



**HAL**  
open science

# Cryo-EM characterisation of an enterobacterial stress response system

Matthew Jessop

► **To cite this version:**

Matthew Jessop. Cryo-EM characterisation of an enterobacterial stress response system. Human health and pathology. Université Grenoble Alpes [2020-..], 2020. English. NNT : 2020GRALV004 . tel-03187407

**HAL Id: tel-03187407**

**<https://theses.hal.science/tel-03187407>**

Submitted on 1 Apr 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## THÈSE

Pour obtenir le grade de

### **DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES**

Spécialité : Biologie Structurale et Nanobiologie

Arrêté ministériel : 25 mai 2016

Présentée par

**Matthew JESSOP**

Thèse dirigée par Irina GUTSCHE

préparée au sein du **Laboratoire Institut de Biologie Structurale**  
dans l'**École Doctorale Chimie et Sciences du Vivant**

### **Cryo-EM characterisation of an enterobacterial stress response system**

### **Caractérisation d'un système entérobactérien de réponse au stress par cryo-EM**

Thèse soutenue publiquement le **19 février 2020**,  
devant le jury composé de :

**Madame MALENE RINGKJØBING JENSEN**

DIRECTRICE DE RECHERCHE, CNRS DELEGATION ALPES  
Présidente

**Monsieur HELGO SCHMIDT**

CHARGE DE RECHERCHE HDR, CNRS DELEGATION ALSACE  
Rapporteur

**Monsieur WOJTEK GALEJ**

DOCTEUR-CHERCHEUR, EMBL GRENOBLE  
Examineur

**Madame SLAVICA JONIC**

DIRECTRICE DE RECHERCHE, CNRS DELEGATION PARIS-CENTRE  
Rapporteuse

**Madame PETRA WENDLER**

PROFESSEUR, UNIVERSITE DE POTSDAM - ALLEMAGNE  
Examinatrice

**Madame IRINA GUTSCHE**

DIRECTRICE DE RECHERCHE, CNRS DELEGATION ALPES  
Directrice de thèse





## i. Acknowledgements

A PhD thesis is not an individual effort, but is rather the result of hours of collaboration, teaching, work, and discussion with many other people. I'm forever grateful to all those who have helped me, inspired me, made me think, and given advice over the course of my PhD.

To Irina, Jan, Ambroise, Alister, Karine, Angélique, Clarissa, Maria, Megghane, Isai, Diego and Ladislav – not only were you wonderful colleagues, but amazing friends. I couldn't have wished for a better group of people to work with. Whether I worked with you for a few months or for the whole four years, you all contributed to a great team atmosphere and made coming to work enjoyable.

Irina, thank you for all of your guidance over these few years. I couldn't have asked for a better mentor during my PhD, and I appreciate everything you've done for me. You always had time to listen to my ideas and answer questions, and to offer invaluable advice and suggestions. You encouraged me when things weren't going well, and celebrated with me when we were successful. You taught me what it means to be a microscopist, I'll miss our discussions and working together.

Jan, it's been a pleasure to work so closely with you over the past few years, I think we made a great team. I'll miss our BLI days and synchrotron sessions, and I'll miss working on papers and travelling to conferences with you and Irina. Thanks for being a great friend, I've learned so much from you and I'm looking forward to seeing where your scientific career goes.

Maria, thank you for the many hours you spent with me at the Polara and the Glacios screening grids and collecting data. I really enjoyed working with you, and even when we had frustrating or unsuccessful sessions you were always happy to try again and persevere. I'm grateful for your patience and your expertise, and a large part of the success of this thesis is thanks to you.

Angélique and Karine, thank you both for all of your help with carrying out experiments, for your work purifying proteins, and for all of the work you do behind the scenes in the lab that keep everything running smoothly. Angélique, I really appreciate all of your help with various administrative things over the years and particularly with organising the thesis defence, it really made a big difference at a stressful time to have someone I could rely on helping out.

Ambroise, thanks for all of the help you gave me during the PhD, and for continuing to teach me image processing. No matter when I asked, you always had the time to explain things to me, or to help with problems I had. Thanks as well for suggesting that I should come to Grenoble in the first place - it's because of you that I ended up where I am today!

Ali, thanks for being the person I could always count on to go for a coffee break! I'll miss our chats – not only the science-related ones, but also the ones about life in general. In those times where I was feeling frustrated, you'd always listen and I really appreciate it.

Thanks also to all the people from the MEM group who have helped me over the past few years, particularly Guy, Daphna, and Emmanuelle for your help with microscopes and the maintenance of the EM platform. Thanks as well to Leandro for your help with many image processing-related questions, and to H el ene and Beno ıt for the coffee breaks and for being great people to work with.

I'd like to thank the members of my thesis advisory committee – Hugues Nury, Christine Cavazza and Montse Soler-L opez. Thank you for taking the time to familiarise yourselves with my project and for your advice, I really appreciated your input. In addition, thanks to Montse and Gabriele Giachin for being wonderful collaborators. I really enjoyed working with you, and I'm proud of the manuscript we produced together.

To the members of the jury – thank you for the time spent reading my thesis, for your comments and suggestions, and for all of the discussion during the defence. Particularly to Slavica Jonic and Helgo Schmidt, thank you for your detailed reports on the thesis manuscript and for your kind words.

To the many friends I've made and people I've met in Grenoble, both through science and in the community, thanks for making these four years unforgettable. Keep in touch, and know that wherever in the world you end up, you'll always have a place to stay.

Finally, thank you to my parents. It hasn't always been straightforward living so far away from home, but your constant love and support has made it easier and is so appreciated. I wouldn't be where I am today without your continual encouragement and belief in me. Thank you for everything.

## ii. Abstract

The field of electron microscopy has undergone rapid transformation in the past few years. It has become one of the most powerful techniques for investigating macromolecular complexes, providing unprecedented structural insights into the fundamental processes of cellular life. In this thesis, I use cryo-electron microscopy and negative stain electron microscopy to investigate two biological systems. The first is an enterobacterial protein triad formed of the *E. coli* proteins LdcI, involved in the acid stress response, and RavA and ViaA, which sensitise *E. coli* to aminoglycosides. I show that the AAA+ ATPase RavA exists in two distinct conformational states, yielding insight into its ATPase mechanism and revealing unexpected mechanistic similarities to the well-characterised unfoldase ClpX. I also explore the effects of fluorescent protein tags on the structure and function of LdcI, in order to facilitate super-resolution fluorescence microscopy. The second complex is the mitochondrial Complex I assembly complex, which is composed of ACAD9, ECSIT and NDUFAF1 and is involved in the maturation of respiratory Complex I in mitochondria. I investigate the ACAD9-ECSIT<sub>CTD</sub> subcomplex using cryo-electron microscopy, providing insights into the structure of ACAD9, the location of the ECSIT<sub>CTD</sub> binding site and the release of the FAD cofactor of ACAD9 upon ECSIT<sub>CTD</sub> binding. Finally, in conjunction with biochemical and biophysical analysis, I place the structural information presented in this thesis in a biological context and lay a platform for future studies of both protein complexes.

**Keywords:** Cryo-EM, AAA+ ATPase, MoxR, RavA, LdcI, MCI complex.



### iii. Résumé

Le domaine de la microscopie électronique a connu une transformation rapide au cours des dernières années. Cette technique est devenue l'une des plus puissantes pour l'étude des complexes macromoléculaires, fournissant des informations structurales sans précédent sur les processus fondamentaux de la vie cellulaire. Dans cette thèse, j'utilise la cryo-microscopie électronique et la microscopie électronique à coloration négative pour étudier deux systèmes biologiques. Le premier est une triade de protéines entérobactériennes formée de la protéine LdcI de *Escherichia coli*, impliquée dans la réponse au stress acide, et RavA et ViaA, qui sensibilisent *E. coli* aux aminoglycosides. Je montre que l'ATPase AAA+ RavA existe dans deux états conformationnels distincts, ce qui permet de mieux comprendre son mécanisme d'hydrolyse de l'ATP et révèle des similarités mécanistiques inattendues avec l'unfoldase ClpX bien caractérisée. J'explore également les effets des marqueurs de protéines fluorescentes sur la structure et la fonction de LdcI, afin de faciliter la microscopie de fluorescence à super-résolution. Le second système est le complexe d'assemblage du Complexe I mitochondrial. Celui-ci est composé d'ACAD9, d'ECSIT et de NDUFAF1 et qui est impliqué dans la maturation du Complexe respiratoire I chez les mitochondries. J'étudie le sous-complexe ACAD9-ECSIT<sub>CTD</sub> à l'aide de la cryo-microscopie électronique, ce qui me permet de mieux comprendre la structure d'ACAD9, l'emplacement du site de liaison d'ECSIT<sub>CTD</sub> et la libération du cofacteur FAD d'ACAD9 lors de la liaison d'ECSIT<sub>CTD</sub>. Enfin, en combinaison avec des analyses biochimiques et biophysiques, je place les informations structurales présentées dans cette thèse dans un contexte biologique et pose une base de futures études des deux complexes protéiques.

**Mots-clés :** Cryo-EM, AAA+ ATPase, MoxR, RavA, LdcI, complexe MCI A.



## iv. Table of Contents

i. Acknowledgements .....	3
ii. Abstract .....	5
iii. Résumé.....	7
iv. Table of Contents .....	9
v. List of Figures.....	13
I. Introduction .....	17
Preface .....	18
1. Principles and developments in cryo-EM .....	20
1.1. Historical background .....	20
1.2. Principles of cryo-EM .....	24
1.3. The image processing pipeline for single-particle electron microscopy ..	32
1.4. Modern electron microscopy .....	35
1.5. Image processing developments – a new set of tools in the microscopist’s toolbox .....	43
1.6. The next revolution – sample preparation .....	55
1.7. Future directions.....	60
2. Evolutionary history and structural features of AAA+ ATPases.....	64
2.1. ATP – a universal energy currency .....	65
2.2. P-loop NTPases.....	66
2.3. ASCE division ATPases.....	71
2.4. The AAA+ superfamily.....	75
2.5. AAA+ ATPase clades, and defining an ATPase mechanism.....	76
2.6. Structure and function of the AAA+ clades .....	81
3. The MoxR AAA+ ATPase RavA.....	99
3.1. The MoxR family.....	99
3.2. The <i>ravAviaA</i> operon .....	100

3.3.	RavA interacts with the inducible lysine decarboxylase LdcI.....	100
3.4.	RavA and LdcI X-ray crystal structures .....	103
3.5.	The cryo-EM structure of the LdcI-RavA cage .....	106
3.6.	The biological role of the LdcI-RavA cage .....	108
3.7.	The goal of the thesis.....	109
II.	Results .....	111
4.	Manuscript on RavA and the LdcI-RavA cage .....	112
5.	Towards the cryo-EM characterisation of RavA .....	139
5.1.	BC2-tagged RavA .....	140
5.2.	Stabilising the RavA hexamer – ATPase mutants .....	143
5.3.	RavA + ADP – additional image processing steps.....	146
5.4.	RavA + ATPγS dataset.....	149
5.5.	Stabilising the substrate-bound conformation .....	151
6.	Manuscript on LdcI .....	153
7.	RavA fluorescent fusions – a cautionary tale .....	190
7.1.	Expression and purification of fluorescently-tagged RavA .....	191
7.2.	Negative stain EM .....	193
7.3.	MALLS .....	195
7.4.	Cryo-EM .....	196
7.5.	Mass spectrometry and N-terminal sequencing .....	200
7.6.	Arn-believable .....	201
8.	Manuscript on the MCIA Complex.....	204
9.	Towards the structural characterisation of the MCIA complex .....	251
9.1.	Negative stain EM of the ACAD9-ECSIT-NDUFAF1 triple complex.....	252
9.2.	ACAD9-ECSIT <sub>CTD</sub> sample preparation and cryo-EM.....	255
9.3.	Cryo-EM on ACAD9 alone with CHAPSO.....	262
III.	Discussion.....	265
10.	Perspectives and future directions .....	266

10.1. Improving the RavA reconstruction – overcoming preferential orientation .....	266
10.2. The search for the biological function of the LdcI-RavA-ViaA triad .....	267
10.3. Functional implications of LdcI stack formation .....	268
10.4. MCIa complex structural characterisation.....	269
IV. Bibliography .....	271
V. Appendix .....	290



## v. List of Figures

<b>Figure 1.1.</b>	Early light microscopy.....	20
<b>Figure 1.2.</b>	Comparison between light and electron microscopes.....	21
<b>Figure 1.3.</b>	Early negative stain electron microscopy.....	22
<b>Figure 1.4.</b>	Negative stain EM micrograph of human papillomavirus.....	22
<b>Figure 1.5.</b>	Cryo-EM micrograph of Semliki Forest virus.....	23
<b>Figure 1.6.</b>	The effect of defocusing on cryo-EM image contrast.....	25
<b>Figure 1.7.</b>	The effect of defocussing on the CTF.....	26
<b>Figure 1.8.</b>	Radiation damage under the electron beam.....	27
<b>Figure 1.9.</b>	Common experimental setup for freezing cryo-EM grids.....	28
<b>Figure 1.10.</b>	Averaging of noisy particle images in cryo-EM.....	30
<b>Figure 1.11.</b>	The central slice theorem.....	31
<b>Figure 1.12.</b>	B-factor sharpening of cryo-EM maps.....	34
<b>Figure 1.13.</b>	Scanning electron microscopy of three different electron sources...	35
<b>Figure 1.14.</b>	DQE comparison between different electron detectors.....	37
<b>Figure 1.15.</b>	The effects of motion correction on cryo-EM images.....	38
<b>Figure 1.16.</b>	Automatic data collection set-up in EPU.....	40
<b>Figure 1.17.</b>	New techniques for high-throughput data collection.....	41
<b>Figure 1.18.</b>	The effect of phase plates on cryo-EM image contrast.....	42
<b>Figure 1.19.</b>	Limits on resolution by errors in defocus estimation.....	44
<b>Figure 1.20.</b>	Random conical tilt for generating initial models.....	47
<b>Figure 1.21.</b>	Overview of signal subtraction.....	50
<b>Figure 1.22.</b>	Multi-body refinement in RELION.....	51
<b>Figure 1.23.</b>	Non-uniform refinement in CryoSPARC.....	53
<b>Figure 1.24.</b>	Variation in ice thickness across a cryo-EM grid.....	56
<b>Figure 1.25.</b>	The Spotiton system for freezing cryo-EM grids.....	57
<b>Figure 1.26.</b>	Preferential orientation of particles at the pair-water interface.....	58
<b>Figure 1.27.</b>	The effect of denoising on cryo-EM image contrast.....	61
<b>Figure 1.28.</b>	On-the-fly processing in WARP.....	62
<b>Figure 1.29.</b>	The CryoFLARE graphical user interface.....	63
<b>Figure 2.1.</b>	Structures of ATP and GTP.....	65

<b>Figure 2.2.</b>	$\beta$ -strand arrangement in KG and ASCE ATPases.....	67
<b>Figure 2.3.</b>	NTP binding site in KG and ASCE ATPases.....	68
<b>Figure 2.4.</b>	Conservation of catalytic residues in p97, Ras and Rap1GAP.....	70
<b>Figure 2.5.</b>	Structural features of ASCE superfamilies.....	71
<b>Figure 2.6.</b>	Cryo-EM structure of FtsK.....	72
<b>Figure 2.7.</b>	Structural comparison of selected STAND ATPases.....	74
<b>Figure 2.8.</b>	AAA+ subdomain arrangement in the ClpX monomer.....	75
<b>Figure 2.9.</b>	Defining structural features of the seven AAA+ ATPase clades.....	77
<b>Figure 2.10.</b>	Cryo-EM structures of VAT in closed and split ring states.....	78
<b>Figure 2.11.</b>	Summary of the 'hand-over-hand' ATPase mechanism.....	79
<b>Figure 2.12.</b>	Structure of the RFC clamp loader.....	81
<b>Figure 2.13.</b>	Mechanism of clamp loader action.....	82
<b>Figure 2.14.</b>	Structures of bacterial initiator ATPases.....	83
<b>Figure 2.15.</b>	Mechanism of DnaB loading by DnaC.....	84
<b>Figure 2.16.</b>	Structure of the yeast ORC/Cdc6-Cdt1/MCM/DNA complex.....	85
<b>Figure 2.17.</b>	Mechanism of MCM loading by ORC/Cdc6.....	86
<b>Figure 2.18.</b>	Structures of spastin and ClpB bound to substrate.....	88
<b>Figure 2.19.</b>	Structure of the LTag helicase bound to double-stranded DNA.....	89
<b>Figure 2.20.</b>	Structure of the E1 helicase bound to single-stranded DNA.....	90
<b>Figure 2.21.</b>	Structure of the ClpXP complex.....	91
<b>Figure 2.22.</b>	Analysis of PSI-inserts and H2-inserts in clade 5 ATPases.....	93
<b>Figure 2.23.</b>	Structures of the NtrC1 heptamer and the McrBC complex.....	95
<b>Figure 2.24.</b>	Monomer structures of NtrC1 and RavA.....	96
<b>Figure 2.25.</b>	Structure of human dynein.....	97
<b>Figure 3.1.</b>	Conversion of lysine to cadaverine by LdcI.....	100
<b>Figure 3.2.</b>	Negative stain EM characterisation of LdcI.....	101
<b>Figure 3.3.</b>	Negative stain EM characterisation of the LdcI-RavA complex.....	102
<b>Figure 3.4.</b>	X-ray crystal structure of the RavA monomer.....	103
<b>Figure 3.5.</b>	Negative stain EM analysis of the RavA hexamer.....	104
<b>Figure 3.6.</b>	X-ray crystal structure of the LdcI decamer.....	105
<b>Figure 3.7.</b>	Proposed architecture of the LdcI-RavA complex (2010).....	105
<b>Figure 3.8.</b>	Cryo-EM structure of the LdcI-RavA complex (2014).....	106

<b>Figure 5.1.</b>	Negative stain EM micrograph of RavA-BC2.....	141
<b>Figure 5.2.</b>	Negative stain EM 2D class averages of RavA-BC2.....	141
<b>Figure 5.3.</b>	Purification of RavA-K52Q.....	143
<b>Figure 5.4.</b>	Negative stain EM micrographs and classes of RavA-K52Q.....	144
<b>Figure 5.5.</b>	Negative stain EM 2D class averages of wild-type RavA.....	145
<b>Figure 5.6.</b>	Cryo-EM map of RavA from untilted micrographs.....	146
<b>Figure 5.7.</b>	Anisotropic cryo-EM map of RavA in the C2-symmetric state.....	147
<b>Figure 5.8.</b>	Comparison between anisotropic and isotropic RavA maps.....	148
<b>Figure 5.9.</b>	Cryo-EM maps of RavA in the presence of ATP $\gamma$ S.....	150
<b>Figure 5.10.</b>	Cryo-EM micrograph and classes of RavA + ATP $\gamma$ S + casein.....	152
<b>Figure 7.1.</b>	Purification of Dendra2(T69A)-RavA.....	192
<b>Figure 7.2.</b>	Negative stain EM micrographs of fluorescently-tagged RavA.....	193
<b>Figure 7.3.</b>	Negative stain EM classes of fluorescently-tagged RavA.....	194
<b>Figure 7.4.</b>	Negative stain EM reconstructions of fluorescently-tagged RavA...	195
<b>Figure 7.5.</b>	Cryo-EM micrograph of Dendra2(T69A)-RavA.....	197
<b>Figure 7.6.</b>	Cryo-EM class averages and map of Dendra2(T69A)-RavA.....	198
<b>Figure 7.7.</b>	Fitting of the RavA structure in the Dendra2(T69A)-RavA map.....	199
<b>Figure 7.8.</b>	N-terminal sequencing of the Dendra2(T69A)-RavA sample.....	200
<b>Figure 7.9.</b>	Fitting of the ArnA structure in the Dendra2(T69A)-RavA map.....	202
<b>Figure 7.10.</b>	Comparison between RavA and ArnA monomer structures.....	203
<b>Figure 9.1.</b>	Negative stain EM micrograph and classes of the MCIA complex...	252
<b>Figure 9.2.</b>	Schematic of ultracentrifugation gradients used in GraFix.....	253
<b>Figure 9.3.</b>	Negative stain EM of the MCIA complex after GraFix.....	254
<b>Figure 9.4.</b>	Negative stain EM micrograph of ACAD9-ECSIT <sub>CTD</sub> .....	255
<b>Figure 9.5.</b>	Cryo-EM micrograph of ACAD9-ECSIT <sub>CTD</sub> .....	257
<b>Figure 9.6.</b>	Cryo-EM 2D classes of ACAD9-ECSIT <sub>CTD</sub> from RELION.....	257
<b>Figure 9.7.</b>	Cryo-EM map of ACAD9-ECSIT <sub>CTD</sub> from RELION.....	258
<b>Figure 9.8.</b>	Cryo-EM 2D classes of ACAD9-ECSIT <sub>CTD</sub> from CryoSPARC.....	259
<b>Figure 9.9.</b>	Image processing steps for the ACAD9-ECSIT <sub>CTD</sub> core map.....	261
<b>Figure 9.10.</b>	Cryo-EM micrograph of ACAD9 in the presence of CHAPSO.....	263
<b>Figure 9.11.</b>	Cryo-EM 2D classes of ACAD9 in the presence of CHAPSO.....	263

*“What are the most central and fundamental problems of biology today? They are questions like: What is the sequence of bases in the DNA? What happens when you have a mutation? How is the base order in the DNA connected to the order of amino acids in the protein? What is the structure of the RNA; is it single-chain or double-chain, and how is it related in its order of bases to the DNA? What is the organization of the microsomes? How are proteins synthesized?”*

...

*It is very easy to answer many of these fundamental biological questions; you just look at the thing! You will see the order of bases in the chain; you will see the structure of the microsome. Unfortunately, the present microscope sees at a scale which is just a bit too crude. Make the microscope one hundred times more powerful, and many problems of biology would be made very much easier.”*

– Richard Feynman, 1959

# I. INTRODUCTION

# Preface

The core focus of my thesis has been on the *E. coli* protein RavA, which the Gutsche team has been working on for more than a decade. RavA is a member of the AAA+ ATPase superfamily, which is spread throughout the three kingdoms of life. RavA interacts with its binding partner LdcI to form a unique macromolecular cage which possesses a large central cavity. Based on both structural and functional studies, the team proposed this cage as a novel macromolecular chaperone. Along with a third protein ViaA, which is encoded by the *ravAviaA* operon, the LdcI-RavA cage was later proposed by our collaborator Walid Houry at the University of Toronto as being a chaperone involved in the maturation of iron-sulphur (Fe-S) cluster biogenesis systems and of respiratory Complex I.

My thesis began in October 2016 with the overall goal of investigating the role of the LdcI-RavA-ViaA triad in the maturation of Fe-S cluster biogenesis proteins, which is currently a poorly-understood process. The aim was to trap the cage with substrates bound in the central cavity for structural and functional analysis. During my first year in the lab however, we were unable to confirm a direct interaction between RavA and Fe-S cluster biogenesis proteins or respiratory Complex I.

Therefore, the main focus of the thesis was reoriented towards structural studies of LdcI-RavA cage itself, and especially on RavA. The 2014 cryo-EM structure of the LdcI-RavA cage published by the team revealed the overall architecture of the complex, but did not reveal how RavA would hydrolyse ATP to fulfil its proposed chaperone function.

Indeed, until 2016 the mechanism of AAA+ ATPase action was poorly understood. From 2016, the dramatic improvement of cryo-EM as a technique led to rapid advancement in the understanding of how AAA+ ATPases function. Rather than forming symmetric closed rings, it was revealed that most AAA+ proteins form asymmetric spirals. During the course of my thesis, I used cryo-EM to characterise RavA and saw that RavA too forms an asymmetric spiral, but also forms an unusual two-fold symmetric closed ring, which has implications for the ATPase mechanism.

In order to understand how the recent developments in cryo-EM contributed to the new understanding of the AAA+ ATPases, I will discuss the fundamentals of cryo-EM as a technique in Chapter 1 and summarise the recent technical breakthroughs which have enabled this sudden progression.

I will then present an overview of AAA+ ATPases and related proteins in Chapter 2, in order to place RavA in its evolutionary context. I discuss the key structural features of

AAA+ ATPases and related proteins, and present recent work investigating the mechanism of these diverse proteins. This sets the scene for the discussion of the RavA ATPase mechanism presented in Chapter 4.

Finally, I will introduce RavA and its binding partners in Chapter 3, and summarise the previous work of the lab on the structural and functional characterisation of this AAA+ ATPase, and of the LdcI-RavA-ViaA triad.

In addition, during my thesis I worked on a mitochondrial triad composed of the proteins NDUFAF1, ACAD9 and ECSIT which is involved in the maturation of respiratory Complex I and which is implicated in neurodegenerative disease in humans. Because of the expected analogies between this triad and the LdcI-RavA-ViaA triad at the time, we started a collaboration with Montserrat Soler-López at the European Synchrotron Radiation Facility in Grenoble to characterise the mitochondrial triad by electron microscopy.

The Results section is structured around three manuscripts. Chapter 4 contains a manuscript on the structural characterisation of RavA alone and in the context of the LdcI-RavA cage, which has recently been published in *Communications Biology*. Chapter 6 presents a manuscript focussed on LdcI which is currently under review. This manuscript investigates the cellular localisation of LdcI in *E. coli* under stress conditions, setting the stage for future studies on the visualisation of the LdcI-RavA-ViaA triad by super-resolution fluorescence microscopy. For this project, I purified and characterised fluorescently-tagged LdcI fusion proteins to investigate the structural and functional effects of fluorescent tags. Chapter 8 contains a manuscript, currently under review, on the mitochondrial Complex I assembly triad for which I contributed cryo-EM analysis of the ACAD9-ECSIT<sub>CTD</sub> subcomplex. Chapters 5, 7 and 9 present additional results which were not included in the three manuscripts.

Finally, Chapter 10 contains a brief discussion to provide some perspective on the results presented here. While exciting progress has been made since 2016 on the characterisation of the LdcI-RavA triad, there is still ongoing work in the lab that will bring new insights into the biological role of this unusual macromolecular machine.

# Chapter 1. Principles and developments in cryo-EM

## 1.1. Historical background

Microscopy has been used to investigate biological material for hundreds of years. The first light microscopes were developed in the 17th century – Antonie van Leeuwenhoek used these to investigate the fine structure of tissue and microscopic organisms (Lane, 2015; van Leeuwenhoek, 1677, 1682, 1684). Although these microscopes were relatively simple, they allowed the investigation of structural details of blood cells, capillaries and muscle fibres, and enabled the discovery of both protists and bacteria (Figure 1.1).

