glutamic acids are important in coordinating potassiums and thus the phosphates of the dsRNA. Therefore translocation through the pore would slow down because the phosphates are moving in the opposite direction to the potassiums, towards the negatively charged electrode. Subsequently there was an increase in speed of translocation in the DA/EA mutant due to less resistance from the flow of potassium ions. For the DA/EA/KA/RA mutant, there was only a moderate reduction in the number of ions to an average of 12 potassium ions ($-0.03$ V nm$^{-1}$: $12.9 \pm 2.4$; $-0.05$ V nm$^{-1}$: $12.1 \pm 2.4$), and this correlated with the moderate increase in translocation speed of dsRNA. The mutations had very little impact on the RMSD, the SASA, and the base-pairing of the dsRNA during translocation. The speed of dsRNA translocation, measured by the movement of the centre-of-mass through sections of the pore, was calculated and plotted (see Appendix Section A.15) but was not a useful metric for assessing dsRNA movement.

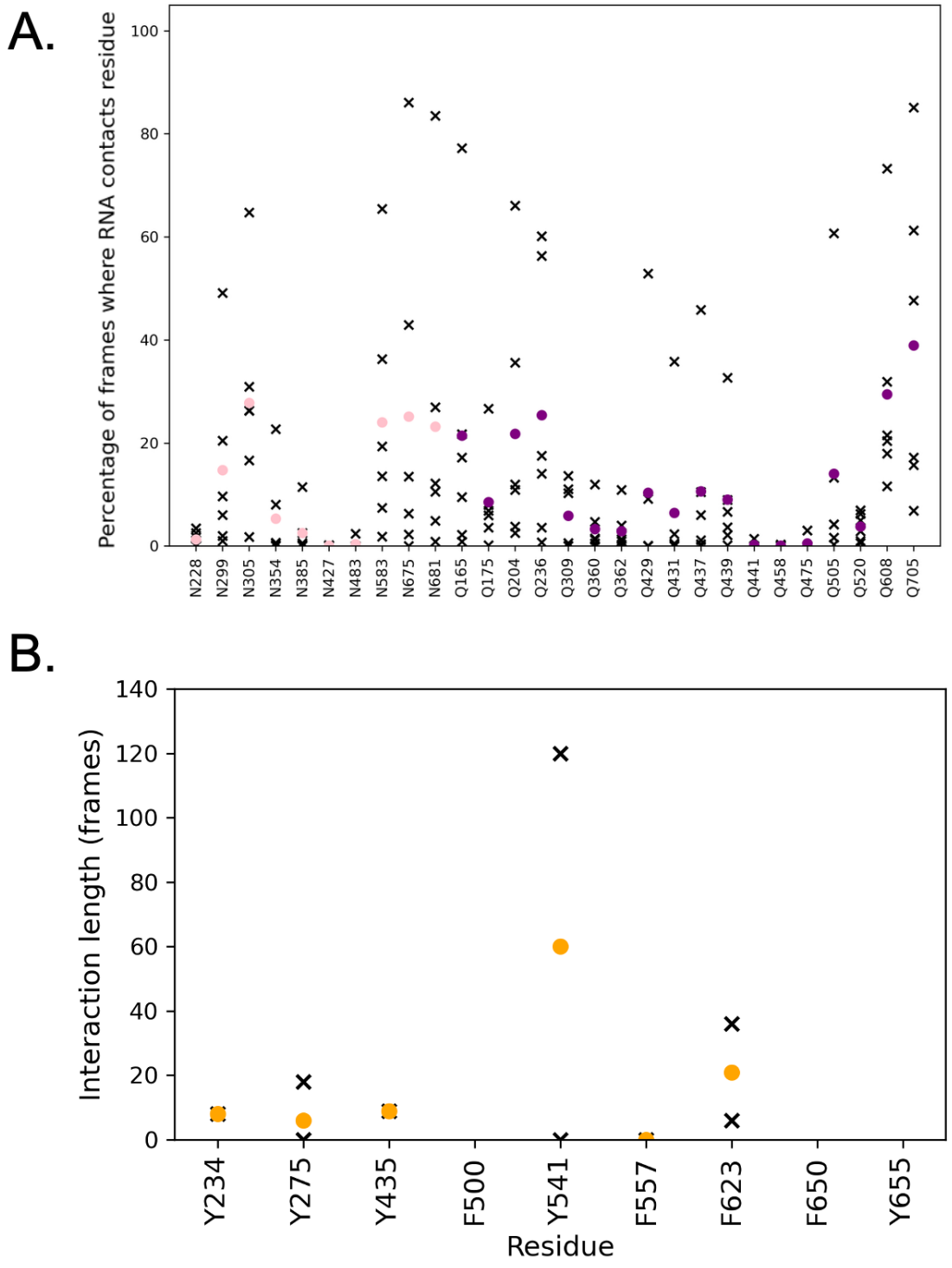



**Figure 3.25:** A. Percentage of frames where RNA contacts polar residues in DA/EA/KA/RA pore. Xs indicate individual simulations, circles indicate average percentage (pink: asparagine; purple: glutamine).
B. Length of pi-stacking interactions between RNA and aromatic residues in DA/EA/KA/RA pore. Xs indicate individual simulations, circles indicate average interaction length.

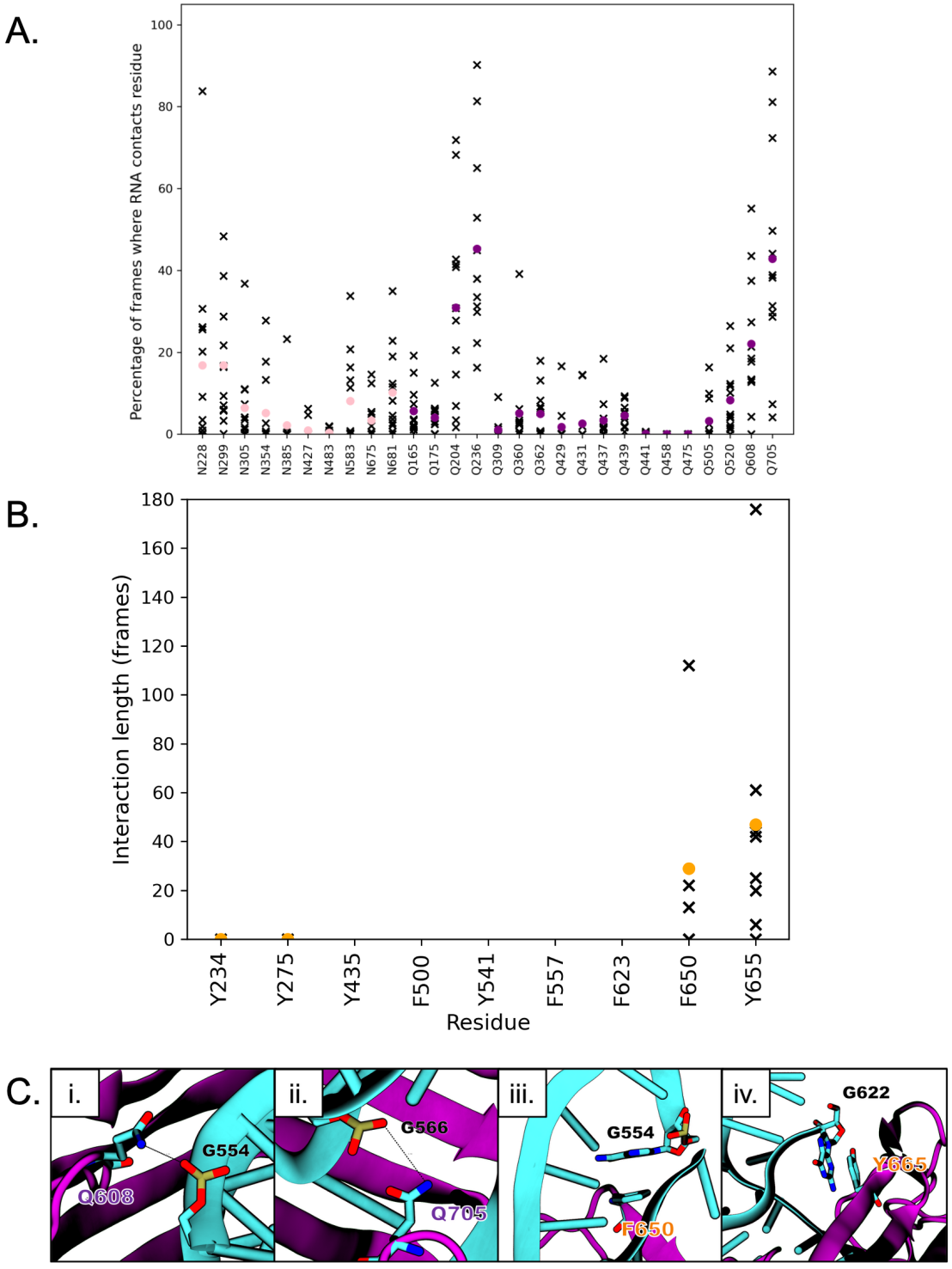



**Figure 3.26:** A. Percentage of frames where RNA contacts polar residues in WT pore. Xs indicate individual simulations, circles indicate average percentage (pink: asparagine; purple: glutamine).
B. Length of pi-stacking interactions between RNA and aromatic residues in WT pore. Xs indicate individual simulations, circles indicate average interaction length.
C. Interactions between RNA and WT pore. Residue side chain and RNA phosphate group shown in licorice style, without hydrogen atoms. i. GLN608; ii. GLN705; iii. PHE650; iv. TYR655.

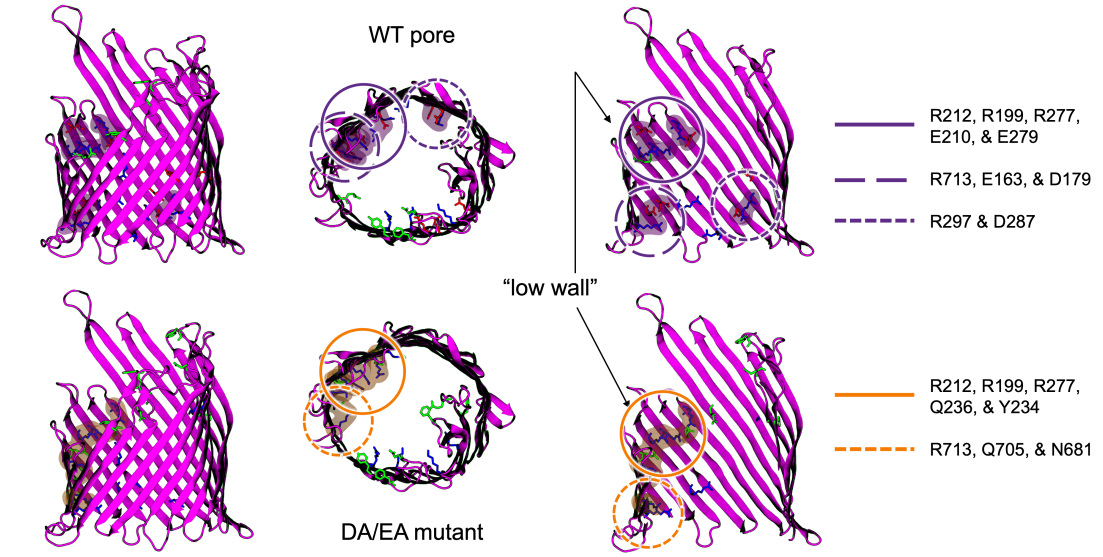



**Figure 3.27:** Positions of the residues indicated as important for interacting with the dsRNA, with those involved in clusters highlighted in purple (top, WT) or orange (bottom, DA/EA). The positions are very similar. Blue residues: positively charged. Red residues: negatively charged. Green residues: polar residues, or pi-stacking residues.

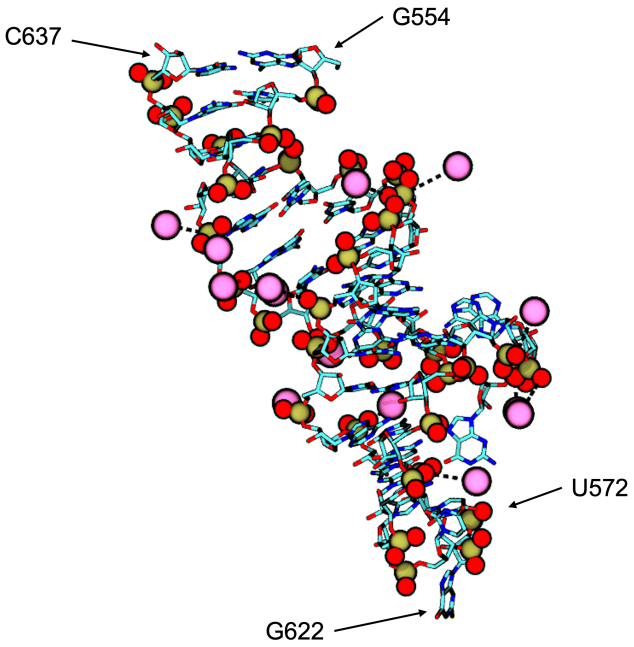



**Figure 3.28:** Example of dsRNA inside a WT nanopore, showing just the dsRNA and the potassium ions within 5 Å of the P, O1P, and O2P atoms comprising the phosphate moiety.

#### 3.3.9.1 Lack of dsRNA translocation in the KA/RA mutant

The results in subsection 3.3.9 may help explain why translocation did not occur through the KA/RA mutant. Without the repulsion of potassium ions by the positively charged lysine and arginine sidechains, more potassium ions can associate with the dsRNA phosphate backbone, and this leads to the dsRNA being pulled back out by the flow of potassium ions towards the negative side of the membrane in the opposite direction to the translocation direction of dsRNA. There is likely a balance to be found between having enough positively charged sidechains to allow translocation, as the simulations with the KA/RA mutant show that having no residues with positively charged sidechains prevents translocation, and not having so many that translocation occurs too quickly for a change in current to be detected as is observed in Table 3.5 where translocation occurs in under 15 ns on average under an $-0.05$ V electric field. Rapid ejection of a polynucleotide from a nanopore was also observed in previous simulation studies of single-stranded DNA through $\alpha$-hemolysin (Bond *et al.*, 2011). This was seen in a mutant pore with a more negatively charged electrostatic surface, in agreement with the results of subsection 3.3.5, though Bond *et al.* (2011) achieved this by mutating a methionine to an aspartic acid rather than by removing positively charged residues.

### 3.3.10 Translocation in the opposite direction

To investigate the importance of the identified residues, dsRNA was also pulled through the pore in the opposite direction, using an electric field reversed in the $z$-direction. If a residue interacted with the dsRNA in $> 20\%$ of the time when in the pore during simulations in both directions, it was designated as important. A notably higher percentage of translocation events was observed in all conditions when compared to the original simulations (see Table 3.5), and translocations occurred in a shorter amount of time, as shown in Table 3.6.

There was good agreement in the findings from the previous results with those from the reverse translocation simulations. In the WT pore, the interactions with lysine and arginine residues that appeared in both forward and reverse

**Table 3.6:** Time taken, in nanoseconds, for the dsRNA to fully translocate through the pore if applicable.

| **Pore** | **Replicas** | **Electric Field (V nm$^{-1}$)** | **Number of translocations** | **Crossing times (ns) and *average*** |
|---|---|---|---|---|
| WT | 6 | 0.05 | 6 | 19.09, 16.41, 5.96, 12.50, 19.63, 6.04 *13.21 ± 5.63* |
| WT | 6 | 0.03 | 2 | 39.77, 28.91 *34.34 ± 5.43* |
| DA/EA | 6 | 0.05 | 6 | 2.39, 3.31, 2.49, 4.35, 5.15, 2.62 *3.39 ± 1.04* |
| DA/EA | 6 | 0.03 | 6 | 34.40, 35.74, 15.59, 11.07, 4.29, 19.02 *20.02 ± 11.56* |

translocations are LYS575, ARG193, ARG212, ARG222, ARG277, ARG637, and ARG713, with LYS526, ARG199, and ARG297 showing up additionally. ASP179, GLU163, and GLU279 were identified in both, with ASP287, ASP358, GLU210, GLU522, GLU571, and GLU573 appearing additionally. No pi-stacking was seen in the reverse translocation simulations in the WT pore, but GLN608 and GLN705 interactions were seen.

In the DA/EA mutant, the conserved lysine and arginine residues were LYS545, LYS575, ARG193, ARG212, ARG277, ARG637, ARG663, and ARG713. As with the reverse translocation in the WT pore, LYS526 and ARG199 appeared as important interactions, and ARG297 only appeared in this mutant. ASN305, GLN236, GLN505, and GLN608 were conserved interactions from the forward DA/EA mutant simulations, and ASN385 was discovered as an additional interaction. The examples of pi-stacking observed were TYR655 (seen previously, Figure 3.26) and PHE623 (not seen previously).

It is worth noting that because translocations happened more quickly, there was less data on interactions since the dsRNA spent fewer frames within the pore. This may be due to the initial position of the threaded dsRNA (Figure 3.29). However, the prominent clusters of residues identified in the earlier simulations also appeared here, lending credence to these results.

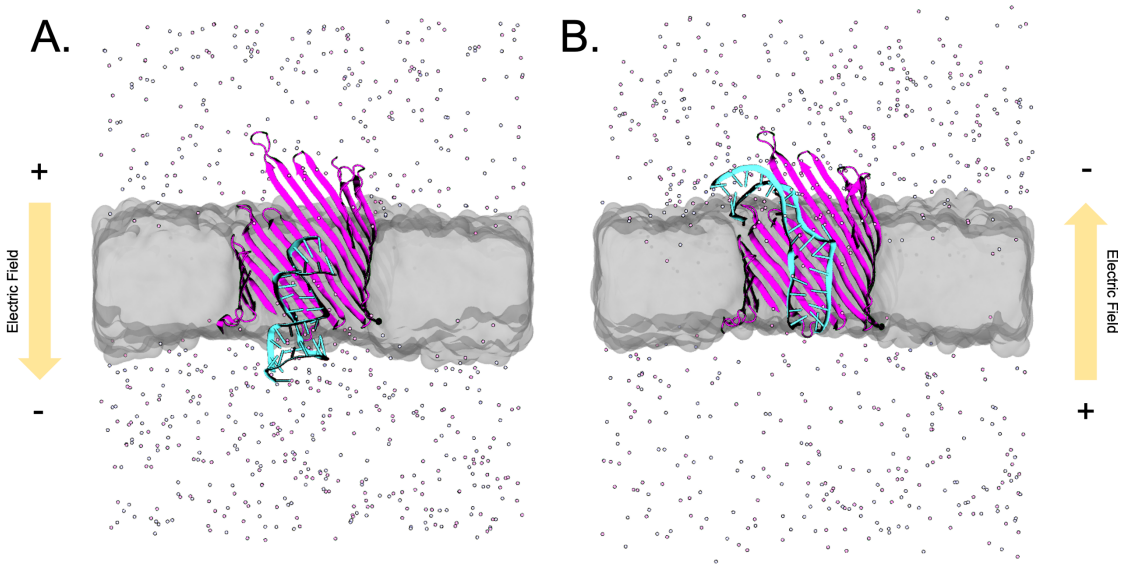



**Figure 3.29:** A. Initial threaded position of dsRNA in simulations for subsection 3.3.5. B. Initial threaded position of dsRNA in simulations for subsection 3.3.10.

As well as the general agreement on important residues, the trend involving ion association with the phosphate backbone was also maintained. On average, there were fewer potassium ions associated with the dsRNA backbone in the DA/EA mutant compared to the WT FhuA, which had the same outcome of faster translocation through the DA/EA mutant pore at both voltages tested.

### 3.3.11 Towards nanopore sequencing

When dsRNA translocates through the pore, there is an impact on the flux of ions through the pore and thus the current. The results showed that when occluded the DA/EA pore experienced a large, 250 pA increase in current. Uniquely, there was a reduction in current observed when dsRNA occluded the WT pore using a $-0.03$ V nm$^{-1}$ electric field, reducing to 7 pA from between 17 and 21 pA for the unoccluded pore at comparable voltages.

Simulations where the dsRNA was position-restrained in the pore were performed for 500 ns. One simulation was performed for each of WT, DA/EA mutant, and DA/EA/KA/RA mutant, at both $-0.05$ V nm$^{-1}$ and $-0.03$ V nm$^{-1}$ electric fields. The current across the pore was calculated for each system for this

period (Figure 3.30). The biggest difference between the occluded and non-occluded currents was observed with the WT pore, followed by the DA/EA/KA/RA mutant. The small difference for the DA/EA mutant suggests that the increase in movement of counter ions (potassiums) is not helpful for obtaining a distinguishable change in current when the pore is occluded.

The data in this section is not sufficient to determine what the difference in current might be for the presence of different base pairs in the pore. MD simulations of nanopores are not currently used to determine what the specific changes in current might be with the presence of different base pairs, instead focusing on insights on molecular interactions (Zhang and Ding, 2025). Thus, work towards identifying the bases present in the pore has not been a focus of this work.

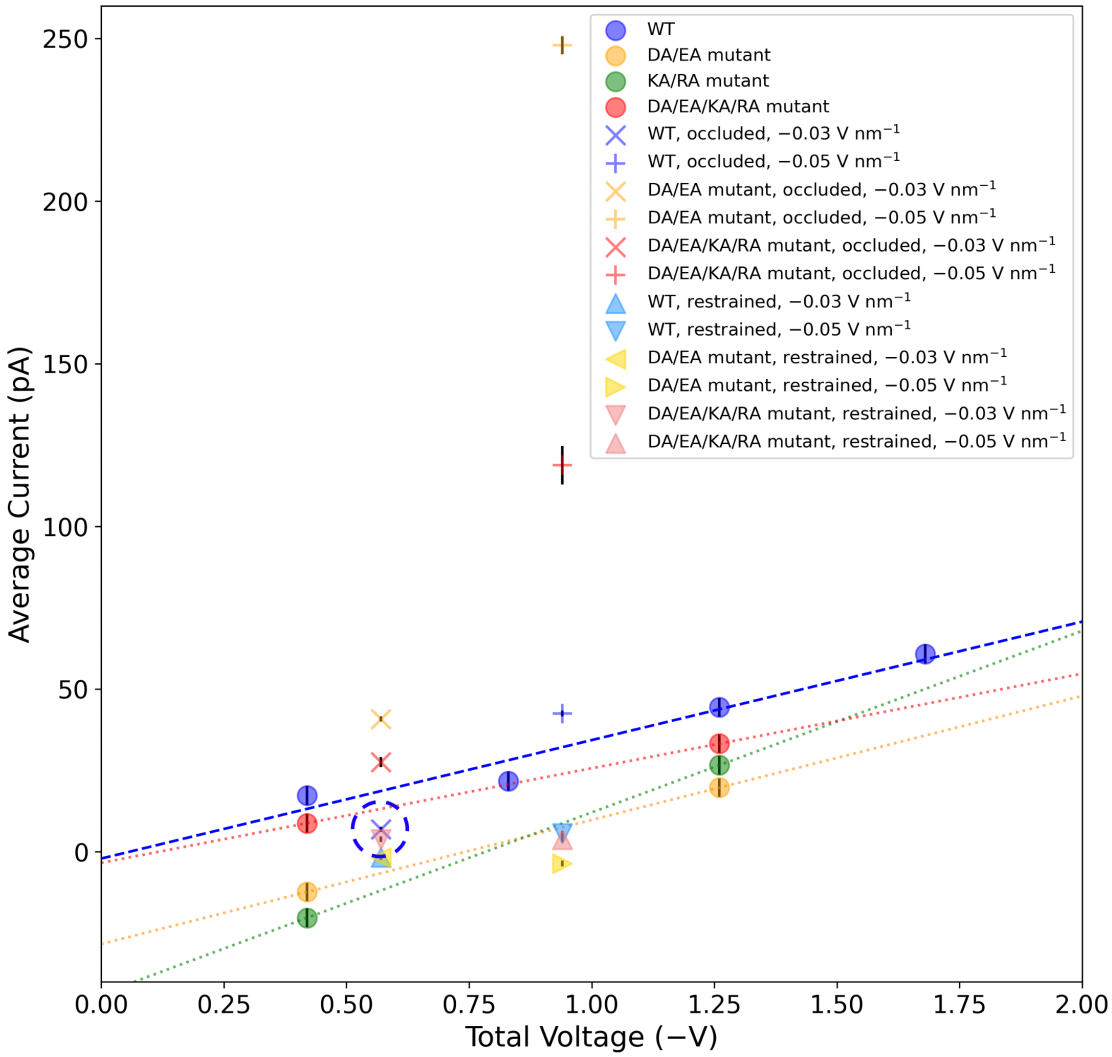



**Figure 3.30:** Plot of current against voltage for non-occluded (lines), occluded (+s and xs), and occluded-restrained (triangles) pores. Standard deviation shown as vertical black lines. Data point showing decrease in current when dsRNA occluded the WT pore in $-0.03$ V nm$^{-1}$ electric field circled.

## 3.4 Discussion

### 3.4.1 Choice of force field

The vast majority of the simulations performed in this work were 50 ns in length. This is a short simulation time, and while this reduced the computational cost it limits the interpretation of the data. Longer simulations using the CHARMM36 force field would likely see non-physical behaviour of the dsRNA, as dsRNA is not stable using CHARMM36 on the scale of hundreds of nanoseconds. The simulations could therefore be repeated using the most accurate AMBER force field for RNA, which would enable longer simulations to be run at a lower applied electric field (Kührová *et al.*, 2019). This would potentially be more representative of an experimental set up. However, the AMBER force fields are still limited in their representation of RNA, so it is difficult to know how much the results would be improved by this (Nithin *et al.*, 2025).

### 3.4.2 Effect of using C-$\alpha$ restraints in simulations

One of the major limitations of the work presented here is the use of C-$\alpha$ restraints on the backbone, which is evidently non-physiological. Their use was justified to isolate the changes caused by the chosen mutations from changes caused by the stochastic fluctuations in structure, both of which could also impact translocation speed, and thus identify design principles regarding charged amino acid residues. Additionally, the experimental data from Niedzwiecki *et al.* (2012) measuring current across FhuA showed that the pore demonstrates spontaneous gating, likely due to the remaining extracellular loops and long $\beta$-strands having the flexibility to collapse and obstruct the extracellular side of the pore. This transient occlusion of the pore has also been seen in MD simulations, and this behaviour makes it difficult to interpret how small changes to the pore affect the performance of the pore (Bond *et al.*, 2007; Chen *et al.*, 2008). Therefore, to investigate only the changes caused by the mutations of charged amino acids, the use of position restraints allowed for the removal of spontaneous gating as a confounding factor.

Further work could be done to see if the same results were observed in simulations without the applied restraints and to see if there is a correlation between RMSD of the pore and the speed of dsRNA translocation. This could have design implications such as requiring a method of reducing the flexibility of the pore to make the translocation of dsRNA more controlled and predictable.

### 3.4.3 Comparison of results to previous experimental data

The engineering of MspA to improve its potential as a DNA-sensing nanopore involved the removal of four aspartic acid residues by mutating them to asparagine residues, which removed the energetic barrier caused by the presence of those negative charges (Butler *et al.*, 2008). The work on MspA utilised a potential of mean force analysis to identify the barrier. Thus, the results of this work agree with the experimental data of Butler *et al.* (2008), with the removal of aspartic acid and glutamic acid residues in FhuA increasing the speed of translocation of dsRNA because an energetic barrier has been decreased.

Studies of the $\alpha$-helical pore protein Fragaceatoxin C (FraC) showed that the balance of positive and negatively charged residues is important in successful translocation of ssDNA, with a key mutation involving an aspartic acid residue being mutated to an arginine (Wloka *et al.*, 2016). This corresponds with the results in subsection 3.3.5, which demonstrate that the removal of positively charged residues is prohibitive to dsRNA translocation in FhuA. The engineered FraC also required distal mutations of residues to acidic residues for optimal translocation, suggesting that this may need to be considered in further FhuA designs.

### 3.4.4 Alternative engineered FhuA nanopores

Wolfe *et al.* (2016) proposed an alternative redesign of FhuA involving the removal of five extracellular loops and the removal of a number of negatively charged residues on the strands, turns, and loops of the protein. Compared to the design by Niedzwiecki *et al.* (2012), Wolfe *et al.* (2016) also propose the removal of L10.

The behaviour of this pore includes blockades of current that are not observed in other pore redesigns, and the impact of these on dsRNA is unknown.

The work presented here, combined with previous studies, suggests that FhuA is an excellent candidate for the development of a larger-diameter nanopore, and dsRNA may be a suitable substrate.

### 3.4.5 Mechanism of effect of mutations on dsRNA translocation through FhuA

Based on this work, I propose an overall mechanism for how different mutations affect the translocation of dsRNA (Figure 3.31).

Panel A shows the system set-up, with the flow of potassium (rose) and chloride (lilac) ions shown by the arrows following the electric field. The FhuA residues highlighted in blue and red are as follows: ARG212, ARG199, ARG277 (top cluster, blue); GLU210, GLU279 (top cluster, red); ARG713 (bottom cluster, blue); and GLU163, ASP179 (bottom cluster, red).

Panel B. i. through iii. follow the dsRNA translocating from the initial insertion into WT FhuA to the FhuA beginning to escape. The panels show interactions between the phosphates of the dsRNA and the blue-highlighted residues. There are also interactions between the phosphates of the dsRNA and potassium ions, which interact with the red-highlighted residues. The interactions with the blue residues allow the dsRNA to overcome the pull of the potassium ions and move towards the positive side of the membrane along the charge gradient.

Panel C shows the DA/EA pore through which translocation can happen much faster, as there are no longer any acidic (red) residues to increase the flow of potassium ions and counter the movement of the dsRNA. Additionally, these mutations lead to an increased flow of chloride ions, which helps move the dsRNA towards the positive side of the membrane. In Panel D, with the KA/RA pore, the increased flow of potassium ions is insurmountable for the dsRNA, which cannot translocate. There are no basic (blue) residues with which the dsRNA can make attractive electrostatic interactions, therefore the dsRNA is repelled from the pore.

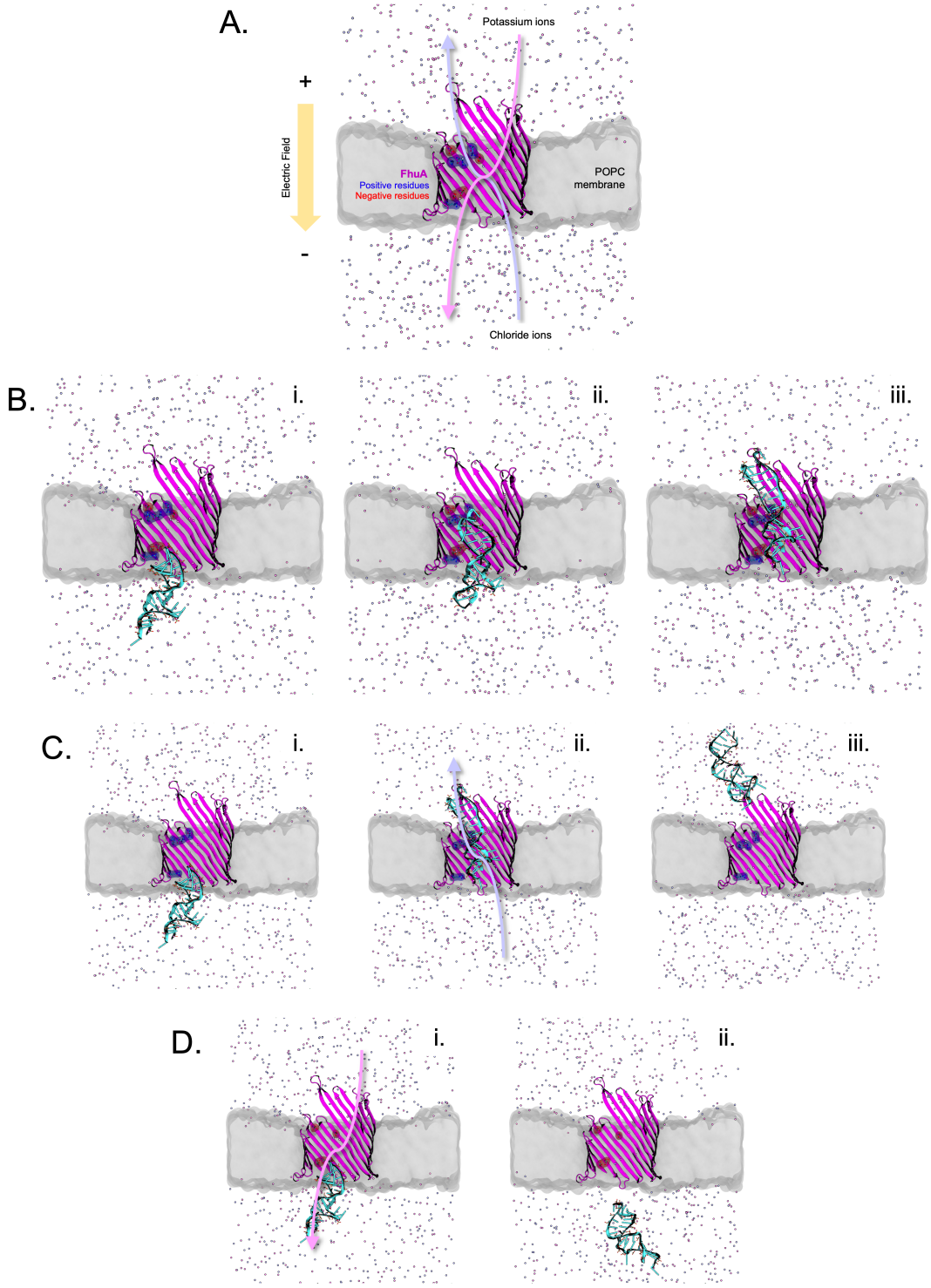



**Figure 3.31:** A. Set-up of FhuA as a nanopore, with the direction of ion flow shown by arrows.
B. Translocation of dsRNA in the WT pore.
C. Translocation of dsRNA in DA/EA pore, which occur faster than in the WT pore.
D. (non)-Translocation of dsRNA in KA/RA pore.

Overall, my results suggest that the ideal nanopore would have fewer acidic residues to slow down the movement of dsRNA, but still requires some to help anchor the dsRNA as it translocates through the pore. Based on these simulations, the rate of dsRNA movement through the nanopore is on the scale of millions of base pairs per second, compared to the ideal of 450 bases per second for DNA in current sequencing (Wang *et al.*, 2021b). Some of this can be accounted for by the higher total voltage applied to the system, but it is likely that other methods of reducing dsRNA translocation speed will need to be utilised, such as the addition of bulky amino acids to create further constriction at specific points (Rattu *et al.*, 2021a). As the movement of dsRNA is heavily impacted by the movement of ions, the design will also have to account for the specific ion concentration being used, as ion flux is dependent on ion concentration and therefore the speed of translocation is also dependent on ion concentration.

## 3.5 Conclusions

The data in this chapter contributes to the overall understanding of important design principles required when engineering nanopores for the translocation and interrogation of biological macromolecules.

Through the use of MD simulations, I have shown that the engineered FhuA ΔC/Δ4L is of a suitable size for the translocation of dsRNA, which has a more variable $xy$ cross-sectional area than dsDNA. This means that other factors than size are likely to be important when creating a constriction point, as it needs to lead to changes in translocation while not inhibiting the movement of dsRNA samples that do not have complimentary base-pairing for the full sequence. The pore has to be larger than the average $xy$ cross-sectional area of an RNA base pair to accommodate secondary structures. RNA is also more flexible in structure, and therefore minimising this in the pore will lead to less variable translocation and data that is less noisy.

When the dsRNA is present in the pore for tens of nanoseconds, there is a decrease in the calculated current when compared to the current across the non-occluded pore. This means the presence of dsRNA in the pore can theoretically be detected. At present, it is not possible to predict how the current would fluctuate dependent on the specific base pair or RNA secondary structure in the pore. Experimental data would be required to investigate this further. This is a more complex problem than deconvoluting the current signal from nanopores sequencing ssDNA, as there are a large number of permutations when considering canonical and non-canonical base pairs, in addition to bulges and other secondary structure features. Therefore this is also likely to require software to interpret the changes in signal and calculate the sequence and structure.

It is possible to affect the speed of translocation of dsRNA by changing the electrostatic surface of the pore of FhuA, as dsRNA has a high negative charge. These changes also impact the flux of charged ions moving through the pore, and thus the current in the system. Removing all of the negatively charged amino

acids lining the pore leads to a decrease in the flux of positive ions, an increase in the flux of negative ions, and an increase in translocation speed of dsRNA. Removing all of the positively charged amino acids lining the pore leads to an increase in the flux of positive ions, a decrease in the flux of negative ions, and the abolition of the ability of dsRNA to translocate through FhuA. Removing all of the charged amino acids lining the pore leads to only a moderate change in the flux of ions, similar to that observed when removing all of the negatively charged amino acids. This moderate change is likely due to the slightly higher number of negatively charged residues in the WT pore sequence. A number of other polar amino acids also contribute to the electrostatic surface and are also likely to modulate the speed of dsRNA translocation. Pi-stacking interactions with dsRNA bases not involved in canonical base-pairing (making them more accessible to the surrounding environment) are also likely to be involved, as these interactions have been observed in the simulations in this study.

Changing charged residues to alanines also changes the internal $xy$ cross-sectional area of the pore. However, as the polar residues investigated (glutamine and asparagine) and the aromatic residues (phenylalanine, tyrosine, and tryptophan) are also large and solvent accessible when the charged residues are present, increases in interactions with these residues are likely linked to the change in electrostatic profile and not the change in pore shape.

In conclusion, I have shown that ion flux and direct electrostatic interactions between the pore and the dsRNA impact the speed of movement of dsRNA. Therefore, when considering the effect of ion flux as a component of controlling the speed of dsRNA translocation, the electrostatic surface of the pore must also be included as a contributing factor. This work suggests that to slow down translocation events, allowing for detectable and interpretable changes in current to be observed, having more negatively charged amino acids than positively charged amino acids would be a suitable starting point.

## 3.6 Future directions

The conclusions of this work lead to some immediate suggestions for future research. This work has solely removed charged amino acids from their initial positions in the protein. Understanding the importance of the position of different charged amino acids could be achieved in various ways. For example, by creating charged amino acid rings, changing the positions of clusters of amino acids, and probing the partial mutations to wild-type clusters currently identified as important to translocation. In addition, the positions of charged amino acids could be reversed, with negatively charged amino acids being replaced by positively charged amino acids and vice versa. These permutations could be guided by the positions of charged amino acids in successful nanopore designs.

Additionally, this work could be extended by using different fragments of dsRNA, with a variety of secondary structures and lengths. This may affect how the dsRNA translocates, as some structures may cause more or less occlusion depending on the $xy$ cross-sectional area of the RNA at that point. This would impact the feasibility of using FhuA for the purpose of RNA secondary structure characterisation. The final studies for this would require experimental data, to examine if the design improvements suggested by MD simulations are effective and sufficient for dsRNA translocation at a rate amenable to the detection and decoding of current fluctuations in a nanopore system.

# 4

# Structural studies of the ribosome of *Nitrosopumilus maritimus*

## Contents


**4.1 Introduction . . . . . 122**
4.1.1 Ribosome structure and function . . . . . 123
4.1.2 History of structural studies of the ribosome . . . . . 126
4.1.3 Variation of ribosomes across the domains of life . . . . 128
4.1.4 Ribosomes in *N. maritimus* cells . . . . . 131
4.1.5 Cryo-electron microscopy and model building . . . . . 132
4.1.6 Aim of this chapter . . . . . 134
**4.2 Materials and Methods . . . . . 135**
4.2.1 Production of *N. maritimus* ribosomes . . . . . 135
4.2.2 Cryo-EM . . . . . 137
4.2.3 Model building . . . . . 140
4.2.4 Other analyses . . . . . 141
**4.3 Results . . . . . 142**
4.3.1 50S map resolution and features . . . . . 142
4.3.2 70S map resolution and features . . . . . 142
4.3.3 Universal proteins in the LSU . . . . . 145
4.3.4 Archaea-eukarya proteins in the LSU . . . . . 159
4.3.5 rRNA model building in the LSU . . . . . 168
4.3.6 Unmodelled density in the LSU . . . . . 171
**4.4 Discussion . . . . . 174**
4.4.1 Ribosomal protein content . . . . . 174
4.4.2 Structural features . . . . . 176
4.4.3 Auto-building . . . . . 181
**4.5 Conclusions . . . . . 182**
**4.6 Future work . . . . . 184**

# Statement of Authorship

The work in this chapter was conducted by the doctoral candidate, with the exception of the production of data as described in subsection 4.2.1 (growth of cells), subsubsection 4.2.2.1 (preparation of grids), and subsubsection 4.2.2.2 (data collection) that was performed by Lennart Pagani (Utrecht University and William Dunn School of Pathology, University of Oxford), Sofie van Dorst (Utrecht University and William Dunn School of Pathology, University of Oxford), and Andriko von Kügelgen (William Dunn School of Pathology, University of Oxford, and MRC Laboratory for Molecular Biology, Cambridge, UK). The work in subsubsection 4.2.2.3, subsubsection 4.2.3.1, and subsubsection 4.2.3.3 (ribosome modelling) was performed in collaboration Lennart Pagani (built 600 residues of 23S rRNA), Sofie van Dorst (worked on refinement of small subunit), and Andriko von Kügelgen (worked on refinement of small subunit) as detailed in those sections, and is covered in subsection 4.3.1, subsection 4.3.2, and subsection 4.3.5. All other work (data processing, modelling, refinement, bioinformatics) was conducted by the doctoral candidate.

## 4.1 Introduction

Ribosomes are a fundamental component of all living cells, as they are required for the assembly of all proteins. They are the link between the nucleic acid code and the amino acid code, enabling the language of RNA to be turned into the language of proteins. Hence, their role is named translation. In the cell, proteins are continuously turned over, meaning that there needs to be a constant activity of the ribosomes to replenish the contents of the cell and maintain a steady state, or growth phase. Thus, the function of ribosomes is of vital importance, and this underlies our desire to understand them.

## 4.1.1 Ribosome structure and function

### 4.1.1.1 Ribosome composition

Ribosomes are composed of a large and a small subunit, both of which comprise long strands of RNA (rRNA) and a number of accessory proteins (ribosomal proteins, or r-proteins) (Bashan *et al.*, 2000; Brimacombe *et al.*, 2000; Ban *et al.*, 2000b). This makes the complex a ribonucleoprotein (RNP). The subunits are known both by their relative size (large and small) and by their sedimentation coefficient in Svedberg (S). In prokaryotes, the large subunit (LSU), 50S, consists of the 23S rRNA, the 5S rRNA, and at least 19 proteins (Ramakrishnan *et al.*, 2000; Yusupov *et al.*, 2001). This subunit contains the A, P, and E sites of the ribosome (Frank *et al.*, 1995). The small subunit (SSU), 30S, consists of the 16S rRNA and at least 15 proteins in prokaryotes (see Figure 4.1). This subunit is responsible for the binding of the mRNA molecule and the presentation and recognition of the codon, allowing translation of the nucleic acid sequence into amino acids (Frank *et al.*, 2000; Ogle *et al.*, 2001). Together, the prokaryotic LSU and SSU form the 70S ribosome.

The eukaryotic ribosomal subunits are both larger, and form a larger total complex, 80S, which is demonstrated by the larger sedimentation coefficient giving them their names. The eukaryotic LSU, 60S, is formed from three rRNA molecules and at least 47 proteins; the SSU, 40S, consists of the 18S rRNA and at least 33 proteins (Ben-Shem *et al.*, 2010). Many of these proteins that are additionally present in eukaryotic ribosomes are required for nuclear export of the rRNA molecules (Ruland *et al.*, 2021).

The ribosome also has a "stalk", which is important for recruiting translation factors. The stalk is formed from more unstructured rRNA and a number of proteins, and it is highly flexible. The stalk extends out from the 50S subunit; however, detailed structural information is lacking due to the flexibility of the segment.

### 4.1.1.2 Role of the ribosome

The overall role of the ribosome is to catalyse the formation of peptide bonds between amino acids, in the order encoded by an mRNA molecule. The first step

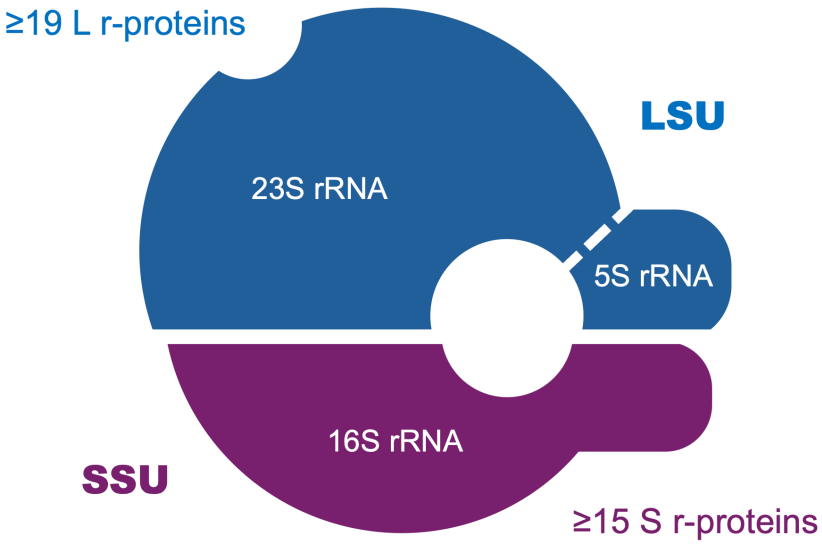



**Figure 4.1:** Overview of prokaryotic (70S) ribosome composition, depicted as a simple cartoon. The circular gap in between the LSU and SSU represents where the mRNA threads through the ribosome, and the cognate tRNA is matched up to the presented triplet codon. The cut out at the top left of the LSU represents the polypeptide exit tunnel. The dashed line in the LSU represents the segment based around the 5S rRNA and associated proteins. The stalk is not represented in this cartoon.

of translation is the ribosome binding to an mRNA molecule. The start codon is aligned in the ribosome A site, and a $tRNA^{MET}$ molecule is recruited. The ribosome moves along the mRNA by one triplet codon, with the first tRNA moving to the P site, allowing the next complementary charged tRNA to bind in the A site. After interrogating the complementarity between the codon and the tRNA, the peptide bond formation is then catalysed between the amino acids on the tRNA molecules at the peptidyl transferase centre (PTC). The empty $tRNA^{MET}$ molecule is then released from the E site. By continuing to shift along the mRNA and recruit more tRNAs, the nascent amino acid chain is extended and extruded by the ribosome through the exit tunnel (see Figure 4.2).

Both biochemical and structural approaches have been used to understand the structure and function of the ribosome, and there is a large body of research that has contributed to our identification and understanding of a universal architecture of ribosomes across all domains of life (Petrov *et al.*, 2014; Pape *et al.*, 2000;

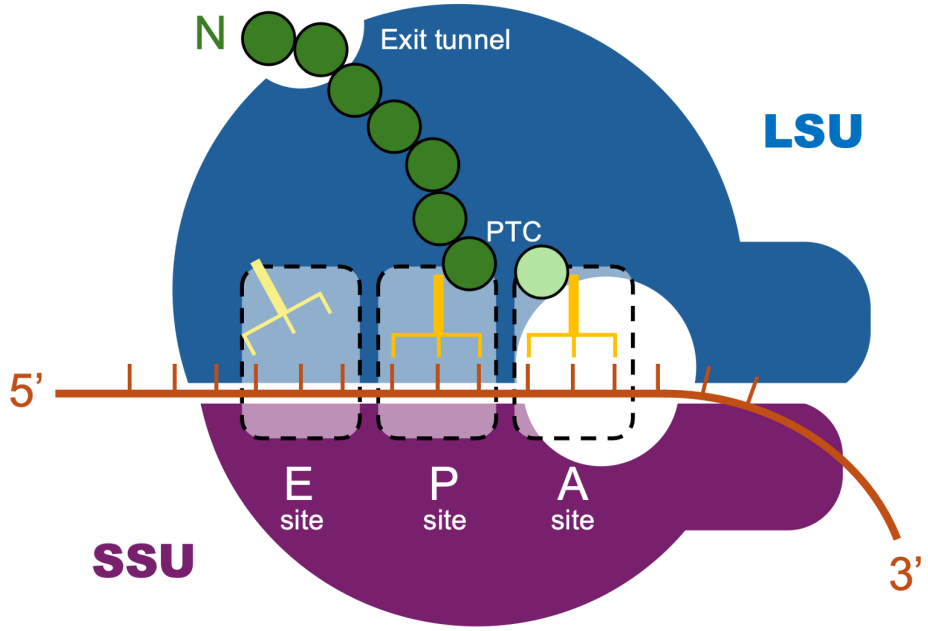



**Figure 4.2:** Overview of ribosome function. mRNA (orange) binds to the SSU, and the LSU forms a ribosome particle ready to translate. In the A site, a triplet codon is presented, and charged tRNA molecules (pale orange) are tried until a cognate tRNA is found. This allows movement of the ribosome along the mRNA, placing the mRNA-tRNA pair in the P site. Another cognate tRNA is then found for the codon now in the A site, and the peptidyl transferase centre (PTC) catalyses the formation of a peptide bond between the amino acids (green circles) on the charged tRNAs in the A and P site. This leaves the tRNA in the P site without an amino acid, and it moves to the E site and exits as it is now not charged (pale yellow). The tRNA in the A site, with the extending peptide chain, then moves to the P site and the ribosome continues to elongate the peptide chain and translocate along the mRNA. The peptide chain exits the ribosome through the protein-lined exit tunnel.

Ramakrishnan, 2002; Noller, 2024).

#### 4.1.1.3 Importance of RNA

All of the stages of peptide bond formation in ribosomes are catalysed by the nucleic acids, with no involvement from the proteins (Noller *et al.*, 1992; Green and Noller, 1997; Nissen *et al.*, 2000). This makes the ribosome a ribozyme, and a "living fossil" reminder of the RNA world from which early life evolved (Cech, 2000; Petrov *et al.*, 2015; Bowman *et al.*, 2020). There are no proteins within 10 Å of the PTC, providing more evidence that the RNA must be solely involved (Rodnina *et al.*,

2005; Beringer and Rodnina, 2007). Furthermore, it is the specific 3D structure of the rRNA that allows this function. A highly conserved adenine residue acts as a general acid-base catalyst, leading to a concerted proton shuttle mechanism that forms the peptide bond between the amino acid on the tRNA in the A site, and the peptide chain on the (then released) tRNA in the P site (Rodnina *et al.*, 2006; Beringer and Rodnina, 2007; Leung *et al.*, 2011).

#### 4.1.1.4 Role of the proteins

The role of ribosomal proteins (r-proteins) is not fully understood, although some are known to play non-ribosomal roles in eukaryotic cells (Kang *et al.*, 2021; Warner and McIntosh, 2009; Razi and Ortega, 2017). While r-proteins are not strictly necessary, as they do not participate in the formation of peptide bonds, they are nevertheless important to the function of the ribosome (Nikolay *et al.*, 2015; Nikolaeva *et al.*, 2021). Some proteins are involved in ribosomal assembly, such as those that chaperone the 5S rRNA before it assembles into the large subunit (Korepanov *et al.*, 2007; Gamalinda and Woolford Jr, 2015; Huang *et al.*, 2020). Others are known to play roles in co- and post-translational protein folding on the ribosome surface (Aseev *et al.*, 2024; Ahn *et al.*, 2022).

### 4.1.2 History of structural studies of the ribosome

The "dense granules" observed in cells by George Palade in 1955 were soon named ribosomes, and it was quickly determined that ribosomes were the site of protein synthesis in cells (Palade, 1955; Roberts, 1958). Palade was awarded one third of the 1974 Nobel Prize for Physiology or Medicine for this work, and it was only the start of a wide range of investigations into the structure and function of the ribosome (Porter and Novikoff, 1974).

#### 4.1.2.1 The first atomic structures

X-ray crystallography has been the dominant method for determining molecular structures at atomic resolution since the structure of myoglobin was solved by Kendrew *et al.* (1958), and thus the first step towards solving the structure of

the ribosome was the production of a diffracting ribosome crystal by Yonath *et al.* (1987). After many lower-resolution structures, the structure of the 50S subunit from the archaeon *Haloarcula marismortui* was solved to 2.4 Å by Ban *et al.* (2000a), and the structure of the 30S subunit from the bacteria *Thermus thermophilus* was solved to 3.0 Å by Wimberly *et al.* (2000). This enabled the atomic structure of the 70S ribosome from *T. thermophilus* to be solved by Yusupov *et al.* (2001), by using the previously solved structures to build atomic detail into the 5.5 Å map. The combination of these efforts was rewarded with the 2009 Nobel Prize for Chemistry, as it enabled an exponential growth in the structural and functional understanding of the ribosome (Van Noorden, 2009).

#### 4.1.2.2 Cryo-EM structures of ribosomes

The cryo-electron microscopy (cryo-EM) resolution revolution was partly spurred on by the work of Bai *et al.* (2013) on the previously-solved *T. thermophilus* ribosome structure. At over 2 MDa, a ribosome is easily identifiable in micrographs as it is well over the minimum size for aligning particles above noise, making it an ideal subject for electron microscopy (Henderson, 1995). The large number of phosphorous atoms in the RNA increases the contrast due to the higher atomic mass. Additionally, the conserved and structured core of the ribosome makes an easy target for structure determination, as features will be well-resolved after averaging across many particles that are in the same conformation. Since the resolution revolution, cryo-EM has become a favoured method for solving ribosome structures as there is no requirement for sample crystallisation (Jobe *et al.*, 2019). This has led to a relative explosion in ribosome structures, mostly focused on a small number of bacterial species and yeast. However, getting complete ribosome structures remains a challenge due to the flexible ribosomal stalk, and the flexibility of peripheral proteins. This is particularly clear in the low number of full eukaryotic and archaeal ribosome structures.

#### 4.1.2.3 Archaeal ribosome structures

Since the structure of the 50S subunit of *H. marismortui* was solved in 2000, very few archaeal ribosome structures have been published. The most notable

of these are detailed here, and were all solved using cryo-EM. The 50S subunit of *Methanothermobacter thermautotrophicus* was solved to 6.6 Å (Greber *et al.*, 2012). Nürenberg-Goloub *et al.* (2020) solved the 30S subunit of *Thermococcus celer* to 2.8 Å, by purifying a complex of 30S with a ribosomal recycling factor. The *Pyrococcus abyssi* 30S subunit, bound to mRNA, was solved to 2.7 Å by Kazan *et al.* (2022). The work of Sas-Chen *et al.* (2020) showed the cryo-EM structure of the 70S ribosome from *Thermococcus kodakarensis* (2.55 Å). In 2023, Wang *et al.* (2023b) resolved 10 states of the *Saccharolobus acidocaldarius* 70S ribosome, with resolutions ranging between 2.7 and 5.7 Å. The 70S structures of both *Pyrobaculum calidifontis* (2.36 Å) and *Pyrococcus furiosus* (2.84 Å) were solved in 2025 (Nissley *et al.*, 2025; Garcia-Campos *et al.*, 2025). This research shows a steady increase in interest of the ribosomes from this domain, and how they fit in with the tree of life connecting from LUCA to eukaryotes.

### 4.1.3 Variation of ribosomes across the domains of life

#### 4.1.3.1 The minimal ribosome

Even before the structure of the ribosome was solved, proteomic and genomic analysis was being performed to determine the "minimal ribosome", the segments of the rRNA and the r-proteins that were conserved across all domains of life and were thus assumed to be essential for ribosomal structure and function. This was first performed by Mears *et al.* (2002), based solely on genomic data, and was later extended by Mushegian (2005) using structural data to investigate which r-proteins interacted with the minimal sections of rRNA. This method allowed the determination of proteins that were minimal for the domain incorporating Archaea and Eukarya, as well as those which were universal.

#### 4.1.3.2 Comparing rRNA sequence and structure

16S rRNA sequencing has long been used as a means of identifying an unknown species and constructing evolutionary relationships between species (Woese *et al.*, 1975; Clarridge, 2004; Johnson *et al.*, 2019). Although it is not an infallible method,

there is nevertheless a clear relationship between rRNA sequence conservation and the genetic relationship between species. The work by Mears *et al.* (2002) identified the highly conserved sequences within the rRNA genes, and it has also since been shown that this links to many conserved 3D rRNA folds (Bernier *et al.*, 2018; Rivas, 2021). The importance of the rRNA segment in the ribosome function is indicated by the degree of conservation of each segment.

Expansion segments are locations within the rRNA genes that are commonly expanded in archaea and eukaryotes, and exist in between conserved segments. These expansions are 5 to 10 nucleotides in most archaea, and far larger in eukaryotes (Fujii *et al.*, 2018; Penev *et al.*, 2020; Rauscher and Polacek, 2024).

Work by Mokdad *et al.* (2006) identified G/U as a common wobble pair in the ribosome. These interactions are important in the rRNA tertiary structure due to their role in long-range contacts between distant helices, and many of them are highly conserved.

#### 4.1.3.3 Comparing protein content

By comparing the distribution of r-proteins across domains, kingdoms, and phylums, researchers have been able to establish a universal common core of proteins (Melnikov *et al.*, 2012). This includes 19 proteins in the LSU, and 15 proteins in the SSU.

The naming of ribosomal proteins is based on the numbering for *E. coli* ribosomes, and r-proteins in other species that were homologous were given the same number (Wittmann *et al.*, 1971; Wittmann-Liebold, 1986; Ban *et al.*, 2000b). R-proteins without homology in *E. coli* were given different numbers, without comparison to other species. Others did not follow this convention. The differing names caused difficulty in determining relationships between proteins and thus Ban *et al.* (2014) proposed a naming convention that allowed for distinction between proteins that were universal (prefixed uL, part of the universal common core), those found only in bacteria (bL), those found in archaea and eukarya (aL), and those found only in eukarya (eL). This system allowed for easier comparison between

ribosomes. It should be noted that not all bL, aL, and eL proteins are found in all phyla within those domains.

There are no ribosomal proteins that are found in bacterial and archaeal ribosomes and not in eukaryotic ribosomes (Lecompte *et al.*, 2002). Also, there are none that are unique to all archaea that are not found in eukarya (Bowman *et al.*, 2020). This was a big piece of evidence in support of the two-branch model of the tree of life, in which the domain of Eukarya does not branch off at the last universal common ancestor (LUCA) and instead evolve later from the archaeal branch, within the Asgardarchaea (Hug *et al.*, 2016; Williams *et al.*, 2020; Kovacs *et al.*, 2018). This relationship is mapped in Figure 4.3.

There are some ribosomal proteins that are widespread (although not universal) across archaea, such as those which are common to the Crenarchaeota phylum (Feng *et al.*, 2014).

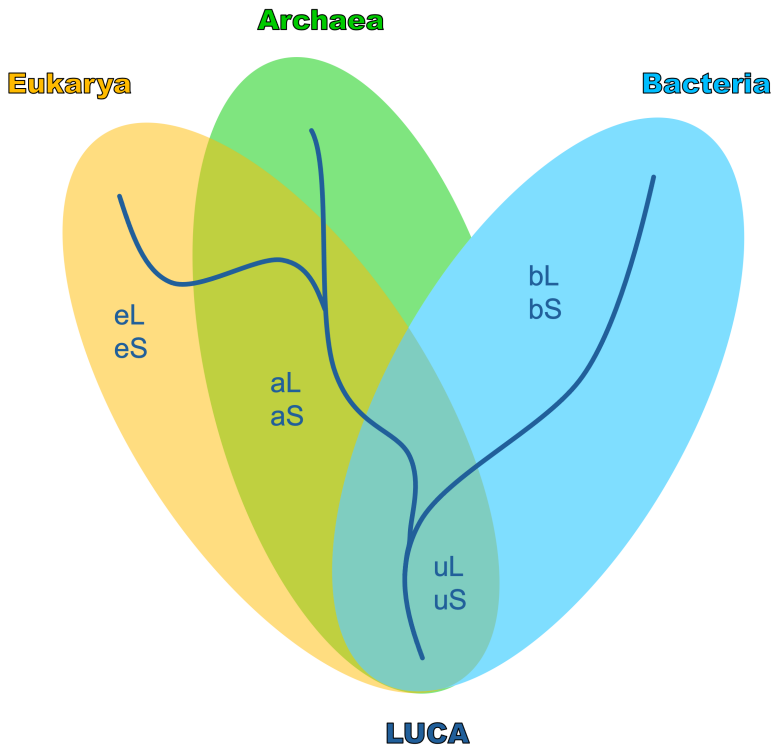



**Figure 4.3:** The division of conserved ribosomal proteins across domains of life, with ovals showing the groups of common proteins between each type of ribosome. Adapted from Kovacs *et al.* (2018).

### 4.1.4 Ribosomes in *N. maritimus* cells

#### 4.1.4.1 *Nitrosopumilus maritimus*

*Nitrosopumilus maritimus* was discovered and isolated from sediment in an aquarium tank by Könneke *et al.* (2005). The genome was sequenced by Richardson and Stahl (2007), and then later work by Walker *et al.* (2010) allowed for the classification of the species as an archaeon of the phylum Thermoproteota. The cells are only 0.2 µm in diameter, making *N. maritimus* one of the smallest known free-living organisms. *N. maritimus* is ubiquitous in seawater, albeit in very low densities. They are vital to the global nitrogen cycle as a member of the class Nitrososphaeria, part of the ammonia-oxidising archaea (AOA). AOA have the ability to oxidise ammonia to nitrite, even at very low ammonia concentrations (see Figure 4.4). As *N. maritimus* cells live in oxygen-depleted habitats, the oxygen required for this ammonia oxidation must be produced by another metabolic pathway within the cell, which is thought to involve the production of oxygen and dinitrogen (Kraft *et al.*, 2022).

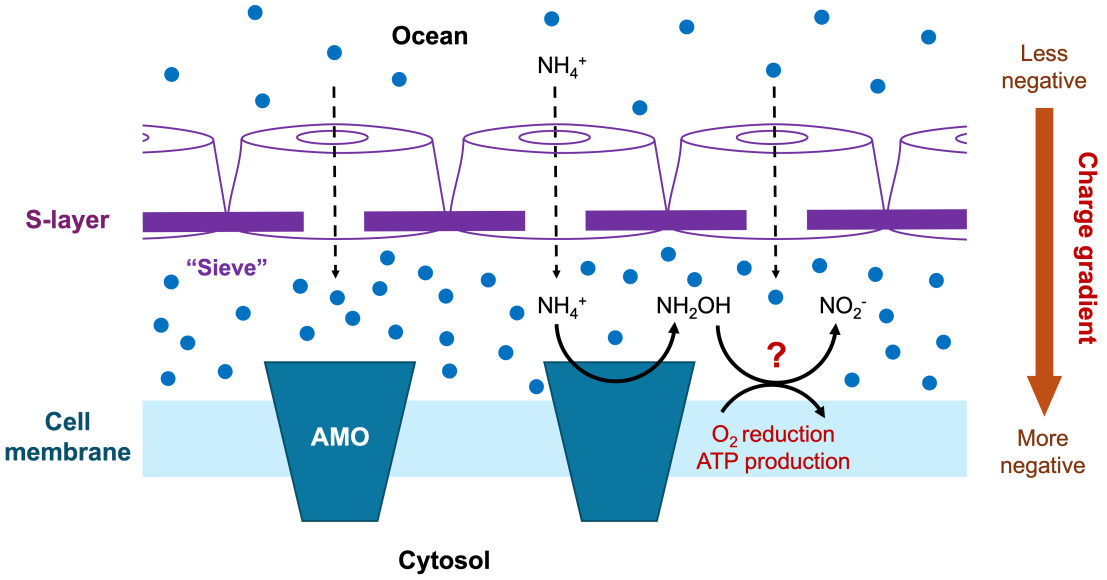



**Figure 4.4:** Ammonia oxidation pathway in AOAs, specifically *N. maritimus*. AMO is the ammonia monooxygenase protein, which catalyses the first reaction of ammonia ($NH_4^+$ to $NH_2OH$). The protein involved in the conversion to nitrate is unknown. Blue circles represent ammonia molecules, being sieved from the ocean by the charge gradient across the S-layer protein. Adapted from von Kügelgen *et al.* (2024).

#### 4.1.4.2 *N. maritimus* ribosomes

For other research in this thesis, whole-cell tomograms of *N. maritimus* were taken for the *in situ* determination of the S-layer structure, as shown in Chapter 5, subsection 5.3.1 (von Kügelgen *et al.*, 2024). These tomograms contained a large quantity of data that was not utilised in the S-layer research, and thus using this data was a logical step. Ribosomes were easily visible in the whole-cell tomograms (Figure 4.5). As few archaeal ribosome structures exist in the literature, studying the *N. maritimus* ribosome filled a research gap. Initial work on a sub-tomogram averaging structure was performed outside of the work presented here, and the results suggested that a single-particle approach would be the most appropriate starting point for structural studies.

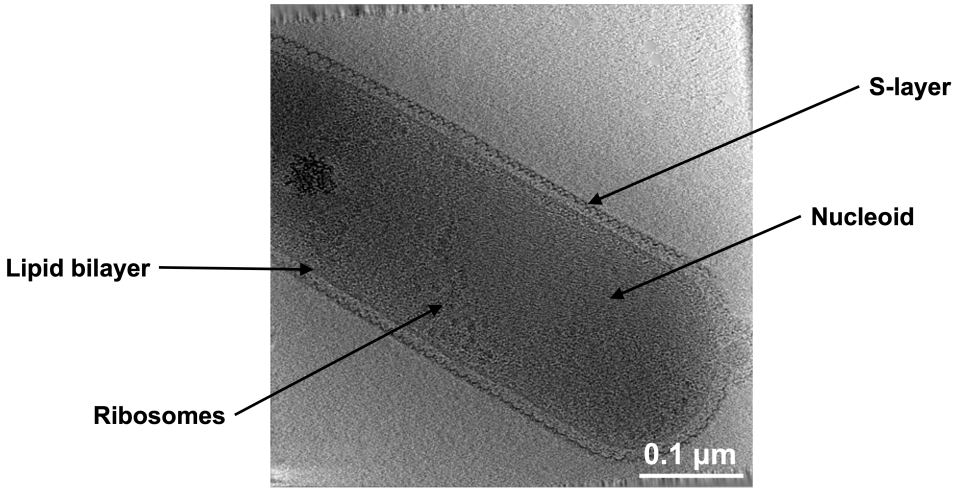



**Figure 4.5:** Tomogram slice of a *N. maritimus* cell, with observable features labelled.

### 4.1.5 Cryo-electron microscopy and model building

#### 4.1.5.1 Cryo-electron microscopy

The cryo-EM resolution revolution has led to cryo-EM becoming a major method for the atomic-resolution structure determination of biological molecules (see Chapter 1, subsection 1.2.1) (Kühlbrandt, 2014). As mentioned above (subsubsection 4.1.2.2), this has involved resolving the structures of ribosomes. Ribosomes are highly

abundant in cells, making sample preparation for cryo-EM simpler than for subjects that require overexpression. Additionally, when considering the biological context, ribosomes are large enough to see and identify in tomograms without labelling (Russo *et al.*, 2022). This makes cryo-EM a more favourable method for studying ribosomes, and ribosomes a favourable target for cryo-EM and cryo-ET (Razi *et al.*, 2017).

#### 4.1.5.2 ModelAngelo

ModelAngelo was developed by Jamali *et al.* (2024) as a novel machine-learning method of sequence-led and *de novo* building of protein and nucleic acid models into high-resolution cryo-EM structures. It uses a graph neural network to build up structures from fragments in cryo-EM maps with a resolution equal to or better than 3.5 Å. By providing sequence data, the fragments can be identified as specific proteins with a high level of accuracy and connected. It is also able to build models without a sequence provided, albeit with lower accuracy. ModelAngelo has been shown to outperform humans in model building, allowing time to be spent on structure refinement and interpretation (Singh, 2024).

#### 4.1.5.3 AlphaFold2

One of the most transformative tools in recent structural biology, AlphaFold2 (AF2) provides predicted protein structures for any protein sequence (Jumper *et al.*, 2021; Bertoline *et al.*, 2023). As it is trained on the PDB, which contains r-protein structures, it is able to predict the structure of many r-proteins with a high degree of accuracy. It has also been shown to accurately predict novel folds outside of the training set. However, AF2 struggles with intrinsically disordered regions of proteins, or those where structures may be dependent on a ligand or particular environment. This means that flexible regions of r-proteins, such as those on the surface of the ribosome, are likely to have low accuracy when modelled by AF2 (Agarwal and McShan, 2024). Additionally, AF2 produces several models, even though proteins are dynamic and have many functional conformations (Perrakis and Sixma, 2021).

#### 4.1.5.4 Predicting RNA structure

While protein structural prediction has increased in power in recent years, the prediction of tertiary RNA structure remains a challenge, as mentioned in Chapter 3, subsection 3.1.3. RNA accounts for around 60% of the 70S ribosome, and performs all of the catalytic functions, so knowledge of the rRNA structure is of vital importance to understanding the function of the ribosome (Alberts *et al.*, 2002). RoseTTAFoldNA (Baek *et al.*, 2024) has made a big step into the journey of accurate RNA structural prediction, however increases in the power of such methods is likely to depend on having an increase in the amount of data available for training (Li *et al.*, 2020; Vicens and Kieft, 2022).

### 4.1.6 Aim of this chapter

This chapter investigates the structure of the ribosome from *N. maritimus* using cryo-EM. This will allow the confirmation of the positioning and structure of both universal and peripheral r-proteins, and the structure of the rRNA core. By using a variety of techniques to build a model, the benefits and limitations of each method can be evaluated in the context of building large biological complexes into density maps of variable resolution. Additionally, the presence or absence of r-proteins can be compared to the structures of other arterial ribosomes, allowing another part of the ribosomal tree of life to be filled in for the domain of archaea. Of the other published archaeal ribosome structures, *N. maritimus* is in the same phylum as two of them, but a different class. The other four archaea are in a different kingdom. This makes *N. maritimus* evolutionarily distant from these archaeal species. For this work, I will be referring to ribosomal proteins according to the names following the guidelines from Ban *et al.* (2014) and Kovacs *et al.* (2018).

## 4.2 Materials and Methods

### 4.2.1 Production of *N. maritimus* ribosomes

#### 4.2.1.1 Cultivation of *N. maritimus* cells

Performed by members of the Bharat group: Lennart Pagani (Utrecht University and William Dunn School of Pathology, University of Oxford), Sofie van Dorst (Utrecht University and William Dunn School of Pathology, University of Oxford), and Andriko von Kügelgen (William Dunn School of Pathology, University of Oxford and MRC Laboratory for Molecular Biology, Cambridge, UK). A culture of *N. maritimus* (SCM1 strain) was kindly provided by Dr. Felix Elling and Prof. Ann Pearson (Harvard University). Cells were grown in the dark at 28 °C, without shaking, following previously published protocols (Leavitt *et al.*, 2023; Martens-Habbena *et al.*, 2009). The cells were only handled at room temperature, and were only in the light for short periods of time. The growth of cells was monitored using a previously reported nitrite detection assay (Könneke *et al.*, 2005). Cells were propagated by transferring 5% volume during the late log phase or the early stationary phase (approximately 14 days) in fresh modified synthetic *Crenarchaeota* (SCM) media. SCM media was prepared by dissolving 0.44 M NaCl, 20 mM $MgSO_4{\cdot}6H_2O$, 25 mM $MgCl_2{\cdot}6H_2O$, 10 mM $CaCl_2{\cdot}2H_2O$, and 0.84 S, $NH_4Cl$, and $KH_2PO_4$ were prepared and autoclaved, and added to the media with final concentrations of 1 mM, 10 mM, 1 mM, and 14.7 µM respectively. FeNaEDTA was filter-sterilised and added to the media with a final concentration of 7.5 µM. Finally, 1 mL of trace elements solution (100 mM HCl, 0.5 mM $H_3BO_3$, 0.5 $MnCl_2{\cdot}4H_2O$, 0.8 mM $CoCl_2{\cdot}6H_2O$, 0.1 mM $NiCl_2{\cdot}6H_2O$, 0.01 mM $CuCl_2{\cdot}2H_2O$, 0.5 mM $ZnSO_4{\cdot}7H_2O$, 0.15 mM $Na_2MoO_4{\cdot}2H_2O$) was added per litre of media. Cell growth was followed using the Griess assay, by measuring the nitrite concentration with a spectrophotometer. The assay was performed by making serial dilutions of the culture, to ensure the accurate measurement of nitrite concentrations below 200 µM. Griess reagent (0.54 mM N-(1-naphthyl)ethylenediamine, 5.8 mM sulphanilamide, 1.6 M orthophosphoric acid) was added to the dilution series with a 1:1 ratio by

volume, and subsequently incubated in the dark for 30 min. The absorbance at 540 nm was measured using a plate reader, and the nitrite concentration was then determined through a standard curve.

#### 4.2.1.2 Isolation of the ribosome

Performed by Lennart Pagani and Andriko von Kügelgen. 4 L of *N. maritimus* (SCM1 strain) cells were harvested by centrifuging at 15,000 *g* at 4 °C for 30 min, in four 1 L batches. Samples were henceforth continuously kept on ice.

The media was decanted, and the cell pellets were resuspended in 1.5 mL buffer (20 mM pH 7.5 Tris, 300 mM $NH_4Cl$, 10 mM $MgCl_2$, 0.5 mM EDTA, 6 mM $\beta$-mercaptoethanol, 10 u/mL RNase inhibitor (SUPERase·In$^{TM}$, Invitrogen), 1 protease cocktail tablet (cOmplete$^{TM}$, Roche), 2.5 µg/mL DNaseI (Sigma Aldrich)) to a concentration of approximately 3 mg/mL (w/v). The resuspensions were then collectively pelleted by centrifugation as previously. Cells were lysed either by sonication or high-pressure homogenisation, after being resuspended in 6 mL of lysis buffer (100 mM $NH_4Cl$, 100 mM KOAc, 20 mM pH 8.5 Tris, 40 mM $MgCl_2$, 0.1 mM $CaCl_2$, 2 mM dithiothreitol, 2 protease cocktail tablets (cOmplete$^{TM}$, Roche), 50 µg/mL DNaseI (Sigma Aldrich)). Cell lysis by sonication was achieved through 10 rounds of sonicating at 15 kHz for 30 s, followed by 30 s of cooling the sample without sonicating. Homogenisation at 2,000 psi for 1.5 min was used as an alternative for cell lysis. Cell debris was pelleted by centrifuging at 30,000 *g* at 4 °C for 45 min. The supernatant was subsequently ultracentrifuged at 100,000 *g* at 4 °C for 1 hour, resulting in a pellet that contained the ribosome (predominantly the LSU alone).

Performed by Lennart Pagani, Sofie van Dorst, and Andriko von Kügelgen. An alternative resuspension buffer (20 mM pH 8.5 Tris, 100 mM $NH_4Cl$, 25 mM $MgCl_2$, 100 mM KOAc, 0.1 mM $CaCl_2$, 2 mM dithiothreitol, 10 u/mL RNase inhibitor (SUPERase·In$^{TM}$, Invitrogen), 2 protease cocktail tablets (cOmplete$^{TM}$, Roche), 50 µg/mL DNaseI (Sigma Aldrich)) was used in an attempt to isolate the entire ribosomal complex by promoting subunit association. The crude ribosome

pellet was purified as described above. The samples were ultracentrifuged at 100,000 *g* at 4 °C for 1.5 hours. This yielded 70S ribosome particles.

### 4.2.2 Cryo-EM

#### 4.2.2.1 Grid preparation

Performed by Lennart Pagani, Sofie van Dorst, and Andriko von Kügelgen. Grids were prepared by applying 2.5 µL of sample to grids (Quantifoil R2/1 Cu 300 mesh, graphene oxide coated) that were glow discharged. The sample was allowed to adsorb for 60 s, blotted for 3 s at a force of −5, before the grid was plunge-frozen into liquid ethane using the Vitrobot Mark IV (ThermoFisher). The blotting chamber was maintained at 100% humidity and 10 °C. Vitrified grids were clipped into autogrid rings and stored in liquid nitrogen.

#### 4.2.2.2 Data collection

Performed by Lennart Pagani, Sofie van Dorst, and Andriko von Kügelgen. Cryo-EM data for single-particle analysis (SPA) was collected on an FEI Titan Krios (ThermoFisher) operating at 300 kV, equipped with a Gatan K3 direct electron detector and a quantum energy filter with 20 keV slit width, using EPU software. 5226 movies were collected of the *N. maritimus* ribosome LSU, with each movie consisting of 40 micrograph frames with a total exposure of approximately 53 $e^-/Å^2$. 6059 movies of the 70S ribosome were collected with 3.5 s exposure at a total dose of approximately 43 $e^-/Å^2$ over 40 frames. Micrographs were captured with defoci between −0.85 µm and 2.0 µm, and a sampling pixel size of 0.55 Å in counting super-resolution mode. AIFS collection mode was used to track, and later correct for, image beam shift-induced aberrations.

#### 4.2.2.3 Single-particle analysis

Single-particle analysis (SPA) of the LSU was performed in collaboration with Lennart Pagani following the RELION 3.1 pipeline (Scheres, 2012); SPA of the 70S ribosome was performed in collaboration with Sofie van Dorst and Andriko von Kügelgen in RELION 4.0 beta (Kimanius *et al.*, 2021). Movies were motion

corrected and dose weighted using MotionCor2 (Zheng *et al.*, 2017b). The contrast transfer functions (CTF) of the micrographs were estimated using CTFFIND4, and 4895 micrographs were selected after visual assessment (Rohou and Grigorieff, 2015). The micrographs were binned to give a Fourier-cropped pixel size of 1.10 Å, equivalent to the physical pixel size. Ribosome or LSU particles were picked using the Topaz automated particle-picking pipeline; firstly, approximately 650 particles were picked manually by and used to train Topaz, and Topaz was subsequently used for automated particle-picking (Bepler *et al.*, 2019). All newly-picked particles were then put through two rounds of reference-free 2D classification, and one round of 3D classification with a previously determined negative stain EM map used as a reference (unpublished, Pagani, 2021). For the LSU, this led to approximately 200k particles being selected for further alignment using 3D auto refinement; for the 70S ribosome, approximately 61k particles were selected.

After CTF refinement and Bayesian polishing (Zivanov *et al.*, 2019), post-processing of the LSU map yielded a resolution of 2.34 Å. Particles were unbinned to surpass physical Nyquist and utilise the super-resolution data, then the refinements were reiterated (Feathers *et al.*, 2021). This produced a final map of the LSU at a resolution of 2.30 Å, according to the gold-standard FSC criterion of 0.143 (Henderson, 2013). For the 70S ribosome, a final map was calculated with a global resolution of 3.3 Å. After focused multi-body refinement (Nakane and Scheres, 2021), differences in local resolution were obtained using RELION, showing that the LSU from the 70S ribosome yielded a local resolution of 3.1 Å, and the SSU a local resolution of 3.6 Å. The processing pipeline for the 70S ribosome is shown in Figure 4.6.

#### 4.2.2.4 Visualisation

UCSF Chimera X (Pettersen *et al.*, 2004; Goddard *et al.*, 2007, 2018; Pettersen *et al.*, 2021), Coot (Emsley *et al.*, 2010), and PyMOL (PyMOL, 2015) were used for visualisation and figure preparation.

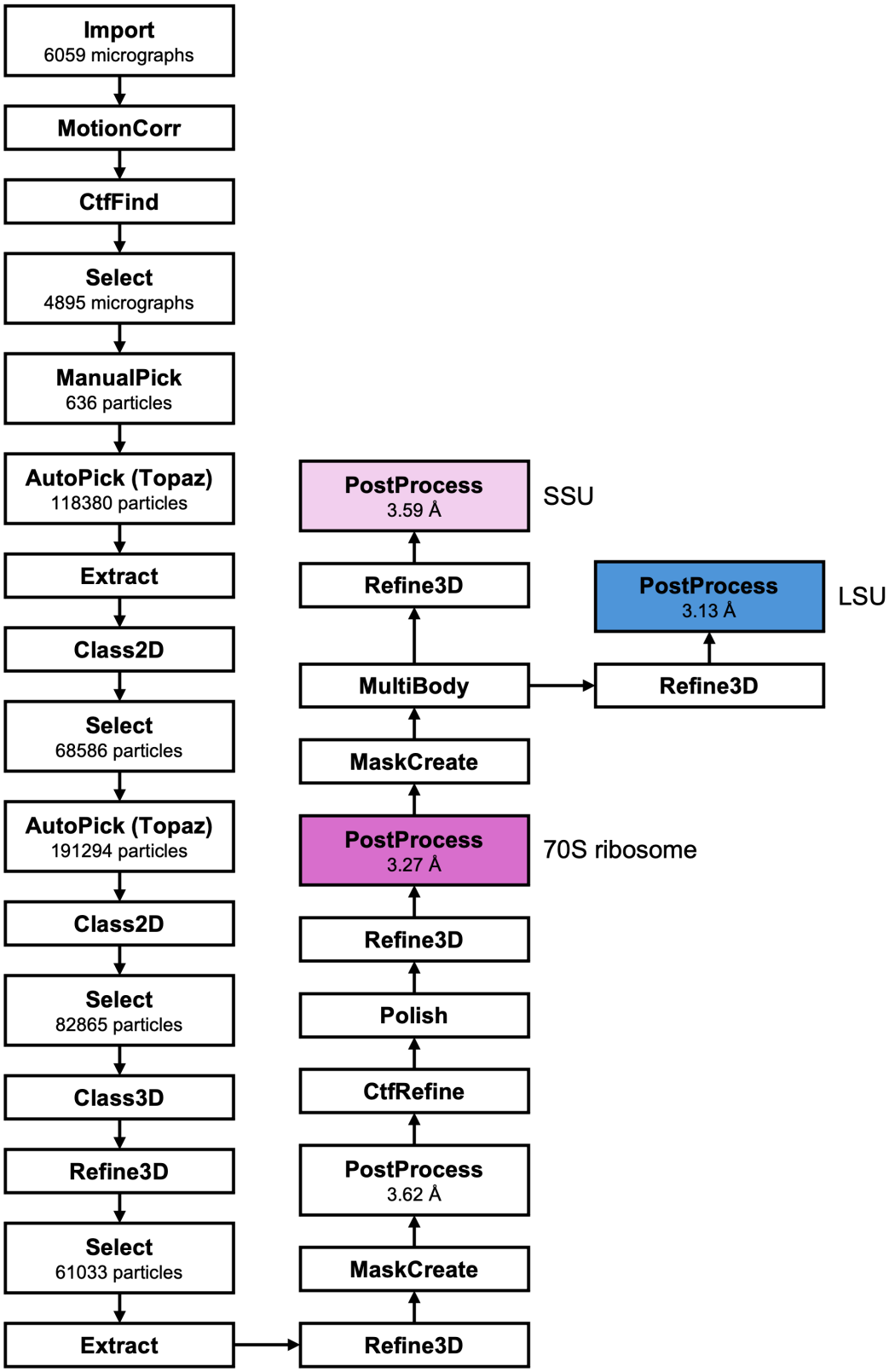



**Figure 4.6:** The RELION processing pipeline, with particle numbers and resolutions from the 70S ribosome dataset.

### 4.2.3 Model building

#### 4.2.3.1 Homology models

Performed in collaboration with Lennart Pagani. The ribosome structure from the archaea *Thermococcus kodakarensis* (PDB 6TH6, Sas-Chen *et al.* (2020)) was used to generate an initial model of the *N. maritimus* ribosome by aligning the *T. kodakarensis* and *N. maritimus* maps in PHENIX and fitting the *T. kodakarensis* model to the *N. maritimus* map (Vagin and Teplyakov, 1997; Adams *et al.*, 2010; Winn *et al.*, 2011). Where individual proteins were used as homology models, the existing structure was aligned with the homology model (mostly based on placement in the *T. kodakarensis* ribosome) using the PyMOL alignment wizard (PyMOL, 2015). Additionally, published structures of archaeal ribosomes from *Haloarcula marismortui* (PDB 1JJ2), *Pyrococcus horikoshii* (PDB 3A1Y), *Pyrococcus abyssi* (PDB 7HZG), and *Thermococcus celer* (PDB 6TMF) were used (Klein *et al.*, 2004; Naganuma *et al.*, 2010; Coureux *et al.*, 2016, 2020; Nürenberg-Goloub *et al.*, 2020). Sequence alignment for rRNAs and ribosomal proteins were performed using NCBI BLAST (Altschul *et al.*, 1990; Altschul, 1993; Altschul *et al.*, 1997; Zhang *et al.*, 2000; Johnson *et al.*, 2008), using sequences obtained from the UniProt and NCBI databases (Walker *et al.*, 2010).

#### 4.2.3.2 Structure prediction

As there are a number of differences in protein content between the *T. kodakarensis* and *N. maritimus* ribosomes, additional methods were required to create starting points for model building. An alpha version of ModelAngelo was first employed to build protein structures into the LSU density (Jamali *et al.*, 2022). The published version of ModelAngelo was later utilised to fill in additional portions of unassigned density in both the LSU map and the 70S map (Jamali *et al.*, 2024).

AlphaFold2 was used to generate protein structures where the homology model and ModelAngelo did not produce helpful starting points (Jumper *et al.*, 2021). Placement of these predicted structures was placed into the map using

alignment to a homology model, based on identification of related proteins using BLAST. This was performed in PyMOL (PyMOL, 2015).

To help build the rRNA, ClustalOmega (Sievers *et al.*, 2011), R2DT (Sweeney *et al.*, 2021), and DeepFoldRNA (Pearce *et al.*, 2022) were used.

#### 4.2.3.3 Model building

Model building was performed manually in Coot (Emsley *et al.*, 2010). Where the sequence could not be unambiguously identified, protein chains were built as poly-alanines and nucleic acid chains were built as poly-cytosines. Initial building of 23S rRNA (approximately 600 residues, centred on the high-conserved core) in the LSU map was performed by Lennart Pagani. Models were built without hydrogen atoms.

#### 4.2.3.4 Model refinement and validation

Refinement and validation of the LSU and 70S models was performed with PHENIX, within the CCP4 suite (Adams *et al.*, 2010; Liebschner *et al.*, 2019; Winn *et al.*, 2011). Hydrogens were added by PHENIX during the refinement process. The final model statistics are provided in Appendix Section B.7, Table B.1, and report the model-to-map correlation coefficient, $CC_{mask}$ (Jiang and Brünger, 1994; Afonine *et al.*, 2018a,b). Protein annotations are taken from UniProt and HAMAP (Pedruzzi *et al.*, 2015; The UniProt Consortium, 2021), and rRNA domain annotations are taken from Petrov *et al.* (2013).

### 4.2.4 Other analyses

A phylogenetic tree was generated using PhyloT and the Interactive Tree of Life (Letunic and Bork, 2007, 2011; Letunic, 2015). Data for the tree was taken from the NCBI taxonomy database (Sayers *et al.*, 2019; Schoch *et al.*, 2020). Structural alignment of zinc-binding proteins and aL8 with aL30 was performed using PyMOL (PyMOL, 2015). Sequence alignment of aL8 and aL30 was performed using BLAST (Altschul *et al.*, 1990, 1997; Johnson *et al.*, 2008). Low-pass filtering of maps was performed in UCSF Chimera X (Pettersen *et al.*, 2004; Goddard *et al.*, 2007, 2018; Pettersen *et al.*, 2021).

## 4.3 Results

### 4.3.1 50S map resolution and features

Single-particle cryo-EM was performed on ribosomes purified from *N. maritimus* cells. After picking particles, and iterative rounds of 2D and 3D classification (see Appendix Section B.1, Figure B.1), the structure of a 50S subunit was resolved from 255k particles, to a resolution of 3.1 Å. Further CTF refinements and Bayesian polishing allowed the structure to be improved to a resolution of 2.4 Å. The density map can be seen in Figure 4.7, and the FSC curve is in Appendix Section B.3. The map and model can be downloaded from Section B.2.

Upon building the model into the structure, it was quickly observed that this was not a complete 50S subunit. A number of components were not able to be observed, and it was determined that the 5S rRNA and some associated proteins were not present in the particles. The ribosomal stalk was also not observed, however this was expected as the stalk is extremely flexible, and is not routinely observed in ribosomal structures. The model-in-map FSC is provided in Appendix Section B.4.

In order to fully understand the ribosome of *N. maritimus*, the purification protocol was improved to ensure that the complete ribosome was enriched and was the dominant particle in the next round of cryo-EM.

### 4.3.2 70S map resolution and features

After work to improve the purification protocol to favour the stability of the 70S ribosome, SPA was performed on the new sample. This did yield an increased proportion of intact ribosomes, which were then picked from the micrographs. Iterative rounds of 2D classification to exclude 50S particles and 3D classification of picked particles led to 60.1k particles forming the final set. The global resolution of the resulting structure was 3.3 Å, which could be refined to show that the LSU had a resolution of 3.1 Å and the SSU a resolution of 3.6 Å (shown in Appendix Section B.5). As the resolution for the SSU was not able to be improved further than this with multibody refinement and focused masking, it was decided that

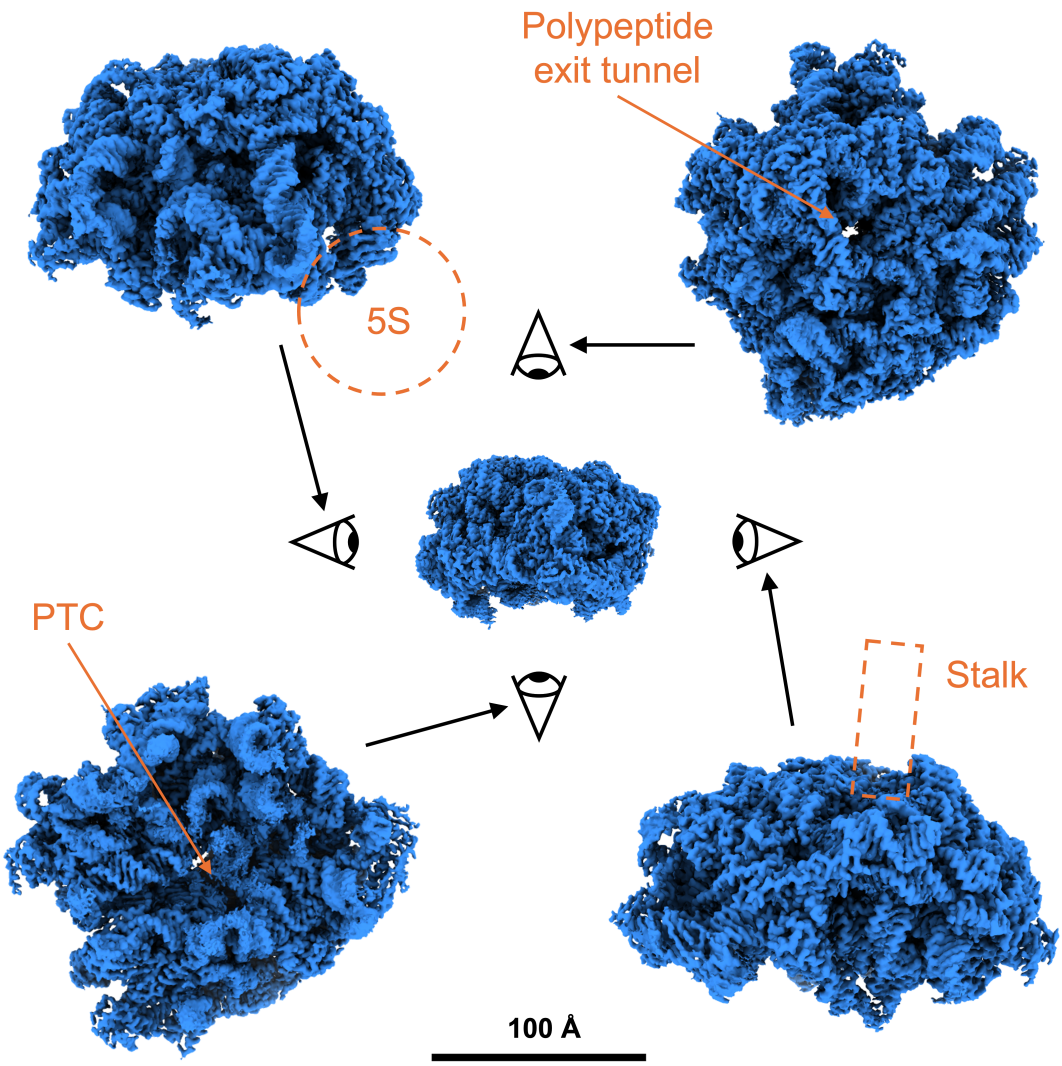



**Figure 4.7:** The density map of the 50S subunit. Four views are shown, as indicated by the eye icons.

model building efforts would focus on the LSU, and comparing the structure of the LSU in the 70S map with the structure from the 50S map (Figure 4.8). This is because 3.6 Å is above the limit for which ModelAngelo can confidently model. The map-in-model FSC for the LSU is provided in Section B.6. The map and model for the LSU can be downloaded from Section B.2.

#### 4.3.2.1 30S subunit model

The genome of *N. maritimus* has 29 proteins annotated as being part of the SSU, as well as the 16S rRNA. 18 of these proteins were able to be built into the map

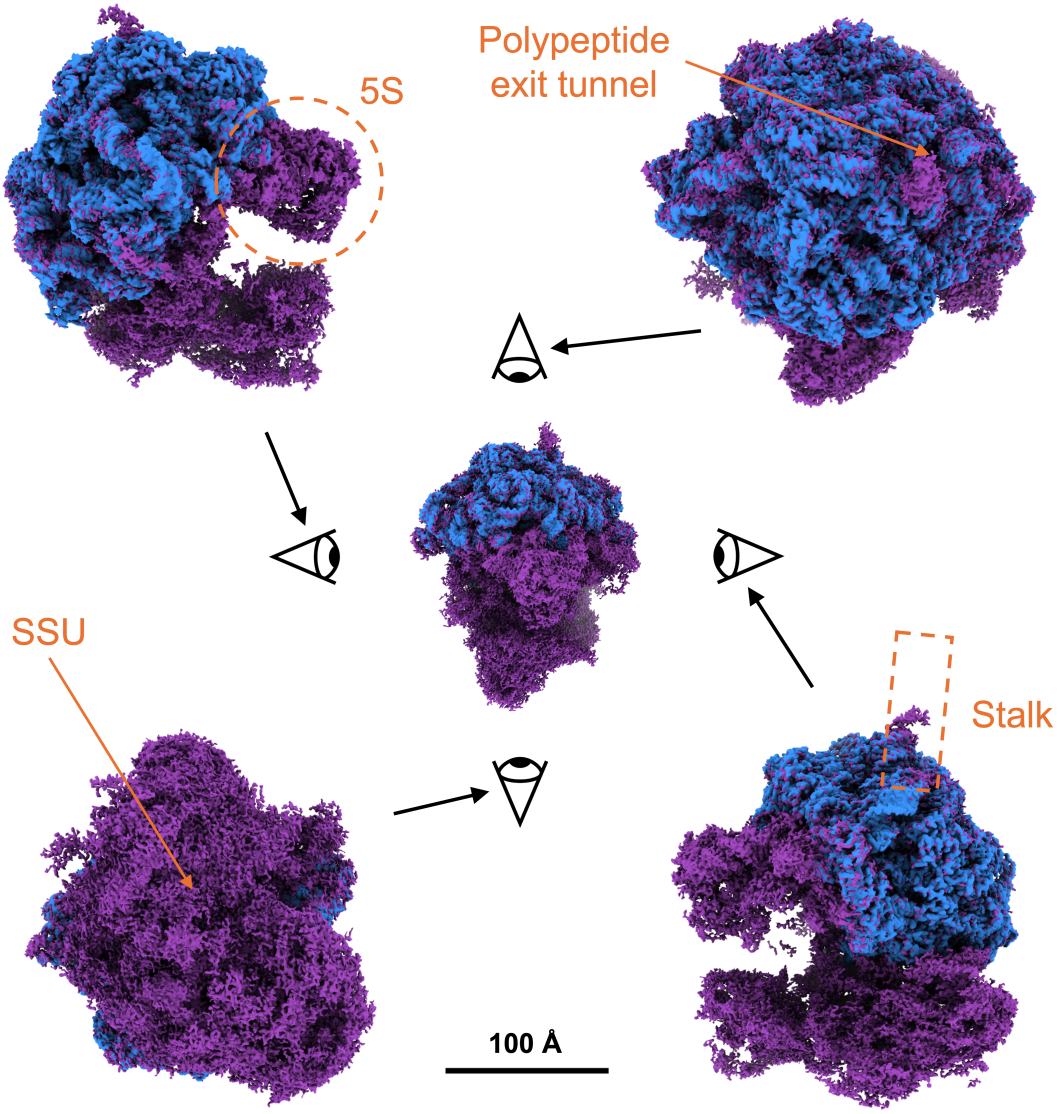



**Figure 4.8:** The density map of the 70S ribosome (purple), with the 50S map (blue) fitted. Features highlighted in Figure 4.7 highlighted on the full structure.

with varying degrees of confidence, but generally the resolution limited what could be built and therefore interpreted about this structure. An overview of what was built into each map is shown in Figure 4.9.

#### 4.3.2.2 Model validation

Model validation for the LSU can be found in Appendix Section B.7, Table B.1. Data for the SSU is included in Appendix Section B.8, Table B.2.

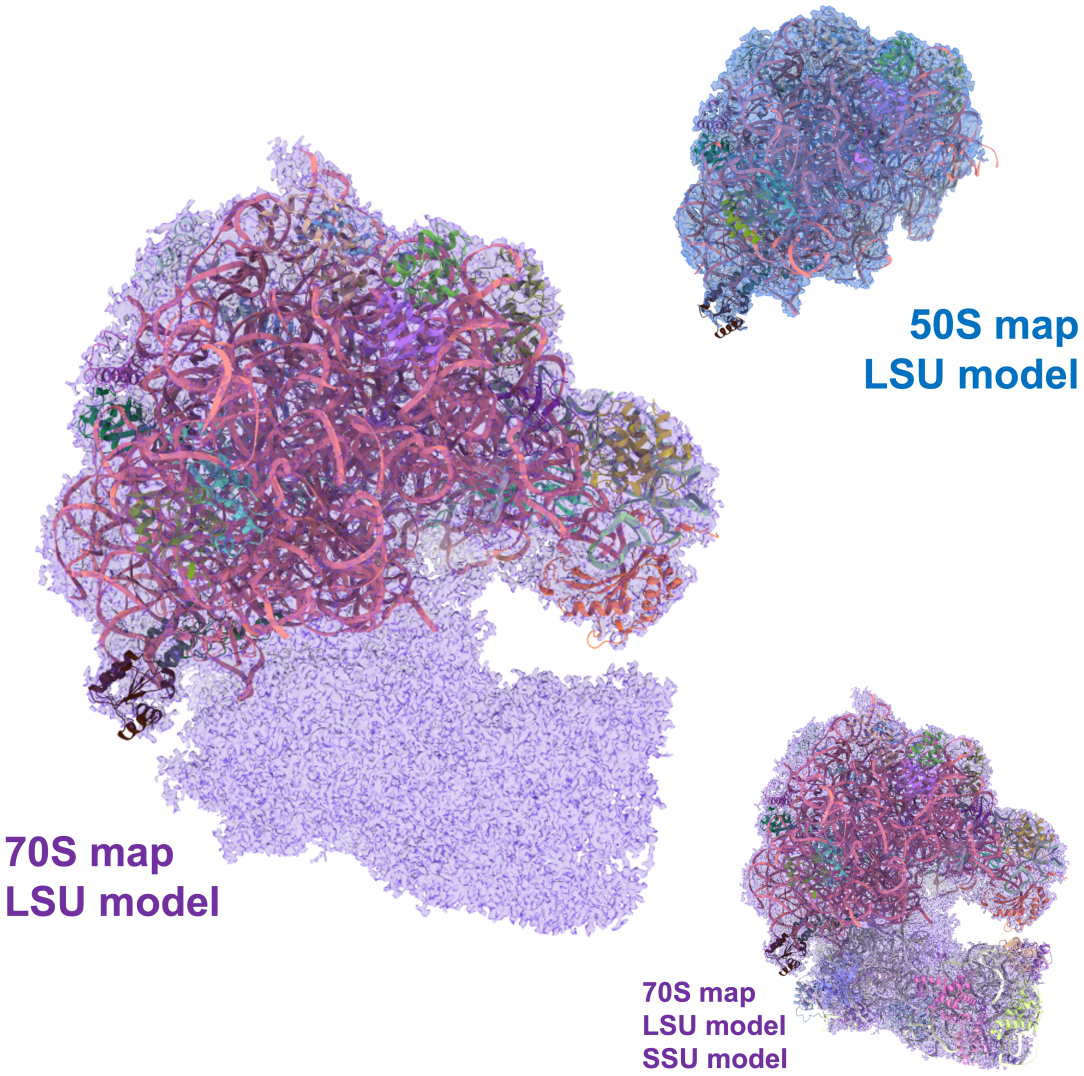



**Figure 4.9:** The density map of the 70S ribosome, with the LSU proteins and rRNAs built in. Also shown are the 50S map with the LSU built, and the 70S map with the LSU and SSU built.

### 4.3.3 Universal proteins in the LSU

There are 19 ribosomal proteins in the large subunit, which are universally conserved across all domains of life. All of these have been identified in the *N. maritimus* genome. An overview is provided in Table 4.1, along with data on their presence in both the 50S map and the 70S map.

**Table 4.1:** Table showing the name, UniProt ID, and percentage of sequence built into each map for the universal r-proteins. Where a protein was built into both maps, the RMSD between the two models was calculated.

| Name | UniProt ID | 50S map % built | 70S map % built | RMSD (Å) |
|---|---|---|---|---|
| uL1 | A9A579 | n/a | n/a | n/a |
| uL2 | A9A141 | 72% | 94% | 0.793 |
| uL3 | A9A5J4 | 99% | 99% | 0.723 |
| uL4 | A9A5J3 | 97% | 99% | 0.823 |
| uL5 | A9A5I2 | n/a | 99% | n/a |
| uL6 | A9A5H9 | 95% | 97% | 0.753 |
| uL10 | A9A578 | n/a | n/a | n/a |
| uL11 | A9A582 | n/a | n/a | n/a |
| uL13 | A9A5C3 | 97% | 97% | 0.429 |
| uL14 | A9A5I5 | 94% | 97% | 0.493 |
| uL15 | A9A598 | 62% | 90% | 0.611 |
| uL16 | A9A5E6 | n/a | 91% | n/a |
| uL18 | A9A595 | n/a | 99% | n/a |
| uL22 | A9A5J0 | 99% | 99% | 0.550 |
| uL23 | A9A5J2 | 100% | 95% | 0.492 |
| uL24 | A9A5I4 | 70% | 71% | 0.849 |
| uL29 | A9A5I8 | 96% | 97% | 0.927 |
| uL30 | A9A597 | 99% | 99% | 0.517 |
| uL33 | A9A5Z7 | n/a | 99% | n/a |

#### 4.3.3.1 uL1

uL1 is a protein that directly binds to the 23S rRNA, and is thought to be involved in E site tRNA release (HAMAP MF_01318_A). No position for uL1 was predicted by ModelAngelo in either map. uL1 was also not built into the *T. kodakarensis* map by Sas-Chen *et al.* (2020), so an X-ray crystallography structure of the *Saccharolobus acidocaldarius* uL1 bound to a fragment of the 23S rRNA (PDB 1MZP) was aligned by the 23S rRNA with the *T. kodakarensis* ribosome model (Nikulin *et al.*, 2003). An AlphaFold2 model based on the *N. maritimus* uL1 sequence was then aligned to the uL1 protein from *S. acidocaldarius*, suggesting an approximate location for uL1, however there was no clear density in either processed cryo-EM map (see Figure 4.10). This result is not unexpected, because the uL1 protuberance has limited contacts with the rest of the 50S subunit and is very flexible. There is also no density observed for the three rRNA helices bound by uL1 in either processed map.

The protuberance does appear when the map is low-pass filtered to $2\sigma$, showing it is still present in the purified ribosome, but the position of the protein and rRNA relative to the majority of the 50S subunit is not resolved to high resolution.

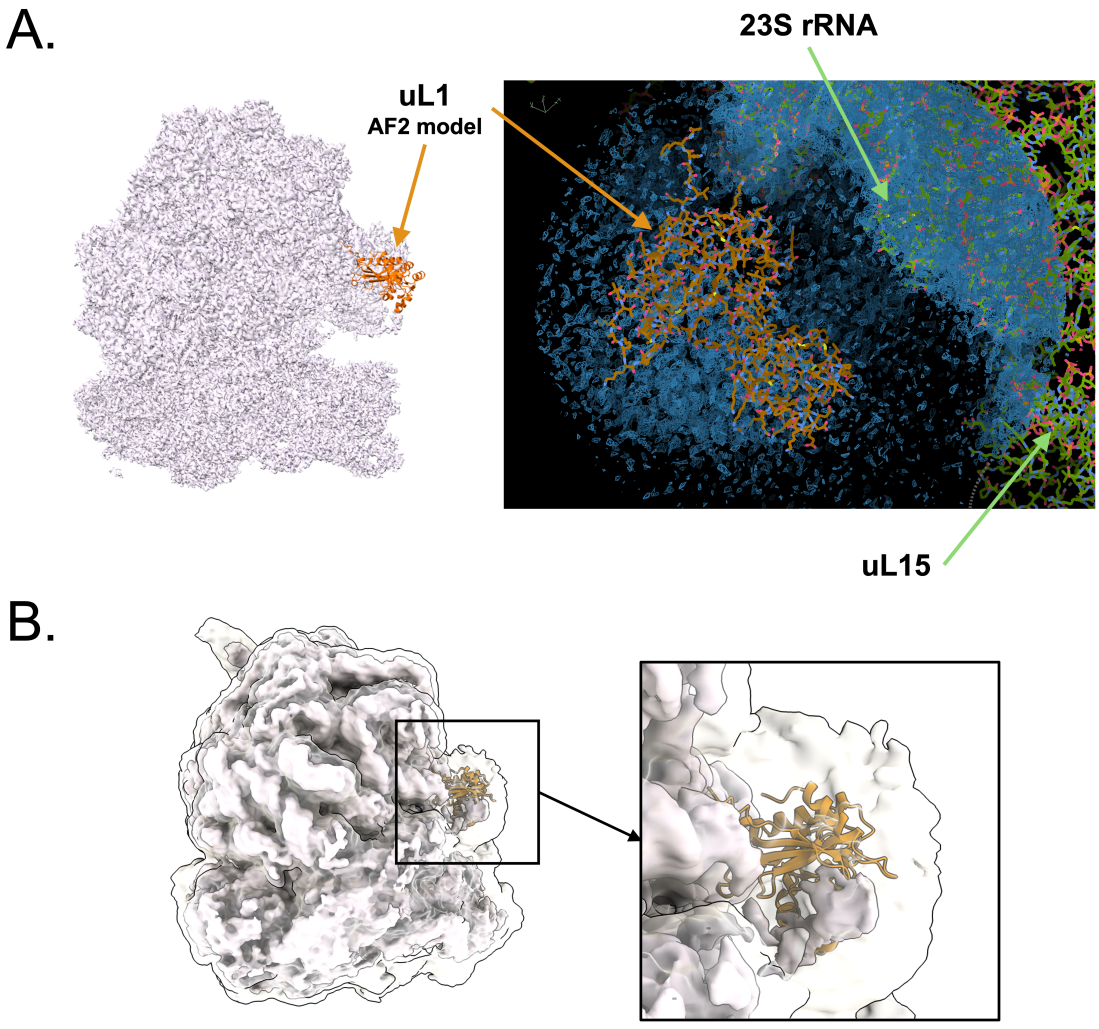



**Figure 4.10:** Ribosomal protein uL1. A. The AF2 model, aligned with the predicted placement of uL1 based on the bound 23S rRNA fragment from *S. acidocaldarius*. No density is seen in the map.
B. $2\sigma$ (transparent, outlined) and $1.5\sigma$ low-pass filtered maps of the 70S ribosome, showing the uL1 protuberance appears more strongly in the $2\sigma$ filtered map.

#### 4.3.3.2 uL2

uL2 is a protein required for the association of the large and small subunits, as well as for tRNA binding (HAMAP MF_01320_A). In the 70S ribosome, uL2 makes several

contacts with the 16S rRNA in the SSU. In the 50S structure, uL2 was modelled from F17 until R192, which represents approximately 72% of residues. This was able to be extended from G2 to K241 in the 70S map, representing 94% of residues, as shown in Figure 4.11. uL2 is seen to make contact with aL43 and three domains of the 23S rRNA: domain II (H35a, H38), domain IV (H62, H69), and domain V (H84).

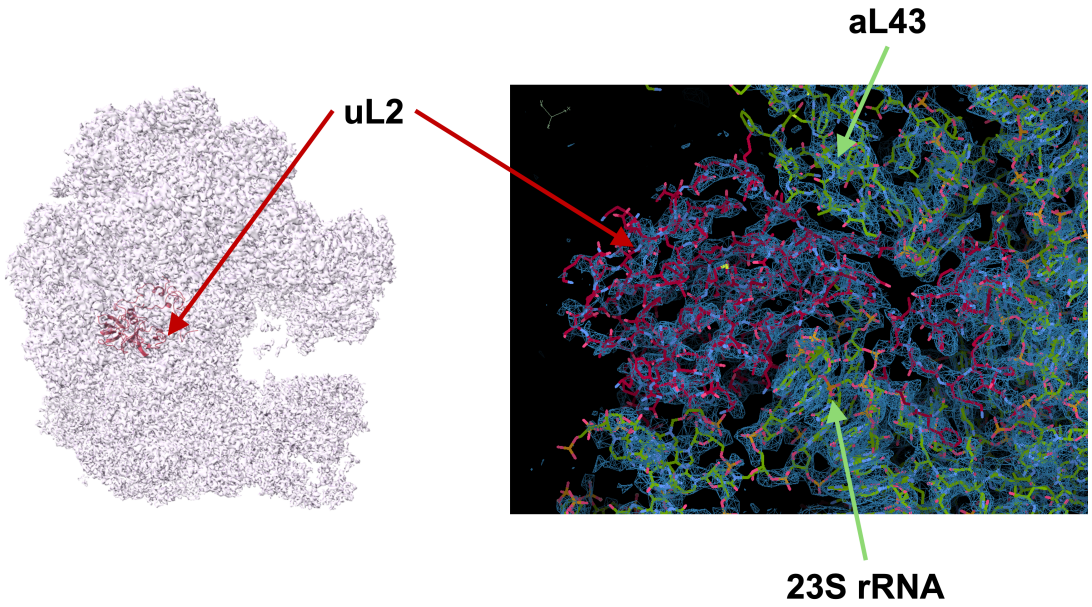



**Figure 4.11:** Ribosomal protein uL2, showing the position of the protein in the 70S complex at the interface between LSU and SSU on the left. The model is built into the 70S density on the right.

#### 4.3.3.3 uL3

uL3 is one of the primary rRNA-binding proteins, binding directly towards the 3' end of the 23S rRNA. It has an important role in nucleating assembly of the LSU (HAMAP MF_01325_A). The protein was modelled into both maps from K5 to 331V, representing approximately 99% of residues, and the binding at the 3' end of the rRNA is seen in the contacts with domains 0 (H73), IV (H65), V (H75, H76, H78), and VI (H94, H97, H99, H100, H101). Figure 4.12 shows the model built into the 70S map.

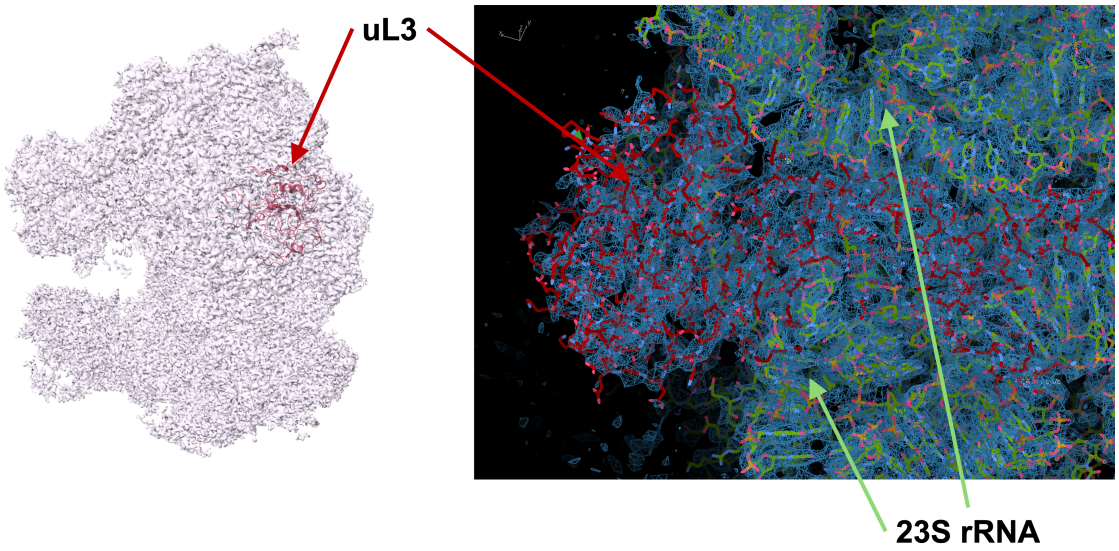



**Figure 4.12:** Ribosomal protein uL3, shown in the context of the 70S ribosome on the left. The model built into the 70S map is shown on the right.

#### 4.3.3.4 uL4

uL4 forms part of the polypeptide exit tunnel, and is another of the primary rRNA-binding proteins near the 5' end of the 23S rRNA that associates in the early stages of 50S assembly. In the assembled ribosome, it makes multiple contacts with different domains of the 23S rRNA(HAMAP MF_01328_A). Most of uL4 was modelled in the 50S map, from T2 to S61 and from G69 to V270, which represents approximately 97% of residues. The missing segment of seven residues fall in an "empty" part of the map, suggesting it is highly flexible, but are seen and modelled in the 70S map (see Figure 4.13). Contacts are seen with the 23S rRNA domain I (H18, H23 H24), domain II (H27, H33, H35), and domain III (H50, H52). uL4 also interacts with uL24, another protein forming the polypeptide exit tunnel (see Figure 4.23).

#### 4.3.3.5 uL5, uL18, and uL30

uL5 is a protein involved in binding the 5S rRNA and mediating the attachment of the 5S RNP to the LSU, where it forms part of the central protuberance. This protein is also involved in contacting S13 in the SSU, forming part of the bridge between the two proteins (HAMAP MF_01333_A).

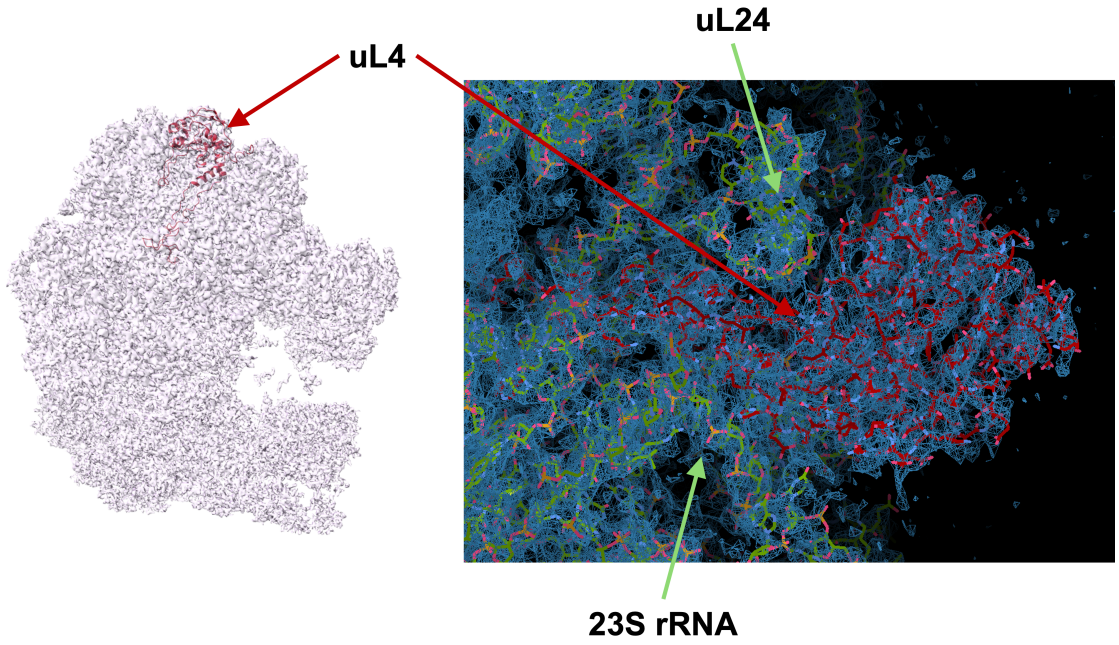



**Figure 4.13:** Ribosomal protein uL4, shown in the context of the 70S ribosome on the left, at the top of the LSU around the polypeptide exit tunnel. The model is shown built into the map on the right, next to uL24.

uL18 is involved in binding the 5S rRNA and mediating its attachment to the LSU (HAMAP MF_01337_A).

uL5 and uL18 are not seen in the 50S map, but are present and built into the 70S map albeit with low confidence for everything but the backbone. Figure 4.14 shows the proteins built into the 70S map.

uL30 (HAMAP MF_01371_A) is modelled into both maps, from N3 to I155. This represents approximately 99% of residues. uL30 has interactions with both the 23S rRNA and the 5S rRNA, whereas uL5 and uL18 only contact the 5S rRNA. Thus the loss of 5S rRNA from the LSU means that those two proteins also cannot associate with the LSU.

A "scoop" of the model, based around the 5S rRNA, was taken from the 70S model. This represents the 5S RNA, which is visualised in Figure 4.15. Five proteins are seen interacting with the 5S rRNA, and these are uL5, uL18, uL30 along with uL16 and aL21.

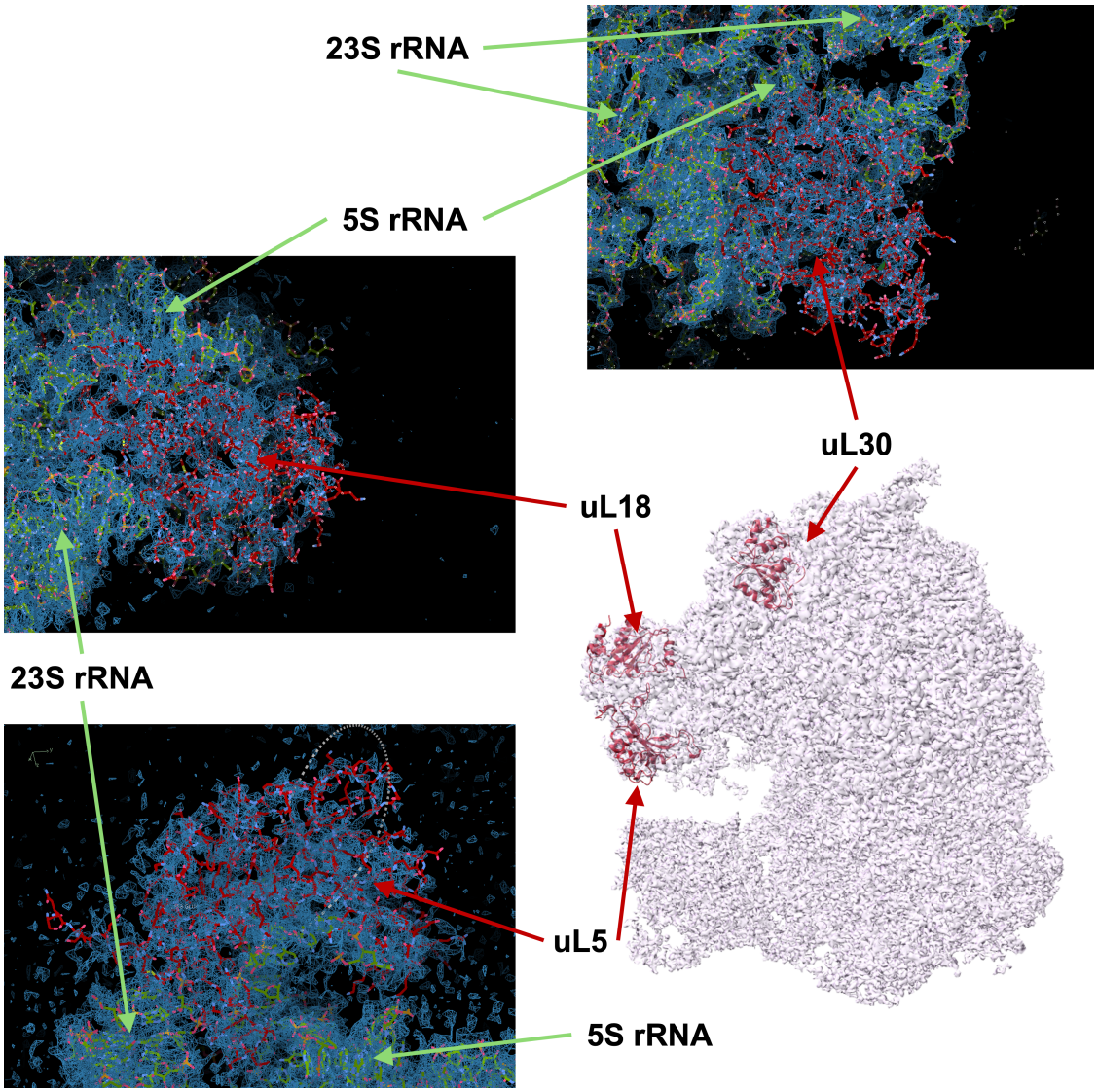



**Figure 4.14:** Ribosomal proteins uL5, uL18, and uL30 shown in context of the 70S ribosome, and shown modelled into the 70S density. All make contact with both the 23S rRNA and the 5S rRNA.

#### 4.3.3.6 uL6

uL6 is a protein important to the secondary structure of the 23S rRNA, located near the subunit interface in the base of the aLP1 stalk and the PTC (HAMAP MF_01365_A). It was modelled into the 50S density from E7 to I183, representing approximately 95% of residues, and only four extra residues were built in the 70S map. The model built into the 70S map is shown in Figure 4.16.

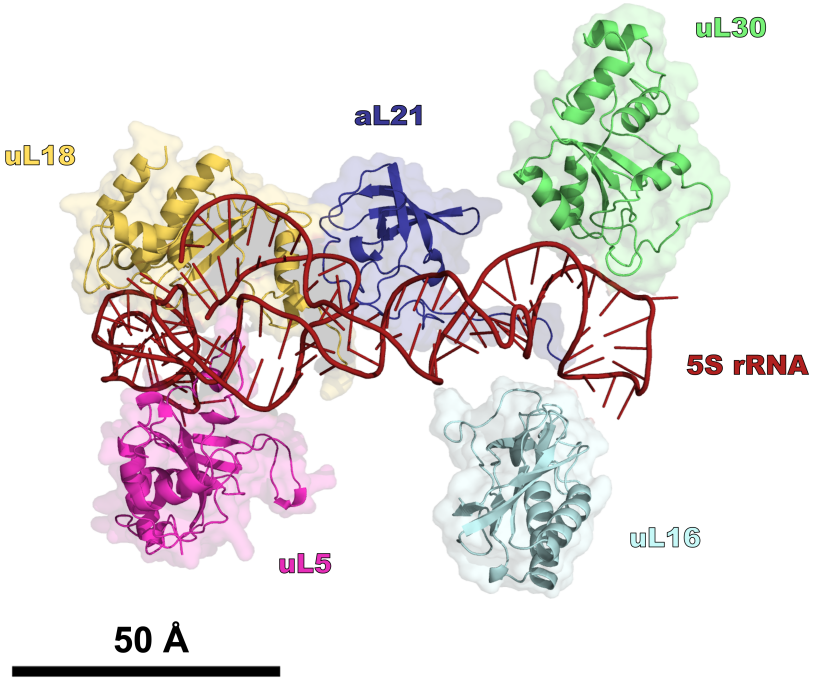



**Figure 4.15:** The 5S ribonucleoprotein, containing the 5S rRNA, and the r-proteins uL5, uL16, uL18, uL30, and aL21.

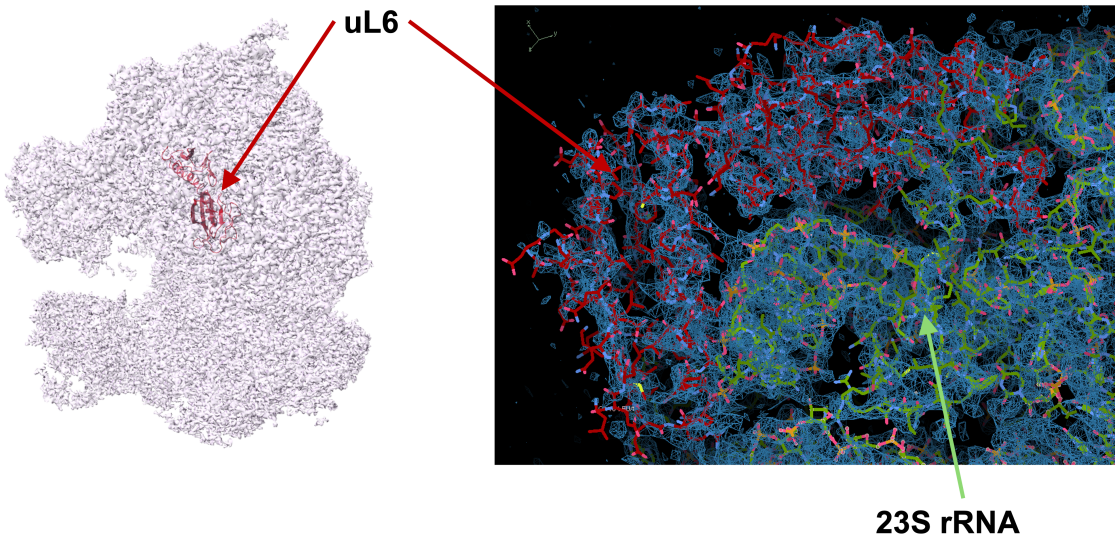



**Figure 4.16:** Ribosomal protein uL6, shown in the context of the 70S ribosome on the left, and the model built into the 70S density on the map with extensive interactions with the 23S rRNA.

#### 4.3.3.7 uL10 and uL11

uL10 forms part of the flexible ribosomal stalk, playing a central role in interactions with translation factors (HAMAP MF_00280). uL11 also forms part of the ribosomal stalk (HAMAP MF_00736). No density is seen in the either map for the stalk, excepting a small protrusion of rRNA. A putative model of the ribosomal stalk was generated using AlphaFold2 predictions and homology models of the ribosomal stalk from *Methanocaldococcus jannaschii* (PDB 5DAR, Gabdulkhakov *et al.* (2024)) and *Pyrococcus horikoshii* (PDB 3A1Y, Naganuma *et al.* (2010)) and is shown in Figure 4.17.

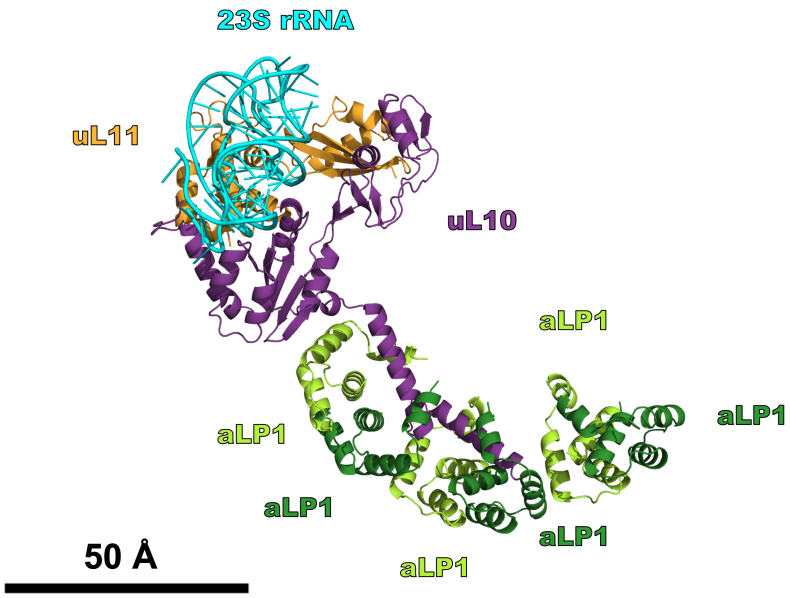



**Figure 4.17:** A putative model of the ribosomal stalk, created using PDB 5DAR and 3A1Y as homology models.

#### 4.3.3.8 uL13

uL13 is one of the early assembly proteins of the LSU (HAMAP MF_01366). Though reportedly not seen to bind rRNA directly, interactions are seen in the *N. maritimus* ribosome. This includes interactions with interactions with 23S rRNA domains 0 (H26), II (H29 (shown in Figure 4.18), H42), and VI (H101). The

protein was modelled into both maps from E5 to E154, representing approximately 97% of the residues.

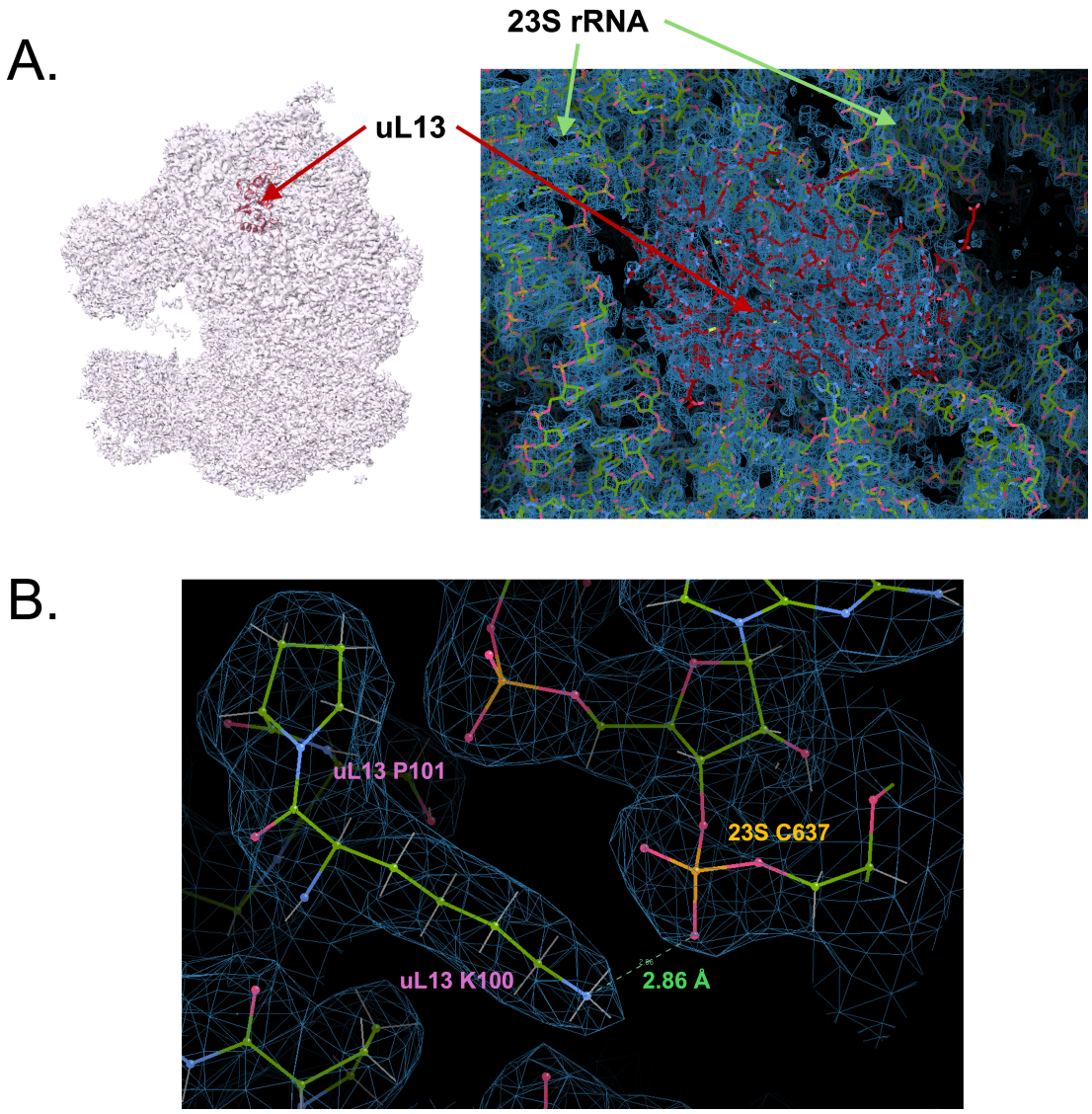



**Figure 4.18:** A. Ribosomal protein uL13, shown in the context of the 70S ribosome structure on the left, and built into the 70S density on the right.
B. One of the contacts between uL13 and 23S rRNA, differing from previous structures that have not observed uL13 interacting directly with rRNA.

#### 4.3.3.9 uL14

uL14 is a protein that forms part of two inter-subunit bridges in the 70S ribosome, and binds to the 23S rRNA (HAMAP MF_01367). The protein was modelled into

the 50S map from V9 to V140, which represents approximately 94% of residues. The first eight residues are not seen, and are theoretically close to residues I106 to E111 that have ambiguously placed sidechains. This is partially confirmed by the 70S map as shown in Figure 4.19, in which three more additional residues were modelled (backbone only).

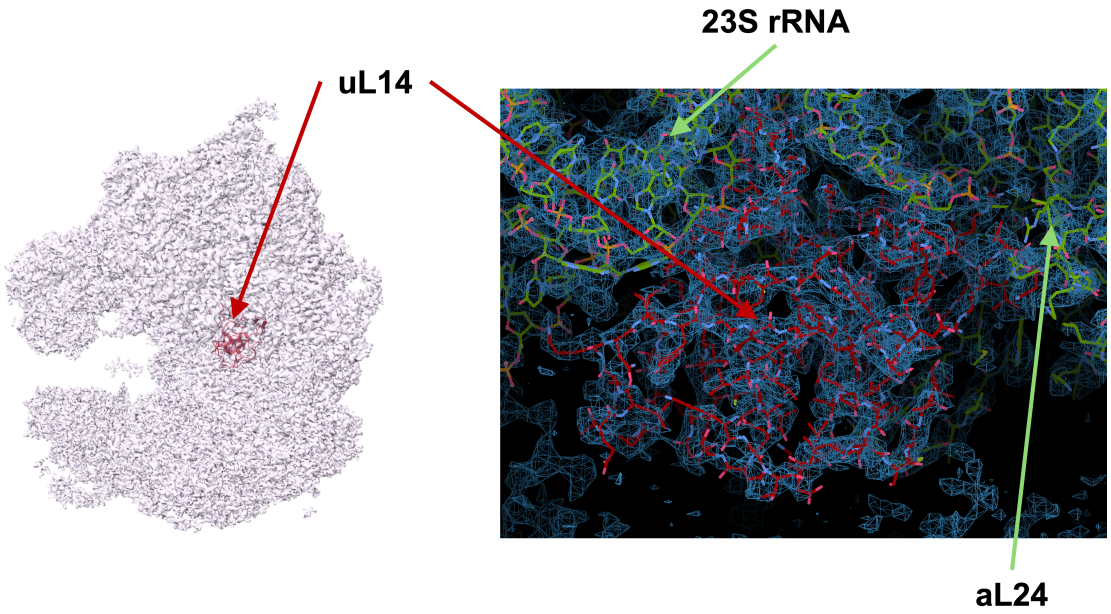



**Figure 4.19:** Ribosomal protein uL14, shown in the context of the 70S ribosome at the interface between the LSU and SSU. On the right, the model is shown built into the density, with interactions with aL24 and the 23S rRNA.

#### 4.3.3.10 uL15

uL15 is a 23S rRNA-binding protein (HAMAP MF_01341). Two portions of uL15 were built into the 50S map: A2 to V22, and P62 to I137. This represents approximately 62% of residues in total. Most of the missing residues could be modelled into the 70S map unambiguously, and the entire backbone of this section could be traced as shown in Figure 4.20. The fifteen C-terminal residues were not able to be modelled in either map, and are likely to be extremely flexible due to their location on the ribosome periphery.

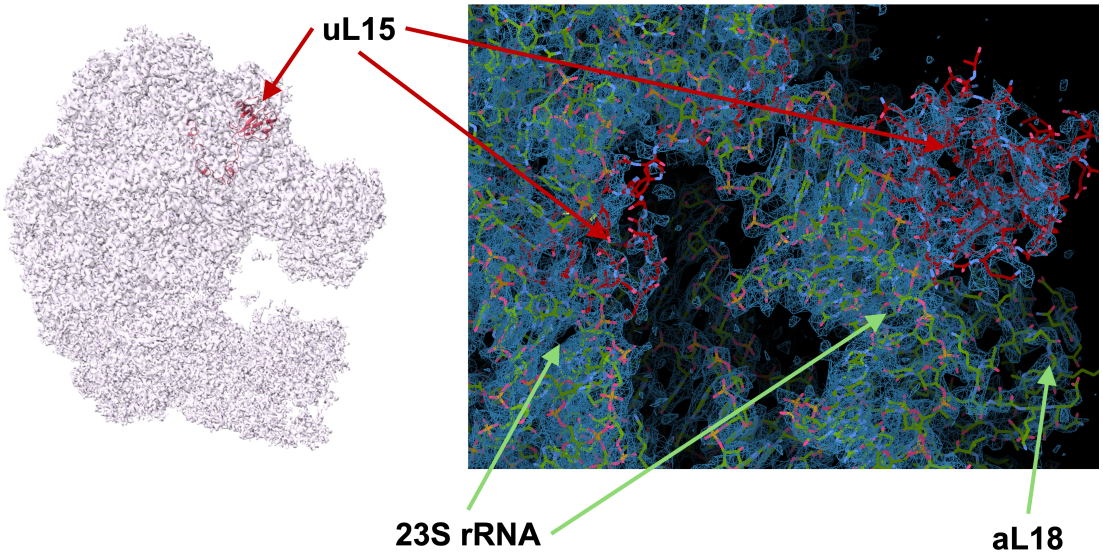



**Figure 4.20:** Ribosomal protein uL15, shown in the context of the 70S ribosome. The C-terminal of the protein is at the ribosome periphery, and the 40 residues missing in the 50S map are near at the interface with the 5S RNP. uL15 is shown in the 70S density on the right, with poorer resolution on the right of the image. Interactions are shown with aL18 and the 23S rRNA.

#### 4.3.3.11 uL16

uL16 is a protein that in archaea is often classed as L10e, despite the homology to the bacterial L16 (HAMAP MF_01342). This may be due to their location in the genome, as uL16 is not found in the same r-protein operon in archaea as bacteria (Wang *et al.*, 2009). uL16 is only seen in the 70S map, binding to H90 in domain V of the 23S rRNA (see Figure 4.21).

#### 4.3.3.12 uL22, uL23, uL24, and uL29

uL22 makes multiple contacts with different domains of the 23S rRNA in the assembled LSU, forming part of the polypeptide exit tunnel. The globular domain is located near the outside of the LSU, while an extended $\beta$-hairpin is towards the centre of the ribosome (HAMAP MF_01331_A). These observations are seen in the model in both maps, which covered 99% of residues.

uL23 is a 23S rRNA-binding protein also involved in lining the polypeptide exit tunnel (HAMAP MF_01369_A). All residues were modelled in both maps.

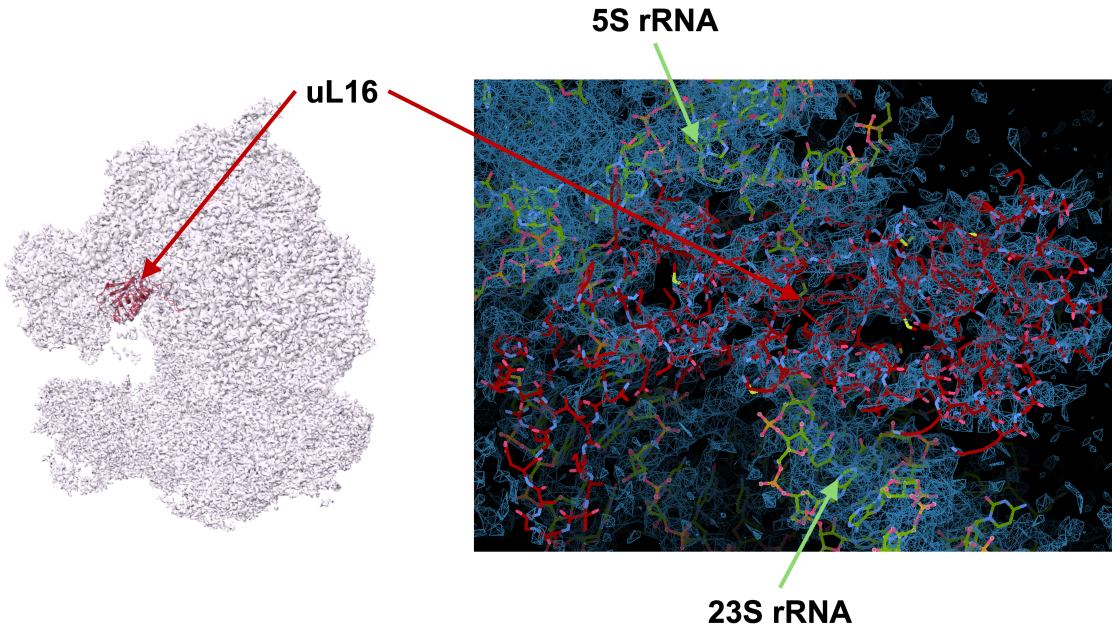



**Figure 4.21:** Ribosomal protein uL16, shown in the context of the 70S ribosome in proximity to the 5S rRNA. The model is shown built into the 70S map on the right, between the 23S and 5S rRNAs.

uL24 is one of two assembly initiator proteins that binds directly to the 5' end of the 23S rRNA, nucleating LSU assembly. It also forms part of the polypeptide exit tunnel (HAMAP MF_01326_A). However in *N. maritimus*, uL13 seems to bind the very 5' end of the 23S rRNA, with uL24 binding around residues 80–100 and 300–320. In the 50S map, a model was built for M1 to P117, representing approximately 70% of the residues. The remaining 50 residues are not seen in the 50S map, and are on the edge of the LSU at the polypeptide exit tunnel, suggesting flexibility to allow accommodation of the nascent polypeptide.

uL29 is a small protein, with approximately 96% of residues being built into both maps showing that it is involved in lining the end of the polypeptide exit channel (HAMAP MF_00374).

uL22, uL23, uL24, and uL29 are shown built into the 70S map in Figure 4.22.

The above proteins have been shown by Ahn *et al.* (2022) to be involved in co-translational folding of nascent chains by modulating the freedom of the chains as they exit the ribosome. The structure of the *N. maritimus* ribosomal

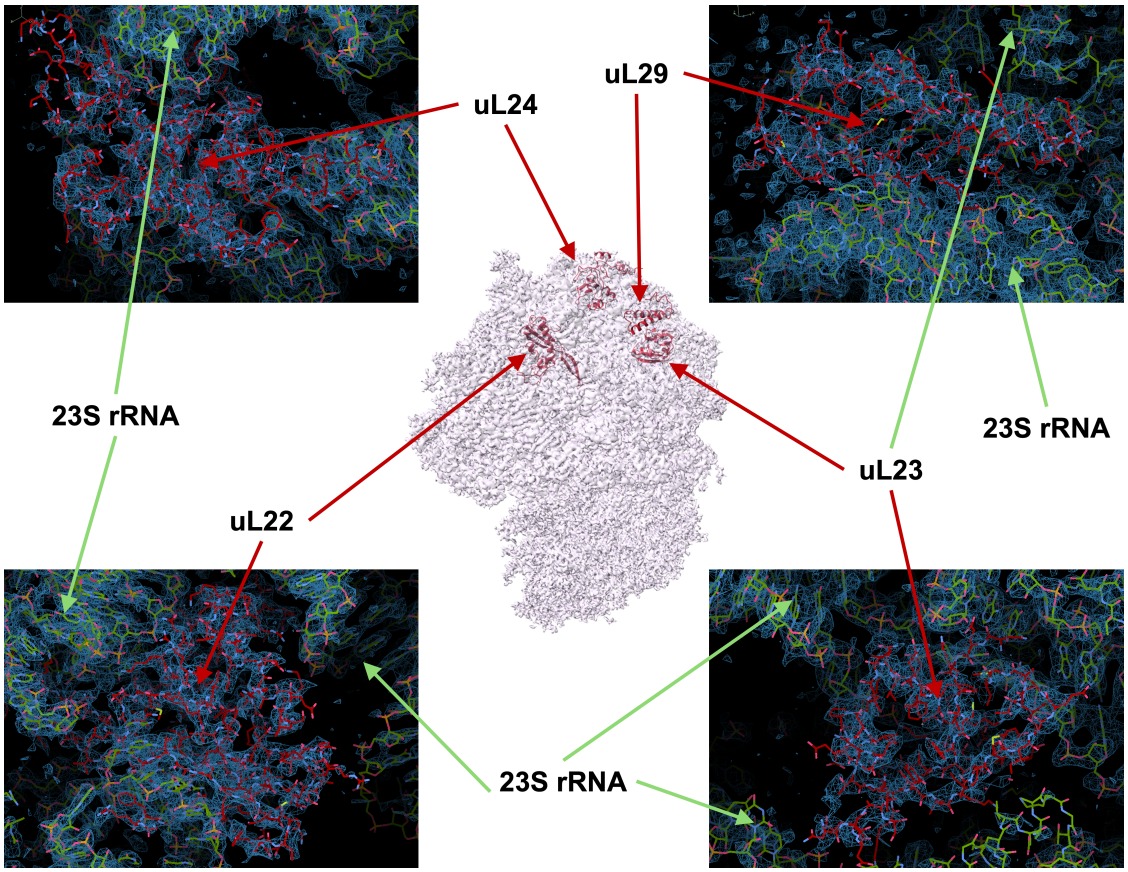



**Figure 4.22:** Ribosomal proteins uL22, uL23, uL24, and uL29, shown in context of the 70S ribosome at the polypeptide exit tunnel. The models are shown built into the 70S map, with interactions with the 23S rRNA, and an interaction between uL23 and uL29.

polypeptide exit channel agrees with those findings based on the positions of each protein relative to the exit, and the position of the flexible loops involved in the freedom of nascent chains (Figure 4.23).

#### 4.3.3.13 uL33

uL33 is a small protein predicted to bind zinc due to the presence of four cysteine residues (HAMAP MF_01476). It has previously been known as aL42, however work by Kovacs *et al.* (2018) showed that the protein was homologous to bacterial L33, and therefore it was renamed although the gene name retains the previous school of thought. In the 70S map, 99% of the residues are able to be modelled (see Figure 4.24). No density exists in the same placement in the 50S map, however ModelAngelo predicted an alternative placement on the LSU (Figure 4.24). Only a

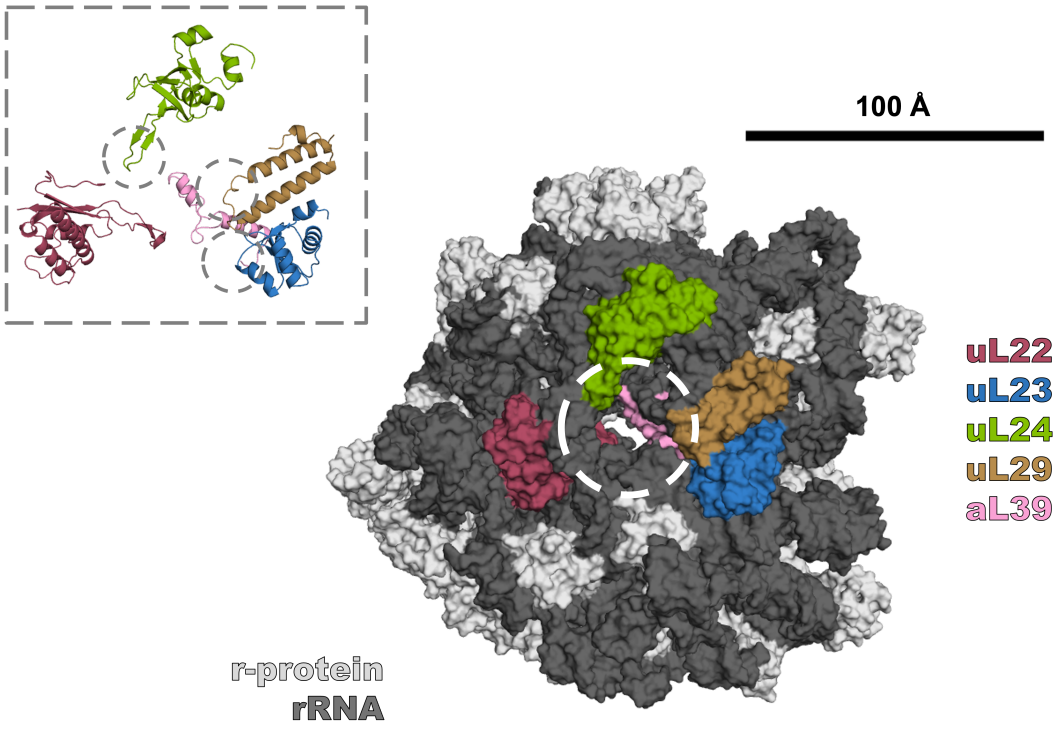



**Figure 4.23:** Surface model of the LSU, showing proteins forming the end of polypeptide exit tunnel. The white dotted circle indicates the location of the exit tunnel, surrounded by rRNA (dark grey) and uL22 (maroon), uL23 (blue), uL24 (green), uL29 (brown), and aL39 (pink). A small segment of uL22 can be seen further towards the centre of the ribosome, towards the beginning of the exit tunnel.
The dark grey dotted inset shows the cartoon structure of uL22, uL23, uL24, uL29, and aL39, with the circles highlighting the flexible loops.

small number of residues were modelled, so no confident assertions can be made.

### 4.3.4 Archaea-eukarya proteins in the LSU

There are 22 ribosomal proteins in the large subunit, which are conserved across the domains of Archaea and Eukarya. Not all of these proteins are essential, as they are not found in all archaeal or eukaryotic ribosomes. *N. maritimus* encodes what is likely to be a prudent set of r-proteins due to the relatively small size of the genome at 1.65 Mbp (Edmondson *et al.*, 2009; Walker *et al.*, 2010). An overview is provided in Table 4.2, along with data on their presence in both the 50S map and the 70S map.

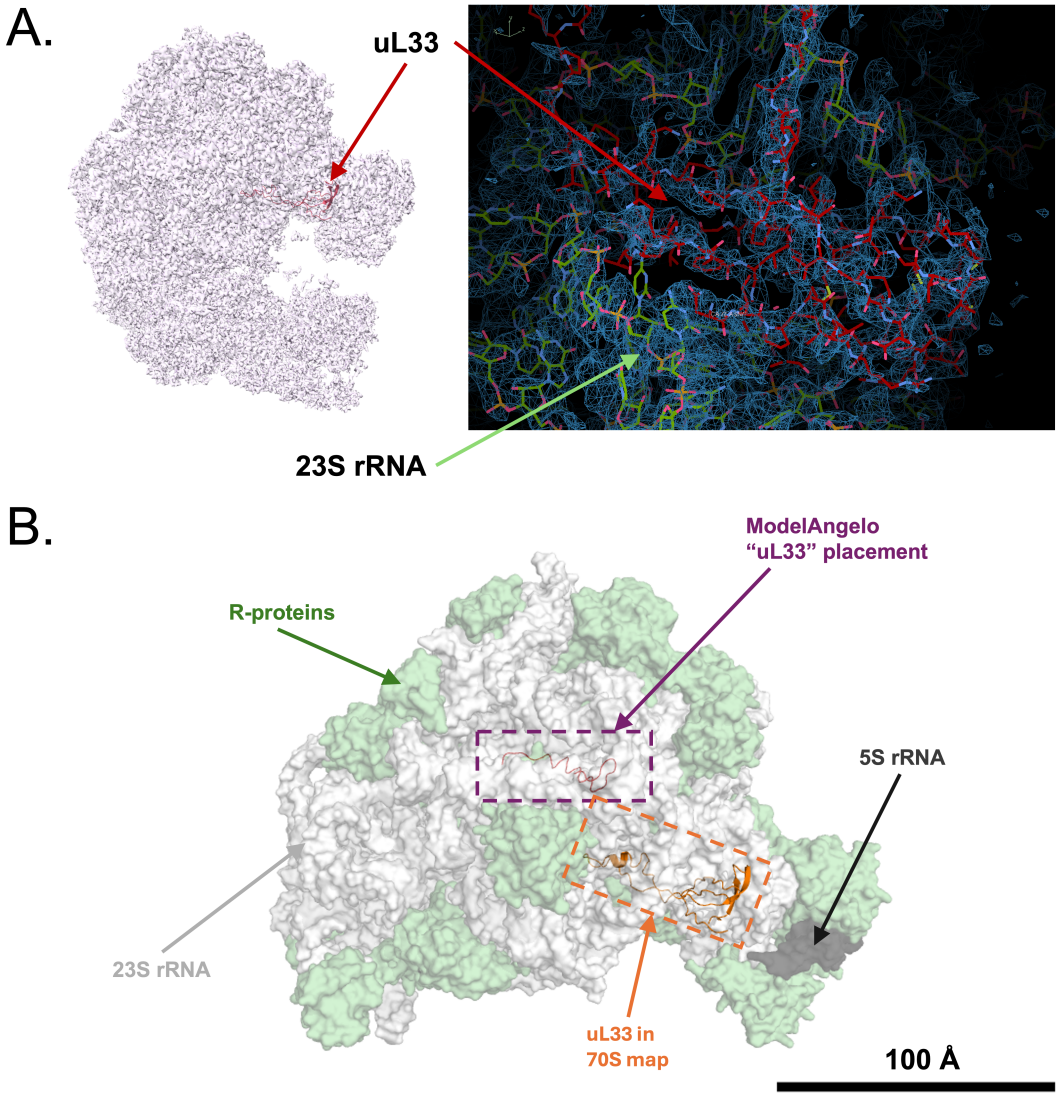



**Figure 4.24:** A. Ribosomal protein uL33, shown in the context of the 70S ribosome on the left, and built into the 70S density on the right.
B. The placement of uL33 (orange) in the ribosome as built into the 70S map, in the 50S map. The purple fragment is the ModelAngelo-predicted position of 13 residues of uL33 in the 50S map, with the backbone extended a further 18 residues.

**Table 4.2:** Table showing the name, UniProt ID, and percentage of sequence built into each map for the archaea-eukarya r-proteins. Where a protein was built into both maps, the RMSD between the two models was calculated.
* indicates that Mushegian (2005) identified the protein as being part of the minimal ribosome.

| Name | UniProt ID | 50S map % built | 70S map % built | RMSD (Å) |
|---|---|---|---|---|
| aL8 | A9A2Z3 | 93% | 97% | 0.892 |
| aL13 | n/a | n/a | n/a | n/a |
| aL14 | n/a | n/a | n/a | n/a |
| aL15* | A9A1H1 | 97% | 98% | 0.510 |
| aL18 | A9A5C4 | 99% | 99% | 0.554 |
| aL19 | A9A587 | 99% | 98% | 0.711 |
| aL20 | n/a | n/a | n/a | n/a |
| aL21* | A9A3L8 | n/a | 99% | n/a |
| aL24* | A9A2Z5 | 95% | 95% | 0.622 |
| aL30 | A9A464 | 97% | 88% | 1.153 |
| aL31* | A9A1I0 | 43% | 44% | 0.466 |
| aL32* | A9A588 | 96% | 98% | 0.655 |
| aL33 | n/a | n/a | n/a | n/a |
| aL37* | A9A2Y8 | 87% | 98% | 0.701 |
| aL38 | n/a | n/a | n/a | n/a |
| aL39* | A9A1I1 | 98% | 98% | 0.454 |
| aL40 | A9A4P3 | n/a | 98% | n/a |
| aL41 | n/a | n/a | n/a | n/a |
| aL43 | A9A5D1 | 91% | 98% | 0.896 |
| aLP1 | A9A571 | n/a | n/a | n/a |
| aLP2 | n/a | n/a | n/a | n/a |

#### 4.3.4.1 aL8

aL8 is a multi-functional RNA-binding protein that recognises the K-turn motif in rRNA as well as other nucleoproteins (HAMAP MF_00326). Two copies of aL8 (referred to as L7Ae) were modelled into the ribosome of *T. kodakarensis*. However, upon inspection of the model and map, the density of one copy resembles RNA. Only one copy of aL8 has been identified in the density maps of the *N. maritimus* ribosome, and is modelled from A4 to G123 in the 50S map. This represents approximately 93% of the residues. The sidechains for residues between L102 and V111 are not visible in the density, suggesting they are very mobile, and this is also observed in the 70S map as shown in Figure 4.25. This makes sense given the placement of aL8 on the subunit surface. Work by Daume *et al.* (2017) shows that aL8 can bind

a wide variety of RNA substrates, which may explain why it can be observed in multiple locations on the ribosomal surface in other archaeal ribosome structures.

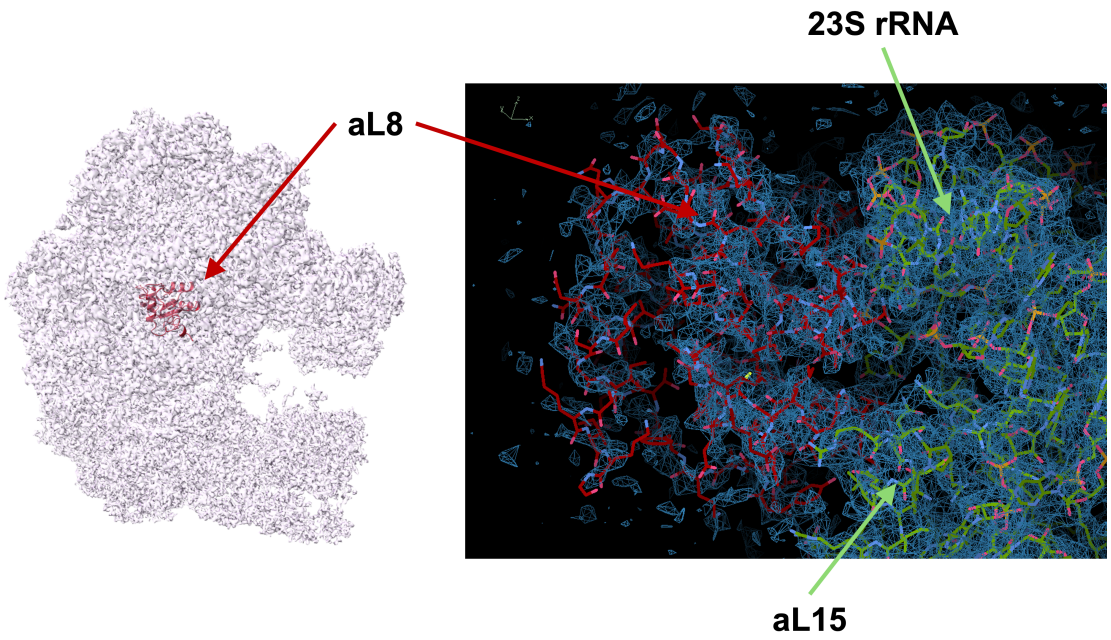



**Figure 4.25:** Ribosomal protein aL8, shown in the context of the 70S ribosome on the left and built into the 70S density on the right. Interactions are seen with aL15 and the 23S rRNA.

#### 4.3.4.2 aL15, aL18, aL19, and aL21

aL15 is a ribosomal protein (HAMAP MF_00256). In both maps, it is modelled from P5 to A154, covering approximately 98% of residues. Fernández-Fernández *et al.* (2023) showed that, in yeast, aL15 is important in 60S subunit assembly, and the presence of aL15 in the 50S map suggests that the protein is stably incorporated into the LSU.

aL18 is modelled into both maps from T2 to G115, covering approximately 99% of the residues (HAMAP MF_00329). In eukaryotic cells, Luan *et al.* (2022) showed that aL18 plays a role in cell senescence, but no research exists on the roles of aL18 in archaea.

aL19 is a 23S rRNA-binding protein (HAMAP MF_01475). The protein is modelled into both maps from V2 to K150, which represents approximately 99%

of residues. In eukaryotic ribosomes, work by Kisly *et al.* (2016) suggests it is important in the inter-subunit bridge, however the extended C-terminal $\alpha$-helix does not exist in *N. maritimus*, so this function may not apply in archaea.

aL21 is ribosomal protein about which very little is known, particularly in archaea (HAMAP MF_00369). It is not seen in the 50S map, however in the 70S map it is seen interacting with the 5S rRNA (see Figure 4.15). This has not been previously reported.

The models of aL15, aL18, aL19, and aL21 are shown in Figure 4.26.

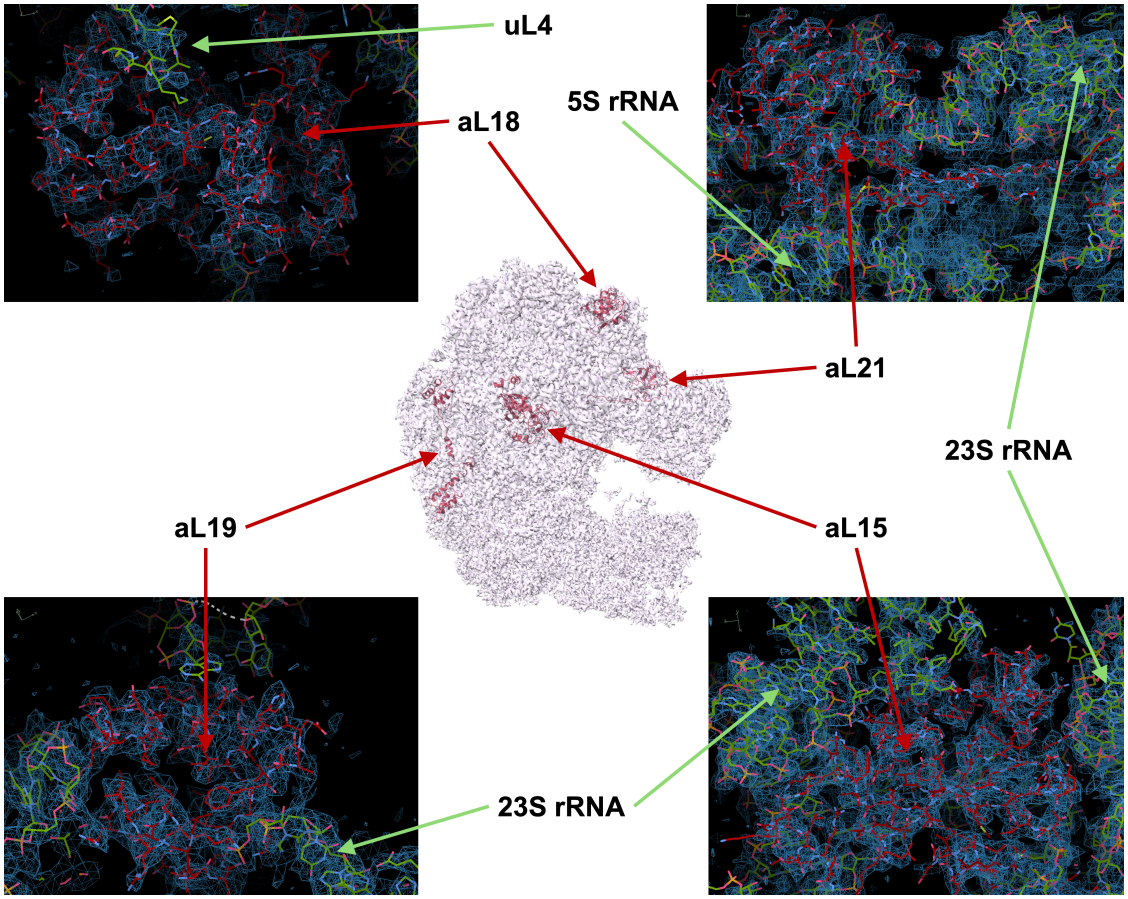



**Figure 4.26:** Ribosomal proteins aL15, aL18, aL19, and aL21 shown in the context of the 70S density. The models are shown built into the 70S map, and interactions with the 23S rRNA are indicated as well as between aL21 and the 5S rRNA, and aL18 and uL4.

#### 4.3.4.3 aL24, aL37, aL40, and aL43

aL24 is a zinc-binding protein that binds to the 23S rRNA (HAMAP MF_00773). Four cysteine residues coordinate one zinc ion. In both maps, S2 to R64 are modelled, which represents approximately 95% of residues.

aL37 (HAMAP MF_00547) is another zinc-binding protein. In the 50S map, residues M8 to K14 are tentatively modelled, and residues K15 to 55T are confidently modelled, whereas all the residues from V2 to 55T are modelled in the 70S map.

aL40 (HAMAP MF_00788) is a small protein predicted to bind zinc due to the presence of four cysteine residues, and is modelled from P2 to K55 in the 70S map albeit with less confidence at the N- and C-termini. It is not seen in the 50S map.

aL43 is a zinc-binding protein that binds the 23S rRNA (HAMAP MF_00327). It binds one zinc ion, as shown in Figure 4.28 A. It is also called L37Ae. It is present in the 50S map, and modelled from K6 to K69, representing approximately 91% of residues, and this is extended to 99% of the residues in the 70S map.

Figure 4.27 shows the models of aL24, aL37, aL40, and aL43.

Other than binding a zinc ion by four co-ordinating cysteine residues, these proteins (along with uL33) share no structural homology, as shown in Figure 4.28 B.

#### 4.3.4.4 aL30

aL30 (HAMAP MF_00481). It is modelled into the 50S map from K3 to K98, with residues to E104 modelled with ambiguous sidechains. This represents approximately 97% of residues. The map of aL30 is much lower quality in the 70S map, and the model cannot be built as accurately. The model is shown built into both maps for comparison in Figure 4.29.

Previous work from Wang *et al.* (2009) characterised aL30 as distinct from aL8, though it had previously been documented as a second copy of aL8 in a different operon. The structures of the two proteins from the *N. maritimus* ribosome were compared, as shown in Figure 4.30, and clearly show very similar architecture in both secondary and tertiary structure formed of five $\alpha$-helices and four $\beta$-strands. This is supported by the Encyclopedia of Domains, which classes both proteins in

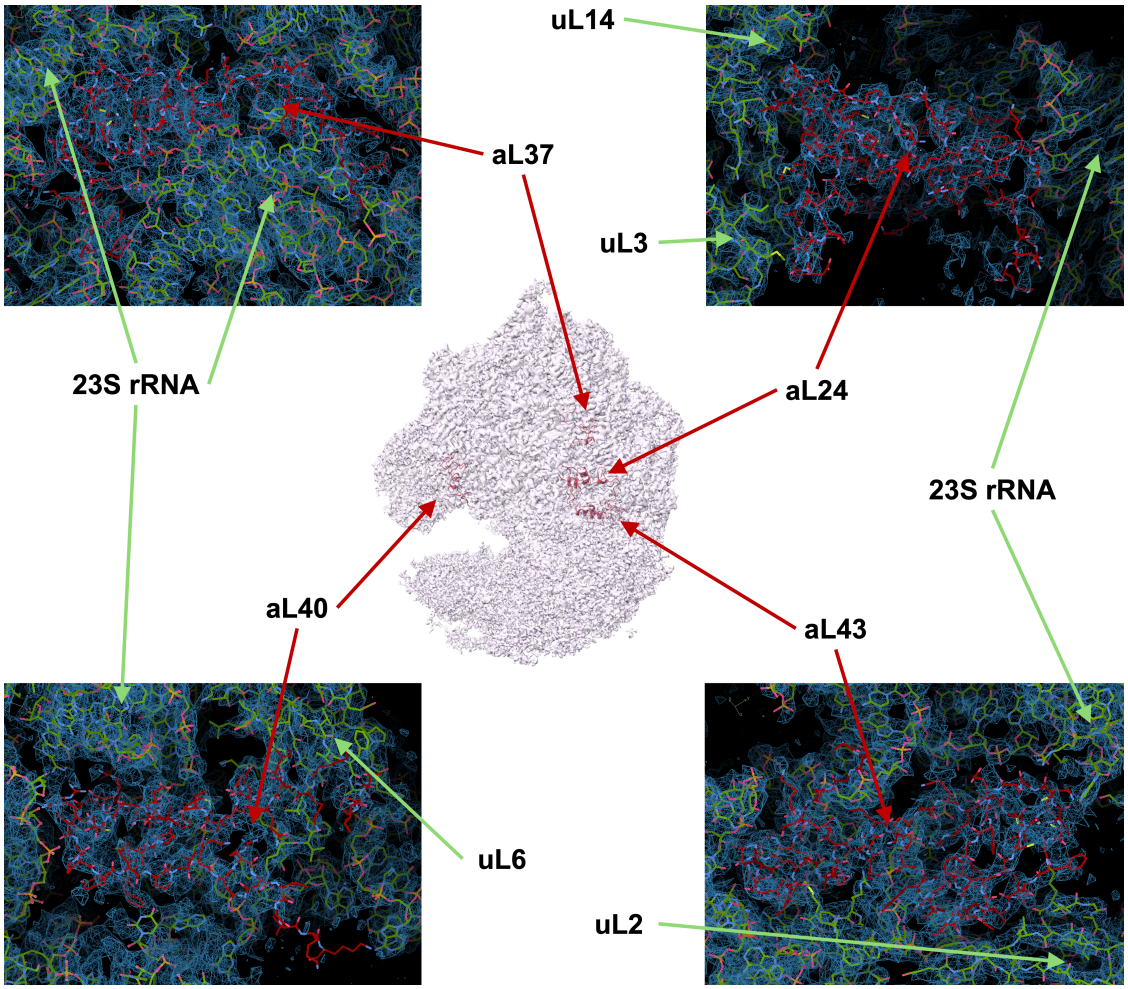



**Figure 4.27:** Ribosomal proteins aL24, aL37, aL40, and aL43, shown in the context of the 70S density. The models are shown built into the 70S map, including interactions with other ribosomal proteins and the 23S rRNA.

the same domain superfamily, CATH Superfamily 3.30.1330.30 (Lau *et al.*, 2024). However, they are highly divergent in their primary sequence, with an Expectation value score of $3 \times 10^{-4}$ calculated by BLAST (Altschul *et al.*, 1990, 1997), which indicates only low to moderate significance in alignment.