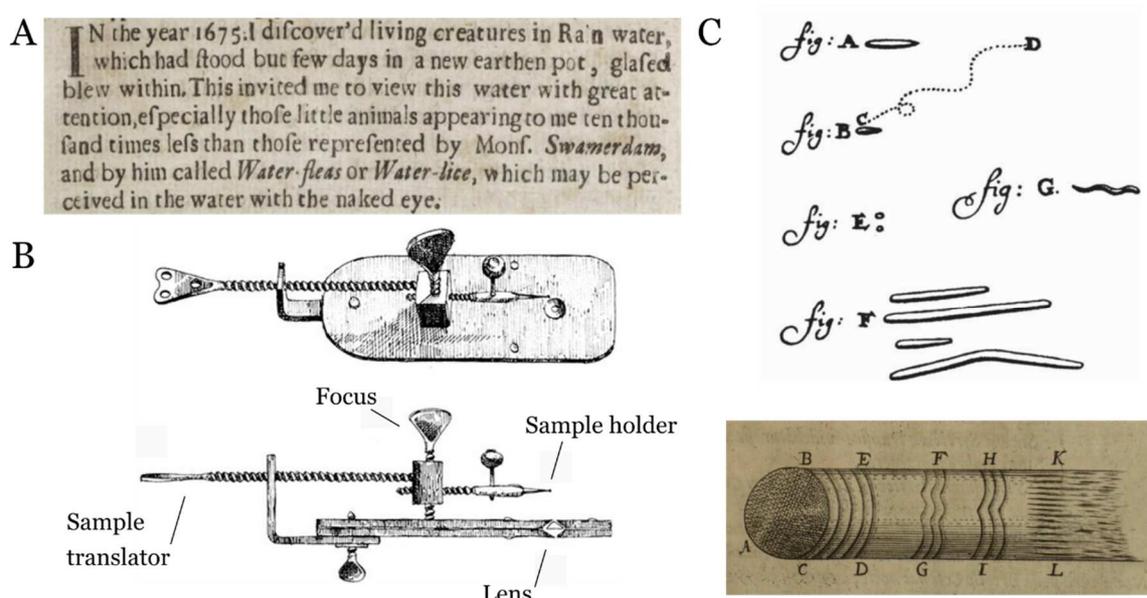


Figure 1.1. Early light microscopy by Antonie van Leeuwenhoek. **A)** Extract from van Leeuwenhoek, 1677, which described bacteria and protists for the first time. **B)** van Leeuwenhoek's light microscope design. Adapted from [www.biologydiscussion.com](http://www.biologydiscussion.com). **C)** van Leeuwenhoek's drawings of bacteria (above) and fish muscle fibres (below). Originally published in van Leeuwenhoek, 1684 (adapted from Lane, 2015) and van Leeuwenhoek, 1682 respectively.

Light microscopy has since been used extensively to characterise biological systems, yielding vast amounts of information on cellular ultrastructure. However, the maximum resolution of light microscopy is fundamentally limited by the wavelength of visible light (Rayleigh, 1896). The resolution limit for the light microscope is approximately 200 nm, or half the wavelength of visible light. This means that the direct visualisation of fundamental cellular processes, such as enzymatic reactions, DNA replication and protein synthesis, is impossible with the light microscope.

Using electrons overcomes this limit – the wavelength of an electron accelerated at 300 kV is approximately  $0.02 \text{ \AA}$ , or five orders of magnitude smaller than that of light (Zhang and Zhou, 2011). Development of the electron microscope was carried out by Ernst Ruska and Max Knoll in the 1930s, with the first images at a resolution better than light microscopes achieved in 1933. The overall design of the electron microscope is similar to that of the light microscope, as shown in Figure 1.2. Instead of glass lenses, electromagnetic lenses are used to focus a beam of electrons emitted from a cathode ray tube onto a sample, with an image formed on a fluorescent screen (Ruska, 1934, 1987). This technique is termed transmission electron microscopy (TEM), as images are formed by electrons transmitted through an extremely thin sample. Despite the  $\sim 0.02 \text{ \AA}$  wavelength of a 300 kV electron beam, the resolution of the electron microscope is practically limited to  $\sim 0.5 \text{ \AA}$  because of diffraction of the electron beam on the edges of the electromagnetic lenses and objective aperture (Smith, 2008; Williams and Carter, 2009).

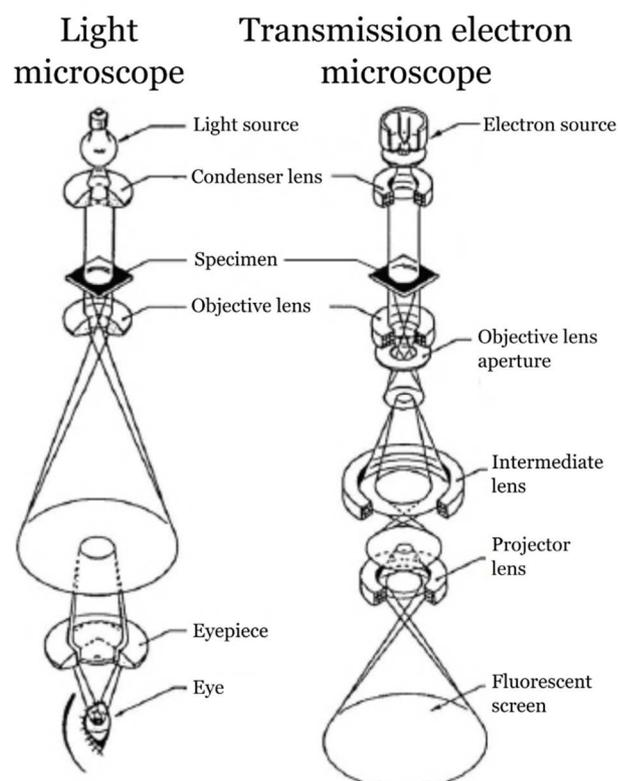


Figure 1.2. Comparison between the design of the light microscope and electron microscope. Adapted from [www.americanlaboratory.com](http://www.americanlaboratory.com).

Early on it was recognised that using such a microscope to study biological materials would be difficult as the high-energy electron beam would quickly destroy cells and organic matter (Marton, 1934). In addition, the column through which the electron beam passes has to be kept under vacuum as molecules in the air scatter electrons. The vacuum causes dehydration of biological samples when they enter the column, destroying

their structural integrity (Cheng, 2018; Lepault and Pitt, 1984). Because of these factors, the resolution of images from electron microscopy of biological objects is not limited by the microscope itself, but rather by the sample. One way of mitigating the structural damage caused by the electron beam and high vacuum is using heavy metal stains. In the 1940s the use of stains such as osmic acid and phosphotungstic acid allowed the visualisation of fine structural details in muscle fibres, as shown in Figure 1.3.

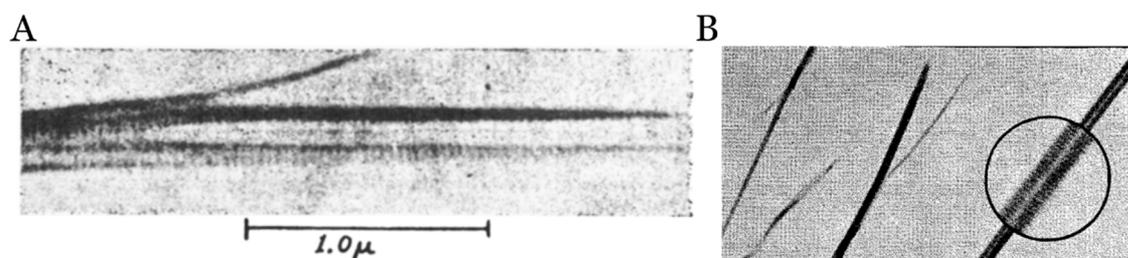


Figure 1.3. Visualisation of striations in clam muscle fibres by electron microscopy, stained with **A)** osmic acid and **B)** phosphotungstic acid. Adapted from Jakus et al., 1944 and Hall et al., 1945.

Heavy metal stains have been particularly useful in the study of macromolecules. Negative stain electron microscopy makes use of heavy metal stains to coat protein complexes – areas where stain accumulates appear dark, such as cavities in or around a protein, and the protein itself appears lighter (Figure 1.4) (Brenner and Horne, 1959). Such stains increase the contrast of images (see section 1.2.1) and are resistant to high electron doses. This technique was used to characterise viruses in the 1960s and 1970s (Figure 1.4), providing insight into their structure for the first time (Brenner and Horne, 1959; Crowther et al., 1970; Horne, 1973; Klug and Finch, 1965). However, the use of stains can cause artefacts such as sample flattening, and limits the resolution of images to 15-20 Å (Cheng et al., 2015). In addition, because they form a coat around the protein they only allow visualisation of surface details and not the internal structure.

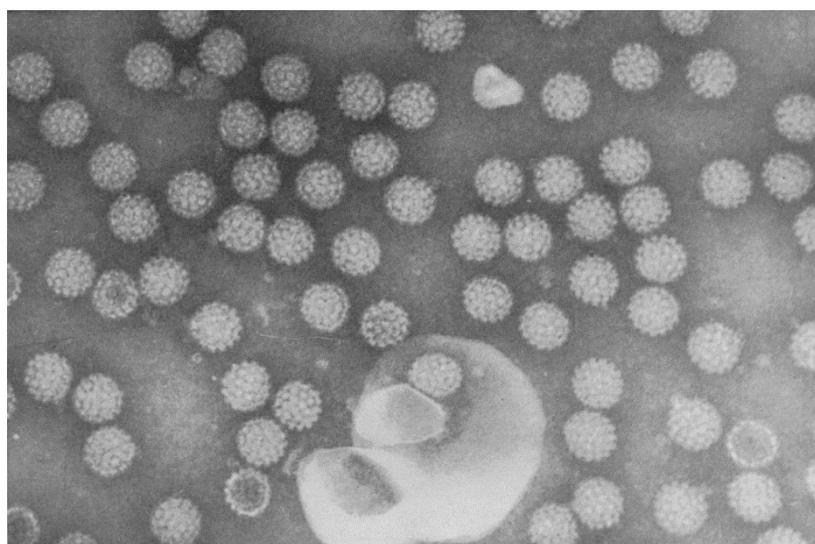


Figure 1.4. Negative stain micrograph of human papillomavirus showing features of the viral capsid. Adapted from Klug and Finch, 1965.

The key technical development that enabled direct imaging of native, unstained protein and the preservation of high-resolution structural features came in the 1980s (Adrian et al., 1984; Dubochet and McDowell, 1981; Dubochet et al., 1982). Rapidly cooling the sample to cryogenic temperatures preserves structural features while avoiding the artefacts associated with negative stain electron microscopy and is termed cryo electron microscopy (cryo-EM), as discussed in section 1.2.3. An example of a cryo-EM micrograph of viruses is shown in Figure 1.5.

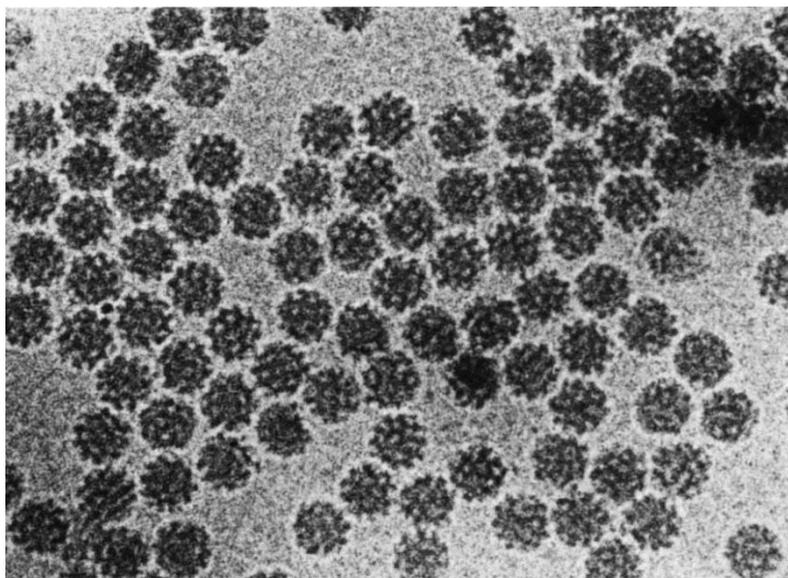


Figure 1.5. Cryo-EM image of Semliki Forest virus, adapted from Adrian et al., 1984. The contrast is inverted compared to negatively stained images, as shown in Figure 1.4.

This development enabled the rapid expansion of cryo-EM as a tool to investigate the structure of macromolecules. There are two main techniques used in cryo-EM of biological samples – single-particle electron microscopy, and electron tomography. Single-particle electron microscopy involves the collection of many two-dimensional (2D) images of isolated macromolecules in various orientations, and computationally calculating a three-dimensional (3D) volume from these 2D images. In contrast, electron tomography involves the imaging of a single object from many angles and the reconstruction of a volume from the series of tilted images. Subvolumes corresponding to individual particles can then be extracted from tomograms and averaged, which is termed subtomogram averaging.

The fields of single-particle EM and electron tomography (in particular subtomogram averaging) have progressed recently to the point where both techniques are able to determine the 3D structures of macromolecules to high resolution, revealing the atomic details of fundamental cellular pathways and processes (Danev et al., 2019). In the context of this thesis only single-particle electron microscopy will be considered – the principles of this technique and recent developments in the field will be discussed in the following sections.

## 1.2. Principles of cryo-EM

### 1.2.1. Image formation and contrast

There are two types of contrast which contribute to image formation in the microscope – amplitude contrast and phase contrast (Dubochet et al., 1988; Erickson and Klug, 1970). Amplitude contrast occurs due to the differential scattering of electrons by regions of the sample with different mass (Williams and Carter, 2009). This can occur either because of increased thickness of the sample at a given point, or because of an increased scattering cross-section of an atom at a given point. Because cryo-EM is usually carried out on thin specimens (only a few hundred nanometres thick) that are essentially transparent to the electron beam (Danev, 2019), amplitude contrast primarily comes from differences in scattering cross-sections. The difference in the scattering cross-section of low-mass ‘biological’ atoms (C, H, N and O) which primarily make up proteins and the surrounding vitreous ice (H<sub>2</sub>O) is low (Orlova and Saibil, 2011). Because of this, amplitude contrast contributes only a small amount (around 7%) towards the total contrast in cryo-EM images of biological objects (Henderson, 1995; Lepault and Pitt, 1984).

Amplitude contrast can be increased by using a heavy metal stain, such as uranyl acetate (Hall et al., 1945; Williams and Carter, 2009). Atoms such as uranium and osmium have much higher scattering cross-sections and so increase the contrast significantly. This principle is used in negative stain electron microscopy, as discussed in section 1.1. However, the artefacts associated with negative stain electron microscopy and resolution limit because of the use of stain mean that the technique cannot be used to investigate protein structure at high resolution.

In cryo-EM amplitude contrast is much lower, and phase contrast is instead the primary contributor to the total contrast in images (Danev, 2019; Dubochet et al., 1988; Erickson and Klug, 1970). Phase contrast occurs when the electrons that are elastically scattered by the sample (see section 1.2.3) and the unscattered electron wave arrive at the detector out of phase, resulting in interference (Erickson and Klug, 1970; Unwin, 1971). When images are collected exactly at focus, the scattered and unscattered electrons arrive in phase. Therefore images are usually collected slightly under focus, by up to around 3  $\mu\text{m}$ , which is known as defocussing (Erickson and Klug, 1970). This introduces an additional phase shift in the scattered wave compared to the unscattered wave, resulting in modulation of the amplitude of the unscattered wave and therefore in increased contrast. The effect of defocussing on image contrast is shown in Figure 1.6.

However, defocussing introduces aberrations into the collected images. Because of imperfections in electromagnetic lenses and the effect of defocussing, the image of the biological object is not a perfect representation of the object, but is modulated by a function termed the contrast transfer function (CTF), as discussed in the following section.

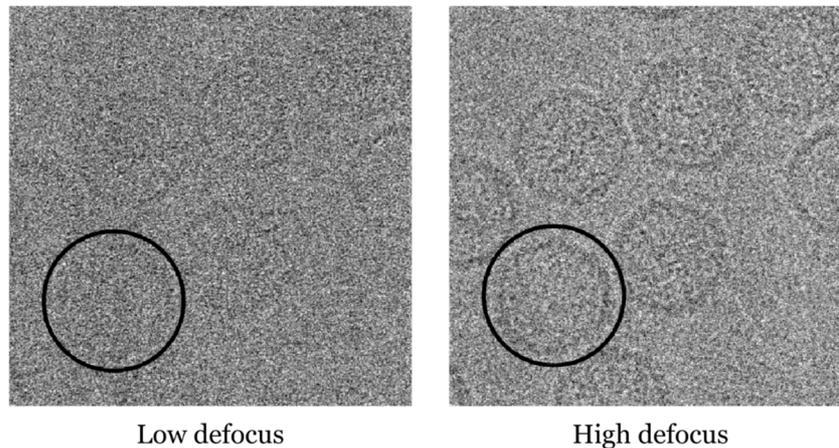


Figure 1.6. The effect of defocus on image contrast. Both low defocus (left) and high defocus (right) images are of the same P22 bacteriophage particles, with one particle circled. At lower defocus particles are difficult to distinguish, whereas the increased contrast at higher defocus makes it easier to visualise particles. Adapted from Thuman-Commike and Chiu, 2000.

### 1.2.2. The Contrast Transfer Function (CTF)

In a perfect microscope, an image should be a faithful representation of the imaged object. However, the point spread function of the electron microscope, which arises because of imperfections in the electromagnetic lenses and diffraction of the electron beam at the edges of the lenses, modulates the image of an object. The Fourier transform of the point spread function is the contrast transfer function (CTF), an oscillating function which describes the amount of information present in an image as a function of spatial frequency (Orlova and Saibil, 2011). Taking the Fourier transform of a micrograph allows visualisation of the CTF as shown in Figure 1.7A.

During imaging of an object, electrons are scattered through multiple scattering angles from a single point. Electrons scattered at lower angles contain information corresponding to lower-resolution features, whereas highly scattered electrons contain higher-resolution information. When images are collected with a slight defocus, electrons scattered at different angles arrive at the detector with differing amounts of phase shift relative to the unscattered beam. Interference between unscattered and scattered electrons will have differing effects on the amplitude of the resulting wave depending on the scattering angle. For some scattering angles (i.e. at certain spatial frequencies), interference will produce maximum contrast, whereas at others there will be a complete

loss of contrast (Erickson and Klug, 1970). The exact spatial frequencies where this occurs depends on the amount of defocus used, as shown in Figure 1.7B. The phases of the image are also inverted every time the CTF crosses zero, and when this happens contrast is also inverted. Finally, there is also a dampening of information at higher frequencies – this dampening is known as the CTF envelope (Orlova and Saibil, 2011).

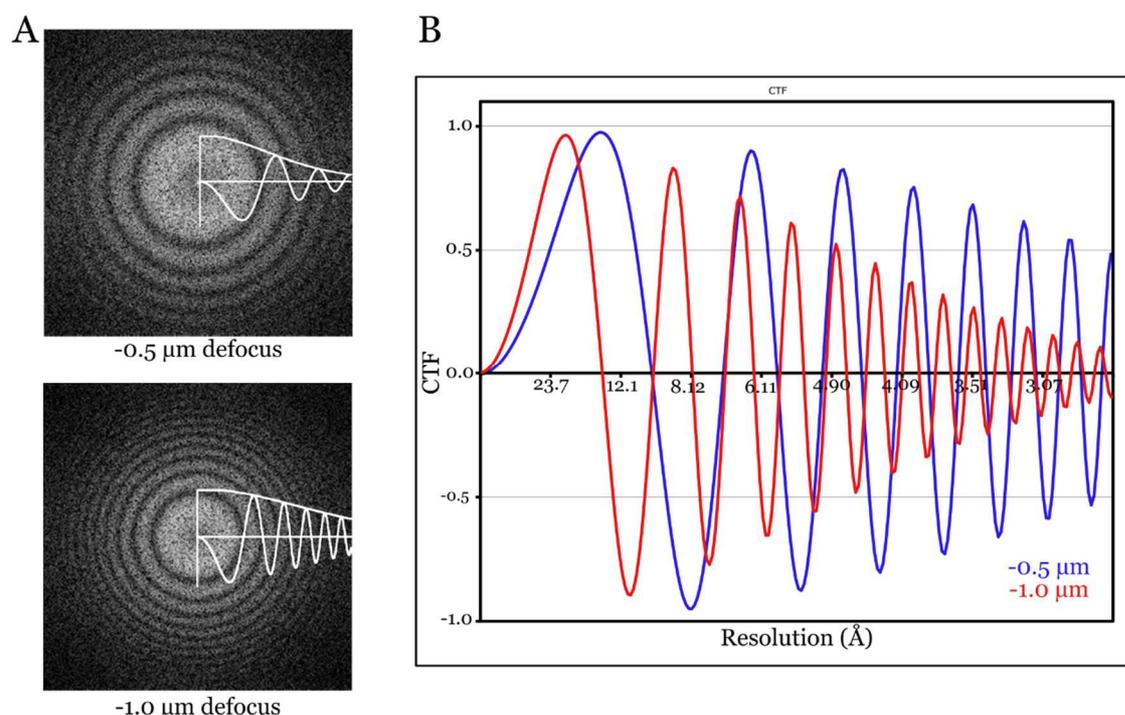


Figure 1.7. **A)** The effect of defocussing on the CTF shown in the Fourier transform of cryo-EM micrographs at defocus values of  $-0.5\ \mu\text{m}$  (above) and  $-1.0\ \mu\text{m}$  (below). Spatial frequencies at which the CTF is zero appear as dark rings termed Thon rings (Thon, 1966). A 1D fit of the CTF is shown over each Fourier transform. Adapted from Orlova and Saibil, 2011. **B)** Superimposed CTFs for defocus values of  $-0.5\ \mu\text{m}$  (blue) and  $-1.0\ \mu\text{m}$  (red) at 300 kV. At higher defocus, there is higher signal at lower resolution resulting in increased contrast, but the amplitude fall-off is greater and higher-resolution features are dampened. CTFs were simulated using the CTF simulator tool at [www.c-cina.unibas.ch/tools/soft/ctf-simulation/](http://www.c-cina.unibas.ch/tools/soft/ctf-simulation/).

Because of this, images must be corrected computationally to compensate for the CTF. CTF correction is carried out by flipping inverted phases and by boosting higher spatial frequencies to compensate for amplitude fall-off at higher resolution (Mindell and Grigorieff, 2003; Orlova and Saibil, 2011). Although this CTF correction can correct for inverted phases and amplitude fall-off, information cannot be recovered from spatial frequencies at which the CTF is zero (Orlova and Saibil, 2011). Therefore, images are collected at multiple defoci during data acquisition. At different defoci, zeros in the CTF will be at different spatial frequencies (as shown in Figure 1.7B). The ensemble of images from the microscope will therefore contain information about the object for all spatial frequencies, and images at multiple defoci are then averaged together to allow faithful reconstruction of the 3D structure.

The idea of averaging is key to using cryo-EM for the structural determination of macromolecules – not only because of the CTF, but also because of the low signal inherent to images collected with low electron doses, as discussed in section 1.2.4.

### 1.2.3. Radiation damage and vitrification

When an electron arrives at a sample, most of the time it passes through unscattered, without interacting with the sample (Erickson and Klug, 1970). Some electrons do interact with the sample, and undergo either elastic scattering, where no energy is lost, or inelastic scattering, in which some energy is transferred to the sample (Dubochet et al., 1988; Orlova and Saibil, 2011). Inelastic scattering damages the sample, resulting in the breakage of chemical bonds and degradation of the sample structure. Examples of such structural damage include the cleavage of disulphide bonds and decarboxylation of acidic amino acids (Hattne et al., 2018). At high doses, radiation damage can be visually seen as boiling or ‘bubbling’ of the sample under the beam due to the release of hydrogen gas (Dubochet et al., 1982; Meents et al., 2010), as shown in Figure 1.8. This damage affects high-resolution features (such as amino acid side chains) first, and starts to occur even after very short exposures to the beam (Glaeser, 1971; Hattne et al., 2018; Knappek and Dubochet, 1980). Because of this, radiation damage is one of the key limiting factors for determining the structure of proteins using electron microscopy (Henderson, 1995).

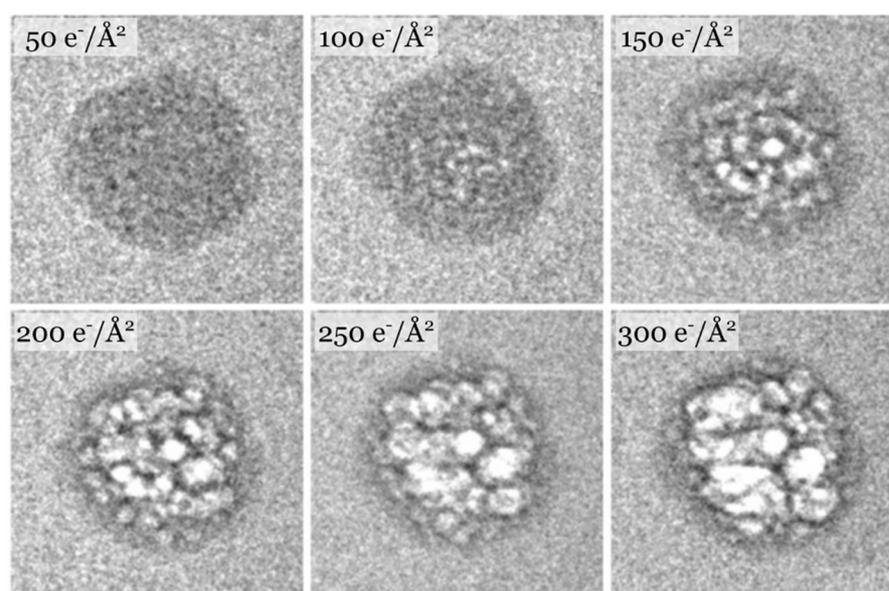


Figure 1.8. Images of HIV nucleocapsid protein at increasing electron doses. At higher doses, radiation damage is visible as bubbling. Adapted from Fontana et al., 2015.

The idea of the “intense cooling” of biological samples to overcome radiation damage was already recognised in the 1930s (Marton, 1934). Reducing the temperature slows down the processes of sample degradation by inelastic scattering, allowing longer exposures before information is lost (Knappek and Dubochet, 1980; Taylor and Glaeser, 1976). However, at cryogenic temperatures, the water surrounding biological samples forms crystalline ice, which scatters the electron beam and destroys sample ultrastructure (Dubochet et al., 1988). Freezing the sample rapidly instead results in vitreous or amorphous ice, where water molecules are oriented randomly and the ultrastructure of biological samples is preserved (Adrian et al., 1984; Dubochet and McDowell, 1981; Taylor and Glaeser, 1976). This allows the imaging of native, hydrated biological specimens while retaining high-resolution information, and is termed cryo-electron microscopy (cryo-EM).