#### 4.3.4.5 aL31

aL31 (HAMAP MF_00410) is only partially visible in both maps. Residues L5 to E85 are modelled, which is approximately 43% of residues, and there is no visible density for the remaining 102 residues. The location of aL31, on the ribosomal surface, could explain why much of it is unseen as it may be unstructured as

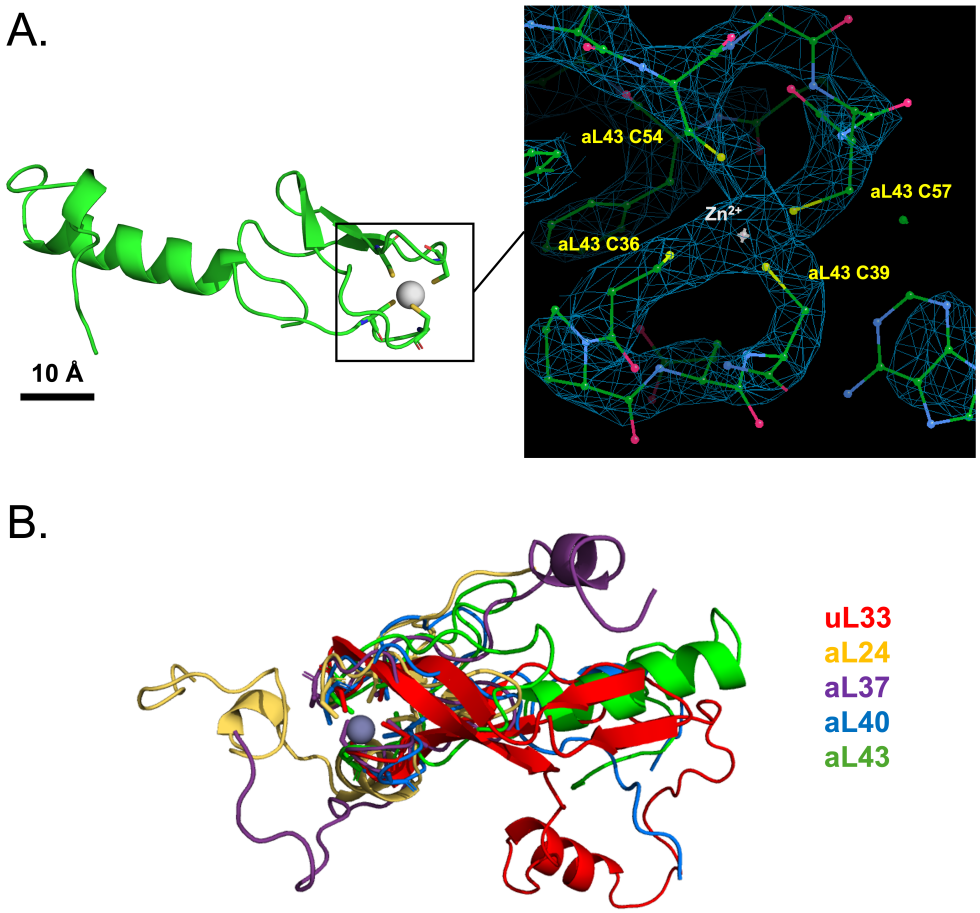



**Figure 4.28:** A. Coordination of a $Zn^{2+}$ ion by four cysteine residues in aL43 (C36, C39, C54, and C57), with the cartoon structure on the left and the model built into the 50S map on the right.
B. The structures of the five zinc-binding proteins in the LSU, aligned by zinc-binding cysteine residues.

predicted by the low confidence score from the AlphaFold2 structure C-terminus as shown in Figure 4.31. The role of this protein is unknown.

#### 4.3.4.6 aL32

aL32 (HAMAP MF_00810) is modelled into both maps from P2 to A129, representing approximately 96% of residues (model in 70S map shown in Figure 4.32). Not much is known about the role of this protein, aside from being present in the LSU.

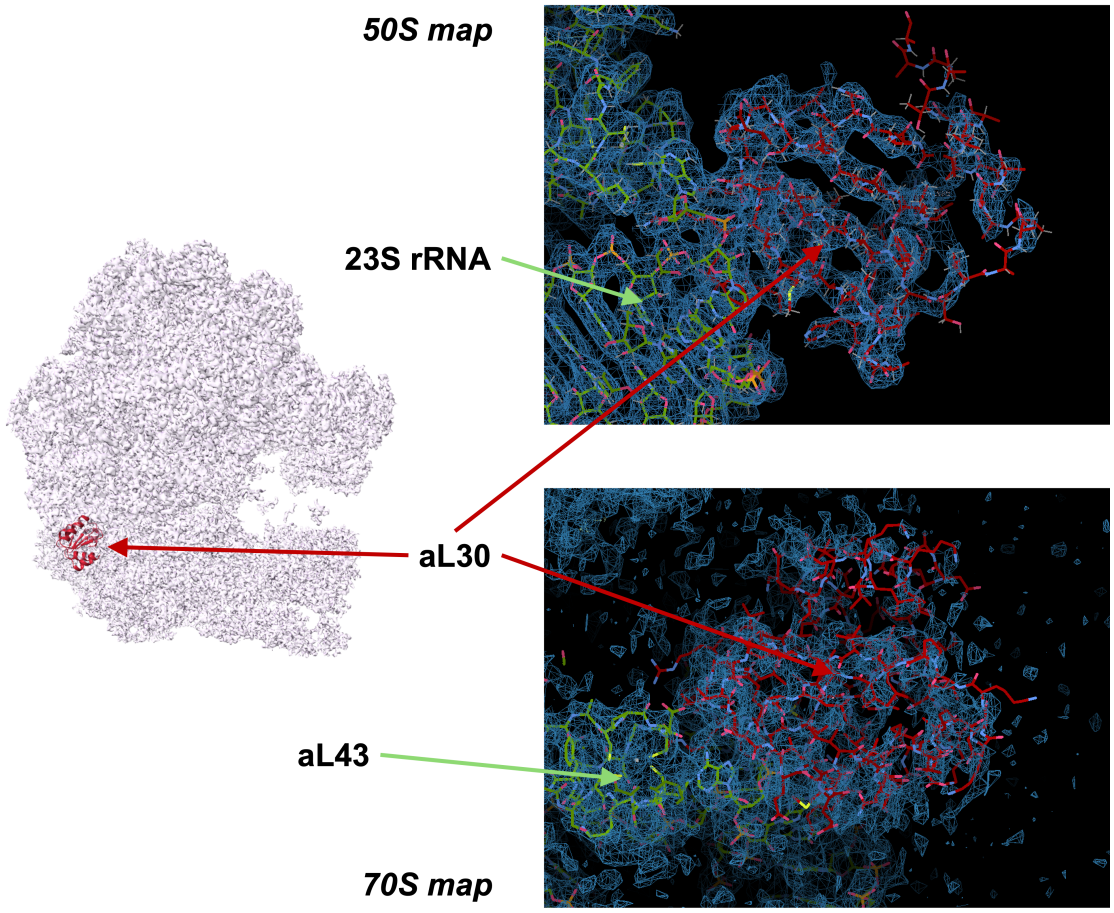



**Figure 4.29:** Ribosomal protein aL30 is shown in the context of the 70S ribosome. The model is shown built into both the 50S map (above) and the lower-resolution 70S map (below), clearly showing that the protein is better resolved in the 50S map.

#### 4.3.4.7 aL39

aL39 (HAMAP MF_00629) is a small protein, modelled in both maps from A2 to G52 (see Figure 4.33 for the model in the 70S map). This represents approximately 98% of residues. It forms part of the lining of the polypeptide exit tunnel, as seen in Figure 4.23, and agrees with work by Micic *et al.* (2022), which shows the importance of aL39 in protein folding.

#### 4.3.4.8 aLP1

aLP1, like uL11 and uL10, is part of the ribosomal stalk (HAMAP MF_01478). It is also known as L12. While no density was visible in either map, six copies

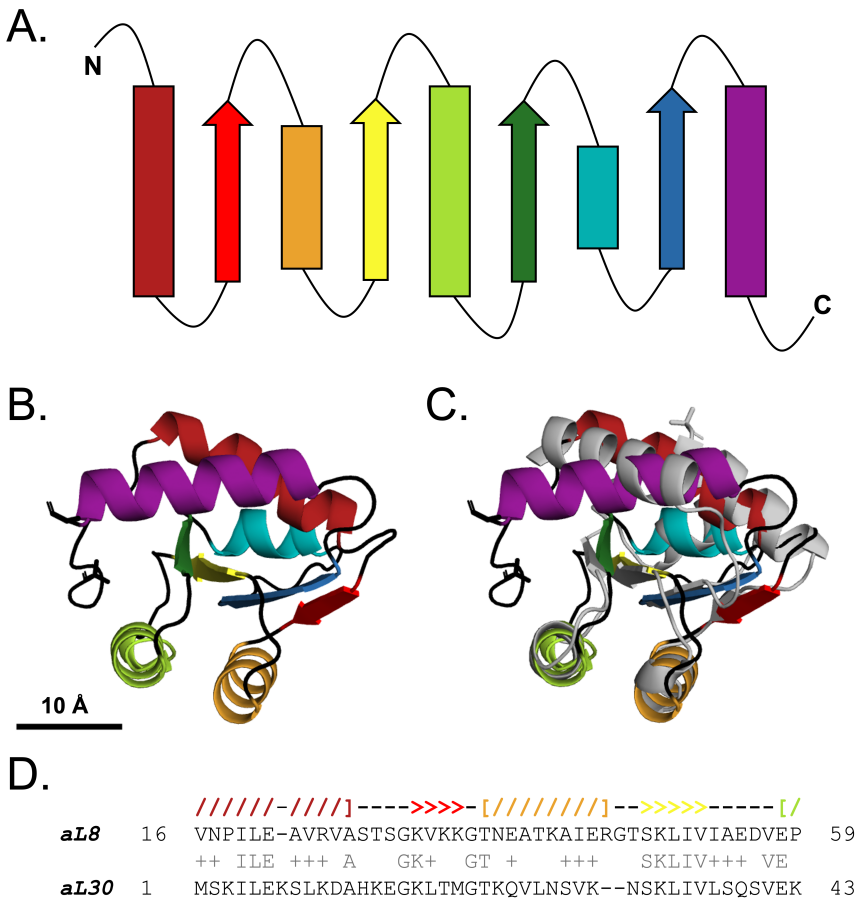



**Figure 4.30:** A. Secondary structure in aL8 and aL30.
B. Tertiary structure of aL8, coloured as in A.
C. The structure of aL30 (light grey), aligned with aL8.
D. Alignment of aL8 and aL30 sequences, from BLAST (Altschul *et al.*, 1990, 1997). Identities: 15/45 (33%), Positives: 28/45 (62%), Gaps: 3/45 (6%).

were putatively placed in the homology model in Figure 4.17.

### 4.3.5 rRNA model building in the LSU

The 5S rRNA was not present in the 50S map, explaining the lack of 5S-binding proteins identified in the map. The 5S RNP involves the 5S rRNA, and r-proteins uL5, uL16, uL18, uL30, and aL21, which was determined by building the model of the 5S rRNA into the 70S map (Figure 4.34 shows the map, Figure 4.15 shows the model of the full RNP).

Approximately 80% of the residues of the 23S rRNA were able to be modelled

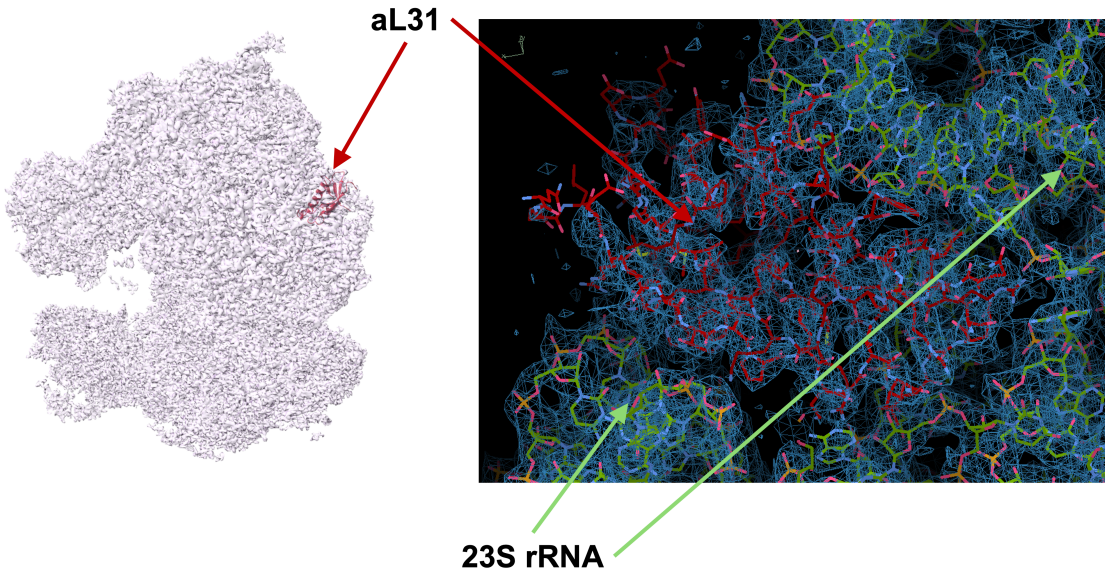



**Figure 4.31:** Ribosomal protein aL31 shown in the context of the 70S ribosome, on the periphery of the structure. The protein model is shown in the 70S density on the right, including interactions with the 23S rRNA.

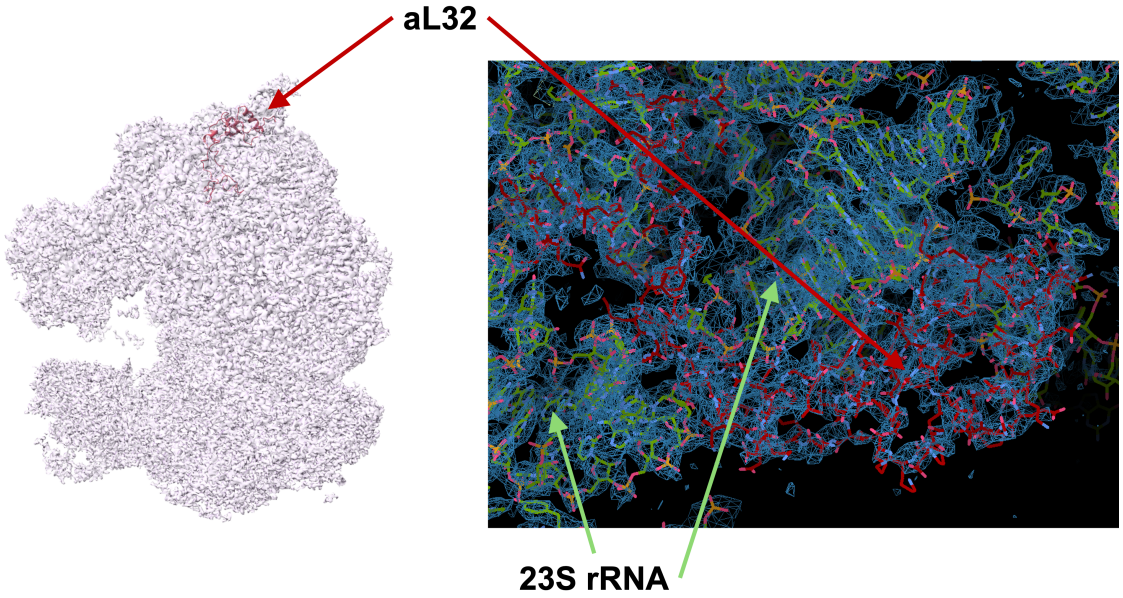



**Figure 4.32:** Ribosomal protein aL32 shown in the context of the 70S ribosome on the left, and built into the 70S density on the right.

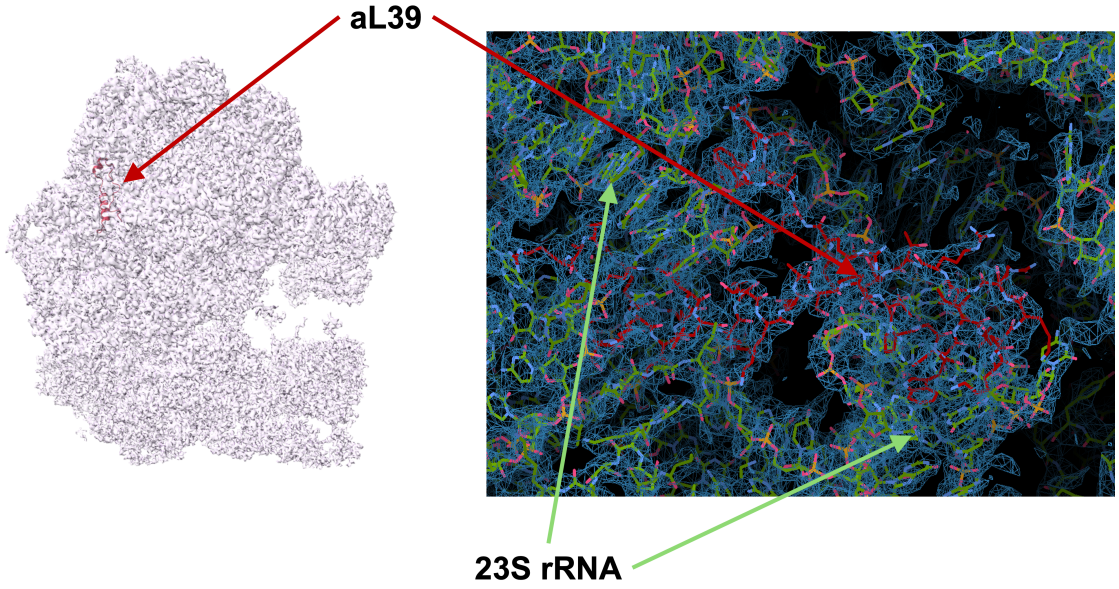



**Figure 4.33:** Ribosomal protein aL39 shown in the context of the 70S ribosome on the left, and built into the 70S density on the right.

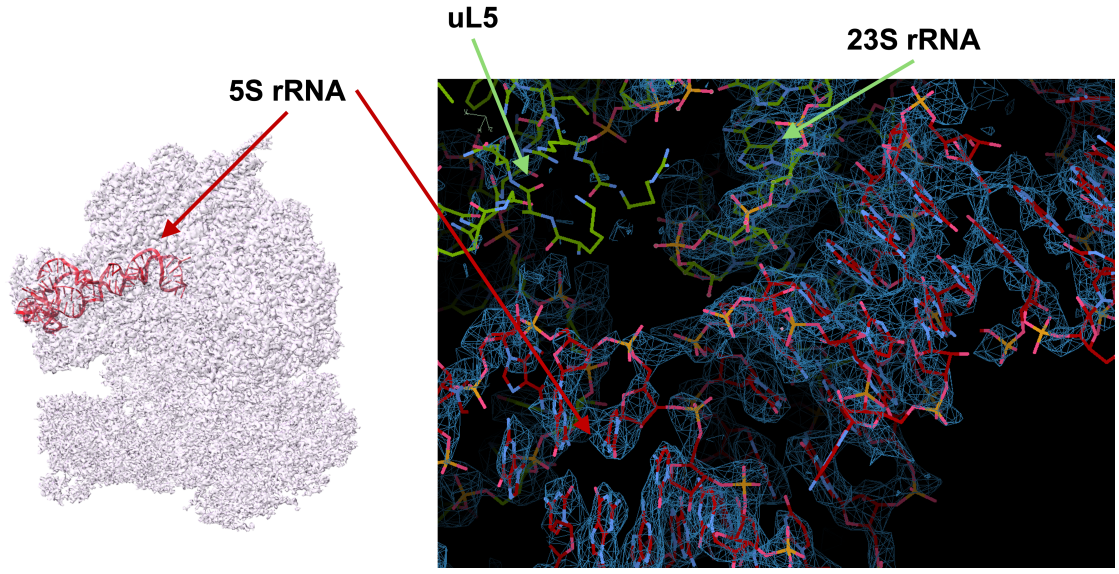



**Figure 4.34:** Ribosomal RNA 5S, shown in the context of the 70S ribosome. Part of the rRNA model is shown built into the 70S density on the right, including an interaction with uL5 and the 23S rRNA.

in both maps, shown in Figure 4.35 for the 70S map. By using the domain architecture defined by Petrov *et al.* (2013), the missing 23S rRNA structures could be identified. Helices H26a and H72 are not visible from domain 0, which is the central core of the 23S rRNA and therefore was expected to be all resolved. H27, H28, H29, H31, H40, H45, and H46 are not visible from domain II, and helices H80, H81, and H88 are not visible from domain V. As most of the rRNA was able to be resolved, this agrees with the assessment that the rRNA is "rock-like" (Assmann *et al.*, 2023), being highly structured and allowing the structure to be solved by cryo-EM with relatively few gaps.

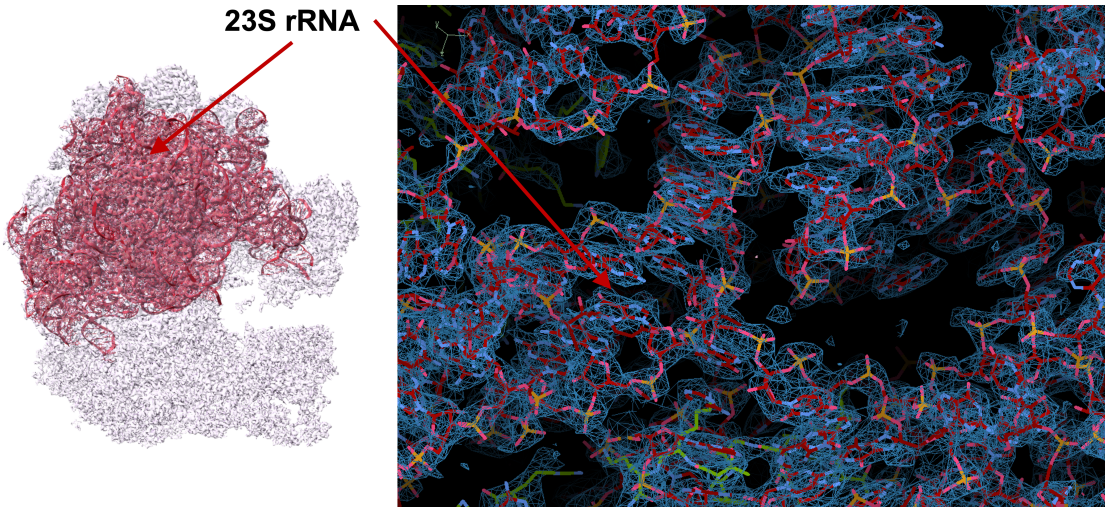



**Figure 4.35:** Ribosomal RNA 23S, shown dominating the LSU in the 70S ribosome structure. Part of the 23S rRNA model built into the 70S density is down on the right.

### 4.3.6 Unmodelled density in the LSU

There is one section of clear protein density that has not been successfully modelled, despite clearly being defined and ModelAngelo predicting the presence of protein. 21 amino acids were built into the map by ModelAngelo *de novo*, however these were not found in any protein in the known *N. maritimus* ribosome or wider proteome (Figure 4.36). A different run of ModelAngelo built 13 residues of uL33 in this position (see subsubsection 4.3.3.13), but there was a clear mismatch between the density and the sidechains of the following residues in the uL33 sequence.

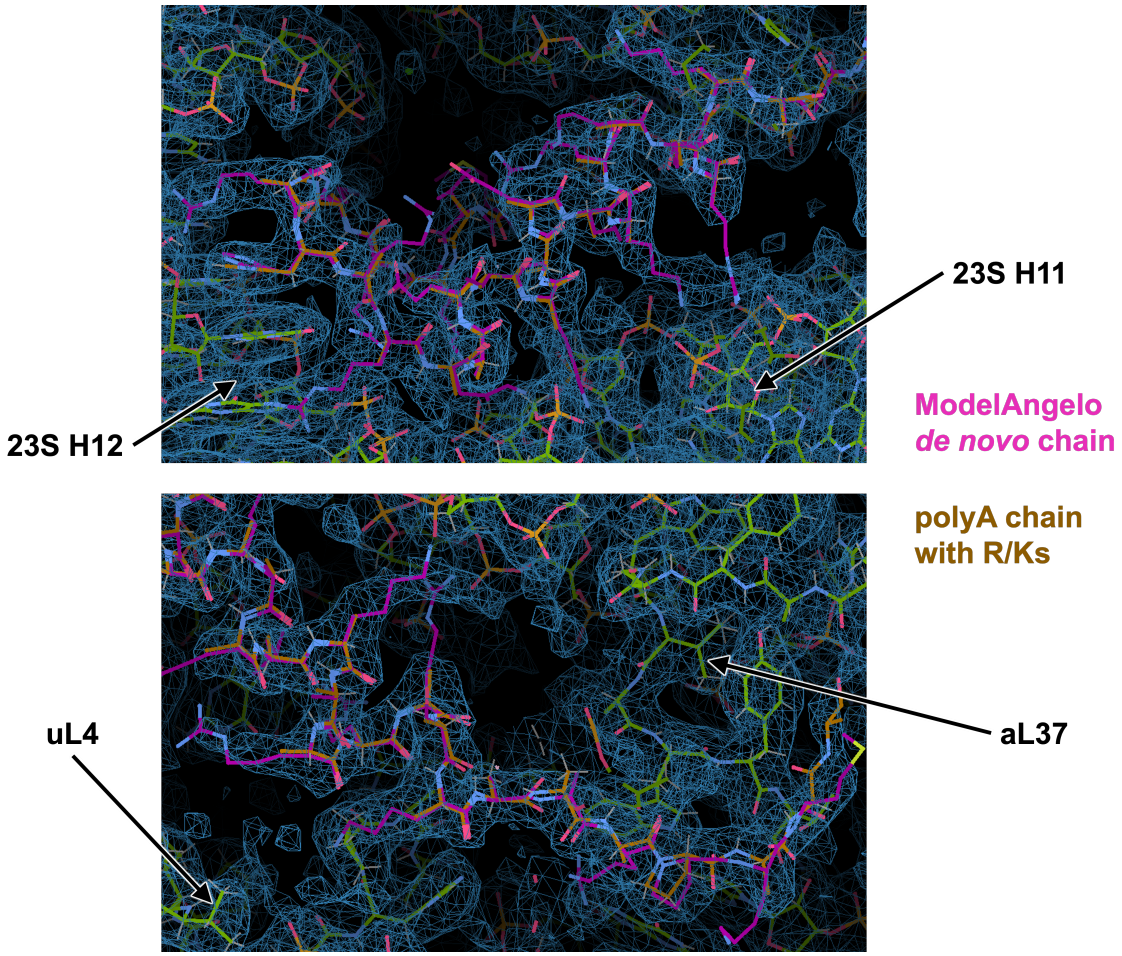



**Figure 4.36:** Two models of amino acid chains in a region of protein density that was not successfully modelled using known protein sequences from *N. maritimus* ribosomal proteins.

Several segments of rRNA density are unmodelled, where ModelAngelo was not able to determine an RNA sequence (Figure 4.37). By looking at the positioning of the box 1. section of density relative to the rest of the ribosome, it appeared that the RNA would either be a 5' or a 3' extension of the predicted 23S rRNA, which did not make sense. A putative model of poly-cytosine was built into the map. Boxes 2. and 3. are also clear extensions of the rRNA chain, however the density was of low quality and programs were unable to refine the model in the map. In box 2., a poly-cytosine fragment was built in one half of the helix, because the sequence was not perfectly complimentary and could not be located in the missing segments of the sequence.

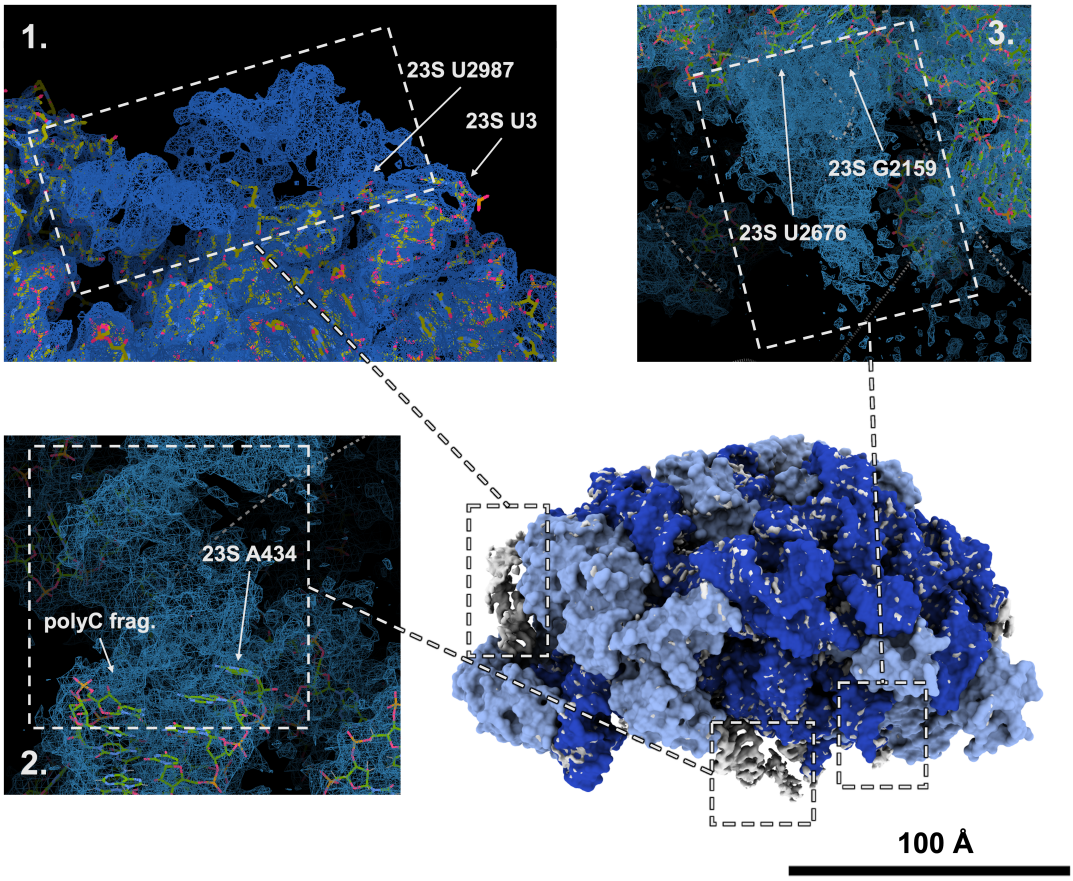



**Figure 4.37:** 1., 2., and 3. are major segments of RNA density that were unable to be modelled, shown in the context of the 50S map. In the structure, light blue indicates protein, and dark blue indicates rRNA, shown in surface representation. Grey indicates unmodelled density.

Other segments of unmodelled density were identified using a difference map. Mostly these consist of unconnected fragments on the ribosomal surface, which are potentially the loops of proteins where a backbone is unable to be traced. Unmodelled density may also represent bound ions or water molecules within the structure.

# 4.4 Discussion

## 4.4.1 Ribosomal protein content

### 4.4.1.1 Proteins encoded in the genome

The genome of *N. maritimus* encodes for 33 LSU proteins, which can be divided into the 19 universal LSU proteins and the 14 that are widespread in archaea and eukaryotes (see Table 4.3). Of the latter, 7 are in contact with the minimal rRNA as defined by Mears *et al.* (2002). This leaves 7 that are not defined in either the minimal rRNA contact or Min1 group of r-proteins described by Mushegian (2005). Therefore, this suggests that the 7 r-proteins that are in contact with the minimal rRNA are more likely to be widely conserved across the domains of Archaea and Eukarya, and the other 7 are more varied by phylum. The work of Kisly and Tamm (2023) found that there are also several functional riboprotein homologues between archaeal and eukaryotic ribosomes, with very low sequence similarity. Therefore, more understanding of the relative positions of riboproteins in ribosomes across domains will help further elucidate evolutionary relationships between ribosomes, as positional and structural data can be used where traditional methods of identifying homology by sequence comparison are unsuccessful.

There are a number of other known archaea-eukarya proteins that are not found in the genome of *N. maritimus* when performing a homology search. These are aL13, aL14, aL20, aL33, aL34, aL38, aL41, and aLP2.

### 4.4.1.2 Comparison to published archaeal ribosome structures

Four proteins are present in the *T. kodakarensis* ribosome LSU structure, which were not present in the *N. maritimus* genome: aL14, aL20, aL34, and aL41 (Sas-Chen *et al.*, 2020). The genome also encodes for aL33.

No proteins were present in the *H. marismortui* LSU structure that were not also found in *N. maritimus* (Klein *et al.*, 2004). The genome, however, does encode aL20.

**Table 4.3:** R-protein content overview. All of the r-proteins identified in the *N. maritimus* genome are listed. Min1 is defined by Mushegian (2005) as proteins which are both universal and in contact with the minimal rRNA that was defined by Mears *et al.* (2002).

| **Protein** | **Universal** | **Min1** | **Min rRNA** | **50S map** | **70S map** |
|---|---|---|---|---|---|
| uL1 | ✓ | | | | |
| uL2 | ✓ | ✓ | ✓ | ✓ | ✓ |
| uL3 | ✓ | ✓ | ✓ | ✓ | ✓ |
| uL4 | ✓ | ✓ | ✓ | ✓ | ✓ |
| uL5 | ✓ | | | | ✓ |
| uL6 | ✓ | ✓ | ✓ | ✓ | ✓ |
| aL8 | | | | ✓ | ✓ |
| uL10 | ✓ | ✓ | ✓ | | |
| uL11 | ✓ | ✓ | ✓ | | |
| uL13 | ✓ | ✓ | ✓ | ✓ | ✓ |
| uL14 | ✓ | ✓ | ✓ | ✓ | ✓ |
| uL15 | ✓ | ✓ | ✓ | ✓ | ✓ |
| aL15 | | | ✓ | ✓ | ✓ |
| uL16 | ✓ | ✓ | ✓ | | ✓ |
| uL18 | ✓ | | | | ✓ |
| aL18 | | | | ✓ | ✓ |
| aL19 | | | | ✓ | ✓ |
| aL21 | | | ✓ | | ✓ |
| uL22 | ✓ | ✓ | ✓ | ✓ | ✓ |
| uL23 | ✓ | | | ✓ | ✓ |
| uL24 | ✓ | | | ✓ | ✓ |
| aL24 | | | ✓ | ✓ | ✓ |
| uL29 | ✓ | | | ✓ | ✓ |
| uL30 | ✓ | | | ✓ | ✓ |
| aL30 | | | | ✓ | ✓ |
| aL31 | | | ✓ | ✓ | ✓ |
| aL32 | | | ✓ | ✓ | ✓ |
| uL33 | ✓ | | | | |
| aL37 | | | ✓ | ✓ | ✓ |
| aL39 | | | ✓ | ✓ | ✓ |
| aL40 | | | | | ✓ |
| aL43 | | | | ✓ | ✓ |
| aLP1 | | | | | |

The *P. furiosus* LSU structure has aL14, aL20, aL33, aL34, and aL41 in addition to the proteins in *N. maritimus* (Armache *et al.*, 2013; Garcia-Campos *et al.*, 2025).

The structure of the *M. thermautotrophicus* LSU has aL14 and aL20, as well as translation initiation factor 6 present in the LSU, which are not seen in the *N. maritimus* genome (Greber *et al.*, 2012). The genome also encodes aL34.

Three proteins, aL14, aL20, and aL34, are present in the *S. acidocaldarius* LSU and not in *N. maritimus* (Wang *et al.*, 2023b). Additionally, the structure has elongation factor 2 and three unnamed proteins with the UniProt accession codes Q4J9G6, Q4J953, and F2Z6C4. BLAST searches returned no results for the first two unnamed proteins, and suggested "L47A" for F2Z6C4. None returned any homology in the *N. maritimus* genome.

The *P. calidifontis* LSU structure has aL13, aL14, aL20, aL34, and aL38, as well as two unnamed proteins with the UniProt accession codes A3MTM7 and A3MWU0 (Nissley *et al.*, 2025). Neither of those proteins were identified in the genomes of *N. maritimus* or *S. acidocaldarius*.

The evolutionary relationships between the above species, *N. maritimus*, and the bacteria *T. thermophilus* are shown in Figure 4.38. *N. maritimus* is the only archaea without aL20, suggesting it was lost at least by the time the class Nitrososphaeria separated from the other classes. aL33 is not in any of the archaea studied in the kingdom Thermoproteati, which also points to an earlier loss, however this is a very small sampling of genomes and a more systematic, genomic study would be needed to pinpoint where the proteins were lost or gained.

None of the archaeal species for which ribosome structures exist encode for aLP2 in the genomes.

## 4.4.2 Structural features

### 4.4.2.1 Comparison of LSU model between maps

Where a protein was able to be built into both maps, the models were aligned and the RMSD was calculated in PyMOL (results shown in Table 4.1, Table 4.2). The

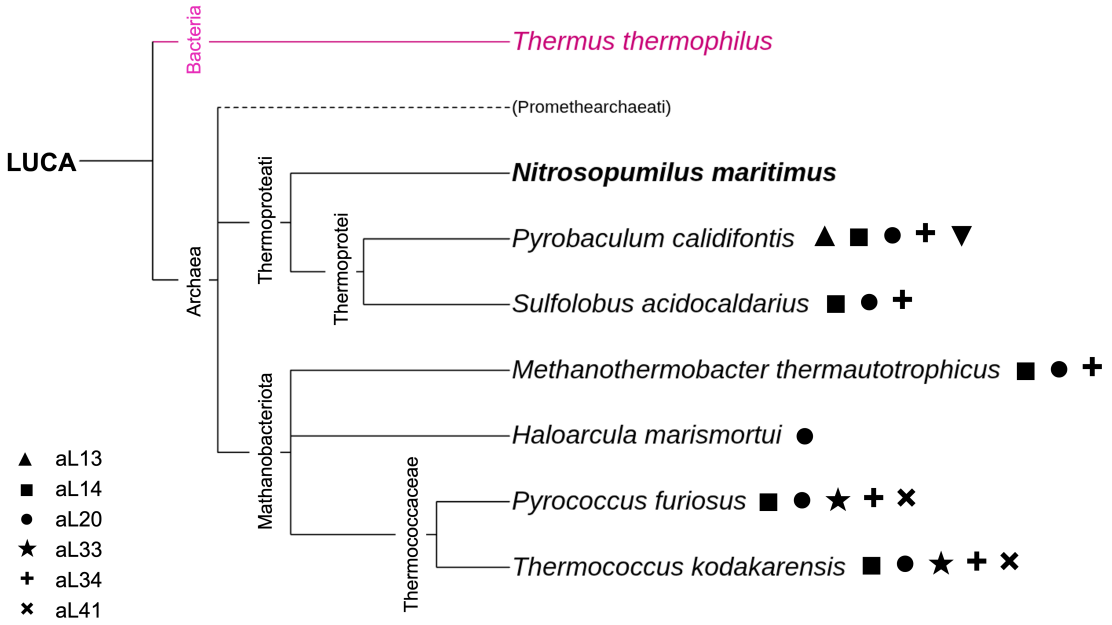



**Figure 4.38:** aL proteins mapped on a phylogenetic tree, generated using NCBI taxonomy in phyloT (Letunic, 2015; Sayers *et al.*, 2019).

RMSD for every protein was $< 1$ Å, with the exception of aL30. This suggests that the conformations of the proteins are extremely similar between the two maps, and that the proteins do not change in conformation when the 70S ribosome assembles. However, it is likely that the segments of proteins that would be changing conformation would also be flexible in the map of the 50S particle. This suggests that it is the proteins that cannot be modelled into the 50S map, or have an increase in the percentage built between the maps (such as uL15), are the ones undergoing more major conformational changes. By taking the proteins that are not seen as much in the 50S map as in the 70S map, this gives a list of proteins that are most likely to be changing conformation upon 70S assembly: uL2, uL5, uL15, uL16, uL18, uL33, aL21, and aL40.

#### 4.4.2.2 rRNA density

A low-pass filter at 15 Å resolution was applied to the cryo-EM maps, which allowed some more density to be seen at a lower resolution. This primarily allowed regions of unmodelled rRNA to be identified. However, it was not clear how these

segments of rRNA would connect to the modelled regions, meaning that specific sequences could not be built. RNA structure prediction software, for both 2D and 3D, was unable to provide solutions that allowed the building of anything other than poly-cytosine chains. Not all of these were built due to the map quality, as some of the models could not be refined in Coot. Examples of modelled and unmodelled density are shown in Figure 4.39.

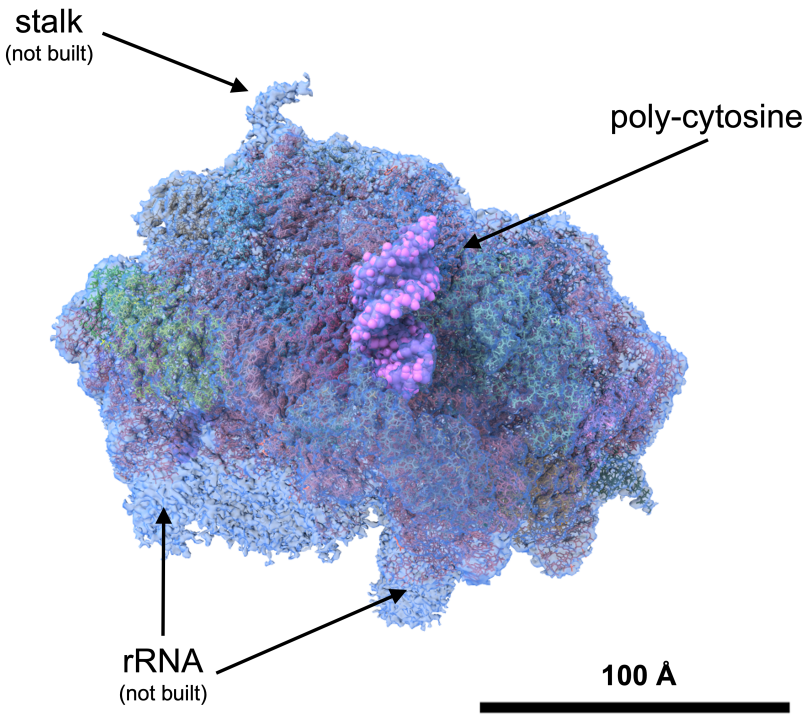



**Figure 4.39:** Some density in the 50S map was modelled with poly-cytosine, shown in purple spheres. Other regions could not be modelled successfully with a filler chain.

One area of unbuilt rRNA density is a particular mystery: Figure 4.37, box 1. A number of potential sequences were tested, including the residues before and after the 23S rRNA gene in the genome. Neither the preceding or proceeding residues fitted into the density, when considering the pattern of purines and pyrimidines. Also, as the area was seen in both maps, it did not make sense that it was a pre-processed 23S rRNA. It was also considered that the rRNA trace had been started incorrectly and was offset, however there was no satisfying solution when attempting

to retrace the backbone through this additional density, and ModelAngelo was confident in the original positioning.

#### 4.4.2.3 Visibility of proteins

Once the ribosome was established as a ribozyme, there arose the question of what role the proteins had to play, given their necessity in ribosomal function (Noller *et al.*, 1992). The r-proteins are involved in the structural stability of the ribosome, but are not simply structural blocks. If they were just rigid bodies, they would be visible in structural methods such as cryo-EM. The fact that proteins such as uL10 and uL11 are never seen in 50S structures indicates that proteins are involved in dynamic roles of the ribosome such as nascent polypeptide folding (Moore, 2009). Also, not all of the proteins involved in interactions with the minimal RNA are seen; much of the conserved rRNA is flexible, and thus the proteins associated with these rRNA segments are also flexible, with evolutionary pressure to maintain this. This points to the rRNA and proteins working in tandem for some roles associated with ribosomal function.

#### 4.4.2.4 The SSU

The ribosome SSU is known to be dynamic relative to the LSU, due to the nature of translocation of the ribosome along mRNA (Hassan *et al.*, 2023). As the ribosomes purified from *N. maritimus* were not stalled in a specific state, it can be assumed that the SSU could adopt a number of conformations relative to the LSU. These would be averaged out in SPA, thus worsening the resolution of the SSU relative to the LSU. Multi-body refinement was used to mask and align particles on the SSU in the 70S map, but were unable to improve the resolution beyond 3.6 Å. The local resolution of the ribosome structure is shown in Figure 4.40. Further improvements to the structure of the SSU would start with better masking of the SSU for alignment using multi-body refinement, allowing the particles to be aligned only on the SSU and ignoring the LSU. Masks based on previously-solved SSU structures would allow the specific selection of SSU proteins and rRNA. The collection of more data would also help improve the resolution of the SSU as more particles can contribute to the

structure, and could allow an alternative method of resolution improvement where the ribosome particles are separated by conformation. Finally, the optimisation of the ribosome purification method could yield 70S particles in only one conformation.

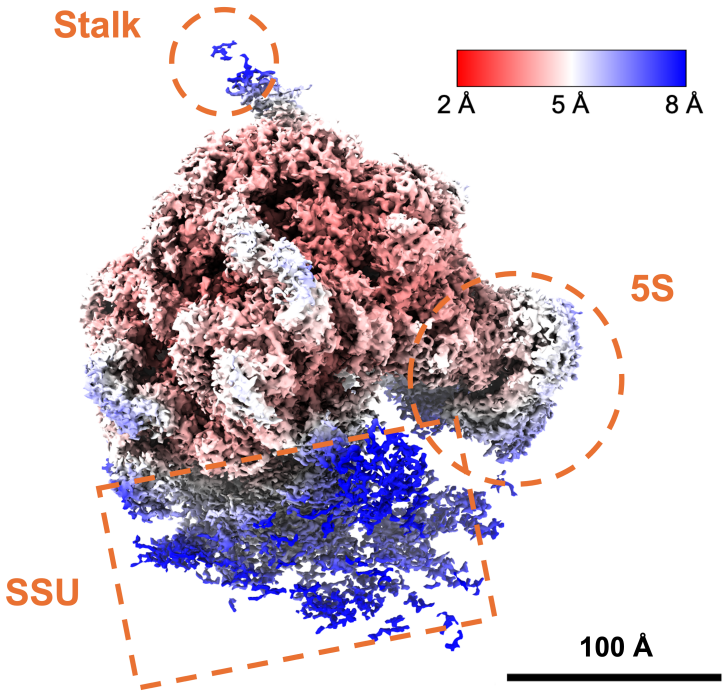



**Figure 4.40:** Map of the *N. maritimus* 70S ribosome, coloured by local resolution. The SSU is clearly at a poorer resolution than the LSU.

The ribosome purification method was based on published methods for ribosomes from other species. The first round of purification did not produce complete 70S ribosome particles, suggesting that the buffer used did not allow for the large and small subunits to stay associated and also caused the dissociation of the 5S RNP. Further improving the purification method, by considering the internal cellular environment of *N. maritimus*, would allow for more stable 70S particles to be isolated, and could be combined with factors to stall the ribosome in a specific state to elucidate different conformations more easily. Novel purification strategies such as RAPPL (Erath *et al.*, 2025), or ribosome-specific workflows leveraging recent software improvements such as cryoPRISM (May *et al.*, 2026), could also enable the improvement in data quality.

### 4.4.3 Auto-building

#### 4.4.3.1 AlphaFold2

AF2 accurately predicted the structures of many of the r-proteins in the *N. maritimus* LSU. Where the protein had any length of loop or disorder, such as uL24 or aL21, the outcome was less successful, and recognised as such in the Predicted Local Distance Difference Test (pLDDT). The use of AF2-predicted structures, aligned with a previously-solved archaeal ribosome structure, proved to be a useful homology model. However, it did require some knowledge on where the protein would be located to be an aid in building.

#### 4.4.3.2 ModelAngelo

The vast majority of the LSU was able to be modelled using ModelAngelo in both cryo-EM maps. The software was provided with the predicted sequences of the rRNA and r-proteins from the *N. maritimus* genome, which enabled the fast and accurate building of most of the residues that were visible in the LSU at resolutions better than 3.5 Å. In all cases, the model built by ModelAngelo matched what was built manually, even using early pre-publication versions of the software. This aided in the development of ModelAngelo by providing statistics for the comparison of human-built models and machine-built models (Jamali *et al.*, 2022).

## 4.5 Conclusions

The data in this chapter provides a model for the structure of the large subunit of the ribosome from *Nitrosopumilus maritimus*, an archaeon that is important to the global nitrogen cycle.

The published archaeal ribosome structures are phylogenetically distant from *N. maritimus*, being in the same kingdom as two other species but with no closer relations. The cryo-EM structures of the *N. maritimus* 70S ribosome and a 50S particle from this research provide insight into the structure of this ribosome, adding to the small body of literature and data on archaeal ribosomal subunits. This adds another data point for the presence or absence of specific ribosomal proteins, helping map the change in protein composition and potentially leading to a better understanding of the role of particular r-proteins.

High-resolution structures, such as the ones in this research, allow for the detailed investigation of the structure of biological molecules. In these structures, density can be seen for bound metal ions such as the zinc ions that are coordinated by four cysteine residues, a motif found in five proteins within the LSU. Much of the backbone of the rRNA can be unambiguously placed, even outside of double-stranded rRNA structures. However, at the resolutions of these structures, it can often be difficult to identify the nucleic acid residue more specifically than as a purine or a pyrimidine.

Additionally, the work presented here exemplifies one of the pitfalls of structure determination using cryo-EM: heterogeneous resolution. Most notably, there are large discrepancies in the resolution of the LSU and the SSU in the 70S map. While the overall resolution of the map can be calculated as 3.4 Å, the SSU has an average resolution of only 3.8 Å, which is not sufficient for confident atomic model building. A number of r-proteins and a reasonable amount of the 16S rRNA was built, but there are large areas of the map that remain intractable. For the LSU, while the centre of the subunit was well-resolved, the periphery of the subunit proved more of a challenge in many areas. The ribosomal stalk is not visible in

the high-resolution map, and only a short section of it appears at low resolutions. This is not unexpected, compared to previous ribosome structures. A number of proteins that are present in the genome and have been resolved in other ribosome structures are also not seen in the *N. maritimus* ribosome structure, such as uL1. Protein domains that are predicted to be intrinsically disordered are also not seen. Around 20% of the 23S rRNA is not well-resolved in the maps.

The identity of the particle that led to the 50S map is unknown. This LSU-like particle does not contain the 5S rRNA or the associated r-proteins, and two other proteins are also missing; it is clear that there is no density present for them. The origin and reason behind the abundance of this particle remain to be discovered.

Two powerful new structural tools, AlphaFold2 and ModelAngelo, were developed and released during the course of this research. AlphaFold2, far from completely solving the protein folding problem, has nevertheless revolutionised structural biology (Bowman, 2024). By comparing the AF2-predicted protein structures to those already built in the ribosome, it was determined that AF2 was accurate and a useful starting point for modelling r-proteins, once the correct location for the protein was identified. ModelAngelo does not have the issue of identifying protein location, as it builds models into cryo-EM maps using a graph neural network approach. This research aided a small part in the development of the software, enabling initial testing on protein-nucleic acid complexes. In turn, the software aided in the building of both r-proteins and rRNA, particularly where a protein did not exist in the homology model. Additionally, it was able to build in domains of proteins that did not have strong secondary structure and were therefore not well-predicted by AF2, but were visible in the density.

Overall, the work presented here has determined a novel archaeal ribosome structure, allowing for the comparison to other published archaeal ribosome structures and filling in the gaps of a poorly-understood domain.

## 4.6 Future work

With the ribosome being such a large complex, there are many potential avenues for further investigation. The most obvious is to improve the resolution of the 30S subunit, and therefore complete the model of the SSU. This will require the collection of more single-particle data, and more focused refinement based on the SSU. With the model of the SSU built, it will allow the inspection of the inter-subunit interactions, and the determination of a complete set of protein-protein and protein-nucleic acid interactions, as carried out in Timsit *et al.* (2021). This will also provide the description of the SSU orientation relative to the LSU, allowing the state of the ribosome to be identified.

Additionally, the biological context of the 50S structure could be investigated, to see if it is an intermediate in the assembly process or an artefact of the ribosome purification process. The dissociation of the 5S RNP from the complex suggests weaker interactions with the LSU, and a number of proteins are also missing, which could relate to the strength of the association with the ribosome. Interestingly, aL30 seemed to have a stronger association with the LSU when the SSU was not present, which could link to the promiscuity of proteins like aL30 and aL8 as discovered in Daume *et al.* (2017).

The existence of polysomes, which is the binding of multiple ribosomes to one mRNA molecule, has been known for a long time (Warner *et al.*, 1963). Polysomes are part of the greater question about the distribution of ribosomes throughout cells both with and without organelles. In the initial tomograms of *N. maritimus* (see subsubsection 4.1.4.2 or Chapter 5), ribosomes were not seen as being uniformly distributed in the cytoplasm. Instead, some were observed either in clusters resembling polysomes or in close proximity to the cell membrane. Research to classify the ribosomes in the tomograms by their location could lead to sub-tomogram structures being solved for each class, allowing for the comparison in the structural features between each.

Ribosomes are known to hibernate, requiring hibernation factors and the unusual conformation of subunits (McLaren *et al.*, 2023; Helena-Bueno *et al.*, 2024; Nissley *et al.*, 2025). Tomograms of cells may also reveal the conformation of hibernating ribosomes in *N. maritimus*, which could be compared and contrasted with those of other recent research.

Molecular dynamics simulations of ribosomes have previously been used to elucidate steps of translation (Arenz *et al.*, 2016; Šponer *et al.*, 2018; Świderek *et al.*, 2015; Levi *et al.*, 2019). This could also be an avenue for investigating how the ribosome of *N. maritimus* interacts in a polysome, or with the unique ether-linked lipid bilayer found in archaea.

# 5

# Structural studies of archaeal S-layers

## Contents


**5.1 Introduction . . . . . 187**
5.1.1 S-layers . . . . . 187
5.1.2 Archaea . . . . . 190
5.1.3 Previous work referenced in this chapter . . . . . 194
5.1.4 Aims of this chapter . . . . . 196
**5.2 Materials and Methods . . . . . 198**
5.2.1 *N. maritimus* . . . . . 198
5.2.2 *S. solfataricus* . . . . . 200
5.2.3 *P. oguniense* and *P. arsenaticum* . . . . . 201
5.2.4 *H. volcanii* . . . . . 202
5.2.5 Other analysis . . . . . 204
**5.3 Results . . . . . 205**
5.3.1 Structural studies of the S-layer of *N. maritimus* . . . . 205
5.3.2 Structural studies of the S-layer of *S. solfataricus* . . . . 210
5.3.3 Structural studies of the S-layers of *P. oguniense* and *P. arsenaticum* . . . . . 217
5.3.4 MD simulations of the *H. volcanii* S-layer . . . . . 227
**5.4 Discussion . . . . . 240**
5.4.1 Structural comparisons . . . . . 240
5.4.2 Sequence comparisons . . . . . 244
5.4.3 Using MD to predict identity of ions in cryo-EM density 247
**5.5 Conclusions . . . . . 249**
**5.6 Future work . . . . . 252**

# Statement of Authorship

The work in this chapter was conducted by the doctoral candidate, with the exception of the production of data as described in subsubsection 5.2.1.1 and subsubsection 5.2.1.2, which was performed by Lennart Pagani (Utrecht University and Sir William Dunn School of Pathology, University of Oxford), Sofie van Dorst (Utrecht University and Sir William Dunn School of Pathology, University of Oxford), and Andriko von Kügelgen (Sir William Dunn School of Pathology, University of Oxford, and MRC Laboratory for Molecular Biology, Cambridge, UK). The work in subsubsection 5.2.1.3 and subsubsection 5.2.1.4 was performed in collaboration with Andriko von Kügelgen as detailed in those sections, and is covered in subsubsection 5.3.1.3. The author would like to acknowledge Dr. Mart Krupovic (Institut Pasteur) for the samples of *S. solfataricus*, *P. oguniense*, and *P. arsenaticum*, and Professor Phill Stansfeld (University of Warwick) for the expertise and support during the work in subsection 5.2.4 and subsection 5.3.4. All other work was conducted by the doctoral candidate.

# 5.1 Introduction

## 5.1.1 S-layers

Prokaryotes have evolved a number of strategies to enable the maintenance of cell shape and protection from the environment that are vitally important for cell survival. Gram-positive and Gram-negative bacteria both possess a peptidoglycan cell wall, with one or two lipid membranes respectively. Another solution that more than 80% of archaea and 50% of bacteria have evolved is having a proteinaceous, paracrystalline lattice on the surface of the cell, outside of the membrane, known as an S-layer (Sleytr *et al.*, 2014; Johnston *et al.*, 2024). Depending on the species, the S-layer may anchor to the cell envelope in one of three ways (Isbilir *et al.*, 2025). Anchoring domains within a bipartite S-layer protein structure are the most common, and often this anchoring mechanism is also found in other surface proteins such as pilins. There may be attachments to the polymeric substance (such

as pseudomurein in *Methanothermus fervidus* (Arbing *et al.*, 2012)) coating the membrane. A lipid modification to the S-layer protein on the N- or C-terminus to anchor to the membrane is utilised in species such as *Haloferax volcanii* (Abdul Halim *et al.*, 2016). The architecture of archaeal S-layers is shown in Figure 5.1. They can also serve as filters for nutrients, though the additional roles likely varies by species (Engelhardt, 2007).

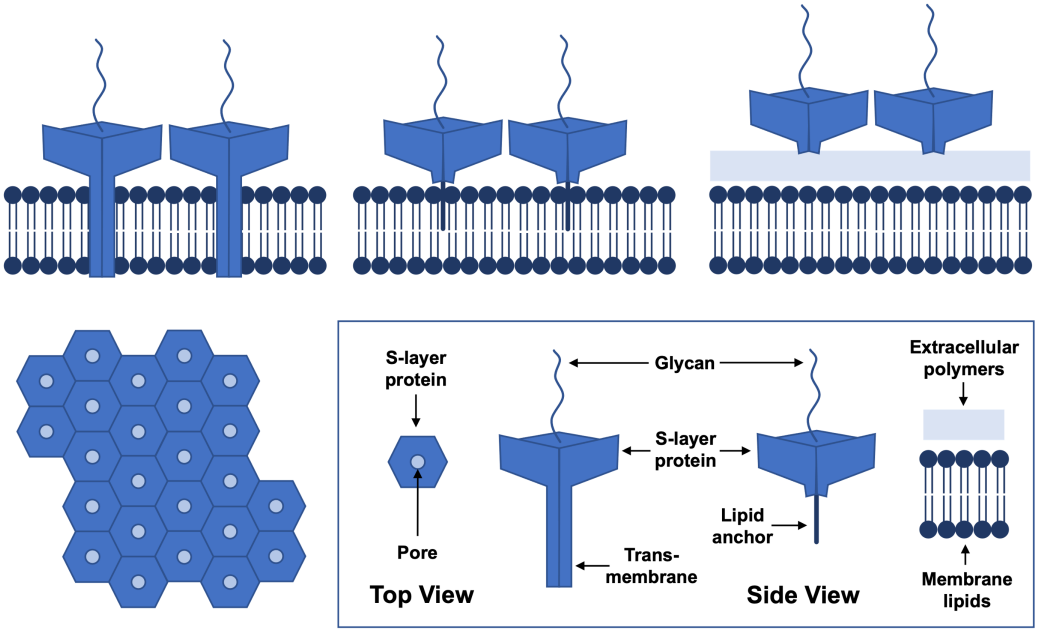



**Figure 5.1:** Overview of archaeal S-layer composition, showing the ways in which an S-layer can attach or interact with the cell membrane.

S-layers are composed of thousands of copies of one or two proteins that come together to form a symmetric arrangement of pores. Most often, this is hexagonal (p3 or p6) symmetry (Landsberg and Hankamer, 2007). Due to the geometrical constraints of covering an enclosed 3D object with hexagons, pentameric defects must exist (Goldberg, 1937), which can either be gaps with no protein present, or a pentamer form of the S-layer subunit as was observed by von Kügelgen *et al.* (2021). Some S-layers have p2 or p4 symmetry, but are not part of the work presented here.

Before archaea had been identified as a separate domain from bacteria by Woese and Fox (1977), the first archaeal S-layer was visualised using electron

microscopy by Houwink (1956). More low-resolution images were taken of archaeal S-layers through the 1990s (Phipps *et al.*, 1990, 1991), with the vast majority of species investigated possessing an S-layer; very few archaeal genera were identified where not all members had S-layers. In some archaeal species, the S-layer is the only cell envelope constituent outside of the plasma membrane, suggesting the S-layer is the earliest cell wall structure to have evolved (Albers and Meyer, 2011).

#### 5.1.1.1 Structural work on S-layers

Structural studies of S-layers have shown that the proteins are diverse in structure, easily observed by the variety of ways in which the lattice has been achieved: varied sizes of the repeating unit, the size or capping of a potential pore, the presence or absence of a pentameric version of the S-layer, and the existence of a second protein component to make the repeating unit (Rodrigues-Oliveira *et al.*, 2017; Grill-Walcher and Schäffer, 2025). The first archaeal S-layer structure at atomistic resolution was determined by X-ray crystallography (Jing *et al.*, 2002), however since then only two partial structures for archaeal S-layer proteins have been solved using this method. Currently cryo-electron microscopy, specifically tomography, is a popular method for obtaining mid- to high-resolution structures of S-layer proteins, both archaeal and bacterial (von Kügelgen *et al.*, 2020).

The work by Johnston *et al.* (2024) utilised AlphaFold2 multimer, combined with planar symmetry approaches, to predict the structure of over 150 diverse S-layers from archaea and bacteria. This allowed them to support earlier theories that S-layers did not evolve from a single common ancestor, but instead arose through multiple, independent evolutionary events as posited by Bharat *et al.* (2021). The theory gives an explanation for why S-layers from different phyla and kingdoms are very dissimilar, but not why S-layers in closely-related species can be very dissimilar.

#### 5.1.1.2 Divergence of sequence and structure

It has been observed since the first S-layer sequences were compared that there is poor sequence homology between S-layers, and this curiosity was systematically investigated by Bharat *et al.* (2021). They showed that the structure of archaeal

S-layer proteins did not always cluster along phylogenetic lines, and that even closely related species in the same genera could demonstrate a surprising lack of homology.

Despite this divergence, Bharat *et al.* (2021) also showed that there are many commonalities between S-layer structures, such as their bipartite structure with a segment for lattice formation and a segment for cell envelope attachment. It is this latter segment that often exhibits higher sequence homology. Also, they showed that S-layer proteins are often dominated by $\beta$-strand-heavy structures, such as immunoglobulin-like $\beta$-sandwiches. This was further explored by Isbilir *et al.* (2025).

#### 5.1.1.3 Molecular dynamics of S-layers

At the time of writing, there are four published papers that employ molecular dynamics simulations as part of an integrative structural investigation of an S-layer structure (von Kügelgen *et al.*, 2020, 2024; Herdman *et al.*, 2022; Sogues *et al.*, 2024). These have proven the usefulness of simulations in investigating the inter- and intra-molecular interactions S-layers are involved in, such as the work of von Kügelgen *et al.* (2024) where simulations showed how the S-layer concentrated ammonium molecules at the membrane (see subsubsection 4.1.4.1). Sogues *et al.* (2024) used simulations to investigate the role of interfacial contacts in the stability of a bacterial S-layer, validating and augmenting their experimental findings. Herdman *et al.* (2022) utilised simulations to map the positions and stability of $Ca^{2+}$ ions in the *C. crescentus* S-layer at a high resolution, to support structural and biochemical data.

### 5.1.2 Archaea

All archaea are prokaryotes, with some compositional similarities to bacteria but with several distinguishing features. These differences include having cell membranes formed of the ether-linked lipids discovered by Sehgal *et al.* (1962) rather than ester-linked lipids, and a distinct 16S rRNA and ribosomal protein composition (as discussed in Chapter 4, subsubsection 4.1.2.3). Both of these features underpinned the work of Woese and Fox (1977), which allowed the archaea to be untangled from the domain of Bacteria.

As mentioned in Chapter 4, the current tree of life posits that eukaryotes evolved from a branch of archaea, and archaea share a number of features with eukaryotes (Forterre, 2015). Despite this interesting place in evolution, relatively little is known or understood about the wide diversity of archaeal species (Dance, 2021). One reason for this is the difficulty in culturing many extremophilic archaeal species, due to their challenging growth requirements that include very high temperatures and hyper-saline conditions. Other archaeal species have poorly-understood nutrient requirements that are usually provided by the unique balance of other species in their environment. However, there are several archaeal species that are more easily culturable, and these have become model organisms where their genomes have also been tractable (Leigh *et al.*, 2011).

The archaeal species studied in the work presented here are introduced below, and the phylogenetic relationships between them are shown in Figure 5.2.

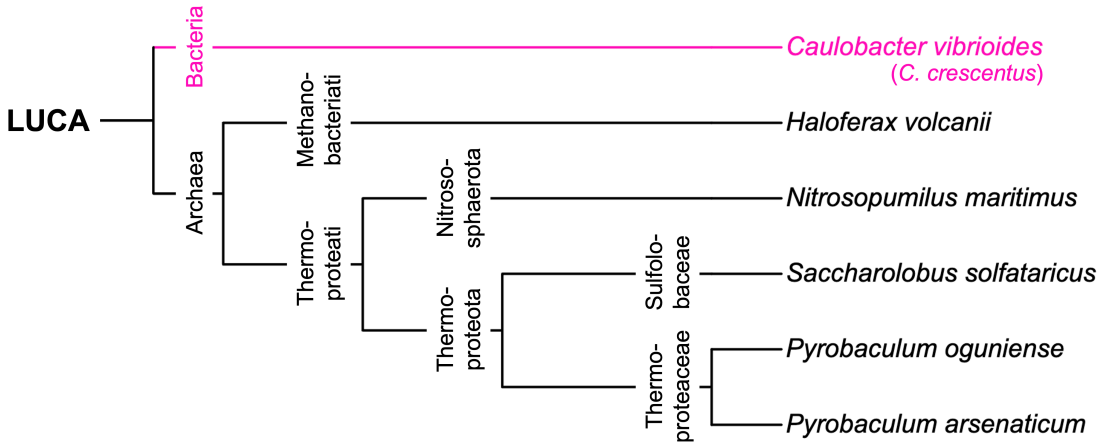



**Figure 5.2:** Phylogenetic tree of the archaeal species for which S-layers were studied in this work. Also shown is *C. vibroides*, also known as *C. crescentus*, a bacteria with a well-studied S-layer referenced above (von Kügelgen *et al.*, 2020).

#### 5.1.2.1 *Nitrosopumilus maritimus*

At 0.2 µm in diameter, *Nitrosopumilus maritimus* is one of the smallest known free-living organisms in the world, and a sea-dwelling member of the AOA (see Chapter 4, subsubsection 4.1.4.1 and Figure 4.4). The genome was first sequenced

by Richardson and Stahl (2007), and the analysis of the proteome by Qin *et al.* (2017) allowed the identification of two genes that were predicted to be S-layer proteins: Nmar_1547 and Nmar_1201. A very low-resolution structure (40 Å) of the S-layer was previously published from cryo-ET data of cells (Li *et al.*, 2018). As *N. maritimus* cells are so small, and cryo-EM technology has improved, the full width of the cell can now be seen in electron micrographs taken at a magnification that is theoretically sufficient for high-resolution structural studies. This development led to the renewed interest in the species as a structural target, despite the challenge of culturing a species which is growth-limited at very low density.

#### 5.1.2.2 *Saccharolobus solfataricus*

Initially classified as *Sulfolobus solfataricus* when discovered in the Solfatara volcano by Zillig *et al.* (1980), *S. solfaraticus* was reclassified by Sakai and Kurosawa (2018) as a member of the *Saccharolobus* genus. The cells are roughly spherical, can grow on sulphur and a variety of organic compounds, and are known to produce extracellular vesicles as part of their cell division cycle (Liu *et al.*, 2021). They grow optimally at 75 °C. *S. solfataricus* is considered to be a model organism for the phylum Crenarchaeota. The genome was sequenced by She *et al.* (2001) and the S-layer genes, SlaA and SlaB, were identified by Veith *et al.* (2009). A low-resolution (16 Å) structure of the S-layer was published by Gambelli *et al.* (2019) (see Figure 5.3), but no high-resolution structure has been published.

#### 5.1.2.3 *Pyrobaculum oguniense* and *P. arsenaticum*

*Pyrobaculum oguniense* are rod-shaped cells that grow optimally at 90 °C and neutral pH, and were first discovered in a Japanese hot spring (Sako *et al.*, 2001). The only published work on the S-layer is from Rensen *et al.* (2015), who observed hexagonal structures forming and breaking through the S-layer when the cells were exposed to radiation. The genome of *P. oguniense* was sequenced by Bernick *et al.* (2012), and the genes predicted to encode the S-layer proteins have the UniProt IDs H6Q8U3 and H6Q8S4.

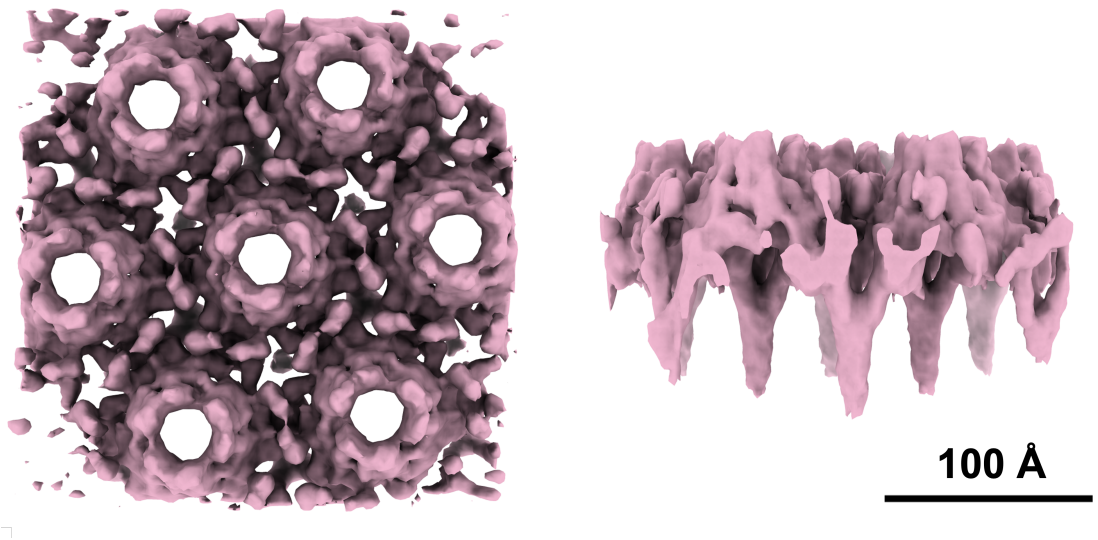



**Figure 5.3:** *S. solfataricus* S-layer structure at low resolution from Gambelli *et al.* (2019).

*P. arsenaticum* was first isolated from a hot spring in Naples, Italy, and grows best on inorganic electron acceptors such as sulphur and arsenate (Huber *et al.*, 2000). It was the first recorded arsenic-oxidising species, and this role of heavy metal-fixation is very important in its ecosystem. Like *P. oguniense*, *P. arsenaticum* are rod-shaped cells that grow at high temperatures (95 °C) and around neutral pH. The genome was sequenced by Lowe and Richardson (2007), and the two predicted S-layer genes have the UniProt IDs A0A7L4PCH8 and A0A7L4P6S3.

*P. oguniense* and *P. arsenaticum* are the two most closely related species in the *Pyrobaculum* genus. Their genomes are mostly structurally syntenic, and where genes are shared, they have an average 96% nucleotide identity. *P. oguniense* has a genome that is 15% larger than that of *P. arsenaticum*, though the origin of this difference is unknown.

Work by Phipps *et al.* (1991) showed the hexagonal symmetry of the S-layer of another member of the genus, *P. organotrophum*, suggesting this would also be the symmetry of S-layers of other related species. Their work also showed that this hexagonal symmetry remained even after S-layer sheets were obtained using harsh isolation methods. This demonstrates that the S-layer is highly stable and resilient to high temperatures, which is a requirement for the extreme environments

in which *Pyrobaculum* species live. However, it is not known how the S-layer protein achieves these qualities.

#### 5.1.2.4 *Haloferax volcanii*

*Haloferax volcanii* is an extremophile that can be isolated from hypersaline environments such as the Dead Sea. Mullakhanbhai and Larsen (1975) described *H. volcanii* and named the organism after B. E. Volcani, the researcher who first identified living organisms in the Dead Sea in the 1930s (Oren and Ventosa, 1999). They are typically described as having a "dished crisp" shape. After being proposed as a potential model organism for archaea by Soppa (2006), the genome was sequenced by Hartman *et al.* (2010). The S-layer gene had previously been isolated, and early structural studies were performed on the S-layer before the structure was solved (Sumper *et al.*, 1990; Rodrigues-Oliveira *et al.*, 2019). These studies revealed glycosylation sites on the S-layer, as well as the structural stability of the protein under a range of temperatures, pHs, and salt concentrations.

### 5.1.3 Previous work referenced in this chapter

Gambelli *et al.* (2024) published the structure of the S-layer from *Saccharolobus acidocaldarius*, a species in the same genus as *S. solfataricus*. This structure was solved from a single-particle dataset of a dimer of the SlaA protein, and tomography data of S-layer sheets. They were able to build a model of five domains of SlaA, and fit predicted models of the other domains and SlaB into the low-resolution sub-tomogram averaging structure. A monomer from the single-particle structure is shown in Figure 5.4, aligned with a model of the full SlaA protein predicted by AlphaFold3 (Abramson *et al.*, 2024).

The S-layer structure of *H. volcanii* was published by von Kügelgen *et al.* (2021), from single-particle cryo-EM and cryo-ET data. The hexameric and pentameric structures were both solved, and are shown in Figure 5.5. In the density map of the hexamer, many areas were identified as likely ion binding sites, however it was not possible to determine the identity of the occupying ions. This is

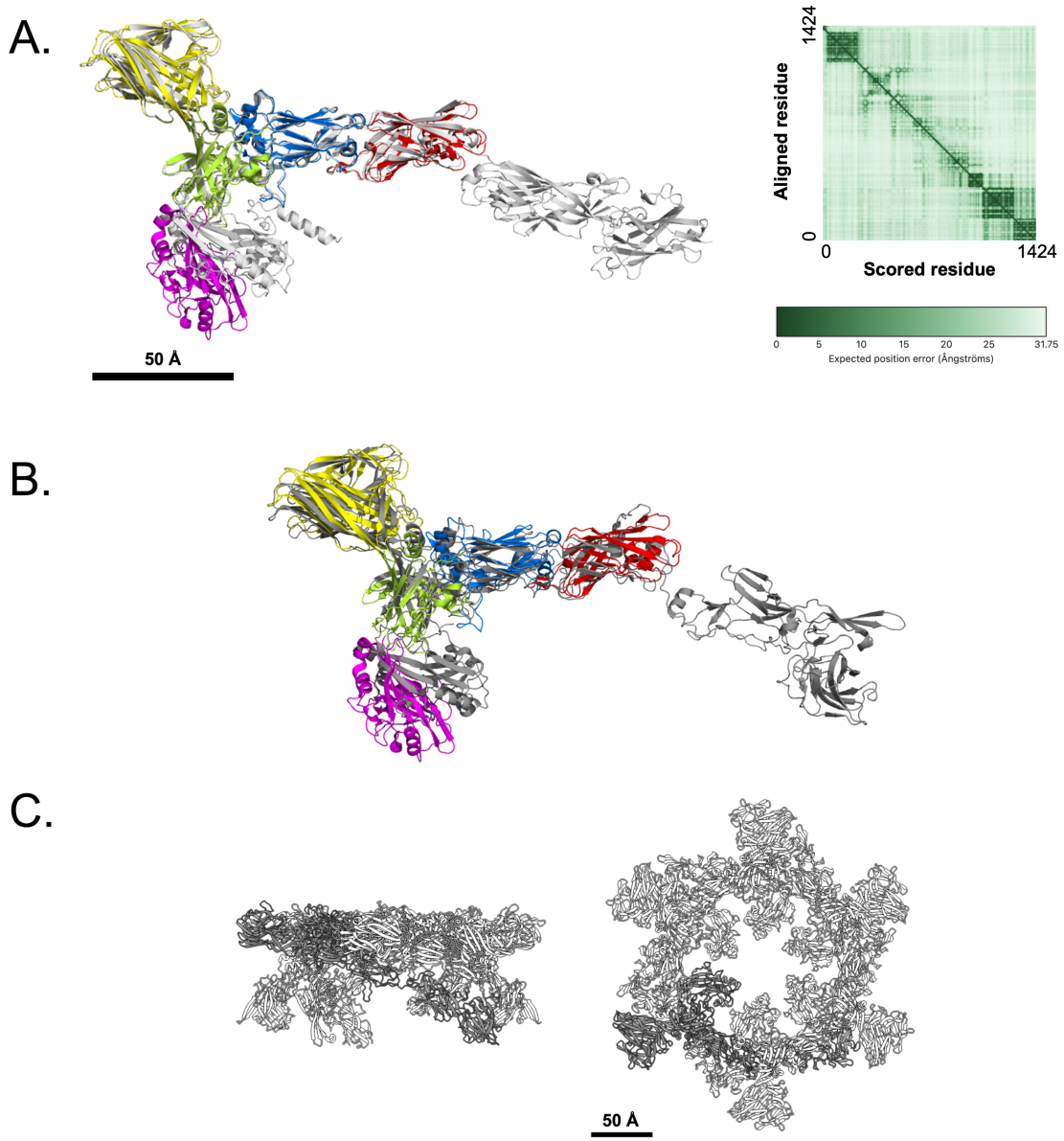



**Figure 5.4:** The monomer from the S-layer structure of *S. acidocaldarius* (Gambelli *et al.*, 2024), showing domain 1 (pink), domain 2 (green), domain 3 (yellow), domain 4 (blue), and domain 5 (red).
A. The monomer aligned with the AlphaFold3 structure of the monomer shown in white, which also includes an N-terminal leader sequence, and domains 6 and 7. Predicted Alignment Error (PAE) scores shown on the right.
B. The monomer structure aligned with one monomer of the hexameric model shown in grey, also published by Gambelli *et al.* (2024).
C. The hexameric model, with one monomer shown in grey.

both because of the 3.5 Å resolution of the structure, and the limited ability of cryo-EM to determine the identity of ions even at very high resolutions. Biochemical work had elucidated the importance of the divalent cations $Mg^{2+}$ and $Ca^{2+}$ to the stability of the structure, but it is not possible to discriminate between the two in this cryo-EM structure alone.

### 5.1.4 Aims of this chapter

This chapter uses cryo-ET and molecular dynamics to investigate the structures of archaeal S-layers. First, the structure of the S-layer of *N. maritimus* is solved *in situ* from tomograms of whole cells. This is combined with a single-particle approach from collaborators to achieve high resolution. The structure of the S-layer of *S. solfataricus* is solved from tomograms of vesicles, and compared to a published structure of the S-layer of *S. acidocaldarius*, giving insight into the conservation (or lack thereof) of sequence and structure within the *Saccharolobus* genus. The S-layers of two syntenic species from the *Pyrobaculum* genus, *P. oguniense* and *P. arsenaticum*, are compared at low-resolution. Finally, molecular dynamics simulations are used to try and elucidate the identity of divalent cations in binding sites observed in the previously published structure of the *H. volcanii* S-layer (von Kügelgen *et al.*, 2021), demonstrating the use of computational methods in an integrative approach to structural biology.

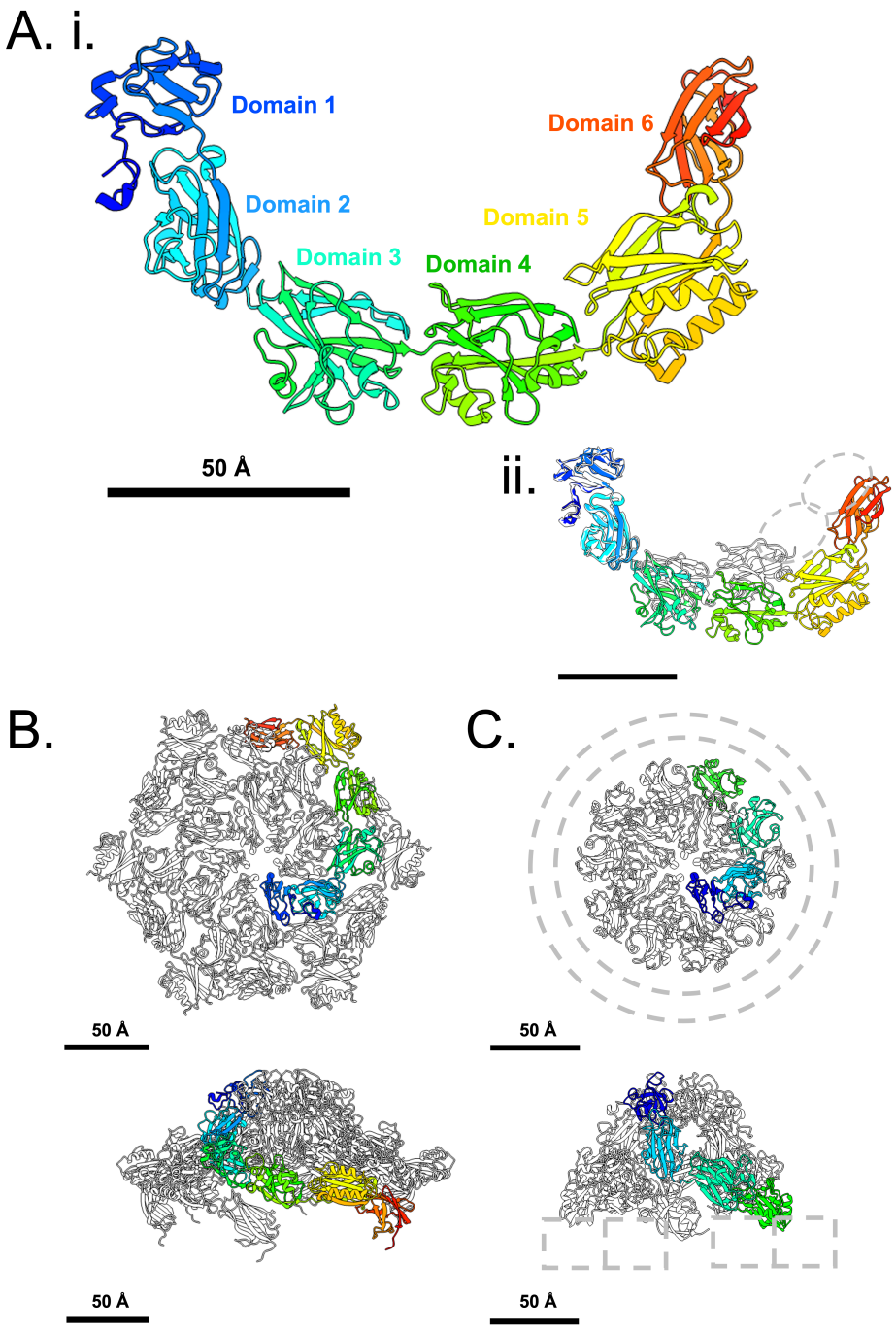



**Figure 5.5:** The S-layer structure of *H. volcanii*, from von Kügelgen *et al.* (2021).
A. The monomer of the S-layer, comprising six domains. i. shows the arrangement in the hexamer, and ii. overlays this with the angle of domain 2 to domain 3 and domain 3 to domain 4 in the pentamer. The pentameric structure does not include domain 5 and domain 6, and their approximate locations are indicated by the dotted grey lines.
B. The monomer within the hexameric structure, from above (top) and a side view (bottom). The N-terminus is at the centre of hexameric symmetry.
C. The monomer within the pentameric structure, from above (top) and a side view (bottom). Dotted grey lines indicate approximate locations of domain 5 and domain 6.

## 5.2 Materials and Methods

### 5.2.1 *N. maritimus*

#### 5.2.1.1 Cultivation of cells

Performed as described in Chapter 4, subsection 4.2.1, by Lennart Pagani, Sofie van Dorst, and Andriko von Kügelgen. A culture of *N. maritimus* (SCM1 strain) was kindly provided by Dr. Felix Elling and Prof. Ann Pearson (Harvard University).

#### 5.2.1.2 Grid preparation

Performed by Lennart Pagani, Sofie van Dorst, and Andriko von Kügelgen. For cryo-ET, *N. maritimus* cells were taken from culture at the mid-log phase, and harvested by centrifugation at 15,000 *g* at 4 °C for 30 min. The cells were then resuspended in 30 µL of SCM and a 1:10 ratio of 10 nm colloidal gold (CMC, Utrecht) was added. The sample was immediately applied to freshly glow-discharged grids (Quantifoil, R2/2 Cu/Rhu 200 mesh) using a Vitrobot Mark IV (ThermoFisher), adsorbed for 60 s, blotted for 4 s at a force of −15, and plunge-frozen in liquid ethane. The chamber was maintained at 100% humidity and 10 °C. As in Chapter 4, subsubsection 4.2.2.1, the grids were clipped in autorings and stored in liquid nitrogen.

#### 5.2.1.3 Data collection

Performed in collaboration with Andriko von Kügelgen, as described in von Kügelgen *et al.* (2021). Cryo-ET data was collected on an equivalent FEI Titan Krios microscope equipped with a Gatan K2 direct electron detector. In total, 160 tilt series were recorded using a dose-symmetric tilt scheme (Hagen *et al.*, 2017). Each tilt series was recorded between ±60° using increments of 3°. Each tilt image was exposed for 0.9 s and dose-fractioned into 10 frames, giving a total dose of 121 $e^-/Å^2$ applied across the entire series. Series were collected at a pixel size of 1.3 Å with defoci ranging between −2 µm and −5 µm.

#### 5.2.1.4 Data analysis

Performed in collaboration with Andriko von Kügelgen, and also described in von Kügelgen *et al.* (2021). Tomogram reconstructions were performed in IMOD, by alignment of the tilt series using gold fiducials (Kremer *et al.*, Feb; Mastronarde and Held, 2017). Defoci for each of the tilt images were estimated using CTFFIND4 to calculate estimated CTFs in combination with the microscope parameters (Rohou and Grigorieff, 2015), and the tomograms were reconstructed using 3D-CTF correction of the entire tomographic volume (Turoňová *et al.*, 2017). After duplicated particles were removed, 40,000 particle coordinates were imported with estimated CTFs into scripts written in MATLAB based on the AV3 and TOM toolbox scripts for extraction and alignment (Nickell *et al.*, 2005; Förster and Hegerl, 2007; Förster *et al.*, 2008). Mask creation for structure refinement, and calculations of structure resolution were performed using RELION3.0 (Zivanov *et al.*, 2018). This yielded a structure at 5.7 Å resolution.

Further work to refine the structure was performed by Andriko von Kügelgen. Roughly aligned particles were imported into RELION4.0 (Zivanov *et al.*, 2022). The sub-tomogram coordinates and Euler angles were then refined iteratively as described in Zivanov *et al.* (2022). Initially a six-fold (C6) symmetry was imposed on the map, resulting in a 3.4 Å structure. Upon visual inspection, broken symmetry was identified at the centre of the hexamer, and a relaxation of the symmetry to two-fold (C2) improved the structure to a final resolution of 3.3 Å.

Tomograms for visualisation were generated using the simultaneous iterative reconstruction technique (SIRT), implemented in Tomo3D (Pelt *et al.*, 2015; Agulleiro and Fernandez, 2015).

#### 5.2.1.5 Model fitting

The model built by von Kügelgen *et al.* (2021) using a single-particle structure was fitted into the sub-tomogram averaging map using ChimeraX (Goddard *et al.*, 2018).

### 5.2.2 *S. solfataricus*

#### 5.2.2.1 Sample and grid preparation

Samples of *S. solfataricus* cells and vesicles were kindly provided by Dr. Mart Krupovic and Dr. Virginija Cvirkaite-Krupovic (Institut Pasteur). Treatment of both samples was the same, and as follows. The samples were mixed with 10 nm colloidal gold (CMC, Utrecht) at a ratio of 1:10, and immediately applied to glow-discharged grids (Quantifoil, R3.5/1 Cu/Rh 200 mesh) using a Vitrobot Mark IV (ThermoFisher). The grids were then blotted for 3 s at a force of −15, and plunge-frozen in liquid ethane. The chamber was maintained at 100% humidity and 10 °C. Grids were then clipped in autorings and stored in liquid nitrogen until imaging.

#### 5.2.2.2 Data collection

Preliminary images were collected on a Talos Arctica fitted with a TFS Falcon 4i direct electron detector for both cells and vesicles. Data was collected at 45,000x magnification. After screening, it was decided that only the vesicles would be used for high-resolution cryo-ET. Tomograms of the vesicles were collected on an FEI Titan Krios microscope with a Gatan K2 direct electron detector. Six tilt series were recorded using a dose-symmetric tilt scheme (Hagen *et al.*, 2017). Data was collected between ±60° in increments of 3°, and each tilt-image was exposed for 1.5 s dose-fractioned across 3 frames. The total dose per series was 105 $e^{-}/Å^{2}$, at a pixel size of 2.262 Å. Defoci ranged between −2.5 µm and −5 µm.

#### 5.2.2.3 Data analysis

Tomogram reconstructions were performed in IMOD, by alignment of the tilt series using gold fiducials (Kremer *et al.*, Feb; Mastronarde and Held, 2017). Defoci for each of the tilt images were estimated using CTFFIND4 to calculate estimated CTFs in combination with the microscope parameters (Rohou and Grigorieff, 2015), and the tomograms were reconstructed using 3D-CTF correction of the entire tomographic volume (Turoňová *et al.*, 2017). Sub-tomograms were extracted and aligned using scripts written in MATLAB based on the AV3 and TOM toolbox

scripts (Nickell *et al.*, 2005; Förster and Hegerl, 2007; Förster *et al.*, 2008). Mask creation for structure refinement, and calculations of structure resolution were performed using RELION3.0 (Zivanov *et al.*, 2018).

Alignment of the *S. solfataricus* and *S. acidocaldarius* SlaA sequences was performed using BLAST (Altschul *et al.*, 1990, 1997; Johnson *et al.*, 2008) and COBALT (Papadopoulos and Agarwala, 2007). Predicted structures were generated using AlphaFold2 in ColabFold (Jumper *et al.*, 2021; Mirdita *et al.*, 2022) and AlphaFold3 (Abramson *et al.*, 2024). PAE plots generated using the PAE viewer webserver (Elfmann and Stülke, 2023). Models were aligned using PyMOL (PyMOL, 2015).

#### 5.2.2.4 Model fitting

The structure and model built by Gambelli *et al.* (2024) were fitted into the sub-tomogram averaging map using ChimeraX (Goddard *et al.*, 2018).

### 5.2.3 *P. oguniense* and *P. arsenaticum*

#### 5.2.3.1 Sample and grid preparation

Samples of *P. oguniense* and *P. arsenaticum* cells were kindly provided by Dr. Mart Krupovic and Dr. Virginija Cvirkaite-Krupovic (Institut Pasteur). Treatment of both samples was the same, and as follows. The samples were mixed with 10 nm colloidal gold (CMC, Utrecht) at a ratio of 1:10, and immediately applied to glow-discharged grids (Quantifoil, R3.5/1 Cu/Rh 200 mesh) using a Vitrobot Mark IV (ThermoFisher). The grids were then blotted for 3 s at a force of −15, and plunge-frozen in liquid ethane. The chamber was maintained at 100% humidity and 10 °C. Grids were then clipped in autorings and stored in liquid nitrogen until imaging.

#### 5.2.3.2 Data collection

Preliminary images were collected on a Talos Arctica fitted with a TFS Falcon 4i direct electron detector for both cells and vesicles. Data was collected at 45,000x magnification. Cryo-ET data was collected on an FEI Titan Krios microscope with a Gatan K2 direct electron detector. Four tilt series were recorded using a

dose-symmetric tilt scheme (Hagen *et al.*, 2017), between ±60° in increments of 1°. Each tilt-image was exposed for 3 s and dose-fractioned across 10 frames, giving a total dose of 140 $e^{-}/Å^{2}$ applied to the entire series. The pixel size was 5.527 Å, and defoci ranged between −6 µm and −8 µm.

#### 5.2.3.3 Data analysis

Tomogram reconstructions were performed in IMOD, by alignment of the tilt series using gold fiducials (Kremer *et al.*, Feb; Mastronarde and Held, 2017). Defoci for each of the tilt images were estimated using CTFFIND4 to calculate estimated CTFs in combination with the microscope parameters (Rohou and Grigorieff, 2015), and the tomograms were reconstructed using 3D-CTF correction of the entire tomographic volume (Turoňová *et al.*, 2017). Sub-tomograms were extracted and aligned using scripts written in MATLAB based on the AV3 and TOM toolbox scripts (Nickell *et al.*, 2005; Förster and Hegerl, 2007; Förster *et al.*, 2008). Mask creation for structure refinement, and calculations of structure resolution were performed using RELION3.0 (Zivanov *et al.*, 2018).

Alignment of the *P. oguniense* and *P. arsenaticum* S-layer sequences was performed using BLAST (Altschul *et al.*, 1990, 1997; Johnson *et al.*, 2008) and COBALT (Papadopoulos and Agarwala, 2007). Predicted structures were taken from the AlphaFold2 database (Jumper *et al.*, 2021; Varadi *et al.*, 2022). PAE plots generated using the PAE viewer webserver (Elfmann and Stülke, 2023). Models were aligned using PyMOL (PyMOL, 2015). Transmembrane helix prediction was performed with TMHMM (Sonnhammer *et al.*, 1998).

### 5.2.4 *H. volcanii*

#### 5.2.4.1 Model preparation

Performed in collaboration with Professor Phill Stansfeld (University of Warwick). The published model of the S-layer hexamer from von Kügelgen *et al.* (2021), PDB 7PTR, was placed in a hexameric box with periodic boundary conditions applied in all dimensions. This was achieved by applying a triclinic box with dimensions 180 x

180 x 130 Å, and angles 90°, 90°, and 60°, and then visualising using the GROMACS trjconv tool and the compact option under the -ur flag (Bekker *et al.*, 1993).

The molecules of the glycan $\beta$-D-glucopyranose were removed from the model, as the exact chemical identity of the glycans on the S-layer are unknown. The $Ca^{2+}$ ions were kept, removed, or replaced as required by the simulations. The published model did not include residues 760–827, and were not modelled into simulations. This is because the residues were not needed to complete the tiling of the S-layer units, and are likely involved in membrane-anchoring.

After preparation, the protein structure was energy minimised, and then equilibration was performed on the solvated structure independently for each simulation run.

#### 5.2.4.2 Simulation set-up

The S-layer was solvated with TIP3P water and cations as shown in Table 5.1 for the 6 set-ups (Berendsen *et al.*, 1981; Price and Brooks, 2004). Chloride ions were added to ensure the overall system was neutrally charged.

**Table 5.1:** *H. volcanii* S-layer simulation set-ups.

| **Setup** | **No. of $Ca^{2+}$ ions** | **No. of $Mg^{2+}$ ions** | **No. of $Na^{+}$ ions** |
|---|---|---|---|
| 1 | 0 | 0 | 1045 |
| 2 | 36 | 0 | 1045 |
| 3 | 36 | 808 | 0 |
| 4 | 0 | 808 | 0 |
| 5 | 808 | 36 | 0 |
| 6 | 808 | 0 | 0 |

Simulations were performed using the GROMACS 2020.3 engine (Bekker *et al.*, 1993; Berendsen *et al.*, 1995; Lindahl *et al.*, 2001; Van Der Spoel *et al.*, 2005; Hess *et al.*, 2008; Pronk *et al.*, 2013; Abraham *et al.*, 2015; Szilárd *et al.*, 2015), using the CHARMM36m force field (MacKerell Jr, 2006; Best *et al.*, 2012; Huang *et al.*, 2017). Each set-up was equilibrated 3 times from the same starting coordinates, first with position restraints of 1000 kJ $mol^{-1}$ $nm^{-2}$ applied to the protein, and

then removing those restraints. These equilibrated systems were confirmed to have stabilised temperature, pressure, and overall energy.

Production simulations were performed for 100 ns, with an integration step size of 2 fs. Temperature was maintained at 310 K by the V-rescale thermostat (Bussi *et al.*, 2007), and isotropic pressure maintained at 1 bar by the Parrinello-Rahman barostat (Parrinello and Rahman, 1981). Non-bonded interactions were cut-off at 1.2 nm. Electrostatics were treated using the particle mesh Ewald method (Darden *et al.*, 1993). Hydrogen bonds and constraints were managed by the LINCS algorithm (Hess *et al.*, 1998).

#### 5.2.4.3 Analysis

Analysis was performed using the MDAnalysis module for Python (Theobald, 2005; Liu *et al.*, 2010; Michaud-Agrawal *et al.*, 2011; Gowers *et al.*, 2016). Other Python packages used were NumPy, Pandas, MatPlotLib, and SciPy (Hunter, 2007; McKinney, 2011; van der Walt *et al.*, 2011; Virtanen *et al.*, 2020). VMD was used for calculating density maps of divalent cations using the VolMap tool, and for visualisation (Humphrey *et al.*, 1996). ChimeraX was used to visualise the density maps, after they had been symmetrised in RELION3.0 (Pettersen *et al.*, 2021; Zivanov *et al.*, 2018). Electrostatic surface calculations were performed with the ABPS package for PyMOL (PyMOL, 2015; Jurrus *et al.*, 2018). Python scripts were developed by the author specifically for this project. Custom contact analysis scripts used a cutoff of 5 Å.

### 5.2.5 Other analysis

A phylogenetic tree was generated using PhyloT and the Interactive Tree of Life (Letunic and Bork, 2007, 2011; Letunic, 2015). Data for the tree was taken from the NCBI taxonomy database (Sayers *et al.*, 2019; Schoch *et al.*, 2020).

## 5.3 Results

### 5.3.1 Structural studies of the S-layer of *N. maritimus*

The work in this subsection contributed to sub-tomogram averaging section of the paper 'Membraneless channels sieve cations in ammonia-oxidizing marine archaea', von Kügelgen *et al.* (2024).

#### 5.3.1.1 Tomograms

160 tomograms were taken of *N. maritimus* cells at a pixel size of 1.33 Å, with slices of one tomogram shown in Figure 5.6. This pixel size gives a theoretical highest resolution of 2.66 Å for sub-tomogram averaging structures, after accounting for the Nyquist criterion (Nyquist, 1928).

The cells were at a low density on the grids, even after concentrating the *N. maritimus* culture, meaning that only one cell was present per tomogram (shown in Appendix Section C.1).

#### 5.3.1.2 Sub-tomogram averaging structure

Sub-tomogram averaging was performed on the tomograms, following the procedure described in subsubsection 5.2.1.4, to produce a 5.5 Å structure of the S-layer. This resolution was not sufficient for atomic model building, but the architecture of the S-layer can be seen in the structure as shown in Figure 5.7. It is clear that the structure is heavily glycosylated, and the hexagonal symmetry is easily visible. Each S-layer unit has a cavity on the underside.

#### 5.3.1.3 Single-particle structure and published model

A single-particle structure (2.7 Å) and higher-resolution sub-tomogram structure (3.3–4.5 Å) were solved by Andriko von Kügelgen, which were published in von Kügelgen *et al.* (2024) and shown in Figure 5.8. Models were built into both structures, and aligning the models gives an RMSD of 2.21 Å, which indicates the structures are highly similar. Differences can be accounted for in part by the differences in curvature between the sheets used for SPA, and the cell surface for

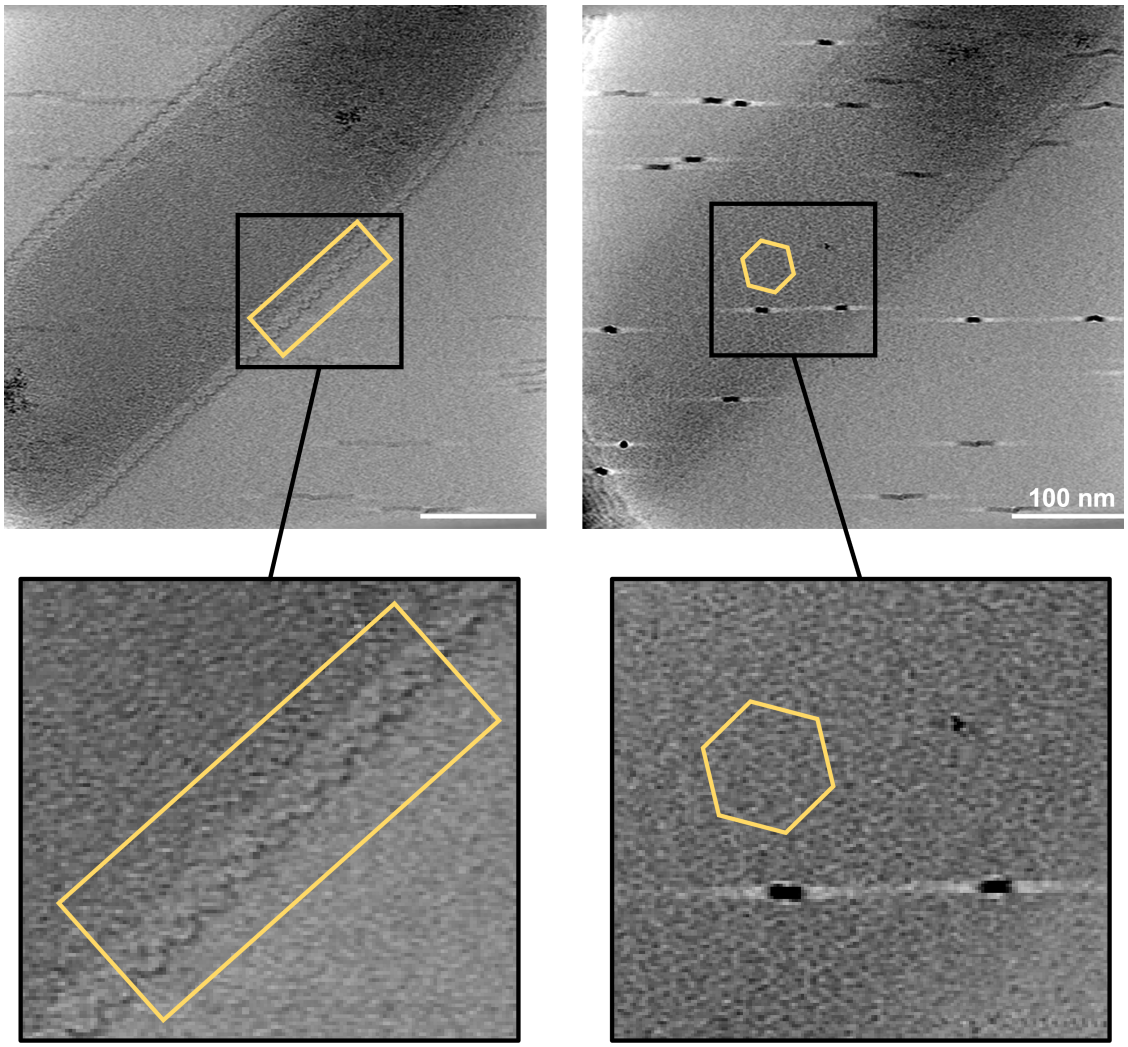
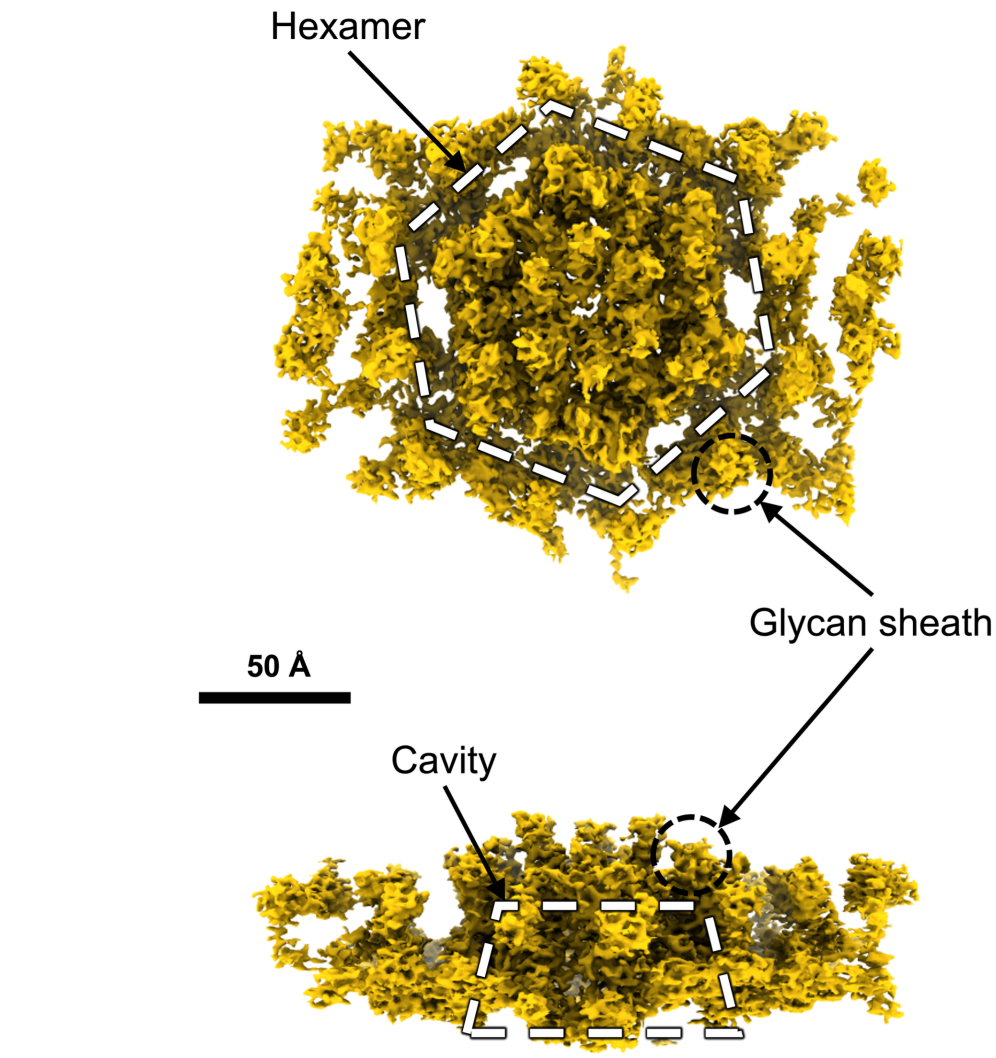



**Figure 5.6:** Tomogram slices of *N. maritimus* cells. In the slice on the left, the ridges outside of the membrane can be clearly seen, which is the S-layer. In the slice on the right, the tessellating hexagonal pattern of the S-layer can be seen.

STA. In the SPA structure, 9 of the 10 predicted domains were resolved, with only the C-terminal domain missing. The STA structure was unable to resolve the last 2 of the predicted domains. Interestingly, the structure revealed that the centre of each hexamer has p2 symmetry, with the pore being elongated rather than roughly circular. This can be seen in the structure shown in Figure 5.7, however the resolution was insufficient to recognise it as a feature.



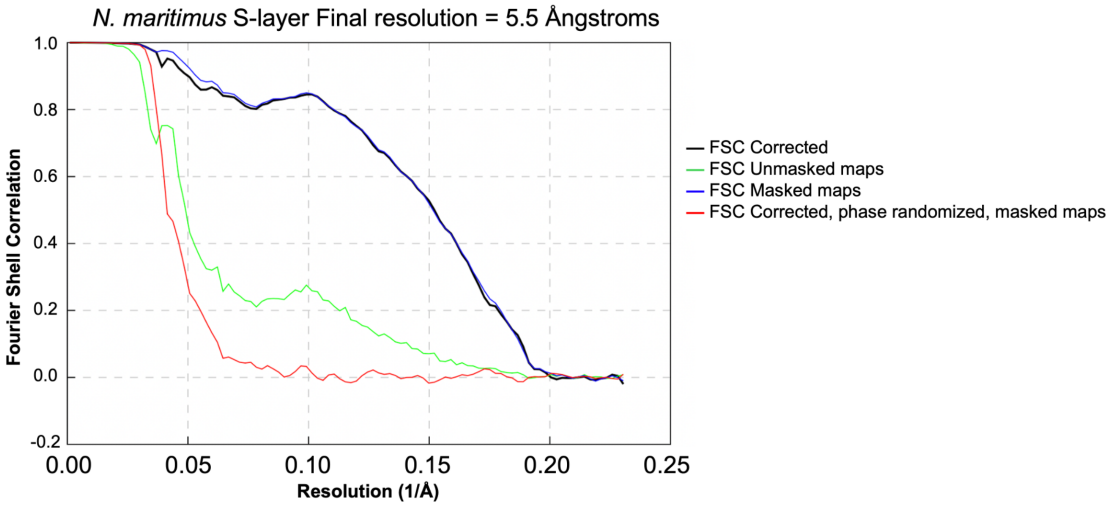




**Figure 5.7:** 5.5 Å structure of *N. maritimus* S-layer, shown from the outside (top) and the side (bottom). Notable features are indicated. FSC curve is shown below the structure.

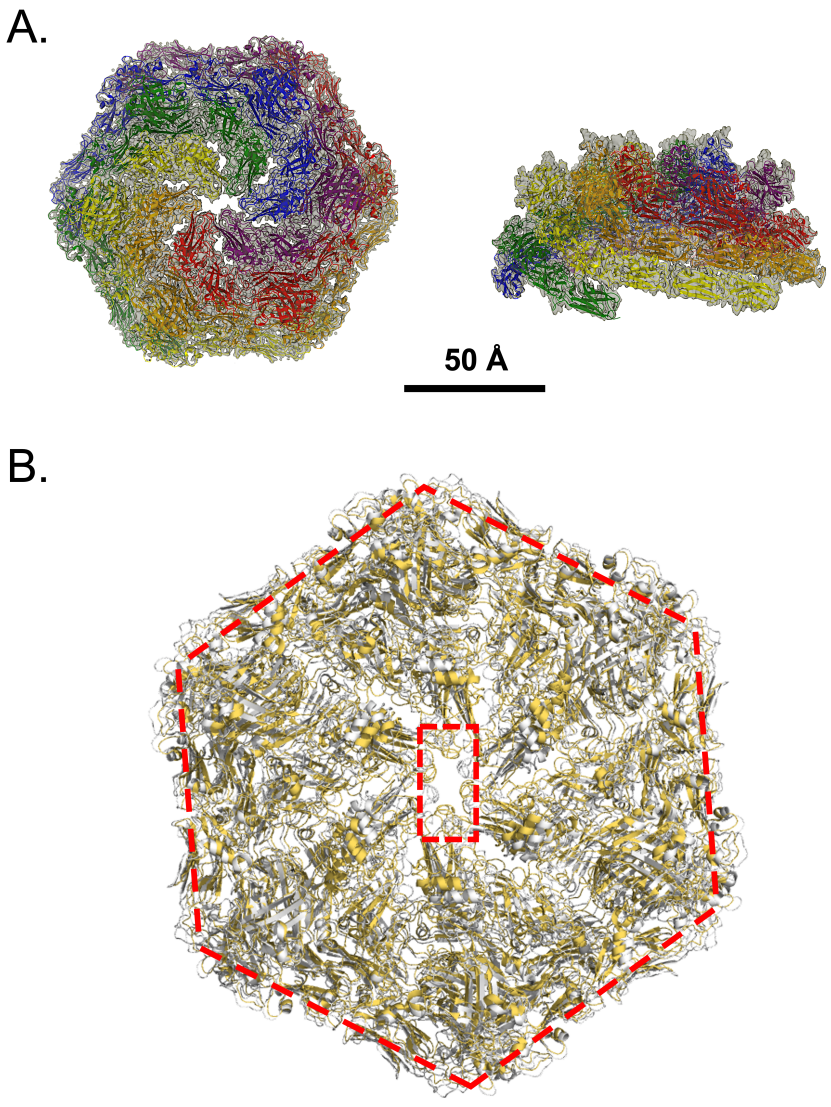



**Figure 5.8:** Published structure of *N. maritimus* S-layer from von Kügelgen *et al.* (2024).
A. Single-particle structure (grey) with model fitted. Each chain is coloured, showing how the domains curve in a "cashew-like" way around the cavity.
B. Alignment of the single-particle structure (grey) with the tomogram structure (yellow), both published in von Kügelgen *et al.* (2024). The hexagon shows the p6 symmetry, and the rectangle indicates the p2 symmetry of the central pore.

#### 5.3.1.4 Fitting the model into the sub-tomogram averaging structure

To assess the accuracy of the 5.5 Å structure presented in this research, the SPA model from von Kügelgen *et al.* (2024) was placed into the structure. The model fitted well, as shown in Figure 5.9, with the density that is not modelled representing the glycans and the last Ig-like domain of the S-layer protein.

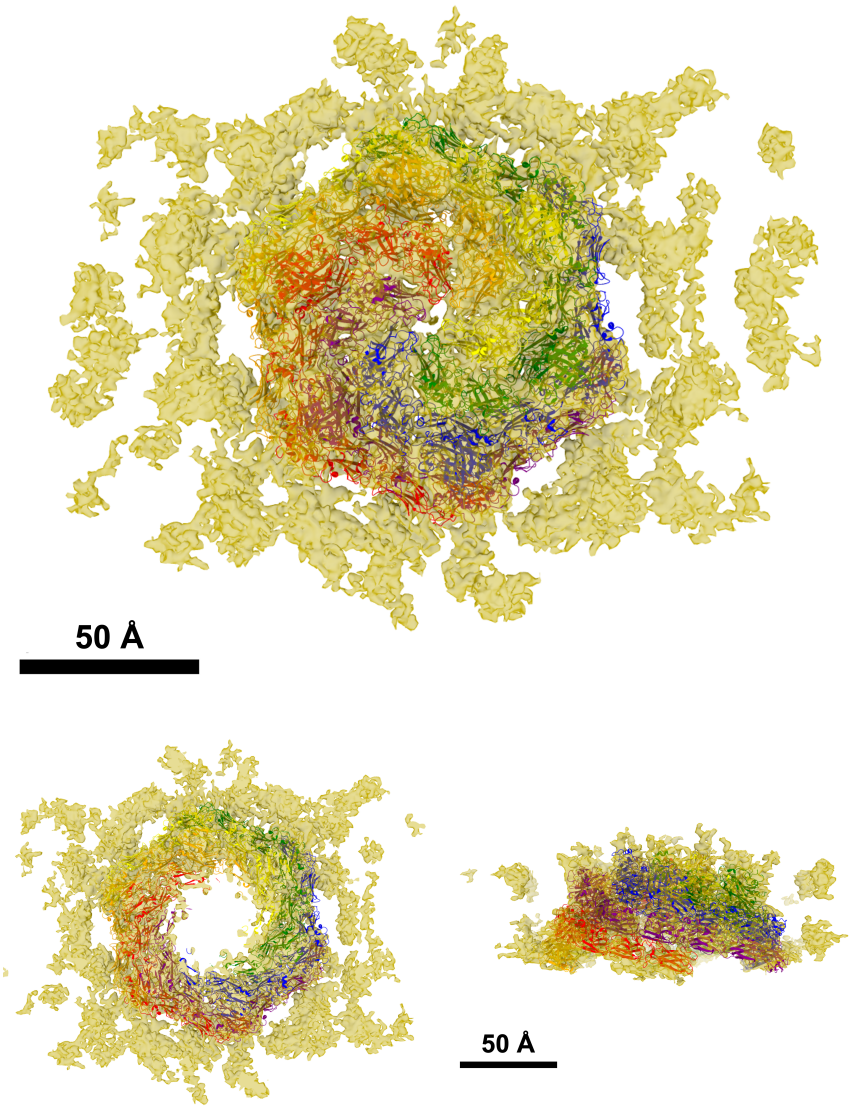



**Figure 5.9:** Model (rainbow, cartoon representation) fitted to STA structure (yellow) of *N. maritimus* S-layer. Porous nature of the S-layer can be observed as the gaps between the hexamer and the density representing the next hexamers. The entirety of a single copy of the hexameric model is represented in the structure. The hexamer interfaces between symmetry-related copies can also be seen. Bottom left shows a slice through the structure and model.

### 5.3.2 Structural studies of the S-layer of *S. solfataricus*

#### 5.3.2.1 Tomograms of vesicles

Micrographs of *S. solfataricus* cells were taken to assess if they were a suitable target for cryo-ET studies of the S-layer (Figure 5.10). However, the cells were too thick to observe top views of the S-layer, and thus a different approach was required.

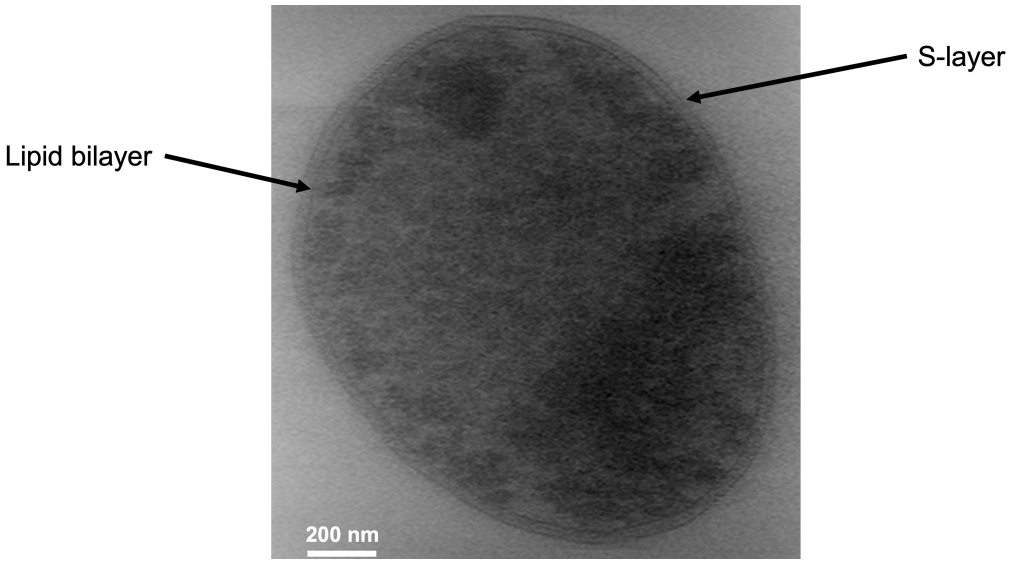



**Figure 5.10:** Micrograph of an *S. solfataricus* cell. The S-layer is discernible outside the lipid bilayer, but the cell is thick and appears dark, meaning the top views of the S-layer are not visible.

*S. solfataricus* cells secrete vesicles, which are coated in S-layer and can be easily isolated from culture by centrifugation (Liu *et al.*, 2021). These vesicles are much smaller than the cells, and tomographs of the vesicles clearly showed all views of the S-layer (Figure 5.11). By starting with a spherical model of coordinates, sub-tomograms were extracted, aligned, and iteratively refined to improve the position and angle of the sub-tomograms, each centred on a copy of the S-layer hexamer. This allowed the extraction of sub-tomograms covering the majority of the views of the S-layer.

Pentameric defects were identifiable in the vesicles, however there were too few to consider solving an STA pentamer structure.

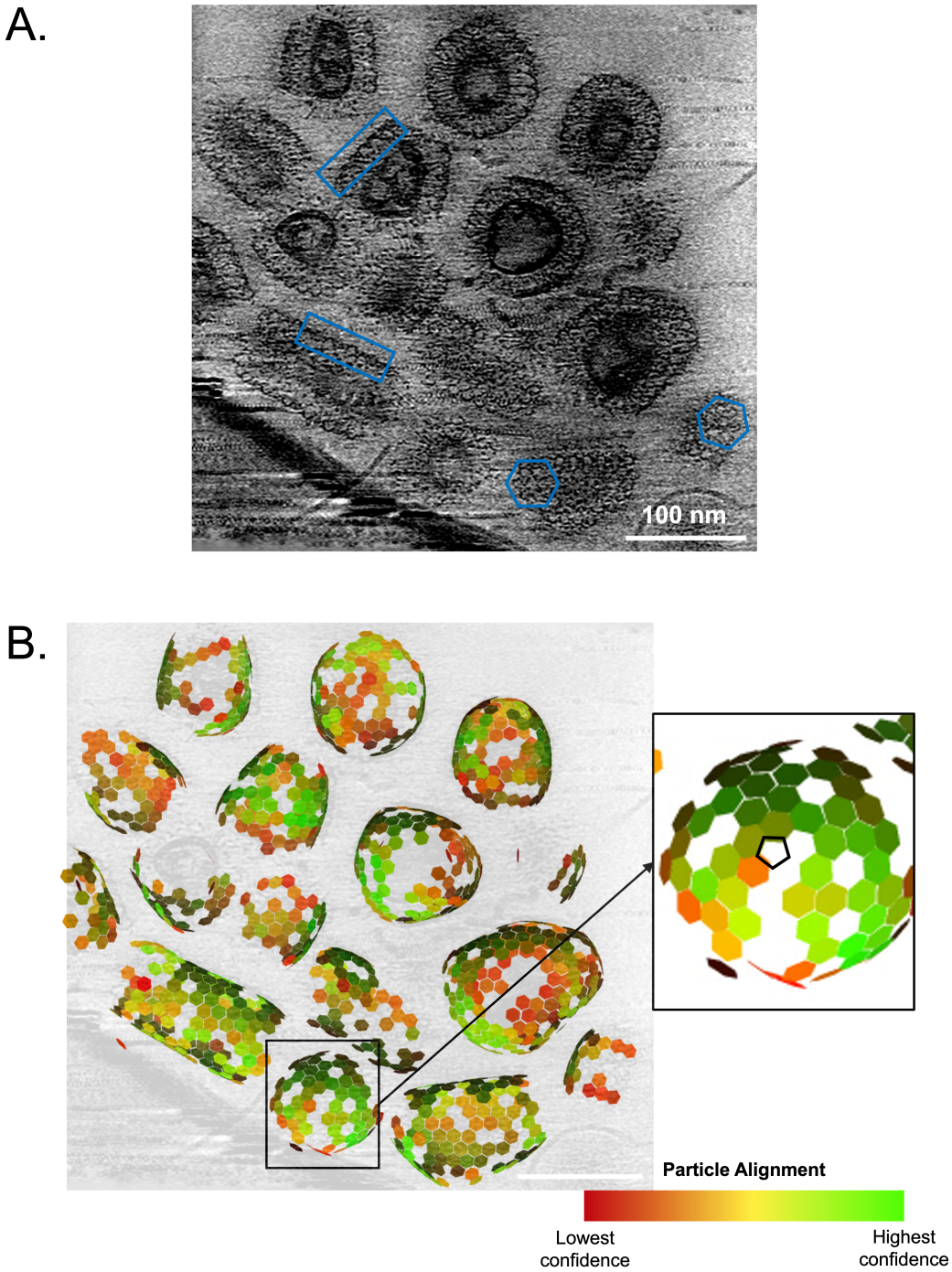



**Figure 5.11:** A. A slice of a tomogram of vesicles from *S. solfataricus*. The S-layer can clearly be seen on the outside of the vesicles, and the hexagonal lattice can be observed. B. S-layer particles picked from the tomogram of vesicles above by iterative refinement of position and alignment of extracted sub-tomograms. Particles coloured by confidence metric, which is calculated by the AV3 toolbox as a comparison between the sub-tomogram represented by the particle and the average structure using cross-correlation. A vesicle with an identifiable pentameric defect is highlighted.

#### 5.3.2.2 Sub-tomogram averaging structure

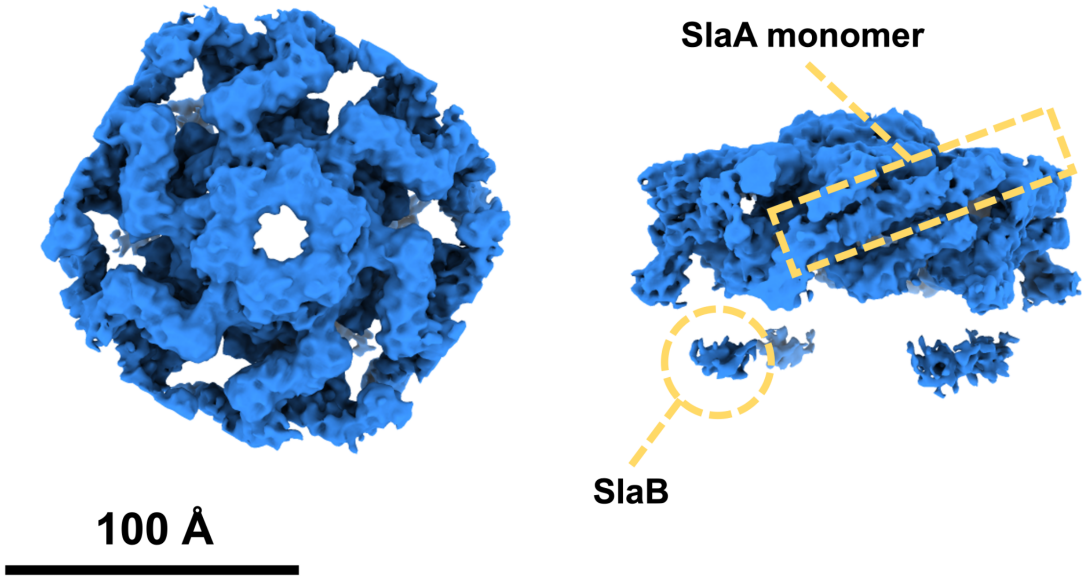
Sub-tomogram averaging of the S-layer on vesicles yielded a structure with a resolution 9.9 Å, without symmetry applied, which is shown in Figure 5.12. The S-layer of *Saccharolobus* species is formed of two proteins, SlaA and SlaB. SlaA forms the lattice and connects to the more flexible SlaB, which then connects to the cell membrane. Only a small part of the density at higher resolution represents SlaB, with SlaA being much better resolved. The structure shows large pores in the S-layer, and centres of both three-fold and six-fold symmetry in the p6 lattice. The structure, while only at medium-resolution, still allows for the identification of major features of the S-layer, which suggests that STA is a valuable part of the structure-solving process.

This structure was compared to the structure published by Gambelli *et al.* (2019), as shown in Figure 5.13. There is good agreement in the structures, with better $z$-resolution in particular in the structure solved in the research presented here compared to the published structure.

$z$-slices through the S-layer structure allowed the creation of a schematic of SlaA (Figure 5.14). This enables the internal structure of SlaA to be followed, such as the way the "arms" of each monomer extend down from the central pore. Edges of three-fold and six-fold rotational symmetry are shown.

#### 5.3.2.3 Structure and model of *S. acidocaldarius* S-layer

Gambelli *et al.* (2024) published the structure of the *S. acidocaldarius* SlaA, with the single-particle structure coming from dimers of SlaA. This was then fitted to a hexamer they produced from sub-tomogram averaging of S-layer-coated vesicles. These structures are shown in Figure 5.15. The S-layer appears to have the same gross overall architecture as the *S. solfataricus* S-layer. Previous comparisons by Gambelli *et al.* (2019) between *S. solfataricus* and *S. acidocaldarius* suggested that the *S. acidocaldarius* S-layer subunit was around 10% larger than that of *S. solfataricus*, with a different architecture around the pore.



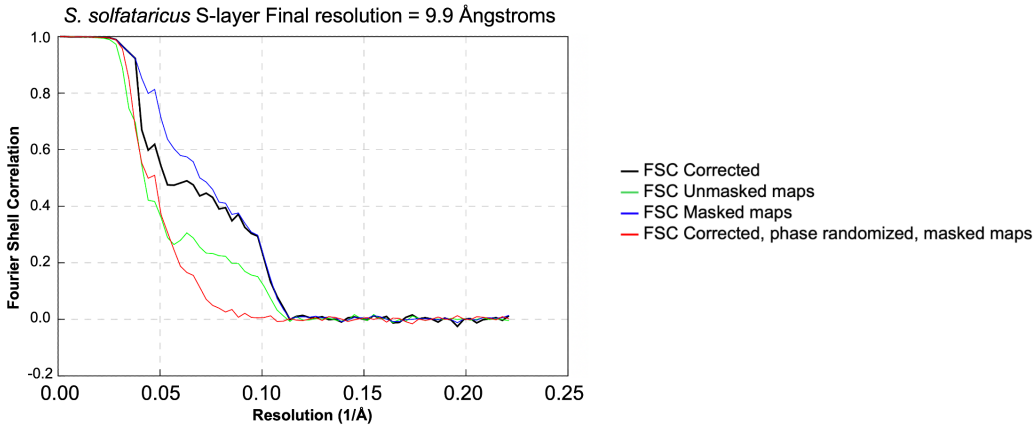




**Figure 5.12:** Structure of the *S. solfataricus* S-layer, shown from above (left) and the side (right). Each SlaA monomer extends out from the central pore in an "arm" with a sharp angle, and domains can be seen extending down towards the densities presumed to be SlaB. FSC curve shown below the structure.

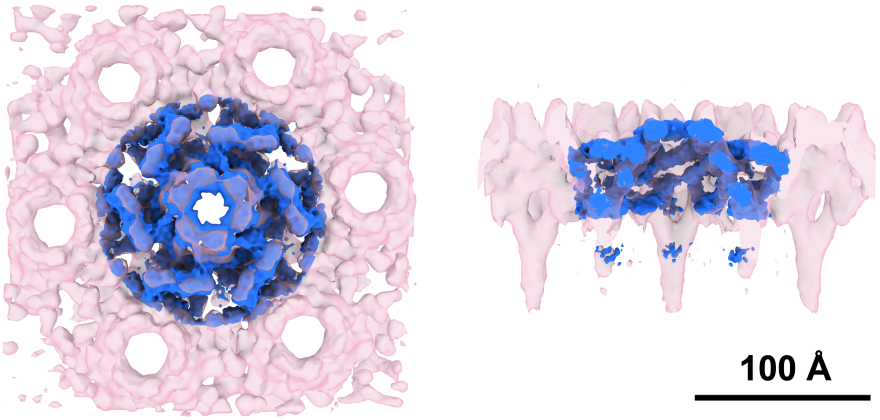



**Figure 5.13:** Comparison of structures of the *S. solfataricus* S-layer. Structure from Figure 5.12 in blue, structure from Gambelli *et al.* (2019) in pink.

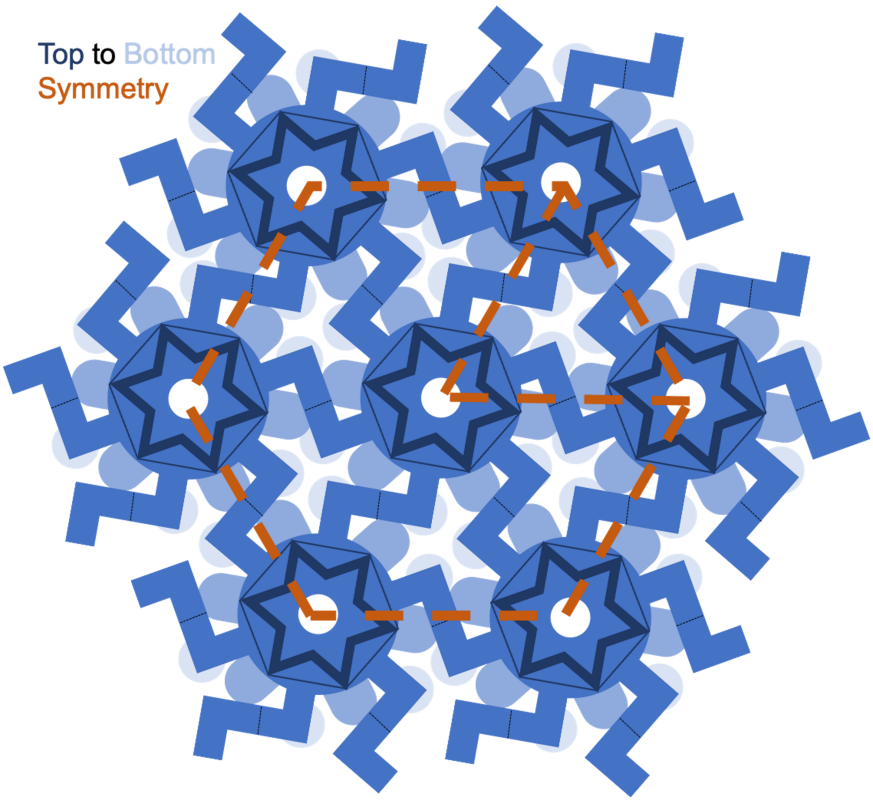



**Figure 5.14:** Schematic of the *S. solfataricus* S-layer, showing the symmetric units and how the hexamers interact with the lattice.

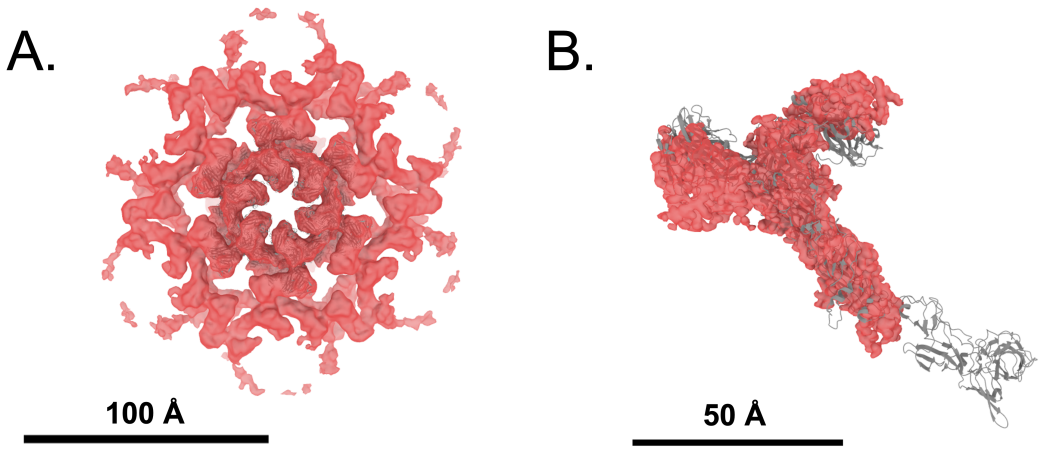



**Figure 5.15:** The *S. acidocaldarius* S-layer structure from Gambelli *et al.* (2024).
A. *S. acidocaldarius* sub-tomogram structure (red) with the hexamer model (grey, cartoon) built in.
B. *S. acidocaldarius* single-particle structure (red) fitted to a monomer from the hexamer model (grey, cartoon).

#### 5.3.2.4 Alignment of *S. solfataricus* and *S. acidocaldarius* SlaA sequences

BLAST was used to align the SlaA sequences from *S. solfataricus* and *S. acidocaldarius* (Altschul *et al.*, 1990, 1997; Johnson *et al.*, 2008). The Encyclopedia of Domains was used to identify domains in both structures (Lau *et al.*, 2024), based on the structures generated in the AlphaFold2 database (Varadi *et al.*, 2022). By aligning the structures and the sequences, it was found that the homology was strongest in domains 5, 6, and 7, which are furthest from the central S-layer pore (Figure 5.16). There was some homology in the sequences comprising parts of domains 2 and 4, however this had a much lower alignment score. The alignments can be found in Appendix Section C.2, Figure C.2 and Figure C.3.

#### 5.3.2.5 Comparing *Saccharolobus* SlaA structures

The STA structures of SlaA from both *Saccharolobus* species were aligned using ChimeraX (Goddard *et al.*, 2018; Pettersen *et al.*, 2021). This was then used to try and fit the model of the *S. acidocaldarius* SlaA into the structure of the *S. solfaraticus* S-layer. The results of these attempts are shown in Figure 5.17, and show that the

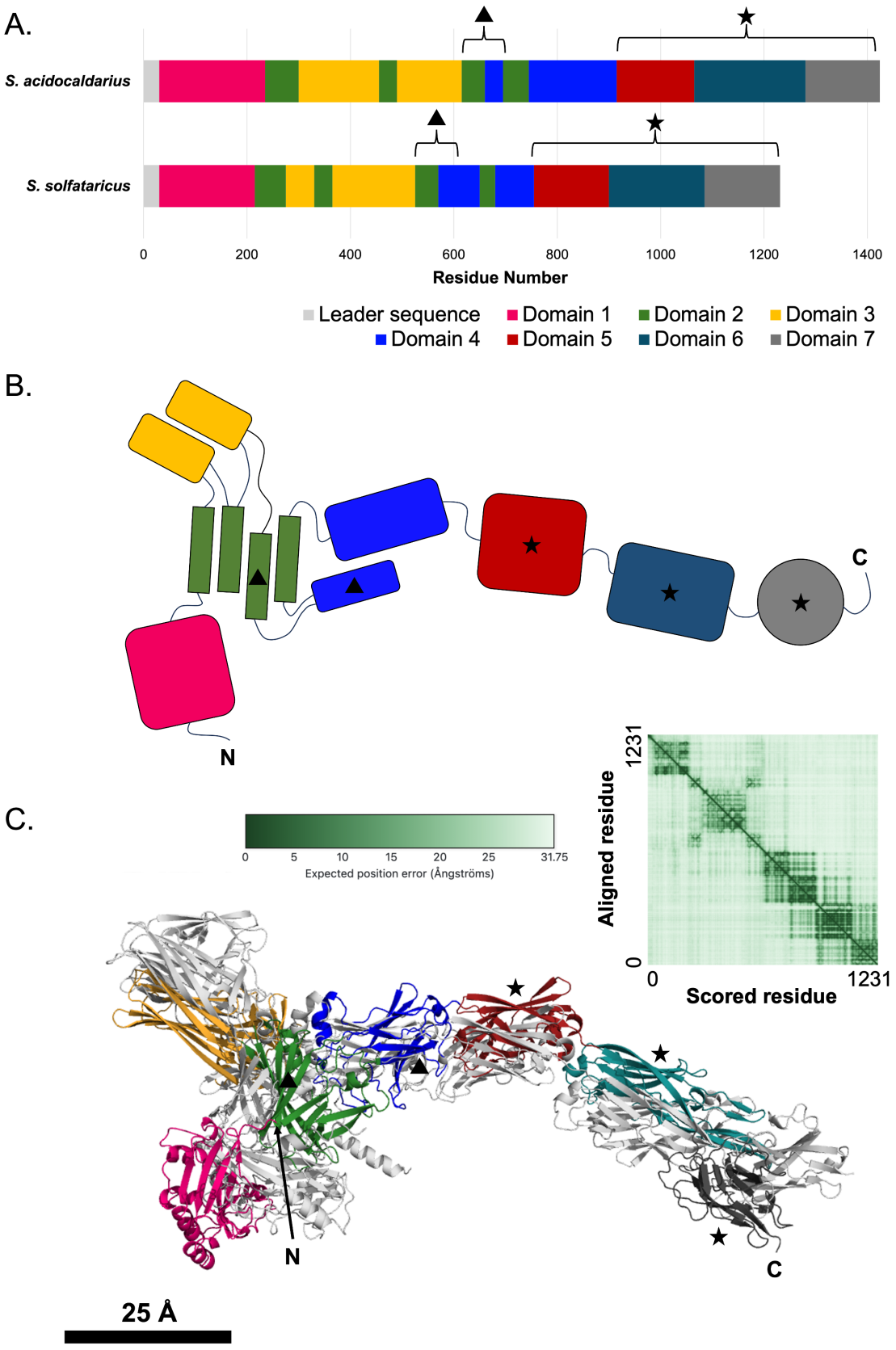



**Figure 5.16:** (Caption on next page.)

**Figure 5.16:** Comparison of the *S. solfataricus* and *S. acidocaldarius* SlaA sequences and models.
A. Comparison of the sequences of the SlaA proteins from the two species, indicating domain architecture. Triangle and star symbols indicate the regions of homology identified by BLAST as shown in Appendix Section C.2, Figure C.2 B. and C. respectively. Triangle represents sequence with low homology ($E = 1 \times 10^{-5}$), star represents sequence with high homology ($E = 4 \times 10^{-36}$).
B. Domain architecture of the *Saccharolobus* S-layer protein SlaA, including homology labelling as indicated in A.
C. The SlaA structure as predicted by AlphaFold2 for *S. solfataricus* (in colour, by domain as in A. and B.), aligned with the AlphaFold2 prediction of the *S. acidocaldarius* SlaA (white). PAE of the AF model shown to the top right of the structure.

alignment is poor. These results, combined with those shown in Figure 5.16, suggest that the differences in the sequence and structure for the first 4 domains of SlaA lead to the very different arrangement of domains around the central pore (Figure 5.17). This is also supported by the previous work of Gambelli *et al.* (2019). These findings are in agreement with previous work that discussed the low conservation of sequence even between highly-related species, which has consequences for the structure even if the gross architecture is similar (Bharat *et al.*, 2021).

### 5.3.3 Structural studies of the S-layers of *P. oguniense* and *P. arsenaticum*

#### 5.3.3.1 Tomograms of *P. oguniense*

Four tomograms of *P. oguniense* cells were taken at a pixel size of 5.527 Å (Figure 5.18). While this was not sufficient for sub-nanometre resolution, a medium-resolution structure was achievable for the S-layer. S-layer particles were visible in the tomograms of cells, both from the top and from the side, and picking of particles started with a cylinder of points placed in a hexagonal lattice. The cells are rod-shaped, however the vitrification process compressed the shape of the cells in the $z$-dimension, meaning that the top and bottom of the cells were furthest in distance from an ideal cylinder. This had the effect of the top and bottom having the worst alignment of particles picked during sub-tomogram averaging. To counter this, a MATLAB script was written to extend the lattice of aligned particles and

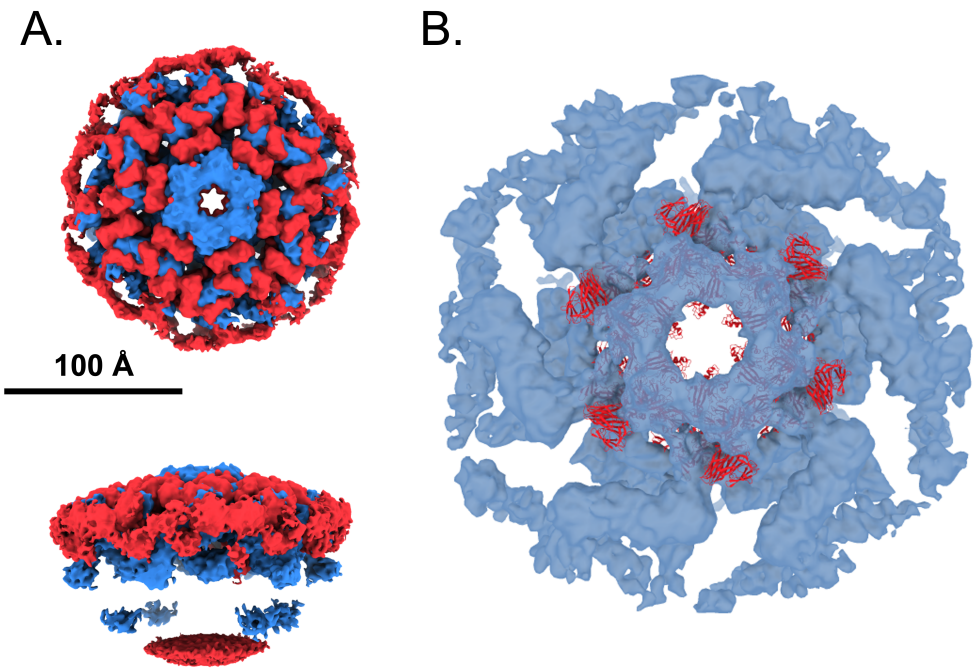



**Figure 5.17:** Comparison of the *S. solfataricus* and *S. acidocaldarius* structures, with alignment performed with ChimeraX (Goddard *et al.*, 2018; Pettersen *et al.*, 2021).
A. *S. acidocaldarius* sub-tomogram structure (red) from Gambelli *et al.* (2024) aligned to the *S. solfataricus* sub-tomogram structure (blue). The arms of the monomers do not align between the species.
B. *S. acidocaldarius* sub-tomogram averaging model (red, cartoon) from Gambelli *et al.* (2024) fitted to the *S. solfataricus* sub-tomogram structure (blue). Much of the model does not fit into the density, suggesting the structures are distinct.

improve the coverage of the cell with picked particles. For this sample, the number of sub-tomograms was increased by roughly 30% using this method.

#### 5.3.3.2 Sub-tomogram averaging structure of *P. oguniense* S-layer

The sub-tomogram averaging structure of the *P. oguiense* S-layer was resolved to 22 Å, and is shown in Figure 5.19. The S-layer had p6 symmetry, a central stalk potentially connecting to the membrane, and a cap that sat above the central stalk. The resolution was insufficient to determine any fold or secondary structure elements, or distinguish between monomers in the hexamer. The incidental finding of a second layer above the membrane, as shown in Figure 5.19, suggests a new avenue for biological study, outside the remit of the research presented in this thesis.

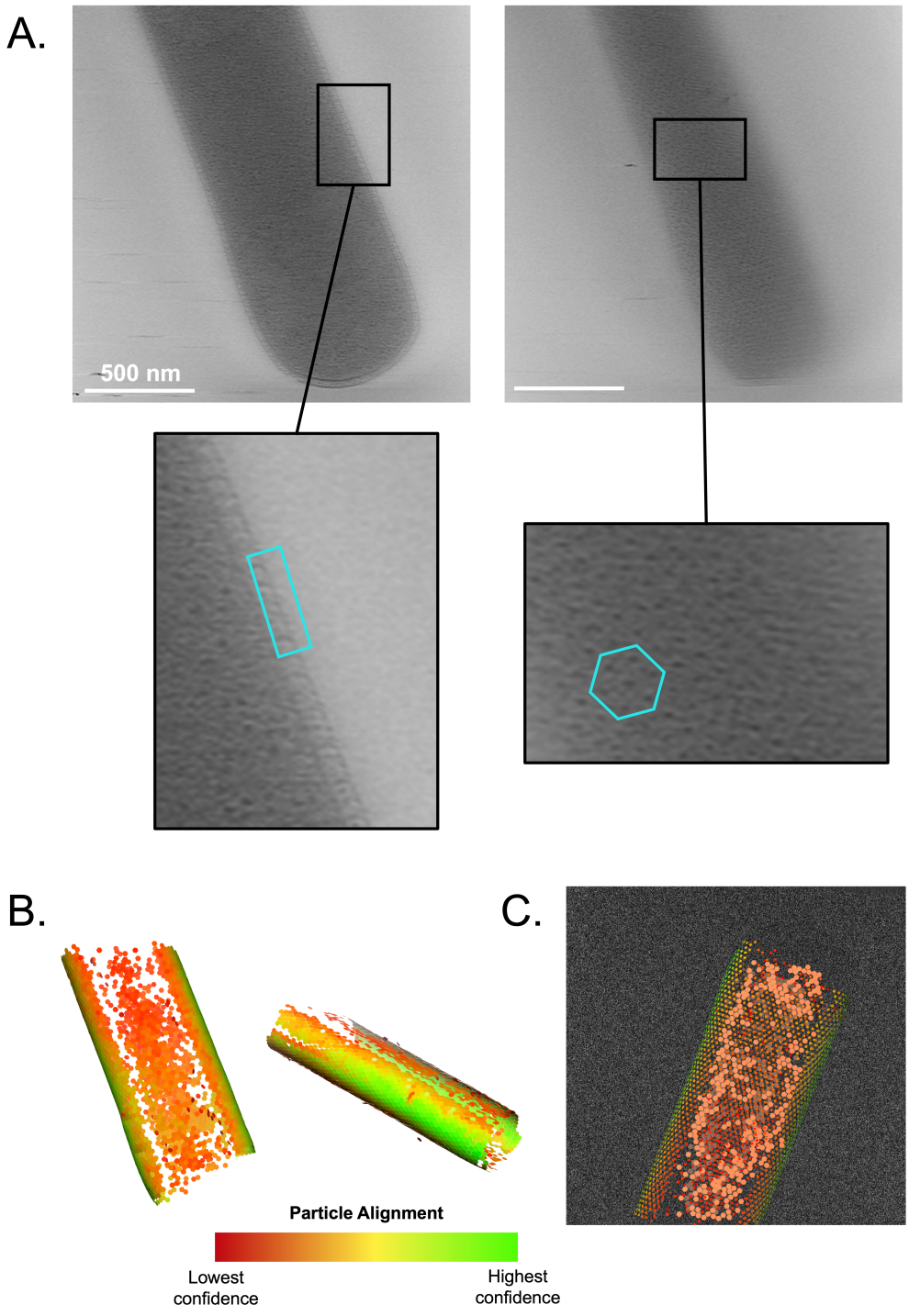



**Figure 5.18:** A. Two slices of a tomogram of a *P. oguniense* cell. The S-layer is highlighted, both a side view (rectangle) and a top view (hexagon) of the lattice.
B. Particle picking of the S-layer. Particles coloured by confidence metric, which is calculated by the AV3 toolbox as a comparison between the sub-tomogram represented by the particle and the average structure using cross-correlation.
C. The increase in lattice coverage of the cell after the use of a custom script to extend from the built lattice.

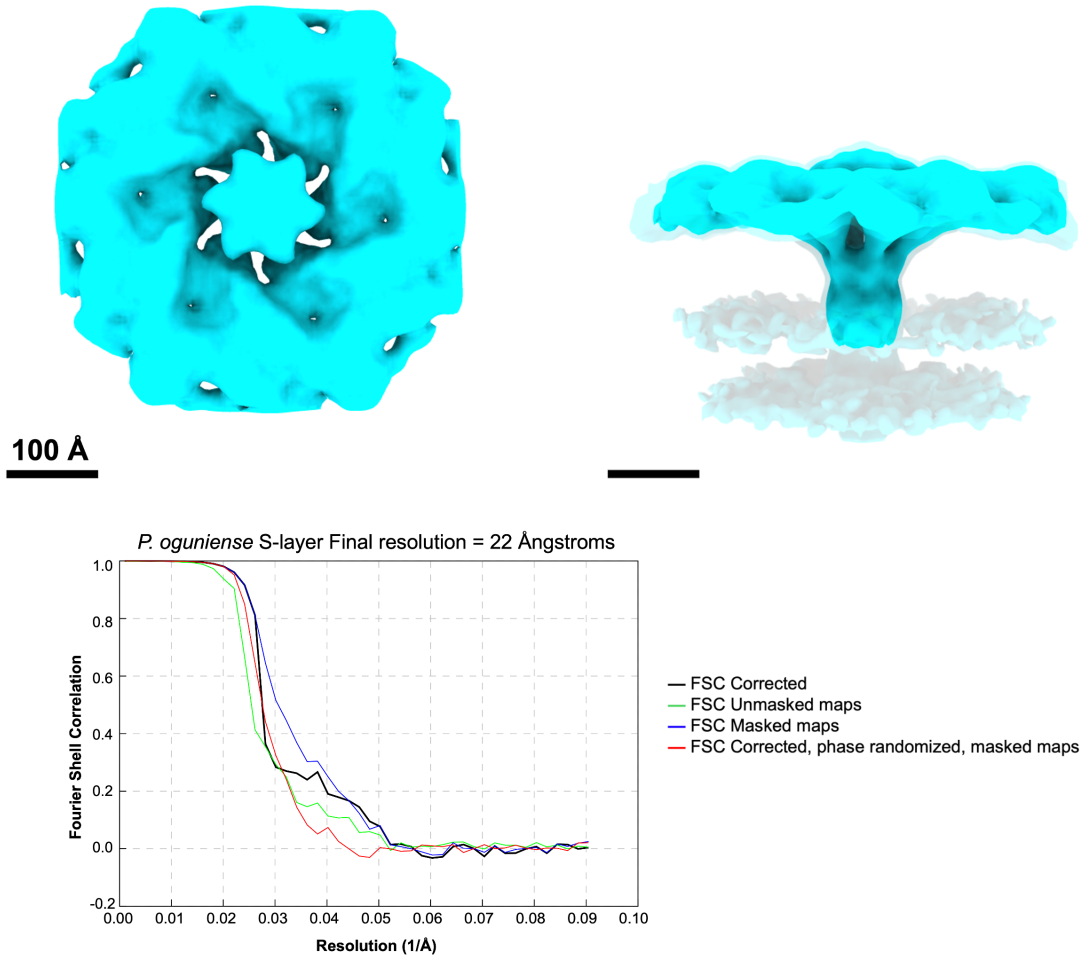



**Figure 5.19:** Sub-tomogram averaging structure of the *P. oguniense* S-layer from whole cells. Top view is shown on the left, with a clear density capping the central pore. The structure is not symmetrised, and displays p6 symmetry. Side view is shown on the right, projected on top of a lower-resolution structure indicating the position of the membrane (bottom) and an unknown layer (middle) with unknown chemical composition. The S-layer forms a stalk-like structure down to the membrane. FSC curve shown below the structure.

#### 5.3.3.3 Tomograms of *P. arsenaticum*

Four tomograms of *P. arsenaticum* cells were taken at a pixel size of 5.527 Å, as with the *P. oguniense* cells (see Figure 5.20). The same protocol was followed for locating sub-tomogram volumes to extract the S-layer particles, however even after extending the lattice, the coverage of the cell surface was poorer for *P. arsenaticum*.

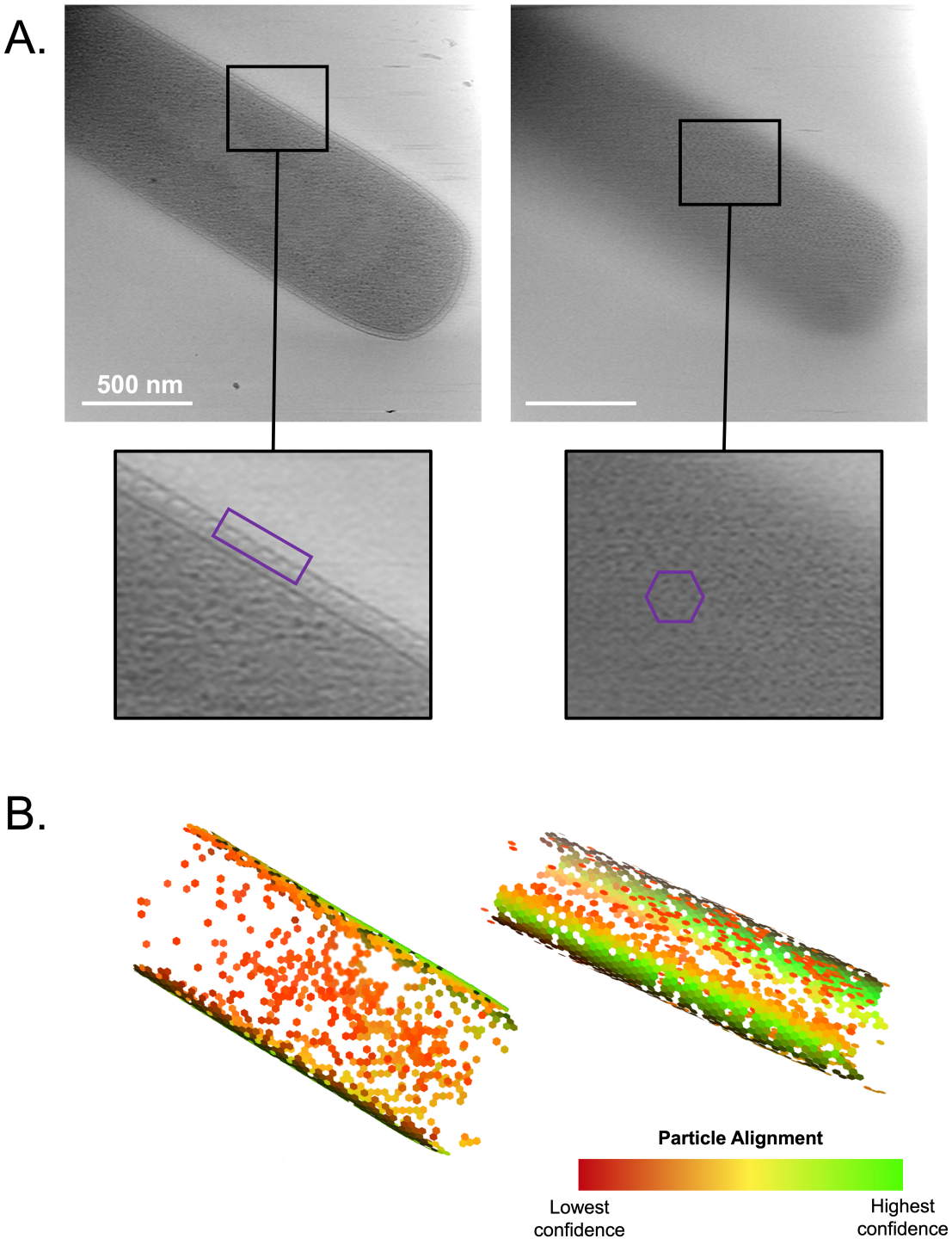



**Figure 5.20:** A. Two slices of a tomogram of a *P. arsenaticum* cell. The S-layer is highlighted, both a side view (rectangle) and a top view (hexagon) of the lattice.
B. Particle picking of the S-layer. Particles coloured by confidence metric, which is calculated by the AV3 toolbox as a comparison between the sub-tomogram represented by the particle and the average structure using cross-correlation.

#### 5.3.3.4 Sub-tomogram averaging structure of *P. arsenaticum* S-layer

Sub-tomogram averaging yielded an S-layer structure with a resolution of 16 Å, shown in Figure 5.21. The S-layer had p6 symmetry, although the monomers could not be distinguished within the hexamer. The central stalk of this S-layer more clearly extended down to the membrane, and in some parts six-fold symmetry could also be observed in the stalk. Glycan-like densities could be observed extending from the top of the S-layer, though not from the cap over the central pore.

#### 5.3.3.5 Comparing S-layer sequences

Two S-layer sequences are annotated in the genomes of *P. oguniense* and *P. arsenaticum* in UniProt. It should be noted that there is no evidence that confirms these are expressed S-layer proteins in either species. Nevertheless, BLAST alignment was performed for each pair of sequences, to identify regions of homology. Multiple sequence alignment was performed using COBALT across all four sequences to identify conserved residues. The results are summarised in Table 5.2, with additional data provided in Appendix Section C.3. BLAST alignment showed that the two most similar proteins were A0A7L4P6S3 from *P. arsenaticum* and H6Q8S4 from *P. oguniense*, which had 97% positive identity and the same number of residues. A0A7L4PCH8 from *P. arsenaticum* was the least similar to the other proteins, being around 100 residues longer and having only a small section of alignment with relatively low identity. H6Q8U3 from *P. oguniense* was very similar to H6Q8S4 and A0A7L4P6S3 from *P. arsenaticum*.

After considering the BLAST results, COBALT was run on the three most similar sequences (A0A7L4P6S3, H6Q8U3, and H6Q8S4; see Appendix Section C.4). The majority of the sequences were identical, with more variation in the H6Q8U3 sequence compared to A0A7L4P6S3 and H6Q8S4. The consequence of this variation in sequence is unknown.

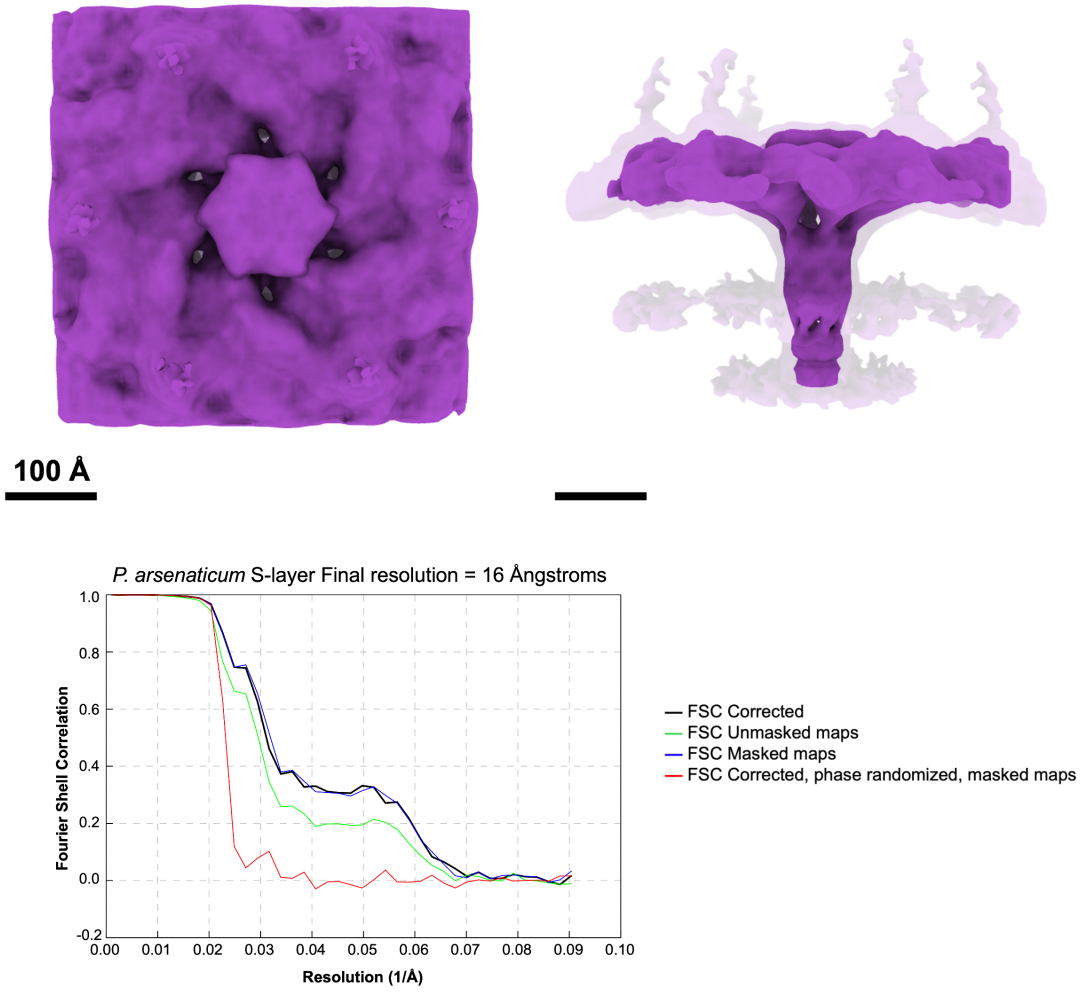



**Figure 5.21:** Sub-tomogram averaging structure of the *P. arsenaticum* S-layer from whole cells. Top view is shown on the left, with a clear density capping the central pore. The structure is not symmetrised, and displays p6 symmetry. Side view is shown on the right, projected on top of a lower-resolution structure indicating the position of the membrane (bottom) and an unknown layer (middle) with unknown chemical composition. The S-layer forms a stalk-like structure down to the membrane. Each monomer also has density extending upwards from the S-layer, which may be glycans. FSC curve shown below the structure.

**Table 5.2:** BLAST alignment of *P. arsenaticum* and *P. oguniense* S-layer protein sequences. Text in red indicates that COBALT could not run on the pair of sequences, and that alignment was poor. Text in green indicates that alignment is very good, and COBALT was run. % Positive identity is the percentage of the aligned sequence where the residue is the same in both sequences, or chemically similar.

| ***Species*** UniProt ID (No. of residues) | ***P. arsenaticum*** A0A7L4PCH8 (791 residues) | ***P. arsenaticum*** A0A7L4P6S3 (688 residues) | ***P. oguniense*** H6Q8U3 (702 residues) | ***P. oguniense*** H6Q8S4 (688 residues) |
|---|---|---|---|---|
| ***P. arsenaticum*** A0A7L4PCH8 (791 residues) | % Positive identity (of alignment length) | 57% positives (of 49 aligned) | no significant alignment found | 57% positives (of 49 aligned) |
| ***P. arsenaticum*** A0A7L4P6S3 (688 residues) | | | 77% positives (of 627 aligned) | 97% positives (of 688 aligned) |
| ***P. oguniense*** H6Q8U3 (702 residues) | | | | 75% positives (of 661 aligned) |
| ***P. oguniense*** H6Q8S4 (688 residues) | | | | |

#### 5.3.3.6 Comparing S-layer models

Models of the putative S-layer sequences for both *Pyrobaculum* species studied were taken from the AlphaFold2 database (Varadi *et al.*, 2022). This provided predictions for the structure of the monomers, and are shown in Figure 5.22. Notably, the PAE scores show that while AF2 is confident in the prediction of the fold of each domain, there is less confidence in the inter-domain contacts. Attempts were made to generate hexamer, trimer, and dimer models using AlphaFold2 in Colabfold and AlphaFold3, however these were unsuccessful at producing models that resembled a lattice and did not identify any potential interfaces between monomers (Jumper *et al.*, 2021; Mirdita *et al.*, 2022; Abramson *et al.*, 2024).

Three of the models have six domains that are dominated by $\beta$-strands. A0A7L4PCH8, which was shown to have poor sequence alignment with the other three sequences (see Table 5.2), has seven domains. By calculating RMSD between the domains of each model, it was also observed that none of the domains of A0A7L4PCH8 shared significant structural similarity to any of the other models, whereas all of the other models had $< 0.7$ Å RMSD between pairs of domains. (For example, domain 1 of A0A7L4PP6S3 to domain 1 of H6Q8S4 has an RMSD

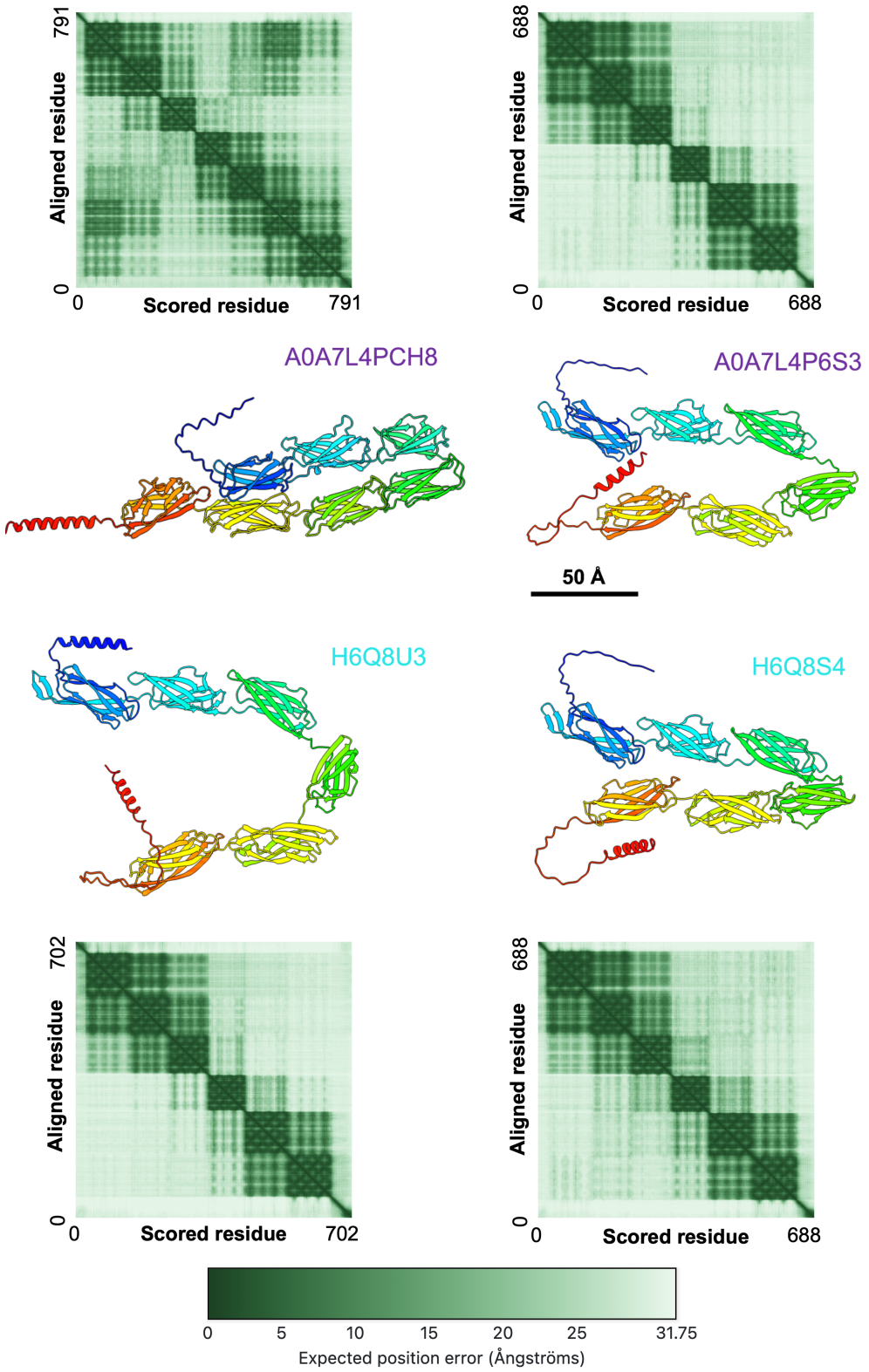



**Figure 5.22:** AF2 models of *Pyrobaculum* S-layer monomers, coloured from N- (blue) to C-termini (red). Models are labelled by UniProt identifier, with *P. arsenaticum* in purple and *P. oguniense* in cyan. All models have a C-terminal $\alpha$-helix. PAE plots shown above or below each model.

of 0.174 Å; see Appendix Section C.5). This suggests that A0A7L4PCH8 may produce a significantly different lattice.

All of the models include a C-terminal $\alpha$-helix. The sequences were passed through TMHMM (Sonnhammer *et al.*, 1998), which predicted that these helices are transmembrane and therefore it is likely that they are responsible for anchoring the S-layer to the cell membrane. The N-terminal helix or disordered region is likely to be a signal peptide. Results from TMHMM are in Appendix Section C.6.

The overall shape of each monomer is not likely to be representative of the shape within the hexamer, as the links between the domains are flexible and inter-monomer contacts are not accounted for in these models.

#### 5.3.3.7 Comparing S-layer structures

The S-layer structures from *P. oguniense* and *P. arsenaticum*, solved by sub-tomogram averaging, were compared. By treating both densities as half-maps, a resolution could be calculated using an FSC curve, giving a resolution to which the two structures were the same (see Appendix Section C.7). This was calculated to be 25 Å, which suggests a high level of similarity as this is close to the resolution of the *P. oguniense* structure. It is likely that the two S-layers have very similar architectures, and that differences in sequence do not make a difference to the fold of the proteins.

Additionally, two horizontal densities are seen in both structures below the lattice, through which the stalk passes. The bottom layer is the membrane, and the layer above is of unknown composition, but is more clearly defined in *P. oguniense*. The unknown density is not consistent with protein, and is not well-defined in the whole-cell tomograms suggesting it is also not lipid.

The S-layer structures do not give any indication to the structural features underpinning the ability of the S-layer to maintain stability in the 95 °C environments of these *Pyrobaculum* species, and higher-resolution structures would be needed to begin to identify these.

### 5.3.4 Molecular dynamics simulations of the *H. volcanii* S-layer

Six different S-layer simulation systems were set up as shown in Figure 5.23, varying the presence and concentration of the divalent cations added into the simulation box (as detailed in Table 5.1). Each system was simulated three times from independent solvent starting coordinates, and the energy was minimised and the system allowed to equilibrate for 10 ns before the simulations that are reported upon below.