Samples for electron microscopy are prepared by the application of a protein-containing solution to a thin grid. Practically, vitrification is achieved by freezing cryo-EM grids in a cryogen such as ethane or propane, which is kept cool by surrounding liquid nitrogen (Dubochet and McDowell, 1981; Dubochet et al., 1985). For vitrification to occur,

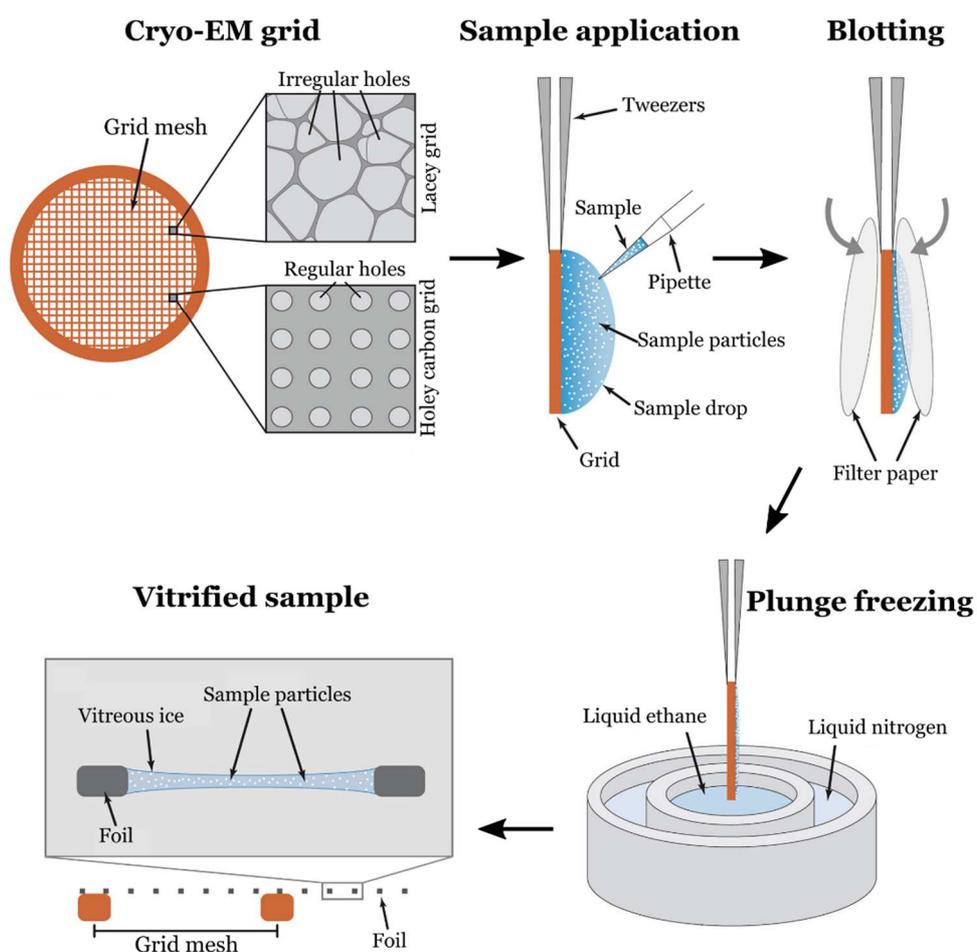


Figure 1.9. Common experimental setup for freezing cryo-EM grids. Adapted from Sgro and Costa, 2018.

water is cooled extremely rapidly ( $\sim 10^{-4}$  s) to prevent crystalline ice from forming (Dubochet et al., 1988). Nitrogen does not conduct heat away from the sample rapidly enough for vitrification, whereas propane and ethane have a thermal conductivity around 300-500 times higher (Dobro et al., 2010).

The most commonly-used experimental setup for preparing samples for cryo-EM is shown in Figure 1.9. 2-3  $\mu\text{L}$  of sample is pipetted onto a cryo-EM grid, which is 3 mm in diameter. Grids are usually made of copper or gold, and have a fine mesh of thin grid bars. Across the mesh is a thin film (foil) with holes, usually made from carbon or gold, across which a vitrified layer of ice will be suspended. These holes can either be regularly spaced and sized, or irregular. After sample application to the grid, most of the liquid (more than 99%) is blotted away with filter paper – this ensures that a layer of ice thin enough for TEM is formed. The grid is quickly plunged into the cryogen and subsequently transferred into liquid nitrogen and stored for imaging.

#### 1.2.4. Single-particle electron microscopy, and the concept of averaging

Even at cryogenic temperatures, radiation damage accumulates quickly. Therefore in order to preserve the high-resolution information in images of proteins, imaging has to be carried out with low electron doses (Glaeser, 2013). This results in images with a very low signal-to-noise ratio, such that there is very little information present in a single particle image (Dubochet et al., 1985). Therefore, many thousands of noisy images of proteins are collected and averaged. This is known as single-particle EM – the ‘single-particle’ refers to images of single isolated proteins in random orientations, as opposed to a crystal of proteins. Although each particle image contains very little signal and substantial noise, by aligning and averaging many images of proteins lying in the same orientation the signal reinforces and the noise cancels out (Frank, 1975). This principle is demonstrated in Figure 1.10 in 2D, but also applies in 3D. Theoretically, it was estimated that only around 10,000 particles would be needed to calculate protein structures at atomic resolution (Henderson, 1995). In practise this number is usually much higher, with tens or hundreds of thousands of particle images often being used to calculate protein structures using cryo-EM (Tan et al., 2016).

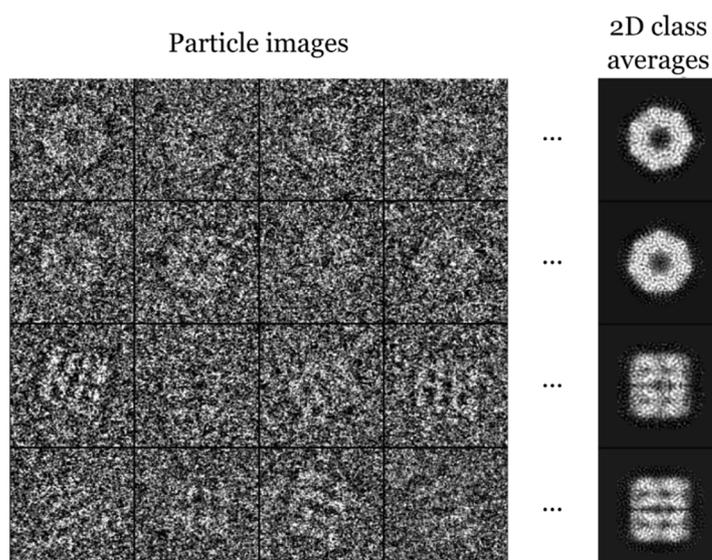


Figure 1.10. Averaging many noisy particle images of GroEL in the same orientation results in increased signal. Adapted from Ludtke et al., 2004.

### 1.2.5. From 2D to 3D – the central slice theorem

TEM of thin specimens produces projection images – the depth of focus in the microscope is much larger than the depth of the specimen (except for samples such as very large viruses), and so the entire sample is in focus (Crowther et al., 1970; Jensen and Kornberg, 2000; De Rosier and Klug, 1968). Therefore, projection images contain all of the 3D information of an object from a particular viewing direction, corresponding to the orientation that the object is lying in with respect to the electron beam. In order to reconstruct the full 3D volume, projection images from all viewing directions are required (Crowther, 1971; Penczek, 2010; De Rosier and Klug, 1968).

Computational reconstruction of a 3D model from 2D projection images makes use of the central slice theorem (Bracewell, 1956; Crowther, 1971) as shown in Figure 1.11. The central slice theorem states that the Fourier transform of a 2D projection of an object corresponds to a central slice through the Fourier transform of that 3D object. By determining the relative orientations of individual projection images, the 2D Fourier transforms of the images can be placed to build up the full 3D Fourier transform of the object. Carrying out an inverse Fourier transform of the 3D Fourier transform can then be done to calculate the original 3D structure of the object.

In practise, 2D projection images from cryo-EM have a low signal-to-noise ratio and so determining the exact orientation of each 2D Fourier slice can be difficult. The process of determining the orientation of each particle image is therefore carried out iteratively, as discussed in the following section.

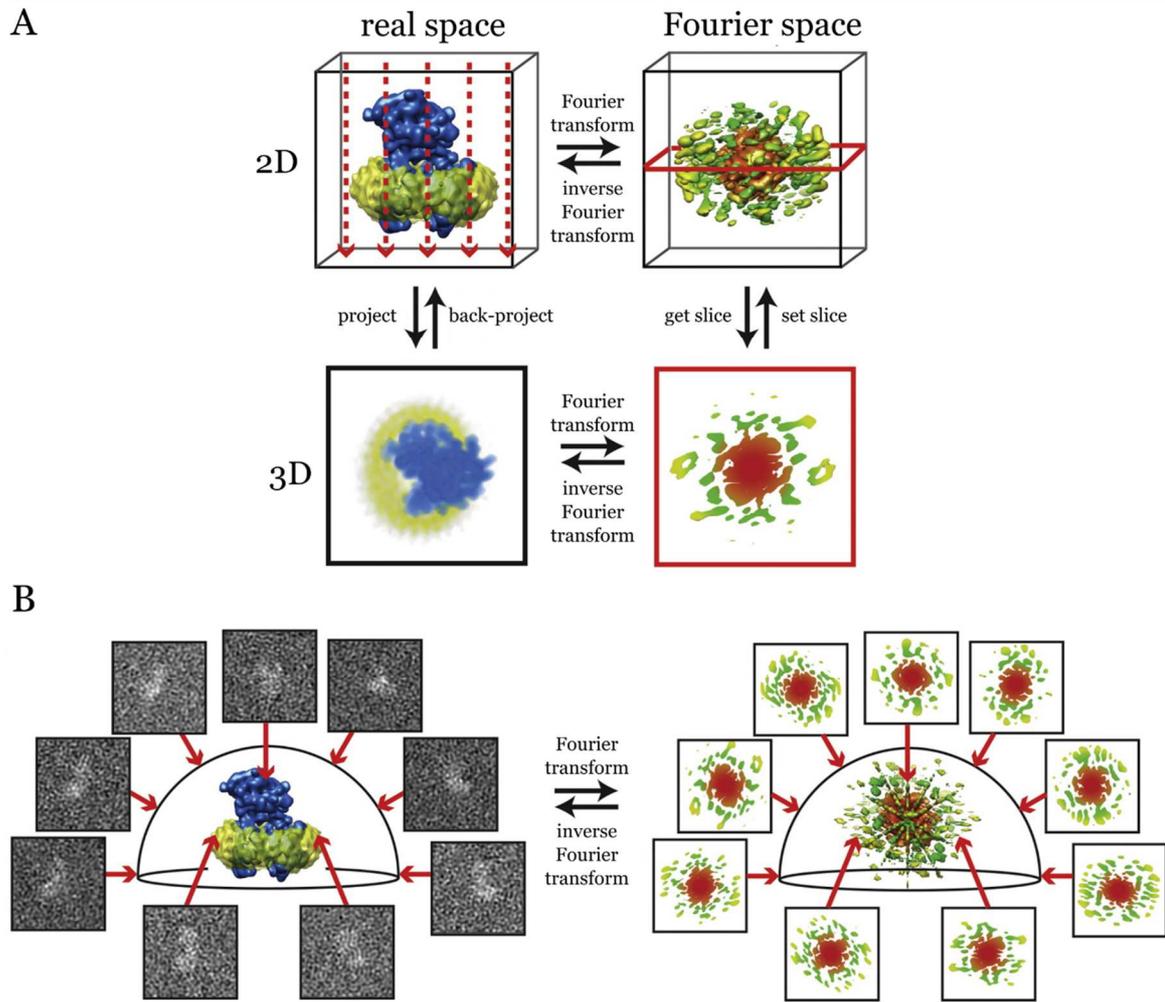


Figure 1.11. The central slice theorem for electron microscopy, adapted from (Nogales and Scheres, 2015). **A)** Projecting a 3D object (top left) in the direction shown by the red arrows yields a 2D projection image. The Fourier transform of the projection image corresponds to a central slice through the Fourier transform of the original 3D object. **B)** Diagram showing how the central slice theorem can be used to build up a 3D reconstruction of an object from 2D projection images, by filling up the 3D Fourier space with 2D Fourier slices. The colour in the Fourier space corresponds to spatial frequency – low frequencies in the centre are coloured red, higher frequencies further out are coloured green and yellow.

### 1.3. The image processing pipeline for single-particle electron microscopy

The basic pipeline for single-particle electron microscopy consists of several steps which are carried out to produce a 3D map from 2D micrographs. There were several key software packages developed in the 1990s and 2000s for processing cryo-EM data, including SPIDER (Shaikh et al., 2008), SPARX (Hohn et al., 2007), EMAN and EMAN2 (Ludtke et al., 1999; Tang et al., 2007), IMAGIC (Van Heel et al., 1996), FREALIGN (Grigorieff, 2007) and XMIPP (Scheres et al., 2005). Although the algorithms used by these software packages differed, the same basic steps were conserved between packages. In modern electron microscopy this core pipeline is conserved, although there are many additional steps that can be carried out as discussed in section 1.5.

On electron microscopes equipped with modern detectors, micrographs are collected as movies with multiple frames per image – this allows for the correction of drift which occurs during imaging (see section 1.4.1 for details). CTF estimation is carried out on motion-corrected cryo-EM images using software packages such as CTFFIND3 (Mindell and Grigorieff, 2003). These packages estimate the CTF on Fourier transforms of 2D micrographs using the signal from Thon rings (see Figure 1.7 and section 1.2.2), which are radially averaged to overcome the low signal-to-noise ratio.

Particles are then picked from images, which can be done manually (a slow and labour-intensive process), semi-automatically or fully automatically. Automatic picking software is often reference-based – 2D class averages or projections of a 3D model from a previous dataset can be used as templates, and particles are selected based on peaks in cross-correlation with templates. References are usually low-pass filtered for picking to avoid model bias (Cheng et al., 2015). Selected particles are then extracted from micrographs to form stacks of boxed particle images. Extracted particles are subjected to 2D classification – similar images are aligned and averaged as shown in Figure 1.10, which allows the removal of bad particles and assessment of the sample quality and homogeneity.

In order to calculate a 3D reconstruction of the protein, an initial reference volume is needed for aligning particle images to. This initial reference can come from an existing structure of a homologous protein, which is filtered to low resolution to avoid model bias (Cheng et al., 2015). Traditionally it was difficult to calculate a good initial model *ab initio*, and various strategies such as random conical tilt (Radermacher et al., 1987) or the common lines approach (Van Heel, 1987) were employed. Recently, *ab initio* model generation algorithms have significantly improved and are no longer as much of a limiting step in the image processing pipeline (see section 1.5.3). 3D classification against the initial

model can then be employed to separate out heterogeneity in 3D – this can yield multiple reconstructions corresponding to different conformational or compositional states.

Selected particles after 3D classification are then subjected to 3D refinement against a reference volume – either the previously-determined initial model, or a volume resulting from 3D classification. During 3D refinement, the data is split into two half-datasets which are refined independently against two separate 3D reference volumes. During refinement, the independent volumes are aligned at low resolution to prevent the two half-maps diverging too far. This approach is termed “gold-standard refinement” and avoids overfitting and inflated resolution estimates (Scheres and Chen, 2012).

After each iteration during refinement, particle images are back-projected and an updated volume is calculated for each half-dataset. The two half-maps are aligned at low resolution, and are compared by Fourier Shell Correlation (FSC) – the cross-correlation between the half-maps is calculated at each spatial frequency and plotted. At low resolution, the half-maps will be very similar and the FSC is high. At increasing resolutions the FSC drops as high-resolution features become less consistent between the half-maps due to the fall-off in signal at high spatial frequencies. The nominal resolution of the reconstruction is set to the spatial frequency at which the FSC between half-maps drops below 0.143 (Rosenthal and Henderson, 2003). Volumes are filtered to this resolution, and another round of refinement then takes place. As the reconstruction improves at each iteration the accuracy of assigning angles to particle images also increases. 3D refinement is iteratively performed until the angular accuracy and estimated resolution do not improve, and half-maps are then combined to give the final 3D reconstruction.

The final reconstruction is then sharpened to boost high-resolution frequencies – a negative B-factor is applied to compensate for amplitude fall-off at high resolution (Rosenthal and Henderson, 2003). However, over-sharpening (the application of an inappropriately large B-factor) can introduce noise into the map and can result in discontinuity in the density. The effects of sharpening on the final map are shown in Figure 1.12.

Structural interpretation of the final map is carried out differently according to the resolution of the map. For high-resolution maps (better than  $\sim 4$  Å), atomic models can be built *de novo* into the density and subsequently refined using software such as COOT and PHENIX (Adams et al., 2010; Emsley et al., 2010). For low-resolution maps, fitting of pre-existing crystal structures of the protein of interest or closely-related homologs into the cryo-EM envelope is carried out using software packages like Chimera (Pettersen et al., 2004). For medium-resolution maps, software packages like IMODFIT (López-Blanco and

Chacón, 2013) and Flex-EM (Topf et al., 2008) can be used to the flexibly fit crystal structures into medium-resolution maps, where secondary structures are visible.

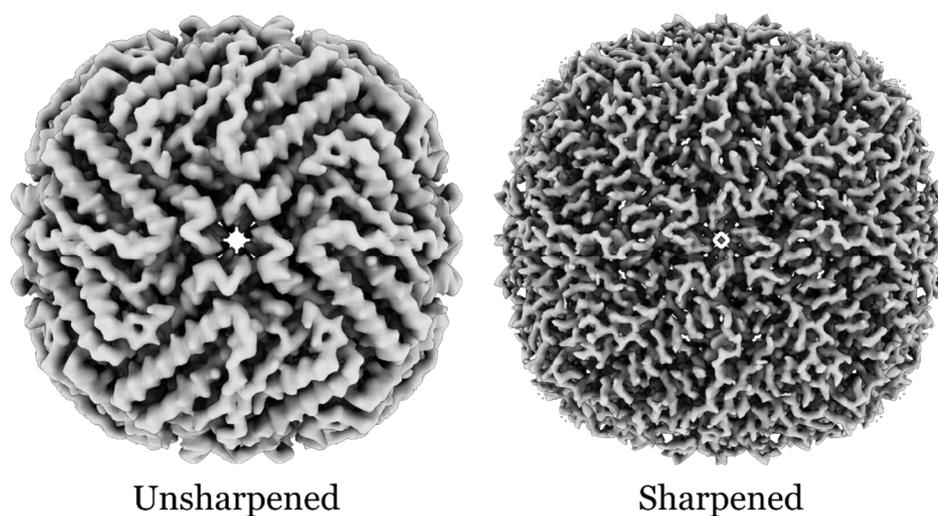


Figure 1.12. 3D reconstruction of apoferritin at a resolution of 2.57 Å before (left) and after (right) sharpening with a B-factor of  $-167 \text{ \AA}^2$ . Sharpening during post-processing boosts high-resolution features, improving the quality and appearance of the map. Maps courtesy of Ambroise Desfosses.

## 1.4. Modern electron microscopy

Although the overall microscope design has not changed significantly since its inception, the development of modern electron microscopes has significantly improved electron microscopy as a method for determining the structure of biological macromolecules (Henderson, 2015; Schröder, 2015).

One of the key developments to electron microscopes has been the improvement of the electron source (Henderson, 2015). The original electron source in microscopes was a tungsten filament heated to  $\sim 3000^{\circ}\text{C}$ . This was first replaced by a crystal of  $\text{LaB}_6$ , which improved the coherence of the electron beam as it possessed a smaller tip. More recently the field emission gun (FEG) was developed, improving the brightness and coherence of the beam even further. Figure 1.13 shows a comparison between the three different electron sources, highlighting the differences in tip size. The tip of a FEG is only 3-20 nm in size, which is significantly smaller than the 5  $\mu\text{m}$  tip of a  $\text{LaB}_6$  crystal and 30  $\mu\text{m}$  tungsten filament tip.

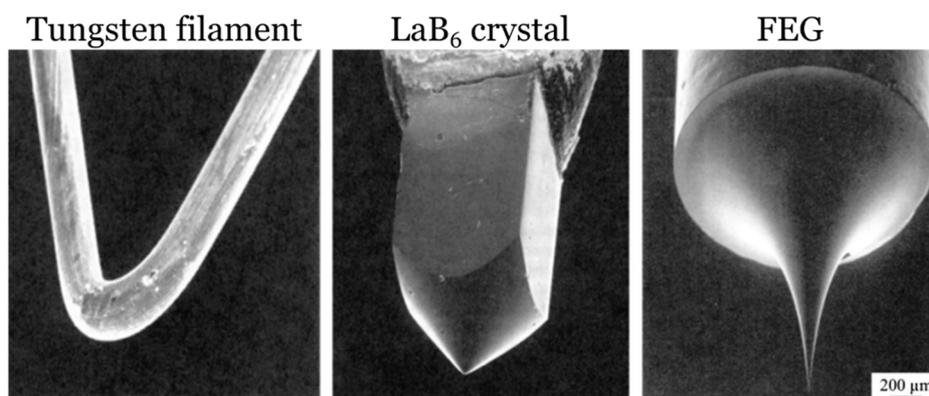


Figure 1.13. Scanning electron microscopy images of the tips of three electron sources – tungsten filament,  $\text{LaB}_6$  crystal and field emission gun. Adapted from Henning and Adhikari, 2017.

Another key development has been in the stability of microscope stages (Henderson, 2015). On high-end microscopes, such as Titan Krios, Talos Arctica and Glacios microscopes (Thermo Fisher) and CRYO ARM 300 and CRYO ARM 200 microscopes (JEOL), the traditional ‘side-entry’ holder has been replaced by small cassettes that hold clipped grids isolated in the centre of the microscope column. With side-entry holders, grids in the centre of the column were connected by a metal rod to a liquid nitrogen dewar on the outside of the microscope for cooling. This caused vibrations at the level of the stage, increasing sample drift and reducing the quality of images collected. This vibration has now been eliminated with the advent of cassettes – grids are isolated in the column from external vibration, and as a consequence the stage is much more stable.

Previously, electron microscopy was a relatively niche technique used by a small community to study mostly large, symmetric protein complexes (Stuart et al., 2016). Today, it is firmly established as mainstream technique which is being adopted by more users each year to study a wide variety of proteins at high resolution (Danev et al., 2019; Eisenstein, 2015). Three key developments which have facilitated this will be discussed in the following sections. The development of direct electron detectors, which significantly out-perform CCD cameras and film as a method for recording images, has been crucial to the rapid advancement of cryo-EM (McMullan et al., 2016). Phase plates have also been developed to improve the contrast of images, which have especially given success in determining the structure of small proteins (Fan et al., 2019; Khoshouei et al., 2017). In addition, modern electron microscopes are faster and easier to operate, with many steps in data collection now automated. Because of this, it is now not strictly necessary to be an expert user to collect high-resolution datasets (Cheng, 2018).

#### 1.4.1. Direct electron detectors

Perhaps the largest contributing factor to the dramatic improvement in cryo-EM in the past decade is the development of direct electron detectors (Danev et al., 2019; McMullan et al., 2016). Historically, images from electron microscopes were recorded on photographic film. Although these had very large fields of view, the throughput of data collection was low – only a few dozen images could be collected in a single session, and each image had to subsequently be scanned and digitised (Baldwin et al., 2018). The introduction of charge coupled device (CCD) cameras improved the throughput of data collection (see section 1.4.2) and allowed instant visualisation of images, however this came at the expense of image quality.

An ideal detector is able to record signal for every spatial frequency equally well without introducing noise (McMullan et al., 2009). The detector quantum efficiency (DQE) is a measure of how the signal-to-noise ratio is degraded (i.e. how much noise is added by detector) at every spatial frequency during detection. Ideally, the DQE would be 1 at every spatial frequency until Nyquist, which is the maximum spatial frequency able to be recorded at a given pixel size (McMullan et al., 2009). In reality, there is a fall-off in the DQE at higher frequencies because of errors in detection – this limits the resolution of images, particularly at the low electron doses used in cryo-EM.

At low accelerating voltages CCD cameras outperform photographic film (McMullan et al., 2009). However at 300 kV, the voltage typically used for high-resolution electron microscopy, the DQE of film is significantly better than CCD cameras as shown in

Figure 1.14. At low spatial frequencies the DQE is similar for film and CCD cameras, but at higher frequencies the DQE is much higher for film. In addition, the field of view is significantly larger for film compared to CCD cameras, and so for the same number of particle images at a given magnification significantly fewer micrographs need to be recorded. For these reasons, film was used to record images for high-resolution electron microscopy until recently (Ge et al., 2015; Guo and Jiang, 2014; Gutsche et al., 2015).

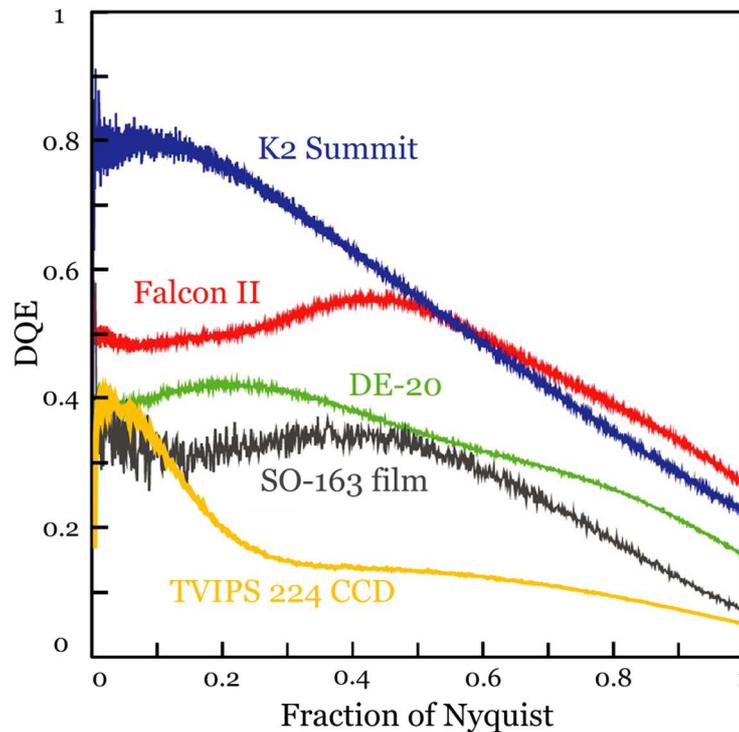


Figure 1.14. Comparison of DQE between three common direct electron detectors (K2 Summit, Falcon II and DE-20), film (SO-163) and a CCD camera (TVIPS 224) at 300 kV as a function of Nyquist. Adapted from McMullan et al., 2009 and McMullan et al., 2016.

In the late 2000s, new generation electron detectors were developed that revolutionised the field of cryo-EM. CCD cameras require the conversion of electrons into photons for detection, whereas the new generation detectors can detect individual electrons directly and are therefore termed direct electron detectors (McMullan et al., 2009). These detectors use semiconductor technology and are much thinner, resulting in reduced backscattering in the silicon support matrix which introduces noise into images (McMullan et al., 2016). Because of this, direct detectors have a much higher DQE than both film and CCD, as shown in Figure 1.14, and have now become the dominant type of detector used for electron microscopy (Bai et al., 2015a; McMullan et al., 2016).

In addition, direct detectors have a much higher frame rate than film or CCD cameras and are able to capture multiple frames per second. Rather than recording single images, they are able to record ‘movies’ – crucially, this allows the correction for movement of the sample that occurs during image acquisition, both from stage drift and beam-

induced sample deformation (Brilot et al., 2012; Ripstein and Rubinstein, 2016). This restores high-resolution information which is lost during sample movement due to blurring of the image, allowing dramatic improvement in the resolution of information present in cryo-EM images. Correcting for such motion is carried out during image processing, and dedicated software packages have been developed for this purpose (Zheng et al., 2017). The effects of motion correction are shown in Figure 1.15.