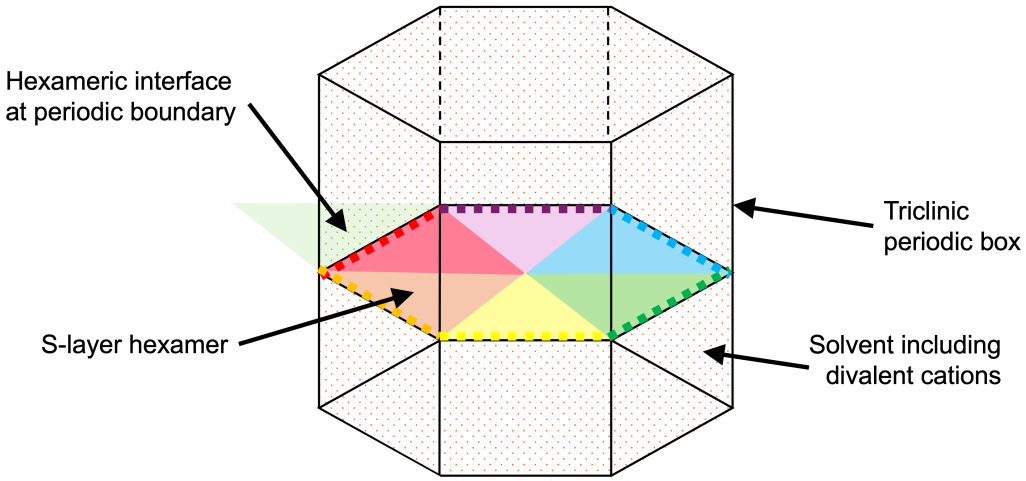



**Figure 5.23:** Simulation setup for *H. volcanii* S-layer. Hexamer shown in triclinic box, with the S-layer extending to each periodic boundary in $x$ and $y$. This allows for the effect of a periodic lattice of hexamers, as shown by the additional pale green segment to the left of the red segment.

#### 5.3.4.1 Stability of S-layer structure in mono- or divalent cations

It was reported from experiments that $Ca^{2+}$ was required for S-layer formation, along with a high concentration of $Mg^{2+}$ in the growth medium (von Kügelgen *et al.*, 2021). To confirm this computationally, simulations of the S-layer were performed where there were no divalent cations present, only NaCl. After 100 ns of simulation, it was observed that the S-layer had undergone significant structural deformation, with the monomers separating from each other at multiple interfaces. The RMSD was calculated and averaged across the three simulations, confirming

this result. This data agrees with the experimental work, showing the S-layer is not stable without divalent cations (see Figure 5.24).

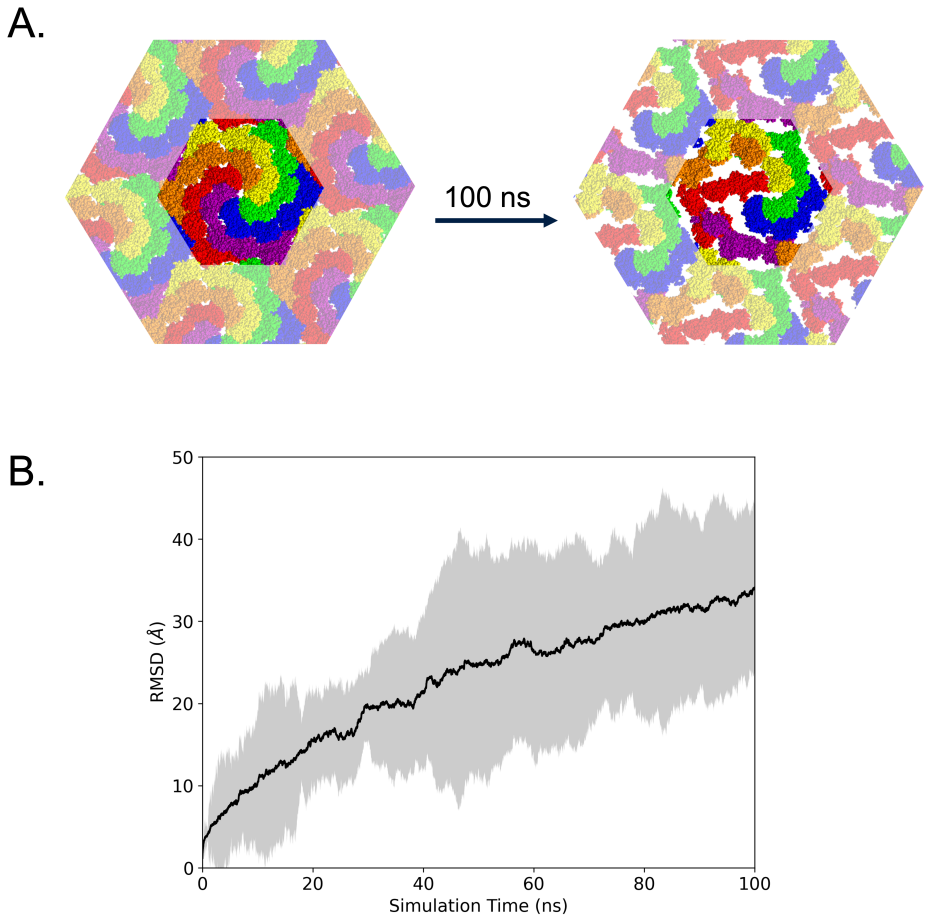



**Figure 5.24:** Structural deformation of S-layer without divalent cations.
A. Start and end of a simulation of the *H. volcanii* S-layer with no divalent cations present.
B. RMSD plot showing the average and standard deviation across the three simulations.

The same analysis was applied to the simulations of the S-layer with either $Ca^{2+}$ or $Mg^{2+}$ present. In all cases, the average RMSD plateaued under 4.5 Å, and reached this plateau within 20 ns. Additionally, the structure was never visibly deformed, and projecting the RMSF onto the structure did not identify any interfaces that consistently broke down across all the simulations (example shown in Figure 5.25).

As a small number of $Ca^{2+}$ ions were placed in the published structure (6 per monomer), simulations with these as the only divalent cations were performed,

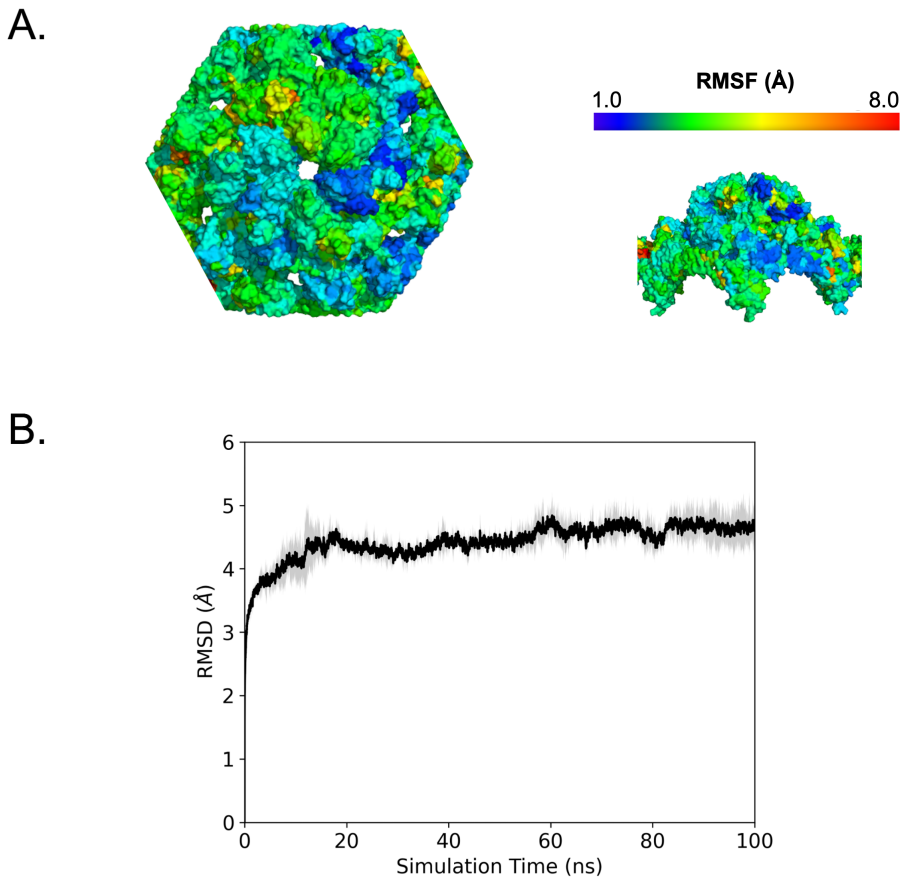



**Figure 5.25:** A. RMSF projected onto the S-layer structure, from one simulation with high concentration of $Mg^{2+}$ and the published $Ca^{2+}$ ions.
B. RMSD plot showing the average and standard deviation across the three simulations with high concentration of $Mg^{2+}$ and the published $Ca^{2+}$ ions.

and the RMSD was calculated. The RMSD increased to a similar level as in the simulations with no divalent cations (both around 32 Å at 100 ns, on average), and there was visual deformation of the structure (see Figure 5.26). This suggests that the ions in the published structure are not sufficient for the stability of the *H. volcanii* S-layer.

#### 5.3.4.2 Identification of potential binding sites

Divalent cations like $Mg^{2+}$ and $Ca^{2+}$ are most commonly coordinated by the negatively charged residues aspartate (ASP, D) and glutamate (GLU, E), then by backbone carbonyl oxygen atoms, followed by other oxygens in functional groups

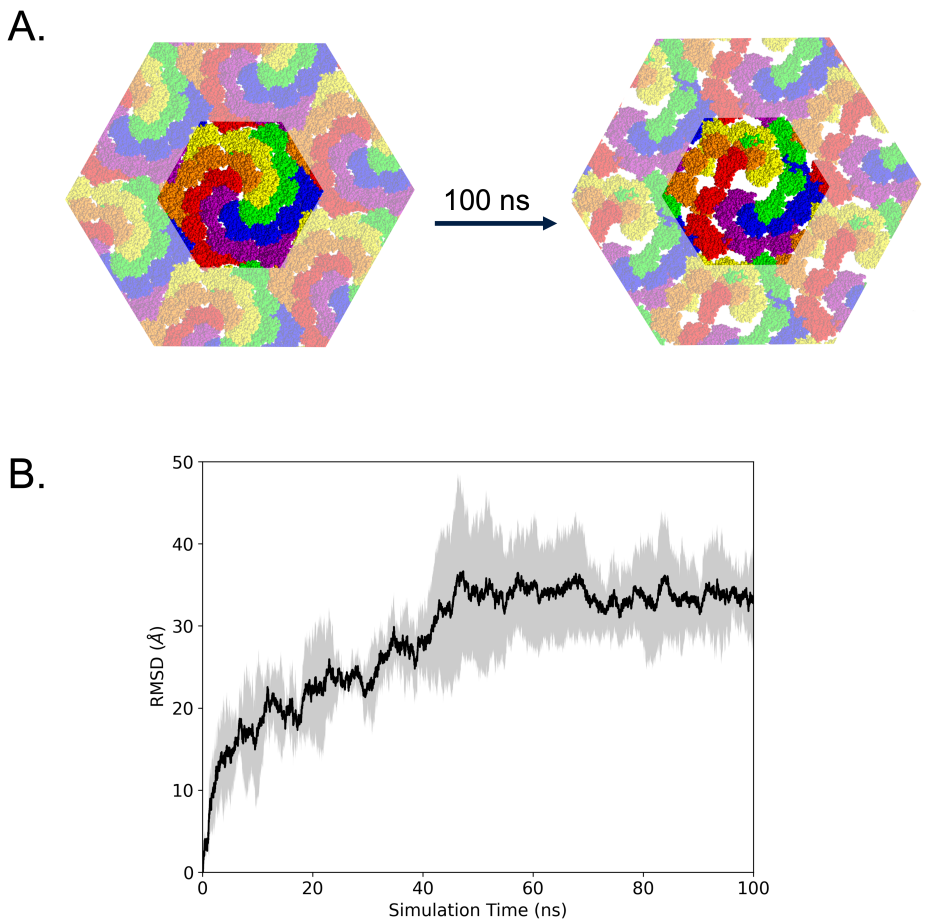



**Figure 5.26:** Structural deformation of S-layer with only published divalent cations.
A. Start and end of a simulation of the *H. volcanii* S-layer with only the $Ca^{2+}$ ions present in the published structure.
B. RMSD plot showing the average and standard deviation across the three simulations.

(Harding, 2004). The search for potential binding sites was undertaken without taking into account the densities seen in the cryo-EM structure, to see if the sites could be reproduced using simulation data. Thus, contact analysis was performed on the last 50 ns of each simulation to identify pairs or trios of residues that were all in contact (defined as one of the oxygen atoms in the functional group or backbone being within 5 Å of the ion) with the same divalent cation for at least 0.1 ns continuously, in at least one repeat of one simulation. Additionally, as the S-layer is a hexamer, there were six copies of each binding site in each simulation. Every discrete incidence of the coordination of a divalent cation was recorded. Then the

list of residue pairs/trios was filtered to remove any that appeared fewer than 20 times. This resulted in a list of 104 potential divalent cation binding sites per monomer in the *H. volcanii* S-layer, which are detailed in Appendix Section C.8.

#### 5.3.4.3 Location of potential binding sites

APBS electrostatics calculations, visualised in Figure 5.27, showed that the whole surface of the S-layer was negatively charged, which suggested there will be a large number of cation binding sites (Jurrus *et al.*, 2018). The small blue patch is an arginine residue near the C-terminus of the protein model, where biologically there would be more residues present for the domains of the protein that link to the membrane.

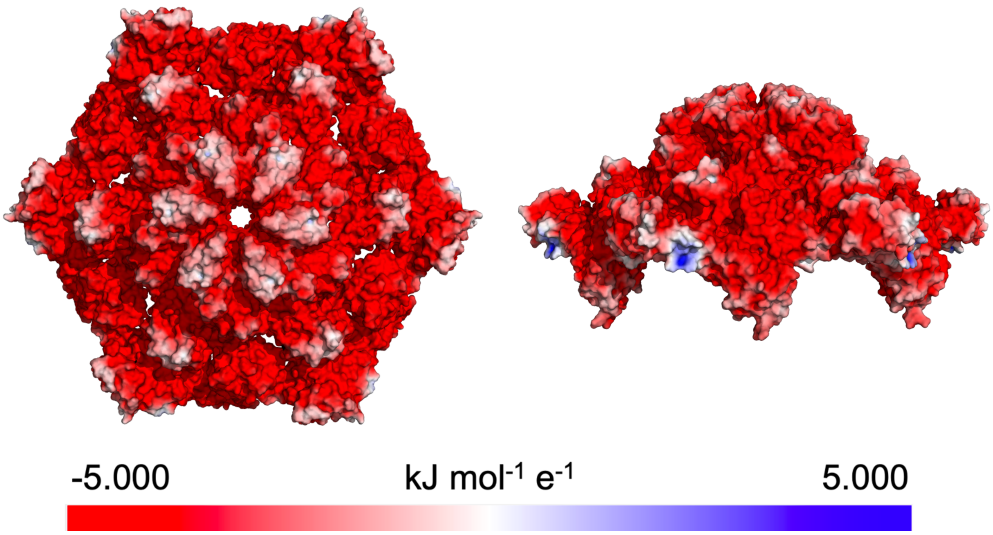



**Figure 5.27:** Visualisation of the electrostatic surface of the *H. volcanii* S-layer. Positive electrostatic potential shown in blue representing positively charged residues, negative electrostatic potential shown in red representing negatively charged residues.

VolMap in VMD was used to calculate the ion densities on the protein across the combined simulation trajectories, separated by ion type (Humphrey *et al.*, 1996). The last 50 ns of each simulation was used. p6 symmetry was then applied to the maps using RELION (Kimanius *et al.*, 2021). The maps were visualised in ChimeraX, overlaid with the protein structure, and are shown in Figure 5.28 (Pettersen *et al.*, 2021). Slices of the VolMap densities are shown in Appendix

Section C.9. The maps show that both $Ca^{2+}$ and $Mg^{2+}$ ions decorate the outside and inside of the hexameric structure, suggesting the high number of predicted binding sites is likely to be accurate.

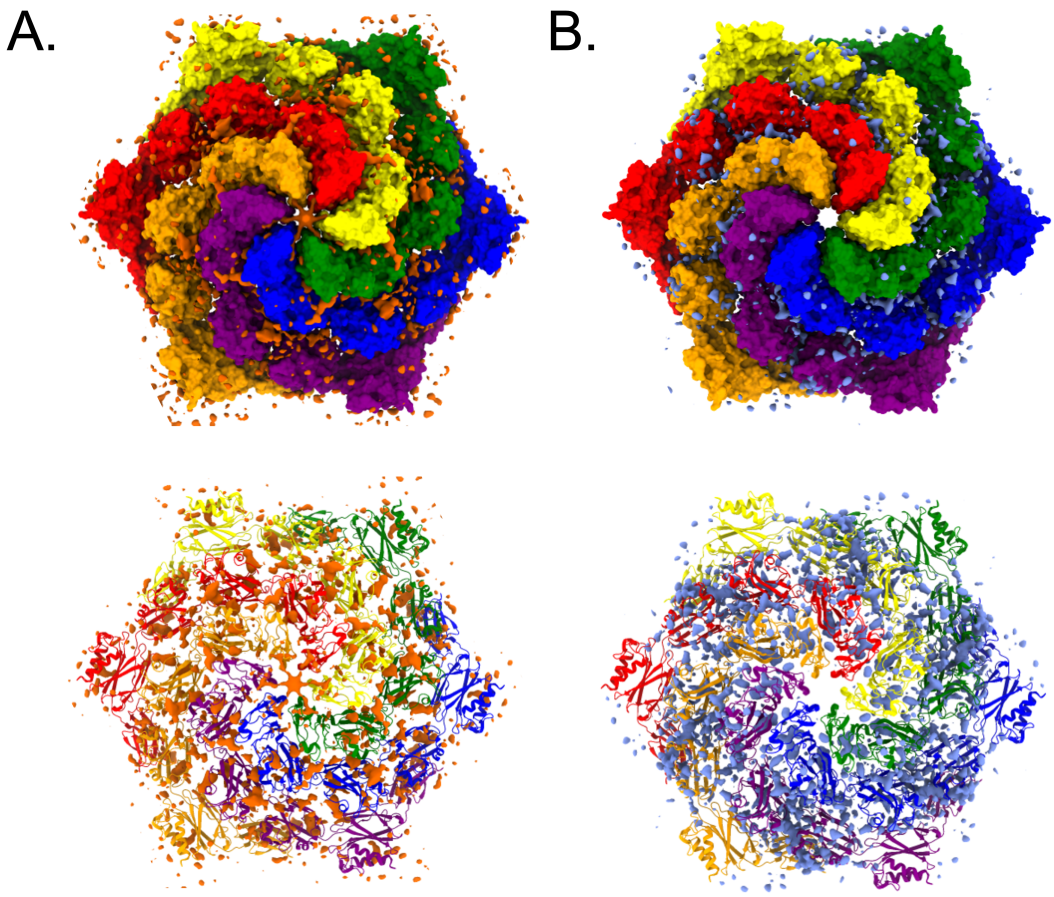



**Figure 5.28:** Ion density around the S-layer structure.
A. $Ca^{2+}$ density shown on S-layer with surface representation (top) and cartoon representation (bottom) to show binding sites on the surface and within the protein.
B. $Mg^{2+}$ density shown on S-layer with surface representation (top) and cartoon representation (bottom).

#### 5.3.4.4 Calcium and magnesium occupancy

The occupancy of every potential binding site was calculated for both divalent cation species, combining data from all of the simulations and accounting for all six copies of the monomer in each simulation box (see Appendix Section C.10). Sites with occupancy under 10% of simulation time for both ions were discarded

(such as site 73), as well as sites that were duplicates (such as 5 and 6; site 5 was discarded). The remaining were classified into intra- and inter-monomer, and inter-hexamer coordination sites. In total, 14 sites were not intra-monomer, with 12 of these having a preference for $Ca^{2+}$ and 2 having no statistically significant difference between occupancies of the ions.

73 binding sites were identified per subunit, including the 6 that had been modelled in the starting structure. These were then divided into four categories based on the highest occupancy between $Ca^{2+}$ and $Mg^{2+}$: 75–100% classed as high occupancy sites; 50–75% classed as medium occupancy cites; 25–50% and 10–25% classed as low and lowest occupancy sites. The ion species with the highest occupancy was built into the site, and then the size of the atom was scaled relative to the occupancy class, as visualised in Figure 5.29.

The assignment of ion identity to each binding site is shown in Table 5.3. If there was no statistical difference between the occupancy of both ions, other biological context was considered. Firstly, the number of oxygen atoms available for coordination was counted, with more than 6 indicating the site has a preference for $Ca^{2+}$ due to the many-body polarisation effect (Jing *et al.*, 2018). The distance from the oxygen atoms to the ion was calculated because $Ca^{2+}$ and $Mg^{2+}$ have different coordination bond lengths: $Ca^{2+}$ has an average of 3.5 Å to the ligand, and $Mg^{2+}$ has a shorter average distance of 2.1 Å (Cates *et al.*, 2002).

**Table 5.3:** Assignments of ion identity, by percentage of occupancy. Unidentified ions represent sites where there was no statistically significant difference in occupancy for $Ca^{2+}$ and $Mg^{2+}$ ions, and there were no other factors from simulation or structure suggesting one would be physiologically favoured.

| | **$Ca^{2+}$** | **$Mg^{2+}$** | **Unidentified** |
|---|---|---|---|
| **75–100%** | 1, 6, 14, 16, 17, 20, 28, 34, 35, 36, 37, 47, 49, 63, 76, 77, 78, 85, 86, 88, 91, 93, 96, 97, 99, 100, 102, 104 | 18, 58, 80, 83 | 19, 82 |
| **50–75%** | 2, 32, 40, 42, 50, 53, 54 | 27, 48, 71 | 87 |
| **25–50%** | 10, 23, 30, 44, 56, 60, 72, 89, 94 | | 8, 15, 43 |
| **10–25%** | 11, 12, 22, 25, 29, 39, 59, 64, 66, 67, 70, 75, 81, 101, 103 | | 21 |

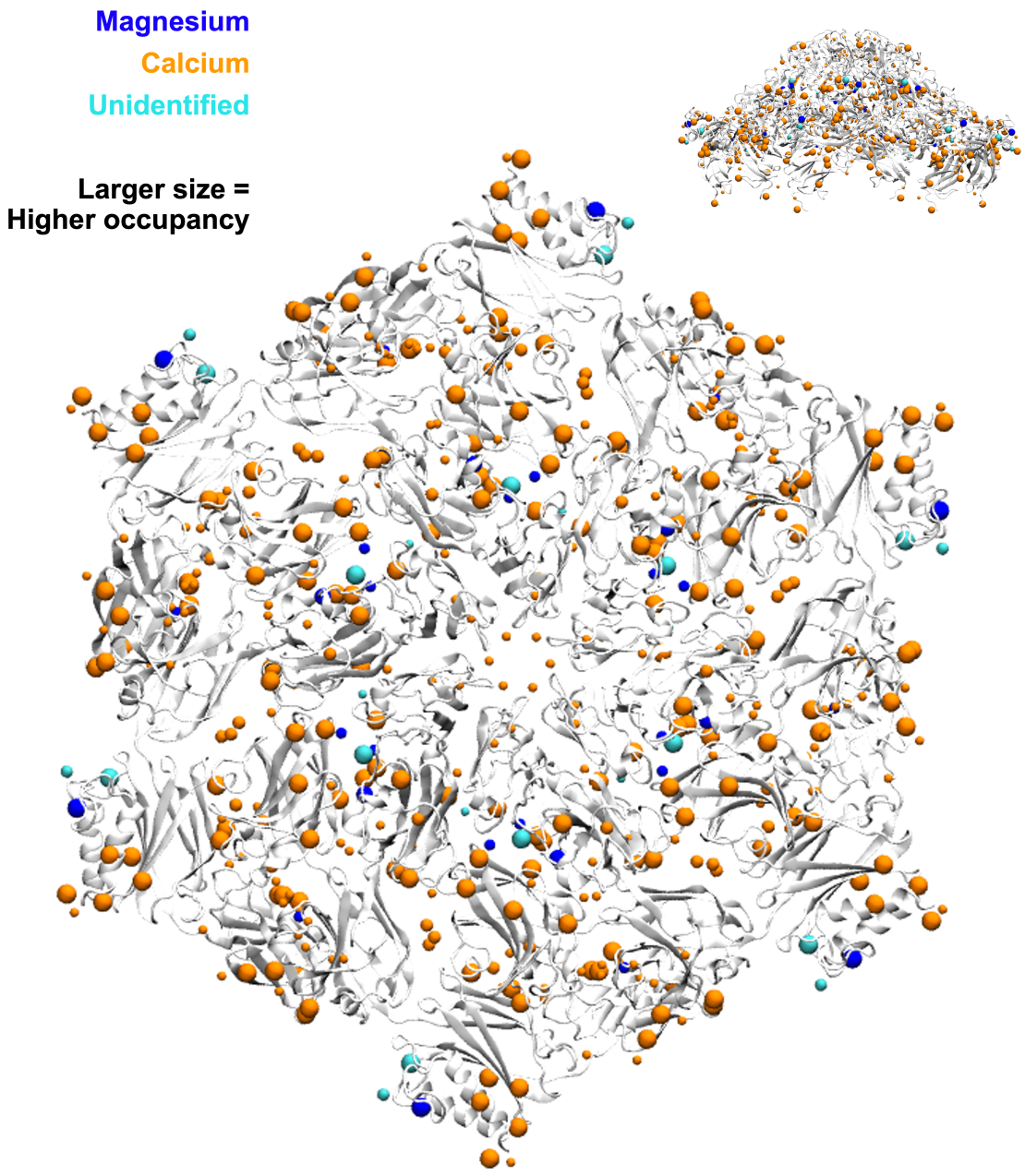



**Figure 5.29:** Divalent cation binding sites in the *H. volcanii* S-layer. Each binding site is represented six times, for each copy of the monomer in the hexameric structure. The size of the ion representation reflects the percentage occupancy observed for that site (alternative representation shown in Appendix Section C.11). Unidentified ions represent sites where there was no statistically significant difference in occupancy for $Ca^{2+}$ and $Mg^{2+}$ ions, and there were no other factors from simulation or structure suggesting one would be physiologically favoured.

#### 5.3.4.5 High occupancy

34 sites were classified as high occupancy, with 28 having a preference for $Ca^{2+}$, 4 for $Mg^{2+}$, and 2 showing no statistical difference between occupancy of both ions. All of the high-occupancy sites are intra-monomer sites. Occupancy data and two exemplifying sites are shown in Figure 5.30.

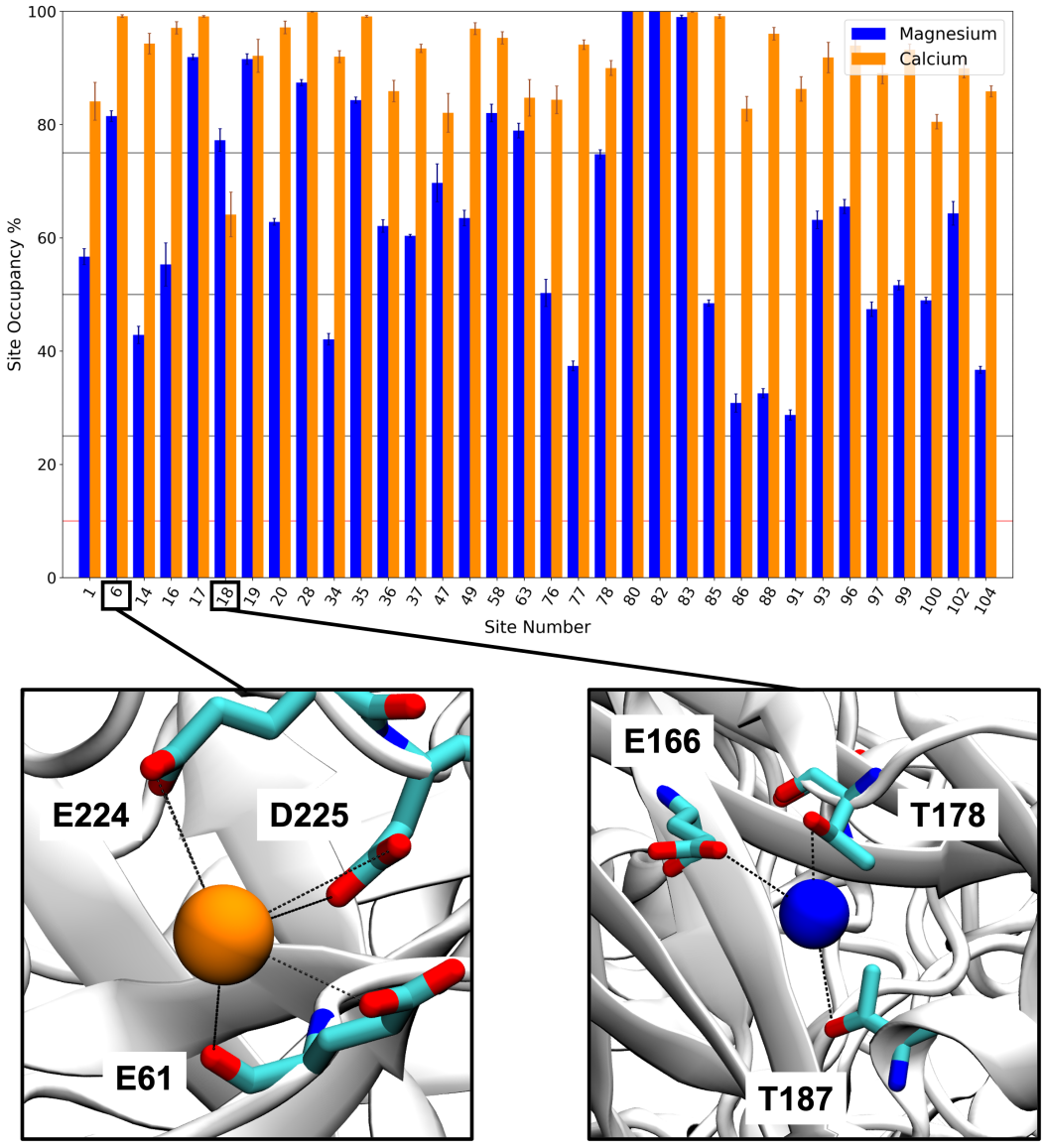



**Figure 5.30:** Graph of occupancy data for the high occupancy binding sites, showing occupancy for $Ca^{2+}$ and $Mg^{2+}$ ions with standard error of the mean. Visualisation of site 6 ($Ca^{2+}$) and site 18 ($Mg^{2+}$) shown, with residues forming the site in licorice representation and the rest of the protein in white cartoon representation.

#### 5.3.4.6 Medium occupancy

11 sites were classified as medium occupancy, with 7 having a preference for $Ca^{2+}$, 3 for $Mg^{2+}$, and 1 showing no statistical difference between occupancy of both ions. Of these medium occupancy sites, 4 are inter-monomer. Occupancy data and two exemplifying sites are shown in Figure 5.31.

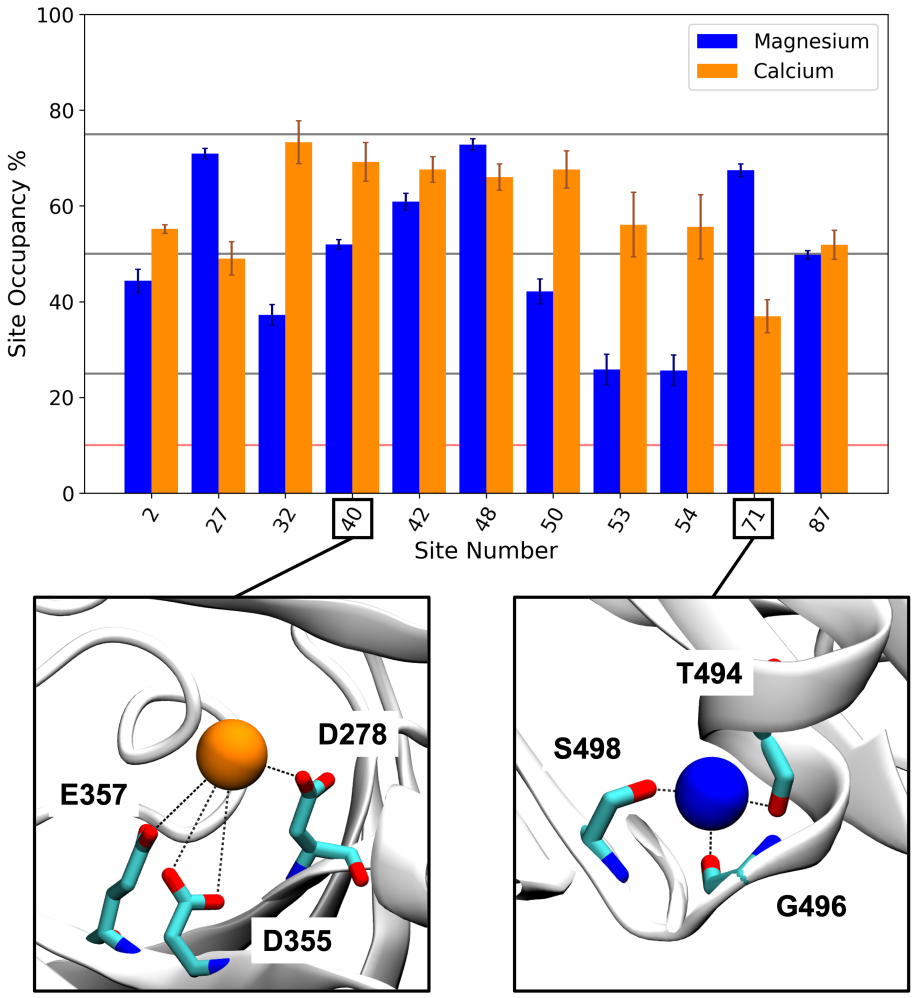



**Figure 5.31:** Graph of occupancy data for the medium occupancy binding sites, showing occupancy for $Ca^{2+}$ and $Mg^{2+}$ ions with standard error of the mean. Visualisation of site 40 ($Ca^{2+}$) and site 71 ($Mg^{2+}$) shown, with residues forming the site in licorice representation and the rest of the protein in white cartoon representation.

#### 5.3.4.7 Low occupancy

12 sites were classified as low occupancy, with 9 having a preference for $Ca^{2+}$ and 3 showing no statistical difference between the occupancy of both ions. 3 of the 12 sites are inter-monomer sites.

16 sites were classified as lowest occupancy, with 15 having a preference for $Ca^{2+}$ and 1 showing no statistical difference between the occupancy of both ions. 7 of the sites are inter-monomer sites. Graphs of both low and lowest occupancies are shown in Figure 5.32.

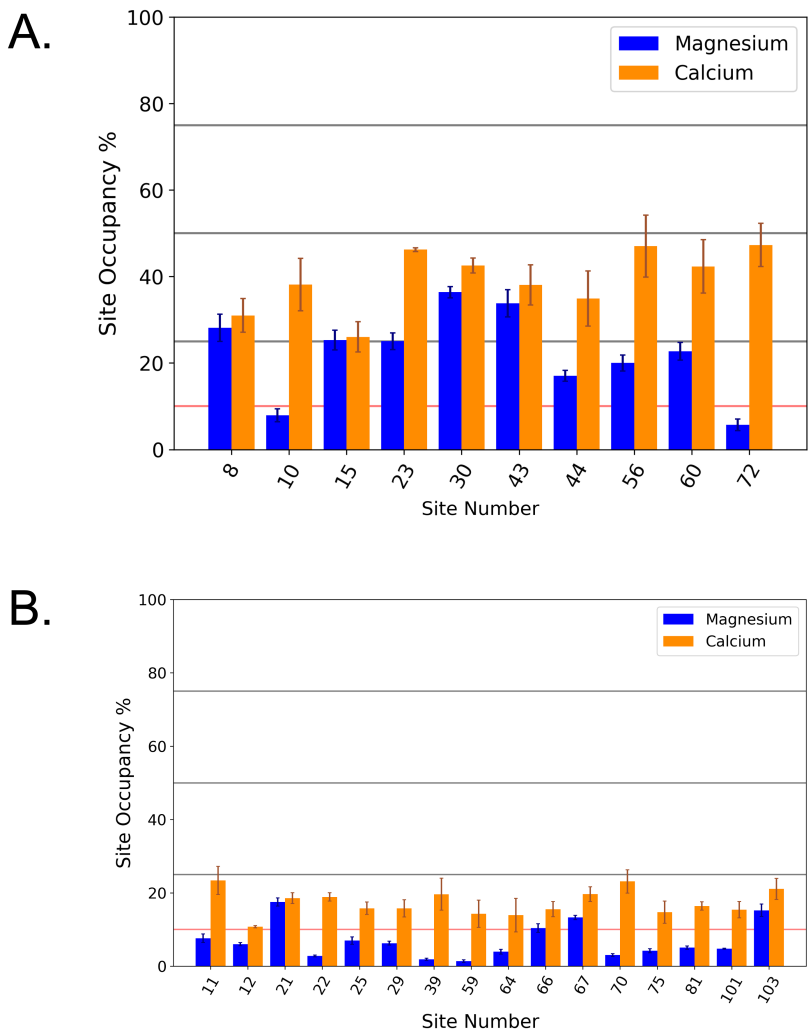



**Figure 5.32:** Low occupancy binding sites. A. 25% to 50%. B. 10% to 25%.

#### 5.3.4.8 Previously published binding sites

The binding sites in the published structure from von Kügelgen *et al.* (2021) were independently identified using the method discussed in subsubsection 5.3.4.4. They were numbered as sites 10, 11, 80, 82, 83, and 103. The graph of occupancy for these sites is shown in Figure 5.33, along with part of the model in the cryo-EM structure. This shows that there can be confidence in the data from the MD simulations, due to the agreement with the original cryo-EM data, and suggests this could be an approach used for other situations where the identity of ions in cryo-EM structures is unclear. Three of the published sites have low occupancy (10, 11, and 103) in simulations and are still visible in the cryo-EM density, however the simulations were able to provide more data on the likely identity of the ions in the sites which could not be resolved from the quality of the map. This was acknowledged by von Kügelgen *et al.* (2021) in their paper.

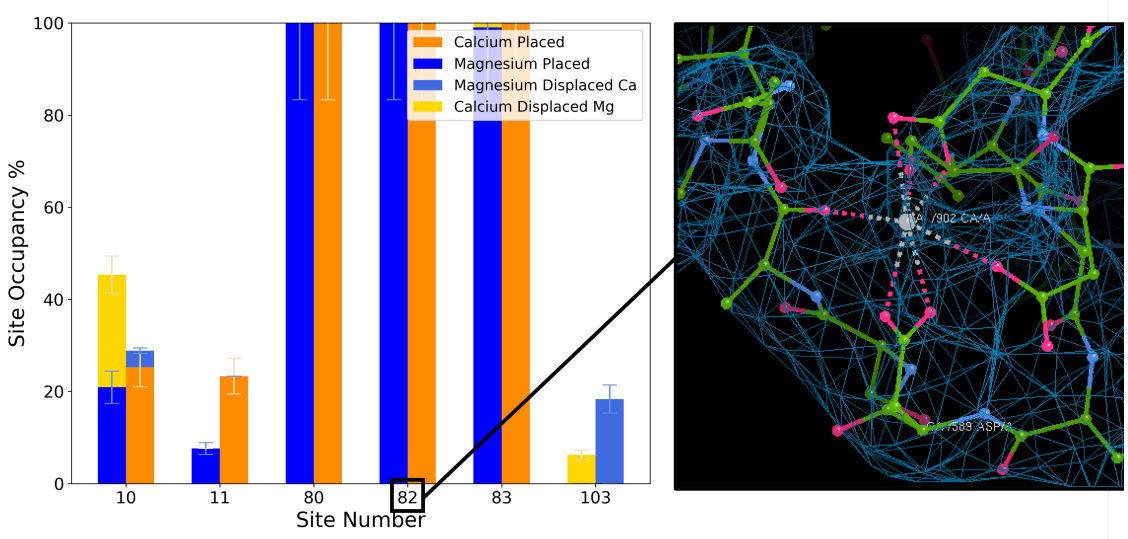



**Figure 5.33:** Occupancy of published binding sites, including simulations where one ion started in a site and was displaced by a different ion species. The model of site 82 from the published structure is shown on the right in the electron density, with the modelled $Ca^{2+}$ and coordinating residues.

#### 5.3.4.9 Validation of binding sites

The sites identified in the work presented here were then located in the original cryo-EM density published by von Kügelgen *et al.* (2021) (see Figure 5.34). In all cases, unmodelled density was observed in the space between the residues forming the site, and the density resembled a coordinated ion. Therefore the simulations were successful at independently finding the binding sites.

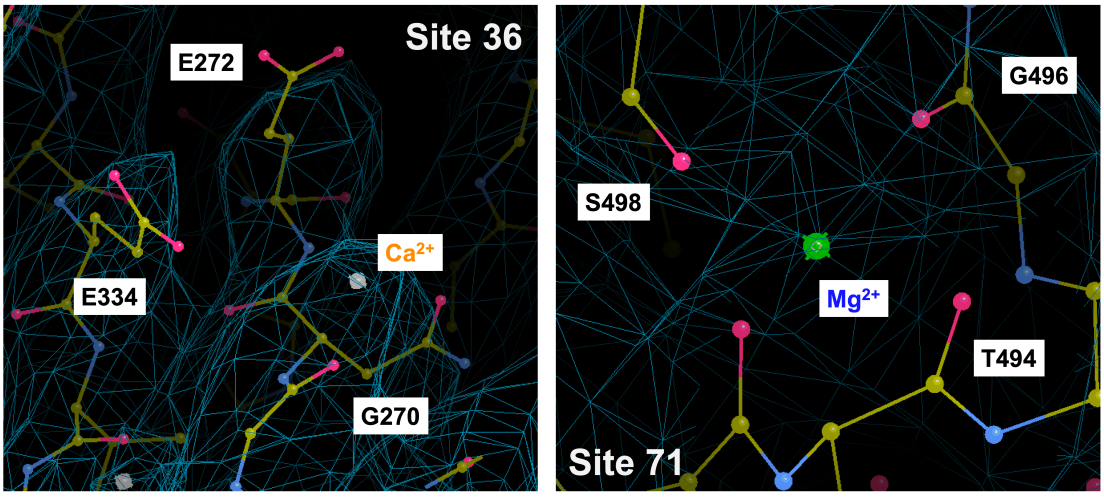



**Figure 5.34:** Building ions into the electron density of the *H. volcanii* S-layer from von Kügelgen *et al.* (2021). Site 36 ($Ca^{2+}$) and site 71 ($Mg^{2+}$) are shown, built into unoccupied density.

Additionally, the metrics used to assign ion preferences to sites that were indistinguishable were then checked against the preferences determined by occupancy percentage. The average coordination number and bond length were lower for $Mg^{2+}$, in agreement with the literature (Cates *et al.*, 2002; Jing *et al.*, 2018).

Many more sites were assigned to have a preference for coordinating $Ca^{2+}$, just as there are many more $Ca^{2+}$-specific binding proteins than $Mg^{2+}$-specific (Tang and Yang, 2013). Also, many of the sites assigned with $Ca^{2+}$ include bidentate coordination of the ion (see Figure 5.31), which has been observed previously (Cates *et al.*, 2002).

## 5.4 Discussion

### 5.4.1 Structural comparisons

#### 5.4.1.1 Ig-like folds dominate S-layer structures

All of the S-layers within the work presented here contain multiple domains that visually appear to have Immunoglobulin-like (Ig-like) folds. The Encyclopedia of Domains database has annotated protein models from the AF2 database by domain, with domains classified according to the CATH database of tertiary structure superfamilies (Orengo *et al.*, 1999; Knudsen and Wiuf, 2010; Varadi *et al.*, 2022; Lau *et al.*, 2024). These databases were used to compare the folds in the domains of the S-layers studied.

The SlaA protein from *S. solfataricus* has two domains, 5 and 6, classified as the Ig-like CATH topology group 2.60.40. Domain 7 is in the CATH topology group 2.70.50 (Coagulation Factor XIII), which is a distorted sandwich architecture opposed to the sandwich architecture of topology 2.60.40. The *S. acidocaldarius* SlaA structure also contains two domains in the CATH topology group 2.60.40, domains 5 and 7. These overlap with the sections of sequence with the highest homology to *S. solfataricus*, as shown in Figure 5.16. The first four domains were not classified by CATH in either species, so this method of comparison cannot be used as an assessment of structural similarity for this portion of the proteins.

*P. arsenaticum* and *P. oguniense* both have two proteins annotated (but unconfirmed) as S-layer proteins in UniProt. All four of these proteins are classified as having six domains in the Encyclopedia of Domains, with all of those domains being part of the Ig-like topology 2.60.40. All but one domain are classified as 2.60.40.10, which is the Immunoglobulin homologous superfamily. Notably, the A0A7L4PCH8 protein from *P. arsenaticum* visually has seven domains, but only six were identified by the Encyclopedia of Domains, with the sixth domain being excluded. Also, domain 5 in this protein is classified as 2.60.40.1930, part of the Macroglobulin homologous superfamily. These results agree with the BLAST alignment results that identified A0A7L4PCH8 as being least similar to the other

S-layer proteins, and also the work of Bharat *et al.* (2021), which found that fold and topology was more conserved than sequence in S-layer proteins.

Six domains were also identified in the *H. volcanii* S-layer monomer (von Kügelgen *et al.*, 2021). One is unclassified, and the other five are part of the Ig-like topology. Similar to the *P. arsenaticum* S-layer protein A0A7L4PCH8, one of the domains is in the Immunoglobulin homologous superfamily, and one is in the Macroglobulin homologous superfamily.

As the S-layer protein from *N. maritimus* is not annotated in UniProt, it does not have an entry in the AF database and is therefore not in the Encyclopedia of Domains. However, as stated in von Kügelgen *et al.* (2024), the protein has ten Ig-like domains.

The dominance of $\beta$-strand-based topologies in S-layers was previously identified in Bharat *et al.* (2021), and the work presented here agrees with these results. The level of identity for the domains being truly Ig-like varies between the S-layers, with those from *Pyrobaculum* being truly homologous to Ig, and those from *S. solfataricus* not all having the sandwich architecture. However, several of the domains are not assigned a CATH group, which makes identifying similarities and differences more difficult.

#### 5.4.1.2 Multiple centres of symmetry in p6 lattices

2D lattices composed of hexagonal subunits with six-fold rotational symmetry must always have centres of p3 and p2 symmetry as well as p6, as seen in Figure 5.35 (Scherrer, 1946; Terzibaschian and Enderlein, 1986). This is the case for all of the S-layers studied here, as S-layers are 2D lattices.

The *N. maritimus* S-layer has pseudo-six-fold symmetry, with the pore at the centre of the hexamer having two-fold symmetry. However, the S-layer does still have six-fold symmetry at the outer edges of the hexamer, and therefore has trimeric and dimeric interfaces, which are centres of p3 and p2 symmetry respectively (von Kügelgen *et al.*, 2024). The two-fold symmetry of the pore was not clear in the sub-tomogram averaging structure shown in subsubsection 5.3.1.2, as higher

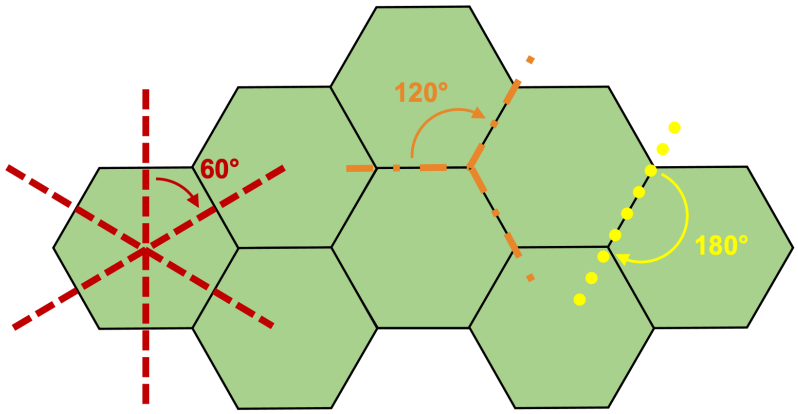



**Figure 5.35:** Rotational symmetry in p6 2D lattices. From left to right: six-fold symmetry (red), three-fold symmetry (orange), two-fold symmetry (yellow).

resolution details were required. Also, the premature application of p6 symmetry to the structure would have obscured this feature.

The S-layers of *S. solfataricus* and *H. volcanii* have true six-fold symmetry, and the atomic models show the details of the trimeric and dimeric interfaces, as detailed by Gambelli *et al.* (2024) and von Kügelgen *et al.* (2021) respectively.

As only lower-resolution structures are available for the *P. oguniense* and *P. arsenaticum* S-layers, the conclusion that they have p6 symmetry cannot be confirmed yet with atomic detail. However, the current structures do not have applied symmetry, and still clearly exhibit features of six-fold rotational symmetry such as the arms coming out from the pore, the pore cap, and the stalk.

#### 5.4.1.3 Impact of surface curvature

During the lifetime of a cell, the membrane has to adopt a wide variety of curvatures (Mukhopadhyay and Wingreen, 2009). This includes both convex and concave curvatures, the most extreme of which are required for membrane invagination (e.g. during cell division). As the S-layer is connected to the membrane, curvature of the membrane is accompanied by curvature of the S-layer (von Kügelgen *et al.*, 2020). Therefore the S-layers of cells must have inherent flexibility to accommodate these changes and curvatures, both within and between symmetric units. In some

archaea, the S-layer plays a direct role in cell morphology including how it relates to cell division (Pum *et al.*, 1991).

The structure of the *S. solfataricus* S-layer solved in the work presented here was determined from tomograms of vesicles, which have a high degree of curvature. This means that the structure will have differences to a structure solved from S-layer sheets, which would predominantly be planar. Also, the vesicles were pleomorphic, meaning the curvature experienced by each S-layer subunit varied. As this likely has consequences for the subunit structure, there would be an impact on the overall resolution of the STA structure, with poorer resolution at areas of flexibility. This resolution heterogeneity could be used to identify flex points, but also may limit the resolution obtainable from S-layers *in situ*, in any species.

For the rod-shaped *Pyrobaculum* S-layers, S-layer subunits from the cylindrical body will have more similar curvature to the subunits around the cell poles, matching the curvature of the membrane (Gill *et al.*, 2015). It may mean that *in situ* structures of S-layers using only subunits in areas of similar curvature produce higher resolution structures, as the angles between domains may subsequently be more similar.

Molecular dynamics is another method for investigating the atomic details of S-layer flexibility to produce curvature. By combining a single-particle high-resolution structure with RMSF data about residue flexibility, domains or interfaces that allow accommodation of cell curvature could be identified and further investigated. This could lead to further insight about the rigidity of S-layers and their role in cell shape and protection from osmotic shock.

#### 5.4.1.4 Pentameric defects

S-layer lattices achieve complete cell coverage by the inclusion of non-hexameric S-layer units such as pentamers or by gaps that are also often pentameric in shape. This is due to the geometric requirements of covering a curved surface with a hexameric lattice (Goldberg, 1937). S-layer defects have been observed in archaea for many years, however the structural basis behind them was not

understood (Pum *et al.*, 1991). It is not readily possible to predict if S-layer monomers will form stable polymers of more or less than six subunits. Therefore, structural studies are often required.

The S-layer of *N. maritimus* contains pentamers, as does the S-layer of *H. volcanii* (von Kügelgen *et al.*, 2021, 2024). Gambelli *et al.* (2024) did not identify pentameric defects in their work on *S. acidocaldarius*, however they were observed in the work presented here in the *S. solfataricus* vesicles (see Figure 5.11). Unfortunately, the limited data collected here was not sufficient to investigate if this pentameric defect was a gap or a pentameric form of SlaA. Also, due to the limited data collection and resolution, pentameric defects were not observed in the S-layers of *P. oguniense* and *P. arsenaticum*. However, the fact that they were not observed does not preclude them from being present as no other forms of defect were identified.

#### 5.4.1.5 Cell envelope composition

The most unexpected find in the STA S-layer structures from *P. arsenaticum* and *P. oguniense* was the presence of an unknown density between the S-layer lattice and the membrane. This density did not have observable order and at the resolution solved, there were no specific contacts seen to the S-layer stalk. Currently there is no literature suggesting what this extracellular polymer may be for *Pyrobaculum*, though some have previously been identified in other archaeal species as being formed from methanochondroitin or glutaminylglycan (Klingl *et al.*, 2019). This novel finding is likely to contribute to the ability of *Pyrobaculum* to survive in extremely high temperature environments.

### 5.4.2 Sequence comparisons

#### 5.4.2.1 Sequence homology between species in the same genus

The sequence and structure of S-layers were compared between pairs of species in two genera, two species of *Saccharolobus* and *Pyrobaculum*. This was to investigate how quickly the sequences are able to diverge, as Bharat *et al.* (2021) showed that, in some cases, even species in the same genus can be highly dissimilar.

The two *Pyrobaculum* species are known to be the most highly related in the genus, and many of the genes that are in both species have highly conserved sequences (Bernick *et al.*, 2012). It was not known if this applied to the S-layer proteins. The work presented here discovered that the structures are highly similar, however atomic structure or even secondary structure determination was not possible. The comparison of the sequences between the two species revealed that the two putative S-layer sequences in *P. oguniense* are highly homologous to one of the S-layer sequences in *P. arsenaticum* but not the other. It is not currently known which protein sequences, if any, form the structures that were solved.

The *S. solfataricus* S-layer sequence and structure presented here were compared to those of *S. acidocaldarius*, as published by Gambelli *et al.* (2024). Sequence alignment revealed that the domains forming the centre of the lattice are not homologous, as shown in Figure 5.36. It is clear that the sections of homology (in black and white) are much less significant to the architecture of the hexamer, which may be a reason for the lack of structural similarity in the S-layer hexamer.

Taken together, these pairs of S-layers demonstrate the broad spectrum of sequence and structure variation of S-layers within species of the same genus. This is very unusual when compared to the conservation of sequence and structure seen in other proteins performing important functions in closely related species. When comparing proteins essential to cell survival between two species in the same genus, it is usually expected that they will have a high degree of homology, and this can often be used to infer evolutionary relationships between species. Therefore it is a surprising result to see how much variation has evolved between the S-layers of *S. acidocaldarius* and *S. solfataricus*, for example, in such a short span of evolutionary time when the function of the protein is essential.

#### 5.4.2.2 Gene redundancy in S-layers

It was reported by von Kügelgen *et al.* (2024) that, of two similar S-layer genes encoded in the *N. maritimus* genome, one was predominantly expressed and the abun-

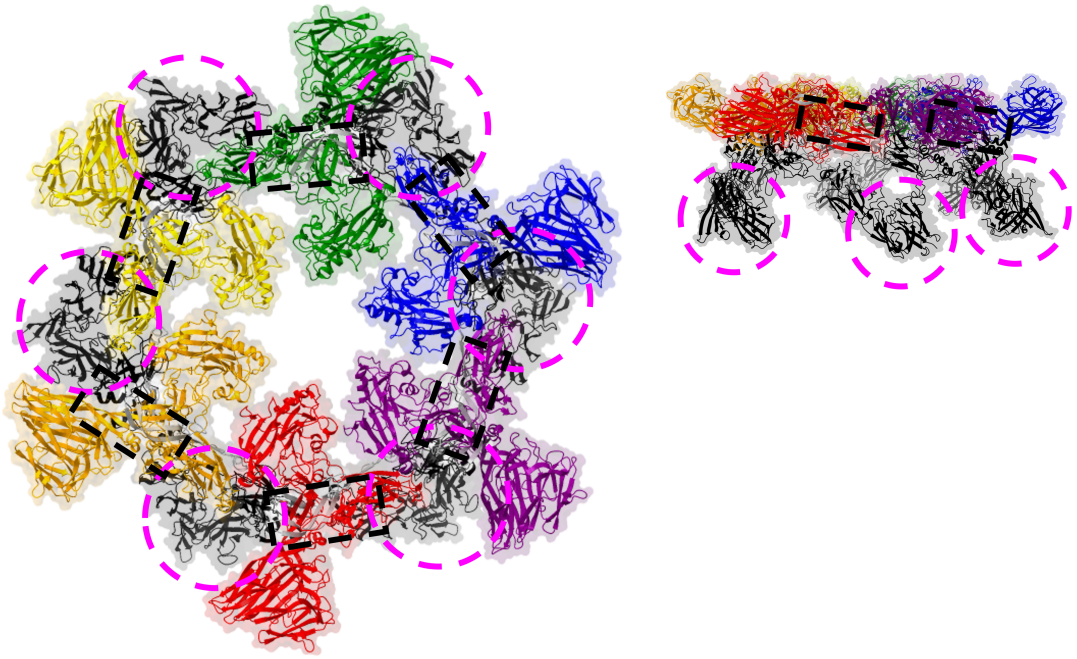


**Figure 5.36:** *S. acidocaldarius* model coloured by homology to *S. solfataricus*. Regions highlighted in pink (black cartoon model) and black (white cartoon model) have homology to *S. solfataricus* S-layer as calculated in Appendix Section C.2. Coloured regions of the chains are not homologous. Right inset shows the side view, and that the black domains do not contribute to the top of the lattice.

dant protein in the S-layer. To determine the abundant protein, transcriptomics data and unambiguous residue sequences identified in the atomic structure were combined.

There are also two annotated S-layer genes in each of the *Pyrobaculum* species studied here. However, such transcriptomics data does not exist, and sufficient structural resolution was not achieved, so this method could not be used to confirm or determine the abundant S-layer protein in these species. By examining the data that is available, some conclusions can nevertheless be tentatively drawn. One of the sequences from *P. arsenaticum*, A0A7L4PCH8, is not similar to the other S-layer sequence in that species, nor the two from *P. oguniense*. It is likely that A0A7L4PCH8 would have a different structure to the other proteins. As the two *Pyrobaculum* S-layer structures determined are similar to each other to the resolution of nanometres, A0A7L4PCH8 is not likely to be the dominant protein in *P. arsenaticum*. Therefore, A0A7L4P6S3 may be the dominant protein in *P. arsenaticum* of the two putatively annotated in UniProt. The sequence H6Q8S4

from *P. oguniense* is 97% similar to A0A7L4P6S3, and as the two species are so closely related, it may be inferred that this could be the dominant protein in *P. oguniense* (Bernick *et al.*, 2012). These conclusions must be heavily caveated by the lack of experimental evidence, and it is possible that the original assumption that these are the only S-layer proteins in the genomes of the two species is incorrect.

### 5.4.3 Using MD to predict identity of ions in cryo-EM density

The work of Cates *et al.* (2002) simulated the protein parvalbumin and showed coordination sites exchanging $Ca^{2+}$ and $Mg^{2+}$. This was also observed in the simulations of the S-layer of *H. volcanii*, with both cations being able to displace each other in several binding sites.

One of the major limitations of the simulations presented in this work is a consequence of the force field used, CHARMM36m. For computational efficiency, atoms are generally treated as spheres with the same property across the sphere and a hard edge defined by the VdW radius. However, atoms do not actually behave as a sphere following classical mechanics, and $Ca^{2+}$ is a soft metal ion that can be polarised (Lightstone *et al.*, 2005; de Courcy *et al.*, 2010). This property is not captured in the classical force field used by the study, and thus the primary difference between the two divalent cations studied is the radius and consequently the charge density. Therefore, results must be interpreted and conclusions caveated by the $Ca^{2+}$ ions lacking polarisability.

The simulations were able to identify a large number of divalent cation binding sites. While it would be presumptive to conclude that this is an exhaustive list, the method has vastly increased the number of binding sites compared to those identified by von Kügelgen *et al.* (2021). The sites identified are comprised of amino acids which are typically found in $Ca^{2+}$ and $Mg^{2+}$ binding sites in other proteins. Such sites commonly have the same amino acids, and vary primarily in the number of amino acids coordinating the ion and the distance from the ion to

the coordinating group. Thus, this work has found sites which agree with divalent cation sites previously found in literature.

Simulations were also performed as part of the work contributing to von Kügelgen *et al.* (2024), though not as part of this research. The simulations were able to elucidate the role of charge distribution in the movement of ammonium ions across the *N. maritimus* S-layer, and this work was supported by experimental and structural studies. This shows that molecular dynamics simulations can be a valuable part of the studies of S-layers, providing time-resolved data in unique detail. Future experimental studies may be able to confirm the results of the simulations of the *H. volcanii* S-layer.

## 5.5 Conclusions

This chapter presents four novel cryo-electron tomography structures of archaeal S-layers, from *Nitrosopumilus maritimus*, *Saccharolobus solfataricus*, *Pyrobaculum oguniense*, and *Pyrobaculum aresnaticum*. The structures provide insight into the diversity in architecture within this class of proteins across the domain. Additionally, data is presented that enhances the understanding of the role of divalent cations in the S-layer of *Haloferax volcanii* using molecular dynamics simulations.

The species studied are from two kingdoms of archaea, with the novel structures all coming from species in the kingdom Thermoproteati. The *Pyrobaculum* and *Saccharalobus* species all thrive in high temperature environments, and the two genera have different morphologies. This leads to different structural requirements for the S-layer compared to the *Nitrosopumilus* S-layer, as the small marine-dwelling archaeon lives in much colder temperatures and the cell has higher curvatures due to its small size and rod shape. *H. volcanii* is in the kingdom Methanobacteriati.

Cryo-electron tomography and sub-tomogram averaging produced medium-resolution structures of the S-layers studied. While these were not at a sufficient resolution for atomic model building, the structures still allowed for comparison of gross S-layer architecture between species. The 9.9 Å structure of the *S. solfaraticus* S-layer was sufficient to identify major structural differences with the S-layer structure from *S. acidocaldarius*. Also, the 22 Å and 16 Å resolution structures from *P. oguniense* and *P. arsenaticum* respectively were of a sufficient resolution to observe that S-layers have a very similar architecture. For the *N. maritimus* S-layer, the 5.7 Å structure did show the two-fold symmetry of the central pore in hindsight, but solving the high-resolution structure was required to understand it (von Kügelgen *et al.*, 2024).

In many cases, lower-resolution structures can be modelled by using sources such as AlphaFold and approximating the fit. However, without customisation, AlphaFold2 and 3 do not successfully produce lattice models, as hexameric units rely on interactions with other units for their structure. Models of the monomers lack

accuracy in interdomain relationships, and this makes the AF models unreliable as a source of models for fitting into the structures. The work of Johnston *et al.* (2024) may be able to solve this problem, as their approach included information about rotational symmetry to produce more plausible models. However, the individual domains predicted by AF are likely to be accurate due to the prevalence of Ig-like architecture and the strength of AF in reproducing folds.

Bioinformatic analysis, using BLAST and multiple sequence alignment, was performed to help explain the differences and similarities in structure linked to the protein sequences. This analysis revealed that, in the two *Pyrobaculum* species, there are two putative proteins that could form the S-layer (as there are in *N. maritimus*, and not in the form of a two-component S-layer as in *Saccharolobus*), and one gene had 97% homology between the two species. One of the proteins in *P. arsenaticum* was not homologous to the other proteins, and visually had 7 domains in the AF structure. All of the other proteins had 6 domains. From this information, it was tentatively concluded that the proteins with highest homology were likely to be the dominant proteins in the S-layer.

For *S. solfataricus*, the lack of homology in the SlaA sequence when compared to that from *S. acidocaldarius* explained the lack of structural similarity. The arms extending from the central pore share no significant sequence homology between the two species, suggesting that it is likely the structure and architecture of these sections are different. This was seen in the alignment of the two SlaA structures.

The work on comparing the two *Pyrobaculum* S-layers, and the two *Saccharolobus* S-layers to each other, demonstrated the variation in sequence and structure similarity between S-layers species in the same genus. This agrees with, and adds more data points to, the findings of Bharat *et al.* (2021). Additionally, while all of the S-layers studied have structures dominated by $\beta$-strand-based folds, the overall architecture of the S-layers have significant differences.

Finally, this work used molecular dynamics simulations to build upon the published structure of the *H. volcanii* S-layer from von Kügelgen *et al.* (2021) with divalent cations. Although limited by the classical dynamics force field, the

simulations were able to independently identify coordination sites observed in the structure, and through calculations of occupancy across the simulations make predictions on the preference of the sites for $Ca^{2+}$ or $Mg^{2+}$. These results were able to locate more than 70 sites per subunit, all of which had visible ion-like density in the cryo-EM map.

Overall, this work combined structural, bioinformatic, and computational approaches to give novel insights into the structures of five archaeal S-layers.

## 5.6 Future work

To understand in more detail the differences between the *S. solfataricus* and *S. acidocaldarius* structures, the atomic structure of the *S. solfataricus* S-layer should be solved. This could be achieved by taking cryo-EM data of S-layer sheets, which will also eliminate the problem of varying curvature between the pleomorphic vesicles. After a model is built, a more accurate comparison can be made between the two structures, including identifying if the homologous domains are arranged in the same way.

There are two big questions arising from the *Pyrobaculum* structures. The first is the identity of the protein forming the S-layer, as there are two proteins annotated in the genome in both species. This can be identified either by techniques such as mass spectrometry or by solving the high-resolution structure. Some work has already been started by the Bharat group on determining a single-particle structure of the *P. arsenaticum* S-layer from sheets, however this has not yet been able to achieve high resolution. The second question is discovering the identity of the substance forming the density between the membrane and the S-layer, as is observed in both S-layer structures. This may be an archaeal peptidoglycan such as pseudomurein, or may be another polymer. The interactions of this substance with the S-layer could also subsequently be investigated.

Experimental work to support the results of the *H. volcanii* S-layer simulations could involve testing S-layer stability in a range of ratios of added calcium and magnesium salts, and comparing this to the ratio of $Ca^{2+}$ to $Mg^{2+}$ sites identified. Also, new methods such as electron energy-loss spectroscopy (EELS) may be able to determine the metal ion identities (Pfeil-Gardiner *et al.*, 2024). Further simulations could also be performed to investigate the energetics of the binding sites, and see if one ion gets displaced by one of a different identity and at what rate.

Investigation into how various S-layers adopt a variety of curvatures could be investigated using simulations. This would require larger simulations of multiple hexamers, however it would allow inter- and intra-hexamer and monomer flexibility

to be identified. Force could be applied to simulate positive and negative curvature of the membrane, to study how S-layers adopt and adapt to changes in cell shape. This could have applications in biotechnology as a sturdy, flexible nanomaterial, as well as furthering understanding into how S-layers are involved in cell division.

Further investigations into S-layers in general may aid in the understanding of how extremophiles survive in challenging environments such as high temperatures and high salt concentrations.

# 6
# Discussion and Outlook

Within this thesis, I have presented data showcasing a range of applications for structural and computational biochemistry. By integrating complementary data from different techniques, I have shown that the interactions within complex biological complexes can be elucidated on an atomic scale. I have also shown that high-resolution structures can be complimented by lower-resolution structures to give biological context and direct further experiments.

## 6.1 Molecular dynamics

Since the logical step between determining the structure and understanding the function of a biological molecule is to study how it moves, molecular dynamics is an ideally-placed method for integrative structural biology studies (Grandori, 2023). MD is also not limited to post-structural research due to the ability to perform *in silico* mutations, and the harnessing of machine learning methods such as AlphaFold, which can provide plausible models before the experimental work has been completed (Akdel *et al.*, 2022).

The work in Chapter 3 shows that screening of mutations for impacts on function can be done computationally. The translocation of double-stranded RNA through the protein nanopore FhuA is impacted by the distribution of charge through the lumen of the pore (subsection 3.3.4), and analysis showed that this is partly due to the change in the flow of ion species through the pore (subsubsection 3.3.4.4). This atomic understanding of the process allows for more thoughtful design of nanopores for RNA translocation when performed *in vitro*, making the experimental design process more data-driven and therefore efficient (subsection 3.4.5).

Both Chapter 3 and Chapter 5, subsection 5.3.4, focus on the charge-based interactions driving the processes of translocation and S-layer stability, respectively. In the latter, the divalent cations $Ca^{2+}$ and $Mg^{2+}$ are known experimentally to be vital to the S-layer structure, however the exact interactions were not able to be determined solely from the cryo-electron microscopy structure in von Kügelgen *et al.* (2021). This is because it is not possible to determine the identity of atoms in cryo-EM structures, and thus ion densities for both $Ca^{2+}$ and $Mg^{2+}$ appear identical. MD was therefore a suitable method for solving this problem. The simulations were able to independently identify divalent cation binding sites observed in the structure, and then provide likely identities for the species occupying each (subsubsection 5.3.4.4 and subsubsection 5.3.4.9).

A limitation of the classical molecular dynamics simulations performed in the work presented here is the lack of consideration for non-uniform, non-static properties of ions. In particular, the properties of $Ca^{2+}$ are not well-described due to the lack of atom polarisability. However, simulations of $Ca^{2+}$ ions in S-layers have been used previously and the results were in agreement with experimental data, suggesting that this approach is valuable when used in conjunction with other data sources (Herdman *et al.*, 2022).

All of the molecular dynamics research presented here uses the all-atom force field CHARMM36m, which provides atomic detail at the expense of accessing longer simulation times. By using coarse-grained force fields, it is possible to extend the simulation time, thus potentially allowing simulations to sample longer-scale events. Some preliminary simulations were prepared in coarse-grain for the research in Chapter 3, however the applied electric field caused significant deformations in the coarse-grain membrane and this approach was not further pursued (see subsubsection 3.3.6.2 and Appendix Section A.11). Simulations of parts of the ribosome model in Chapter 4 were also prepared in coarse-grain, focusing on the 5S ribonucleoparticle (see Figure 4.15). These were unsuccessful due to the nucleic acid structure being unable to be equilibrated. This highlights the current problems in nucleic acid representations in the coarse-grain force field MARTINI (Marrink

*et al.*, 2007; de Jong *et al.*, 2013; Uusitalo *et al.*, 2017). Another method to extend the simulations in Chapter 3 would be through the use of a different force field, such as the AMBER force field. RNA is more stable in the AMBER force field, however it is still limited to the range of hundreds of nanoseconds. This would allow longer simulations to be run, albeit with a higher computational cost.

## 6.2 Cryo-electron microscopy

The improvements in software and hardware for cryo-EM have enabled it to become a much more popular tool for solving the structures of biological macromolecules (Kühlbrandt, 2014). In the research presented here, both medium- and high-resolution structures were solved, allowing valuable insights into S-layer and ribosome structures from a range of archaeal species.

The development of tools such as ModelAngelo has increased the speed of model building into high resolution structures, which was previously a time-consuming process in structure determination (Jamali *et al.*, 2024). The use of machine learning allows more time to be spent on structural analysis after refining the model, and interpreting the functional implications of the structural features. Further experiments can then be performed to probe the structure-function relationship. An alpha version of ModelAngelo was employed alongside model building into the *N. maritimus* ribosome, showing that the software could tackle large complexes of both protein and nucleic acids, at resolutions better than 3.5 Å (Chapter 4, subsubsection 4.4.3.2). This exceeds what is possible for a researcher to identify by eye, and is also far faster than manual model building techniques. The impact of this is that researchers can allocate their time to other avenues and generate complementary data to give greater structural understanding, and it allows model building to become more accessible to inexperienced users thus allowing them to integrate structural biology into their studies.

Cryo-electron tomography allows structures to be solved *in situ*. This gives cellular context to the structures, potentially showing interacting partners and subcellular location, and allowing better interpretation of function. Tomography

was used to solve *in situ* structures of archaeal S-layers in Chapter 5. While the resolution achieved was not high, the structures did provide additional information to what might be determined using single-particle methods. This includes the discovery of an extracellular polymeric substance in *P. oguniense* and *P. arsenaticum* between the S-layer lattice and the membrane (Figure 5.19 and Figure 5.21), and identification of pentameric defects in the S-layer coating the *S. solfataricus* vesicles (Figure 5.11). Also, the tomograms of *N. maritimus* cells have visible ribosomes (Chapter 4, subsubsection 4.1.4.2 and Figure 4.5), which can be classified into membrane-bound, polysomes, and cytoplasmic. This data could be used in the future to determine if there are structural differences between these classes of ribosomes.

While single-particle approaches to structural determination have often produced only one average structure, many cryo-EM samples contain particles in a range of conformations. This leads to structural heterogeneity, which can decrease the local resolution of structures in areas of variability. In many cases, this heterogeneity is around the areas important in protein function, which means that the global resolution may not actually be good enough for functional interpretation. This can lead to overfitting and over-interpretation of a model (Gilles and Singer, 2025). However, with improvements in methods in separating particles by heterogeneity and interpolating a range of structures, the presence of multiple conformations in a sample will lead to an increase in the information determined from such datasets (Zhong *et al.*, 2021; Kimanius and Schwab, 2024; Gilles and Singer, 2025).

It is not possible to determine the identity of atoms in cryo-EM density maps. For example, the identity of ions coordinated in the *H. volcanii* S-layer structure could not be determined from the density map alone, and two species of divalent cations were known to be essential to the structure. Complementary methods were used in Chapter 5, subsection 5.3.4, to give insight to their identities.

## 6.3 Bioinformatics tools

Structure determination can be complemented by the judicious use of bioinformatics tools (Gutmanas *et al.*, 2013; Paiva *et al.*, 2022; Soares and Lousa, 2025). This

is particularly the case where the structures being solved are not from well-studied species, such as the *N. maritimus* ribosome in Chapter 4 or the S-layers studied in Chapter 5.

There are currently a limited number of published archaeal ribosome structures. Therefore the prior information available for model building the *N. maritimus* ribosome in Chapter 4 had to be used selectively. BLAST comparisons of protein sequences allowed for the most appropriate models to be chosen for building and analysing the structures of ribosomal proteins. Previous bioinformatic work focusing on the genetic relationships between r-proteins in different domains of life allowed homologous genes to be identified, meaning the process of identifying which proteins were present in the genome was much more efficient (e.g. subsubsection 4.3.3.11, subsubsection 4.3.3.13). AlphaFold could then be used to predict models of these proteins, providing a high-quality starting point for model building into the structure (subsection 4.3.3 and subsection 4.3.4). This approach could also be applied to the rRNA, as large parts of the rRNA sequences and therefore structures are highly conserved throughout all ribosomes. However, where the structure was not known, model prediction was less successful (subsection 4.3.5). This is because there is much less data available on RNA structures in general, meaning that machine learning models have less training data and subsequently their outputs are more uncertain.

For the *S. solfataricus* S-layer, sequence comparison with the SlaA protein from *S. acidocaldarius* offered a quick explanation to the difference in structural architecture (Chapter 5, subsubsection 5.3.2.4 and subsection 5.4.2). The protein domains with the most structural identity align with the homologous sections of sequence. Bioinformatics tools were also used to examine the putative *Pyrobaculum* S-layer protein sequences, however the speculative results are of limited use without experimental confirmation (Chapter 5, subsubsection 5.3.3.5 and subsection 5.4.2). There is also extensive bioinformatics analysis underpinning the work by Bharat *et al.* (2021).

The usefulness of bioinformatics tools is related to the confidence in the data source, and the application of the tool (Wankowicz, 2025).

## 6.4 Integrative structural biology

The data from molecular dynamics simulations, cryo-electron micrographs, and bioinformatics tools were synergistic for the structural biology studies detailed in this thesis. For example, in Chapter 4, bioinformatics tools were used to aid the building of the ribosome model into the cryo-EM structure, and were also used subsequently to analyse the r-protein composition of the *Nitrosopumilus maritimus* ribosome compared to other published archaeal ribosomes. The cryo-EM structure and model can then contribute to bioinformatics databases with observed interactions and structural features of the r-proteins. The S-layer structures in Chapter 5 were studied with both molecular dynamics simulations and cryo-EM. MD simulations extended the model of the *Haloferax volcanii* S-layer ((von Kügelgen *et al.*, 2021)) by identifying divalent cation binding sites and suggesting the preference for $Ca^{2+}$ or $Mg^{2+}$. These binding sites agreed with what was observed in the electron density from the cryo-EM structure, and the simulations complemented the cryo-EM work as they could determine the ion species. The comparison of two S-layer cryo-EM structures from *Saccharolobus acidocaldarius* ((Gambelli *et al.*, 2024)) and *Sacchorolobus solfataricus* was aided by bioinformatics tools that identified regions of the protein sequence with significant divergence. Overall, an integrative approach was key to furthering the understanding of the systems studied in this thesis.

## 6.5 Outlook

This thesis has utilised different structural and related methods to gain understanding of the structure and function of several biological complexes. Structural studies are increasingly supported by a range of experimental and computational methods, with the results integrated to give a more rounded understanding of the subject (Rout and Sali, 2019; Koukos and Bonvin, 2020; Schwalbe *et al.*, 2024). With the advent of machine learning methods to speed up parts of the process, such as model building, more time can be devoted to orthogonal studies. This allows researchers to work towards a better understanding of the rules that govern the

relationship between structure and function. An increased interest in the study of archaeal biology is helping to improve the understanding of the evolutionary relationship between the last universal common ancestor and eukaryotes, adding another perspective to the question of how multicellularity evolved (Adam *et al.*, 2017; Dance, 2021). The study of diverse organisms will continue to produce unknown features in structures such as the S-layer from *Pyrobaculum* which then further our understanding of the variety of approaches which have evolved to allow extremophiles to exist in every biological niche on the planet. It is important that the study of these organisms remains holistic in order to best understand the cellular biology, including the structure-function relationship of unique proteins that may also have future biotechnological use. The work presented here contributes to the concerted effort of the scientific community to close the knowledge gaps in areas such as RNA structure, ribosome evolution, and protein stability in extreme conditions.

# Appendices

# A

# Translocation of double-stranded RNA through a nanopore

## Contents


**A.1 Using convex hulls to calculate cross-sectional area . . 263**
**A.2 dsRNA at different concentrations of KCl . . . . . 265**
**A.3 dsRNA hydrogen bonding between base pairs in simulation of solution . . . . . 266**
**A.4 Simulation of other fragments of ribosomal RNA . . . 267**
**A.5 Principal component analysis of dsRNA in water . . . 268**
**A.6 Membrane curvature . . . . . 269**
**A.7 Simulations of translocation without C-$\alpha$ restraints . . 270**
**A.8 Ion conductance through FhuA . . . . . 271**
**A.9 Translocation at different concentrations of KCl . . . . 273**
**A.10 dsRNA H-bonding between base pairs in simulation of translocation . . . . . 274**
**A.11 Coarse-grained simulations . . . . . 275**
**A.12 Principal component analysis of dsRNA in FhuA . . . 277**
**A.13 dsRNA interactions with the membrane . . . . . 277**
**A.14 Ion "helter skelter" . . . . . 279**
**A.15 dsRNA velocity in the pore . . . . . 280**

## A.1 Using convex hulls to calculate cross-sectional area

The Convex Hull algorithm as implemented by SciPy was utilised to more methodically obtain a cross-sectional area of the dsRNA along the length as an extension to the use of HOLE and PoreAnalyser (Barber *et al.*, 1996). Utilising the trajectories of solvent and dsRNA translocating through the pore,the $x$- and $y$-coordinates attained by the dsRNA or water molecules were binned by 1 Å $z$-slices through the pore for each simulation frame. A convex hull was calculated for these points, which was then extended by the largest VdW radius for atoms in the molecule according to the force field. This was then used to give another estimate for cross-sectional area along $z$. It is likely that this is an overestimate as the hull is convex, and the largest radius value is used when only a small number of atoms have this applied. This means that it is difficult to draw conclusions from this data as the calculated results do not match up with what is observed in simulation (Figure A.1).

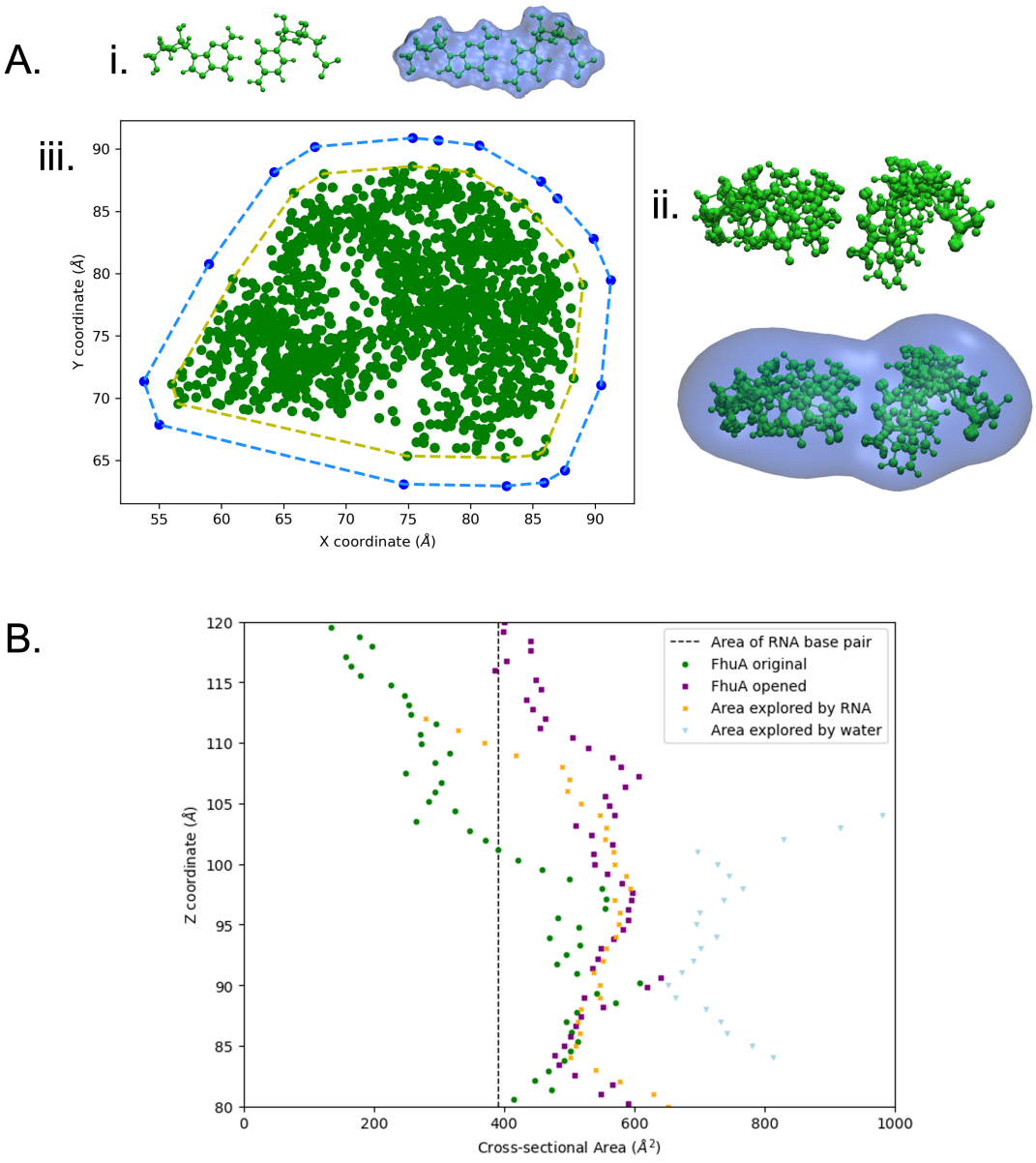



**Figure A.1:** A. i. Ball-and-stick representation of RNA pair, and surface representation of the VdW radii of those atoms. ii. Overlaid ball-and-stick representations of RNA pair over 10 frames, and the cumulative VdW radii of those atoms. iii. Graph of $x$- and $y$-coordinates of an RNA pair over 100 frames in green, the calculated convex hull in lime, and the addition of the VdW radius around the convex hull.
B. Plot of cross-sectional area of FhuA across the $z$-coordinate of the simulation box, and the cross-sectional area being explored by dsRNA through those coordinates.

## A.2 dsRNA at different concentrations of KCl

Simulations of fragment 2 were also run at 0.05 M, 0.3 M, and 0.5 M KCl. No distinct differences in the behaviour or structure of the dsRNA was observed throughout the simulations, other than the number of ions associated with the backbone (Figure A.2).

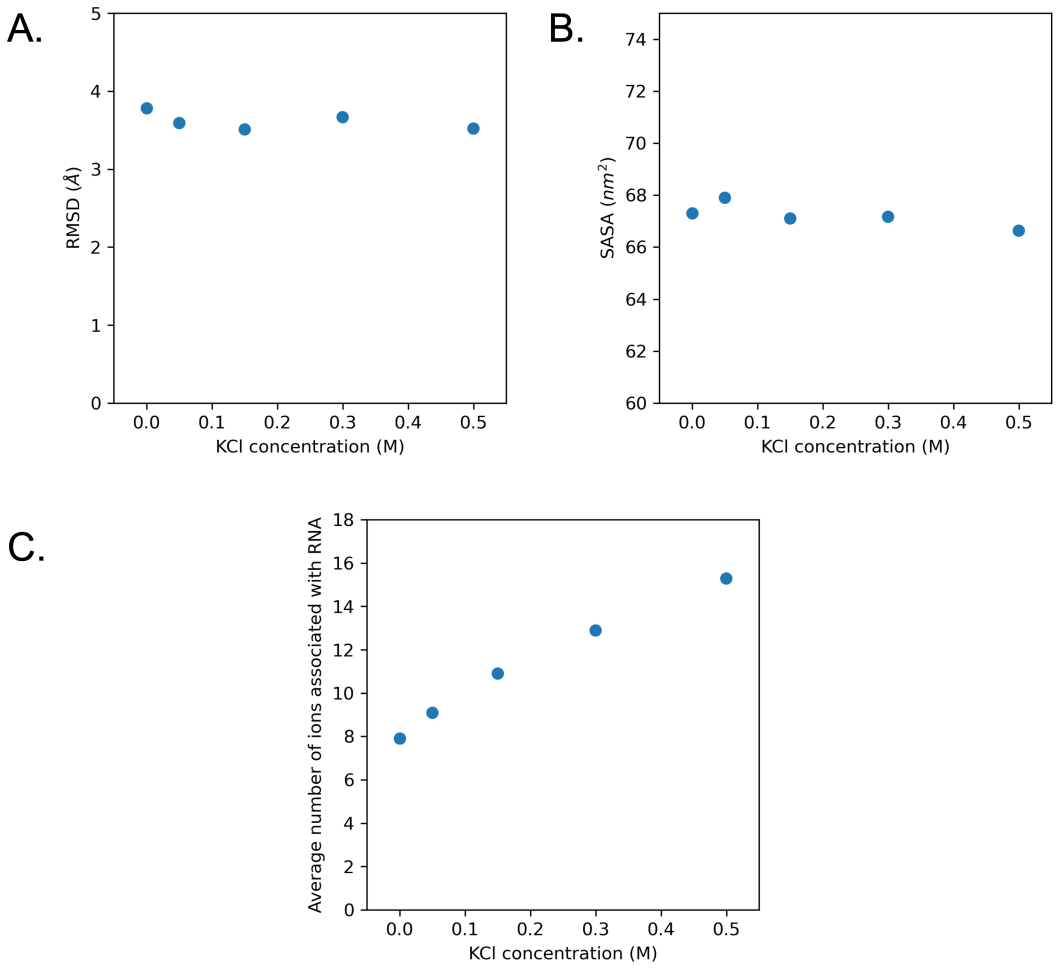



**Figure A.2:** A. Plot of average RMSD for simulations of RNA at varying concentrations of KCl.
B. Plot of average SASA for simulations of RNA at varying concentrations of KCl.
C. Plot of average number of potassium ions associated with RNA at varying concentrations of KCl.

## A.3 dsRNA hydrogen bonding between base pairs in simulation of solution

The hydrogen bonding between base pairs in the dsRNA was tracked across the simulations of dsRNA in water, and the percentage of frames where $> 50\%$ of the hydrogen bonds were maintained was calculated (Figure A.3). Only one pair, the wobble pair G568:U626, had $< 90\%$ maintenance of hydrogen bonding.

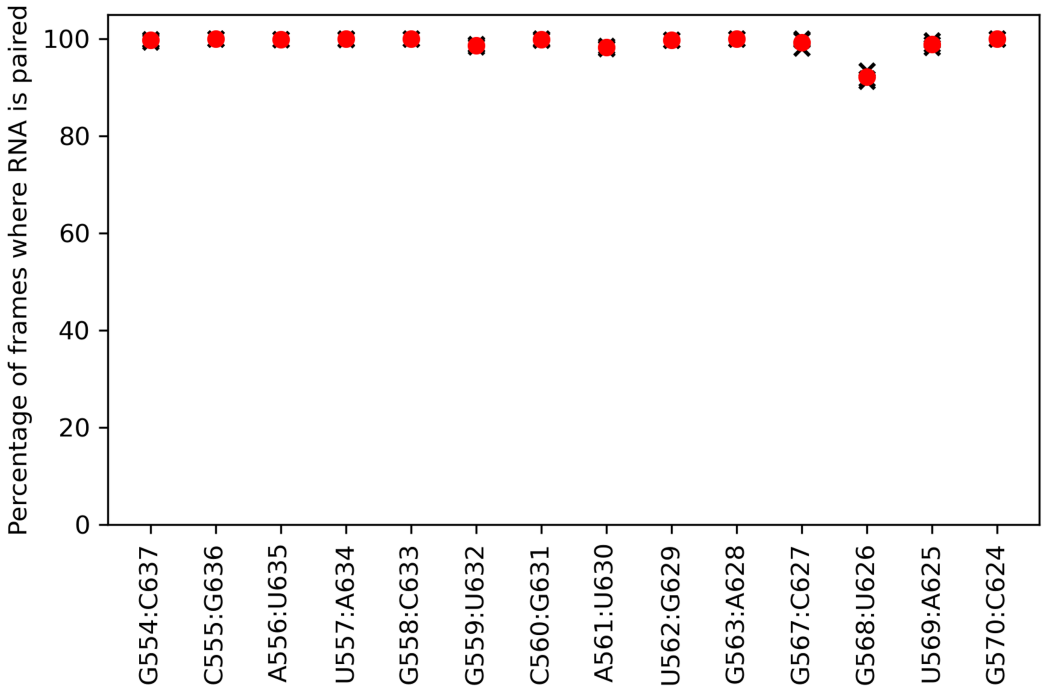



**Figure A.3:** Percentage of frames where RNA base pair maintains $> 50\%$ of the hydrogen bonds, across all the simulations of dsRNA in water.

## A.4 Simulation of other fragments of ribosomal RNA

RNA structures were stable in all conditions tested, and any fragment would have been suitable for testing translocation (Figure A.4).

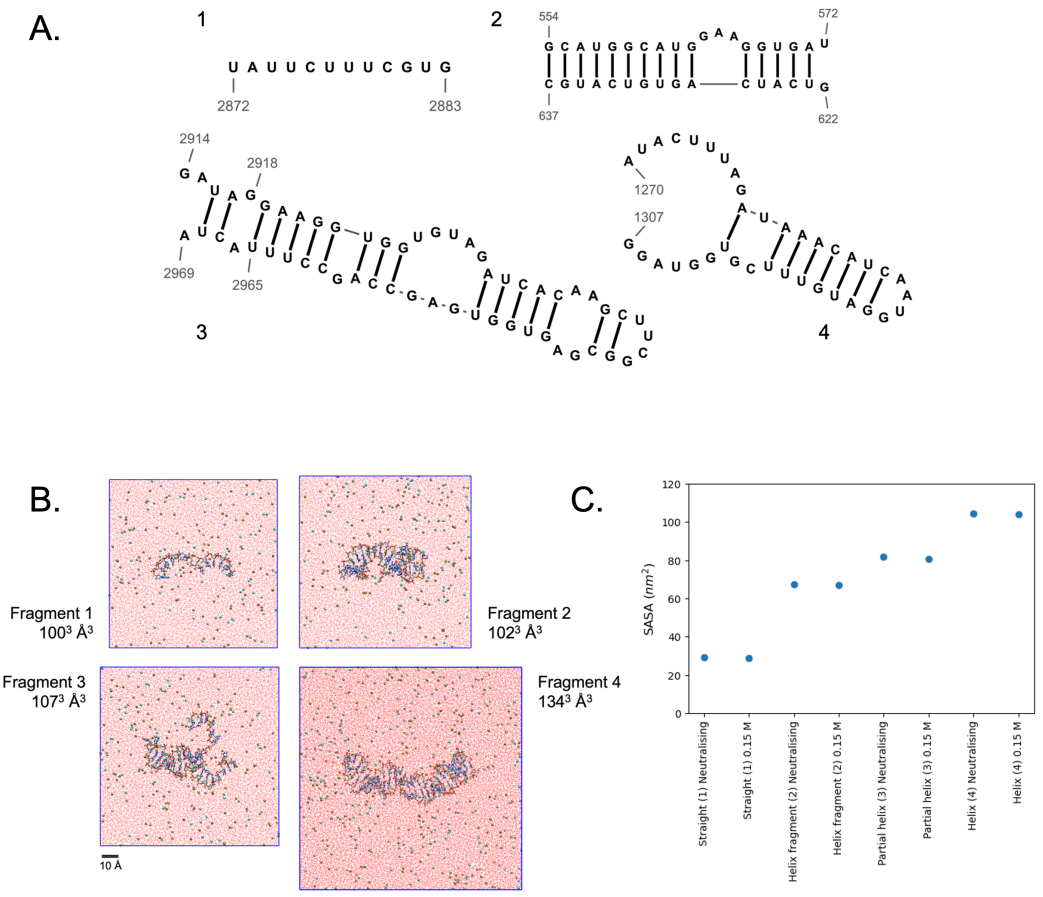



**Figure A.4:** A. The 2D structures of the dsRNA fragments shown in Figure 3.6.
B. Comparison of the simulations of dsRNA fragments in water.
C. Plot of SASA for each simulation, comparing neutralising $K^+$ and 0.15 M KCl.

## A.5 Principal component analysis of dsRNA in water

The cumulative variance for the eigenvectors calculated from performing PCA on simulations of dsRNA in water was plotted in Figure A.5. The first ten eigenvectors covered $> 90\%$ of the variance.

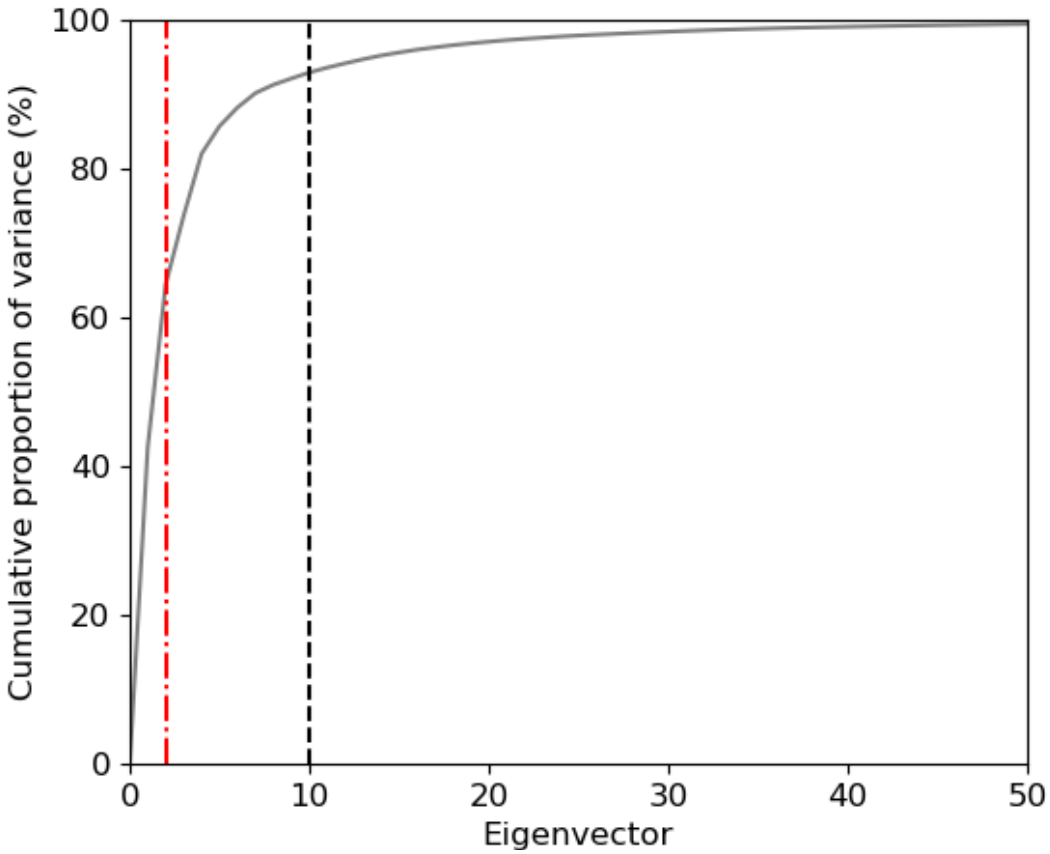



**Figure A.5:** Graph of the cumulative variance covered by the eigenvectors calculated when performing PCA on the dsRNA backbone conformations for simulations of dsRNA in water.

## A.6 Membrane curvature

When an electric field is applied to the system, the overall charge of FhuA leads to the movement of the protein in the $z$-direction, which leads to a measurable membrane curvature. The direction and magnitude of this curvature depends on the size of the electric field applied and on the overall charge of FhuA (Figure A.6).

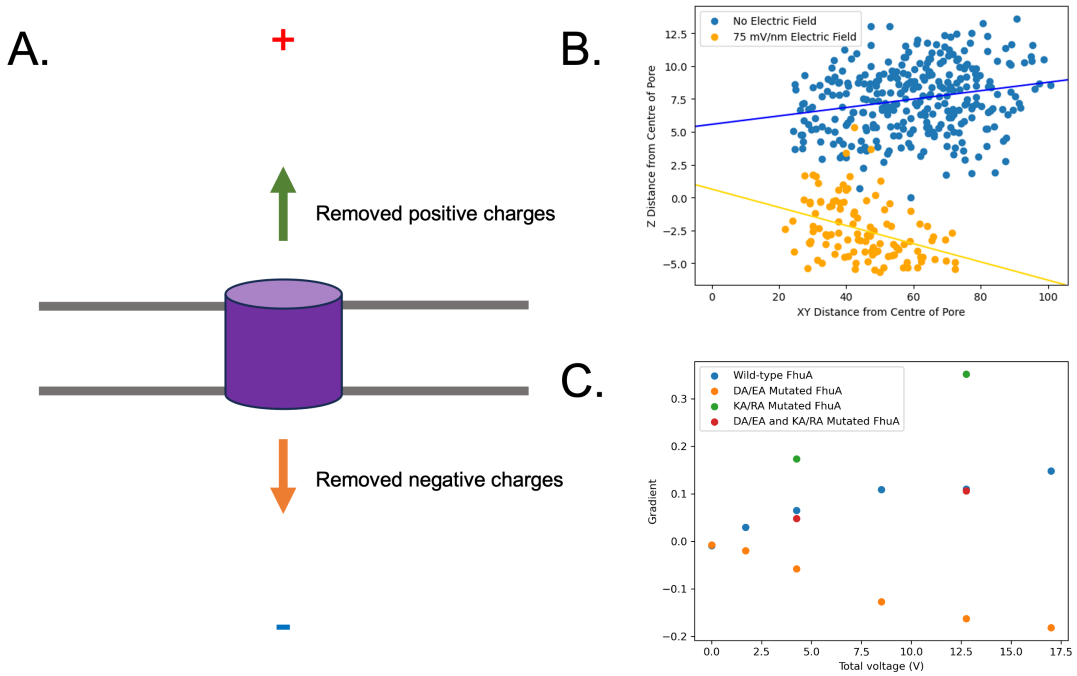



**Figure A.6:** A. Diagram of the effect of mutations to FhuA on the direction of membrane curvature.
B. Plot of lipid headgroup distances from the centre of the pore in $xy$ and $z$-axes.
C. Plot of curvature, calculated from the gradient of regression lines from B., against applied electric field.

## A.7 Simulations of translocation without C-$\alpha$ restraints

Simulations were performed of dsRNA translocation through WT and DA/EA mutant FhuA, without position restraints on the C-$\alpha$ atoms (see Table A.1). Overview of these results is presented in the graphs in Figure A.7.

**Table A.1:** Simulation set-ups for simulations of FhuA without dsRNA present.

| **Pore** | **Restraints** | **Replicas** | **Electric field (V nm$^{-1}$)** | **Box $z$-dimension (nm)** | **Electric field (total, V)** |
|---|---|---|---|---|---|
| WT | none | 3 | −0.05 | 19 | −0.95 |
| DA/EA | none | 3 | −0.05 | 19 | −0.95 |

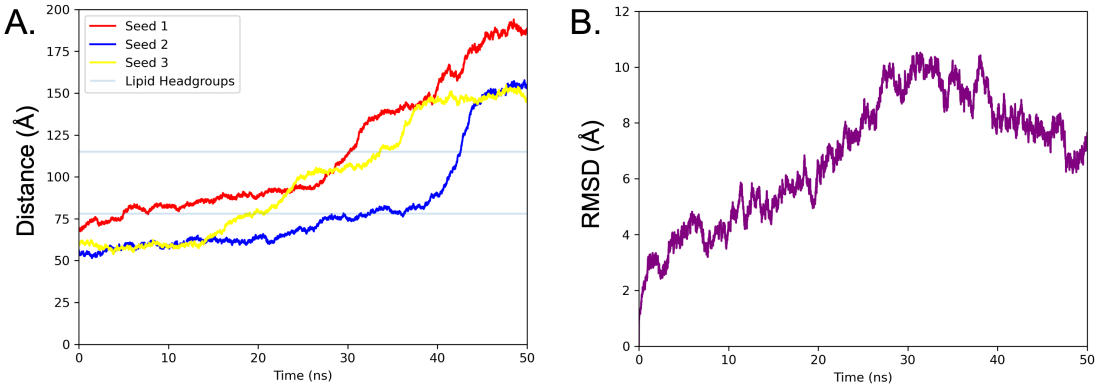



**Figure A.7:** A. Plot of RNA centre-of-mass over time for three simulations of dsRNA translocating through unrestrained, WT FhuA with 0.15 M KCl and −0.05 V nm $^{-1}$ applied electric field.
B. Plot of RMSD of FhuA over the course of simulation seed 3 shown in A.

## A.8 Ion conductance through FhuA

Plots of potassium and chloride ion conductance, and of the median number of frames needed for ions to cross the membrane, which were then combined to calculate ion speed in Figure 3.16 (Figure A.8).

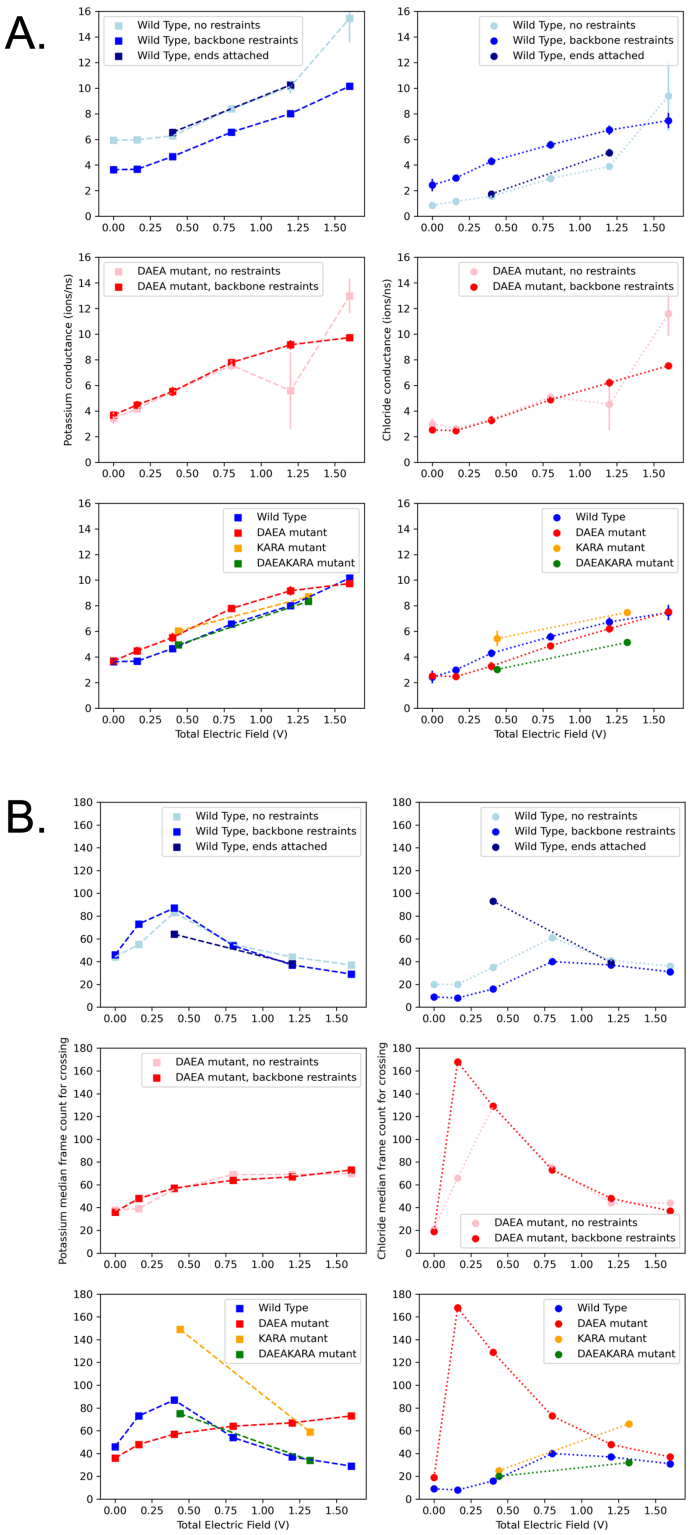



**Figure A.8:** A. Potassium and chloride ion conductance.
B. Median number of frames taken for potassium and chloride ions to cross the membrane.

## A.9 Translocation at different concentrations of KCl

Initial simulations of dsRNA translocating through the unrestrained pore were performed at 0.05 M, 0.3 M, and 0.5 M potassium chloride (Figure A.9). Translocation occurred in all cases where a 0.05 V $nm^{-1}$ electric field was applied (Table A.2). Additionally, simulations of dsRNA translocating through the unrestrained pore with 0.3 M KCl were performed with an electric field of 0.02 V $nm^{-1}$, with no cases of the dsRNA translocating fully.

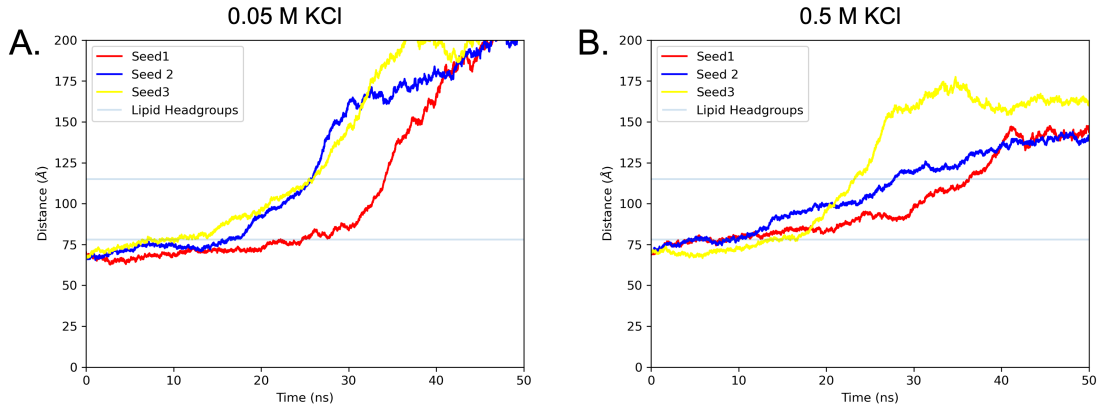



**Figure A.9:** A. $z$-coordinate of the RNA centre-of-mass over time, showing translocation through unrestrained FhuA with 0.05 M KCl and at 0.05 V $nm^{-1}$ applied electric field. A. $z$-coordinate of the RNA centre-of-mass over time, showing translocation through unrestrained FhuA with 0.5 M KCl and at 0.05 V $nm^{-1}$ applied electric field.

**Table A.2:** Time taken, in nanoseconds, for the dsRNA to fully translocate through the pore at a range of KCl concentrations.

| **Pore** | **Replicas** | **KCl concentration (M)** | **Number of translocations** | **Crossing times (ns) and *average*** |
|---|---|---|---|---|
| WT | 3 | 0.00 | 3 | 22.5, 48.3, 39.1<br>*36.6 ± 10.7* |
| WT | 3 | 0.05 | 3 | 34.1, 25.7, 25.4<br>*28.4 ± 4.0* |
| WT | 3 | 0.30 | 3 | 20.6, 34.0, 23.8<br>*26.1 ± 5.7* |
| WT | 3 | 0.50 | 3 | 36.3, 27.7, 23.2<br>*29.1 ± 5.4* |

## A.10 dsRNA H-bonding between base pairs in simulation of translocation

The hydrogen bonding between base pairs in the dsRNA was tracked across the simulations of dsRNA translocating through each FhuA, and the percentage of frames where > 50% of the hydrogen bonds were maintained was calculated (Figure A.10).

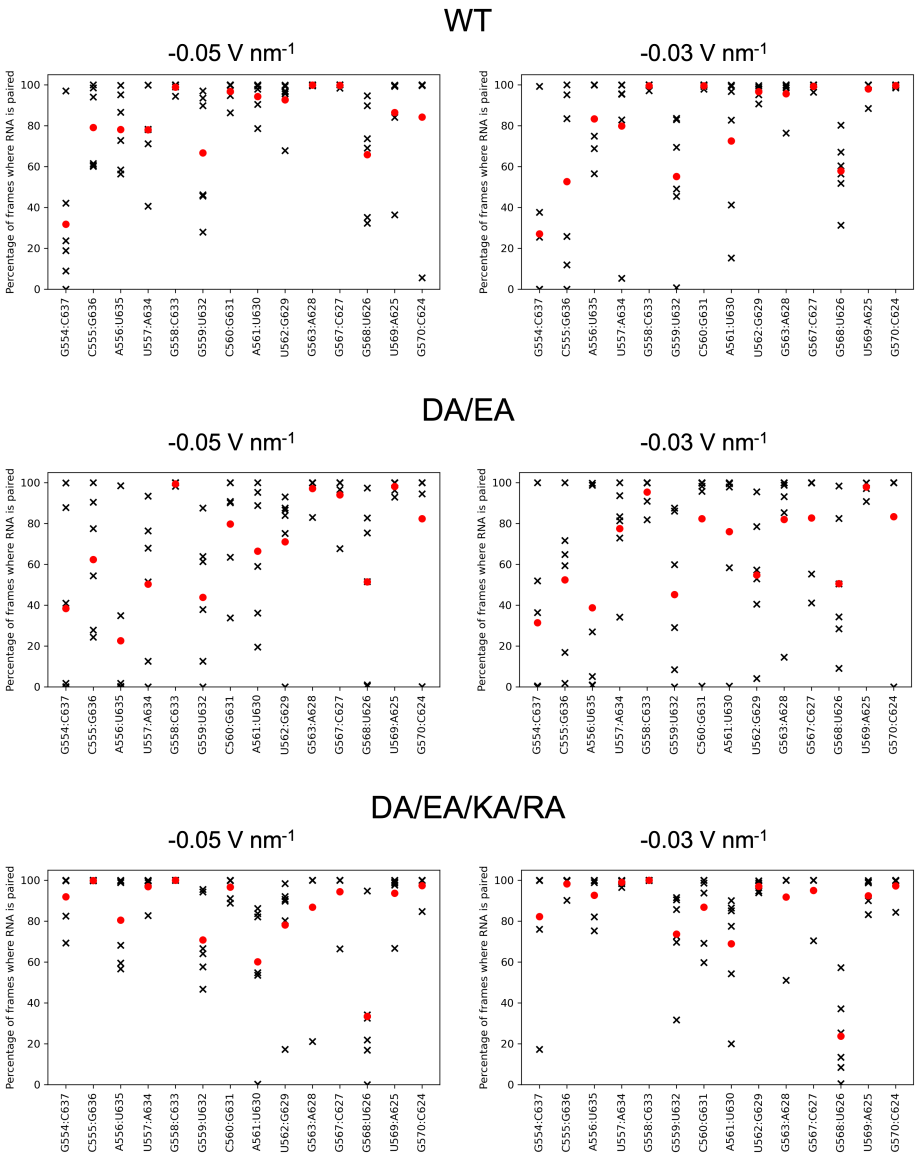



**Figure A.10:** Percentage of frames where RNA base pair maintains > 50% of the hydrogen bonds, across all the simulations of dsRNA translocating through FhuA.

## A.11 Coarse-grained simulations

A few simulations were run using MARTINI 2.2, a coarse-grained force field (Marrink *et al.*, 2007; de Jong *et al.*, 2013; Uusitalo *et al.*, 2017). In order to successfully run the simulations, restraints had to be applied to the lipid headgroups to prevent the membrane from being torn apart even under very low electric fields. This meant that ions could travel through the membrane throughout the simulation. In all simulations, the dsRNA translocated. In the DA/EA/KA/RA mutant, several translocations were observed with the dsRNA spontaneously re-entering the pore, until it approached the pore perpendicularly and was trapped (Figure A.11).

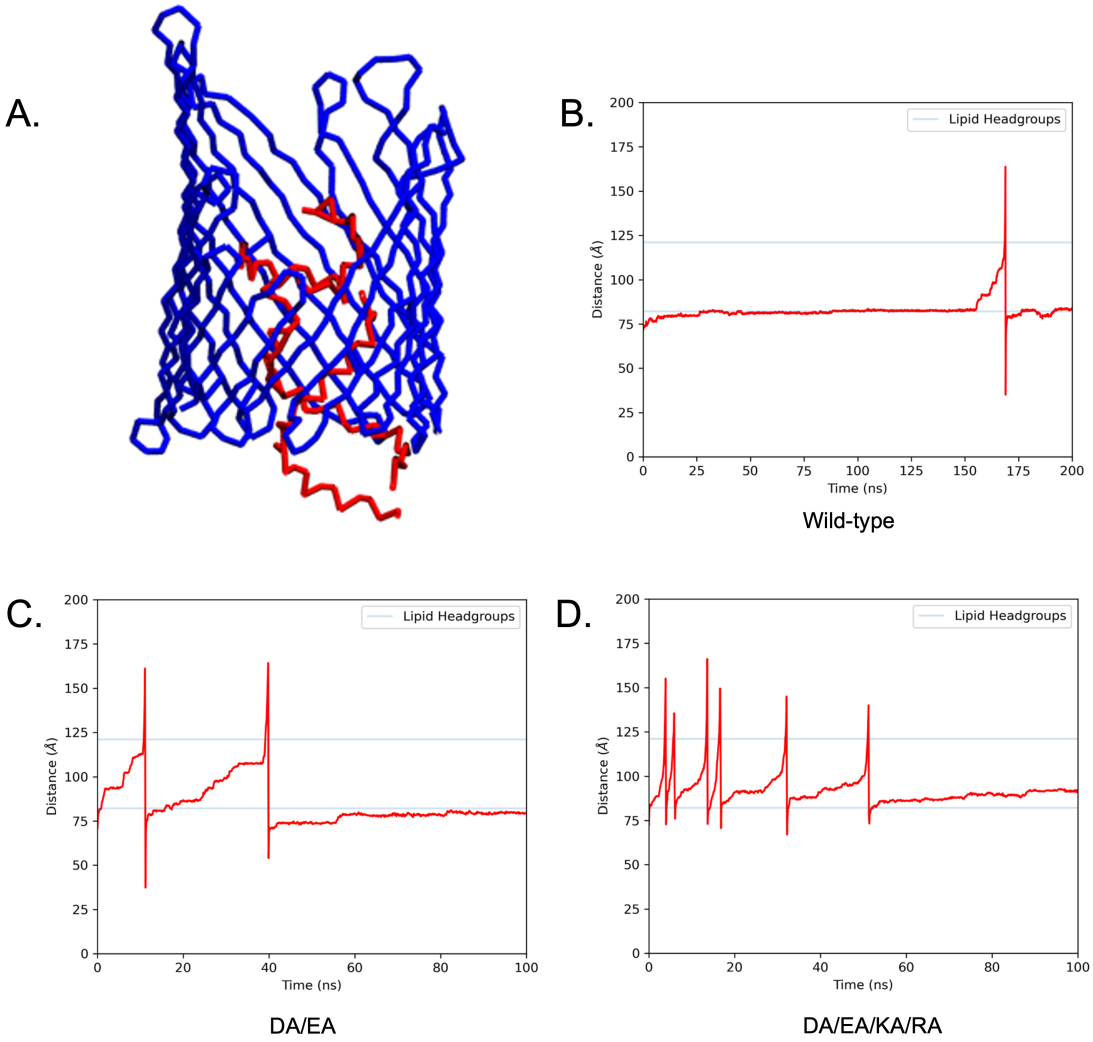



**Figure A.11:** A. Representation of backbone beads of dsRNA and FhuA.
B. Position in $z$ of the centre-of-mass of the dsRNA molecule over time in simulation with WT FhuA.
C. Position in $z$ of the centre-of-mass of the dsRNA molecule over time in simulation with DA/EA FhuA.
D. Position in $z$ of the centre-of-mass of the dsRNA molecule over time in simulation with DA/EA/KA/RA FhuA.

## A.12 Principal component analysis of dsRNA in FhuA

The cumulative variance for the eigenvectors calculated from performing PCA on simulations of dsRNA in FhuA was plotted in Figure A.12. The first ten eigenvectors covered $< 65\%$ of the variance.

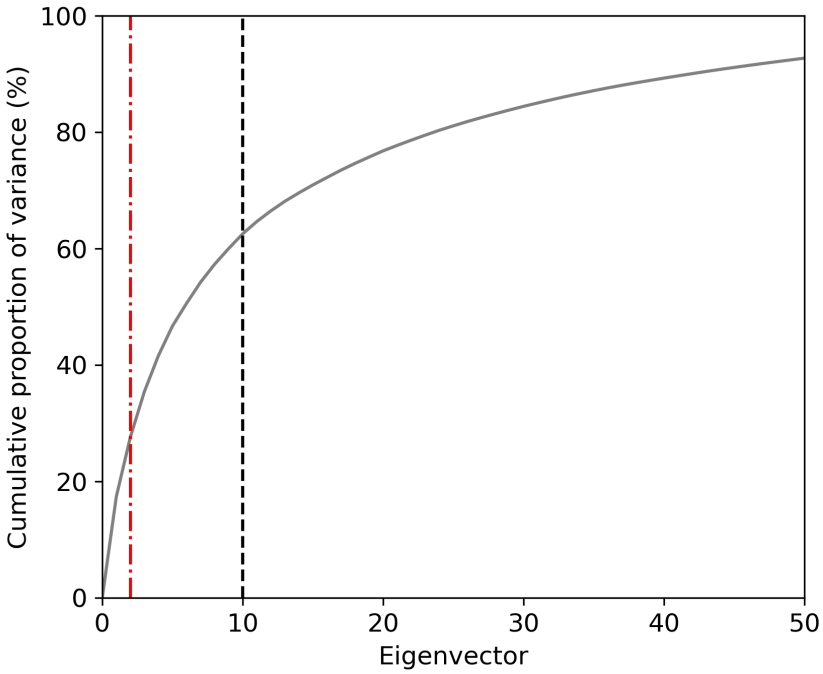
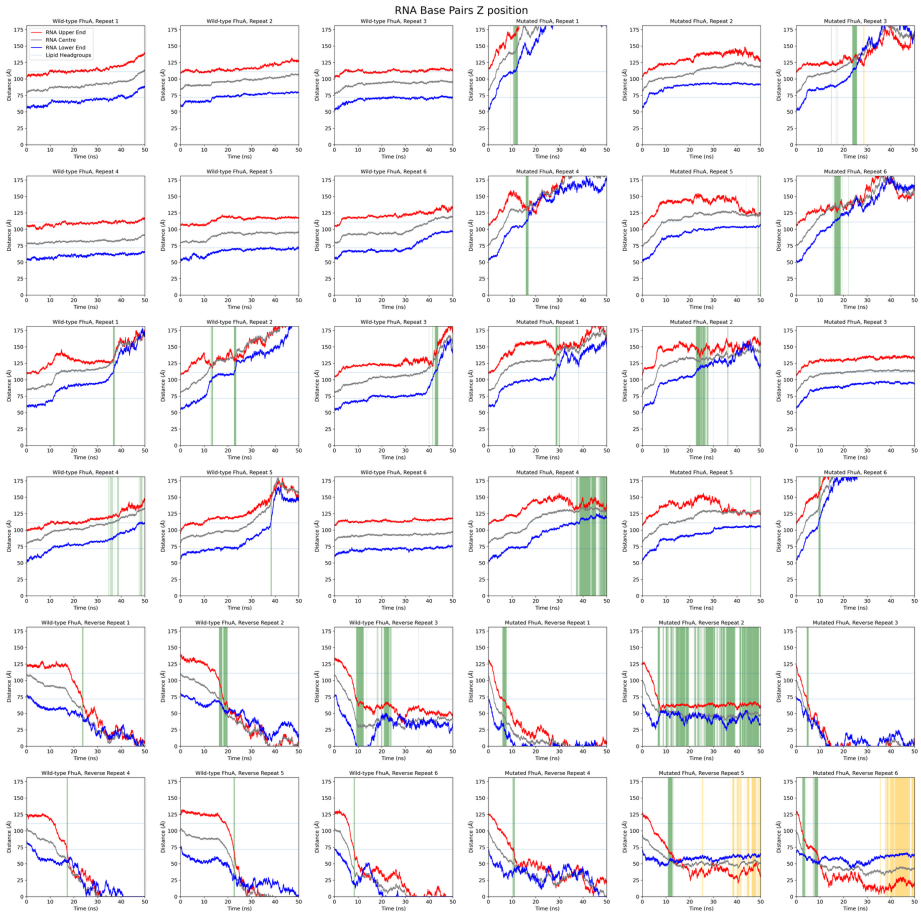



**Figure A.12:** Graph of the cumulative variance covered by the eigenvectors calculated when performing PCA on the dsRNA backbone conformations for simulations of dsRNA in FhuA.

## A.13 dsRNA interactions with the membrane

In some simulations, the dsRNA exited the pore and interacted with the headgroups of the POPC lipids, maintaining this interaction for an extended period (Figure A.13). This impacted the ability of the remainder of the dsRNA to translocate.

A.




B.

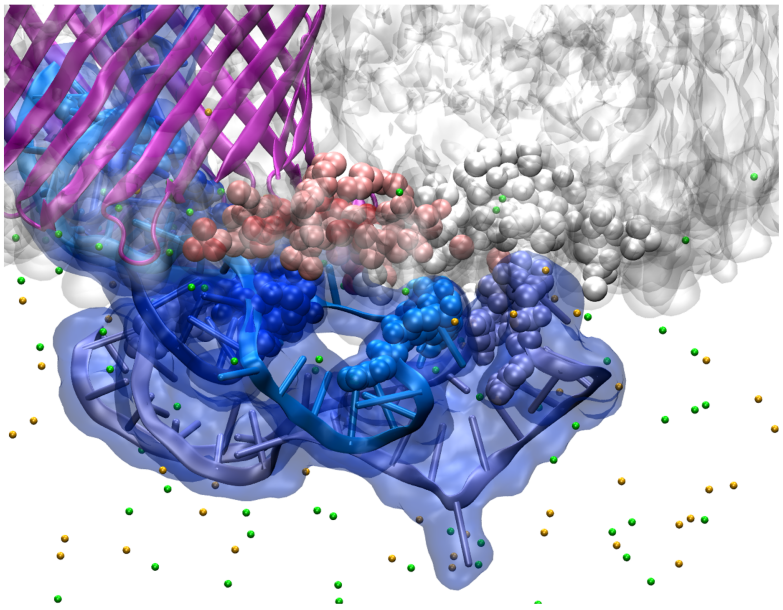


**Figure A.13:** *(Caption on next page.)*

**Figure A.13:** A. Position of RNA pair plotted against time. Red: $z$-coordinate of the centre-of-mass of the upper RNA pair; Grey: $z$-coordinate of the centre-of-mass of the dsRNA molecule; Blue: $z$-coordinate of the centre-of-mass of the lower RNA pair. Green vertical lines represent frames where the last RNA pair to exit the pore is in contact with the POPC membrane after translocation. Yellow vertical lines represent frames where the first RNA pair to exit the pore is in contact with the POPC membrane, causing an inversion of the RNA.
B. Overlay of several frames, showing that the first pair of RNA bases that exits the pore contacts the POPC membrane, and then maintains those interactions. First POPC molecules interacted with shown in red; then pink; final frame in white.

## A.14 Ion "helter skelter"

Potassium ions interact with consecutive phosphate groups as the ions pass through the pore. This happens both when the dsRNA is stationary, and when it is translocating (Figure A.14). Visualising this produces a path resembling a helter skelter.

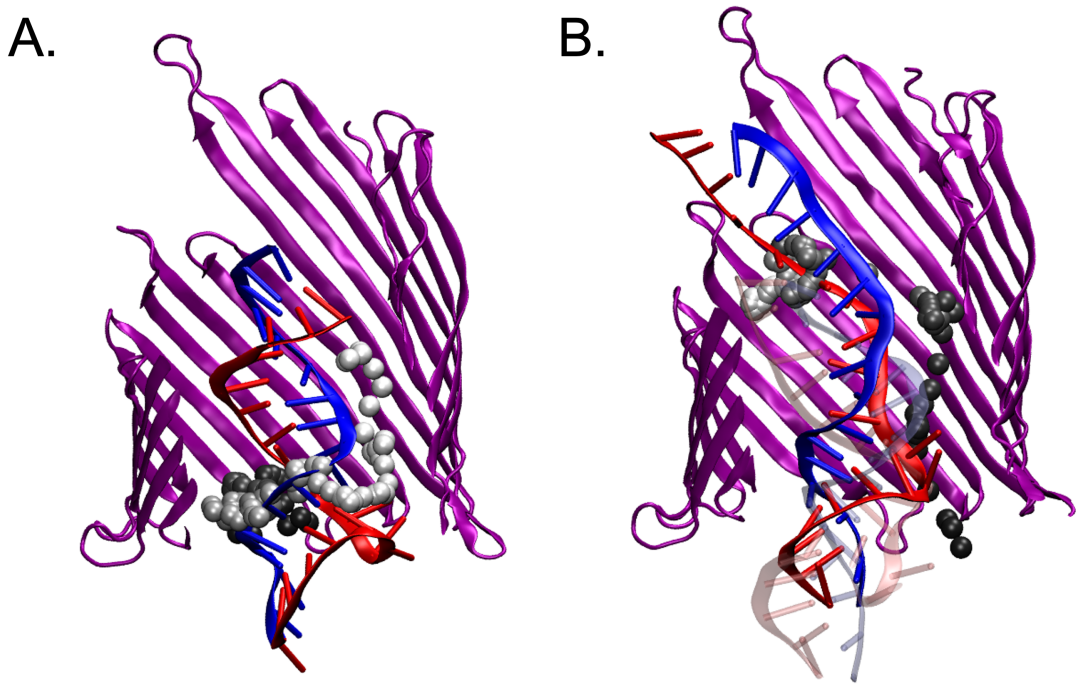



**Figure A.14:** A. Path of one potassium ion (first frame in white, final frame in black) following the backbone of a stationary dsRNA molecule in a FhuA pore.
B. Path of one potassium ion (first frame in white, final frame in black) following the backbone of a dsRNA molecule translocating in the opposite direction through a FhuA pore.

## A.15 dsRNA velocity in the pore

In an effort to characterise the movement of the dsRNA through the pore, the membrane-spanning section of FhuA was divided into five equally-sized segments along the $z$-axis. The time taken for the centre-of-mass of dsRNA to pass through each segment was calculated and plotted, to investigate if there was a particular segment through which the dsRNA moved particularly quickly or slowly (Figure A.15). No discernible pattern was identified.

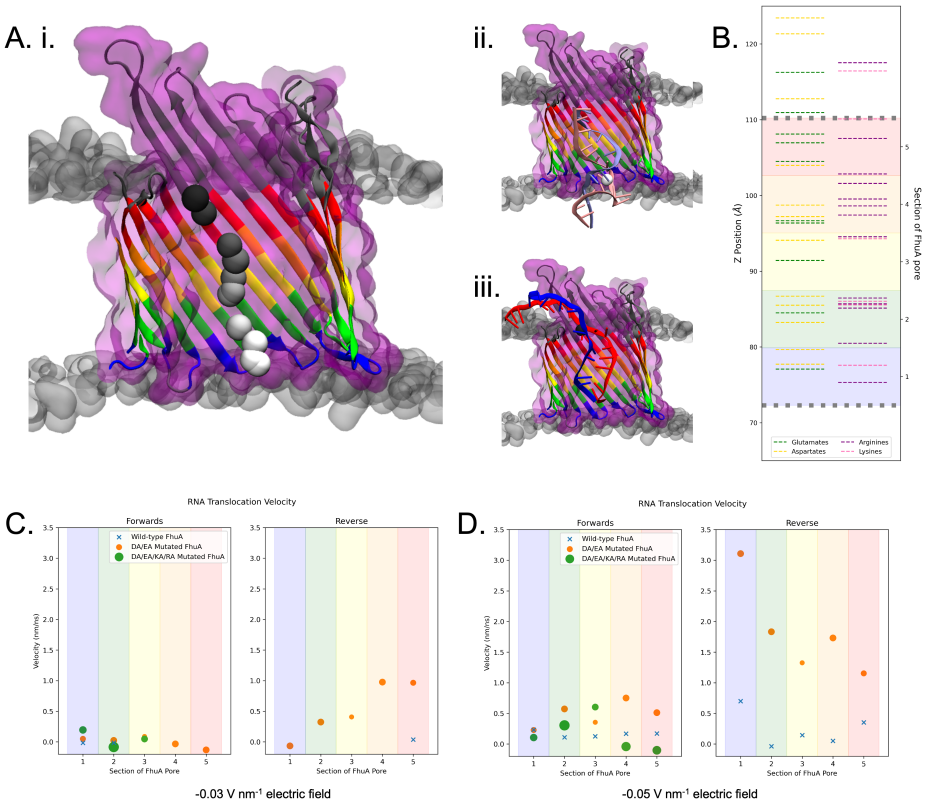



**Figure A.15:** A. FhuA coloured by $z$ segment. Sphere represents centre-of-mass of the dsRNA molecule over 10 frames, starting with the white sphere and ending with the black sphere (i.). ii. The representative RNA structure at the first frame. iii. The representative RNA structure at the last frame.
B. Positions of charged residues in $z$, showing number in each horizontal segment.
C. Translocation speed of centre-of-mass of RNA, divided by horizontal segment, in $-0.03$ V nm$^{-1}$ electric field. The size of the point is relative to the number of mutations made in that segment from the WT FhuA.
D. Translocation speed of centre-of-mass of RNA, divided by horizontal segment, in $-0.05$ V nm$^{-1}$ electric field. The size of the point is relative to the number of mutations made in that segment from the WT FhuA.

# B

# The ribosome of *Nitrosopumilus maritimus*

## Contents


**B.1 Ribosome particle picking . . . . . 282**
**B.2 Ribosome map and model data . . . . . 283**
**B.3 50S map FSC . . . . . 283**
**B.4 50S model-in-map FSC . . . . . 284**
**B.5 70S map FSC . . . . . 285**
**B.6 70S model-in-map FSC . . . . . 286**
**B.7 LSU Model validation . . . . . 287**
**B.8 Model validation of the SSU . . . . . 288**

## B.1 Ribosome particle picking

Particles were initially picked manually, from micrographs at a range of defoci. These particles were then passed to Topaz, and after training, particles were picked automatically. RELION was then used to classify the ribosome particles in 2D. Images showing manual picking and 2D classification are shown in Figure B.1.

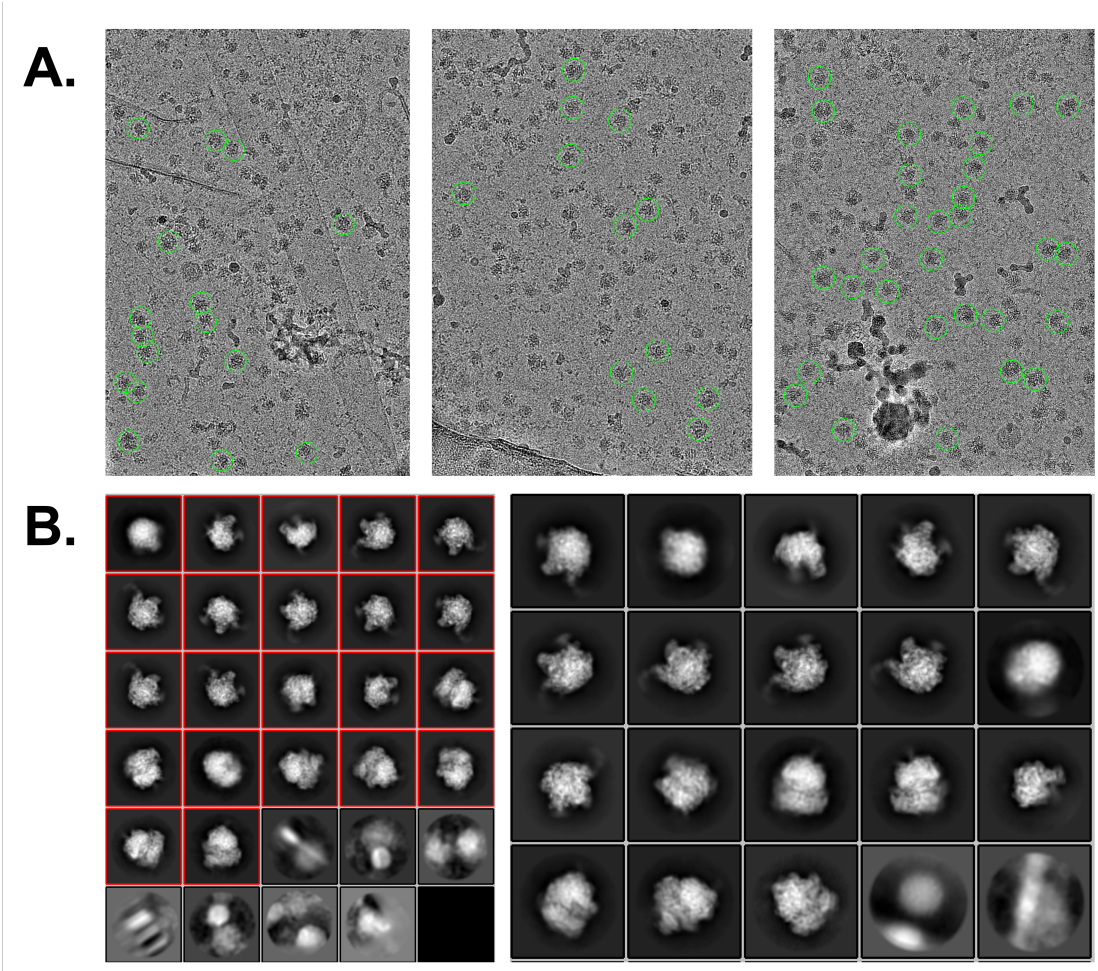



**Figure B.1:** A. Picking ribosome particles using Topaz.
B. 2D classification of ribosomes in RELION.

## B.2 Ribosome map and model data

The maps and models referred to in Chapter 4 can be downloaded here (https://tinyurl.com/zford-data-thesis).

## B.3 50S map FSC

The resolution of the 50S map was calculated using the gold-standard FSC value of 0.143, with the curve shown in Figure B.2.

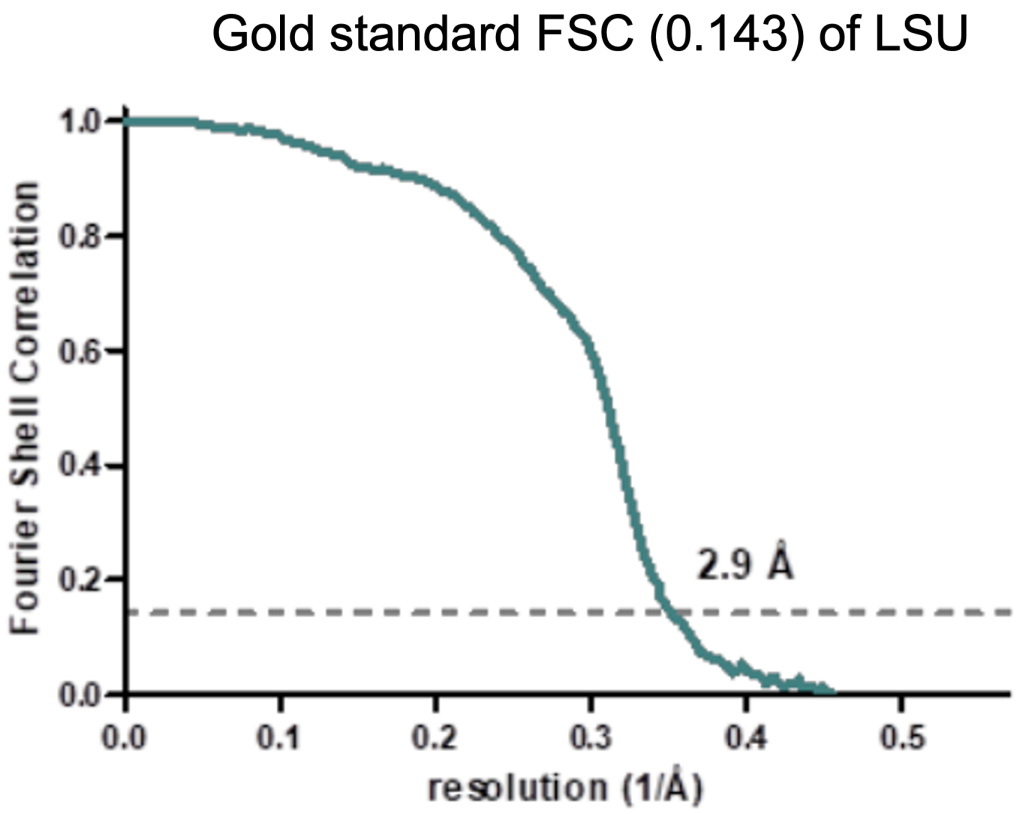



**Figure B.2:** 50S map FSC.

## B.4 50S model-in-map FSC

The resolution of the 50S map was calculated using the gold-standard FSC value of 0.143, with the curve shown in Figure B.3.

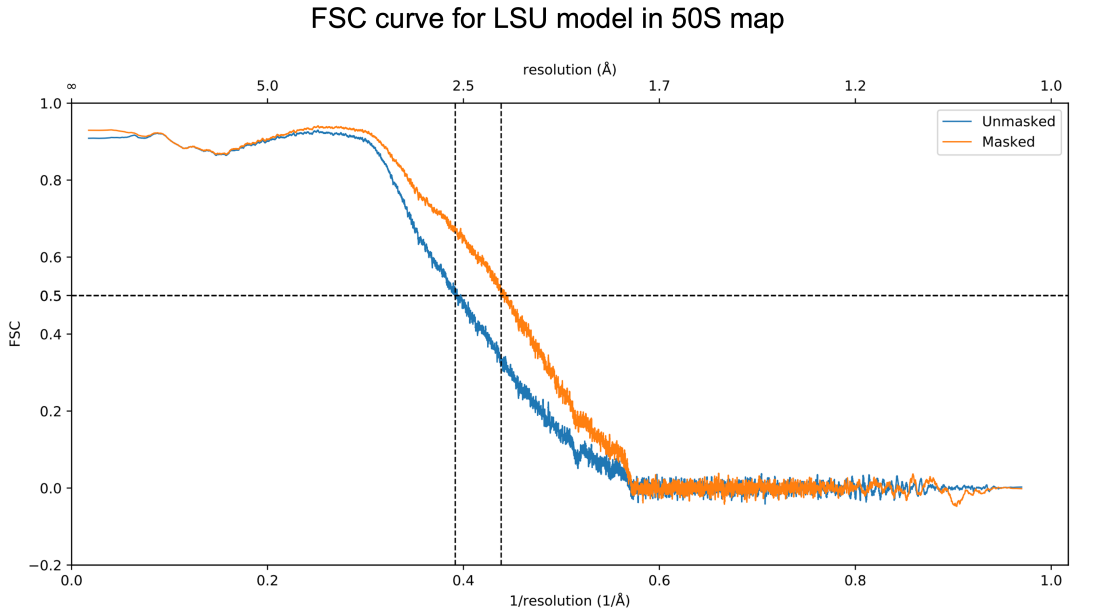



**Figure B.3:** 50S model-in-map FSC.

## B.5 70S map FSC

The resolution of the large and small subunits in the 70S map was calculated using the gold-standard FSC value of 0.143, with the curves shown in Figure B.4.

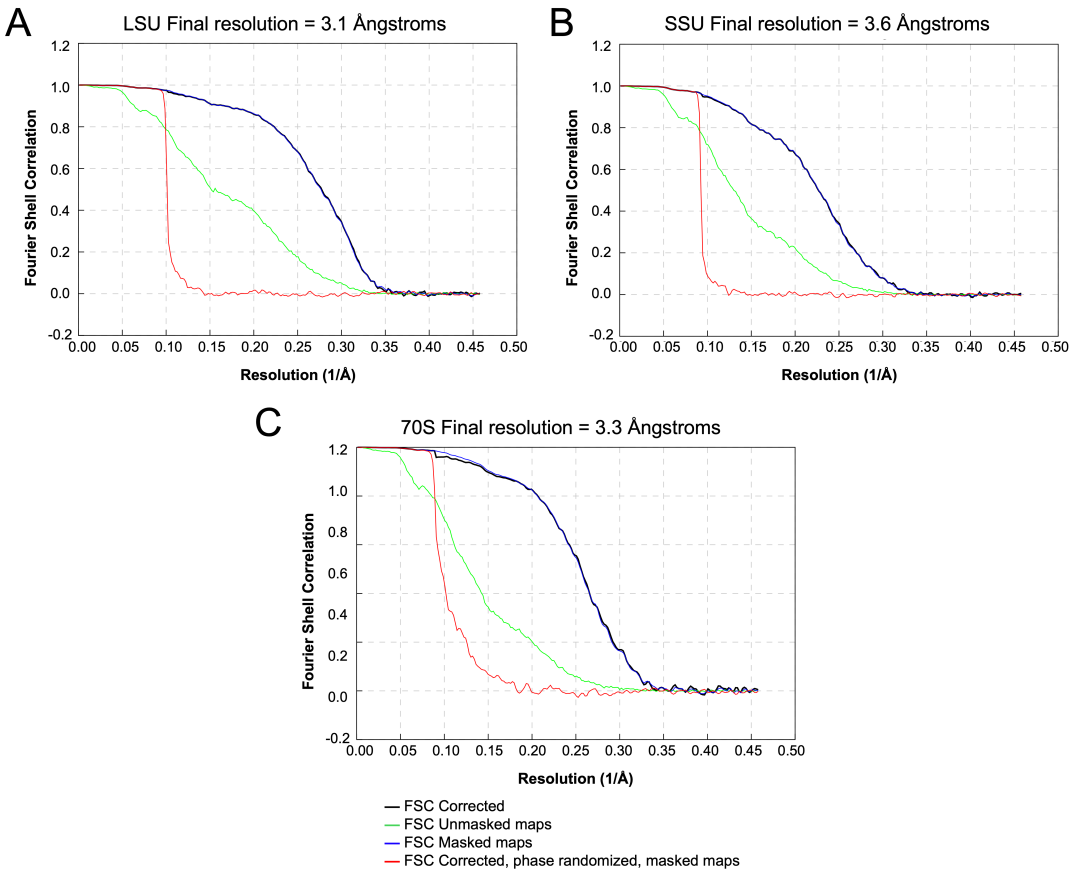



**Figure B.4:** A. 70S map LSU FSC.
B. 70S map SSU FSC.
C. 70S map overall FSC.

## B.6 70S model-in-map FSC

The resolution of the large and small subunits in the 70S map was calculated using the gold-standard FSC value of 0.143, with the curves shown in Figure B.5.

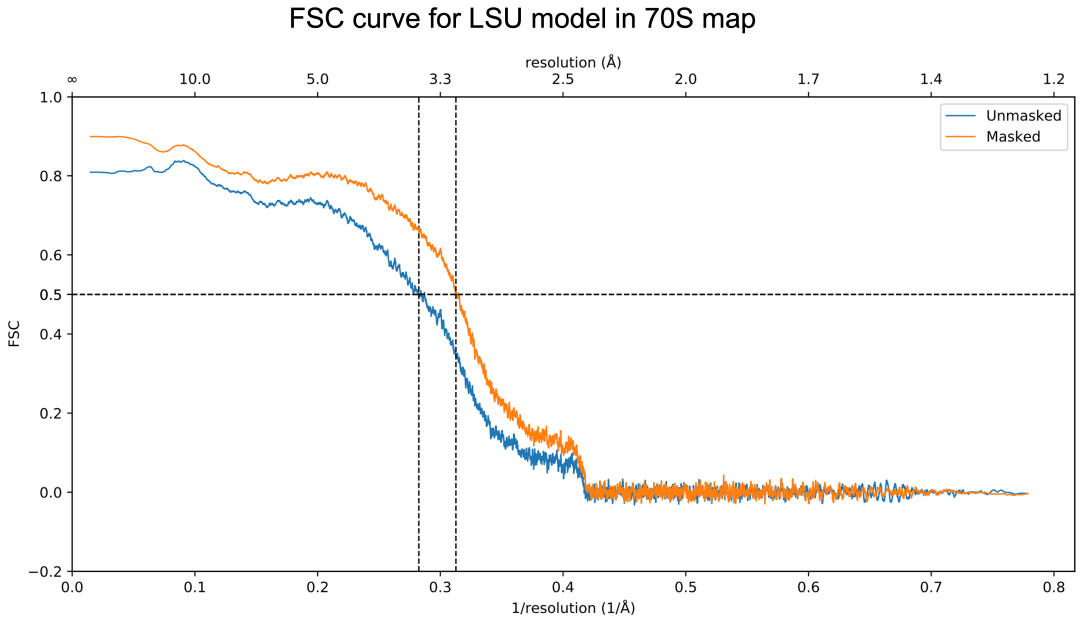



**Figure B.5:** 70S model-in-map FSC.

## B.7 LSU Model validation

PHENIX was used to run refinement and validation on the LSU models. Key statistics are summarised below in Table B.1.

**Table B.1:** Model validation in PHENIX using MolProbity, for the 50S models in both maps.

| | **50S map 50S model** | **70S map 50S model** |
|---|---|---|
| *Number of residues (Protein, nucleotide)* | 3018, 2376 | 3847, 2824 |
| *Clashes* | 215 overlapping atom pairs | 1579 overlapping atom pairs |
| *CaBLAM (Outliers %, Disfavoured %, Cα outliers %)* | 1.33, 5.91, 0.25 | 5.76, 14.90, 1.29 |
| *Cβ* | 1 deviation | 0 deviations |
| *Cis/Twisted* | 7 cis prolines | 2 cis prolines<br>6 twisted peptides |
| *Rotamers* | 18 outliers | 239 outliers |
| *Ramachandran* | 0 outliers | 38 outliers |
| *Geometry restraints (Bond length, angle, dihedral angle)* | 0, 3, 37 outliers | 0, 79, 96 outliers |
| *Correlation coefficients ($CC_{mask}$)* | 0.89 | 0.75 |

## B.8 Model validation of the SSU

Full refinement has not been performed in the 70S map yet, due to the incomplete building of the small subunit meaning that interactions with this subunit are not present and cannot be accounted for. However, the current statistics for the SSU model are shown in Table B.2.

**Table B.2:** Model validation in PHENIX using MolProbity, for the 30S model in the 70S map.

| | **70S map 30S model** |
|---|---|
| *Number of residues (Protein, nucleotide)* | 2586, 1139 |
| *Clashes* | 1489 overlapping atom pairs |
| *CaBLAM (Outliers %, Disfavoured %, Cα outliers %)* | 7.57, 18.50, 2.23 |
| *Cβ* | 0 outliers |
| *Cis/Twisted* | 2 cis conformations 5 twisted residues |
| *Rotamers* | 164 outliers |
| *Ramachandran* | 19 outliers |
| *Geometry restraints (Bond length, angle, dihedral angle)* | 20, 61, 63 |
| *Correlation coefficients ($CC_{mask}$)* | 0.66 |

# C

# Structural studies of archaeal S-layers

## Contents


**C.1 Low density of *N. maritimus* cells on grids . . . . . 290**
**C.2 *Saccharolobus* SlaA sequence alignment . . . . . 290**
**C.3 *Pyrobaculum* S-layer sequence alignment . . . . . 290**
**C.4 Selective *Pyrobaculum* S-layer sequence alignment . . 295**
**C.5 RMSD between *Pyrobaculum* S-layer domains . . . . . 296**
**C.6 TMHMM for *Pyrobaculum* S-layer sequences . . . . . 297**
**C.7 *Pyrobaculum* S-layers structure comparison . . . . . 298**
**C.8 *H. volcanii* potential cation binding sites . . . . . 299**
**C.9 *H. volcanii* VolMap density calculations . . . . . 302**
**C.10 Percentage *H. volcanii* binding site occupancies . . . . 304**
**C.11 *H. volcanii* modelled binding site occupancies . . . . . 306**

## C.1 Low density of *N. maritimus* cells on grids

Low magnification images of *N.maritimus* cells on grids showed that there was a low density of cells (Figure C.1).

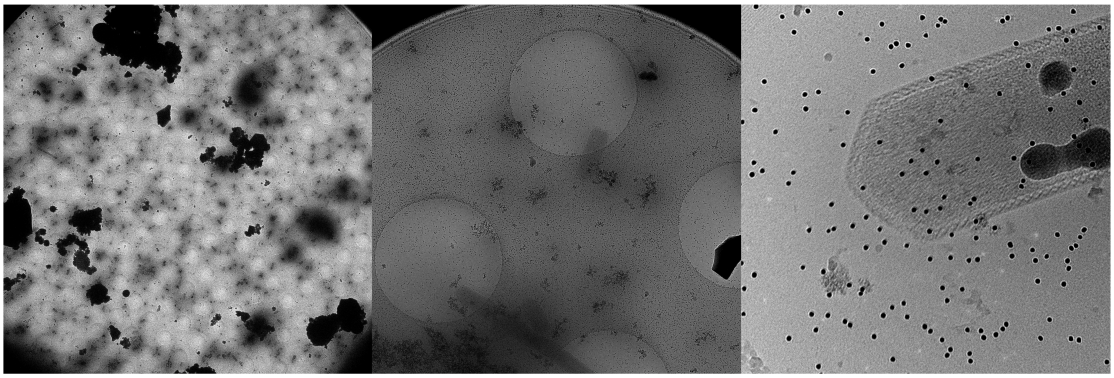


**Figure C.1:** Arctica screening of *N. maritimus* cells.

## C.2 *Saccharolobus* SlaA sequence alignment

The sequences of the S-layers from *S. solfataricus* and *S. acidocaldarius* were compared using BLAST. The results are shown in Figure C.2.

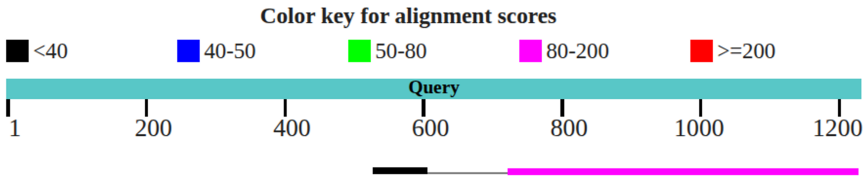
The sequences of the S-layers from *S. solfataricus* and *S. acidocaldarius* were compared using COBALT. The results are shown in Figure C.3.

## C.3 *Pyrobaculum* S-layer sequence alignment

The sequences of the S-layers from *P. oguniense* and *P. arsenaticum* were compared using BLAST. The results are shown in Figure C.4.

The sequences of the S-layers from *P. oguniense* and *P. arsenaticum* were compared altogether using COBALT. The results are shown in Figure C.5.

A.




B.

| Score | Expect | Method | Identities (X) | Positives (+) | Gaps (--) |
|---|---|---|---|---|---|
| **37.0 bits** (84) | 1e-05 | Compositional matrix adjust. | 27/81 (33%) | 36/81 (44%) | 3/81 (3%) |

```
S. solf  528   FIVERLGYIYLVTFHIWGPSATVTVTGVDYHGTQIA--LGSFRAYVALPSFYTLPKTPLV  585
               F V    YI LV F +WG   +VTV+  D    +++  LG F   V  PS  T P
S. acid  615   FTVPSTSYIALVDFGLWGNLTSVTVSAYDTVNNKLSVNLGYFYGIVIPPSISTAPYN-YQ  673

S. solf  586   GLTCDNMYTLAQVSDAGSVLE  606
                  C N Y    + D  +VL+
S. acid  674   NFICPNNYVTVTIYDPDAVLD  694
```

C.

| Score | Expect | Method | Identities (X) | Positives (+) | Gaps(--) |
|---|---|---|---|---|---|
| **136 bits** (342) | 4e-36 | Compositional matrix adjust. | 170/553 (31%) | 260/553 (47%) | 59/553 (10%) |

```
S. solf  722   TKIVMWYVSPDYAAYFYYSSTGQYQTSNVTLTFANVTPELVMPQ-VFPVGKLYMPFGIND  780
               + I + Y + DYA   Y+ STGQY+    T++  NVT  L  P    P+ +L +P  +++
S. acid  881   SNIKLTYTTTDYAVLHYFYSTGQYKVFK-TVSVPNVTANLYFPSSTTPLYQLSVPLYLSE  939

S. solf  781   PYY--VFYSSISIGPQSGV-----------IEAVENGFNVGNITSITVEL-NGMNESIVL  826
               PYY     + I +G                + AV+    +G I SI+V L NG    I L
S. acid  940   PYYGSPLPTYIGLGTNGTSLWNSPNYVLFGVSAVQQ--YLGFIKSISVTLSNGTTVVIPL  997

S. solf  827   SPLNVSKLLISSNLGEVSQCSPLFTTTLFNISALASLLGLPNAAALNGSYLYVTYHDIIS  886
               +  N+  L       E+  C+  F   + +I+ L  LL L N   LN S L VTYHD ++
S. acid  998   TTSNMQTLFPQLVGQELQACNGTFQFGI-SITGLEKLLNL-NVQQLNNSILSVTYHDYVT  1055

S. solf  887   GAYVTNRTLLVVGQFYVMPPTTPGSVEWILTA-KYI-NATTGIPVEISYAVVQQPSAKVV  944
               G  +T  T LV      +     G VE++LTA  Y  N T   P  I+  VV+QP
S. acid  1056  GETLTATTKLVALSTLSLVAKGAGVVEFLLTAYPYTGNITFAPPWFIAENVVKQPFMTYS  1115

S. solf  945   DINAAN---ASITQIQVTGVKIVSKYATVTVMYNPSNASTIVY-MNGMFVASYGGNLISS  1000
               D+  A    ++I  +    + +V      +V YN ++  T++  + G  VA+  GN++ +
S. acid  1116  DLQFAKTNPSAILSLSTVNITVVGLGGKASVYYNSTSGQTVITNIYGQTVATLSGNVLPT  1175

S. solf  1001  LSETSTF-GVYYGAVVNLYVATGTLSS-------------PNGTMYVVLGSHKVAVGTAN  1046
               L+E +   G + G++    V   T+                NG++ +VL     ++G A
S. acid  1176  LTELAAGNGTFTGSLQFTIVPNNTVVQIPSSLTKTSFAVYTNGSLAIVLNGKAYSLGPAG  1235

S. solf  1047  LYTYAGYHFGPYTALPLVSNVT--FTVQDPVTHATLTGQTTLGAFNNTPIRLAPLGV--S  1102
               L+      F  YT   + +N T   TV D V   T T Q  + A N TPIRLAP  V
S. acid  1236  LFLLP---FVTYTGSAIGANATAIITVSDGV--GTSTTQVPITAENFTPIRLAPFQVPAQ  1290

S. solf  1103  IPQTAQYKVFYYYSTPLVLSPTSQYIVLSVTSVISYPYPFYIETVSFLG--YNVTTGTPV  1160
               +P     K+ Y Y+  +V++P  Q  + + VTS++ YP  F I+   +    +NV TG+P
S. acid  1291  VPLPNAPKLKYEYNGSIVITPQQQVLKIYVTSILPYPQEFQIQAFVYEASQFNVHTGSPT  1350

S. solf  1161  PGTPAFQTVYS-----PSLGPGVVLQVPVQSYQFISLST-PSEPHTVVMFAVPFAGGPAI  1214
                    F   YS     P+LG G  +   +   Q   +S  P+  + +V+ AVPFAGGP +
S. acid  1351  AAPVYFS--YSAVRAYPALGIGTSVPNLLVYVQLQGISNLPAGKYVIVLSAVPFAGGPVL  1408

S. solf  1215  SLYPTFLVYANVT  1227
               S YP  L++ NVT
S. acid  1409  SEYPAQLIFTNVT  1421
```

**Figure C.2:** BLAST alignment of SlaA sequences from *S. solfataricus* and *S. acidocaldarius*.
A. Summary of BLAST alignments.
B. BLAST alignment shown in black in A.
C. BLAST alignment shown in pink in A.




**Figure C.3:** COBALT alignment of SlaA sequences from *S. solfataricus* and *S. acidocaldarius*. Red indicates residues are highly conserved, blue indicates lower conservation of sequence.

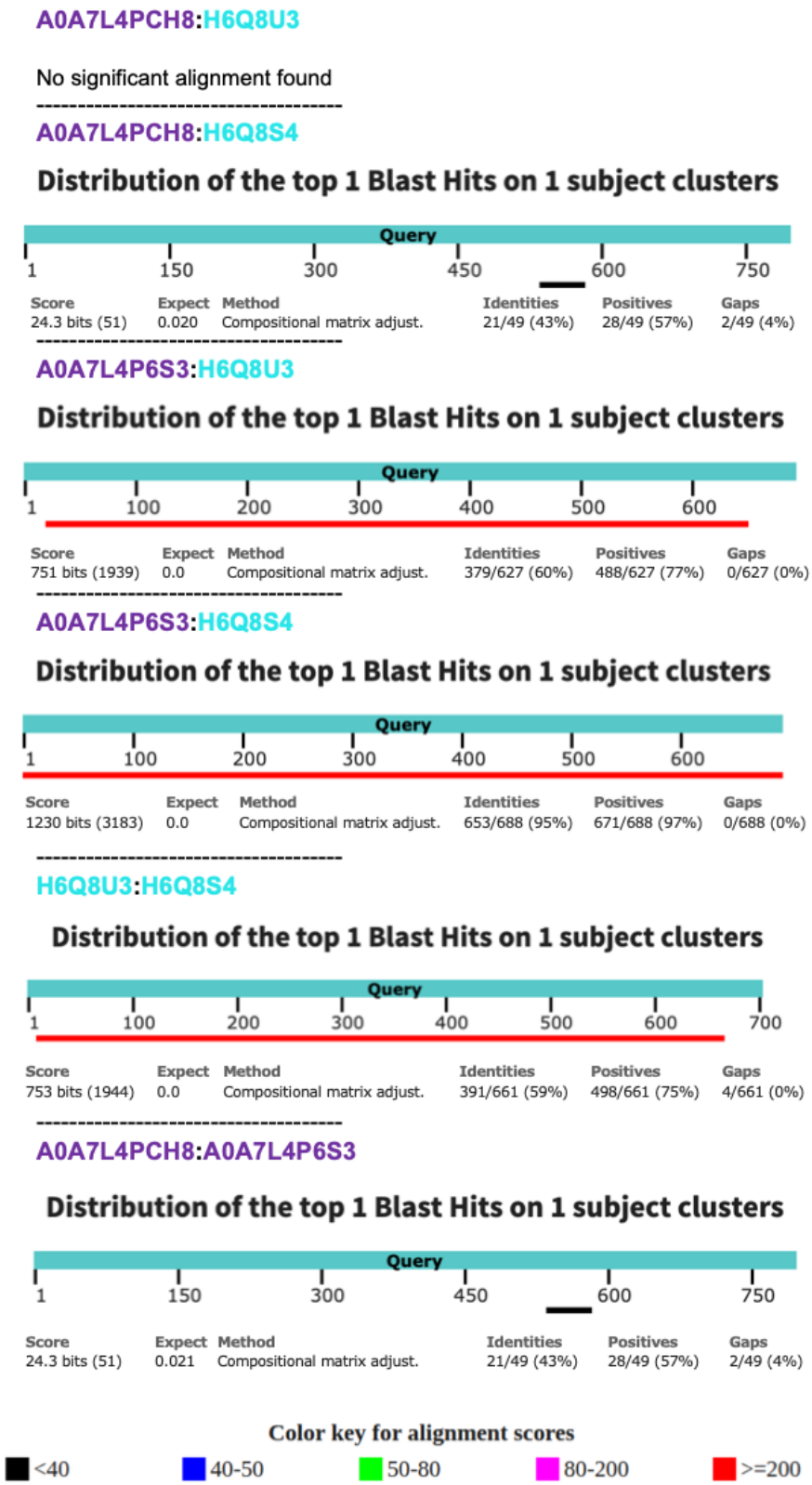



**Figure C.4:** BLAST alignment of S-layer sequences from *P. oguniense* and *P. arsenaticum*.

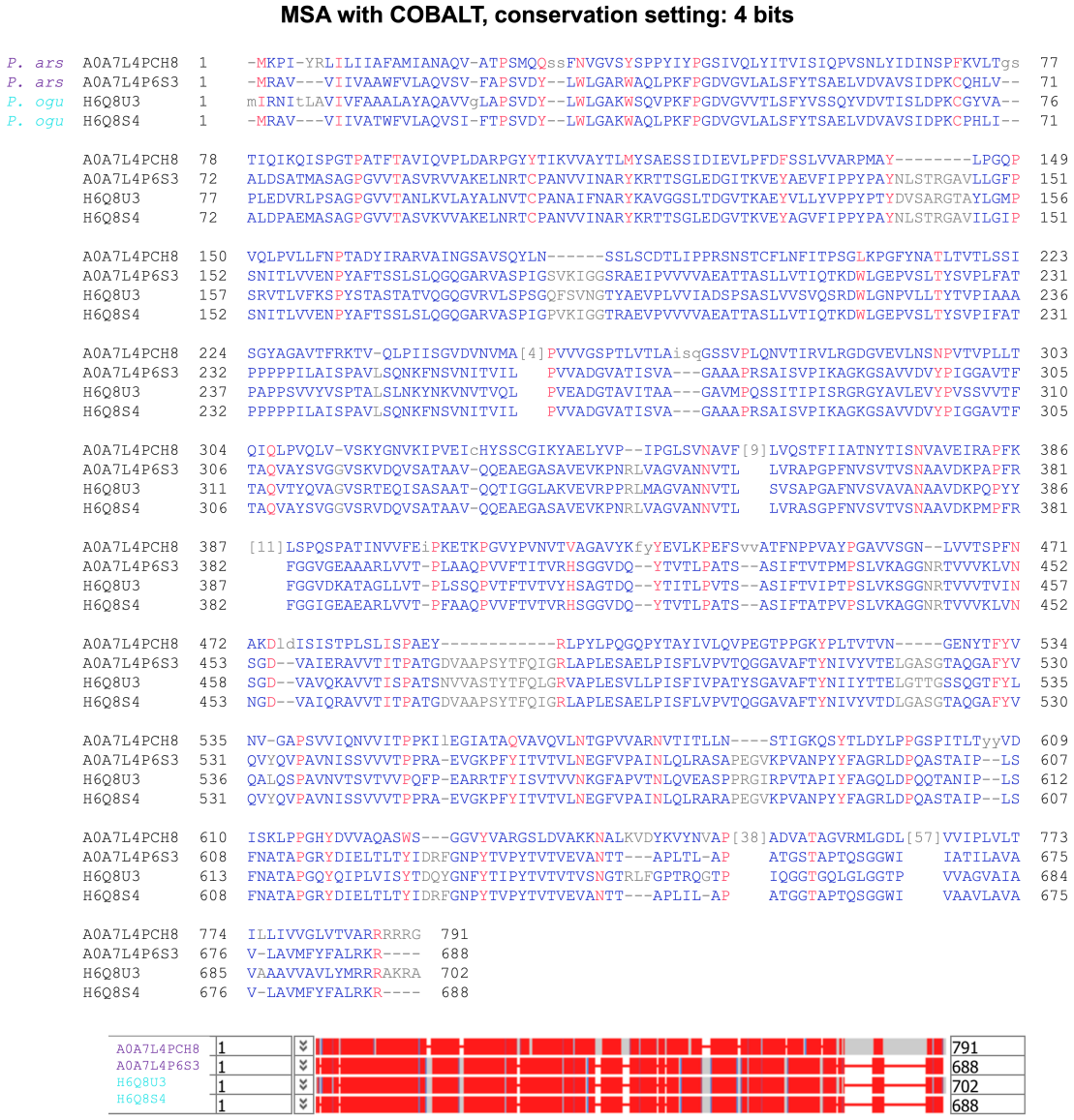



**Figure C.5:** COBALT MSA alignment of S-layer sequences from *P. oguniense* and *P. arsenaticum*. Red indicates residues are highly conserved, blue indicates lower conservation of sequence.

## C.4 Selective *Pyrobaculum* S-layer sequence alignment

Three sequences of the S-layers from *P. oguniense* (H6Q8U3 and H6Q8S4) and *P. arsenaticum* (A0A7L4P6S3) were compared using COBALT. The results are shown in Figure C.6.

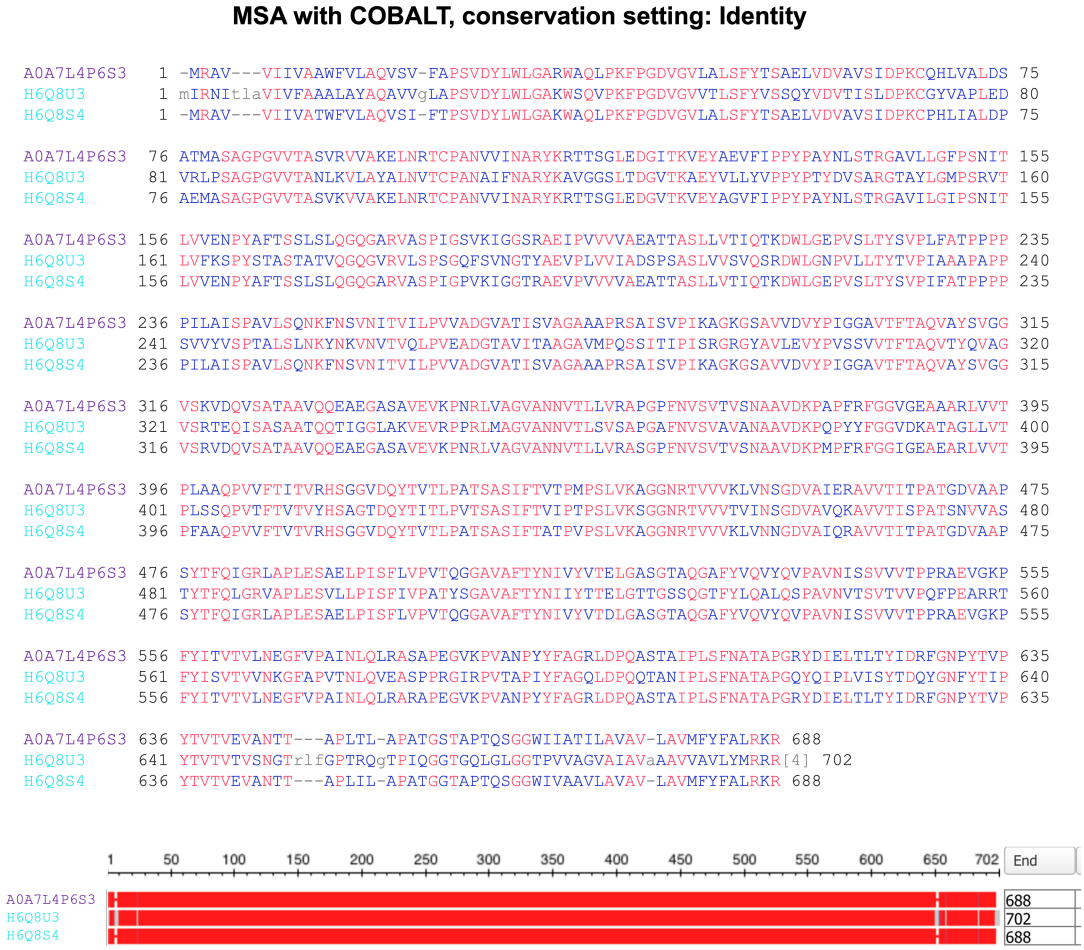



**Figure C.6:** COBALT MSA alignment of S-layer S-layers from *P. oguniense* (H6Q8U3 and H6Q8S4) and *P. arsenaticum* (A0A7L4P6S3). Red indicates residues are identical, blue indicates lower conservation of sequence.

## C.5 RMSD between *Pyrobaculum* S-layer domains

RMSD was calculated between the domains of the predicted structures of the four *Pyrobaculum* S-layer sequences. These are shown in Table C.1, Table C.2, and Table C.3.

**Table C.1:** RMSD calculated by domain, against H6Q8S4. Numbers in brackets indicate that alignment of the H6Q8S4 domain was to a different domain of the other protein.

| | **RMSD by domain, against H6Q8S4** (Å) | | |
|---|---|---|---|
| | **H6Q8U3** | **A0A7L4P6S3** | **A0A7L4PCH8** |
| *1* | 0.516 | 0.174 | 18.596 |
| *2* | 0.053 | 0.184 | 24.793 |
| *3* | 0.415 | 0.179 | 9.946 (6) |
| *4* | 0.245 | 0.210 | 3.891 (6) |
| *5* | 0.668 | 0.586 | 7.251 (6) |
| *6* | 0.484 | 0.170 | 1.628 |

**Table C.2:** RMSD calculated by domain, against H6Q8U3. Numbers in brackets indicate that alignment of the H6Q8U3 domain was to a different domain of the other protein.

| | **RMSD by domain, against H6Q8U3** (Å) | |
|---|---|---|
| | **A0A7L4P6S3** | **A0A7L4PCH8** |
| *1* | 0.522 | 21.821 |
| *2* | 0.561 | 24.709 |
| *3* | 0.344 | 17.139 |
| *4* | 0.213 | 9.200 (6) |
| *5* | 0.362 | 6.314 (7) |
| *6* | 0.389 | 2.661 (1) |

**Table C.3:** RMSD calculated by domain, against A0A7L4P6S3. Numbers in brackets indicate that alignment of the A0A7L4P6S3 domain was to a different domain of the other protein.

| | **RMSD by domain, against A0A7L4P6S3** (Å) |
|---|---|
| | **A0A7L4PCH8** |
| *1* | 18.355 |
| *2* | 23.597 |
| *3* | 9.964 (6) |
| *4* | 10.663 (6) |
| *5* | 7.673 (6) |
| *6* | 1.699 |

## C.6 TMHMM for *Pyrobaculum* S-layer sequences

The S-layer sequences were analysed through TMHMM to identify likely transmembrane helices, outputs shown in Figure C.7.