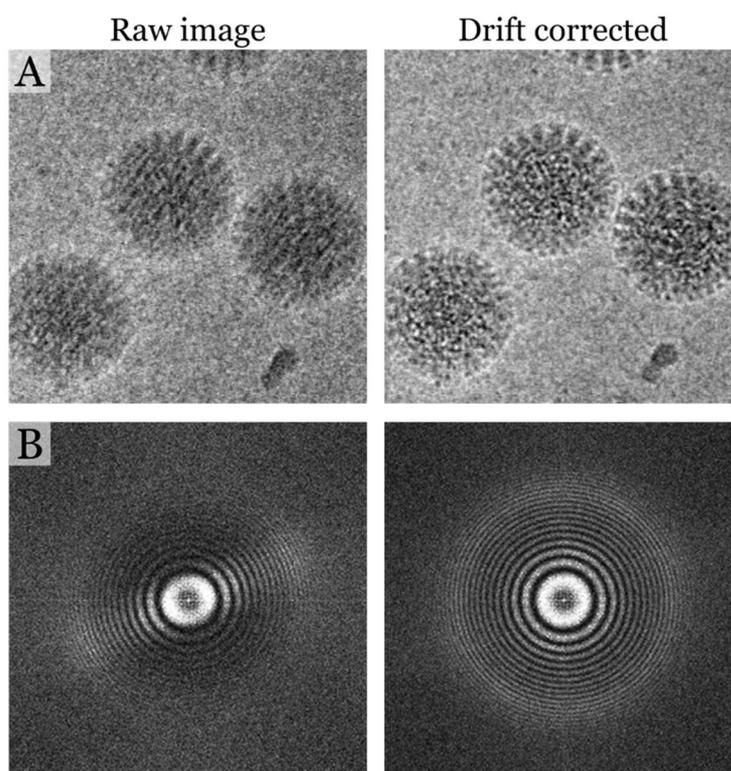


Figure 1.15. The effects of motion correction on **A**) cryo-EM images and **B**) the Fourier transform of cryo-EM images. Motion correction removes blurring associated with specimen movement under the electron beam and restores high-resolution information in images. Adapted from Brilot et al., 2012 and Scherer et al., 2014.

Collecting cryo-EM movies also allows the differential weighting of information in different frames based on spatial frequency. Because radiation damage affects higher-resolution features first (see section 1.2.3), later frames have higher accumulated electron doses and so have degraded signal at high-resolution. However, signal corresponding to low-resolution features is still useful in these later frames, especially for particle visualisation and picking. A dose-weighting scheme has been developed that allows high-resolution information to be down-weighted in later frames, allowing these frames to be retained during image processing without affecting high-resolution information (Scheres, 2014). Although the first two frames of cryo-EM movies contain the highest-resolution information, they are usually discarded as the movement of the sample under the beam in these initial frames is significantly higher than in later frames. This is often attributed to charging effects and doming of the ice layer under the beam (Glaeser, 2016; Scheres, 2014).

Algorithms have also been developed to carry out motion correction on a per-particles basis, as individual particles can have significantly different motion from one another during imaging (Bartesaghi et al., 2018; Scheres, 2014; Zivanov et al., 2019). Additionally, in the most recent algorithm termed Bayesian polishing in RELION-3, the first two frames of movies can be retained. Such algorithms improve the achievable resolution from cryo-EM datasets, and have directly contributed to the recent ability of cryo-EM to resolve structures to resolutions higher than 3 Å (Danev et al., 2019).

### 1.4.2. Automated data collection and improved throughput

When electron microscopy images were collected on film, the data collection process was slow and labour-intensive. When CCD cameras were introduced, despite the fact that images were of lower quality than those recorded on film (see section 1.4.1), they were widely adopted because many more images could be collected in a single session, and much more quickly (Rosenthal, 2015). Crucially, they also allowed the development of software packages for automated data collection, such as Leginon and SerialEM, which could carry out steps such as focussing, setting the eucentric height, correcting for astigmatism, and imaging (Carragher et al., 2000; Mastronarde, 2005; Tan et al., 2016).

Since the introduction of direct electron detectors, it has been possible to rapidly collect many cryo-EM movies in an automated manner, such that it is now feasible to achieve high-resolution reconstructions in under 24 hours (Baldwin et al., 2018). In recent years, software for automatic data collection has improved significantly and packages are now accompanied by user-friendly graphical interfaces to aid microscope alignment and data collection by non-expert users.

A brief overview of automatic data acquisition, as implemented in EPU (Thermo Fisher, Thompson et al., 2019), is summarised in Figure 1.16. A low-magnification montage is first taken of the whole grid, creating an atlas to screen for areas with good ice. Images at medium magnification are then taken of individual grid squares, to allow the selection of target holes for imaging. The data acquisition template is then set up at higher magnification – an area for focussing near the target hole is selected, and the number and position of exposures for each hole are defined. During data collection, the defocus is continuously varied within a user-defined range, and automatic correction for astigmatism is carried out. In this way, once a data collection is launched, imaging can be carried out on selected areas of the grid without any user input.

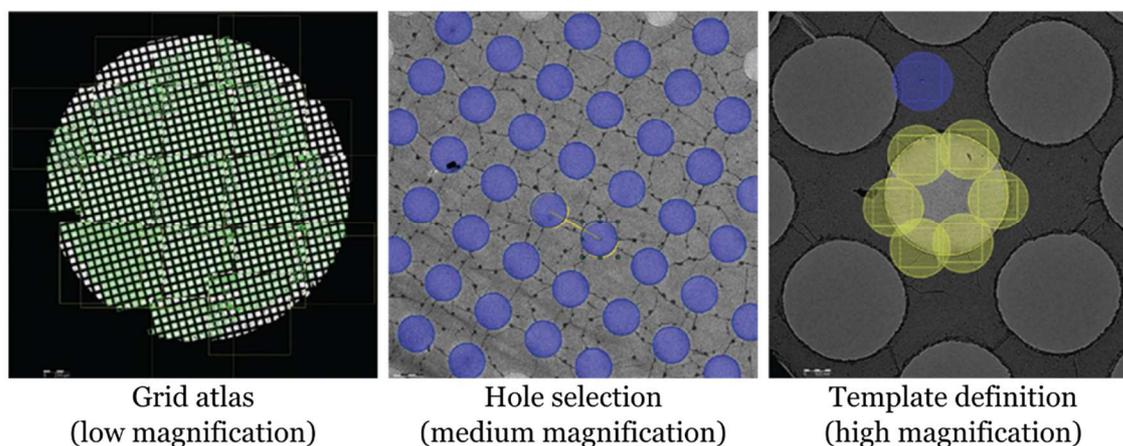


Figure 1.16. Automatic data collection set-up in EPU (Thermo Fisher Scientific). A low-magnification atlas providing an overview of the grid is shown on the left. A medium-magnification grid square image is shown in the middle, with target holes highlighted in blue. Template definition is shown on the right – the blue circle indicates the focussing area, while yellow circles indicate regions targeted for imaging in single hole.

As well as automatic data collection software, there are also changes to microscope hardware facilitating the increased throughput of cryo-EM imaging. On high-end microscopes, such as Titan Krios, Talos Arctica and Glacios microscopes (Thermo Fisher) and CRYO ARM 300 and CRYO ARM 200 microscopes (JEOL), up to 12 grids can be loaded at a single time and rapid switching between grids is possible, significantly reducing the amount of time needed to screen samples before data collection.

Data collection strategies are constantly evolving to improve throughput. One of the longest steps during data acquisition is the wait for drift to settle after physically moving the microscope stage, which can take up to 30 s (Cheng et al., 2018; Danev et al., 2019). Instead of physically moving the stage to each new imaging location, the current practise is to optically shift the electron beam, which does not require a wait time for the stage to settle before imaging (Cheng et al., 2018). This allows the imaging of several locations not only within a single hole, but also on neighbouring holes as shown in Figure 1.17A. This can increase the throughput of data collection up to 10 times, but can also introduce artefacts due to the tilted beam (Cheng et al., 2018; Danev et al., 2019). However, these can be corrected during the data collection process, as described by Wu et al., 2019 or as implemented in the latest version of EPU as ‘aberration-free image shift’ (Konings et al., 2019), or *a posteriori* corrected computationally (as discussed in section 1.5.5). In addition, the recently-described ‘fringe-free illumination’ scheme (Konings et al., 2019) allows for smaller beam diameters to be used due to the elimination of Fresnel fringes at the beam edge, and therefore many more images per hole can be collected (Figure 1.17B). Taken together, these advances mean that it will not be long before several hundred images can be taken with only a few movements of the stage, rapidly increasing the throughput of data collection and meaning that entire datasets can be collected in a single day.

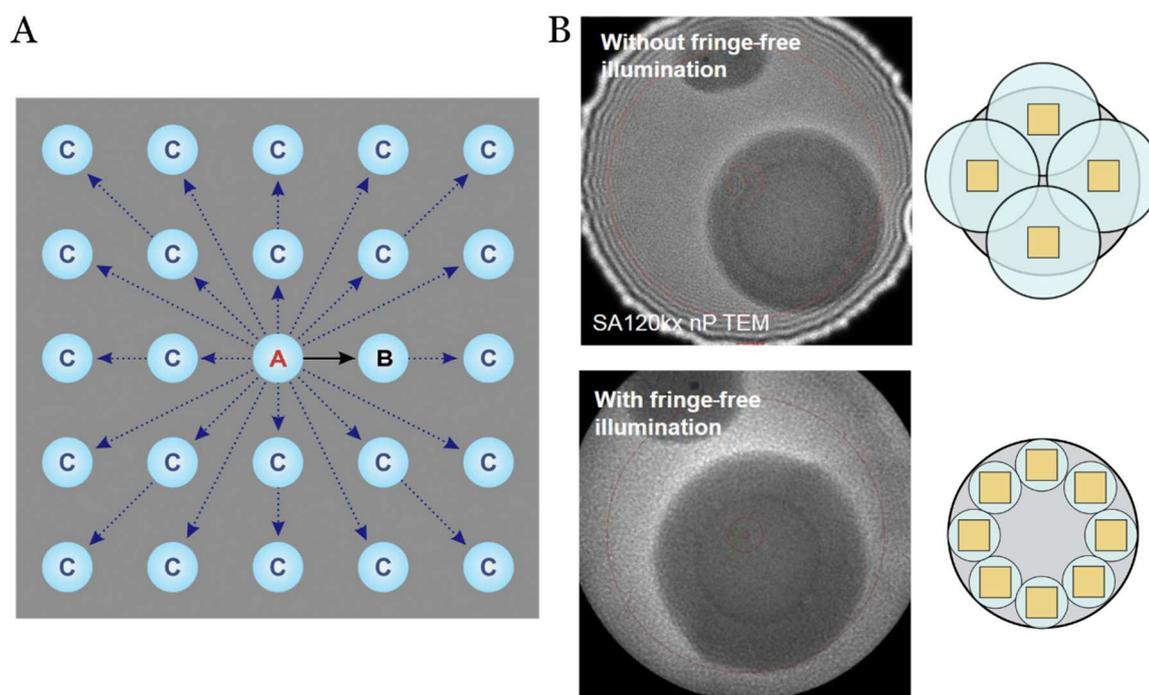


Figure 1.17. **A)** Scheme showing how beam shift can be used to image many areas on a cryo-EM grid without moving the stage. Instead of mechanically moving the stage from one hole 'A' to hole 'B', the beam is optically sifted to image many nearby holes 'C', dramatically increasing the number of images that can be collected with each movement of the stage. Adapted from Danev et al., 2019. **B)** Left - comparison between conventional imaging (above) and 'fringe-free illumination' (below). Fringe-free imaging eliminates the fringes at the edge of the electron beam. This means that many more images (represented by orange squares on the right) can be collected in a single hole (grey circle), as the diameter of the electron beam (blue circles) can be significantly reduced. Adapted from Konings et al., 2019, Themo Fisher.

### 1.4.3. Phase plates

As discussed in section 1.2.1, phase contrast is the primary contributor to contrast in cryo-EM images. There is little contrast in images collected at focus – therefore, images are collected with a defocus that increases the contrast but introduces aberrations due to the CTF (section 1.2.2). Phase plates have been developed as a way of imaging at low defocus or at focus with increased contrast. They introduce a differential phase shift in the unscattered electron beam compared to electrons that are scattered from the sample, which results in additional phase contrast in images (Danev and Baumeister, 2017). The most widely-adopted example of a phase plate is the Volta phase plate (VPP), which consists of a thin amorphous carbon film and is placed at the back-focal plane in the electron microscope (Danev and Baumeister, 2017; Danev et al., 2014). When the unscattered electron beam (which contains the majority of incident electrons) interacts with the phase plate it induces a local charging of the carbon film, which creates an additional phase shift in the unscattered electrons (Wang and Fan, 2019). This can dramatically increase the contrast of cryo-EM images, as shown in Figure 1.18.

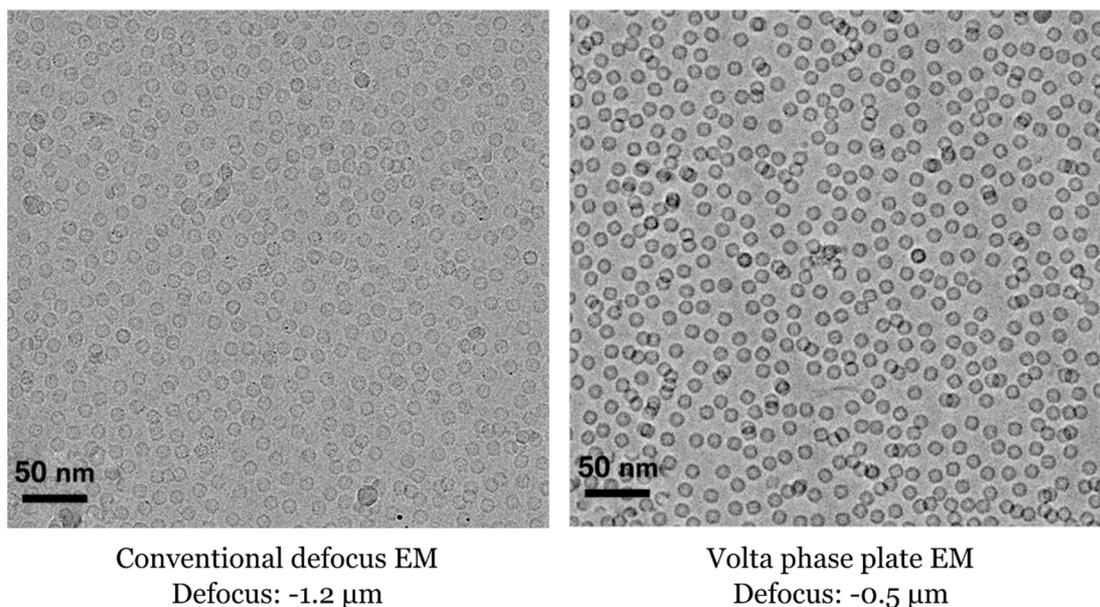


Figure 1.18. Comparison between conventional defocus cryo-EM (left) and Volta phase plate cryo-EM (right) on apoferritin. Despite the higher defocus used in the conventional cryo-EM image, the contrast is much higher when imaging with the phase plate. Adapted from Wang and Fan, 2019.

Proteins smaller than  $\sim 200$  kDa have historically been difficult to study using cryo-EM, because of the low contrast in images. The VPP opened up the possibility of imaging such proteins and held great promise for improving the feasibility of studying difficult targets (Danev et al., 2019; Fan et al., 2019; Khoshouei et al., 2016, 2017). However, recent results have demonstrated that using a conventional defocus-based approach is actually sufficient for reaching high resolution even for complexes with a molecular weight of less than 100 kDa (Herzik et al., 2019; Zhang et al., 2019). In addition, imaging with the VPP can be technically challenging and complicates the image processing process. The phase plate itself also reduces the information content of images by up to 18% because of interaction between electrons and the carbon film (Danev et al., 2019). Because of this, there has been a recent shift away from imaging with phase plates.

However, a recent paper has detailed the use of a laser phase plate in transmission electron microscopy (Schwartz et al., 2019). Instead of a phase shift being generated by the interaction between electrons and a carbon film, the phase shift results from the interaction between electrons and light. This overcomes many of the technical challenges of the VPP, such as loss of information and a changing phase shift over time (Danev et al., 2019; Schwartz et al., 2019). Further development of the laser phase plate may bring improvement, but this is still a few years away from being fully realised (Danev et al., 2019).

## 1.5. Image processing developments – a new set of tools in the microscopist’s toolbox

There have been several key developments from a software point of view that have improved image processing – not only in terms of the achievable resolution, but also in terms of usability, speed, and dealing with flexible proteins.

The software package RELION (Scheres, 2012a) was released in 2012 and rapidly became the most widely-used cryo-EM image processing package (Cheng, 2018). RELION is based on a maximum-likelihood approach – rather than assigning each particle a single orientation during refinement, it is assigned multiple orientations which are weighted according to the probability of each, given the experimental data and the reference map (Scheres, 2016). This was more powerful than earlier approaches, particularly with regard to dealing with structural heterogeneity (Cheng, 2018). In addition, RELION was more user-friendly than previous packages, and helped to open up image processing to non-expert users.

In the past five years there have been many other image processing software packages released which also use the maximum likelihood approach, including CryoSPARC (Punjani et al., 2017), SPHIRE (Moriya et al., 2017) and *cis*TEM (Grant et al., 2018). However, RELION is still the most commonly-used software package – 82% of deposited entries in the Electron Microscopy Data Bank (EMDB) in 2019 were calculated in RELION, compared to 10% for CryoSPARC and less for other packages. The speed of image processing has dramatically improved over the past few years, primarily because of GPU (graphic processing unit) acceleration in several of these software packages (Kimanius et al., 2016; Punjani et al., 2017; Zhang, 2016). This has opened up the ability to carry out image processing ‘on-the-fly’, as discussed in section 1.7.2.

Some of the key developments in cryo-EM image processing, particularly in terms of improving the achievable resolution, are discussed in the following sections.

### 1.5.1. Improved CTF estimation

As discussed in section 1.2.2, correction of the CTF is necessary to determine the structures of macromolecules to high resolution, and errors in the estimation of the CTF limit the achievable resolution of reconstructions from cryo-EM images (see Figure 1.19) (Zhang and Zhou, 2011). Larger errors in defocus estimation are tolerated at lower resolutions, but as resolution increases the precision of defocus estimation can be a limiting

factor. With reconstructions from cryo-EM images now frequently reaching resolutions of 3 Å and above, new ways of estimating the defocus have been developed to more accurately correct the CTF (Danev et al., 2019).

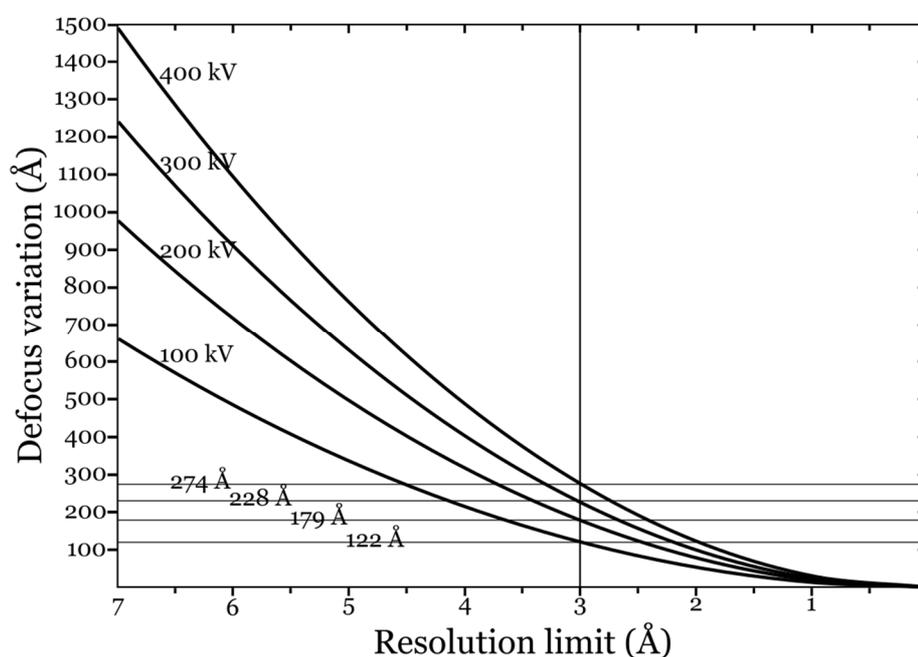


Figure 1.19. Relationship between the error in defocus estimation and the resulting resolution limit at different accelerating voltages (100 kV - 400 kV). At higher resolutions, the maximum tolerated error in defocus estimation decreases. The allowable error in defocus estimation to achieve a 3 Å reconstruction is indicated by the horizontal lines. Adapted from Zhang and Hong Zhou, 2011.

CTF estimation was traditionally carried out on a per-micrograph basis, giving a single defocus estimate for every particle in an image. However, cryo-EM grids are often not perfectly flat in the microscope, and particles frequently lie at different heights in the ice which can be much thicker than the longest particle axis. This can cause local deviations from the global defocus estimation (Noble et al., 2018a; Zhang, 2016). At higher resolutions, correcting for these local defocus differences can improve the resolution of reconstructions (Zhang, 2016; Zivanov et al., 2018).

There are several software packages which have now implemented CTF estimation on a local or per-particle basis, including GCTF (Zhang, 2016), RELION (Zivanov et al., 2018), CryoSPARC (Punjani et al., 2017) and WARP (Tegunov and Cramer, 2019). For example, there are two contrasting approaches implemented in GCTF and RELION-3. Local CTF estimation in GCTF first carries out an initial CTF estimation for the entire micrograph. Using coordinates from picked particles, it locally refines this estimate by estimating the CTF on the area around a particle, taking into account nearby particles (Zhang, 2016). However, this requires a high enough signal-to-noise ratio to see Thon rings in power spectra of patches from around individual particles.

In contrast, the recent CTF refinement algorithm in RELION-3 uses an approach that does not require fitting of Thon rings in power spectra (Zivanov et al., 2018). Rather, a 3D reference structure (usually from the data itself) is projected in the previously-determined viewing direction for each particle, and multiplied by the CTF (as initially estimated) in Fourier space. The Fourier transform of the particle is then compared by cross correlation to the Fourier transformation of the projected volume multiplied by the CTF. The CTF parameters used to modulate the reference projection are then varied to maximise the cross correlation between the two, resulting in a better estimation of the CTF for each particle. This has been shown to improve the resolution of 3D reconstructions more than other algorithms, and is especially useful for cases where data collection has been carried out on tilted grids (see section 1.6.2) (Zivanov et al., 2018).

### 1.5.2. Particle picking

In order to obtain high-resolution reconstructions of proteins, it is regularly necessary to select several hundred thousand noisy particle images from micrographs. There are many different ways of automatically selecting particles in cryo-EM images, many of which are template-based. These often use 2D class averages or projections of 3D volumes as references, which are used to find cross-correlation peaks when moved sequentially over a micrograph. Many image processing packages have their own particle picking algorithm, and several stand-alone particle picking algorithms have been also been developed such as Gautomatch (<http://www.mrc-lmb.cam.ac.uk/kzhang/Gautomatch/>) and Fast Projection Matching (Estrozi and Navaza, 2008).

While these particle picking algorithms have improved greatly over the past few decades, many of the cross-correlation based methods suffer from a high false positive rate because of the low signal-to-noise ratio and so manual inspection of particle picks is often needed (Bepler et al., 2019a; Wagner et al., 2019). Such picking algorithms often struggle to correctly identify particles with unusual shapes, such as elongated particles, or have difficulties in picking particles from sub-optimal data. In addition, selecting particles based on template matching has the danger of introducing template bias, which can cause ‘Einstein from noise’ scenarios in which pure noise particle images are selected based on alignment to features in the template, which can lead to incorrect structures being calculated from the data (Van Heel, 2013; Henderson, 2013; Subramaniam, 2013).

Recently, there have been several new approaches to picking particles that have improved the robustness of particle selection. There are now reference-free methods for selecting particles, which use Gaussian blobs with defined diameters to select particles

without any prior information and are now implemented in several image processing software packages (Grant et al., 2018; Voss et al., 2009; Zivanov et al., 2018). These approaches remove any danger of reference bias, and do not require prior structural information for particle selection unlike template-based methods. However, these methods still suffer from high false positive rates and often pick high-contrast contaminants such as frozen ethane or ice crystals.

There is now a move towards using neural network-based approaches to selecting particles from micrographs, and neural network-based particle picking algorithms have been developed recently, including crYOLO (Wagner et al., 2019) as implemented in the SPHIRE package, Topaz (Bepler et al., 2019a), the BoxNet particle picking algorithm in WARP (Tegunov and Cramer, 2019), amongst others. Such algorithms require training on a small subset of micrographs which have picked particle coordinates, in order to use machine learning to identify patterns in the noisy data. These methods outperform traditional particle picking algorithms in terms of the number of particles recovered from micrographs, the false positive rate and the picking of difficult or non-standard datasets. Further discussion of neural networks applied to cryo-EM is presented in section 1.7.1.

### 1.5.3. Initial model generation

Historically, one of the bottlenecks of cryo-EM image processing was the generation of an initial volume for aligning particle images during refinement. Using an incorrect starting model would often cause refinements to fail to converge upon the true structure, which is often termed as becoming trapped in local minima (Lyumkis et al., 2013; Orlova and Saibil, 2011).

Two key approaches that were historically used to generate initial models are the common lines approach (Van Heel, 1987) and random conical tilt (Radermacher et al., 1987). The common lines approach is based on the fact that every pair of 2D projection images of a 3D object will share a common 1D line projection. In 3D Fourier space, the common line between two 2D projections is the intersecting line between the two Fourier planes (Orlova and Saibil, 2011). With three or more 2D projection images, the relative orientations between the images can be determined and used to fill up Fourier space to generate an initial model. However this process does not always provide optimal results, and is most suited to objects with high symmetry (Orlova and Saibil, 2011).

In contrast, random conical tilt makes use of tilt pairs of images to generate initial models. One image of particles is taken without tilt, and a second image is taken of the

same particles with a high tilt angle of  $45\text{--}60^\circ$  (Radermacher et al., 1987). This provides pairs of particle images for which the relative orientation is known. Particles with the same orientation in the untilted image with respect to the electron beam (but different in-plane rotations in the ice) are classified together. The corresponding tilted images for each particle fill up a cone of orientations as shown in Figure 1.20, allowing partial filling of the 3D Fourier space in and the calculation of an initial model. However, this method is time consuming and difficult, especially for cryo-EM images, and the resulting maps often suffer from artefacts due to the incomplete sampling of Fourier space (Cheng et al., 2015).

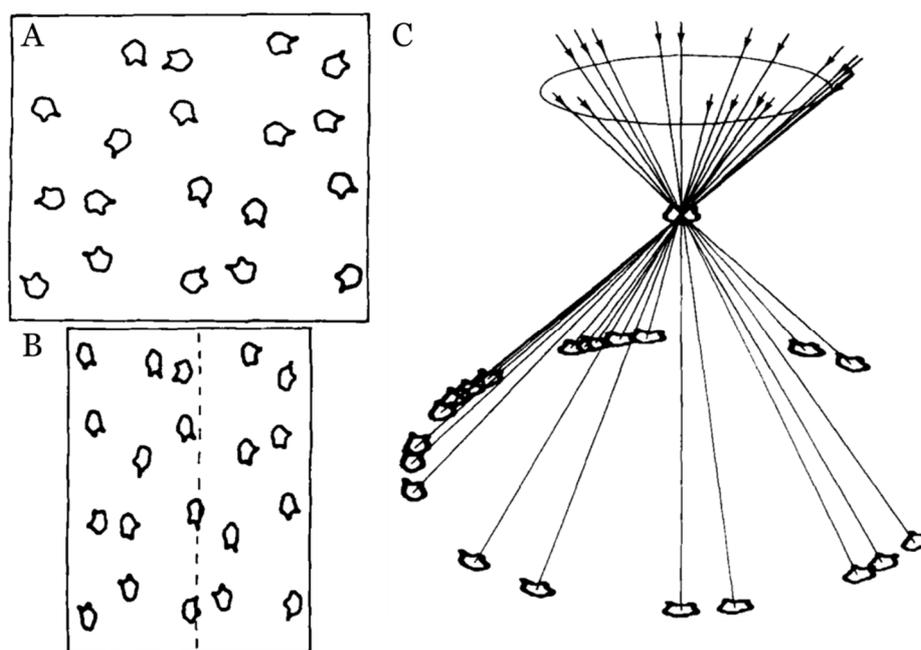


Figure 1.20. Random conical tilt involves the collection of untilted (A) and (B) tilted micrographs of the same particles to generate tilt pairs. The random in-plane rotation of particles in the same orientation allows the filling of a cone of Fourier space, depicted diagrammatically in (C). Adapted from (Radermacher et al., 1987).