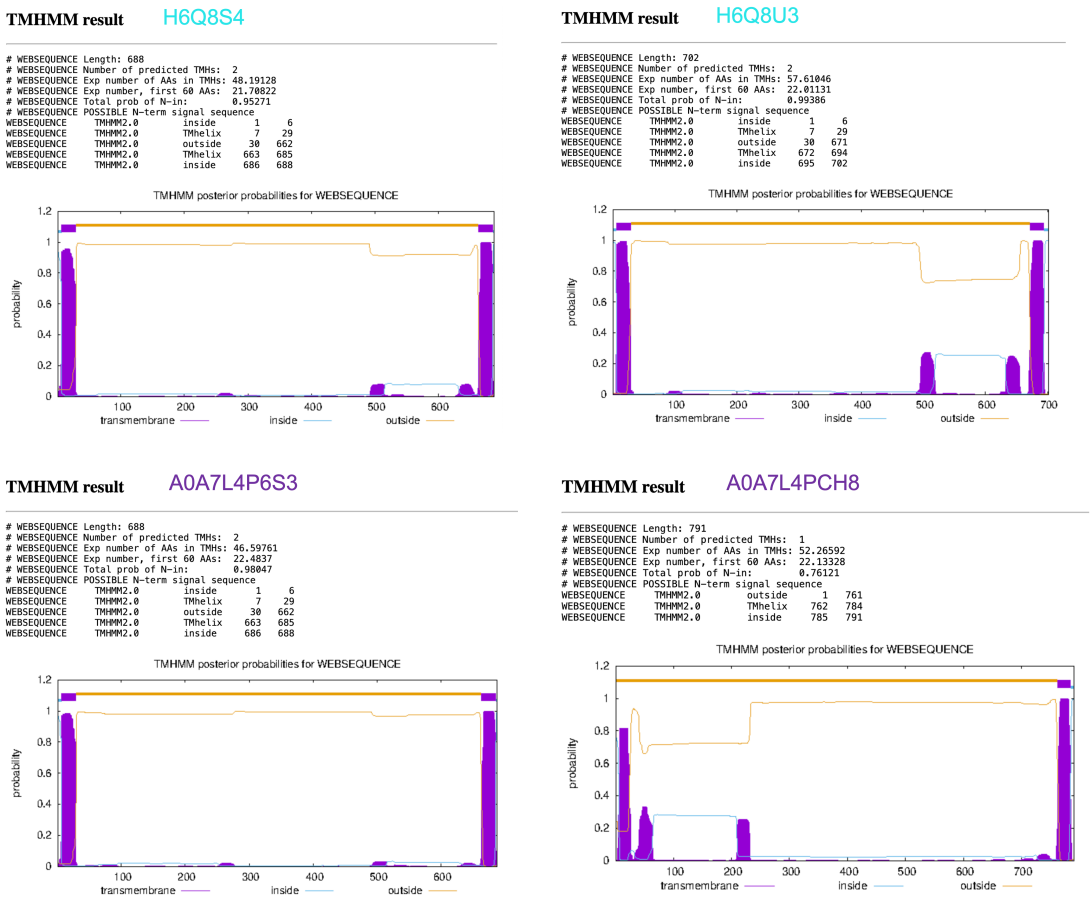



**Figure C.7:** TMHMM analysis of *Pyrobaculum* S-layer sequences, showing that the C-termini are likely to be transmembrane helices, and the N-termini are likely to be signal peptides.

## C.7 *Pyrobaculum* S-layers structure comparison

By treating the densities for the *P. oguniense* and *P. arsenaticum* S-layers as half-maps, a resolution could be calculated using an FSC curve, giving a resolution to which the two structures were the same. This was calculated to be 25 Å, shown in Figure C.8.

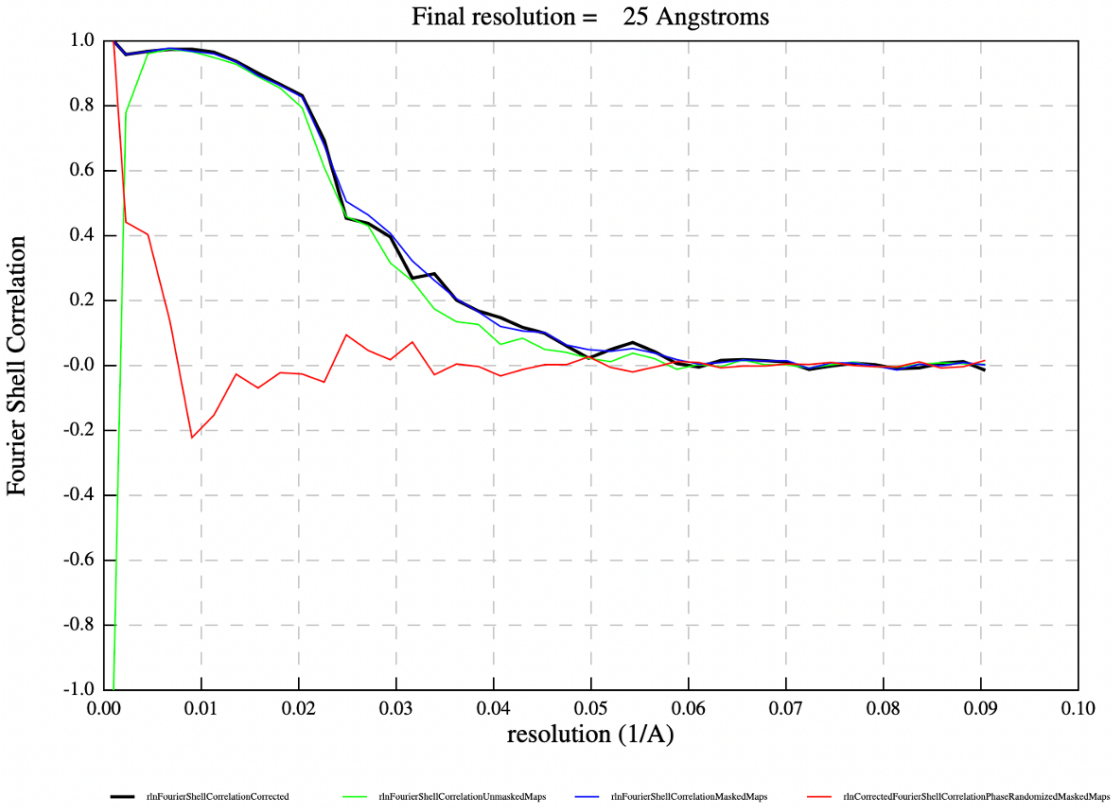



**Figure C.8:** FSC curve for comparing S-layer structures of *P. oguniense* and *P. arsenaticum*.

## C.8 *H. volcanii* potential cation binding sites

104 potential divalent cation binding sites were identified in the *H. volcanii* S-layer structure. The pair or trio of residues used to define the binding site are listed below in Figure C.10, with chain numbering following Figure C.9.

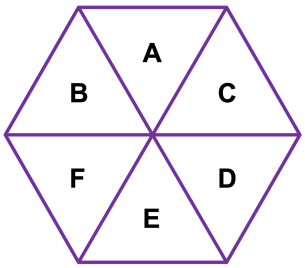



**Figure C.9:** Chain identification in the *H. volcanii* S-layer.

| *Site number* | Residue 1 | Residue 2 | Residue 3 | Ca2+ occupancy (%) | Mg2+ occupancy (%) |
|---|---|---|---|---|---|
| *1* | resid 35 and segid A | resid 36 and segid A | resid 38 and segid A | 84.1 | 56.67 |
| *2* | resid 190 and segid A | resid 40 and segid A | | 55.16 | 44.35 |
| *3* | resid 40 and segid A | resid 160 and segid A | resid 190 and segid A | 6.12 | 3.04 |
| *4* | resid 156 and segid A | resid 61 and segid C | resid 90 and segid C | 2.94 | 1.16 |
| *5* | resid 61 and segid A | resid 225 and segid A | | 99.15 | 81.48 |
| *6* | resid 61 and segid A | resid 225 and segid A | resid 224 and segid A | 99.15 | 81.48 |
| *7* | resid 65 and segid A | resid 193 and segid B | | 1.22 | 0.1 |
| *8* | resid 80 and segid A | resid 80 and segid C | | 31 | 28.15 |
| *9* | resid 90 and segid A | resid 191 and segid B | | 1.44 | 0.32 |
| *10* | resid 98 and segid A | resid 94 and segid A | resid 96 and segid A | 38.12 | 7.91 |
| *11* | resid 106 and segid A | resid 113 and segid A | resid 108 and segid A | 23.35 | 7.62 |
| *12* | resid 230 and segid A | resid 233 and segid A | resid 125 and segid A | 10.77 | 5.99 |
| *13* | resid 127 and segid A | resid 128 and segid A | resid 137 and segid A | 0.09 | 0.49 |
| *14* | resid 233 and segid A | resid 128 and segid A | resid 127 and segid A | 94.29 | 42.85 |
| *15* | resid 127 and segid A | resid 129 and segid A | resid 137 and segid A | 26.05 | 25.3 |
| *16* | resid 147 and segid A | resid 146 and segid A | | 97.08 | 55.3 |
| *17* | resid 221 and segid A | resid 166 and segid A | resid 168 and segid A | 99.09 | 91.93 |
| *18* | resid 166 and segid A | resid 187 and segid A | resid 178 and segid A | 64.12 | 77.25 |
| *19* | resid 166 and segid A | resid 223 and segid A | resid 224 and segid A | 92.14 | 91.55 |
| *20* | resid 170 and segid A | resid 174 and segid A | resid 176 and segid A | 97.15 | 62.81 |
| *21* | resid 179 and segid A | resid 297 and segid B | resid 301 and segid B | 18.56 | 17.51 |
| *22* | resid 210 and segid A | resid 207 and segid A | resid 180 and segid A | 18.87 | 2.8 |
| *23* | resid 190 and segid A | resid 188 and segid A | | 46.22 | 25.05 |
| *24* | resid 209 and segid A | resid 436 and segid B | resid 411 and segid B | 0.23 | 0 |
| *25* | resid 210 and segid A | resid 259 and segid B | resid 345 and segid B | 15.79 | 6.97 |
| *26* | resid 440 and segid A | resid 257 and segid A | resid 212 and segid C | 0.21 | 0.37 |
| *27* | resid 223 and segid A | resid 224 and segid A | resid 227 and segid A | 49.01 | 70.92 |
| *28* | resid 224 and segid A | resid 225 and segid A | resid 227 and segid A | 99.87 | 87.39 |
| *29* | resid 439 and segid A | resid 241 and segid C | resid 243 and segid C | 15.77 | 6.27 |
| *30* | resid 242 and segid A | resid 243 and segid A | resid 244 and segid A | 42.55 | 36.38 |
| *31* | resid 443 and segid A | resid 437 and segid A | resid 242 and segid C | 5.48 | 3.21 |
| *32* | resid 443 and segid A | resid 445 and segid A | resid 244 and segid C | 73.31 | 37.22 |
| *33* | resid 376 and segid A | resid 241 and segid A | | 0 | 0 |
| *34* | resid 250 and segid A | resid 261 and segid A | | 92 | 42.07 |
| *35* | resid 336 and segid A | resid 334 and segid A | resid 335 and segid A | 99.11 | 84.31 |
| *36* | resid 334 and segid A | resid 270 and segid A | resid 272 and segid A | 85.91 | 62.08 |
| *37* | resid 272 and segid A | resid 302 and segid A | resid 303 and segid A | 93.45 | 60.34 |
| *38* | resid 280 and segid A | resid 278 and segid A | resid 566 and segid B | 5.66 | 0.04 |
| *39* | resid 567 and segid A | resid 278 and segid C | resid 327 and segid C | 19.62 | 1.86 |
| *40* | resid 355 and segid A | resid 357 and segid A | resid 278 and segid A | 69.19 | 51.95 |
| *41* | resid 569 and segid A | resid 350 and segid C | resid 280 and segid C | 9.03 | 5.54 |
| *42* | resid 281 and segid A | resid 353 and segid A | resid 352 and segid A | 67.61 | 60.86 |
| *43* | resid 281 and segid A | resid 355 and segid A | resid 357 and segid A | 38.04 | 33.81 |
| *44* | resid 475 and segid A | resid 285 and segid A | resid 283 and segid A | 34.9 | 17.03 |
| *45* | resid 495 and segid A | resid 283 and segid A | | 0.78 | 0.21 |
| *46* | resid 289 and segid A | resid 318 and segid A | resid 288 and segid A | 0.93 | 0.31 |
| *47* | resid 293 and segid A | resid 345 and segid A | | 82.06 | 69.69 |
| *48* | resid 300 and segid A | resid 336 and segid A | resid 339 and segid A | 66.02 | 72.82 |
| *49* | resid 327 and segid A | resid 326 and segid A | resid 324 and segid A | 96.94 | 63.49 |
| *50* | resid 351 and segid A | resid 692 and segid B | | 67.61 | 42.13 |
| *51* | resid 353 and segid A | resid 355 and segid A | resid 566 and segid B | 1.75 | 0.01 |
| *52* | resid 372 and segid A | resid 369 and segid A | resid 359 and segid A | 0.14 | 0.01 |

| *Site number* | Residue 1 | Residue 2 | Residue 3 | Ca2+ occupancy (%) | Mg2+ occupancy (%) |
|---|---|---|---|---|---|
| *53* | resid 372 and segid A | resid 446 and segid B | | 56.09 | 25.8 |
| *54* | resid 372 and segid A | resid 446 and segid B | resid 445 and segid B | 55.62 | 25.63 |
| *55* | resid 383 and segid A | resid 498 and segid A | | 2.3 | 2.45 |
| *56* | resid 383 and segid A | resid 714 and segid B | resid 692 and segid B | 47.04 | 20 |
| *57* | resid 539 and segid A | resid 400 and segid A | resid 559 and segid A | 0.39 | 0.1 |
| *58* | resid 419 and segid A | resid 420 and segid A | | 95.3 | 82.04 |
| *59* | resid 420 and segid A | resid 739 and segid B | resid 699 and segid B | 14.28 | 1.36 |
| *60* | resid 427 and segid A | resid 432 and segid A | resid 485 and segid A | 42.32 | 22.69 |
| *61* | resid 444 and segid A | resid 445 and segid A | resid 428 and segid A | 5.54 | 3.29 |
| *62* | resid 444 and segid A | resid 431 and segid A | | 0 | 0.01 |
| *63* | resid 437 and segid A | resid 439 and segid A | resid 441 and segid A | 84.73 | 78.92 |
| *64* | resid 473 and segid A | resid 474 and segid A | resid 477 and segid A | 13.93 | 3.94 |
| *65* | resid 492 and segid A | resid 495 and segid A | resid 475 and segid A | 2.3 | 0.65 |
| *66* | resid 477 and segid A | resid 481 and segid A | resid 483 and segid A | 15.55 | 10.38 |
| *67* | resid 481 and segid A | resid 483 and segid A | resid 485 and segid A | 19.64 | 13.28 |
| *68* | resid 481 and segid A | resid 483 and segid A | resid 486 and segid A | 0.92 | 0.96 |
| *69* | resid 491 and segid A | resid 741 and segid B | | 3.72 | 2.51 |
| *70* | resid 714 and segid A | resid 494 and segid C | resid 495 and segid C | 23.1 | 3.02 |
| *71* | resid 494 and segid A | resid 496 and segid A | resid 498 and segid A | 36.96 | 67.44 |
| *72* | resid 712 and segid A | resid 498 and segid C | | 47.29 | 5.68 |
| *73* | resid 507 and segid A | resid 640 and segid A | resid 641 and segid A | 1.34 | 0.73 |
| *74* | resid 512 and segid A | resid 647 and segid A | resid 648 and segid A | 0 | 0 |
| *75* | resid 569 and segid A | resid 692 and segid A | resid 530 and segid A | 14.73 | 4.2 |
| *76* | resid 551 and segid A | resid 618 and segid A | resid 619 and segid A | 84.37 | 50.26 |
| *77* | resid 615 and segid A | resid 618 and segid A | resid 552 and segid A | 94.1 | 37.36 |
| *78* | resid 560 and segid A | resid 562 and segid A | resid 564 and segid A | 89.98 | 74.74 |
| *79* | resid 566 and segid A | resid 568 and segid A | | 86.11 | 69.18 |
| *80* | resid 566 and segid A | resid 567 and segid A | resid 568 and segid A | 99.99 | 99.99 |
| *81* | resid 588 and segid A | resid 589 and segid A | | 16.39 | 5.06 |
| *82* | resid 643 and segid A | resid 642 and segid A | resid 589 and segid A | 99.99 | 99.99 |
| *83* | resid 594 and segid A | resid 638 and segid A | resid 634 and segid A | 99.85 | 99.01 |
| *84* | resid 603 and segid A | resid 602 and segid A | resid 596 and segid A | 5.61 | 7.71 |
| *85* | resid 615 and segid A | resid 612 and segid A | resid 611 and segid A | 99.12 | 48.46 |
| *86* | resid 625 and segid A | resid 623 and segid A | resid 626 and segid A | 82.8 | 30.83 |
| *87* | resid 639 and segid A | resid 640 and segid A | | 51.85 | 49.76 |
| *88* | resid 648 and segid A | resid 647 and segid A | resid 649 and segid A | 96.02 | 32.55 |
| *89* | resid 654 and segid A | resid 744 and segid A | | 44.55 | 35.61 |
| *90* | resid 654 and segid A | resid 742 and segid A | resid 744 and segid A | 4.31 | 0.22 |
| *91* | resid 665 and segid A | resid 679 and segid A | resid 680 and segid A | 86.29 | 28.7 |
| *92* | resid 759 and segid A | resid 679 and segid A | resid 665 and segid A | 6.53 | 0.45 |
| *93* | resid 665 and segid A | resid 680 and segid A | resid 725 and segid A | 91.83 | 63.16 |
| *94* | resid 756 and segid A | resid 679 and segid A | resid 677 and segid A | 49.34 | 5.88 |
| *95* | resid 699 and segid A | resid 705 and segid A | resid 707 and segid A | 0.24 | 0.29 |
| *96* | resid 699 and segid A | resid 739 and segid A | resid 741 and segid A | 93.95 | 65.53 |
| *97* | resid 704 and segid A | resid 705 and segid A | resid 706 and segid A | 89.06 | 47.39 |
| *98* | resid 711 and segid A | resid 712 and segid A | resid 725 and segid A | 2.32 | 0.26 |
| *99* | resid 712 and segid A | resid 714 and segid A | resid 715 and segid A | 93.22 | 51.63 |
| *100* | resid 715 and segid A | resid 713 and segid A | resid 723 and segid A | 80.49 | 48.96 |
| *101* | resid 717 and segid A | resid 718 and segid A | resid 719 and segid A | 15.4 | 4.75 |
| *102* | resid 739 and segid A | resid 741 and segid A | resid 746 and segid A | 89.92 | 64.33 |
| *103* | resid 747 and segid A | resid 747 and segid E | resid 748 and segid A | 21.07 | 15.24 |
| *104* | resid 756 and segid A | resid 757 and segid A | resid 759 and segid A | 85.86 | 36.69 |

**Figure C.10:** List of potential divalent cation binding sites identified in the *H. volcanii* S-layer, numbered as they are referred to in the main text. Purple shading indicates the residue is in a different monomer, so the binding site is inter-monomer. Resid: residue number. Segid: chain identifier, as shown in Figure C.9.

## C.9 *H. volcanii* VolMap density calculations

VolMap was used to calculate ion density across the simulations. This could be represented as planes of density in VMD, with red being lowest and blue being the highest. This is shown in Figure C.11.

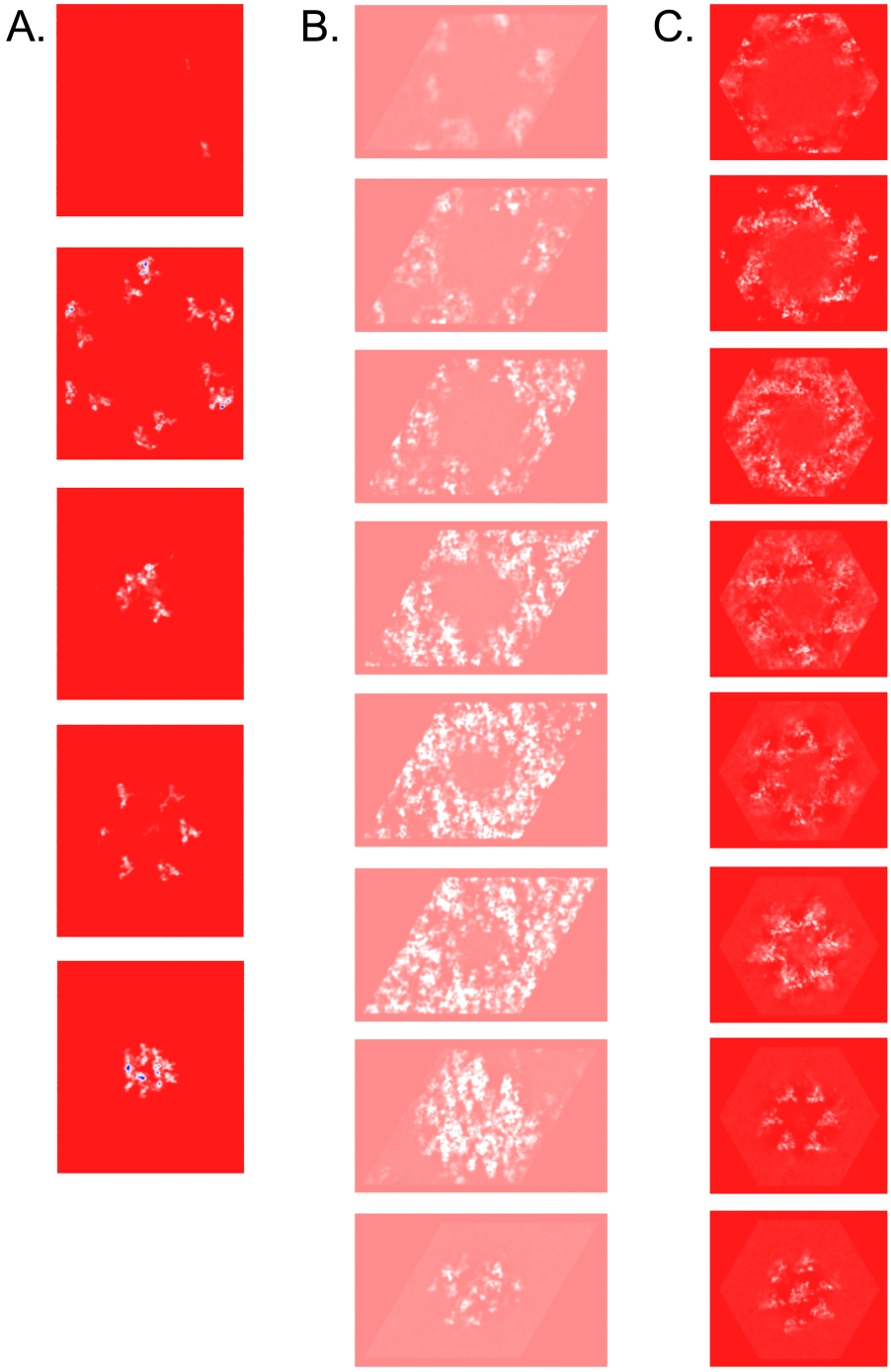
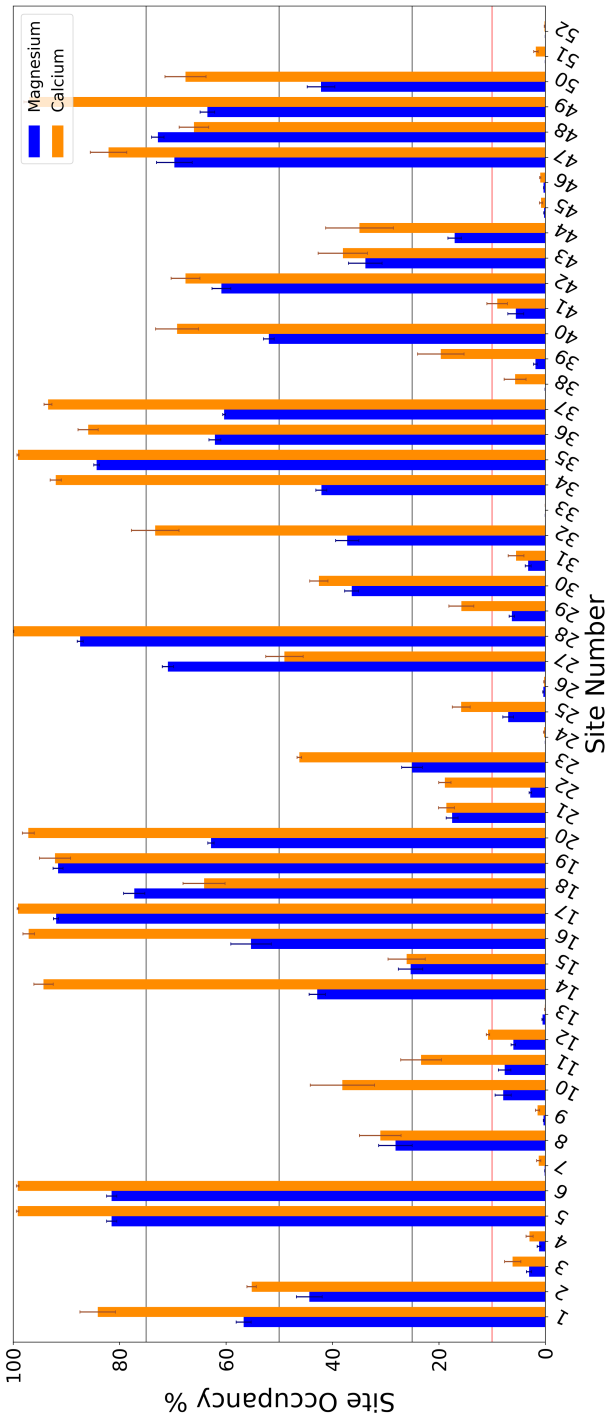



**Figure C.11:** VolMap densities of ions, sliced in $z$ and going from the top of the S-layer to the bottom.
A. Calcium ions that were in the published structure.
B. Calcium ions soaked into the S-layer.
C. Magnesium ions soaked into the S-layer.

## C.10 Percentage *H. volcanii* binding site occupancies

The percentage occupancy of each of the 104 binding sites was calculated for both $Ca^{2+}$ and $Mg^{2+}$. The results are plotted in Figure C.12.



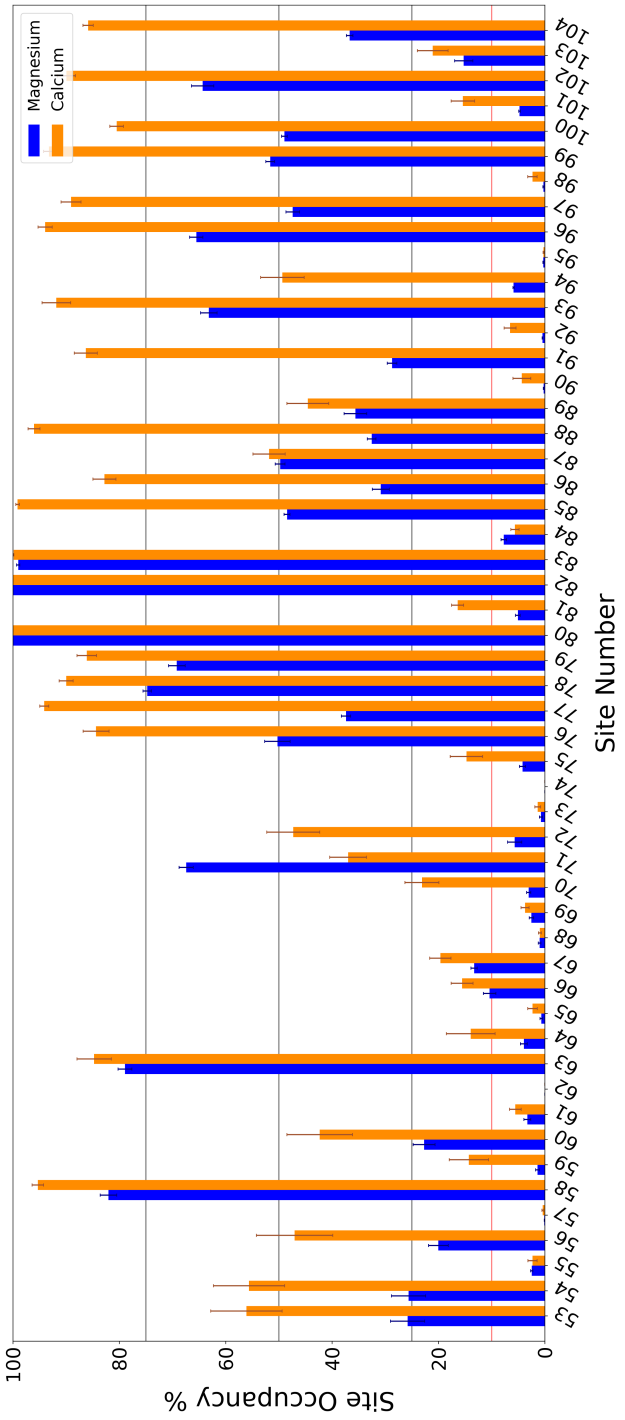



**Figure C.12:** Graph showing the occupancy of each binding site.

## C.11 *H. volcanii* modelled binding site occupancies

The 73 binding sites shown in the structure in Figure 5.29, shown shaded by their percentage occupancy in Figure C.13.

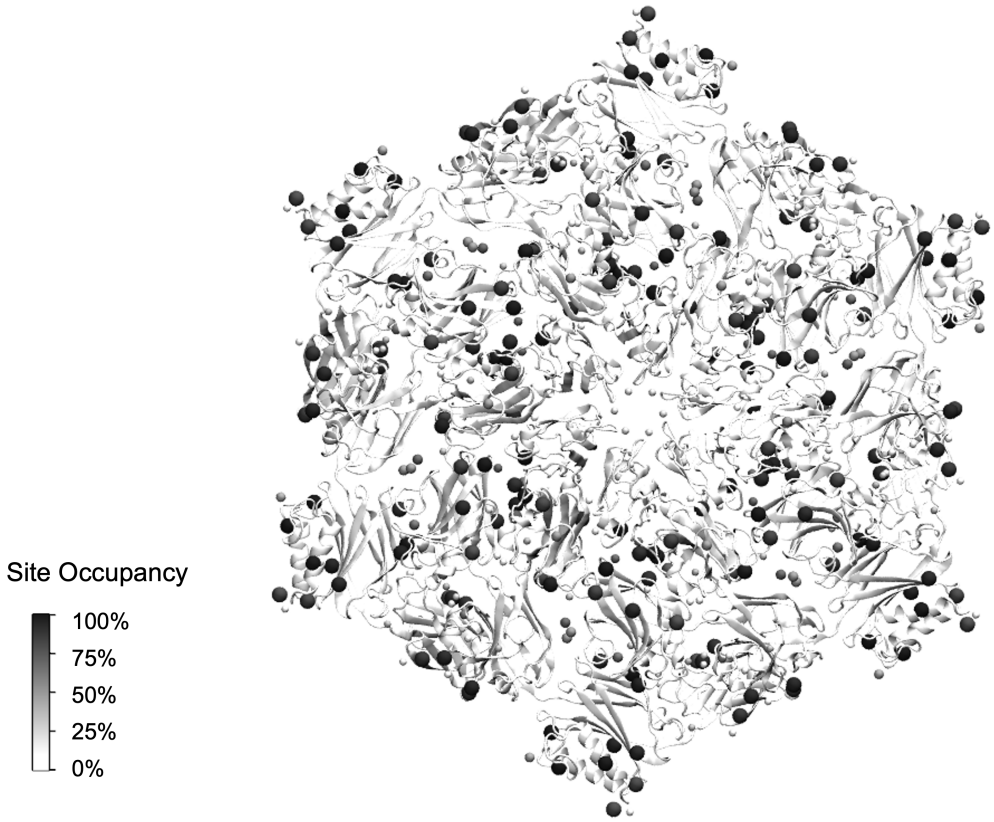



**Figure C.13:** Depiction of percentage occupancy for the ions built into the S-layer structure.

# Bibliography


Abdul Halim, M. F., Karch, K. R., Zhou, Y., Haft, D. H., Garcia, B. A., and Pohlschroder, M. (2016). Permuting the PGF Signature Motif Blocks both Archaeosortase-Dependent C-Terminal Cleavage and Prenyl Lipid Attachment for the Haloferax volcanii S-Layer Glycoprotein. *Journal of Bacteriology*, 198(5):808–815.

Abraham, M. J., Hess, B., van der Spoel, D., Lindahl, E., and the GROMACS development team (2022). *GROMACS User Manual Version 2022*.

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1–2:19–25.

Abrams, C. and Bussi, G. (2014). Enhanced Sampling in Molecular Dynamics Using Metadynamics, Replica-Exchange, and Temperature-Acceleration. *Entropy*, 16(1):163–199.

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., Beattie, C., Bertolli, O., Bridgland, A., Cherepanov, A., Congreve, M., Cowen-Rivers, A. I., Cowie, A., Figurnov, M., Fuchs, F. B., Gladman, H., Jain, R., Khan, Y. A., Low, C. M. R., Perlin, K., Potapenko, A., Savy, P., Singh, S., Stecula, A., Thillaisundaram, A., Tong, C., Yakneen, S., Zhong, E. D., Zielinski, M., Žídek, A., Bapst, V., Kohli, P., Jaderberg, M., Hassabis, D., and Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016):493–500.

Adam, P. S., Borrel, G., Brochier-Armanet, C., and Gribaldo, S. (2017). The growing tree of Archaea: New perspectives on their diversity, evolution and ecology. *The ISME Journal*, 11(11):2407–2425.

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010). PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica. Section D, Biological Crystallography*, 66(Pt 2):213–221.

Aebi, U. and Pollard, T. D. (1987). A glow discharge unit to render electron microscope grids and other surfaces hydrophilic. *Journal of Electron Microscopy Technique*, 7(1):29–33.

Afonine, P. V., Klaholz, B. P., Moriarty, N. W., Poon, B. K., Sobolev, O. V., Terwilliger, T. C., Adams, P. D., and Urzhumtsev, A. (2018a). New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallographica Section D: Structural Biology*, 74(9):814–840.

Afonine, P. V., Poon, B. K., Read, R. J., Sobolev, O. V., Terwilliger, T. C., Urzhumtsev, A., and Adams, P. D. (2018b). Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallographica. Section D, Structural Biology*, 74(Pt 6):531–544.

Agarwal, V. and McShan, A. C. (2024). The power and pitfalls of AlphaFold2 for structure prediction beyond rigid globular proteins. *Nature chemical biology*, 20(8):950–959.

Agulleiro, J.-I. and Fernandez, J.-J. (2015). Tuning the cache memory usage in tomographic reconstruction on standard computers with Advanced Vector eXtensions (AVX). *Data in Brief*, 3:16–20.

Ahn, M., Włodarski, T., Mitropoulou, A., Chan, S. H. S., Sidhu, H., Plessa, E., Becker, T. A., Budisa, N., Waudby, C. A., Beckmann, R., Cassaignau, A. M. E., Cabrita, L. D., and Christodoulou, J. (2022). Modulating co-translational protein folding by rational design and ribosome engineering. *Nature Communications*, 13(1):4243.

Akdel, M., Pires, D. E. V., Pardo, E. P., Jänes, J., Zalevsky, A. O., Mészáros, B., Bryant, P., Good, L. L., Laskowski, R. A., Pozzati, G., Shenoy, A., Zhu, W., Kundrotas, P., Serra, V. R., Rodrigues, C. H. M., Dunham, A. S., Burke, D., Borkakoti, N., Velankar, S., Frost, A., Basquin, J., Lindorff-Larsen, K., Bateman, A., Kajava, A. V., Valencia, A., Ovchinnikov, S., Durairaj, J., Ascher, D. B., Thornton, J. M., Davey, N. E., Stein, A., Elofsson, A., Croll, T. I., and Beltrao, P. (2022). A structural biology community assessment of AlphaFold2 applications. *Nature Structural & Molecular Biology*, 29(11):1056–1067.

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

Al-Hashimi, H. M. and Walter, N. G. (2008). RNA dynamics: It is about time. *Current Opinion in Structural Biology*, 18(3):321–329.

Albers, S.-V. and Meyer, B. H. (2011). The archaeal cell envelope. *Nature Reviews Microbiology*, 9(6):414–426.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). From DNA to RNA. In *Molecular Biology of the Cell. 4th Edition*. Garland Science.

Alder, B. J. and Wainwright, T. E. (1959). Studies in Molecular Dynamics. I. General Method. *The Journal of Chemical Physics*, 31(2):459–466.

Altschul, S. F. (1993). A protein alignment scoring system sensitive at all evolutionary distances. *Journal of Molecular Evolution*, 36(3):290–300.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–410.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, 25(17):3389–3402.

Anfinsen, C. B. (1973). Principles that Govern the Folding of Protein Chains. *Science*, 181(4096):223–230.

Arakawa, T. (2017). Protein–solvent interaction. *Biophysical Reviews*, 10(2):203–208.

Arbing, M. A., Chan, S., Shin, A., Phan, T., Ahn, C. J., Rohlin, L., and Gunsalus, R. P. (2012). Structure of the surface layer of the methanogenic archaean Methanosarcina acetivorans. *Proceedings of the National Academy of Sciences*, 109(29):11812–11817.

Arenz, S., Bock, L. V., Graf, M., Innis, C. A., Beckmann, R., Grubmüller, H., Vaiana, A. C., and Wilson, D. N. (2016). A combined cryo-EM and molecular dynamics approach reveals the mechanism of ErmBL-mediated translation arrest. *Nature Communications*, 7(1):12026.

Armache, J.-P., Anger, A. M., Márquez, V., Franckenberg, S., Fröhlich, T., Villa, E., Berninghausen, O., Thomm, M., Arnold, G. J., Beckmann, R., and Wilson, D. N. (2013). Promiscuous behaviour of archaeal ribosomal proteins: Implications for eukaryotic ribosome evolution. *Nucleic Acids Research*, 41(2):1284–1293.

Aseev, L. V., Koledinskaya, L. S., and Boni, I. V. (2024). Extraribosomal Functions of Bacterial Ribosomal Proteins—An Update, 2023. *International Journal of Molecular Sciences*, 25(5):2957.

Ashkenasy, N., Sánchez-Quesada, J., Ghadiri, M. R., and Bayley, H. (2005). Recognizing a Single Base in an Individual DNA Strand: A Step Toward Nanopore DNA Sequencing. *Angewandte Chemie (International ed. in English)*, 44(9):1401–1404.

Assmann, S. M., Chou, H.-L., and Bevilacqua, P. C. (2023). Rock, scissors, paper: How RNA structure informs function. *The Plant Cell*, 35(6):1671–1707.

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V., van Dijk, A. A., Ebrecht, A. C., Opperman, D. J., Sagmeister, T., Buhlheller, C., Pavkov-Keller, T., Rathinaswamy, M. K., Dalwadi, U., Yip, C. K., Burke, J. E., Garcia, K. C., Grishin, N. V., Adams, P. D., Read, R. J., and Baker, D. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557):871–876.

Baek, M., McHugh, R., Anishchenko, I., Jiang, H., Baker, D., and DiMaio, F. (2024). Accurate prediction of protein–nucleic acid complexes using RoseTTAFoldNA. *Nature Methods*, 21(1):117–121.

Bai, X.-c., Fernandez, I. S., McMullan, G., and Scheres, S. H. (2013). Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *eLife*, 2:e00461.

Bairoch, A., Apweiler, R., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'Donovan, C., Redaschi, N., and Yeh, L.-S. L. (2005). The Universal Protein Resource (UniProt). *Nucleic Acids Research*, 33(suppl_1):D154–D159.

Bakker, H., Bleeker, A., and Mul, P. (1996). Design and performance of an ultra-high-resolution 300 kV microscope. *Ultramicroscopy*, 64(1):17–34.

Ban, N., Beckmann, R., Cate, J. H., Dinman, J. D., Dragon, F., Ellis, S. R., Lafontaine, D. L., Lindahl, L., Liljas, A., Lipton, J. M., McAlear, M. A., Moore, P. B., Noller, H. F., Ortega, J., Panse, V. G., Ramakrishnan, V., Spahn, C. M., Steitz, T. A., Tchorzewski, M., Tollervey, D., Warren, A. J., Williamson, J. R., Wilson, D., Yonath, A., and Yusupov, M. (2014). A new system for naming ribosomal proteins. *Current opinion in structural biology*, 24:165–169.

Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000a). The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science (New York, N.Y.)*, 289(5481):905–920.

Ban, N., Nissen, P., Moore, P. B., and Steitz, T. A. (2000b). Crystal Structure of the Large Ribosomal Subunit at 5-Angstrom Resolution. In *The Ribosome*, chapter 2, pages 11–20. John Wiley & Sons, Ltd.

Barber, C. B., Dobkin, D. P., and Huhdanpaa, H. (1996). The quickhull algorithm for convex hulls. *ACM Trans. Math. Softw.*, 22(4):469–483.

Basdevant, N., Dessaux, D., and Ramirez, R. (2019). Ionic transport through a protein nanopore: A Coarse-Grained Molecular Dynamics Study. *Scientific Reports*, 9(1):15740.

Bashan, A., Pioletti, M., Bartels, H., Janell, D., Schluenzen, F., Gluehmann, M., Levin, I., Harms, J., Hansen, H. A. S., Tocilji, A., Auerbach, T., Avila, H., Simitsopoulou, M., Peretz, M., Bennett, W. S., Agmon, I., Kessler, M., Weinstein, S., Franceschi, F., and Yonath, A. (2000). Identification of Selected Ribosomal Components in Crystallographic Maps of Prokaryotic Ribosomal Subunits at Medium Resolution. In *The Ribosome*, chapter 3, pages 21–33. John Wiley & Sons, Ltd.

Batebi, H., Pérez-Hernández, G., Rahman, S. N., Lan, B., Kamprad, A., Shi, M., Speck, D., Tiemann, J. K. S., Guixà-González, R., Reinhardt, F., Stadler, P. F., Papasergi-Scott, M. M., Skiniotis, G., Scheerer, P., Kobilka, B. K., Mathiesen, J. M., Liu, X., and Hildebrand, P. W. (2024). Mechanistic insights into G-protein coupling with an agonist-bound G-protein-coupled receptor. *Nature Structural & Molecular Biology*, 31(11):1692–1701.

Baum, L. E. and Petrie, T. (1966). Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *The Annals of Mathematical Statistics*, 37(6):1554–1563.

Baumeister, W., Grimm, R., and Walz, J. (1999). Electron tomography of molecules and cells. *Trends in Cell Biology*, 9(2):81–85.

Baxter, W. T., Grassucci, R. A., Gao, H., and Frank, J. (2009). Determination of Signal-to-Noise Ratios and Spectral SNRs in Cryo-EM Low-dose Imaging of Molecules. *Journal of structural biology*, 166(2):126–132.

Bayley, H. (2015). Nanopore sequencing: From imagination to reality. *Clinical Chemistry*, 61(1):25–31.

Beissinger, M. and Buchner, J. (1998). How chaperones fold proteins. *Biological Chemistry*, 379(3):245–259.

Bekker, H., Berendsen, H. J. C., Dijkstra, E., Achterop, S., van Drunen, R., Van Der Spoel, D., Sijbers, A., and Keegstra, H. (1993). Gromacs: A parallel computer for molecular dynamics simulations. *Physics Computing*, 92:252–256.

Ben-Shem, A., Jenner, L., Yusupova, G., and Yusupov, M. (2010). Crystal Structure of the Eukaryotic Ribosome. *Science*, 330(6008):1203–1209.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Wheeler, D. L. (2008). GenBank. *Nucleic Acids Research*, 36(suppl_1):D25–D30.

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

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8):3684–3690.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., and Hermans, J. (1981). Interaction Models for Water in Relation to Protein Hydration. In Pullman, B., editor, *Intermolecular Forces: Proceedings of the Fourteenth Jerusalem Symposium on Quantum Chemistry and Biochemistry Held in Jerusalem, Israel, April 13–16, 1981*, pages 331–342. Springer Netherlands, Dordrecht.

Berendsen, H. J. C., van der Spoel, D., and van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91(1):43–56.

Berezhkovskii, A. M. and Bezrukov, S. M. (2023). Counter-Intuitive Features of Particle Dynamics in Nanopores. *International Journal of Molecular Sciences*, 24(21):15923.

Berezovsky, I. N. and Shakhnovich, E. I. (2005). Physics and evolution of thermophilic adaptation. *Proceedings of the National Academy of Sciences*, 102(36):12742–12747.

Beringer, M. and Rodnina, M. V. (2007). The Ribosomal Peptidyl Transferase. *Molecular Cell*, 26(3):311–321.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, 28(1):235–242.

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

Bernick, D. L., Karplus, K., Lui, L. M., Coker, J. K. C., Murphy, J. N., Chan, P. P., Cozen, A. E., and Lowe, T. M. (2012). Complete genome sequence of Pyrobaculum oguniense. *Standards in Genomic Sciences*, 6(3):336–345.

Bernier, C. R., Petrov, A. S., Kovacs, N. A., Penev, P. I., and Williams, L. D. (2018). Translation: The Universal Structural Core of Life. *Molecular Biology and Evolution*, 35(8):2065–2076.

Bertoline, L. M. F., Lima, A. N., Krieger, J. E., and Teixeira, S. K. (2023). Before and after AlphaFold2: An overview of protein structure prediction. *Frontiers in Bioinformatics*, 3.

Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M., and MacKerell, A. D. (2012). Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone $\phi$, $\psi$ and Side-Chain $\chi 1$ and $\chi 2$ Dihedral Angles. *Journal of Chemical Theory and Computation*, 8(9):3257–3273.

Bharat, T. A., von Kügelgen, A., and Alva, V. (2021). Molecular Logic of Prokaryotic Surface Layer Structures. *Trends in Microbiology*, 29(5):405–415.

Bharat, T. A. M., Russo, C. J., Löwe, J., Passmore, L. A., and Scheres, S. H. W. (2015). Advances in Single-Particle Electron Cryomicroscopy Structure Determination applied to Sub-tomogram Averaging. *Structure*, 23(9):1743–1753.

Bharat, T. A. M. and Scheres, S. H. W. (2016). Resolving macromolecular structures from electron cryo-tomography data using subtomogram averaging in RELION. *Nature Protocols*, 11(11):2054–2065.

Biró, I., Pezeshki, S., Weingart, H., Winterhalter, M., and Kleinekathöfer, U. (2010). Comparing the Temperature-Dependent Conductance of the Two Structurally Similar E. coli Porins OmpC and OmpF. *Biophysical Journal*, 98(9):1830–1839.

Bock, L. V., Igaev, M., and Grubmüller, H. (2024). Single-particle Cryo-EM and molecular dynamics simulations: A perfect match. *Current Opinion in Structural Biology*, 86:102825.

Bock, L. V., Kolář, M. H., and Grubmüller, H. (2018). Molecular simulations of the ribosome and associated translation factors. *Current Opinion in Structural Biology*, 49:27–35.

Böhning, J. and Bharat, T. A. M. (2021). Towards high-throughput in situ structural biology using electron cryotomography. *Progress in Biophysics and Molecular Biology*, 160:97–103.

Boltzmann, L. (1868). *Studien über das Gleichgewicht der lebendigen Kraft zwischen bewegten materiellen Punkten: vorgelegt in der Sitzung am 8. October 1868*. k. und k. Hof- und Staatsdr.

Bond, P. J., Derrick, J. P., and Sansom, M. S. P. (2007). Membrane simulations of OpcA: Gating in the loops? *Biophysical Journal*, 92(2):L23–25.

Bond, P. J., Guy, A. T., Heron, A., Bayley, H., and Khalid, S. (2011). Molecular dynamics simulations of DNA within a nanopore: Arginine-phosphate tethering and a binding/sliding mechanism for translocation. *Biochemistry*, 50(18):3777–3783.

Bonetti, D., Troilo, F., Brunori, M., Longhi, S., and Gianni, S. (2018). How Robust Is the Mechanism of Folding-Upon-Binding for an Intrinsically Disordered Protein? *Biophysical Journal*, 114(8):1889–1894.

Bonhivers, M., Plançon, L., Ghazi, A., Boulanger, A., le Maire, M., Lambert, O., Rigaud, J. L., and Letellier, L. (1998). FhuA, an *Escherichia coli* outer membrane protein with a dual function of transporter and channel which mediates the transport of phage DNA. *Biochimie*, 80(5):363–369.

Bonilla, S. L. and Jang, K. (2024). Challenges, advances, and opportunities in RNA structural biology by Cryo-EM. *Current Opinion in Structural Biology*, 88:102894.

Booth, C. (2012). K2: A Super-Resolution Electron Counting Direct Detection Camera for Cryo-EM. *Microscopy and Microanalysis*, 18(S2):78–79.

Bouvette, J., Liu, H.-F., Du, X., Zhou, Y., Sikkema, A. P., da Fonseca Rezende e Mello, J., Klemm, B. P., Huang, R., Schaaper, R. M., Borgnia, M. J., and Bartesaghi, A. (2021). Beam image-shift accelerated data acquisition for near-atomic resolution single-particle cryo-electron tomography. *Nature Communications*, 12(1):1957.

Bowman, G. R. (2024). AlphaFold and Protein Folding: Not Dead Yet! The Frontier Is Conformational Ensembles. *Annual Review of Biomedical Data Science*, 7(Volume 7, 2024):51–57.

Bowman, J. C., Petrov, A. S., Frenkel-Pinter, M., Penev, P. I., and Williams, L. D. (2020). Root of the Tree: The Significance, Evolution, and Origins of the Ribosome. *Chemical Reviews*, 120(11):4848–4878.

Brandt, S., Heikkonen, J., and Engelhardt, P. (2001). Multiphase Method for Automatic Alignment of Transmission Electron Microscope Images Using Markers. *Journal of Structural Biology*, 133(1):10–22.

Briggs, J. A. (2013). Structural biology in situ—the potential of subtomogram averaging. *Current Opinion in Structural Biology*, 23(2):261–267.

Brimacombe, R., Greuer, B., Mueller, F., Osswald, M., Rinke-Appel, J., and Sommer, I. (2000). Three-Dimensional Organization of the Bacterial Ribosome and Its Subunits: Transition from Low-Resolution Models to High-Resolution Structures. In *The Ribosome*, chapter 14, pages 151–163. John Wiley & Sons, Ltd.

Brion, P. and Westhof, E. (1997). HIERARCHY AND DYNAMICS OF RNA FOLDING. *Annual Review of Biophysics*, 26(Volume 26, 1997):113–137.

Brooks, B. R., Brooks III, C. L., Mackerell Jr., A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. (2009). CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry*, 30(10):1545–1614.

Brown, C. G. and Clarke, J. (2016). Nanopore development at Oxford Nanopore. *Nature Biotechnology*, 34(8):810–811.

Brugè, F., Fornili, S. L., Malenkov, G. G., Palma-Vittorelli, M. B., and Palma, M. U. (1996). Solvent-induced forces on a molecular scale: Non-additivity, modulation and causal relation to hydration. *Chemical Physics Letters*, 254(5):283–291.

Bucciarelli, S., Sayedi, E. S., Osella, S., Trzaskowski, B., Vissing, K. J., Vestergaard, B., and Foderà, V. (2020). Disentangling the role of solvent polarity and protein solvation in folding and self-assembly of $\alpha$-lactalbumin. *Journal of Colloid and Interface Science*, 561:749–761.

Buchner, J. (2019). Molecular chaperones and protein quality control: An introduction to the JBC Reviews thematic series. *The Journal of Biological Chemistry*, 294(6):2074–2075.

Budnik, M., Wawrzyniak, J., Grala, Ł., Kadziński, M., and Szóstak, N. (2024). Deep dive into RNA: A systematic literature review on RNA structure prediction using machine learning methods. *Artificial Intelligence Review*, 57(9):254.

Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1):14101–09.

Butler, T. Z., Pavlenok, M., Derrington, I. M., Niederweis, M., and Gundlach, J. H. (2008). Single-molecule DNA detection with an engineered MspA protein nanopore. *Proceedings of the National Academy of Sciences of the United States of America*, 105(52):20647–20652.

Cao, B., Zhao, Y., Kou, Y., Ni, D., Zhang, X. C., and Huang, Y. (2014). Structure of the nonameric bacterial amyloid secretion channel. *Proceedings of the National Academy of Sciences of the United States of America*, 111(50):E5439–E5444.

Caparotta, M. and Perez, A. (2024). Advancing Molecular Dynamics: Toward Standardization, Integration, and Data Accessibility in Structural Biology. *The Journal of Physical Chemistry B*, 128(10):2219–2227.

Cardone, G., Heymann, J. B., and Steven, A. C. (2013). One number does not fit all: Mapping local variations in resolution in cryo-EM reconstructions. *Journal of structural biology*, 184(2):10.1016/j.jsb.2013.08.002.

Carroni, M. and Saibil, H. R. (2016). Cryo electron microscopy to determine the structure of macromolecular complexes. *Methods (San Diego, Calif.)*, 95:78–85.

Caspy, I., Wang, Z., and Bharat, T. A. (2025). Structural biology inside multicellular specimens using electron cryotomography. *Quarterly reviews of biophysics*, 58:e6.

Cates, M. S., Teodoro, M. L., and Phillips, G. N. (2002). Molecular mechanisms of calcium and magnesium binding to parvalbumin. *Biophysical Journal*, 82(3):1133–1146.

Cech, T. R. (2000). Structural biology. The ribosome is a ribozyme. *Science (New York, N.Y.)*, 289(5481):878–879.

Chen, C., Huang, H., and Wu, C. H. (2017). Protein Bioinformatics Databases and Resources. *Methods in Molecular Biology (Clifton, N.J.)*, 1558:3–39.

Chen, J., Liu, H., Cui, X., Li, Z., and Chen, H.-F. (2022). RNA-Specific Force Field Optimization with CMAP and Reweighting. *Journal of Chemical Information and Modeling*, 62(2):372–385.

Chen, M., Khalid, S., Sansom, M. S. P., and Bayley, H. (2008). Outer membrane protein G: Engineering a quiet pore for biosensing. *Proceedings of the National Academy of Sciences*, 105(17):6272–6277.

Cheng, Y. (2015). Single-particle cryo-EM at crystallographic resolution. *Cell*, 161(3):450–457.

Chérier, D., Patin, D., Blanot, D., Touzé, T., and Barreteau, H. (2021). The Biology of Colicin M and Its Orthologs. *Antibiotics*, 10(9):1109.

Chipot, C. (2024). Recent Advances in Simulation Software and Force Fields: Their Importance in Theoretical and Computational Chemistry and Biophysics. *The Journal of Physical Chemistry B*, 128(49):12023–12026.

Choi, T., Li, Z., Song, G., and Chen, H.-F. (2024). Comprehensive Comparison and Critical Assessment of RNA-Specific Force Fields. *Journal of Chemical Theory and Computation*, 20(6):2676–2688.

Choudhary, A., Maffeo, C., and Aksimentiev, A. (2022). Multi-resolution simulation of DNA transport through large synthetic nanostructures. *Physical Chemistry Chemical Physics*, 24(5):2706–2716.

Chua, E. Y., Mendez, J. H., Rapp, M., Ilca, S. L., Tan, Y. Z., Maruthi, K., Kuang, H., Zimanyi, C. M., Cheng, A., Eng, E. T., Noble, A. J., Potter, C. S., and Carragher, B. (2022). Better, Faster, Cheaper: Recent Advances in Cryo–Electron Microscopy. *Annual review of biochemistry*, 91:1–32.

Clapeyron, B. P. É. (1834). Mémoire sur la puissance motrice de la chaleur. *Journal de l'École polytechnique*, XIV:153–190.

Clarridge, J. E. (2004). Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases. *Clinical Microbiology Reviews*, 17(4):840–862.

Clausius, R. (1870). XVI. On a mechanical theorem applicable to heat. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 40(265):122–127.

Cooper, A. (2000). Heat capacity of hydrogen-bonded networks: An alternative view of protein folding thermodynamics. *Biophysical Chemistry*, 85(1):25–39.

Corcelli, A., Chong, P. L.-G., and Koga, Y. (2012). Lipid Biology of Archaea. *Archaea*, 2012(1):710836.

Corin, K. and Bowie, J. U. (2022). How physical forces drive the process of helical membrane protein folding. *EMBO Reports*, 23(3):e53025.

Coureux, P.-D., Lazennec-Schurdevin, C., Bourcier, S., Mechulam, Y., and Schmitt, E. (2020). Cryo-EM study of an archaeal 30S initiation complex gives insights into evolution of translation initiation. *Communications Biology*, 3:58.

Coureux, P.-D., Lazennec-Schurdevin, C., Monestier, A., Larquet, E., Cladière, L., Klaholz, B. P., Schmitt, E., and Mechulam, Y. (2016). Cryo-EM study of start codon selection during archaeal translation initiation. *Nature Communications*, 7:13366.

Croll, T. I. (2018). ISOLDE: A physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallographica Section D: Structural Biology*, 74(6):519–530.

Crowther, R. A., Amos, L. A., Finch, J. T., De Rosier, D. J., and Klug, A. (1970). Three dimensional reconstructions of spherical viruses by fourier synthesis from electron micrographs. *Nature*, 226(5244):421–425.

Cui, T., Zhou, Y., and Wang, T. (2025). Recent advances in artificial intelligence–driven biomolecular dynamics simulations based on machine learning force fields. *Current Opinion in Structural Biology*, 95:103191.

Dance, A. (2021). The mysterious microbes that gave rise to complex life. *Nature*, 593(7859):328–330.

Darden, T., Perera, L., Li, L., and Pedersen, L. (1999). New tricks for modelers from the crystallography toolkit: The particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure*, 7(3):R55–R60.

Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092.

Daume, M., Uhl, M., Backofen, R., and Randau, L. (2017). RIP-Seq Suggests Translational Regulation by L7Ae in Archaea. *mBio*, 8(4):e00730–17.

Davisson, C. and Germer, L. H. (1927). Diffraction of Electrons by a Crystal of Nickel. *Physical Review*, 30(6):705–740.

Dayhoff, M. O., Schwartz, R., and Orcutt, B. (1979). A model of evolutionary change in proteins. In *Atlas of Protein Sequence and Structure*, volume 5, Supplement 3. National Biomedical Research Foundation, Silver Spring [Md.

de Boer, P., Hoogenboom, J. P., and Giepmans, B. N. G. (2015). Correlated light and electron microscopy: Ultrastructure lights up! *Nature Methods*, 12(6):503–513.

de Broglie, L. (1924). XXXV. A tentative theory of light quanta. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 47(278):446–458.

de Coulomb, C.-A. (1788). *Premier Mémoire sur L'Électricité et le Magnétisme*. De l'imprimerie royale.

de Courcy, B., Pedersen, L. G., Parisel, O., Gresh, N., Silvi, B., Pilmé, J., and Piquemal, J.-P. (2010). Understanding selectivity of hard and soft metal cations within biological systems using the subvalence concept. I. Application to blood coagulation: Direct cation-protein electronic effects vs. indirect interactions through water networks. *Journal of chemical theory and computation*, 6(4):1048–1063.

de Jong, D. H., Singh, G., Bennett, W. F. D., Arnarez, C., Wassenaar, T. A., Schäfer, L. V., Periole, X., Tieleman, D. P., and Marrink, S. J. (2013). Improved Parameters for the Martini Coarse-Grained Protein Force Field. *Journal of Chemical Theory and Computation*, 9(1):687–697.

De Rosier, D. J. and Klug, A. (1968). Reconstruction of Three Dimensional Structures from Electron Micrographs. *Nature*, 217(5124):130–134.

Deamer, D., Akeson, M., and Branton, D. (2016). Three decades of nanopore sequencing. *Nature biotechnology*, 34(5):518–524.

Deka, A. (2023). Analytical Strategies in RNA Purification. pages 95–127.

Dellago, C. and Hummer, G. (2014). Computing Equilibrium Free Energies Using Non-Equilibrium Molecular Dynamics. *Entropy*, 16(1):41–61.

Deuerling, E., Gamerdinger, M., and Kreft, S. G. (2019). Chaperone Interactions at the Ribosome. *Cold Spring Harbor Perspectives in Biology*, 11(11):a033977.

Dickerson, J. L., Lu, P.-H., Hristov, D., Dunin-Borkowski, R. E., and Russo, C. J. (2022). Imaging biological macromolecules in thick specimens: The role of inelastic scattering in cryoEM. *Ultramicroscopy*, 237:113510.

Dickerson, J. L., Naydenova, K., Peet, M. J., Wilson, H., Nandy, B., McMullan, G., Morrison, R., and Russo, C. J. (2025). Reducing the effects of radiation damage in cryo-EM using liquid helium temperatures. *Proceedings of the National Academy of Sciences*, 122(17):e2421538122.

Dill, K. A. (1985). Theory for the folding and stability of globular proteins. *Biochemistry*, 24(6):1501–1509.

Dill, K. A. (1990). Dominant forces in protein folding. *Biochemistry*, 29(31):7133–7155.

Dobson, C. M. and Karplus, M. (1999). The fundamentals of protein folding: Bringing together theory and experiment. *Current Opinion in Structural Biology*, 9(1):92–101.

Doolittle, W. F. (2020). Evolution: Two Domains of Life or Three? *Current Biology*, 30(4):R177–R179.

Dorman, C. J. and Dorman, M. J. (2016). DNA supercoiling is a fundamental regulatory principle in the control of bacterial gene expression. *Biophysical Reviews*, 8(3):209–220.

Dror, R. O., Dirks, R. M., Grossman, J. P., Xu, H., and Shaw, D. E. (2012). Biomolecular simulation: A computational microscope for molecular biology. *Annual Review of Biophysics*, 41:429–452.

Dubochet, J., Adrian, M., Chang, J. J., Homo, J. C., Lepault, J., McDowall, A. W., and Schultz, P. (1988). Cryo-electron microscopy of vitrified specimens. *Quarterly Reviews of Biophysics*, 21(2):129–228.

Duprey, A. and Groisman, E. A. (2021). The regulation of DNA supercoiling across evolution. *Protein Science : A Publication of the Protein Society*, 30(10):2042–2056.

Dyson, H. J. and Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology*, 6(3):197–208.

Edmondson, S. P., Turri, J., Smith, K., Clark, A., and Shriver, J. W. (2009). Structure, Stability, and Flexibility of Ribosomal Protein L14e from Sulfolobus Solfataricus. *Biochemistry*, 48(24):5553–5562.

Egberts, E., Marrink, S. J., and Berendsen, H. J. (1994). Molecular dynamics simulation of a phospholipid membrane. *European biophysics journal: EBJ*, 22(6):423–436.

Egerton, R. F. (2014). Choice of operating voltage for a transmission electron microscope. *Ultramicroscopy*, 145:85–93.

Eisenhaber, F., Lijnzaad, P., Argos, P., Sander, C., and Scharf, M. (1995). The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *Journal of Computational Chemistry*, 16(3):273–284.

Elfmann, C. and Stülke, J. (2023). PAE viewer: A webserver for the interactive visualization of the predicted aligned error for multimer structure predictions and crosslinks. *Nucleic Acids Research*, 51(W1):W404–W410.

Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, 66(4):486–501.

Endriß, F. and Braun, V. (2004). Loop Deletions Indicate Regions Important for FhuA Transport and Receptor Functions in Escherichia coli. *Journal of Bacteriology*, 186(14):4818–4823.

Engelhardt, H. (2007). Are S-layers exoskeletons? The basic function of protein surface layers revisited. *Journal of Structural Biology*, 160(2):115–124.

Erath, J., Kemper, D., Mugo, E., Jacoby, A., Valenzuela, E., Jungers, C. F., Beatty, W. L., Hashem, Y., Jovanovic, M., Djuranovic, S., and Pavlovic Djuranovic, S. (2025). A rapid, simple, and economical method for the isolation of ribosomes and translational machinery for structural and functional studies. *Nature Communications*, 16(1):7185.

Erickson, H. P. and Klug, A. (1971). Measurement and Compensation of Defocusing and Aberrations by Fourier Processing of Electron Micrographs. *Philosophical Transactions of the Royal Society of London Series B*, 261:105–118.

Ewald, P. P. (1921). Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik*, 369(3):253–287.

Feathers, J. R., Spoth, K. A., and Fromme, J. C. (2021). Experimental evaluation of super-resolution imaging and magnification choice in single-particle cryo-EM. *Journal of Structural Biology: X*, 5:100047.

Feng, Y., Song, X., Lin, J., Xuan, J., Cui, Q., and Wang, J. (2014). Structure determination of archaea-specific ribosomal protein L46a reveals a novel protein fold. *Biochemical and Biophysical Research Communications*, 450(1):67–72.

Fernandez, J.-J., Li, S., and Agard, D. A. (2019). Consideration of sample motion in cryo-tomography based on alignment residual interpolation. *Journal of Structural Biology*, 205(3):1–6.

Fernandez, J.-J., Li, S., Bharat, T. A. M., and Agard, D. A. (2018). Cryo-tomography tilt-series alignment with consideration of the beam-induced sample motion. *Journal of Structural Biology*, 202(3):200–209.

Fernández-Fernández, J., Martín-Villanueva, S., Perez-Fernandez, J., and de la Cruz, J. (2023). The Role of Ribosomal Proteins eL15 and eL36 in the Early Steps of Yeast 60S Ribosomal Subunit Assembly. *Journal of Molecular Biology*, 435(24):168321.

Fernandez-Leiro, R. and Scheres, S. H. W. (2017). A pipeline approach to single-particle processing in RELION. *Acta Crystallographica Section D: Structural Biology*, 73(6):496–502.

Fersht, A. R. (2008). From the first protein structures to our current knowledge of protein folding: Delights and scepticisms. *Nature Reviews Molecular Cell Biology*, 9(8):650–654.

Finney, J. L., Gellatly, B. J., Golton, I. C., and Goodfellow, J. (1980). Solvent effects and polar interactions in the structural stability and dynamics of globular proteins. *Biophysical Journal*, 32(1):17–33.

Fogolari, F., Brigo, A., and Molinari, H. (2002). The Poisson–Boltzmann equation for biomolecular electrostatics: A tool for structural biology. *Journal of Molecular Recognition*, 15(6):377–392.

Förster, F. and Hegerl, R. (2007). Structure Determination In Situ by Averaging of Tomograms. In *Methods in Cell Biology*, volume 79, pages 741–767. Academic Press.

Förster, F., Pruggnaller, S., Seybert, A., and Frangakis, A. S. (2008). Classification of cryo-electron sub-tomograms using constrained correlation. *Journal of Structural Biology*, 161(3):276–286.

Forterre, P. (2015). The universal tree of life: An update. *Frontiers in Microbiology*, 6.

Frank, J. (2002). Single-particle imaging of macromolecules by cryo-electron microscopy. *Annual Review of Biophysics and Biomolecular Structure*, 31:303–319.

Frank, J. (2006). *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State*. Oxford University Press.

Frank, J., Penczek, P., Grassucci, R. A., Heagle, A., Spahn, C. M. T., and Agrawal, R. K. (2000). Cryo-Electron Microscopy of the Translational Apparatus: Experimental Evidence for the Paths of mRNA, tRNA, and the Polypeptide Chain. In *The Ribosome*, chapter 5, pages 45–51. John Wiley & Sons, Ltd.

Frank, J., Zhu, J., Penczek, P., Li, Y., Srivastava, S., Verschoor, A., Radermacher, M., Grassucci, R., Lata, R. K., and Agrawal, R. K. (1995). A model of protein synthesis based on cryo-electron microscopy of the E. coli ribosome. *Nature*, 376(6539):441–444.

Frauenfelder, H. and Wolynes, P. G. (1994). Biomolecules: Where the Physics of Complexity and Simplicity Meet. *Physics Today*, 47(2):58–64.

Frenkel, J. (1930). On the Correct Formulation of Pauli's Exclusion Principle. *Nature*, 125(3146):235–236.

Fujii, K., Susanto, T. T., Saurabh, S., and Barna, M. (2018). Decoding the Function of Expansion Segments in Ribosomes. *Molecular cell*, 72(6):1013–1020.e6.

Fung, J. C., Liu, W., de Ruijter, W. J., Chen, H., Abbey, C. K., Sedat, J. W., and Agard, D. A. (1996). Toward Fully Automated High-Resolution Electron Tomography. *Journal of Structural Biology*, 116(1):181–189.

Gabdulkhakov, A. G., Mitroshin, I. V., and Garber, M. B. (2024). CRYSTAL STRUCTURE OF THE BASE OF THE RIBOSOMAL P STALK FROM METHANOCOCCUS JANNASCHII.

Gamalinda, M. and Woolford Jr, J. L. (2015). Paradigms of ribosome synthesis: Lessons learned from ribosomal proteins. *Translation*, 3(1):e975018.

Gambelli, L., McLaren, M., Conners, R., Sanders, K., Gaines, M. C., Clark, L., Gold, V. A., Kattnig, D., Sikora, M., Hanus, C., Isupov, M. N., and Daum, B. (2024). Structure of the two-component S-layer of the archaeon Sulfolobus acidocaldarius. *eLife*, 13:e84617.

Gambelli, L., Meyer, B. H., McLaren, M., Sanders, K., Quax, T. E. F., Gold, V. A. M., Albers, S.-V., and Daum, B. (2019). Architecture and modular assembly of Sulfolobus S-layers revealed by electron cryotomography. *Proceedings of the National Academy of Sciences*, 116(50):25278–25286.

Garcia-Campos, M. A., Nir, R., Reichelt, R., Fluke, K. A., Matzov, D., Iyer, V., Burkhart, B. W., Lui, L., Kustanovich, A., Grünberger, F., Thalalla-Gamage, S., Howpay-Manage, S. A., Gerovac, M., Alexandre, N., Nobe, Y., Nowak, J. S., Perera, M., Apostle, A., Fang, S., Glatt, S., Jona, G., Ferreira-Cerca, S., Vogel, J., Taoka, M., Meier, J. L., Westhof, E., Santangelo, T. J., Grohmann, D., Shalev-Benami, M., and Schwartz, S. (2025). Pan-modification profiling facilitates a cross-evolutionary dissection of the thermoregulated ribosomal epitranscriptome. *Cell*, pages S0092–8674(25)01082–7.

Gianni, S., Morrone, A., Giri, R., and Brunori, M. (2012). A folding-after-binding mechanism describes the recognition between the transactivation domain of c-Myb and the KIX domain of the CREB-binding protein. *Biochemical and Biophysical Research Communications*, 428(2):205–209.

Gill, R. L., Castaing, J.-P., Hsin, J., Tan, I. S., Wang, X., Huang, K. C., Tian, F., and Ramamurthi, K. S. (2015). Structural basis for the geometry-driven localization of a small protein. *Proceedings of the National Academy of Sciences*, 112(15):E1908–E1915.

Gilles, M. A. and Singer, A. (2025). Cryo-EM heterogeneity analysis using regularized covariance estimation and kernel regression. *Proceedings of the National Academy of Sciences*, 122(9):e2419140122.

Glaeser, R. M. (1971). Limitations to significant information in biological electron microscopy as a result of radiation damage. *Journal of Ultrastructure Research*, 36(3):466–482.

Goddard, T. D., Huang, C. C., and Ferrin, T. E. (2007). Visualizing density maps with UCSF Chimera. *Journal of Structural Biology*, 157(1):281–287.

Goddard, T. D., Huang, C. C., Meng, E. C., Pettersen, E. F., Couch, G. S., Morris, J. H., and Ferrin, T. E. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science: A Publication of the Protein Society*, 27(1):14–25.

Goldberg, M. (1937). A Class of Multi-Symmetric Polyhedra. *Tohoku Mathematical Journal*.

Goloubinoff, P., Sassi, A. S., Fauvet, B., Barducci, A., and De Los Rios, P. (2018). Chaperones convert the energy from ATP into the nonequilibrium stabilization of native proteins. *Nature Chemical Biology*, 14(4):388–395.

Göpfrich, K., Li, C.-Y., Mames, I., Bhamidimarri, S. P., Ricci, M., Yoo, J., Mames, A., Ohmann, A., Winterhalter, M., Stulz, E., Aksimentiev, A., and Keyser, U. F. (2016). Ion Channels Made from a Single Membrane-Spanning DNA Duplex. *Nano Letters*, 16(7):4665–4669.

Gowers, R. J., Linke, M., Barnoud, J., Reddy, T. J. E., Melo, M. N., Seyler, S. L., Domański, J., Dotson, D. L., Buchoux, S., Kenney, I. M., and Beckstein, O. (2016). MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. *scipy*.

Grandori, R. (2023). Protein structure and dynamics in the era of integrative structural biology. *Frontiers in Biophysics*, 1.

Greber, B. J., Boehringer, D., Godinic-Mikulcic, V., Crnkovic, A., Ibba, M., Weygand-Durasevic, I., and Ban, N. (2012). Cryo-EM structure of the archaeal 50S ribosomal subunit in complex with initiation factor 6 and implications for ribosome evolution. *Journal of Molecular Biology*, 418(3-4):145–160.

Green, R. and Noller, H. F. (1997). RIBOSOMES AND TRANSLATION. *Annual Review of Biochemistry*, 66(Volume 66, 1997):679–716.

Greenacre, M., Groenen, P. J. F., Hastie, T., D'Enza, A. I., Markos, A., and Tuzhilina, E. (2022). Principal component analysis. *Nature Reviews Methods Primers*, 2(1):100.

Gribaldo, S. and Brochier-Armanet, C. (2006). The origin and evolution of Archaea: A state of the art. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1470):1007–1022.

Grill-Walcher, S. and Schäffer, C. (2025). A new age in structural S-layer biology: Experimental and in silico milestones. *The Journal of Biological Chemistry*, 301(6):110205.

Grotz, K. K., Cruz-León, S., and Schwierz, N. (2021). Optimized Magnesium Force Field Parameters for Biomolecular Simulations with Accurate Solvation, Ion-Binding, and Water-Exchange Properties. *Journal of Chemical Theory and Computation*, 17(4):2530–2540.

Grover, N. (2015). On Using Magnesium and Potassium Ions in RNA Experiments. In Carmichael, G. G., editor, *Regulatory Non-Coding RNAs: Methods and Protocols*, pages 157–163. Springer, New York, NY.

Guth-Metzler, R., Mohamed, A. M., Cowan, E. T., Henning, A., Ito, C., Frenkel-Pinter, M., Wartell, R. M., Glass, J. B., and Williams, L. D. (2023). Goldilocks and RNA: Where Mg2+ concentration is just right. *Nucleic Acids Research*, 51(8):3529–3539.

Gutmanas, A., Oldfield, T. J., Patwardhan, A., Sen, S., Velankar, S., and Kleywegt, G. J. (2013). The role of structural bioinformatics resources in the era of integrative structural biology. *Acta Crystallographica Section D: Biological Crystallography*, 69(5):710–721.

Hagen, W. J. H., Wan, W., and Briggs, J. A. G. (2017). Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *Journal of Structural Biology*, 197(2):191–198.

Halder, A., Data, D., Seelam, P. P., Bhattacharyya, D., and Mitra, A. (2019). Estimating Strengths of Individual Hydrogen Bonds in RNA Base Pairs: Toward a Consensus between Different Computational Approaches. *ACS Omega*, 4(4):7354–7368.

Hamaguchi, T., Maki-Yonekura, S., Naitow, H., Matsuura, Y., Ishikawa, T., and Yonekura, K. (2019). A new cryo-EM system for single particle analysis. *Journal of Structural Biology*, 207(1):40–48.

Harauz, G. and Van Heel, M. (1986). Exact filters for general geometry three dimensional reconstruction. *Optik*, 73(4):146–156.

Harding, M. M. (2004). The architecture of metal coordination groups in proteins. *Acta Crystallographica Section D: Biological Crystallography*, 60(5):849–859.

Hart, R. G. (1968). Electron microscopy of unstained biological material: The polytropic montage. *Science (New York, N.Y.)*, 159(3822):1464–1467.

Hartl, F. U., Bracher, A., and Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. *Nature*, 475(7356):324–332.

Hartman, A. L., Norais, C., Badger, J. H., Delmas, S., Haldenby, S., Madupu, R., Robinson, J., Khouri, H., Ren, Q., Lowe, T. M., Maupin-Furlow, J., Pohlschroder, M., Daniels, C., Pfeiffer, F., Allers, T., and Eisen, J. A. (2010). The Complete Genome Sequence of Haloferax volcanii DS2, a Model Archaeon. *PLoS ONE*, 5(3).

Harvey, S. C., Prabhakaran, M., and McCammon, J. A. (1985). Molecular-dynamics simulation of phenylalanine transfer RNA. I. Methods and general results. *Biopolymers*, 24(7):1169–1188.

Hassabis, D. (2024). Accelerating Scientific Discovery with AI. *The Keio Journal of Medicine*, 73(4):47–47.

Hassan, A., Byju, S., Freitas, F. C., Roc, C., Pender, N., Nguyen, K., Kimbrough, E. M., Mattingly, J. M., Gonzalez Jr., R. L., de Oliveira, R. J., Dunham, C. M., and Whitford, P. C. (2023). Ratchet, swivel, tilt and roll: A complete description of subunit rotation in the ribosome. *Nucleic Acids Research*, 51(2):919–934.

Hayne, C. K., Sekulovski, S., Hurtig, J. E., Stanley, R. E., Trowitzsch, S., and van Hoof, A. (2023). New insights into RNA processing by the eukaryotic tRNA splicing endonuclease. *Journal of Biological Chemistry*, 299(9).

Helena-Bueno, K., Rybak, M. Y., Ekemezie, C. L., Sullivan, R., Brown, C. R., Dingwall, C., Baslé, A., Schneider, C., Connolly, J. P. R., Blaza, J. N., Csörgő, B., Moynihan, P. J., Gagnon, M. G., Hill, C. H., and Melnikov, S. V. (2024). A new family of bacterial ribosome hibernation factors. *Nature*, 626(8001):1125–1132.

Henderson, R. (1995). The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Quarterly Reviews of Biophysics*, 28(2):171–193.

Henderson, R. (2013). Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proceedings of the National Academy of Sciences*, 110(45):18037–18041.

Henderson, R. (2017). From electron crystallography to single-particle cryo-EM. *Acta Crystallographica Section A Foundations and Advances*, 73(a2):C41–C41.

Henderson, R., Sali, A., Baker, M. L., Carragher, B., Devkota, B., Downing, K. H., Egelman, E. H., Feng, Z., Frank, J., Grigorieff, N., Jiang, W., Ludtke, S. J., Medalia, O., Penczek, P. A., Rosenthal, P. B., Rossmann, M. G., Schmid, M. F., Schröder, G. F., Steven, A. C., Stokes, D. L., Westbrook, J. D., Wriggers, W., Yang, H., Young, J., Berman, H. M., Chiu, W., Kleywegt, G. J., and Lawson, C. L. (2012). Outcome of the First Electron Microscopy Validation Task Force Meeting. *Structure*, 20(2):205–214.

Hendrick, J. P. and Hartl, F. U. (1995). The role of molecular chaperones in protein folding. *FASEB journal: official publication of the Federation of American Societies for Experimental Biology*, 9(15):1559–1569.

Heng, J. B., Aksimentiev, A., Ho, C., Marks, P., Grinkova, Y. V., Sligar, S., Schulten, K., and Timp, G. (2006). The Electromechanics of DNA in a Synthetic Nanopore. *Biophysical Journal*, 90(3):1098–1106.

Herdman, M., von Kügelgen, A., Kureisaite-Ciziene, D., Duman, R., El Omari, K., Garman, E. F., Kjaer, A., Kolokouris, D., Löwe, J., Wagner, A., Stansfeld, P. J., and Bharat, T. A. (2022). High-resolution mapping of metal ions reveals principles of surface layer assembly in Caulobacter crescentus cells. *Structure(London, England:1993)*, 30(2):215–228.e5.

Hermann, T. and Patel, D. J. (2000). RNA bulges as architectural and recognition motifs. *Structure*, 8(3):R47–R54.

Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1998). LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472.

Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008). GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447.

Hirst, I. J., Thomas, W. J., Davies, R. A., and Muench, S. P. (2024). CryoEM grid preparation: A closer look at advancements and impact of preparation mode and new approaches. *Biochemical Society Transactions*, 52(3):1529–1537.

Hockney, R. (1970). The potential calculation and some applications. In *Plasma Physics*, volume 9 of *Methods in Computational Physics*, pages 135–211. Academic Press, New York.

Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, M., and Schuster, P. (1994). Fast folding and comparison of RNA secondary structures. *Monatshefte für Chemie / Chemical Monthly*, 125(2):167–188.

Hollingsworth, S. A. and Dror, R. O. (2018). Molecular Dynamics Simulation for All. *Neuron*, 99(6):1129–1143.

Hong, Y., Song, Y., Zhang, Z., and Li, S. (2023). Cryo-Electron Tomography: The Resolution Revolution and a Surge of In Situ Virological Discoveries. *Annual Review of Biophysics*, 52:339–360.

Hoogsteen, K. (1963). The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallographica*, 16(9):907–916.

Hooke, R. (1678). *De Potentia Restitutiva*. John Martyn.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(7):498–520.

Houwink, A. L. (1956). Flagella, gas vacuoles and cell-wall structure in Halobacterium halobium; an electron microscope study. *Journal of General Microbiology*, 15(1):146–150.

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B. L., Grubmüller, H., and MacKerell Jr, A. D. (2017). CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nature Methods*, 14(1):71–73.

Huang, S., Aleksashin, N. A., Loveland, A. B., Klepacki, D., Reier, K., Kefi, A., Szal, T., Remme, J., Jaeger, L., Vázquez-Laslop, N., Korostelev, A. A., and Mankin, A. S. (2020). Ribosome engineering reveals the importance of 5S rRNA autonomy for ribosome assembly. *Nature Communications*, 11(1):2900.

Huber, R., Sacher, M., Vollmann, A., Huber, H., and Rose, D. (2000). Respiration of Arsenate and Selenate by Hyperthermophilic Archaea. *Systematic and Applied Microbiology*, 23(3):305–314.

Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., Butterfield, C. N., Hernsdorf, A. W., Amano, Y., Ise, K., Suzuki, Y., Dudek, N., Relman, D. A., Finstad, K. M., Amundson, R., Thomas, B. C., and Banfield, J. F. (2016). A new view of the tree of life. *Nature Microbiology*, 1(5):16048.

Hull, C. M. and Bevilacqua, P. C. (2016). Discriminating Self and Non-Self by RNA: Roles for RNA Structure, Misfolding, and Modification in Regulating the Innate Immune Sensor PKR. *Accounts of Chemical Research*, 49(6):1242–1249.

Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., De Castro, E., Langendijk-Genevaux, P. S., Pagni, M., and Sigrist, C. J. A. (2006). The PROSITE database. *Nucleic Acids Research*, 34(suppl_1):D227–D230.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38.

Hunter, C. A. and Sanders, J. K. M. (1990). The nature of .pi.-.pi. interactions. *Journal of the American Chemical Society*, 112(14):5525–5534.

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science Engineering*, 9(3):90–95.

Hur, S. (2019). Double-Stranded RNA Sensors and Modulators in Innate Immunity. *Annual review of immunology*, 37:349–375.

Isbilir, B., von Kügelgen, A., Alva, V., and Bharat, T. A. M. (2025). Assembly, architecture and functional roles of microbial surface layers. *Nature Reviews Microbiology*, pages 1–15.

Ithurbide, S., Buan, N., and Schulze, S. (2025). Advancing archaeal research through FAIR resource and data sharing, and inclusive community building. *Communications Biology*, 8(1):519.

Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S., and Miyata, T. (1989). Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proceedings of the National Academy of Sciences of the United States of America*, 86(23):9355–9359.

Jamali, K., Käll, L., Zhang, R., Brown, A., Kimanius, D., and Scheres, S. H. W. (2024). Automated model building and protein identification in cryo-EM maps. *Nature*, 628(8007):450–457.

Jamali, K., Kimanius, D., and Scheres, S. (2022). ModelAngelo: Automated Model Building in Cryo-EM Maps.

Jiang, J. S. and Brünger, A. T. (1994). Protein hydration observed by X-ray diffraction. Solvation properties of penicillopepsin and neuraminidase crystal structures. *Journal of Molecular Biology*, 243(1):100–115.

Jing, H., Takagi, J., Liu, J.-h., Lindgren, S., Zhang, R.-g., Joachimiak, A., Wang, J.-h., and Springer, T. A. (2002). Archaeal Surface Layer Proteins Contain $\beta$ Propeller, PKD, and $\beta$ Helix Domains and Are Related to Metazoan Cell Surface Proteins. *Structure*, 10(10):1453–1464.

Jing, Z., Liu, C., Qi, R., and Ren, P. (2018). Many-body effect determines the selectivity for Ca2+ and Mg2+ in proteins. *Proceedings of the National Academy of Sciences*, 115(32):E7495–E7501.

Jo, S., Kim, T., and Im, W. (2007). Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLOS ONE*, 2(9):e880.