Both random conical tilt and the common lines approach enabled many structures of macromolecules to be determined without prior knowledge. However, they gave inconsistent results and often had difficulty in converging on a correct structure if the initial volume was too far from the real structure (Cheng et al., 2015; Punjani et al., 2017; Scheres, 2012b).

One of the most significant recent improvements in the determination of initial models is the implementation of a stochastic gradient descent algorithm in CryoSPARC (Punjani et al., 2017). This algorithm is much more rapid than traditional methods, and is much better at avoiding local minima in model generation. Rather than using entire particle sets it uses random subsets of the data, and by sampling different subsets of particles at each step there is a higher probability of converging on a true solution, even if the initial volume is very far away from the target structure.

There is now also the possibility of calculating multiple initial models at once, and simultaneously classifying particles between the different classes. This is now implemented in CryoSPARC as *ab initio* reconstruction, and was later incorporated into the initial model generation algorithm of RELION. This approach allows the identification of multiple different 3D species in a mixed population, and performs well in dealing with heterogeneous mixtures.

#### 1.5.4. Dealing with flexibility and partial occupancy

Many proteins are not static in solution, but undergo conformational changes in order to carry out their cellular functions. Conformational heterogeneity can lower the resolution of cryo-EM maps because the averaging of multiple states blurs certain regions during refinement. Compositional heterogeneity, such as partial occupancy at a ligand binding site, is often seen. Such problems are often dealt with by using 3D classification to separate mixed populations of particle images (see section 1.3). However, in cases where the variable region is only a very small part of the whole protein, regular 3D classification can struggle to separate out different states. In addition, discrete classification methods are not well-adapted to cases where there is continuous motion of domains instead of a few defined states (Jonić, 2017; Nakane et al., 2018; Scheres, 2016).

Recently, several new approaches for dealing with flexible regions or partial occupancy have been developed. These include methods that deal with both discrete and continuous variability, and also methods that mitigate over-fitting of noisy regions in the map. These approaches have dramatically improved the way that flexibility in proteins is dealt with during image processing and a few of the most widely-used approaches are discussed in the following sections.

##### 1.5.4.1. *Focussed refinement, classification, and signal subtraction*

3D classification is well-adapted to sorting out large conformational changes in proteins and complexes. In some cases however, structural variation can occur on a much smaller basis, such as movement of a small domain or the presence of a small binding partner, and it can be difficult to resolve these small changes (Scheres, 2016; Serna, 2019). In such cases, alignment is often driven by the larger part of the protein and the density for the smaller part blurs out because of its flexibility (Scheres, 2016). Focussed refinement and focussed classification are two approaches to get around this problem, and both involve the creation of a mask around the region of interest. This mask removes the

contribution of the larger portion of the particle during refinement, meaning that the alignment is instead driven by the smaller, more variable region.

Focussed refinement has been especially useful for studies on large proteins with multiple domains, or on large multi-protein complexes, in which the variable domain is only a small part of the total particle but still has enough signal to allow the assignment of angles to the sub-volume. One example is the use of focussed refinement on the ribosome, which consists of a large and a small subunit. The large subunit often drives alignment of the particle during regular 3D refinement, and masking out the large subunit improves the resolution of the small subunit after refinement (von Loeffelholz et al., 2017; Wong et al., 2014b).

The size limit for carrying out focussed refinement is approximately the same for regular single particle electron microscopy, on the order of ~100-200 kDa (Scheres, 2016). This is because assigning angles to regions smaller than this is difficult due to the limited amount of signal. For such regions, masked classification can instead be carried out. This method makes use of the angles previously determined for each particle by the consensus refinement, and only classifies the region inside the mask. This allows the rapid separation of particles into different classes based on variability in the masked region (Scheres, 2016).

However, particle images being compared to the volume during focussed refinement and classification still contain signal from regions that lie outside the mask, as they are projections of the entire protein and not only the region of interest. This introduces noise into the reconstruction, degrading the quality of reconstruction or classification (Bai et al., 2015b; Scheres, 2016). In order to mitigate this, the signal from the region outside the mask can be subtracted from the particle images, as shown and described in Figure 1.21. This improves the accuracy of masked refinement and classification, and therefore more accurate separation of different conformational states is possible. Such an approach has been used in many cases for improving the structural characterisation of small regions in larger complexes where traditional 3D classification is not sufficient (Desfosses et al., 2019; Fei et al., 2020; Monroe et al., 2017; Zhou et al., 2015b).

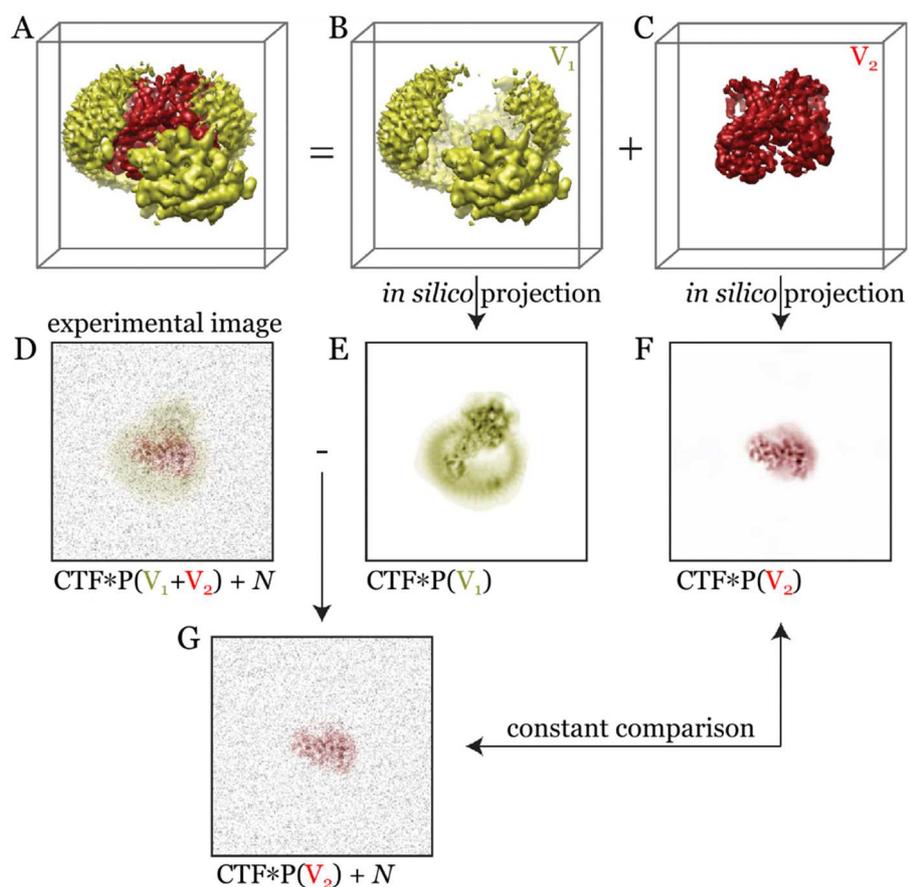


Figure 1.21. Overview of signal subtraction on  $\gamma$ -secretase, adapted from Bai et al., 2015b. A 3D map of the entire complex in **A** is composed of the region to be subtracted ( $V_1$ ) shown in **B** and the region of interest ( $V_2$ ) shown in **C**. A 2D particle image **D** is a projection ( $P$ ) of  $V_1 + V_2$  with added noise ( $N$ ), convolved by the CTF. Signal subtraction is carried out by subtracting an *in silico* projection of  $V_1$  convolved by the CTF **E**, leaving only signal from  $V_2$  and experimental noise in the subtracted image **G**. Therefore during masked refinement and classification, signal from  $V_1$  does not contribute – the assignment of angles is based on comparison between projections of the region of interest shown in **F**) and the subtracted images **G**), rather than **D**).

#### 1.5.4.2. Dealing with continuous variation

Where variation is instead continuous, there are other approaches that are more suitable than masked refinement and classification. One of these is the 3D variability analysis algorithm implemented in CryoSPARC – this makes use of principal component analysis (PCA) to describe continuous motions in 3D. The algorithm requires a consensus 3D refinement from all particles to be carried out, and then automatically finds the regions in a reconstruction that have the highest variance in 3D. This approach calculates linear trajectories (termed eigenvectors) in 3D volumes along which there is significant variation, and outputs ensembles of volumes which describe intermediate states along these trajectories (Punjani and Fleet, 2020). These volumes can then be used to prepare movies for visualisation describing the major 3D variance in the dataset.

The multi-body refinement approach in RELION is particularly suited to studying proteins with multiple domains that move independently to one another (Nakane et al., 2018). A consensus refinement is first carried out and user-defined masks are created around individual domains. In an extension of the signal subtraction approach discussed above, signal from regions outside each masked domain is subtracted from images and each domain is refined in a separate reconstruction. This partial signal subtraction approach is carried out iteratively – for each particle image, the relative orientations of each domain are tracked and signal subtraction improves at each iteration. Multi-body refinement then makes use of PCA to output movies describing the major motions between the different bodies, similarly to 3D variability analysis in CryoSPARC. One example is shown in Figure 1.22 – the spliceosome contains four domains which move independently resulting in smearing of the density, particularly for the helicase and SF3b domains. Multi-body refinement significantly improves the map, allowing the visualisation of secondary structural features in the final reconstruction.

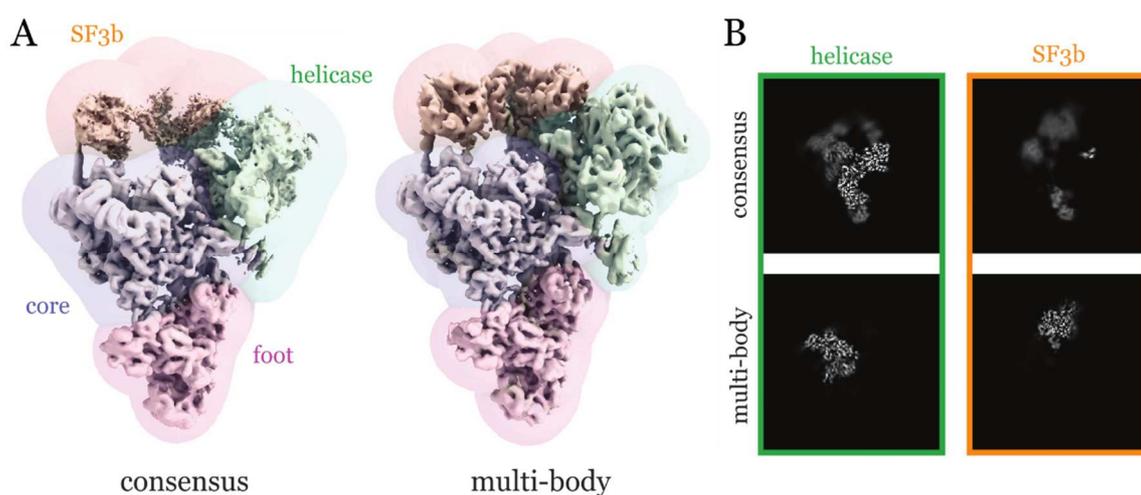


Figure 1.22. Multi-body refinement in RELION of the spliceosome. **A)** Four domains from the consensus refinement of the spliceosome (left) are masked as indicated. The SF3b and helicase domains in particular show poor density in the consensus refinement due to flexibility. The reconstruction after multi-body refinement (right) shows improved density for all regions. **B)** Slices through the consensus reconstruction (above) and multi-body refinement (below) for the helicase and SF3b regions. In the consensus refinement, both domains appear as diffuse density, whereas after multi-body refinement they become well-defined. Adapted from Nakane et al., 2018.

There are other software packages which employ other methods for analysing continuous variation, such as WarpCraft (Schilbach et al., 2017) and HEMNMA (Harastani et al., 2020; Jin et al., 2014) which use normal mode analysis to improve the refinement of flexible proteins. This approach is well-suited to cases where movements are correlated between different regions of the map, but for cases where domains move independently from each other tools like multi-body refinement may be more appropriate (Nakane et al., 2018). The recently-described package CryoDRGN instead uses a neural network approach

for separating out continuously-varying states directly from 2D particle images, without needing an initial model to start from (Zhong et al., 2019).

#### 1.5.4.3. *Non-uniform refinement – local regularisation*

Another algorithm implemented in CryoSPARC termed non-uniform refinement also deals with flexibility and disordered regions in cryo-EM maps, by changing how regularisation is performed during 3D refinement (Punjani et al., 2019). Regularisation is a term which refers to the imposition of prior knowledge during maximum likelihood refinement, as carried out by software packages such as RELION and CryoSPARC. The prior knowledge used to regularise protein structures includes the fact that protein density tends to be continuous or smooth (Scheres, 2012b). This results in the filtering of noise at high spatial frequencies where the signal-to-noise ratio is low (as determined by the global resolution cut-off at an FSC value of 0.143), reducing overfitting and therefore lowering the amount of noise present in 3D maps (Scheres, 2012b).

In many proteins however, different regions of the map have different amounts of disorder or motion, and carrying out regularisation homogeneously across an entire reconstruction may result in some regions being over-regularised and others being under-regularised (Punjani et al., 2019). In regions with high motion or disorder, under-regularisation can result in the appearance of noise in the map (i.e. over-fitting). Conversely, over-regularisation in rigid regions of the map can mean that high-resolution signal is not taken into account and the region is over-smoothed, lowering the resolution.

The non-uniform refinement algorithm implemented in CryoSPARC works by carrying out regularisation differently on different regions of the map. Regularisation is carried out at each iteration on a local rather than a global basis, taking into account neighbouring regions, with the end result that disordered areas have more noise removed without over-smoothing well-structured areas of the particle. This calculation is carried out separately on each independent half-map during refinement (see section 1.3). Non-uniform refinement can improve the reconstruction at each iteration, reducing noise in disordered regions and improving the high-resolution features of rigid regions, which therefore improves the accuracy of particle angle assignments in following iterations.

Non-uniform refinement can be especially useful for membrane proteins which often have large disordered regions in detergent micelles or lipid nanodiscs surrounding transmembrane helices, and can dramatically improve the result of 3D refinement as shown in Figure 1.23. More recently, another algorithm called SIDESPLITTER has been

released, which takes a similar approach to non-uniform refinement in CryoSPARC (Ramlaul et al., 2020).

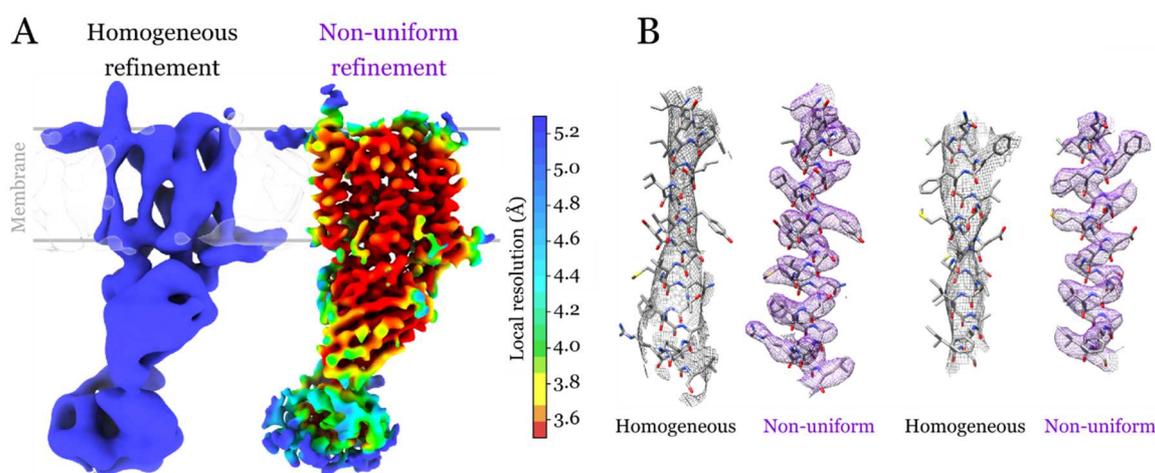


Figure 1.23. Comparison between standard homogeneous refinement and non-uniform refinement of the membrane protein PfCRT in lipid nanodiscs, adapted from Punjani et al., 2019. **A)** Full maps after homogeneous (left) and non-uniform (right) refinement coloured according to local resolution. **B)** Two alpha helices after homogeneous and non-uniform refinement, showing a dramatic improvement in the high-resolution features of the map.

### 1.5.5. Aberration correction

Electron microscopes need to be precisely aligned in order to record high-quality images of protein complexes. One important alignment is the calibration of beam tilt, which ensures that the electron beam is parallel to the optical axis of the microscope (i.e. perpendicular to the sample) (Glaeser et al., 2011). Failure to do so results in coma, which causes phase shifts in the image, particularly at high resolution (Zhang and Zhou, 2011). Beam tilt is also relevant when collecting images in multiple grid holes using image shift, as described in section 1.4.2, as shifting the beam away from the optical axis introduces additional phase errors associated with a tilted beam (Zivanov et al., 2018).

The effects of beam tilt are particularly relevant at resolutions better than  $\sim 3$  Å, and so have not historically been a large limitation (Zhang and Zhou, 2011; Zivanov et al., 2018, 2020). However with the rapid developments in cryo-EM in the past few years, such resolutions are now becoming more frequently attainable and the limitations imposed by aberrations are becoming more and more relevant. Pushing the achievable resolution of cryo-EM is particularly relevant for the pharmaceutical industry in the context of drug discovery, which require very high resolutions to characterise drug binding to protein targets.

Until recently, there was no way to correct for errors in beam tilt alignment once data had been collected. In the past year, there have been computational methods developed to *a posteriori* correct for the errors in phase shifts introduced by beam tilt (Zivanov et al., 2018). This has led to improvements in reconstructions at very high resolutions, and means that the perfect alignment of beam tilt in microscopes is no longer strictly necessary to achieve such resolutions. Particularly as cryo-EM moves into the mainstream, it is important to make the technique robust enough so that data can be collected by non-expert users, and the ability to correct for imperfect microscope alignment during data collection aids progress towards this goal.

More recently, the correction of so-called higher-order or antisymmetrical aberrations has been made possible and is now implemented in RELION-3.1 and CryoSPARC v2.12 (Falcon et al., 2019; Zivanov et al., 2020). These higher-order aberrations also introduce phase shifts into images, but don't require an understanding of the physical basis in order to carry out correction. Correction of such aberrations help to push the achievable resolution of cryo-EM reconstructions even further, with the structure of mouse heavy-chain apoferritin being determined to a resolution of 1.57 Å with the correction for higher-order aberrations (Zivanov et al., 2020).

## 1.6. The next revolution – sample preparation

The current bottleneck in routinely reaching high resolution reconstructions of proteins is not image processing software or microscope limitations, but rather at the level of sample preparation (Danev et al., 2019; Drulyte et al., 2018). The process of freezing grids for cryo-EM has remained virtually unchanged for decades. As discussed in section 1.2.3, it typically involves the application of a few microliters of sample to the grid, blotting of excess sample with filter paper, and plunging into the cryogen (Adrian et al., 1984; Dobro et al., 2010; Taylor and Glaeser, 1976).

While this technique has given many excellent results, it is far from perfect. One key limitation of current sample preparation methods is a lack of ice reproducibility (Armstrong et al., 2019). Very recently, there has been significant work carried out by several groups to design new grid-freezing technologies, which are starting to yield promising results. A second limitation is the nature of proteins to adopt a preferential orientation in the ice (Glaeser and Han, 2017). There is currently a large amount of research investigating ways of mitigating preferential orientation, such as the addition of a detergents to the sample, tilting during data collection, and faster vitrification techniques.

### 1.6.1. Variation in ice thickness

The variability of ice thickness between grids frozen with identical conditions, and even across one grid, is a common problem in the field and results in significant lost time in grid screening and wasted sample. The cause of ice thickness variation was recently investigated using a combination of scanning electron microscopy (SEM) and light microscopy, and revealed that a major contributing factor is the variable arrangement of filter paper fibres when blotting (Armstrong et al., 2019). This causes the blotting of liquid to occur in a non-uniform way, resulting in variations in ice thickness across and between grids. A typical example is shown in Figure 1.24, with a gradient of ice thickness from one side of the grid to the other.

To improve the reproducibility and consistency of ice on cryo-EM grids, several alternative methods for freezing cryo-EM grids are being developed. Many of these rely on the application of much smaller sample volumes to the grid, which do not require the removal of excess liquid. One such system is the VitroJet system, which uses pin printing to deposit a volume less than 1 nL across the surface of a grid, resulting in thin and homogeneous ice layers, before rapidly freezing the grid in twin jets of ethane (Ravelli et

al., 2019). Another is the CryoWriter system, which uses a microcapillary to deposit between 3-20 nL of sample on the grid and then monitors water evaporation to control the thickness of vitrified ice (Arnold et al., 2017). Both systems give reproducibly thin ice layers, improving the efficiency of the sample preparation process and minimising sample waste.

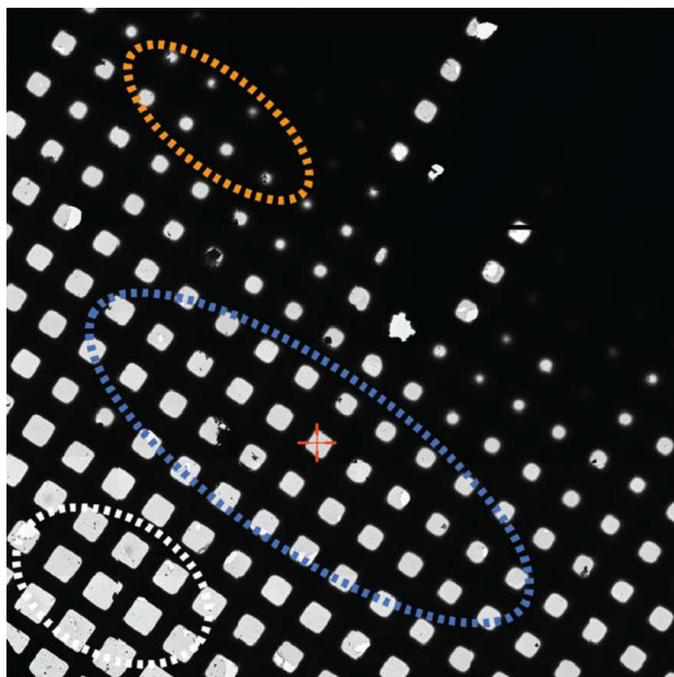


Figure 1.24. Variation in ice thickness across a cryo-EM grid. The area indicated by the orange circle contains ice that is too thick, while the area in the white circle indicates areas with no ice. The grid squares indicated by the blue circle contain ice of appropriate thickness for data collection. Adapted from Thompson et al., 2019.

Other freezing devices have been developed that spray very small amounts of sample onto the cryo-EM grid, which are then plunge-frozen without blotting. These devices make use of a variety of different systems for sample spraying, including microfluidic systems, ultrasonic humidifiers or piezo inkjet dispensers (Feng et al., 2017; Jain et al., 2012; Rubinstein et al., 2019). One device in particular, originally termed Spotiton but now commercialised as chameleon, has been under development for several years and has started to provide exciting results (Dandey et al., 2018; Jain et al., 2012; Noble et al., 2018b). In this system, tiny droplets of sample are sprayed from a piezo as the grid is plunged through the spray into liquid ethane (Figure 1.25A). The grids used are covered in nanowires and are ‘self-wicking’ (Figure 1.25B), meaning that as liquid hits the grid surface, it spreads out rapidly to form a homogeneous thin layer resulting in reproducibly-thin ice (Wei et al., 2018). The further development and commercialisation of such systems over the next few years should dramatically improve the efficiency and reproducibility of grid freezing.

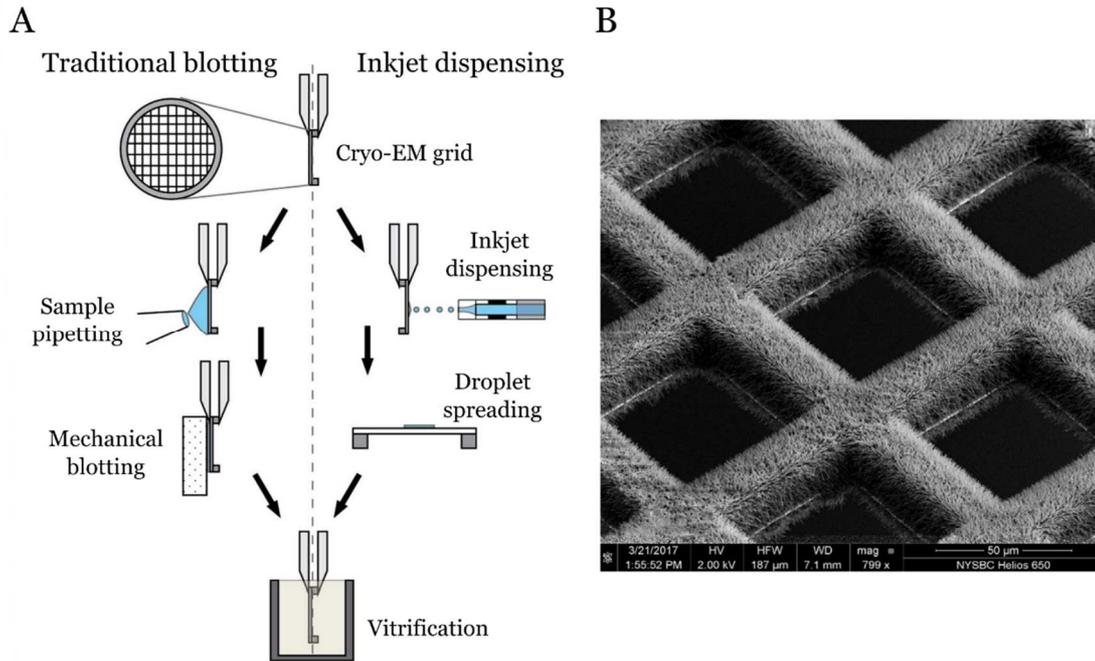


Figure 1.25. The Spotiton system for freezing cryo-EM grids. **A)** Comparison between traditional grid freezing and Spotiton. Adapted from Jain et al., 2012. **B)** SEM image of self-wicking cryo-EM grids showing nanowires on grid bars. Adapted from Wei et al., 2018.