Jo, S., Kim, T., Iyer, V. G., and Im, W. (2008). CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry*, 29(11):1859–1865.

Jobe, A., Liu, Z., Gutierrez-Vargas, C., and Frank, J. (2019). New Insights into Ribosome Structure and Function. *Cold Spring Harbor Perspectives in Biology*, 11(1):a032615.

Johnson, J. S., Spakowicz, D. J., Hong, B.-Y., Petersen, L. M., Demkowicz, P., Chen, L., Leopold, S. R., Hanson, B. M., Agresta, H. O., Gerstein, M., Sodergren, E., and Weinstock, G. M. (2019). Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nature Communications*, 10(1):5029.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T. L. (2008). NCBI BLAST: A better web interface. *Nucleic Acids Research*, 36(Web Server issue):W5–9.

Johnston, E., Isbilir, B., Alva, V., Bharat, T. A. M., and Doye, J. P. K. (2024). Punctuated and continuous structural diversity of S-layers across the prokaryotic tree of life.

Jones, D. T., Taylor, W. R., and Thornton, J. M. (1994). A mutation data matrix for transmembrane proteins. *FEBS letters*, 339(3):269–275.

Jones, J. E. (1924). On the determination of molecular fields.—I. From the variation of the viscosity of a gas with temperature. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 106(738):441–462.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589.

Jung, J., Kobayashi, C., Kasahara, K., Tan, C., Kuroda, A., Minami, K., Ishiduki, S., Nishiki, T., Inoue, H., Ishikawa, Y., Feig, M., and Sugita, Y. (2021). New parallel computing algorithm of molecular dynamics for extremely huge scale biological systems. *Journal of Computational Chemistry*, 42(4):231–241.

Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., Brookes, D. H., Wilson, L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D. W., Dolinsky, T., Konecny, R., Koes, D. R., Nielsen, J. E., Head-Gordon, T., Geng, W., Krasny, R., Wei, G.-W., Holst, M. J., McCammon, J. A., and Baker, N. A. (2018). Improvements to the APBS biomolecular solvation software suite. *Protein Science*, 27(1):112–128.

Kang, J., Brajanovski, N., Chan, K. T., Xuan, J., Pearson, R. B., and Sanij, E. (2021). Ribosomal proteins and human diseases: Molecular mechanisms and targeted therapy. *Signal Transduction and Targeted Therapy*, 6(1):323.

Karplus, M. and Petsko, G. A. (1990). Molecular dynamics simulations in biology. *Nature*, 347(6294):631–639.

Kasianowicz, J. J., Brandin, E., Branton, D., and Deamer, D. W. (1996). Characterization of individual polynucleotide molecules using a membrane channel. *Proceedings of the National Academy of Sciences of the United States of America*, 93(24):13770–13773.

Kasper, J. R. and Park, C. (2015). Ligand binding to a high-energy partially unfolded protein. *Protein Science : A Publication of the Protein Society*, 24(1):129–137.

Kates, M., Wassef, M. K., and Kushner, D. J. (1968). Radioisotopic studies on the biosynthesis of the glyceryl diether lipids of Halobacterium cutirubrum. *Canadian Journal of Biochemistry*, 46(8):971–977.

Kazan, R., Bourgeois, G., Lazennec-Schurdevin, C., Larquet, E., Mechulam, Y., Coureux, P.-D., and Schmitt, E. (2022). Role of aIF5B in archaeal translation initiation. *Nucleic Acids Research*, 50(11):6532–6548.

Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H., and Phillips, D. C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181(4610):662–666.

Khalid, S., Brandner, A. F., Juraschko, N., Newman, K. E., Pedebos, C., Prakaash, D., Smith, I. P. S., Waller, C., and Weerakoon, D. (2023). Computational microbiology of bacteria: Advancements in molecular dynamics simulations. *Structure*, 31(11):1320–1327.

Khrenova, M. G. (2025). Special Issue: "Advanced Research on Molecular Modeling of Protein Structure and Functions". *International Journal of Molecular Sciences*, 26(16):7916.

Kimanius, D., Dong, L., Sharov, G., Nakane, T., and Scheres, S. H. W. (2021). New tools for automated cryo-EM single-particle analysis in RELION-4.0. *The Biochemical Journal*, 478(24):4169–4185.

Kimanius, D. and Schwab, J. (2024). Confronting heterogeneity in cryogenic electron microscopy data: Innovative strategies and future perspectives with data-driven methods. *Current Opinion in Structural Biology*, 86:102815.

Kishore, D., Kundu, S., and Kayastha, A. M. (2012). Thermal, Chemical and pH Induced Denaturation of a Multimeric $\beta$-Galactosidase Reveals Multiple Unfolding Pathways. *PLoS ONE*, 7(11):e50380.

Kisly, I., Gulay, S. P., Mäeorg, U., Dinman, J. D., Remme, J., and Tamm, T. (2016). The functional role of eL19 and eB12 intersubunit bridge in the eukaryotic ribosome. *Journal of molecular biology*, 428(10 Pt B):2203–2216.

Kisly, I. and Tamm, T. (2023). Archaea/eukaryote-specific ribosomal proteins - guardians of a complex structure. *Computational and Structural Biotechnology Journal*, 21:1249–1261.

Klein, D. J., Moore, P. B., and Steitz, T. A. (2004). The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *Journal of Molecular Biology*, 340(1):141–177.

Kleinschmidt, J. H. (2015). Folding of $\beta$-barrel membrane proteins in lipid bilayers — Unassisted and assisted folding and insertion. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1848(9):1927–1943.

Klingl, A., Pickl, C., and Flechsler, J. (2019). Archaeal Cell Walls. *Sub-Cellular Biochemistry*, 92:471–493.

Knapek, E. and Dubochet, J. (1980). Beam damage to organic material is considerably reduced in cryo-electron microscopy. *Journal of Molecular Biology*, 141(2):147–161.

Knappeová, B., Mlýnský, V., Pykal, M., Šponer, J., Banáš, P., Otyepka, M., and Krepl, M. (2024). Comprehensive Assessment of Force-Field Performance in Molecular Dynamics Simulations of DNA/RNA Hybrid Duplexes. *Journal of Chemical Theory and Computation*, 20(15):6917–6929.

Knauer, V., Hegerl, R., and Hoppe, W. (1983). Three-dimensional reconstruction and averaging of 30 S ribosomal subunits of *Escherichia coli* from electron micrographs. *Journal of Molecular Biology*, 163(3):409–430.

Knudsen, M. and Wiuf, C. (2010). The CATH database. *Human Genomics*, 4(3):207–212.

Konings, S. and Bischoff, M. (2020). A Cold Field Emission Gun Optimized for Cryo-EM Applications. *Microscopy and Microanalysis*, 26(S2):566–567.

Könneke, M., Bernhard, A. E., de la Torre, J. R., Walker, C. B., Waterbury, J. B., and Stahl, D. A. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature*, 437(7058):543–546.

Korepanov, A. P., Gongadze, G. M., Garber, M. B., Court, D. L., and Bubunenko, M. G. (2007). Importance of the 5 S rRNA-binding Ribosomal Proteins for Cell Viability and Translation in Escherichia coli. *Journal of molecular biology*, 366(4):1199–1208.

Koster, A. J., Grimm, R., Typke, D., Hegerl, R., Stoschek, A., Walz, J., and Baumeister, W. (1997). Perspectives of Molecular and Cellular Electron Tomography. *Journal of Structural Biology*, 120(3):276–308.

Koukos, P. I. and Bonvin, A. M. J. J. (2020). Integrative Modelling of Biomolecular Complexes. *Journal of Molecular Biology*, 432(9):2861–2881.

Kovacs, N. A., Penev, P. I., Venapally, A., Petrov, A. S., and Williams, L. D. (2018). Circular Permutation Obscures Universality of a Ribosomal Protein. *Journal of Molecular Evolution*, 86(8):581–592.

Kraft, B., Jehmlich, N., Larsen, M., Bristow, L. A., Könneke, M., Thamdrup, B., and Canfield, D. E. (2022). Oxygen and nitrogen production by an ammonia-oxidizing archaeon. *Science*, 375(6576):97–100.

Krause, F. (1937). Das magnetische Elektronenmikroskop und seine Anwendung in der Biologie. *Naturwissenschaften*, 25(51):817–825.

Kremer, J. R., Mastronarde, D. N., and McIntosh, J. R. (1996 Jan-Feb). Computer visualization of three-dimensional image data using IMOD. *Journal of Structural Biology*, 116(1):71–76.

Kucukelbir, A., Sigworth, F. J., and Tagare, H. D. (2014). Quantifying the local resolution of cryo-EM density maps. *Nature Methods*, 11(1):63–65.

Kühlbrandt, W. (2014). The Resolution Revolution. *Science*, 343(6178):1443–1444.

Kührová, P., Best, R. B., Bottaro, S., Bussi, G., Šponer, J., Otyepka, M., and Banáš, P. (2016). Computer Folding of RNA Tetraloops: Identification of Key Force Field Deficiencies. *Journal of chemical theory and computation*, 12(9):4534–4548.

Kührová, P., Mlýnský, V., Zgarbová, M., Krepl, M., Bussi, G., Best, R. B., Otyepka, M., Šponer, J., and Banáš, P. (2019). IMPROVING THE PERFORMANCE OF THE AMBER RNA FORCE FIELD BY TUNING THE HYDROGEN-BONDING INTERACTIONS. *Journal of chemical theory and computation*, 15(5):3288–3305.

Kwon, D. (2025). RNA function follows form – why is it so hard to predict? *Nature*, 639(8056):1106–1108.

Lander, G. C. (2024). Single particle cryo-EM map and model validation: It's not crystal clear. *Current Opinion in Structural Biology*, 89:102918.

Landsberg, M. J. and Hankamer, B. (2007). Symmetry: A guide to its application in 2D electron crystallography. *Journal of Structural Biology*, 160(3):332–343.

Languin-Cattoën, O. and Bussi, G. (2026). RNA Dynamics and Interactions Revealed Through Atomistic Simulations. *Annual Review of Physical Chemistry*.

Laszlo, A. H., Derrington, I. M., and Gundlach, J. H. (2016). MspA nanopore as a single-molecule tool: From sequencing to SPRNT. *Methods (San Diego, Calif.)*, 105:75–89.

Lau, A. M., Bordin, N., Kandathil, S. M., Sillitoe, I., Waman, V. P., Wells, J., Orengo, C. A., and Jones, D. T. (2024). Exploring structural diversity across the protein universe with The Encyclopedia of Domains. *Science*.

Laurent, T. and Carlson, L.-A. (2023). The organization of double-stranded RNA in the chikungunya virus replication organelle. *PLOS Neglected Tropical Diseases*, 17(7):e0011404.

Leach, A. R. (2001). *Molecular Modelling: Principles and Applications*. Pearson Education.

Leavitt, W. D., Kopf, S. H., Weber, Y., Chiu, B., McFarlin, J. M., Elling, F. J., Hoeft-McCann, S., and Pearson, A. (2023). Controls on the hydrogen isotope composition of tetraether lipids in an autotrophic ammonia-oxidizing marine archaeon. *Geochimica et Cosmochimica Acta*, 352:194–210.

Lecompte, O., Ripp, R., Thierry, J.-C., Moras, D., and Poch, O. (2002). Comparative analysis of ribosomal proteins in complete genomes: An example of reductive evolution at the domain scale. *Nucleic Acids Research*, 30(24):5382–5390.

Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., Wei, S., Buckner, J., Jeong, J. C., Qi, Y., Jo, S., Pande, V. S., Case, D. A., Brooks, C. L. I., MacKerell, A. D. J., Klauda, J. B., and Im, W. (2016). CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *Journal of Chemical Theory and Computation*, 12(1):405–413.

Leidenfrost, J. G. (1756). On the fixation of water in diverse fire. In *A Tract about Some Qualities of Common Water*, page 240. Herman Ovenius, Universal Bookseller.

Leigh, J. A., Albers, S.-V., Atomi, H., and Allers, T. (2011). Model organisms for genetics in the domain Archaea: Methanogens, halophiles, Thermococcales and Sulfolobales. *FEMS microbiology reviews*, 35(4):577–608.

Lemieux, S. and Major, F. (2002). RNA canonical and non-canonical base pairing types: A recognition method and complete repertoire. *Nucleic Acids Research*, 30(19):4250–4263.

Leontis, N. B., Stombaugh, J., and Westhof, E. (2002a). The non-Watson–Crick base pairs and their associated isostericity matrices. *Nucleic Acids Research*, 30(16):3497–3531.

Leontis, N. B., Stombaugh, J., and Westhof, E. (2002b). SURVEY AND SUMMARY: The non-Watson–Crick base pairs and their associated isostericity matrices. *Nucleic Acids Research*, 30(16):3497–3531.

Leontis, N. B. and Westhof, E. (2001). Geometric nomenclature and classification of RNA base pairs. *RNA*, 7(4):499–512.

Letunic, I. (2015). phyloT : A phylogenetic tree generator. https://phylot.biobyte.de/.

Letunic, I. and Bork, P. (2007). Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics*, 23(1):127–128.

Letunic, I. and Bork, P. (2011). Interactive Tree Of Life v2: Online annotation and display of phylogenetic trees made easy. *Nucleic Acids Research*, 39(suppl_2):W475–W478.

Leung, E. K. Y., Suslov, N., Tuttle, N., Sengupta, R., and Piccirilli, J. A. (2011). The Mechanism of Peptidyl Transfer Catalysis by the Ribosome. *Annual Review of Biochemistry*, 80(Volume 80, 2011):527–555.

Levi, M., Noel, J. K., and Whitford, P. C. (2019). Studying ribosome dynamics with simplified models. *Methods*, 162–163:128–140.

Levitt, M. (1984). Molecular Dynamics of the DNA Double Helix. In Chagas, C. and Pullman, B., editors, *Specificity in Biological Interactions*, pages 43–57, Dordrecht. Springer Netherlands.

Levitt, M. (2013). Birth and Future of Multiscale Modeling for Macromolecular Systems. *Nobel Lecture*.

Lewis, J. M., Janda, K. E., Kotter, D. B., Grose, J. H., and McCleary, W. R. (2023). Characterization of the Attachment of Three New Coliphages onto the Ferrichrome Transporter FhuA. *Journal of Virology*, 97(7):e00667–23.

Li, B., Cao, Y., Westhof, E., and Miao, Z. (2020). Advances in RNA 3D Structure Modeling Using Experimental Data. *Frontiers in Genetics*, 11.

Li, D. and Ji, B. (2019). Protein conformational transitions coupling with ligand interactions: Simulations from molecules to medicine. *Medicine in Novel Technology and Devices*, 3:100026.

Li, P.-N., Herrmann, J., Tolar, B. B., Poitevin, F., Ramdasi, R., Bargar, J. R., Stahl, D. A., Jensen, G. J., Francis, C. A., Wakatsuki, S., and van den Bedem, H. (2018). Nutrient transport suggests an evolutionary basis for charged archaeal surface layer proteins. *The ISME Journal*, 12(10):2389–2402.

Li, T., He, J., Cao, H., Zhang, Y., Chen, J., Xiao, Y., and Huang, S.-Y. (2024). All-atom RNA structure determination from cryo-EM maps. *Nature Biotechnology*, pages 1–9.

Li, X., Grigorieff, N., and Cheng, Y. (2010). GPU-enabled FREALIGN: Accelerating single particle 3D reconstruction and refinement in Fourier space on graphics processors. *Journal of Structural Biology*, 172(3):407–412.

Li, X., Mooney, P., Zheng, S., Booth, C. R., Braunfeld, M. B., Gubbens, S., Agard, D. A., and Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods*, 10(6):584–590.

Li, Z., Chinnasamy, S., Zhang, Y., and Wei, D.-Q. (2021). Molecular dynamics simulation and binding free energy calculations of microcin J25 binding to the FhuA receptor. *Journal of Biomolecular Structure & Dynamics*, 39(7):2585–2594.

Liebl, K. and Zacharias, M. (2023). The development of nucleic acids force fields: From an unchallenged past to a competitive future. *Biophysical Journal*, 122(14):2841–2851.

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J., and Adams, P. D. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallographica Section D: Structural Biology*, 75(10):861–877.

Lightstone, F. C., Schwegler, E., Allesch, M., Gygi, F., and Galli, G. (2005). A first-principles molecular dynamics study of calcium in water. *Chemphyschem: A European Journal of Chemical Physics and Physical Chemistry*, 6(9):1745–1749.

Lima, T., Auchincloss, A. H., Coudert, E., Keller, G., Michoud, K., Rivoire, C., Bulliard, V., de Castro, E., Lachaize, C., Baratin, D., Phan, I., Bougueleret, L., and Bairoch, A. (2009). HAMAP: A database of completely sequenced microbial proteome sets and manually curated microbial protein families in UniProtKB/Swiss-Prot. *Nucleic Acids Research*, 37(Database issue):D471–D478.

Lindahl, E., Hess, B., and van der Spoel, D. (2001). GROMACS 3.0: A package for molecular simulation and trajectory analysis. *Molecular modeling annual*, 7(8):306–317.

Liu, J., Cvirkaite-Krupovic, V., Commere, P.-H., Yang, Y., Zhou, F., Forterre, P., Shen, Y., and Krupovic, M. (2021). Archaeal extracellular vesicles are produced in an ESCRT-dependent manner and promote gene transfer and nutrient cycling in extreme environments. *The ISME Journal*, 15(10):2892–2905.

Liu, P., Agrafiotis, D. K., and Theobald, D. L. (2010). Fast determination of the optimal rotational matrix for macromolecular superpositions. *Journal of Computational Chemistry*, 31(7):1561–1563.

Locher, K. P., Rees, B., Koebnik, R., Mitschler, A., Moulinier, L., Rosenbusch, J. P., and Moras, D. (1998). Transmembrane Signaling across the Ligand-Gated FhuA Receptor: Crystal Structures of Free and Ferrichrome-Bound States Reveal Allosteric Changes. *Cell*, 95(6):771–778.

London, F. (1930). Zur Theorie und Systematik der Molekularkräfte. *Zeitschrift für Physik*, 63(3):245–279.

Lowe, T. M. and Richardson, P. (2007). Complete sequence of Pyrobaculum arsenaticum DSM 13514. https://www.ebi.ac.uk/ena/data/view/GCA_000016385.1.

Luan, Y., Tang, N., Yang, J., Liu, S., Cheng, C., Wang, Y., Chen, C., Guo, Y.-n., Wang, H., Zhao, W., Zhao, Q., Li, W., Xiang, M., Ju, R., and Xie, Z. (2022). Deficiency of ribosomal proteins reshapes the transcriptional and translational landscape in human cells. *Nucleic Acids Research*, 50(12):6601–6617.

Lucić, V., Förster, F., and Baumeister, W. (2005). Structural studies by electron tomography: From cells to molecules. *Annual Review of Biochemistry*, 74:833–865.

Lučić, V., Rigort, A., and Baumeister, W. (2013). Cryo-electron tomography: The challenge of doing structural biology in situ. *The Journal of Cell Biology*, 202(3):407–419.

Luo, J., Liu, Z., Guo, Y., and Li, M. (2015). A structural dissection of large protein-protein crystal packing contacts. *Scientific Reports*, 5(1):14214.

Macelroy, R. D. (1974). Some comments on the evolution of extremophiles. *BioSystems*, 6:74–75.

MacKenzie, M. and Argyropoulos, C. (2023). An Introduction to Nanopore Sequencing: Past, Present, and Future Considerations. *Micromachines*, 14(2):459.

MacKerell Jr, A. D. (2006). CHARMM force fields, parameterization strategies and future/ongoing force field developments.

Madru, C., Lebaron, S., Blaud, M., Delbos, L., Pipoli, J., Pasmant, E., Réty, S., and Leulliot, N. (2015). Chaperoning 5S RNA assembly. *Genes & Development*, 29(13):1432–1446.

Maglia, G., Heron, A. J., Stoddart, D., Japrung, D., and Bayley, H. (2010). Analysis of single nucleic acid molecules with protein nanopores. *Methods in enzymology*, 475:591–623.

Marchler-Bauer, A., Anderson, J. B., Cherukuri, P. F., DeWeese-Scott, C., Geer, L. Y., Gwadz, M., He, S., Hurwitz, D. I., Jackson, J. D., Ke, Z., Lanczycki, C. J., Liebert, C. A., Liu, C., Lu, F., Marchler, G. H., Mullokandov, M., Shoemaker, B. A., Simonyan, V., Song, J. S., Thiessen, P. A., Yamashita, R. A., Yin, J. J., Zhang, D., and Bryant, S. H. (2005). CDD: A Conserved Domain Database for protein classification. *Nucleic Acids Research*, 33(Database issue):D192–196.

Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P., and de Vries, A. H. (2007). The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *The Journal of Physical Chemistry B*, 111(27):7812–7824.

Martens-Habbena, W., Berube, P. M., Urakawa, H., de la Torre, J. R., and Stahl, D. A. (2009). Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria. *Nature*, 461(7266):976–979.

Martinez-Gutierrez, C. A. and Aylward, F. O. (2022). Genome size distributions in bacteria and archaea are strongly linked to evolutionary history at broad phylogenetic scales. *PLOS Genetics*, 18(5):e1010220.

Marton, L. (1934). Electron Microscopy of Biological Objects. *Nature*, 133(3372):911–911.

Martorana, V., Corongiu, G., and Palma, M. U. (1996). Correlated solvent-induced forces on a protein at single residue resolution: Relation to conformation, stability, dynamics and function. *Chemical Physics Letters*, 254(5):292–301.

Martorana, V., Corongiu, G., and Palma, M. U. (1998). Interaction of explicit solvent with hydrophobic/philic/charged residues of a protein: Residue character vs. conformational context. *Proteins*, 32(2):129–135.

Mastronarde, D. N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. *Journal of Structural Biology*, 152(1):36–51.

Mastronarde, D. N. and Held, S. R. (2017). Automated tilt series alignment and tomographic reconstruction in IMOD. *Journal of Structural Biology*, 197(2):102–113.

Mathé, J., Aksimentiev, A., Nelson, D. R., Schulten, K., and Meller, A. (2005). Orientation discrimination of single-stranded DNA inside the $\alpha$-hemolysin membrane channel. *Proceedings of the National Academy of Sciences of the United States of America*, 102(35):12377–12382.

Maxwell, J. C. (1860). Illustrations of the dynamical theory of gases: On the Motions and Collisions of Perfectly Elastic Spheres. *Philosophical Magazine*, XIX:19–32.

May, M. B., Lopez-Perez, G. S., and Davis, J. H. (2026). Capturing ribosomal structures in cellular extracts with cryoPRISM: A purification-free cryoEM approach reveals novel structural states. *Proceedings of the National Academy of Sciences*, 123(9):e2521210123.

Mayse, L. A. and Movileanu, L. (2023). Gating of $\beta$-Barrel Protein Pores, Porins, and Channels: An Old Problem with New Facets. *International Journal of Molecular Sciences*, 24(15):12095.

McCammon, J. A., Gelin, B. R., and Karplus, M. (1977). Dynamics of folded proteins. *Nature*, 267(5612):585–590.

McGreevy, R., Teo, I., Singharoy, A., and Schulten, K. (2016). Advances in the molecular dynamics flexible fitting method for cryo-EM modeling. *Methods (San Diego, Calif.)*, 100:50–60.

McKinney, W. (2011). Pandas: A Foundational Python Library for Data Analysis and Statistics.

McLaren, M., Conners, R., Isupov, M. N., Gil-Díez, P., Gambelli, L., Gold, V. A. M., Walter, A., Connell, S. R., Williams, B., and Daum, B. (2023). CryoEM reveals that ribosomes in microsporidian spores are locked in a dimeric hibernating state. *Nature Microbiology*, 8(10):1834–1845.

McMullan, G., Clark, A. T., Turchetta, R., and Faruqi, A. R. (2009). Enhanced imaging in low dose electron microscopy using electron counting. *Ultramicroscopy*, 109(12):1411–1416.

Mears, J. A., Cannone, J. J., Stagg, S. M., Gutell, R. R., Agrawal, R. K., and Harvey, S. C. (2002). Modeling a Minimal Ribosome Based on Comparative Sequence Analysis. *Journal of Molecular Biology*, 321(2):215–234.

Melnikov, S., Ben-Shem, A., Garreau de Loubresse, N., Jenner, L., Yusupova, G., and Yusupov, M. (2012). One core, two shells: Bacterial and eukaryotic ribosomes. *Nature Structural & Molecular Biology*, 19(6):560–567.

Michaud-Agrawal, N., Denning, E. J., Woolf, T. B., and Beckstein, O. (2011). MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry*, 32(10):2319–2327.

Micic, J., Rodríguez-Galán, O., Babiano, R., Fitzgerald, F., Fernández-Fernández, J., Zhang, Y., Gao, N., Woolford, Jr, J. L., and de la Cruz, J. (2022). Ribosomal protein eL39 is important for maturation of the nascent polypeptide exit tunnel and proper protein folding during translation. *Nucleic Acids Research*, 50(11):6453–6473.

Minić, S., Annighöfer, B., Hélary, A., Hamdane, D., Hui Bon Hoa, G., Loupiac, C., Brûlet, A., and Combet, S. (2020). Effect of Ligands on HP-Induced Unfolding and Oligomerization of $\beta$-Lactoglobulin. *Biophysical Journal*, 119(11):2262–2274.

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: Making protein folding accessible to all. *Nature Methods*, 19(6):679–682.

Mlýnský, V. and Bussi, G. (2018). Exploring RNA structure and dynamics through enhanced sampling simulations. *Current Opinion in Structural Biology*, 49:63–71.

Modi, N., Winterhalter, M., and Kleinekathöfer, U. (2012). Computational modeling of ion transport through nanopores. *Nanoscale*, 4(20):6166–6180.

Mohammad, M. M., Howard, K. R., and Movileanu, L. (2011). Redesign of a Plugged $\beta$-Barrel Membrane Protein. *Journal of Biological Chemistry*, 286(10):8000–8013.

Mokdad, A., Krasovska, M. V., Sponer, J., and Leontis, N. B. (2006). Structural and evolutionary classification of G/U wobble basepairs in the ribosome. *Nucleic Acids Research*, 34(5):1326–1341.

Möller, J., Schroer, M. A., Erlkamp, M., Grobelny, S., Paulus, M., Tiemeyer, S., Wirkert, F. J., Tolan, M., and Winter, R. (2012). The Effect of Ionic Strength, Temperature, and Pressure on the Interaction Potential of Dense Protein Solutions: From Nonlinear Pressure Response to Protein Crystallization. *Biophysical Journal*, 102(11):2641–2648.

Mollica, L., Bessa, L. M., Hanoulle, X., Jensen, M. R., Blackledge, M., and Schneider, R. (2016). Binding Mechanisms of Intrinsically Disordered Proteins: Theory, Simulation, and Experiment. *Frontiers in Molecular Biosciences*, 3.

Moore, P. B. (1997). Ribosomes: Protein synthesis in slow motion. *Current Biology*, 7(3):R179–R181.

Moore, P. B. (2009). The ribosome returned. *Journal of Biology*, 8(1):8.

Mukhopadhyay, R. and Wingreen, N. S. (2009). Curvature and shape determination of growing bacteria. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 80(6 Pt 1):062901.

Mullakhanbhai, M. F. and Larsen, H. (1975). Halobacterium volcanii spec. nov., a Dead Sea halobacterium with a moderate salt requirement. *Archives of Microbiology*, 104(3):207–214.

Mushegian, A. (2005). Protein content of minimal and ancestral ribosome. *RNA*, 11(9):1400–1406.

Muzard, J., Martinho, M., Mathé, J., Bockelmann, U., and Viasnoff, V. (2010). DNA Translocation and Unzipping through a Nanopore: Some Geometrical Effects. *Biophysical Journal*, 98(10):2170–2178.

Naganuma, T., Nomura, N., Yao, M., Mochizuki, M., Uchiumi, T., and Tanaka, I. (2010). Structural Basis for Translation Factor Recruitment to the Eukaryotic/Archaeal Ribosomes. *The Journal of Biological Chemistry*, 285(7):4747–4756.

Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P. M. G. E., Grigoras, I. T., Malinauskaite, L., Malinauskas, T., Miehling, J., Uchański, T., Yu, L., Karia, D., Pechnikova, E. V., de Jong, E., Keizer, J., Bischoff, M., McCormack, J., Tiemeijer, P., Hardwick, S. W., Chirgadze, D. Y., Murshudov, G., Aricescu, A. R., and Scheres, S. H. W. (2020). Single-particle cryo-EM at atomic resolution. *Nature*, 587(7832):152–156.

Nakane, T. and Scheres, S. H. W. (2021). Multi-body Refinement of Cryo-EM Images in RELION. *Methods in Molecular Biology (Clifton, N.J.)*, 2215:145–160.

Nanda, V. (2019). Building bigger beta-barrels. *eLife*, 8:e44076.

Nature Structural Biology (1998). The 'new' PDB [Editorial]. *Nature Structural Biology*, 5(11):925–926.

Naydenova, K., Jia, P., and Russo, C. J. (2020). Cryo-EM with sub-1 Å specimen movement. *Science (New York, N.Y.)*, 370(6513):223–226.

Naydenova, K. and Russo, C. J. (2017). Measuring the effects of particle orientation to improve the efficiency of electron cryomicroscopy. *Nature Communications*, 8(1):629.

Neidle, S. (2010). *Principles of Nucleic Acid Structure*. Elsevier.

Neselu, K., Wang, B., Rice, W. J., Potter, C. S., Carragher, B., and Chua, E. Y. (2023). Measuring the effects of ice thickness on resolution in single particle cryo-EM. *Journal of Structural Biology: X*, 7:100085.

Newman, M. S. (1955). A notation for the study of certain stereochemical problems. *Journal of Chemical Education*, 32(7):344.

Newton, I. (1687). *Philosophiae Naturalis Principia Mathematica.*

Nguyen, T. T. T., Choi, Y. H., Lee, W.-K., Ji, Y., Chun, E., Kim, Y. H., Lee, J.-E., Jung, H. S., Suh, J. H., Kim, S., and Jin, M. (2023). Tryptophan-dependent and -independent secretions of tryptophanyl- tRNA synthetase mediate innate inflammatory responses. *Cell Reports*, 42(1):111905.

Nickell, S., Förster, F., Linaroudis, A., Net, W. D., Beck, F., Hegerl, R., Baumeister, W., and Plitzko, J. M. (2005). TOM software toolbox: Acquisition and analysis for electron tomography. *Journal of Structural Biology*, 149(3):227–234.

Niedzwiecki, D. J., Mohammad, M. M., and Movileanu, L. (2012). Inspection of the Engineered FhuA $\Delta$C/$\Delta$4L Protein Nanopore by Polymer Exclusion. *Biophysical Journal*, 103(10):2115–2124.

Nikolaeva, D. D., Gelfand, M. S., and Garushyants, S. K. (2021). Simplification of Ribosomes in Bacteria with Tiny Genomes. *Molecular Biology and Evolution*, 38(1):58–66.

Nikolay, R., van den Bruck, D., Achenbach, J., and Nierhaus, K. H. (2015). Ribosomal Proteins: Role in Ribosomal Functions. In *Encyclopedia of Life Sciences*, pages 1–12. John Wiley & Sons, Ltd.

Nikulin, A., Eliseikina, I., Tishchenko, S., Nevskaya, N., Davydova, N., Platonova, O., Piendl, W., Selmer, M., Liljas, A., Drygin, D., Zimmermann, R., Garber, M., and Nikonov, S. (2003). Structure of the L1 protuberance in the ribosome. *Nature Structural Biology*, 10(2):104–108.

Nissen, P., Hansen, J., Ban, N., Moore, P. B., and Steitz, T. A. (2000). The Structural Basis of Ribosome Activity in Peptide Bond Synthesis. *Science*, 289(5481):920–930.

Nissley, A. J., Shulgina, Y., Kivimae, R. W., Downing, B. E., Penev, P. I., Banfield, J. F., Nayak, D. D., and Cate, J. H. D. (2025). Structure of an archaeal ribosome reveals a divergent active site and hibernation factor. *Nature Microbiology*, 10(8):1940–1953.

Nithin, C., Kmiecik, S., Błaszczyk, R., Nowicka, J., and Tuszyńska, I. (2024). Comparative analysis of RNA 3D structure prediction methods: Towards enhanced modeling of RNA–ligand interactions. *Nucleic Acids Research*, 52(13):7465–7486.

Nithin, C., Pilla, S. P., and Kmiecik, S. (2025). When does molecular dynamics improve RNA models? Insights from CASP15 and practical guidelines. *Computational and Structural Biotechnology Journal*, 27:4201–4211.

Noakes, M. T., Brinkerhoff, H., Laszlo, A. H., Derrington, I. M., Langford, K. W., Mount, J. W., Bowman, J. L., Baker, K. S., Doering, K. M., Tickman, B. I., and Gundlach, J. H. (2019). Increasing the accuracy of nanopore DNA sequencing using a time-varying cross membrane voltage. *Nature biotechnology*, 37(6):651–656.

Nogales, E. and Mahamid, J. (2024). Bridging structural and cell biology with cryo-electron microscopy. *Nature*, 628(8006):47–56.

Noller, H. F. (2005). RNA Structure: Reading the Ribosome. *Science*, 309(5740):1508–1514.

Noller, H. F. (2024). The ribosome comes to life. *Cell*, 187(23):6486–6500.

Noller, H. F., Hoffarth, V., and Zimniak, L. (1992). Unusual resistance of peptidyl transferase to protein extraction procedures. *Science (New York, N.Y.)*, 256(5062):1416–1419.

Nürenberg-Goloub, E., Kratzat, H., Heinemann, H., Heuer, A., Kötter, P., Berninghausen, O., Becker, T., Tampé, R., and Beckmann, R. (2020). Molecular analysis of the ribosome recycling factor ABCE1 bound to the 30S post-splitting complex. *The EMBO Journal*, 39(9):e103788.

Nyquist, H. (1928). Certain Topics in Telegraph Transmission Theory. *Transactions of the American Institute of Electrical Engineers*, 47(2):617–644.

O'Brien, E. P., Brooks, B. R., and Thirumalai, D. (2012). Effects of pH on proteins: Predictions for ensemble and single molecule pulling experiments. *Journal of the American Chemical Society*, 134(2):979–987.

Ogle, J. M., Brodersen, D. E., Clemons, W. M., Tarry, M. J., Carter, A. P., and Ramakrishnan, V. (2001). Recognition of Cognate Transfer RNA by the 30S Ribosomal Subunit. *Science*, 292(5518):897–902.

Ohm, G. S. (1827). *Die galvanische Kette, mathematisch bearbeitet*. Riemann.

Olson, W. K., Esguerra, M., Xin, Y., and Lu, X.-J. (2009). New information content in RNA base pairing deduced from quantitative analysis of high-resolution structures. *Methods (San Diego, Calif.)*, 47(3):177–186.

Omori, R., Ichino, S., Yoshikawa, T., Moriyama, Y., and Mitsui, T. (2021). The effects of pore geometry for DNA entering and clogging nanopore. 2021:M71.270.

Oren, A. and Ventosa, A. (1999). Benjamin Elazari Volcani (1915-1999): Sixty-three years of studies of the microbiology of the Dead Sea. *International Microbiology: The Official Journal of the Spanish Society for Microbiology*, 2(3):195–198.

Orengo, C. A., Pearl, F. M. G., Bray, J. E., Todd, A. E., Martin, A. C., Lo Conte, L., and Thornton, J. M. (1999). The CATH Database provides insights into protein structure/function relationships. *Nucleic Acids Research*, 27(1):275–279.

Paiva, V. d. A., Gomes, I. d. S., Monteiro, C. R., Mendonça, M. V., Martins, P. M., Santana, C. A., Gonçalves-Almeida, V., Izidoro, S. C., de Melo-Minardi, R. C., and Silveira, S. d. A. (2022). Protein structural bioinformatics: An overview. *Computers in Biology and Medicine*, 147:105695.

Palade, G. E. (1955). A small particulate component of the cytoplasm. *The Journal of Biophysical and Biochemical Cytology*, 1(1):59–68.

Pande, V. S., Baker, I., Chapman, J., Elmer, S. P., Khaliq, S., Larson, S. M., Rhee, Y. M., Shirts, M. R., Snow, C. D., Sorin, E. J., and Zagrovic, B. (2003). Atomistic protein folding simulations on the submillisecond time scale using worldwide distributed computing. *Biopolymers*, 68(1):91–109.

Papadopoulos, J. S. and Agarwala, R. (2007). COBALT: Constraint-based alignment tool for multiple protein sequences. *Bioinformatics*, 23(9):1073–1079.

Pape, T., Stark, H., Matadeen, R., Orlova, E. V., and Van Heel, M. (2000). Visualization of the Translational Elongation Cycle by Cryo-Electron Microscopy. In *The Ribosome*, chapter 4, pages 35–44. John Wiley & Sons, Ltd.

Parrinello, M. and Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12):7182–7190.

Pauli, W. (1925). Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren. *Zeitschrift für Physik*, 31(1):765–783.

Pearce, R., Omenn, G. S., and Zhang, Y. (2022). De Novo RNA Tertiary Structure Prediction at Atomic Resolution Using Geometric Potentials from Deep Learning.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572.

Pedruzzi, I., Rivoire, C., Auchincloss, A. H., Coudert, E., Keller, G., de Castro, E., Baratin, D., Cuche, B. A., Bougueleret, L., Poux, S., Redaschi, N., Xenarios, I., and Bridge, A. (2015). HAMAP in 2015: Updates to the protein family classification and annotation system. *Nucleic Acids Research*, 43(D1):D1064–D1070.

Pelt, D. M., Batenburg, K. J., King, M., Glick, S., Mueller, K., and Scientific Computing (2015). Accurately approximating algebraic tomographic reconstruction by filtered backprojection. None.

Penev, P. I., Fakhretaha-Aval, S., Patel, V. J., Cannone, J. J., Gutell, R. R., Petrov, A. S., Williams, L. D., and Glass, J. B. (2020). Supersized Ribosomal RNA Expansion Segments in Asgard Archaea. *Genome Biology and Evolution*, 12(10):1694–1710.

Pérez, A., Sponer, J., Jurecka, P., Hobza, P., Luque, F. J., and Orozco, M. (2005). Are the Hydrogen Bonds of RNA (A·U) Stronger Than those of DNA (A·T)? A Quantum Mechanics Study. *Chemistry – A European Journal*, 11(17):5062–5066.

Perilla, J. R., Goh, B. C., Cassidy, C. K., Liu, B., Bernardi, R. C., Rudack, T., Yu, H., Wu, Z., and Schulten, K. (2015). Molecular dynamics simulations of large macromolecular complexes. *Current opinion in structural biology*, 31:64–74.

Perrakis, A. and Sixma, T. K. (2021). AI revolutions in biology. *EMBO reports*, 22(11):e54046.

Perutz, M. F. (1976). STRUCTURE AND MECHANISM OF HAEMOGLOBIN. *British Medical Bulletin*, 32(3):195–208.

Petrov, A. S., Bernier, C. R., Hershkovits, E., Xue, Y., Waterbury, C. C., Hsiao, C., Stepanov, V. G., Gaucher, E. A., Grover, M. A., Harvey, S. C., Hud, N. V., Wartell, R. M., Fox, G. E., and Williams, L. D. (2013). Secondary structure and domain architecture of the 23S and 5S rRNAs. *Nucleic Acids Research*, 41(15):7522–7535.

Petrov, A. S., Bernier, C. R., Hsiao, C., Norris, A. M., Kovacs, N. A., Waterbury, C. C., Stepanov, V. G., Harvey, S. C., Fox, G. E., Wartell, R. M., Hud, N. V., and Williams, L. D. (2014). Evolution of the ribosome at atomic resolution. *Proceedings of the National Academy of Sciences*, 111(28):10251–10256.

Petrov, A. S., Gulen, B., Norris, A. M., Kovacs, N. A., Bernier, C. R., Lanier, K. A., Fox, G. E., Harvey, S. C., Wartell, R. M., Hud, N. V., and Williams, L. D. (2015). History of the ribosome and the origin of translation. *Proceedings of the National Academy of Sciences*, 112(50):15396–15401.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., and Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science: A Publication of the Protein Society*, 30(1):70–82.

Pezeshki, S., Chimerel, C., Bessonov, A. N., Winterhalter, M., and Kleinekathöfer, U. (2009). Understanding Ion Conductance on a Molecular Level: An All-Atom Modeling of the Bacterial Porin OmpF. *Biophysical Journal*, 97(7):1898–1906.

Pfeil-Gardiner, O., Rosa, H. V. D., Riedel, D., Chen, Y. S., Lörks, D., Kükelhan, P., Linck, M., Müller, H., Van Petegem, F., and Murphy, B. J. (2024). Elemental mapping in single-particle reconstructions by reconstructed electron energy-loss analysis. *Nature Methods*, 21(12):2299–2306.

Phipps, B. M., Engelhardt, H., Huber, R., and Baumeister, W. (1990). Three-dimensional structure of the crystalline protein envelope layer of the hyperthermophilic archaebacterium Pyrobaculum islandicum. *Journal of Structural Biology*, 103(2):152–163.

Phipps, B. M., Huber, R., and Baumeister, W. (1991). The cell envelope of the hyperthermophilic archaebacterium Pyrobaculum organotrophum consists of two regularly arrayed protein layers: Three-dimensional structure of the outer layer. *Molecular Microbiology*, 5(2):253–265.

Pidcock, E. and Moore, G. R. (2001). Structural characteristics of protein binding sites for calcium and lanthanide ions. *JBIC Journal of Biological Inorganic Chemistry*, 6(5):479–489.

Pinet, L., Assrir, N., and van Heijenoort, C. (2021). Expanding the Disorder-Function Paradigm in the C-Terminal Tails of Erbbs. *Biomolecules*, 11(11):1690.

Piovesan, D., Profiti, G., Martelli, P. L., and Casadio, R. (2012). The human "magnesome": Detecting magnesium binding sites on human proteins. *BMC Bioinformatics*, 13(Suppl 14):S10.

Porter, K. R. and Novikoff, A. B. (1974). The 1974 Nobel Prize for Physiology or Medicine. *Science*, 186(4163):516–520.

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

Prisant, M. G., Williams, C. J., Chen, V. B., Richardson, J. S., and Richardson, D. C. (2020). New tools in MolProbity validation: CaBLAM for CryoEM backbone, UnDowser to rethink "waters," and NGL Viewer to recapture online 3D graphics. *Protein Science*, 29(1):315–329.

Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., Hess, B., and Lindahl, E. (2013). GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7):845–854.

Ptitsyn, O. (1995). Molten Globule and Protein Folding. In *Advances in Protein Chemistry*, volume 47, pages 83–229. Academic Press.

Pum, D., Messner, P., and Sleytr, U. B. (1991). Role of the S layer in morphogenesis and cell division of the archaebacterium Methanocorpusculum sinense. *Journal of Bacteriology*, 173(21):6865–6873.

PyMOL (2015). The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. Schrödinger, LLC.

Qin, W., Heal, K. R., Ramdasi, R., Kobelt, J. N., Martens-Habbena, W., Bertagnolli, A. D., Amin, S. A., Walker, C. B., Urakawa, H., Könneke, M., Devol, A. H., Moffett, J. W., Armbrust, E. V., Jensen, G. J., Ingalls, A. E., and Stahl, D. A. (2017). Nitrosopumilus maritimus gen. nov., sp. nov., Nitrosopumilus cobalaminigenes sp. nov., Nitrosopumilus oxyclinae sp. nov., and Nitrosopumilus ureiphilus sp. nov., four marine ammonia-oxidizing archaea of the phylum Thaumarchaeota. *International Journal of Systematic and Evolutionary Microbiology*, 67(12):5067–5079.

Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, 7:95–99.

Ramachandran, G. N. and Sasisekharan, V. (1968). Conformation of polypeptides and proteins. *Advances in Protein Chemistry*, 23:283–438.

Ramakrishnan, V. (2002). Ribosome Structure and the Mechanism of Translation. *Cell*, 108(4):557–572.

Ramakrishnan, V., Capel, M. S., Clemons Jr., W. M., May, J. L. C., and Wimberly, B. T. (2000). Progress toward the Crystal Structure of a Bacterial 30S Ribosomal Subunit. In *The Ribosome*, chapter 1, pages 1–9. John Wiley & Sons, Ltd.

Ramalli, S. G., Miles, A. J., Janes, R. W., and Wallace, B. A. (2022). The PCDDB (Protein Circular Dichroism Data Bank): A Bioinformatics Resource for Protein Characterisations and Methods Development. *Journal of Molecular Biology*, 434(11):167441.

Ramlan, E. I. and Zauner, K.-P. (2008). An Extended Dot-Bracket-Notation for Functional Nucleic Acids. pages 75–86. Oesterreichische Computer Gesellschaft.

Rampelotto, P. H. (2013). Extremophiles and Extreme Environments. *Life : Open Access Journal*, 3(3):482–485.

Rattu, P., Belzunces, B., Haynes, T., Skylaris, C.-K., and Khalid, S. (2021a). Translocation of flexible and tensioned ssDNA through in silico designed hydrophobic nanopores with two constrictions. *Nanoscale*, 13(3):1673–1679.

Rattu, P., Glencross, F., Mader, S. L., Skylaris, C.-K., Matthews, S. J., Rouse, S. L., and Khalid, S. (2021b). Atomistic level characterisation of ssDNA translocation through the E. coli proteins CsgG and CsgF for nanopore sequencing. *Computational and Structural Biotechnology Journal*, 19:6417–6430.

Rauscher, R. and Polacek, N. (2024). Ribosomal RNA expansion segments and their role in ribosome biology. *Biochemical Society Transactions*, 52(3):1317–1325.

Razi, A., Britton, R. A., and Ortega, J. (2017). The impact of recent improvements in cryo-electron microscopy technology on the understanding of bacterial ribosome assembly. *Nucleic Acids Research*, 45(3):1027–1040.

Razi, A. and Ortega, J. (2017). Ribosomal Proteins: Their Role in the Assembly, Structure and Function of the Ribosome. In *Encyclopedia of Life Sciences*, pages 1–12. John Wiley & Sons, Ltd.

Redfern, O., Dessailly, B., and Orengo, C. (2008). Exploring the structure and function paradigm. *Current opinion in structural biology*, 18(3):394–402.

Ren, P., Chun, J., Thomas, D. G., Schnieders, M. J., Marucho, M., Zhang, J., and Baker, N. A. (2012). Biomolecular electrostatics and solvation: A computational perspective. *Quarterly reviews of biophysics*, 45(4):427–491.

Rensen, E., Krupovic, M., and Prangishvili, D. (2015). Mysterious hexagonal pyramids on the surface of Pyrobaculum cells. *Biochimie*, 118:365–367.

Richardson, P. and Stahl, D. A. (2007). Complete sequence of Nitrosopumilus maritimus SCM1. https://www.ebi.ac.uk/ena/data/view/GCA_000018465.1.

Rinnenthal, J., Buck, J., Ferner, J., Wacker, A., FÜrtig, B., and Schwalbe, H. (2011). Mapping the Landscape of RNA Dynamics with NMR Spectroscopy. *Accounts of Chemical Research*, 44(12):1292–1301.

Rivas, E. (2021). Evolutionary conservation of RNA sequence and structure. *WIREs RNA*, 12(5):e1649.

Roberts, R. B. (1958). *Microsomal Particles and Protein Synthesis; Papers Presented at the First Symposium of the Biophysical Society, at the Massachusetts Institute of Technology, Cambridge, February 5, 6, and 8, 1958*. Pergamon Press.

Robinson, N. P. (2018). Archaea, from obscurity to superhero microbes: 40 years of surprises and critical biological insights. *Emerging Topics in Life Sciences*, 2(4):453–458.

Rodnina, M., Beringer, M., and Bieling, P. (2005). Ten remarks on peptide bond formation on the ribosome. *Biochemical Society Transactions*, 33(3):493–498.

Rodnina, M. V., Beringer, M., and Wintermeyer, W. (2006). Mechanism of peptide bond formation on the ribosome. *Quarterly Reviews of Biophysics*, 39(3):203–225.

Rodrigues-Oliveira, T., Belmok, A., Vasconcellos, D., Schuster, B., and Kyaw, C. M. (2017). Archaeal S-Layers: Overview and Current State of the Art. *Frontiers in Microbiology*, 8.

Rodrigues-Oliveira, T., Souza, A. A., Kruger, R., Schuster, B., de Freitas, S. M., and Kyaw, C. M. (2019). Environmental factors influence the Haloferax volcanii S-layer protein structure. *PLOS ONE*, 14(5):e0216863.

Rodríguez-Ropero, F. and Fioroni, M. (2012). Structural and dynamical analysis of an engineered FhuA channel protein embedded into a lipid bilayer or a detergent belt. *Journal of Structural Biology*, 177(2):291–301.

Roger, A. J., Susko, E., and Leger, M. M. (2021). Evolution: Reconstructing the Timeline of Eukaryogenesis. *Current Biology*, 31(4):R193–R196.

Rogers, J. M., Steward, A., and Clarke, J. (2013). Folding and Binding of an Intrinsically Disordered Protein: Fast, but Not 'Diffusion-Limited'. *Journal of the American Chemical Society*, 135(4):1415–1422.

Rohou, A. and Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology*, 192(2):216–221.

Rosenthal, P. B. and Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *Journal of Molecular Biology*, 333(4):721–745.

Rossmann, M. G. (2000). Fitting atomic models into electron-microscopy maps. *Acta Crystallographica Section D: Biological Crystallography*, 56(10):1341–1349.

Rout, M. P. and Sali, A. (2019). Principles for Integrative Structural Biology Studies. *Cell*, 177(6):1384–1403.

Roux, B. (2008). The Membrane Potential and its Representation by a Constant Electric Field in Computer Simulations. *Biophysical Journal*, 95(9):4205–4216.

Roy, A., Ward, E., Choi, I., Cosi, M., Edgin, T., Hughes, T. S., Islam, M. S., Khan, A. M., Kolekar, A., Rayl, M., Robinson, I., Sarando, P., Skidmore, E., Swetnam, T. L., Wall, M., Xu, Z., Yung, M. L., Merchant, N., and Wheeler, T. J. (2025). MDRepo—an open data warehouse for community-contributed molecular dynamics simulations of proteins. *Nucleic Acids Research*, 53(D1):D477–D486.

Rozov, A., Khusainov, I., El Omari, K., Duman, R., Mykhaylyk, V., Yusupov, M., Westhof, E., Wagner, A., and Yusupova, G. (2019). Importance of potassium ions for ribosome structure and function revealed by long-wavelength X-ray diffraction. *Nature Communications*, 10(1):2519.

Ruland, J. A., Krüger, A. M., Dörner, K., Bhatia, R., Wirths, S., Poetes, D., Kutay, U., Siebrasse, J. P., and Kubitscheck, U. (2021). Nuclear export of the pre-60S ribosomal subunit through single nuclear pores observed in real time. *Nature Communications*, 12(1):6211.

Ruska, E. (1934a). Über ein magnetisches Objektiv für das Elektronenmikroskop. *Zeitschrift für Physik*, 89(1):90–128.

Ruska, E. (1934b). über Fortschritte im Bau und in der Leistung des magnetischen Elektronenmikroskops. *Zeitschrift für Physik*, 87(9):580–602.

Ruska, E. (1987). The Development of the Electron Microscope and of Electron Microscopy (Nobel Lecture). *Angewandte Chemie International Edition in English*, 26(7):595–605.

Russo, C. J., Dickerson, J. L., and Naydenova, K. (2022). Cryomicroscopy in situ: What is the smallest molecule that can be directly identified without labels in a cell? *Faraday Discussions*, 240:277–302.

Russo, C. J. and Passmore, L. A. (2014). Electron microscopy: Ultrastable gold substrates for electron cryomicroscopy. *Science (New York, N.Y.)*, 346(6215):1377–1380.

Saibil, H. R. (2017). Blob-ology and biology of cryo-EM: An interview with Helen Saibil. *BMC Biology*, 15(1):77.

Sakai, H. D. and Kurosawa, N. (2018). Saccharolobus caldissimus gen. nov., sp. nov., a facultatively anaerobic iron-reducing hyperthermophilic archaeon isolated from an acidic terrestrial hot spring, and reclassification of Sulfolobus solfataricus as Saccharolobus solfataricus comb. nov. and Sulfolobus shibatae as Saccharolobus shibatae comb. nov. *International Journal of Systematic and Evolutionary Microbiology*, 68(4):1271–1278.

Sako, Y., Nunoura, T., and Uchida, A. (2001). Pyrobaculum oguniense sp. nov., a novel facultatively aerobic and hyperthermophilic archaeon growing at up to 97 degrees C. *International Journal of Systematic and Evolutionary Microbiology*, 51(Pt 2):303–309.

Sas-Chen, A., Thomas, J. M., Matzov, D., Taoka, M., Nance, K. D., Nir, R., Bryson, K. M., Shachar, R., Liman, G. L. S., Burkhart, B. W., Gamage, S. T., Nobe, Y., Briney, C. A., Levy, M. J., Fuchs, R. T., Robb, G. B., Hartmann, J., Sharma, S., Lin, Q., Florens, L., Washburn, M. P., Isobe, T., Santangelo, T. J., Shalev-Benami, M., Meier, J. L., and Schwartz, S. (2020). Dynamic RNA acetylation revealed by quantitative cross-evolutionary mapping. *Nature*, 583(7817):638–643.

Sayers, E. W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K. D., and Karsch-Mizrachi, I. (2019). GenBank. *Nucleic Acids Research*, 47(D1):D94–D99.

Scheres, S. H. (2014). Beam-induced motion correction for sub-megadalton cryo-EM particles. *eLife*, 3:e03665.

Scheres, S. H. and Chen, S. (2012). Prevention of overfitting in cryo-EM structure determination. *Nature methods*, 9(9):853–854.

Scheres, S. H. W. (2012). A Bayesian view on cryo-EM structure determination. *Journal of Molecular Biology*, 415(2):406–418.

Scherrer, W. (1946). Die Einlagerung eines regulären Vielecks in ein Gitter. *Elemente der Mathmatik*, 1(6):97–98.

Schoch, C. L., Ciufo, S., Domrachev, M., Hotton, C. L., Kannan, S., Khovanskaya, R., Leipe, D., Mcveigh, R., O'Neill, K., Robbertse, B., Sharma, S., Soussov, V., Sullivan, J. P., Sun, L., Turner, S., and Karsch-Mizrachi, I. (2020). NCBI Taxonomy: A comprehensive update on curation, resources and tools. *Database: The Journal of Biological Databases and Curation*, 2020:baaa062.

Schröder, C. (2017). Proteins in Ionic Liquids: Current Status of Experiments and Simulations. *Topics in Current Chemistry*, 375(2):25.

Schroeder, S. J. (2018). Challenges and approaches to predicting RNA with multiple functional structures. *RNA*, 24(12):1615–1624.

Schwalbe, H., Audergon, P., Haley, N., Amaro, C. A., Agirre, J., Baldus, M., Banci, L., Baumeister, W., Blackledge, M., Carazo, J. M., Carugo, K. D., Celie, P., Felli, I., Hart, D. J., Hauß, T., Lehtiö, L., Lindorff-Larsen, K., Márquez, J., Matagne, A., Pierattelli, R., Rosato, A., Sobott, F., Sreeramulu, S., Steyaert, J., Sussman, J. L., Trantirek, L., Weiss, M. S., and Wilmanns, M. (2024). The future of integrated structural biology. *Structure*.

Sehgal, S. N., Kates, M., and Gibbons, N. E. (1962). Lipids of Halobacterium cutirubrum. *Canadian Journal of Biochemistry and Physiology*, 40:69–81.

Seiferth, D. and Biggin, P. C. (2024). Exploring the influence of pore shape on conductance and permeation. *Biophysical Journal*, 123(18):3107–3119.

Sen, T. Z., Cheng, H., Kloczkowski, A., and Jernigan, R. L. (2006). A Consensus Data Mining secondary structure prediction by combining GOR V and Fragment Database Mining. *Protein Science : A Publication of the Protein Society*, 15(11):2499–2506.

Shaw, D. E., Deneroff, M. M., Dror, R. O., Kuskin, J. S., Larson, R. H., Salmon, J. K., Young, C., Batson, B., Bowers, K. J., Chao, J. C., Eastwood, M. P., Gagliardo, J., Grossman, J. P., Ho, C. R., Ierardi, D. J., Kolossváry, I., Klepeis, J. L., Layman, T., McLeavey, C., Moraes, M. A., Mueller, R., Priest, E. C., Shan, Y., Spengler, J., Theobald, M., Towles, B., and Wang, S. C. (2008). Anton, a special-purpose machine for molecular dynamics simulation. *Commun. ACM*, 51(7):91–97.

Shaw, E. A., Thomas, N. K., Jones, J. D., Abu-Shumays, R. L., Vaaler, A. L., Akeson, M., Koutmou, K. S., Jain, M., and Garcia, D. M. (2024). Combining Nanopore direct RNA sequencing with genetics and mass spectrometry for analysis of T-loop base modifications across 42 yeast tRNA isoacceptors. *Nucleic Acids Research*, 52(19):12074–12092.

She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., Chan-Weiher, C. C.-Y., Clausen, I. G., Curtis, B. A., De Moors, A., Erauso, G., Fletcher, C., Gordon, P. M. K., Heikamp-de Jong, I., Jeffries, A. C., Kozera, C. J.,

Medina, N., Peng, X., Thi-Ngoc, H. P., Redder, P., Schenk, M. E., Theriault, C., Tolstrup, N., Charlebois, R. L., Doolittle, W. F., Duguet, M., Gaasterland, T., Garrett, R. A., Ragan, M. A., Sensen, C. W., and Van der Oost, J. (2001). The complete genome of the crenarchaeon Sulfolobus solfataricus P2. *Proceedings of the National Academy of Sciences of the United States of America*, 98(14):7835–7840.

Sheng, K., Li, N., Rabuck-Gibbons, J. N., Dong, X., Lyumkis, D., and Williamson, J. R. (2023). Assembly landscape for the bacterial large ribosomal subunit. *Nature Communications*, 14(1):5220.

Shirts, M. and Pande, V. S. (2000). Screen Savers of the World Unite! *Science*, 290(5498):1903–1904.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7:539.

Sigworth, F. J. (2016). Principles of cryo-EM single-particle image processing. *Microscopy*, 65(1):57–67.

Sillitoe, I., Bordin, N., Dawson, N., Waman, V. P., Ashford, P., Scholes, H. M., Pang, C. S. M., Woodridge, L., Rauer, C., Sen, N., Abbasian, M., Le Cornu, S., Lam, S. D., Berka, K., Varekova, I. H., Svobodova, R., Lees, J., and Orengo, C. A. (2021). CATH: Increased structural coverage of functional space. *Nucleic Acids Research*, 49(D1):D266–D273.

Singh, A. (2023). Nanopores for sequencing proteins. *Nature Methods*, 20(12):1870–1870.

Singh, A. (2024). Automated atomic modeling in cryo-EM maps. *Nature Methods*, 21(5):749–749.

Sleytr, U. B., Schuster, B., Egelseer, E.-M., and Pum, D. (2014). S-layers: Principles and applications. *Fems Microbiology Reviews*, 38(5):823–864.

Smart, O. S., Neduvelil, J. G., Wang, X., Wallace, B. A., and Sansom, M. S. P. (1996). HOLE: A program for the analysis of the pore dimensions of ion channel structural models. *Journal of Molecular Graphics*, 14(6):354–360.

Smith, J. C. and Roux, B. (2013). Eppur Si Muove! The 2013 Nobel Prize in Chemistry. *Structure*, 21(12):2102–2105.

Soares, C. M. and Lousa, D. (2025). Structural Bioinformatics: Exciting times in a rapidly evolving field. *FEBS Open Bio*, 15(2):200–201.

Sogues, A., Leigh, K., Halingstad, E. V., Van der Verren, S. E., Cecil, A. J., Fioravanti, A., Pak, A. J., Kudryashev, M., and Remaut, H. (2024). Architecture of the Sap S-layer of Bacillus anthracis revealed by integrative structural biology. *Proceedings of the National Academy of Sciences*, 121(51):e2415351121.

Song, L., Hobaugh, M. R., Shustak, C., Cheley, S., Bayley, H., and Gouaux, J. E. (1996). Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science (New York, N.Y.)*, 274(5294):1859–1866.

Sonnhammer, E. L., von Heijne, G., and Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. *Proceedings. International Conference on Intelligent Systems for Molecular Biology*, 6:175–182.

Soppa, J. (2006). From genomes to function: Haloarchaea as model organisms. *Microbiology (Reading, England)*, 152(Pt 3):585–590.

Spang, A., Saw, J. H., Jørgensen, S. L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A. E., van Eijk, R., Schleper, C., Guy, L., and Ettema, T. J. G. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*, 521(7551):173–179.

Šponer, J., Bussi, G., Krepl, M., Banáš, P., Bottaro, S., Cunha, R. A., Gil-Ley, A., Pinamonti, G., Poblete, S., Jurečka, P., Walter, N. G., and Otyepka, M. (2018). RNA Structural Dynamics As Captured by Molecular Simulations: A Comprehensive Overview. *Chemical Reviews*, 118(8):4177–4338.

Šponer, J., Šponer, J. E., Petrov, A. I., and Leontis, N. B. (2010). Quantum Chemical Studies of Nucleic Acids: Can We Construct a Bridge to the RNA Structural Biology and Bioinformatics Communities? *The Journal of Physical Chemistry B*, 114(48):15723–15741.

Srivastava, A., Tiwari, S. P., Miyashita, O., and Tama, F. (2020). Integrative/Hybrid Modeling Approaches for Studying Biomolecules. *Journal of Molecular Biology*, 432(9):2846–2860.

Stairs, C. W. and Ettema, T. J. G. (2020). The Archaeal Roots of the Eukaryotic Dynamic Actin Cytoskeleton. *Current Biology*, 30(10):R521–R526.

Stautz, J., Hellmich, Y., Fuss, M. F., Silberberg, J. M., Devlin, J. R., Stockbridge, R. B., and Hänelt, I. (2021). Molecular Mechanisms for Bacterial Potassium Homeostasis. *Journal of Molecular Biology*, 433(16):166968.

Steiner, L. A. (1959). *The Reduction of the Disulfide Bonds of Ribonuclease.* PhD thesis, Yale University.

Stombaugh, J., Zirbel, C. L., Westhof, E., and Leontis, N. B. (2009). Frequency and isostericity of RNA base pairs. *Nucleic Acids Research*, 37(7):2294–2312.

Sumper, M., Berg, E., Mengele, R., and Strobel, I. (1990). Primary structure and glycosylation of the S-layer protein of Haloferax volcanii. *Journal of Bacteriology*, 172(12):7111–7118.

Svozil, D., Hobza, P., and Šponer, J. (2010). Comparison of Intrinsic Stacking Energies of Ten Unique Dinucleotide Steps in A-RNA and B-DNA Duplexes. Can We Determine Correct Order of Stability by Quantum-Chemical Calculations? *The Journal of Physical Chemistry B*, 114(2):1191–1203.

Swart, M., Fonseca Guerra, C., and Bickelhaupt, F. M. (2004). Hydrogen Bonds of RNA Are Stronger than Those of DNA, but NMR Monitors Only Presence of Methyl Substituent in Uracil/Thymine. *Journal of the American Chemical Society*, 126(51):16718–16719.

Sweeney, B. A., Hoksza, D., Nawrocki, E. P., Ribas, C. E., Madeira, F., Cannone, J. J., Gutell, R., Maddala, A., Meade, C. D., Williams, L. D., Petrov, A. S., Chan, P. P., Lowe, T. M., Finn, R. D., and Petrov, A. I. (2021). R2DT is a framework for predicting and visualising RNA secondary structure using templates. *Nature Communications*, 12(1):3494.

Świderek, K., Marti, S., Tuñón, I., Moliner, V., and Bertran, J. (2015). Peptide Bond Formation Mechanism Catalyzed by Ribosome. *Journal of the American Chemical Society*, 137(37):12024–12034.

Szilárd, P., Abraham, M. J., Kutzner, C., Hess, B., and Lindahl, E. (2015). Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. https://arxiv.org/abs/1506.00716v1.

Tagari, M., Newman, R., Chagoyen, M., Carazo, J. M., and Henrick, K. (2002). New electron microscopy database and deposition system. *Trends in Biochemical Sciences*, 27(11):589.

Tan, Y. Z., Baldwin, P. R., Davis, J. H., Williamson, J. R., Potter, C. S., Carragher, B., and Lyumkis, D. (2017). Addressing Preferred Specimen Orientation in Single-Particle Cryo-EM through Tilting. *Nature methods*, 14(8):793–796.

Tang, S. and Yang, J. J. (2013). Magnesium Binding Sites in Proteins. In Kretsinger, R. H., Uversky, V. N., and Permyakov, E. A., editors, *Encyclopedia of Metalloproteins*, pages 1243–1250. Springer, New York, NY.

Tang, W., Fried, J. P., Tilley, R. D., and Gooding, J. J. (2022). Understanding and modelling the magnitude of the change in current of nanopore sensors. *Chemical Society Reviews*, 51(14):5757–5776.

Taylor, B. (1715). *Methodus incrementorum directa & inversa. Auctore Brook Taylor, ... 1715.*

Terzibaschian, T. and Enderlein, B. (1986). The irreducible representations of the two-dimensional space groups of crystal surfaces. Theory and applications. *physica status solidi (b)*, 133(2):443–461.

The UniProt Consortium (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, 49(D1):D480–D489.

The UniProt Consortium (2025). UniProt: The Universal Protein Knowledgebase in 2025. *Nucleic Acids Research*, 53(D1):D609–D617.

Theobald, D. L. (2005). Rapid calculation of RMSDs using a quaternion-based characteristic polynomial. *Acta Crystallographica. Section A, Foundations of Crystallography*, 61(Pt 4):478–480.

Thirumalai, D. and Hyeon, C. (2005). RNA and Protein Folding: Common Themes and Variations. *Biochemistry*, 44(13):4957–4970.

Thirumalai, D., Klimov, D. K., and Woodson, S. A. (1997). Kinetic partitioning mechanism as a unifying theme in the folding of biomolecules. *Theoretical Chemistry Accounts*, 96(1):14–22.

Thirumalai, D. and Woodson, S. A. (1996). Kinetics of Folding of Proteins and RNA. *Accounts of Chemical Research*, 29(9):433–439.

Thon, F. (1966). Notizen: Zur Defokussierungsabhängigkeit des Phasenkontrastes bei der elektronenmikroskopischen Abbildung. *Zeitschrift für Naturforschung A*, 21(4):476–478.

Timsit, Y., Sergeant-Perthuis, G., and Bennequin, D. (2021). Evolution of ribosomal protein network architectures. *Scientific Reports*, 11(1):625.

Tinoco, I. and Bustamante, C. (1999). How RNA folds. *Journal of Molecular Biology*, 293(2):271–281.

Toader, B., Sigworth, F. J., and Lederman, R. R. (2023). Methods for Cryo-EM Single Particle Reconstruction of Macromolecules Having Continuous Heterogeneity. *Journal of Molecular Biology*, 435(9):168020.

Toyoshima, C. and Unwin, N. (1988). Contrast transfer for frozen-hydrated specimens: Determination from pairs of defocused images. *Ultramicroscopy*, 25(4):279–291.

Trabuco, L. G., Schreiner, E., Gumbart, J., Hsin, J., Villa, E., and Schulten, K. (2011). Applications of the molecular dynamics flexible fitting method. *Journal of structural biology*, 173(3):420–427.

Trabuco, L. G., Villa, E., Mitra, K., Frank, J., and Schulten, K. (2008). Flexible Fitting of Atomic Structures into Electron Microscopy Maps Using Molecular Dynamics. *Structure (London, England : 1993)*, 16(5):673–683.

Trabuco, L. G., Villa, E., Schreiner, E., Harrison, C. B., and Schulten, K. (2009). Molecular Dynamics Flexible Fitting: A practical guide to combine cryo-electron microscopy and X-ray crystallography. *Methods (San Diego, Calif.)*, 49(2):174–180.

Trivedi, R. and Nagarajaram, H. A. (2022). Intrinsically Disordered Proteins: An Overview. *International Journal of Molecular Sciences*, 23(22):14050.

Turoňová, B., Schur, F. K., Wan, W., and Briggs, J. A. (2017). Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4 Å. *Journal of Structural Biology*, 199(3):187–195.

Uehata, T. and Takeuchi, O. (2020). RNA Recognition and Immunity—Innate Immune Sensing and Its Posttranscriptional Regulation Mechanisms. *Cells*, 9(7):1701.

Uusitalo, J. J., Ingólfsson, H. I., Marrink, S. J., and Faustino, I. (2017). Martini Coarse-Grained Force Field: Extension to RNA. *Biophysical Journal*, 113(2):246–256.

Vagin, A. and Teplyakov, A. (1997). MOLREP: An Automated Program for Molecular Replacement. *Journal of Applied Crystallography*, 30(6):1022–1025.

Vakonakis, I. and LiWang, A. C. (2004). N1$\cdots$N3 Hydrogen Bonds of A:U Base Pairs of RNA Are Stronger than Those of A:T Base Pairs of DNA [J. Am. Chem. Soc. 2004, 18, 5688-5689]. *Journal of the American Chemical Society*, 126(22):7152–7152.

Vallat, B., Rose, Y., Piehl, D. W., Duarte, J. M., Bittrich, S., Bi, C., Segura, J., Zalevsky, A., Sekharan, M. R., Webb, B. M., Bhikadiya, C., Voigt, M., Henry, J., Chao, H., Pingale, A., Raymond, D., Sagendorf, J., Peisach, E., Feng, Z., Guranovic, V., Smith, J., Rose, A. S., Young, J. Y., Berman, H. M., Sali, A., and Burley, S. K. (2025). RCSB Protein Data Bank: Delivering integrative structures alongside experimental structures and computed structure models. *Nucleic Acids Research*, page gkaf1187.

van den Berg, B., Silale, A., Baslé, A., Brandner, A. F., Mader, S. L., and Khalid, S. (2022). Structural basis for host recognition and superinfection exclusion by bacteriophage T5. *Proceedings of the National Academy of Sciences of the United States of America*, 119(42):e2211672119.

van der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D. T., Kim, P. M., Kriwacki, R. W., Oldfield, C. J., Pappu, R. V., Tompa, P., Uversky, V. N., Wright, P. E., and Babu, M. M. (2014). Classification of Intrinsically Disordered Regions and Proteins. *Chemical Reviews*, 114(13):6589–6631.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. C. (2005). GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry*, 26(16):1701–1718.

van der Walt, S., Colbert, S. C., and Varoquaux, G. (2011). The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science and Engg.*, 13(2):22–30.

van der Wel, P. C. (2018). New applications of solid-state NMR in structural biology. *Emerging Topics in Life Sciences*, 2(1):57–67.

Van Noorden, R. (2009). Ribosome clinches the chemistry Nobel. *Nature*.

Vangaveti, S., Ranganathan, S. V., and Chen, A. A. (2017). Advances in RNA molecular dynamics: A simulator's guide to RNA force fields. *WIREs RNA*, 8(2):e1396.

Vanommeslaeghe, K. and MacKerell, A. D. (2015). CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. *Biochimica et biophysica acta*, 1850(5):861–871.

Vant, J. W., Sarkar, D., Nguyen, J., Baker, A. T., Vermaas, J. V., and Singharoy, A. (2022). Exploring cryo-electron microscopy with molecular dynamics. *Biochemical Society Transactions*, 50(1):569–581.

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Žídek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., Figurnov, M., Cowie, A., Hobbs, N., Kohli, P., Kleywegt, G., Birney, E., Hassabis, D., and Velankar, S. (2022). AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, 50(D1):D439–D444.

Veith, A., Klingl, A., Zolghadr, B., Lauber, K., Mentele, R., Lottspeich, F., Rachel, R., Albers, S.-V., and Kletzin, A. (2009). Acidianus, Sulfolobus and Metallosphaera surface layers: Structure, composition and gene expression. *Molecular Microbiology*, 73(1):58–72.

Verlet, L. (1967). Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review*, 159(1):98–103.

Vicens, Q. and Kieft, J. S. (2022). Thoughts on how to think (and talk) about RNA structure. *Proceedings of the National Academy of Sciences*, 119(17):e2112677119.

Villa, E., Schaffer, M., Plitzko, J. M., and Baumeister, W. (2013). Opening windows into the cell: Focused-ion-beam milling for cryo-electron tomography. *Current Opinion in Structural Biology*, 23(5):771–777.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., and van Mulbregt, P. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272.

Vlachakis, D., Bencurova, E., Papangelopoulos, N., and Kossida, S. (2014). Current State-of-the-Art Molecular Dynamics Methods and Applications. In *Advances in Protein Chemistry and Structural Biology*, volume 94. Elsevier.

von Heijne, G. (1986). A new method for predicting signal sequence cleavage sites. *Nucleic Acids Research*, 14(11):4683–4690.

von Kügelgen, A., Alva, V., and Bharat, T. A. M. (2021). Complete atomic structure of a native archaeal cell surface. *Cell Reports*, 37(8):110052.

von Kügelgen, A., Cassidy, C. K., van Dorst, S., Pagani, L. L., Batters, C., Ford, Z., Löwe, J., Alva, V., Stansfeld, P. J., and Bharat, T. A. M. (2024). Membraneless channels sieve cations in ammonia-oxidizing marine archaea. *Nature*, 630(8015):230–236.

von Kügelgen, A., Tang, H., Hardy, G. G., Kureisaite-Ciziene, D., Brun, Y. V., Stansfeld, P. J., Robinson, C. V., and Bharat, T. A. (2020). In Situ Structure of an Intact Lipopolysaccharide-Bound Bacterial Surface Layer. *Cell*, 180(2):348–358.e15.

Vuillemot, R., Mirzaei, A., Harastani, M., Hamitouche, I., Fréchin, L., Klaholz, B. P., Miyashita, O., Tama, F., Rouiller, I., and Jonic, S. (2023). MDSPACE: Extracting Continuous Conformational Landscapes from Cryo-EM Single Particle Datasets Using 3D-to-2D Flexible Fitting based on Molecular Dynamics Simulation. *Journal of Molecular Biology*, 435(9):167951.

Walker, C. B., de la Torre, J. R., Klotz, M. G., Urakawa, H., Pinel, N., Arp, D. J., Brochier-Armanet, C., Chain, P. S. G., Chan, P. P., Gollabgir, A., Hemp, J., Hügler, M., Karr, E. A., Könneke, M., Shin, M., Lawton, T. J., Lowe, T., Martens-Habbena, W., Sayavedra-Soto, L. A., Lang, D., Sievert, S. M., Rosenzweig, A. C., Manning, G., and Stahl, D. A. (2010). Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proceedings of the National Academy of Sciences of the United States of America*, 107(19):8818–8823.

Wallin, E. and von Heijne, G. (1998). Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Science: A Publication of the Protein Society*, 7(4):1029–1038.

Wang, J., Dasgupta, I., and Fox, G. E. (2009). Many nonuniversal archaeal ribosomal proteins are found in conserved gene clusters. *Archaea*, 2(4):241–251.

Wang, W., Feng, C., Han, R., Wang, Z., Ye, L., Du, Z., Wei, H., Zhang, F., Peng, Z., and Yang, J. (2023a). trRosettaRNA: Automated prediction of RNA 3D structure with transformer network. *Nature Communications*, 14(1):7266.

Wang, Y., Cao, Z., and Barati Farimani, A. (2021a). Efficient water desalination with graphene nanopores obtained using artificial intelligence. *npj 2D Materials and Applications*, 5(1):1–9.

Wang, Y., Zhao, Y., Bollas, A., Wang, Y., and Au, K. F. (2021b). Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology*, 39(11):1348–1365.

Wang, Y.-H., Dai, H., Zhang, L., Wu, Y., Wang, J., Wang, C., Xu, C.-H., Hou, H., Yang, B., Zhu, Y., Zhang, X., and Zhou, J. (2023b). Cryo-electron microscopy structure and translocation mechanism of the crenarchaeal ribosome. *Nucleic Acids Research*, 51(17):8909–8924.

Wankowicz, S. A. (2025). Ten rules for a structural bioinformatic analysis. *PLOS Computational Biology*, 21(10):e1013094.

Ward, A. B., Sali, A., and Wilson, I. A. (2013). Integrative Structural Biology. *Science (New York, N.Y.)*, 339(6122):913–915.

Warner, J. R., Knopf, P. M., and Rich, A. (1963). A multiple ribosomal structure in protein synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 49(1):122–129.

Warner, J. R. and McIntosh, K. B. (2009). How Common Are Extraribosomal Functions of Ribosomal Proteins? *Molecular Cell*, 34(1):3–11.

Waterston, J. J. and Strutt, J. W. (1892). On the physics of media that are composed of free and perfectly elastic molecules in a state of motion. *Philosophical Transactions of the Royal Society of London. (A.)*, (183):1–79.

Watson, J. D. and Crick, F. H. C. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, 171(4356):737–738.

Weis, W. I., Kahn, R., Fourme, R., Drickamer, K., and Hendrickson, W. A. (1991). Structure of the calcium-dependent lectin domain from a rat mannose-binding protein determined by MAD phasing. *Science (New York, N.Y.)*, 254(5038):1608–1615.

Westhof, E. and Fritsch, V. (2000). RNA folding: Beyond Watson–Crick pairs. *Structure*, 8(3):R55–R65.

Whiddington, R. and Taylor, J. E. (1932). On the photographic action of slow electrons. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 136(830):651–662.

Wiener, N. (1949). *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications*. The MIT Press.

Willems, K., Ruić, D., Lucas, F. L. R., Barman, U., Verellen, N., Hofkens, J., Maglia, G., and Dorpe, P. V. (2020). Accurate modeling of a biological nanopore with an extended continuum framework. *Nanoscale*, 12(32):16775–16795.

Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M., Arendall, W. B., Snoeyink, J., Adams, P. D., Lovell, S. C., Richardson, J. S., and Richardson, D. C. (2018). MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science : A Publication of the Protein Society*, 27(1):293–315.

Williams, T. A., Cox, C. J., Foster, P. G., Szöllősi, G. J., and Embley, T. M. (2020). Phylogenomics provides robust support for a two-domains tree of life. *Nature ecology & evolution*, 4(1):138–147.

Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. *Nature*, 407(6802):327–339.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography*, 67(4):235–242.

Wiswedel, R., Bui, A. T. N., Kim, J., and Lee, M.-K. (2024). Beta-Barrel Nanopores as Diagnostic Sensors: An Engineering Perspective. *Biosensors*, 14(7):345.

Wittmann, H. G., Stöffler, G., Hindennach, I., Kurland, C. G., Randall-Hazelbauer, L., Birge, E. A., Nomura, M., Kaltschmidt, E., Mizushima, S., Traut, R. R., and Bickle, T. A. (1971). Correlation of 30S ribosomal proteins ofEscherichia coli isolated in different laboratories. *Molecular and General Genetics MGG*, 111(4):327–333.

Wittmann-Liebold, B. (1986). Ribosomal Proteins: Their Structure and Evolution. In Hardesty, B. and Kramer, G., editors, *Structure, Function, and Genetics of Ribosomes*, pages 326–361. Springer, New York, NY.

Wloka, C., Mutter, N. L., Soskine, M., and Maglia, G. (2016). Alpha-Helical Fragaceatoxin C Nanopore Engineered for Double-Stranded and Single-Stranded Nucleic Acid Analysis. *Angewandte Chemie International Edition*, 55(40):12494–12498.

Woese, C. R. and Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences*, 74(11):5088–5090.

Woese, C. R., Fox, G. E., Zablen, L., Uchida, T., Bonen, L., Pechman, K., Lewis, B. J., and Stahl, D. (1975). Conservation of primary structure in 16S ribosomal RNA. *Nature*, 254(5495):83–86.

Woese, C. R., Kandler, O., and Wheelis, M. L. (1990). Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America*, 87(12):4576–4579.

Wolfe, A. J., Mohammad, M. M., Thakur, A. K., and Movileanu, L. (2016). Global Redesign of a Native $\beta$-barrel Scaffold. *Biochimica et biophysica acta*, 1858(1):19–29.

Wu, S., Armache, J.-P., and Cheng, Y. (2016). Single-particle cryo-EM data acquisition by using direct electron detection camera. *Microscopy*, 65(1):35–41.

Wu, Y., Shao, W., Yan, M., Wang, Y., Xu, P., Huang, G., Li, X., Gregory, B. D., Yang, J., Wang, H., and Yu, X. (2024). Transfer learning enables identification of multiple types of RNA modifications using nanopore direct RNA sequencing. *Nature Communications*, 15(1):4049.

Ying, Y.-L., Hu, Z.-L., Zhang, S., Qing, Y., Fragasso, A., Maglia, G., Meller, A., Bayley, H., Dekker, C., and Long, Y.-T. (2022). Nanopore-based technologies beyond DNA sequencing. *Nature Nanotechnology*, 17(11):1136–1146.

Yip, K. M., Fischer, N., Paknia, E., Chari, A., and Stark, H. (2020). Atomic-resolution protein structure determination by cryo-EM. *Nature*, 587(7832):157–161.

Yonath, A., Leonard, K. R., and Wittmann, H. G. (1987). A Tunnel in the Large Ribosomal Subunit Revealed by Three-Dimensional Image Reconstruction. *Science*, 236(4803):813–816.

Yu, Y., Wang, J., Shao, Q., Shi, J., and Zhu, W. (2016). The effects of organic solvents on the folding pathway and associated thermodynamics of proteins: A microscopic view. *Scientific Reports*, 6(1):19500.

Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H. D., and Noller, H. F. (2001). Crystal Structure of the Ribosome at 5.5 Å Resolution. *Science*, 292(5518):883–896.

Zadorozhnyi, R., Gronenborn, A. M., and Polenova, T. (2024). Integrative Approaches for Characterizing Protein Dynamics: NMR, CryoEM and Computer Simulations. *Current opinion in structural biology*, 84:102736.

Zaremba-Niedzwiedzka, K., Caceres, E. F., Saw, J. H., Bäckström, D., Juzokaite, L., Vancaester, E., Seitz, K. W., Anantharaman, K., Starnawski, P., Kjeldsen, K. U., Stott, M. B., Nunoura, T., Banfield, J. F., Schramm, A., Baker, B. J., Spang, A., and Ettema, T. J. G. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature*, 541(7637):353–358.

Zemora, G. and Waldsich, C. (2010). RNA folding in living cells. *RNA Biology*, 7(6):634–641.

Zeth, K., Römer, C., Patzer, S. I., and Braun, V. (2008). Crystal Structure of Colicin M, a Novel Phosphatase Specifically Imported by Escherichia coli>. *The Journal of Biological Chemistry*, 283(37):25324–25331.

Zhang, J., Feng, X., Li, M., Liu, Y., Liu, M., Hou, L.-J., and Dong, H.-P. (2025). Deep origin of eukaryotes outside Heimdallarchaeia within Asgardarchaeota. *Nature*, 642(8069):990–998.

Zhang, M., Chen, C., Zhang, Y., and Geng, J. (2022). Biological nanopores for sensing applications. *Proteins: Structure, Function, and Bioinformatics*, 90(10):1786–1799.

Zhang, Y.-S. and Ding, M. (2025). Probing Nanopores: Molecular Dynamics Insights into the Mechanisms of DNA and Protein Translocation through Solid-State and Biological Nanopores. *Soft Matter*.

Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, 7(1-2):203–214.

Zheng, H., Porebski, P. J., Grabowski, M., Cooper, D. R., and Minor, W. (2017a). Databases, Repositories and Other Data Resources in Structural Biology. *Methods in molecular biology (Clifton, N.J.)*, 1607:643–665.

Zheng, S. Q., Palovcak, E., Armache, J.-P., Verba, K. A., Cheng, Y., and Agard, D. A. (2017b). MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods*, 14(4):331–332.

Zhong, E. D., Bepler, T., Berger, B., and Davis, J. H. (2021). CryoDRGN: Reconstruction of heterogeneous cryo-EM structures using neural networks. *Nature methods*, 18(2):176–185.

Zhong, H. and Barrientos, A. (2024). The zinc finger motif in the mitochondrial large ribosomal subunit protein bL36m is essential for optimal yeast mitoribosome assembly and function. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1871(4):119707.

Zhou, W., Qiu, H., Guo, Y., and Guo, W. (2020). Molecular Insights into Distinct Detection Properties of $\alpha$-Hemolysin, MspA, CsgG, and Aerolysin Nanopore Sensors. *The Journal of Physical Chemistry B*, 124(9):1611–1618.

Ziegler, S. J., Mallinson, S. J. B., St. John, P. C., and Bomble, Y. J. (2021). Advances in integrative structural biology: Towards understanding protein complexes in their cellular context. *Computational and Structural Biotechnology Journal*, 19:214–225.

Zillig, W., Stetter, K. O., Wunderl, S., Schulz, W., Priess, H., and Scholz, I. (1980). The Sulfolobus-"Caldariella" group: Taxonomy on the basis of the structure of DNA-dependent RNA polymerases. *Archives of Microbiology*, 125(3):259–269.

Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J., Lindahl, E., and Scheres, S. H. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife*, 7:e42166.

Zivanov, J., Nakane, T., and Scheres, S. H. W. (2019). A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ*, 6(1):5–17.

Zivanov, J., Nakane, T., and Scheres, S. H. W. (2020). Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1. *IUCrJ*, 7(Pt 2):253–267.

Zivanov, J., Otón, J., Ke, Z., von Kügelgen, A., Pyle, E., Qu, K., Morado, D., Castaño-Díez, D., Zanetti, G., Bharat, T. A., Briggs, J. A., and Scheres, S. H. (2022). A Bayesian approach to single-particle electron cryo-tomography in RELION-4.0. *eLife*, 11:e83724.