### 1.6.2. Preferential orientation

In an ideal situation, proteins on a cryo-EM grid are dispersed homogeneously across a grid hole, randomly oriented in centre of a thin vitreous ice layer. However, for the majority of protein samples, this is not the case (D’Imprima et al., 2019; Danev et al., 2019). This can result in stretched cryo-EM density maps due to the anisotropic distribution of views during 3D reconstruction, and therefore incomplete filling of the Fourier space (Chen et al., 2019; Drulyte et al., 2018).

The primary cause of preferential orientation is adsorption of proteins to the air-water interface after sample application to the grid, before the moment of freezing, as shown in Figure 1.26 (Drulyte et al., 2018; Glaeser and Han, 2017; Noble et al., 2018a). A recent study showed that approximately 90% of proteins adhere to the air-water interface when frozen with standard grid freezing techniques (Noble et al., 2018a). One hypothesis for this phenomenon is that proteins preferentially adhere to the air-water interface to expose hydrophobic surfaces, causing a non-random distribution of orientations (Glaeser and Han, 2017; Noble et al., 2018a). This adsorption often causes local denaturation of the protein, further degrading the quality of data (D’Imprima et al., 2019).

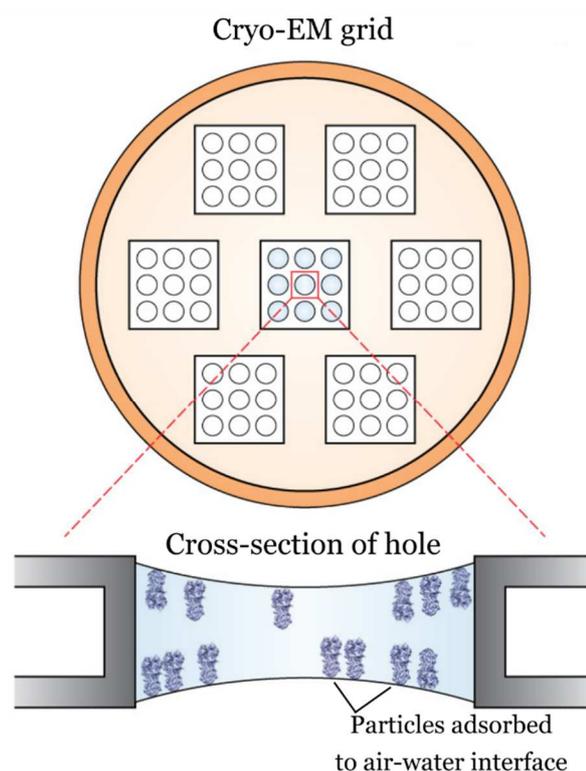


Figure 1.26. Cross-section of a cryo-EM grid hole showing preferential orientation of particles caused by the adsorption to the air-water interface.

There are several possibilities for overcoming preferential orientation in cryo-EM. The simplest method is to collect a percentage of images with a tilted grid, at angles around 30-40° (Zi Tan et al., 2017). This does not change the orientation of proteins in the ice, but increases the coverage of Fourier space during 3D reconstruction. However, this method has its drawbacks – it does not overcome the effects of protein denaturation at the air-water interface, and it does not result in a completely isotropic distribution of views. In addition, tilting the grid increases the thickness of the ice relative to the electron beam, resulting in more inelastic scattering. Despite this, tilting during data collection can dramatically improve the quality and resolution of cryo-EM maps, and is relatively simple to implement (Rivera-Calzada et al., 2017; Su et al., 2017; Zi Tan et al., 2017).

Several other approaches relate to properties of the cryo-EM grid itself. The surface charge of the grid can be modified by glow discharging in the presence of molecules like amylamine, which can sometimes change the distribution of proteins in the ice (Passmore and Russo, 2016). Adding a thin continuous support film made from carbon can also change the distribution of views, as proteins adhere to the film rather than the air-water interface (D’Imprima et al., 2019; Drulyte et al., 2018). However, this can increase the background of images and lower contrast, particularly for small proteins (Passmore and Russo, 2016). Graphene or graphene oxide films can provide an alternative, and do not result in as strong a background as carbon films (Drulyte et al., 2018; Pantelic et al., 2010;

Passmore and Russo, 2016). In addition, there is current research into the development of so-called affinity grids, where proteins specifically bind to molecules on the cryo-EM grid. One example is the streptavidin grid – monolayers of streptavidin crystals are grown on the grid, and randomly-biotinylated proteins bind strongly to streptavidin, which can improve the distribution of orientations (Han et al., 2016; Noble et al., 2018a).

The addition of detergents or surfactants to the sample before freezing can also dramatically mitigate preferential orientation. These molecules can themselves adhere to the air-water interface, preventing protein adsorption and reorienting particles in the ice layer (Chen et al., 2019; Glaeser and Han, 2017). One commonly-used detergent for this purpose is CHAPSO, which has been shown to almost completely eliminate interaction with the air-water interface for bacterial RNA polymerase and can significantly reduce preferential orientation (Chen et al., 2019).

Finally, it has been estimated that in a thin film of liquid, proteins can collide with the air-water interface more than 1000 times per second (Glaeser and Han, 2017; Taylor and Glaeser, 2008). Therefore, minimising the amount of time between the thin film formation (either after traditional blotting or the application of very small sample volumes) and vitrification reduces the chance of proteins adhering to the air-water interface. Several of the recent developments discussed in section 1.6.1 for improving the reproducibility of ice thickness also significantly speed up the grid freezing process, improving the distribution of particle orientations. The typical time from sample application to freezing is on the order of a few seconds, whereas recent grid-freezing devices are able to reduce this time to under 100 ms (Feng et al., 2017; Noble et al., 2018b; Rubinstein et al., 2019). This shorter freezing time can directly result in reduced preferential orientation, and improves the quality of the resulting 3D reconstruction (Noble et al., 2018b). Reducing the freezing time also opens up the possibility of carrying out time-resolved cryo-EM, to investigate transient states at defined time-points during protein activity (Fu et al., 2019).

## 1.7. Future directions

After the initial ‘revolution’ in cryo-EM in the 2010s, the field has expanded rapidly and is now used by an increasing number of labs every year (Danev et al., 2019). While cryo-EM was initially used mostly to characterise large symmetric protein complexes such as viruses, the past decade has seen cryo-EM adopted to study a huge variety of targets, from small asymmetric proteins to dynamic and flexible macromolecules. There are currently a large number of methodological developments being published in the cryo-EM field, both from hardware and image processing perspectives, which will further push the field towards the routine high-resolution characterisation of macromolecules.

In addition, there has been a recent shift towards carrying out cryo-EM on complex mixtures in contrast to traditional approaches which required extremely homogeneous samples. The computational power of image processing packages has increased to the point that it is possible to determine multiple 3D structures of different proteins from a mixture of species (Ho et al., 2019). This can be useful for studying targets which are difficult to purify to homogeneity, or for the investigation of low-affinity complexes directly from the cellular lysate which fall apart upon purification (Schmidli et al., 2019; Yi et al., 2019).

In the following sections, two key future directions are presented which hold great promise for improving the way cryo-EM is carried out, both in terms of image processing and data collection. These are the development of deep-learning algorithms, and the use of on-the-fly processing during data collection.

### 1.7.1. Deep learning algorithms

One of the most exciting directions for the field of cryo-EM currently is the development of algorithms that make use of deep learning approaches for image processing. Such algorithms are based on convolutional neural networks, which are function machines that are optimised to learn the ideal way to take input data to produce an output. This requires a training step in which the input and output are both available.

One example of this being used in the context of cryo-EM is in the development of more reliable particle picking algorithms (see section 1.5.2). A subset of a cryo-EM dataset consisting of micrographs with manually-picked particle locations is used to train a neural network to generate particle positions when given an unseen micrograph as input. These methods are more robust than cross-correlation based approaches, and have given promising results for traditionally challenging datasets (Bepler et al., 2019a).

Another application of deep learning is for the denoising of cryo-EM images. One of the largest limitations in cryo-EM is the low signal-to-noise ratio in images. Several algorithms have been developed which use neural networks to increase the signal-to-noise ratio, which is especially useful for the selection of particles and assessing the quality and homogeneity of images. However, unlike for particle picking in which a ground truth (i.e. particle locations) are available, images with a high signal-to-noise ratio are not available in cryo-EM because of electron dose constraints (Buchholz et al., 2019). Therefore, denoising algorithms make use of the Noise2Noise approach (Lehtinen et al., 2018), which facilitates image restoration from noisy images in the absence of a high signal-to-noise ratio counterpart. Such algorithms can dramatically improve the contrast in cryo-EM images, as shown in Figure 1.27. Noise2Noise approaches have now been implemented in Topaz-Denoise (Bepler et al., 2019b), Cryo-CARE (Buchholz et al., 2019), JANNI in the package SPHIRE (Wagner, 2019), and WARP (Tegunov and Cramer, 2019).

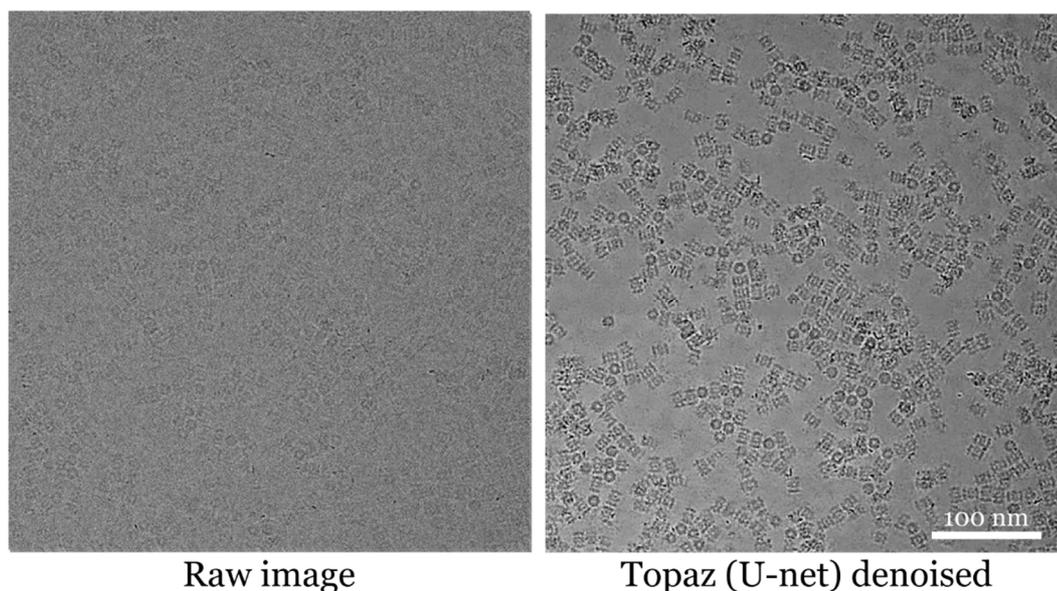


Figure 1.27. Comparison of a cryo-EM micrograph before (left) and after (right) denoising with the Topaz U-net denoising model. Adapted from Bepler et al., 2019b.

Other recent applications of neural networks to cryo-EM image processing include the identification of secondary structural elements in medium-resolution maps, as implemented in Emap2sec (Maddhuri Venkata Subramaniya et al., 2019), the CryoDRGN algorithm for modelling continuous variation in 3D reconstructions (Zhong et al., 2019), the automatic selection of 2D class averages by Cinderella in SPHIRE (Moriya et al., 2017), and the automatic classification of areas for picking on micrographs (Sanchez-garcia et al., 2019). The use of neural networks and deep learning approaches in cryo-EM is still a very new development, but holds significant promise for improving image processing methods over the next few years.

### 1.7.2. On-the-fly image processing

As cryo-EM continues to develop, both from a hardware and software point of view, there will be a further trend towards automation and high-throughput data collection (Danev et al., 2019). From a few years ago, when the number of images collected per day usually ranged from a few hundred to approximately 1,000, it is now possible to collect roughly 5,000 images per day on high-end microscopes (Danev et al., 2019). This number may double over the next few years, meaning that huge amounts of data (several terabytes) will be generated from a single day of data collection (Danev et al., 2019; Thompson et al., 2019).

There is now a push towards ‘on-the-fly’ processing of images while data collection is still running, in order to get quick feedback and avoid collecting data on sub-optimal samples. Several pipelines have now been developed for automatic on-the-fly processing, including those implemented in RELION (Fernandez-Leiro and Scheres, 2017; Zivanov et al., 2018), CryoSPARC (Punjani et al., 2017), WARP (Tegunov and Cramer, 2019) and CryoFLARE (Schenk et al., 2020). One example of such an integrated pipeline, as described by Tegunov and Cramer, 2019, is shown in Figure 1.28. Data is collected using software such as EPU or SerialEM (see section 1.4.2), then imported into WARP for automatic micrograph pre-processing and neural network-based particle picking. Particles are then extracted and continuously imported into CryoSPARC for 2D classification and 3D classification and refinement. This allows the real-time evaluation of data, and enables users to make decisions about data collection parameters during acquisition.

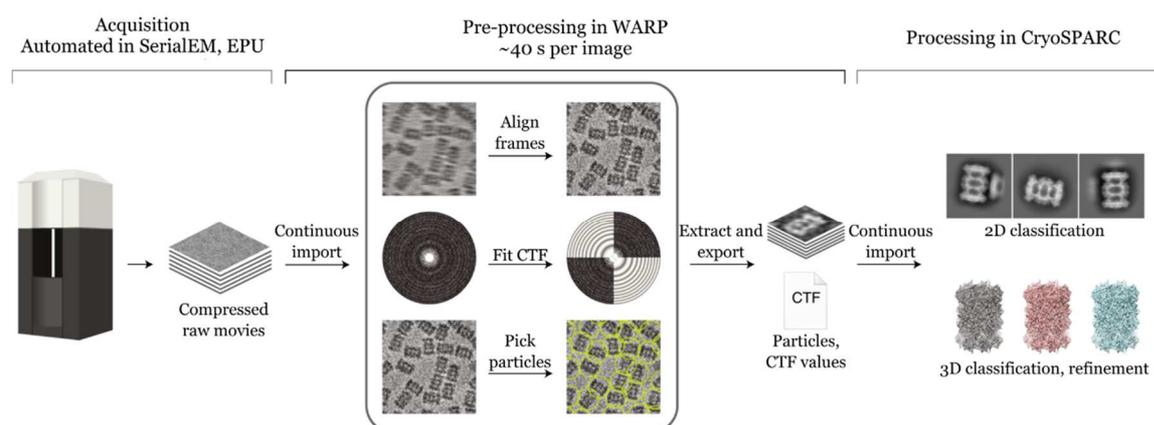


Figure 1.28. Automatic on-the-fly processing pipeline, as implemented in WARP. Adapted from Tegunov and Cramer, 2019.

The recently described platform CryoFLARE even allows for direct interfacing with the microscope data acquisition software, allowing automatic feedback-based control of the microscope during data collection (Schenk et al., 2020). CryoFLARE also allows the plotting of various statistics for each micrograph over their original position on a grid square, such as the estimated defocus or resolution, or number of picked particles, for the easy visualisation of trends in a grid square during data collection (Figure 1.29). As the speed of data collection continues to increase, the biggest challenge will be enabling the most efficient use of microscope time. The ability to gain rapid feedback on data collection as it occurs, as supported by such on-the-fly processing pipelines, will greatly aid this.

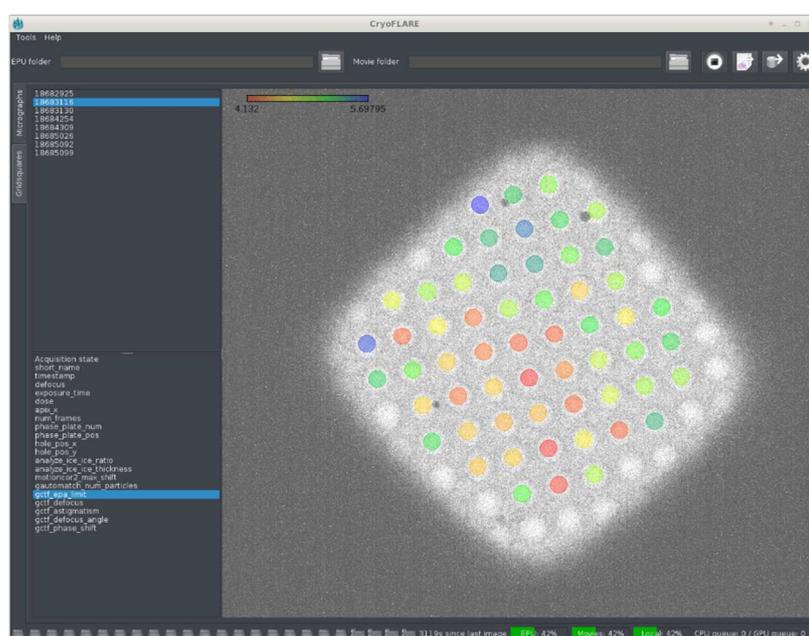


Figure 1.29. The CryoFLARE graphical user interface, showing the ability to map statistics such as the estimated resolution over the location in a grid square the micrograph was collected from. Adapted from Schenk et al., 2020.

The 2010s was a transformative decade for cryo-EM. Now firmly established as a mainstream technique, it was named ‘method of the year’ in 2015, and in 2017 the Nobel Prize in Chemistry was awarded for the development of cryo-EM. As we enter the 2020s, new technological developments will broaden the number of targets which are amenable to characterisation by cryo-EM in terms of size, flexibility, and complexity. In parallel, the move towards automation in data collection and processing, and towards user-friendly software packages will make cryo-EM more attractive to non-specialist users. With the speed at which the field is currently evolving, the 2020s will certainly be an exciting decade in which to be a microscopist.

## Chapter 2. Evolutionary history and structural features of AAA+ ATPases

RavA is a MoxR-family AAA+ ATPase from *E. coli*. AAA+ ATPases are P-loop proteins, which all share a common fold for hydrolysing nucleotides to liberate energy for powering cellular processes. P-loop proteins are classified into two divisions – the KG division, and the ASCE division. The AAA+ ATPase superfamily sits within the ASCE division along with seven other superfamilies.

This chapter will discuss the key structural motifs which are crucial to the hydrolysis of nucleotides in P-loop proteins. I will then summarise the different superfamilies of the ASCE division, to compare and contrast the different ways in which the P-loop fold is used to coordinate a variety of different cellular processes. Finally, I will present the latest work on AAA+ ATPases, and show how the advent of cryo-EM has allowed the definition of a mechanism for ATP hydrolysis amongst the different AAA+ clades. This will set the scene for the investigation of the RavA ATPase mechanism discussed in Chapter 4.

## 2.1. ATP – a universal energy currency

It has long been established that cells use the high energy stored in the phosphate bond for carrying out cellular work (Lipmann, 1941). Nucleoside triphosphates, such as adenosine triphosphate (ATP) and guanosine triphosphate (GTP) consist of a base (adenine or guanine) attached to a sugar (ribose), linked to three phosphate molecules  $\alpha$ ,  $\beta$  and  $\gamma$  (Figure 2.1). Hydrolysis of the high-energy bond between the  $\beta$  and  $\gamma$ -phosphate can be coupled to energy-requiring cellular processes such as mechanical work, the active transport of molecules against a concentration gradient, or thermodynamically unfavourable reactions (Knowles, 1980; Zala et al., 2017).

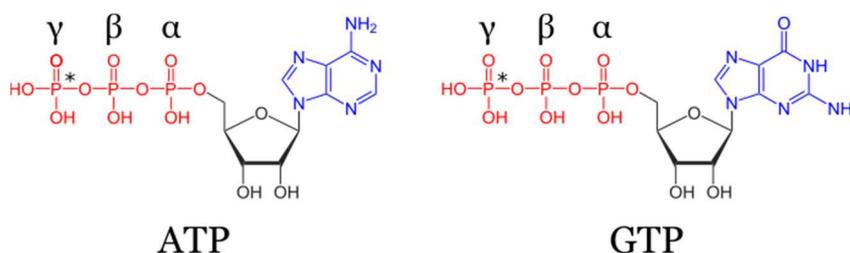


Figure 2.1. Structures of ATP and GTP. Bases (adenine and guanine) are coloured blue, ribose is coloured black and phosphates are coloured red. The  $\alpha$ ,  $\beta$  and  $\gamma$ -phosphates are labelled, and the high-energy  $\beta$ - $\gamma$  phosphate bond is indicated with an asterisk.

ATP is universally conserved amongst all living organisms, and was selected at the beginning of evolution as the major energy-carrying molecule (Plattner and Verkhatsky, 2016). ATP is responsible for providing the energy for most energy-requiring cellular processes, whereas GTP is more commonly used in signalling pathways (Plattner and Verkhatsky, 2016; Vetter and Wittinghofer, 1999; Zala et al., 2017). The central importance of ATP to cellular metabolism is reflected by the fact that ATP-hydrolysing proteins, termed ATPases, are similarly widespread throughout life, and significantly predate the last universal common ancestor of bacteria, archaea and eukaryotes (Kyrpides et al., 1999; Leipe et al., 2002). ATPases catalyse the hydrolysis of the  $\gamma$ -phosphate of ATP, forming adenosine diphosphate (ADP) and releasing energy.

## 2.2. P-loop NTPases

The ability of proteins to bind nucleotides likely evolved several times independently (Saraste et al., 1990; Vetter and Wittinghofer, 1999). Nucleotide binding proteins can be classified according to sequence motifs and structural features, which are related to the mechanism of nucleotide binding (Vetter and Wittinghofer, 1999). One of the most prevalent classes of nucleotide binding proteins is the P-loop (phosphate-binding loop) class of nucleoside triphosphatases (NTPases), which represent approximately 10-18% of total gene products in prokaryotes and eukaryotes (Koonin et al., 2000; Saraste et al., 1990). They are involved in a vast number of cellular processes, across almost all major pathways (Aravind et al., 2004).

All P-loop proteins share a common  $\alpha\beta\alpha$  fold, consisting of five parallel  $\beta$ -strands sandwiched between  $\alpha$ -helices (Leipe et al., 2003; Milner-White et al., 1991). All P-loop proteins possess the conserved Walker A (P-loop) and Walker B sequence motifs, which are responsible for coordinating the nucleotide molecule (Walker et al., 1982). The Walker A/P-loop motif is particularly conserved, and follows the consensus sequence GxxGxGK[S/T], where G = glycine, K = lysine, S = serine, T = threonine, x = any residue (Erzberger and Berger, 2006; Saraste et al., 1990; Walker et al., 1982). The lysine residue is involved in coordination of both the  $\beta$  and  $\gamma$ -phosphate atoms of the nucleotide, and is critical to nucleotide hydrolysis (Fischer and Glockshuber, 1994; Wendler et al., 2012). Similarly, the serine/threonine residue coordinates a  $Mg^{2+}$  cation which in turn interacts with both the  $\beta$  and  $\gamma$ -phosphate atoms of the bound nucleotide.

The Walker B motif, in contrast, is less conserved and can be represented as hhhh[D/E], where h = any hydrophobic residue, D= aspartate and E = glutamate. The acidic aspartate (or less commonly, glutamate) is also involved in the coordination of the  $Mg^{2+}$  ion, and is essential for ATP hydrolysis by P-loop proteins (Iyer et al., 2004a; Leipe et al., 2003; Walker et al., 1982).

The P-loop class can be split into two divisions, according on the arrangement of the five parallel  $\beta$ -strands in the catalytic  $\alpha\beta\alpha$  core (Leipe et al., 2003). One group contains kinases and GTTPases (the so-called ‘KG division), and is typified by a ‘54132’ arrangement of  $\beta$ -strands (Figure 2.2). In this arrangement, the strands containing Walker A and B motifs ( $\beta$ 1 and  $\beta$ 3 respectively) are placed adjacent to one another.

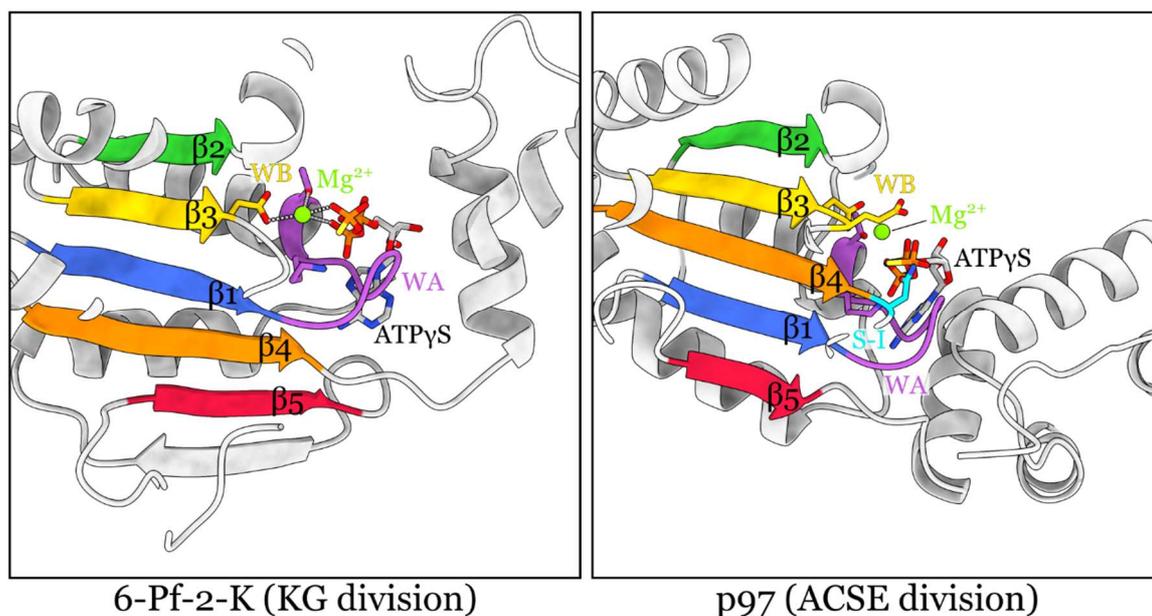


Figure 2.2. Comparison between the arrangement of  $\beta$ -strands and key ATP-binding motifs in KG (6-Pf-2-K, PDB ID: 1BIF) and ASCE (p97, PDB ID: 5C1A, N-terminal domain) P-loop divisions. WA = Walker A, WB = Walker B, S-I = sensor-I.

The second group is more diverse, and contains the AAA+ superfamily of which RavA is a member (Erzberger and Berger, 2006). This group has an inserted  $\beta$ -strand between strands  $\beta_1$  and  $\beta_3$ , such that the arrangement of  $\beta$ -strands is '51432' and the Walker A and B motifs are separated by an additional strand (Erzberger and Berger, 2006; Leipe et al., 2003) (Figure 2.2). This inserted strand  $\beta_4$  usually contains an additional nucleotide-binding motif at its C-terminal tip termed the sensor-I (S-I) motif, which consists of a polar residue (usually asparagine) (Iyer et al., 2004b; Koonin, 1993; Leipe et al., 2003). In addition, the Walker B motif is extended to include a second acidic residue, and usually takes the form hhhhDE. The extra glutamate is important for the hydrolysis of ATP, and acts together with S-I to position a water molecule near the  $\gamma$ -phosphate for nucleophilic attack during hydrolysis (Leipe et al., 2003; Wendler et al., 2012). Because of these two features, this group of proteins is termed the ASCE division (additional strand, catalytic E).

A comparison between key structural features of the nucleotide binding site of KG and ASCE division P-loop proteins is shown in Figure 2.2. 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase (6-Pf-2-K/Fru-2,6-P<sub>2</sub>ase) is a bifunctional enzyme with a kinase domain (6-Pf-2-K) that displays the 54132 arrangement of  $\beta$ -strands characteristic to the KG division (Hasemann et al., 1996). The ASCE-division p97 nucleotide binding site, in contrast, displays the rearranged 51432  $\beta$ -strand order, and the characteristic S-I and Walker B glutamate residues (Hänzelmann and Schindelin, 2016).

### 2.2.1. Arginine fingers

Another key structural feature of the NTPase site is the so-called arginine finger. In P-loop GTPases and ASCE proteins, an arginine residue coordinates the  $\gamma$ -phosphate during hydrolysis of bound nucleotides (Nagy et al., 2016; Ogura et al., 2004). For kinases, there is not a clear conservation of an arginine finger motif (Nagy et al., 2016). Most ASCE proteins form circular oligomers, where the arginine finger of one monomer interacts with the bound nucleotide of a neighbouring monomer. Therefore, oligomerisation is strictly necessary for ATPase activity (Erzberger and Berger, 2006; Ogura et al., 2004). For GTPases, an analogous arginine finger is instead provided by a GTPase-activating protein (GAP), which stimulates GTPase activity by several orders of magnitude (Kötting et al., 2008; Ogura et al., 2004). These *trans*-acting arginine fingers are well-conserved from both a structural and sequence point of view, which reflects their importance to the overall function of these proteins (Nagy et al., 2016; Neuwald et al., 1999; Ogura et al., 2004).

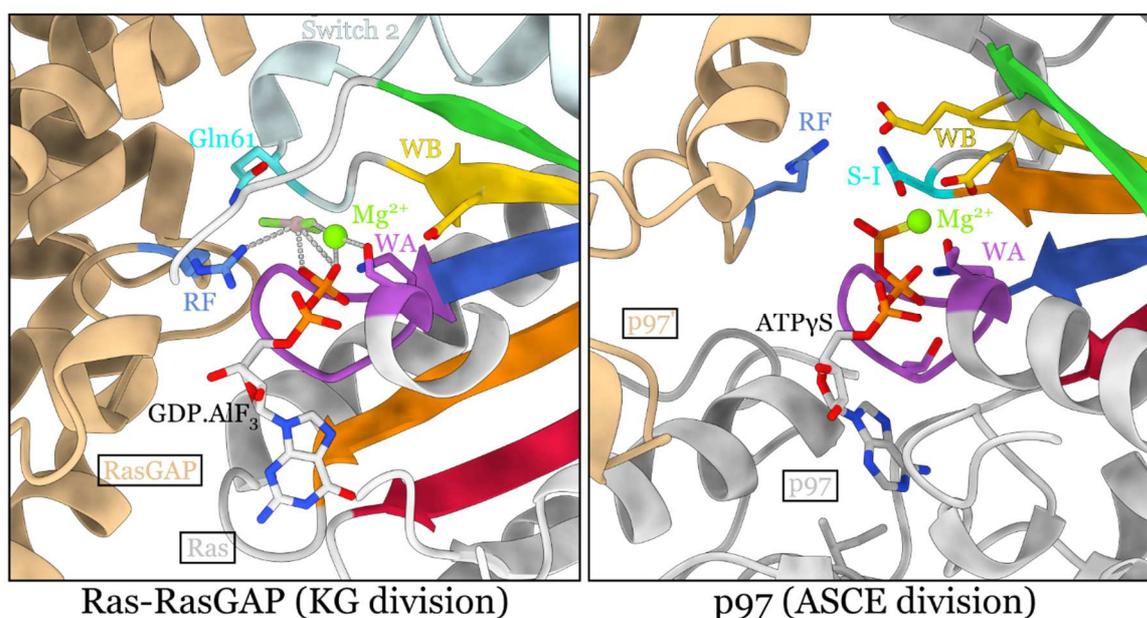


Figure 2.3. Comparison between the NTP binding sites of the KG division Ras-RasGAP (PDB ID: 1WQ1) and the ASCE division p97 (PDB ID: 5C1A, N-terminal domain).  $\beta$ -strands are coloured as in Figure 2.2, key residues and motifs are indicated. WA = Walker A, WB = Walker B, S-I = sensor-I, RF = arginine finger.

A comparison between the ATP binding site of p97 and the GTPase Ras (in complex with its GAP Ras-GAP) is shown in Figure 2.3. The P-loop structure is well-conserved between proteins of the KG and ASCE division. The respective arginine fingers (coming from a neighbouring monomer in p97 and from the GAP in Ras) occupy a similar position, though they are present in different conformations. The nucleotide site between Ras and Ras-GAP is occupied by GDP·AlF<sub>3</sub>, a structural analogue of GTP which mimics the transition state between GTP and GDP. The GAP arginine finger directly coordinates the

AlF<sub>3</sub> molecule (as shown in Figure 2.3), which mimics the position of the  $\gamma$ -phosphate of GTP during hydrolysis (Ahmadian et al., 1997; Scheffzek et al., 1997). In contrast, the nucleotide binding site of p97 is occupied by ATP $\gamma$ S, a non-hydrolysable ATP analogue. The ATP-bound state represents the inactive (or ‘pre-active’) state, and as such, the arginine finger is positioned away from the bound nucleotide (Hänzelmann and Schindelin, 2016). This reflects the fact that arginine fingers are not strictly necessary for nucleotide binding, but are crucial for nucleotide hydrolysis (Nagy et al., 2016; Ogura et al., 2004).

### 2.2.2. Divergence in sequence, conservation in structure

Although the S-I motif asparagine is present only in the ASCE division of NTPases (at a conserved position at the end of  $\beta$ 4) there is an analogous region in certain P-loop GTPases which plays the same functional role (Wendler et al., 2012). This stretch is termed the “switch 2” region (Milburn et al., 1990; Scheffzek et al., 1997). In Ras, for example, the switch 2 Gln61 is positioned in a loop after  $\beta$ 3, and similarly to S-I in ASCE ATPases, coordinates water molecule for nucleophilic attack of the  $\gamma$ -phosphate during hydrolysis (see Figure 2.3 for a comparison between p97 S-I and Ras Gln61 positions) (Resat et al., 2001a). An equivalent switch 2 glutamine residue is also present in Rho, Ran and MglA GTPase families, and is critical for GTP hydrolysis (Bos et al., 2007; Miertzschke et al., 2011; Scheffzek et al., 1997).

In GTPases such as IF-2 and EF-G, the switch 2 glutamine is replaced by a catalytic histidine which plays an equivalent role (Anand et al., 2013). However, not all P-loop GTPases possess a catalytic switch 2 residue. The Ras-like GTPase Rap1 instead has a glutamine to threonine mutation at this position, and a catalytic asparagine residue (termed an ‘asparagine thumb’) is instead provided *in trans* from Rap1GAP in an very similar orientation and position, as shown in Figure 2.4. (Daumke et al., 2004; Scrima et al., 2008). This elegantly illustrates the principle that the 3D structure of proteins are often more conserved than sequence. The catalytic Asn348 of p97, Gln61 of Ras and Asn290 of Rap1GAP all play the same functional role, conserving the mechanism of nucleotide hydrolysis, yet have evolutionarily distinct origins.

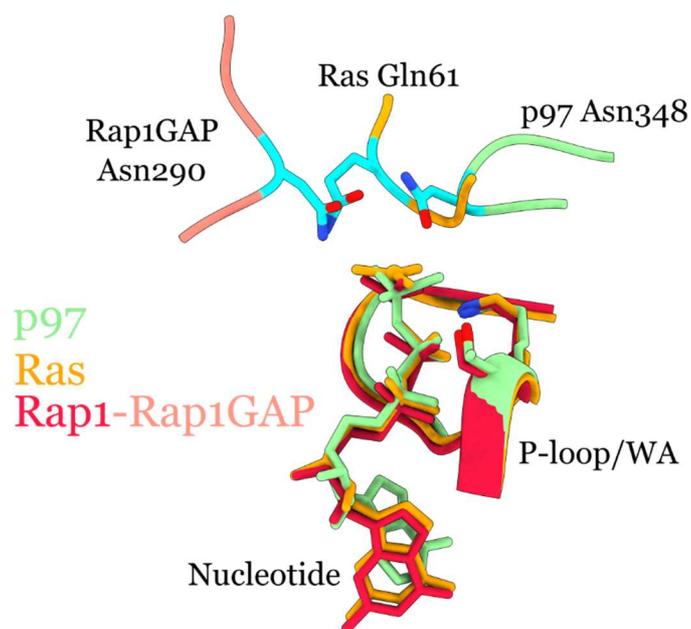


Figure 2.4. Comparison between the S-I motif of the ASCE protein p97 (PDB ID: 5C1A), the switch 2 glutamine of the KG-division Ras (PDB ID: 1WQ1) and Rap1GAP asparagine Asn290 (PDB ID: 3BRW). The P-loop/Walker A motif and bound nucleotide of each are shown.

### 2.3. ASCE division ATPases

ASCE proteins are involved in a wide variety of cellular processes, including nutrient transport, chromosome segregation, viral genome packaging, and protein degradation (Erzberger and Berger, 2006; Iyer et al., 2004a). All ASCE proteins share the same ancestral  $\alpha\beta$  core, consisting of five  $\beta$ -sheets and surrounded by four  $\alpha$ -helices. The ASCE division can be further classified into different superfamilies, according to variations from this conserved core (Erzberger and Berger, 2006). A summary of these groups and their accompanying structural features is shown in Figure 2.5.

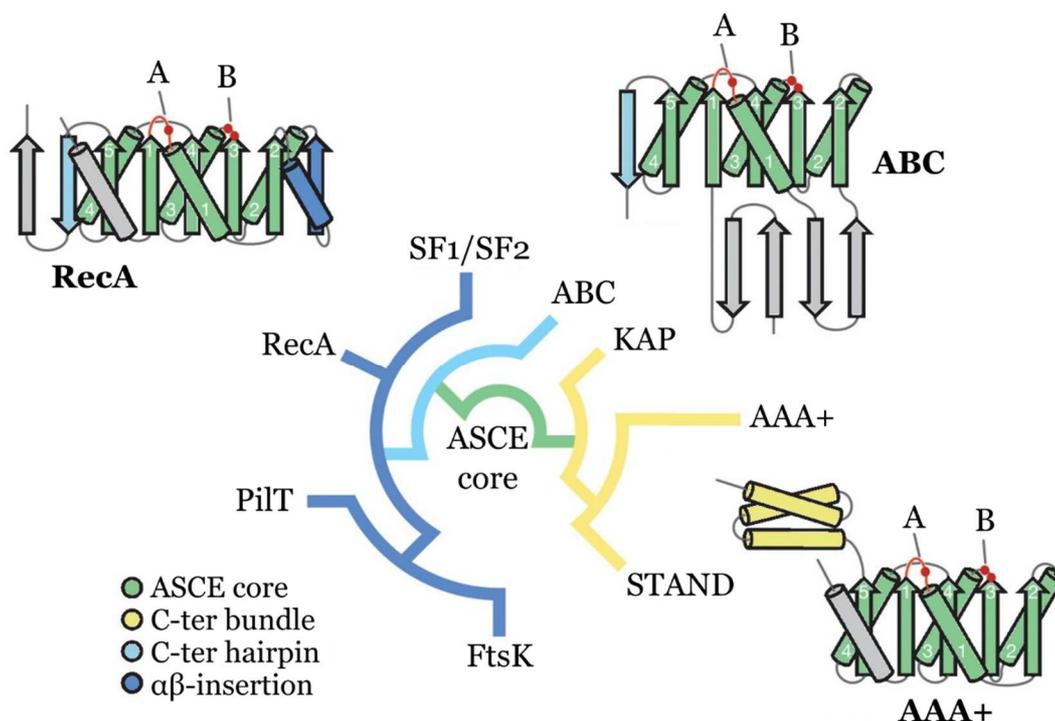


Figure 2.5. Comparison between structural features of ASCE superfamilies. Adapted from Erzberger and Berger, 2006.

The ABC superfamily is characterised by a C-terminal  $\beta$  strand which is inserted to the ancestral ASCE core. The RecA, PilT, FtsK, and SF1/SF2 superfamilies share this insertion and also possess an  $\alpha\beta$ -insertion between  $\beta$ -strand 2 and  $\alpha$ -helix 2. In contrast, the KAP, AAA+ and STAND superfamilies share a C-terminal  $\alpha$ -helical bundle. In AAA+ proteins in particular, this C-terminal bundle plays a distinct role in the ATPase mechanism by contributing to the ATP binding site (see section 2.4).

ASCE proteins usually form cyclic oligomeric rings, most commonly hexameric (Erzberger and Berger, 2006; Lupas and Martin, 2002). There are a few exceptions, such as the ABC superfamily, which usually form homodimers, and SF1/SF2 helicases which

have an internal duplication of the ATPase domain, and are active as monomers or dimers (Davidson et al., 2008; Erzberger and Berger, 2006; Mackintosh and Raney, 2006).

There is substantial functional diversity amongst ASCE proteins. Several superfamilies primarily interact with DNA, in both prokaryotes and eukaryotes. The RecA superfamily is involved in DNA recombination and repair, and RecA members often form helical scaffolds on DNA (Cox, 2007). The SF1/SF2 superfamily do not form hexameric rings, and contains helicases from helicase superfamilies 1 and 2 that unwind and remodel DNA and RNA. The SF2 Swi/Snf helicases, for example, are involved in the remodelling of nucleosome DNA in eukaryotes to modulate transcription (Fairman-Williams et al., 2010).

The FtsK superfamily also contains DNA-interacting translocases and helicases (Blesa et al., 2017; Iyer et al., 2004a). In bacteria, the FtsK protein is recruited by FtsZ, a cell division protein that forms the Z-ring around the dividing cell. FtsK is a DNA translocase, and aids the segregation of chromosomes into daughter cell by pumping DNA through the central pore, powered by ATP hydrolysis (Iyer et al., 2004a; Jean et al., 2019). The cryo-EM structure of the FtsK hexamer in complex with DNA is shown in Figure 2.6, illustrating the general architecture of many ASCE ATPases – a cyclic hexamer surrounding substrate in the central pore.

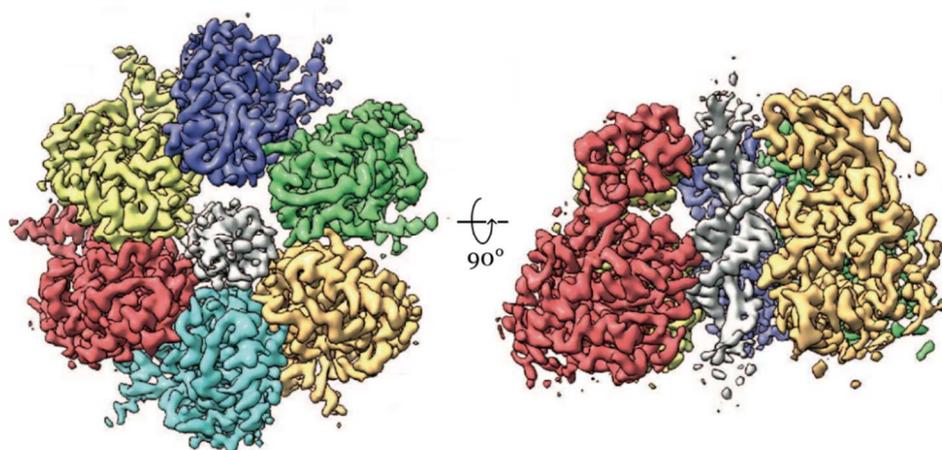


Figure 2.6. Cryo-EM structure of FtsK in complex with DNA viewed from above (left) and the side (right, cut-away view). Individual monomers of the FtsK hexamer are coloured differently, and the bound DNA in the central pore is coloured grey. Adapted from Jean et al., 2019.

The PilT family contains motor proteins which are important for the formation of type IV pilin-like systems in bacteria and archaea (Berry and Pelicic, 2015). Type IV pilins include the type II secretion system and the archaellum, and are often involved in motility or cell attachment to surfaces (Denise et al., 2019). They are composed of repeating pilin subunits, which are polymerised into pili by PilT-like ATPases. Like most other ASCE proteins, PilT members form hexamers and couple the hydrolysis of ATP to mechanical work, in this case promoting the assembly or disassembly of pili (McCallum et al., 2019).

The ABC superfamily is a diverse group of ATPases that transport substrates across membranes (Davidson et al., 2008; Dean and Annilo, 2005). Unlike most other ASCE members, they form membrane-bound homodimers which couple the hydrolysis of ATP to the transport of various molecules, including amino acids, ions, and vitamins. They include multidrug efflux pumps, such as P-glycoprotein, and the chloride channel CFTR, which is mutated in cystic fibrosis (Gadsby et al., 2006; Lin and Yamazaki, 2003).

The KAP, AAA+ and STAND superfamilies all have an inserted C-terminal helical bundle to the ASCE core. The KAP superfamily, named after the proteins Kidins220/ARM5 and PifA, is a small and mostly uncharacterised family of ASCE proteins (Aravind et al., 2004). Most KAP proteins are membrane-bound, similar to ABC transporters, although they can contain up to four transmembrane helices inserted in the nucleotide-binding domain which is unique (Aravind et al., 2004). Although there is currently no structural information on KAP superfamily members, a few have been functionally characterised. The KAP protein Kidins220/ARMS is present in animals and acts as a scaffold protein for neuronal differentiation and development of the vascular system (Bracale et al., 2007; Cai et al., 2017). However, some KAP members appear to have a catalytically inactive ATPase domain, and may play their cellular roles through accessory domains (Aravind et al., 2004).

The STAND (signal transduction ATPases with numerous domains) superfamily is spread across eukaryotes and prokaryotes, and as well as the core ATPase domain have many additional domains (Leipe et al., 2004). These domains range from protein interaction and scaffolding domains to enzymatic domains, and confer diverse functions and oligomerisation properties to STAND-superfamily proteins. Unlike other many other ASCE division proteins, the hydrolysis of ATP by STAND proteins is not linked to mechanical work, but rather to cellular signalling (Marquet and Richet, 2010). For example, the mammalian STAND protein Apaf1 forms the heptameric apoptosome responsible for initiating cellular apoptosis (Leipe et al., 2004; Zhou et al., 2015a). STAND proteins are also involved in immune responses in both plants and animals, through the resistosome and inflammasome respectively. The resistosome is formed by the STAND protein ZAR1, and is pentameric, while the 11-fold symmetric inflammasome is formed by the STAND proteins NLRC4 and NAIP2. The binding of ATP promotes oligomerisation and activation of STAND proteins, whereas the inactive ADP-bound form is monomeric (Marquet and Richet, 2007, 2010; Zhang et al., 2015; Zhou et al., 2015a). Structures of the apoptosome, resistosome and inflammasome are shown in Figure 2.7, highlighting the oligomeric and domain diversity of this unique ASCE superfamily.

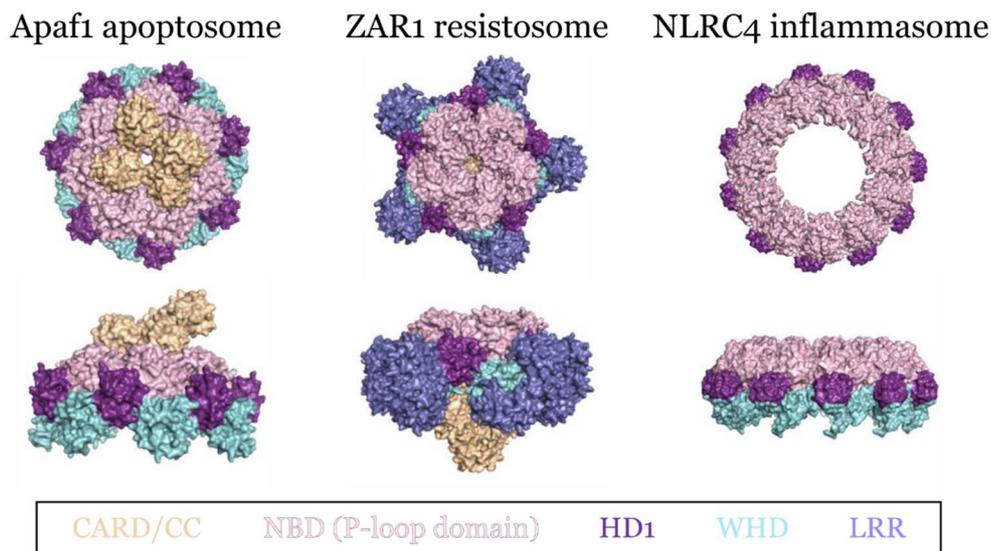


Figure 2.7. Comparison between surface view structures of the Apaf1 apoptosome, ZAR1 resistosome and NLRC4 inflammasome. CARD = caspase activation and recruitment domain (ZAR1 and NLRC4), CC = coiled-coiled domain. NBD = nucleotide-binding domain, containing the ASCE core. HD1 = helical domain 1, WHD = winged helix domain, LRR = leucine-rich repeat domain. Adapted from Wang et al., 2019.

Finally, the AAA+ superfamily is one of the largest and most diverse superfamilies of ASCE ATPases (Ammelburg et al., 2006). These proteins most commonly form hexameric rings that couple the hydrolysis of ATP to mechanical work, and have been extensively studied from both a functional and structural point of view. An overview of the AAA+ superfamily is presented in the next section.

## 2.4. The AAA+ superfamily

RavA is situated in the AAA+ superfamily (ATPases associated with a variety of cellular activities) of ASCE proteins. Originally defined as the AAA family in 1993 based on sequence conservation between putative ATPases, the family was expanded to the AAA+ family after structural studies revealed that conserved features of the AAA family were significantly more widespread than initially thought (Kunau et al., 1993; Neuwald et al., 1999). Most eukaryotes have between 50 and 80 genes coding for AAA+ proteins, and significant diversity in the AAA+ superfamily already existed at the time of the last universal common ancestor of prokaryotes and eukaryotes (Snider et al., 2008).

In addition to the  $\alpha\beta\alpha$  ASCE core, AAA+ ATPases possess a small C-terminal  $\alpha$ -helical bundle (“all- $\alpha$ ” insertion), as shown in Figure 2.5 (Neuwald et al., 1999). The  $\alpha\beta\alpha$  and all- $\alpha$  regions are sometimes referred to as the ‘large AAA+’ and ‘small AAA+’ subdomains respectively. An example is shown in Figure 2.8

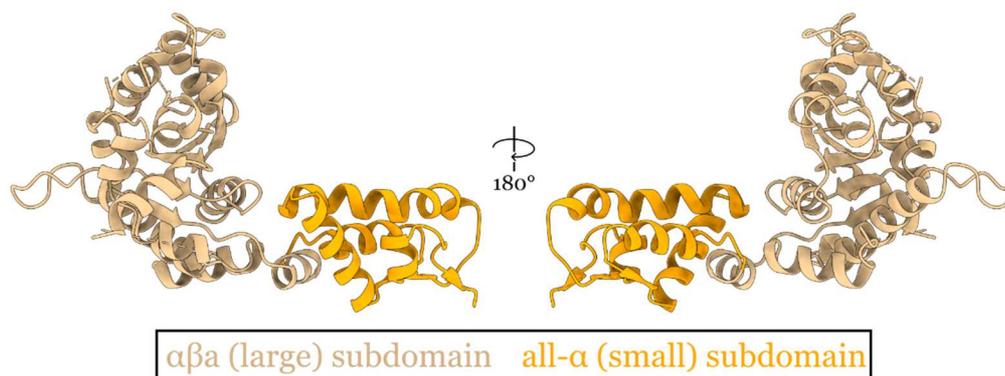


Figure 2.8. Structure of ClpX (PDB ID: 3HWs) showing the canonical arrangement of the  $\alpha\beta\alpha$  (brown) and all- $\alpha$  (orange) subdomains in AAA+ ATPases.

AAA+ ATPases form oligomeric rings, usually hexameric, and the nucleotide binding site sits at the interface between the large and small AAA+ subdomains, with an arginine finger being provided *in trans* from a neighbouring monomer (Sysoeva, 2017; Wendler et al., 2012). The angle between the large and small subdomains changes during ATP hydrolysis, and this conformational change is often linked to mechanical work (Glynn et al., 2009; Miller and Enemark, 2016). Some AAA+ proteins also have an internal duplication of the AAA+ domain, and form hexamers with two stacked ATPase rings (Puchades et al., 2020; Sysoeva, 2017). In addition to the Walker A/B, arginine finger and S-I motifs, most AAA+ ATPases also have a conserved arginine residue termed the sensor-II motif (S-II) in the all- $\alpha$  domain which coordinates the ATP  $\gamma$ -phosphate. Many AAA+ proteins also have inserted N-terminal or C-terminal domains, and these domains are often involved in the interactions with partners and can confer substrate specificity (Erzberger and Berger, 2006; Michalska et al., 2018; Puchades et al., 2020; Zehr et al., 2017).

## 2.5. AAA+ ATPase clades, and defining an ATPase mechanism

Despite the wide diversity of function amongst AAA+ ATPases, the structure of AAA+ proteins is well-conserved. They share a very similar structural core, and small insertions in this core during their evolutionary history have defined subgroups of AAA+ proteins termed clades (Erzberger and Berger, 2006; Iyer et al., 2004b).

There are seven AAA+ clades, as defined by Erzberger and Berger, 2006 and Iyer et al., 2004b, which were based on sequence analysis and some available crystal structures at the time. The key structural features of these clades are summarised in Figure 2.9. Clades 1-3, termed the clamp loader clade, initiator clade, and classical clade respectively, have only small additions to the ancestral AAA+ core. Clades 4-7 are grouped together in the pre-sensor I (PS-I) insert superclade, and share an inserted  $\beta$ -hairpin before the S-I motif. These clades are termed the superfamily III helicase clade, HCLR clade, H2-insert clade and PS-II insert clade respectively.

AAA+ ATPases interact with both protein and nucleic acid substrates, often as a function of which clade they belong to. For example, members of clades 1, 2 and 4 interact with DNA, whereas clade 3 and 5 members usually interact with protein (Erzberger and Berger, 2006; Puchades et al., 2020). RavA is a member of the PS-II insert clade 7, which is comparatively less studied than other clades and contains members with a wide functional diversity (Erzberger and Berger, 2006). Because of this, the general applicability of mechanistic principles revealed for other AAA+ clades to RavA was uncertain, and required investigation.

Significant work has been carried out over the past two decades to investigate the mechanism of AAA+ ATPase function (Arias-Palomo et al., 2019; Puchades et al., 2020). In that time, it has emerged that there are two basic types of AAA+ proteins – switches and motors (Erzberger and Berger, 2006; Kelch, 2016). For members of clades 1 and 2, one cycle of nucleotide binding, hydrolysis and release is coupled with a single event, and these ATPases are therefore referred to as molecular switches. Clade 1 and 2 action triggers the assembly of complexes involved with DNA replication, as discussed in sections 2.6.1 and 2.6.2. This contrasts with members of other clades that employ continuous cycles of ATP hydrolysis which are coupled to mechanical action. This action is often the continuous threading of substrate through the central hexameric pore (Puchades et al., 2020).

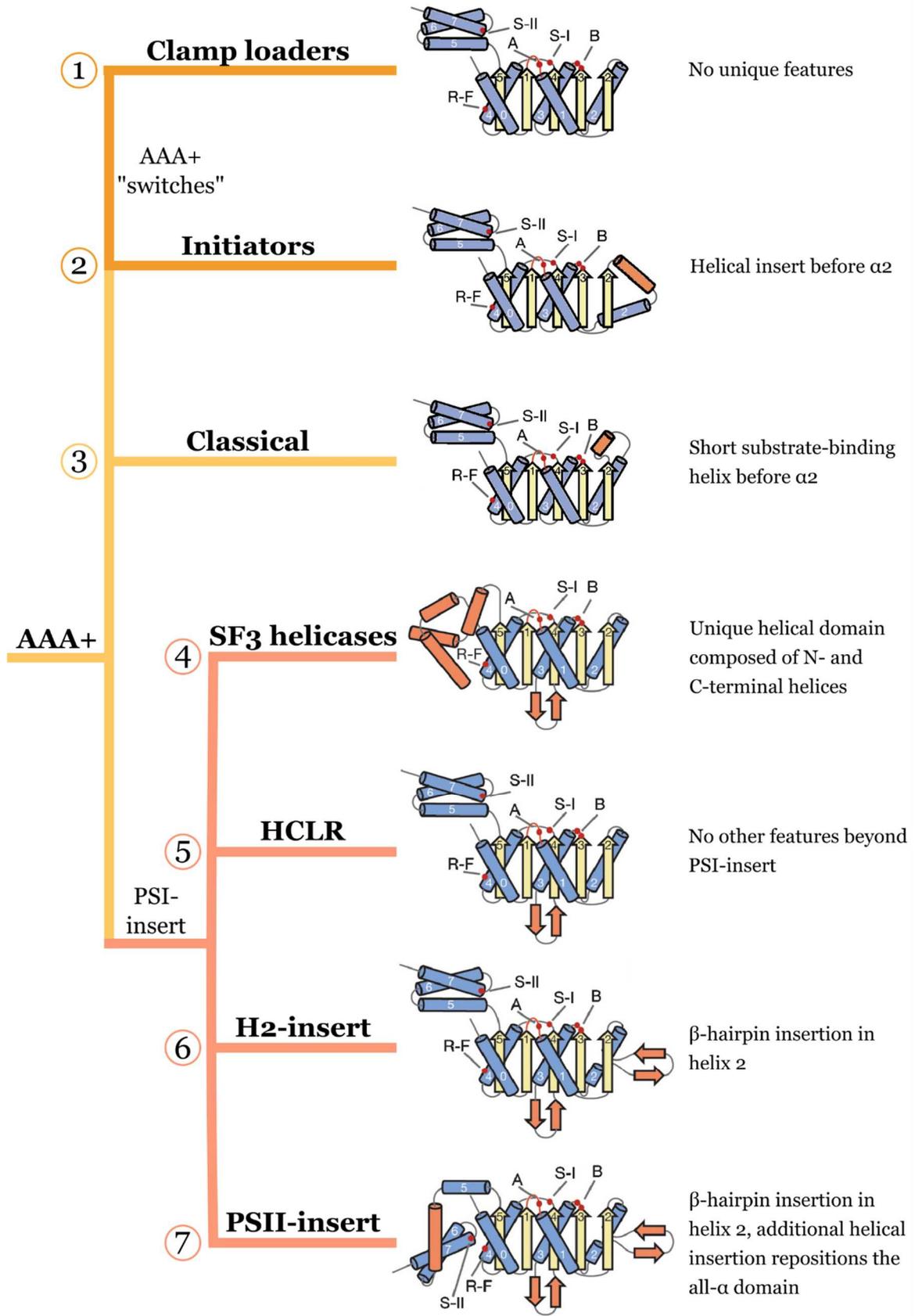


Figure 2.9. Defining structural features of the seven AAA+ clades. A = Walker A, B = Walker B, S-I = sensor-I, S-II = sensor-II, R-F = arginine finger. Adapted from Erzberger and Berger, 2006.

Before 2016, the year I started working on my thesis, the mechanism of AAA+ ATPase action was poorly understood, and it was unclear if the model of substrate translocation through a central pore applied to all AAA+ ATPases (Blok et al., 2015). The prevailing consensus was that the vast majority of AAA+ ATPases from clades 3-7 formed six-fold symmetric closed rings, although there were a few isolated exceptions (Erzberger and Berger, 2006; Lander et al., 2012; Miller and Enemark, 2016).

In 2015 and 2016, the revolution in cryo-EM resulted in a leap forwards in the understanding of the AAA+ mechanism. Previously, most structures solved by cryo-EM had a six-fold rotational symmetry applied during 3D reconstruction. With the advent of direct detectors and improved image processing software (see sections 1.4 and 1.5), several structures were published without the imposition of symmetry that showed a spiral arrangement of AAA+ domains around a central pore for the first time, as shown in Figure 2.10 (Huang et al., 2016; Yokom et al., 2016; Zhao et al., 2015). The formation of a spiral is often linked to the nucleotide state – the flat ring state of VAT shown in Figure 2.10 is with ATP $\gamma$ S bound, while the spiral ring is formed in the ADP-bound state. Although this spiral architecture was described as “surprising” and “unexpected” at the time (Chang et al., 2017; Zehr et al., 2017), many more spiral structures quickly followed which had direct implications for defining the AAA+ ATPase mechanism (Gates et al., 2017; Monroe et al., 2017; Puchades et al., 2017; Ripstein et al., 2017; Su et al., 2017; Zehr et al., 2017).

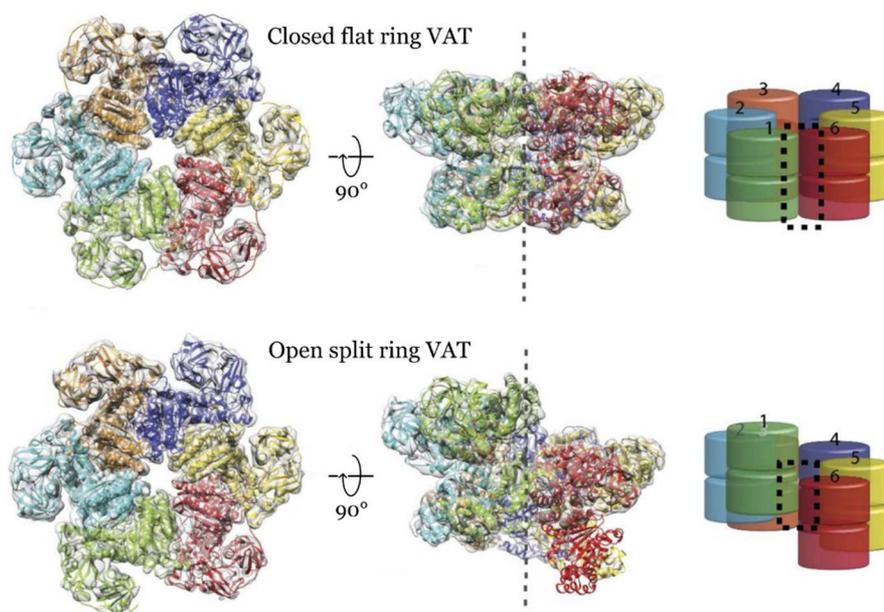


Figure 2.10. The 2016 cryo-EM structure of VAT in closed flat ring (top) and spiral split ring (bottom) conformations. For both conformations, the left image shows a top view, looking down the hexameric pore. The middle image shows a side view, where the two conformations can clearly be distinguished. The right column shows a diagrammatic representation of the two forms, with a dotted box around the interface where the spiral seam opens. Adapted from Huang et al., 2016.

Since then the recent wealth of structural information on AAA+ motors, combined with supporting biochemical data, has allowed the definition of an ATPase mechanism based on the sequential hydrolysis of ATP around the hexameric ring (Puchades et al., 2020). This proposed mechanism is summarised in Figure 2.11. In the ATP-bound state, AAA+ subunits interact tightly with the substrate, either protein or DNA, via loops protruding into the central pore of the hexameric ring. The substrate is translocated through the pore by a ‘hand-over-hand’ motion – during the ATPase cycle, individual AAA+ subunits move downwards in the hexameric spiral, pulling the substrate with them. When a subunit occupies the bottom position of the spiral staircase, they hydrolyse ATP to ADP and disengage with the substrate. Release of ADP induces a conformational rearrangement, and the subunit moves back to the top of the spiral. Subsequent ATP binding induces an interaction with the substrate again, and the cycle repeats. In this way, the sequential movement of substrate-bound pore loops down the spiral, powered by ATP hydrolysis, enables the translocation of substrates through the central hexamer pore.

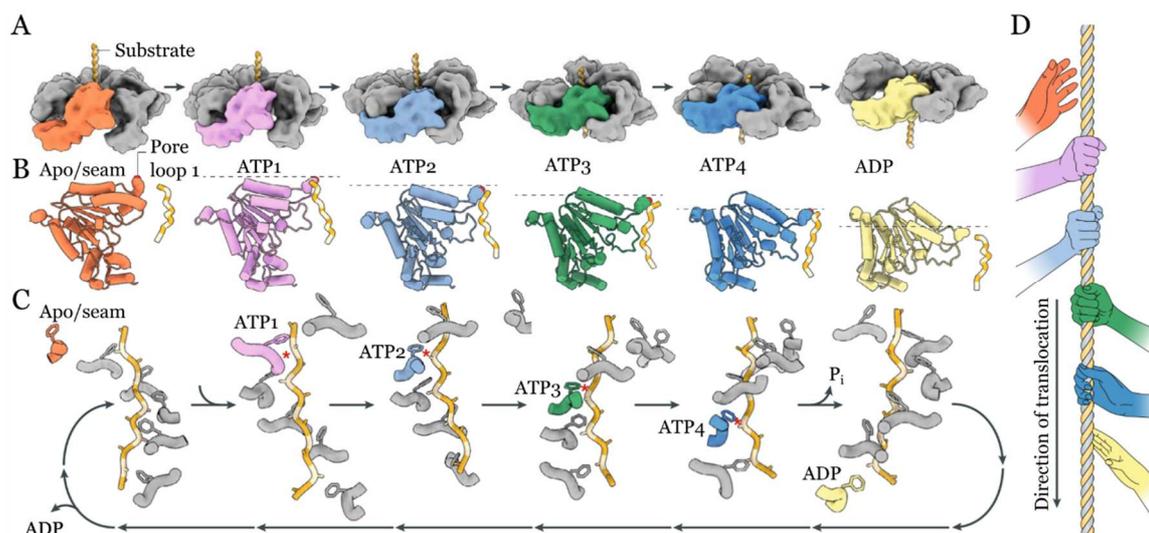


Figure 2.11. Summary of the rotary ‘hand-over-hand’ ATPase mechanism, adapted from (Puchades et al., 2020). **A)** Side view of a spiral AAA+ ATPase, with different colours for subunits occupying different positions in the spiral staircase. **B)** and **C)** Conformation of the AAA+ domain and pore loops in each subunit, with the corresponding nucleotide state of the subunit labelled. The monomer at the top of the spiral is not bound to the substrate, and has no bound nucleotide. Subsequent monomers are bound to ATP and contact the substrate during translocation. The final monomer in the ring is in the post-ATP hydrolysis state and has released the substrate. Release of ADP resets the cycle, and the monomer moves back to the top of the spiral ring. **D)** Cartoon of the ‘hand-over-hand’ mechanism, with each hand representing a pore loop and coloured as in B).

However, the vast majority of AAA+ ATPase cryo-EM structures, especially those with bound protein substrates, have been for members of clade 3 (Puchades et al., 2020). As such, there are still questions about whether a sequential rotary ATPase mechanism is universal, or whether there are other mechanisms of ATP hydrolysis specific to certain clades. One example of this uncertainty is typified by the clade 5 ATPase ClpX. Several studies have suggested that ATP hydrolysis can occur in a stochastic or probabilistic

manner, with non-adjacent nucleotide binding sites able to hydrolyse ATP rather than in a strictly sequential rotary manner (Fei et al., 2020; Glynn et al., 2009; Stinson et al., 2013). Others instead propose that the rotary model is also applicable for ClpX (Ripstein et al., 2020). Because of the limited number of studies on AAA+ ATPases outside of the classical clade, further work needs to be done to understand the general applicability of the sequential ‘around the ring’ ATPase mechanism.

In addition, a very recent publication on the yeast AAA+ ATPase Abo1, which appears to be a member of clade 3 (Zou et al., 2007), provides direct evidence to support a stochastic ATPase mechanism. AFM is used to show that switching between a spiral and closed ring occurs randomly around the ring, contradicting a strictly sequential mechanism (Cho et al., 2019). Therefore despite the recent progress in defining the AAA+ ATPase mechanism, which has particularly been enabled by the recent improvement in cryo-EM, it is clear that there are still unanswered questions about the universality of the hand-over-hand mechanism.

The structural and functional details of the seven different AAA+ clades are presented in the following section, along with pertinent studies that help to define the ATPase mechanism. While studying the recent literature on AAA+ ATPases for writing this chapter I made some structural comparisons between members of different clades. While doing so, I found some apparent inconsistencies in the classification of some ATPases as being members of clade 5 – a few members appear to possess the H2-insert that define clades 6 and 7. In addition, some members of clade 7 appear to lack the clade-defining PSII-insert and appear to be structurally more similar to clade 6 members. As discussed below, perhaps there is now a need to revisit these early classification schemes of AAA+ ATPases in light of the wealth of recent structural information.

## 2.6. Structure and function of the AAA+ clades

### 2.6.1. Clade 1 – Clamp loaders

Clade 1 ATPases have no additional features to the basic ancestral core, and are key regulators of DNA replication (Bowman et al., 2005; Erzberger and Berger, 2006; Iyer et al., 2004b). Clamp loaders are responsible for positioning the DNA clamp, a ring-shaped protein complex that slides along DNA and keeps the DNA polymerase tethered during replication (Marzahn et al., 2014). This improves the processivity of replication – unbinding of DNA polymerase is reduced, which is especially necessary during the non-continuous replication of the lagging strand (Bowman et al., 2005; Miller and Enemark, 2016). Clamp loaders, in addition to the core AAA+ domain (made up of the  $\alpha\beta\alpha$  and all- $\alpha$  subdomains), have a C-terminal oligomerisation domain which forms a collar. As a consequence, the mode of oligomerisation is distinct to that of other clades, and clade 1 AAA+ ATPases form pentameric rings rather than hexamers (Figure 2.12) (Bowman et al., 2005; Marzahn et al., 2014).

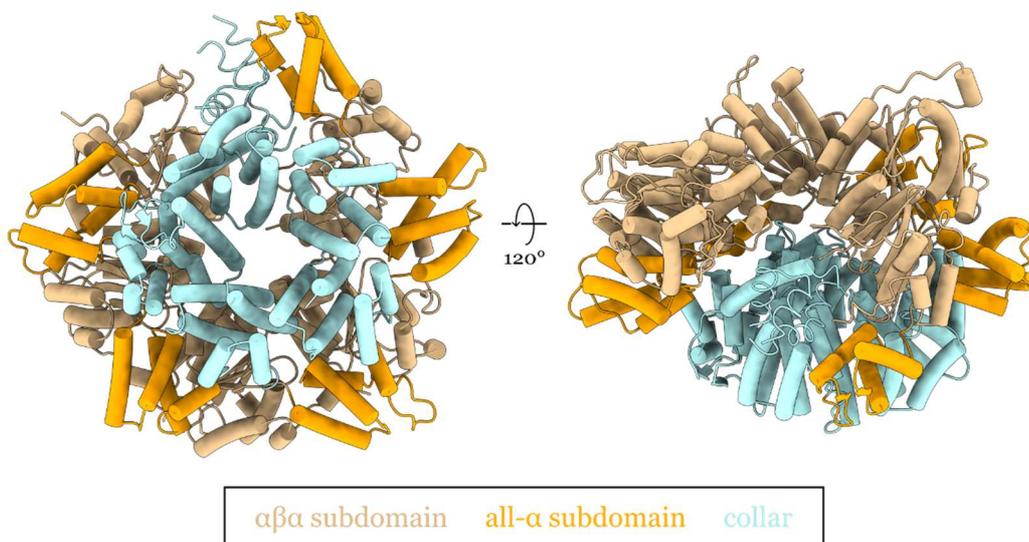


Figure 2.12. Structure the eukaryotic clamp loader RFC from *Saccharomyces cerevisiae* (PDB ID: 1SXJ) shown as cartoons, viewed from the bottom (left) and side (right). The  $\alpha\beta\alpha$  subdomains are coloured in light brown, the all- $\alpha$  subdomains are coloured orange, and the C-terminal oligomerisation collar is coloured in light blue.

One example of a clade 1 clamp loader is Replication Factor C (RFC), which is made up of the five AAA+ subunits RFC1–RFC5 in eukaryotes. The prokaryotic equivalent of RFC is the  $\gamma$  complex of DNA polymerase III, which also forms a heteropentameric ring (Jeruzalmi et al., 2001). RFC recruits the ring-shaped DNA clamp PCNA which aids processive DNA replication by DNA polymerases  $\delta$  and  $\epsilon$  (Resat et al., 2001b). The

mechanism of clamp-loader action is summarised in Figure 2.13. When RFC binds to ATP, the AAA+ domains arrange to form a spiral, with a gap between the first and fifth subunits in the ring. The clamp loader then binds to the PCNA ring and holds it in an open conformation to allow DNA binding by the RFC-PCNA complex (Kelch, 2016; Kelch et al., 2011). Binding to DNA triggers ATP hydrolysis by RFC, by correctly positioning the Walker B catalytic glutamate and the arginine finger at nucleotide binding site (Kelch, 2016; Kelch et al., 2011). This results in closing of the PCNA ring around the DNA strand and ejection of the clamp loader, allowing DNA polymerase to bind and replication to occur (Arias-Palomo et al., 2019; Bowman et al., 2004; Kelch, 2016; Yao and O'Donnell, 2012).

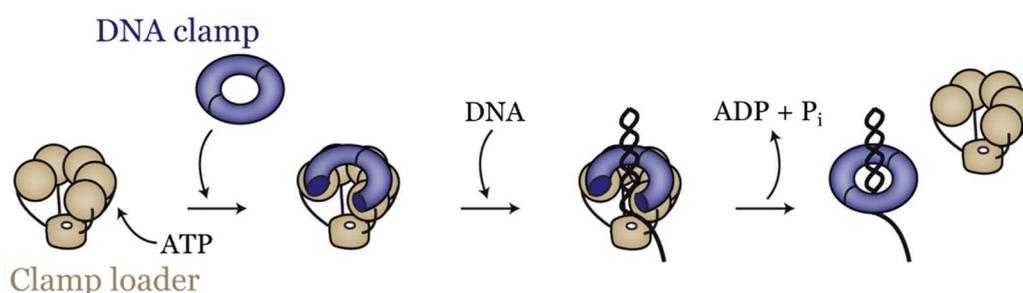


Figure 2.13. Mechanism of clamp loader action. Adapted from Arias-Palomo et al., 2019

The ATP hydrolysis by clamp loaders has been proposed to occur sequentially from one end of the pentameric ring to the other, similarly to the proposed ‘around-the-ring’ ATP hydrolysis mechanism for other AAA+ ATPases (Kelch, 2016; Puchades et al., 2020). However, the mechanical force generated by ATP hydrolysis results in the closing of the clamp on the DNA strand and ejection of the clamp loader, rather than translocation of substrates through a central pore in the AAA+ ring. Clamp loaders therefore act as a ‘switch’ rather than a ‘motor’ – ATP hydrolysis results in dissociation from both the DNA clamp and the DNA strand, and the next hydrolysis cycle requires recruitment of new substrates instead of continuous action on a single substrate.

### 2.6.2. Clade 2 – Initiators

Members of clade 2 possess an  $\alpha$ -helical insertion between the canonical  $\beta 2$  and  $\alpha 2$  of the AAA+ core, termed an initiator-specific motif (ISM) (Figure 2.9). The ISM is usually involved in binding to the DNA strand, and similarly to clade 1 AAA+ ATPases, members of clade 2 are involved with DNA replication, in this case at the level of initiation (Erzberger and Berger, 2006; Iyer et al., 2004b). They recognise origins of replication and aid the formation of the replication complex by loading a helicase onto the DNA strand, analogously to the clamp-loading activity of clade 1 AAA+ ATPases (Iyer et al., 2004b; Miller and Enemark, 2016). Two helicases are loaded onto the origin of replication, which then unwind the DNA strand in both directions allowing replication to occur (Arias-Palomo et al., 2019). There are two key families in clade 2 – the bacterial DnaA/DnaC family, and the Orc/Cdc6 family which are present in both eukaryotes and archaea (Erzberger and Berger, 2006; Iyer et al., 2004b).

In bacteria, the clade 2 ATPases DnaA and DnaC work together to coordinate replication initiation by loading the helicase DnaB, a RecA-type ASCE protein (see section 2.3), onto the DNA strand (Kaguni, 2014). DnaA and DnaC share a common ancestor, but

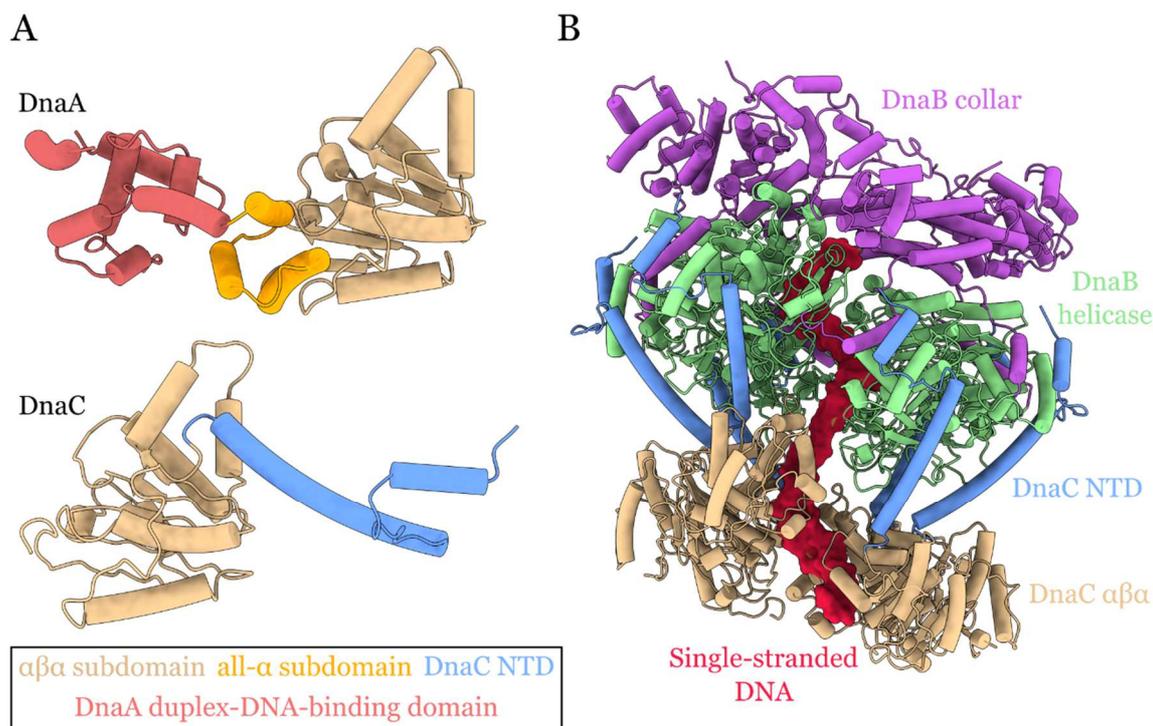


Figure 2.14. **A)** Structures of DnaA from *Aquifex aeolicus* (top, PDB ID: 3R8F) and DnaC from *E. coli* (bottom, PDB ID: 6QEM) shown as cartoons, with domains coloured as indicated. The all- $\alpha$  subdomain and C-terminal DNA-binding domain are only found in DnaA, whereas the N-terminal DnaB-binding domain is only found in DnaC. **B)** Structure of the *E. coli* DnaBC complex bound to single-stranded DNA (PDB ID: 6QEM). DnaB and DnaC spiral rings are coloured in light brown/light blue and light green/violet respectively, while DNA is coloured red.

have diverged in both structure and function, and both contain inserted domains that confer additional functionality (Arias-Palomo et al., 2019; Duderstadt et al., 2011; Kaguni, 2014; Mott et al., 2008). DnaC has lost the C-terminal all- $\alpha$  subdomain which is common to most AAA+ ATPases, and has gained an N-terminal domain (NTD) which interacts with DnaB (Mott et al., 2008). In contrast, DnaA has gained a C-terminal duplex DNA-binding domain (Duderstadt et al., 2011). A comparison between the structure of DnaA and DnaC is shown in Figure 2.14A, highlighting the conservation of the  $\alpha\beta\alpha$ -subdomain and showing the additional inserted domains.

DnaA forms long spiral filaments, and first binds to DNA at the origin of replication and induces local opening of the DNA double strand to form single-stranded DNA (Miller and Enemark, 2016). DnaC forms an open spiral hexamer when bound to ATP. This open spiral recruits the helicase DnaB and induces an extended open spiral conformation in DnaB (Itsathitphaisarn et al., 2012). DnaC loads DnaB onto the single-stranded DNA at the origin of replication, and ATP hydrolysis by DnaC conformationally rearranges DnaB to an active closed spiral around the DNA strand for DNA unwinding (Arias-Palomo et al., 2019). The complex between DnaC, DnaB and single stranded DNA is shown in Figure 2.14B, indicating how the spiral DnaC reaches up to interact with the helicase and stabilises its spiral form.

Unlike the clade 1 clamp loaders, DnaC remains bound to the DNA strand after ATP hydrolysis. The mechanism of helicase loading by DnaC is summarised in Figure 2.15. DnaB then further recruits the clamp loader complex of DNA polymerase III, a clade 1 AAA+ ATPase (Itsathitphaisarn et al., 2012). In this way, both clade 1 and clade 2 bacterial AAA+ ATPases coordinate the initiation of DNA replication.

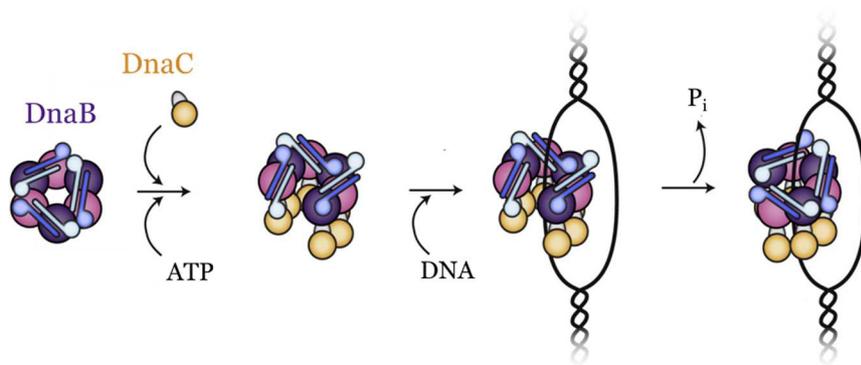


Figure 2.15. Simplified mechanism of DnaB loading by the initiator AAA+ ATPase DnaC. Adapted from Arias-Palomo et al., 2019.

In eukaryotes, the clade 2 ORC/Cdc6 initiator complex is responsible for loading the MCM helicase. The MCM helicase is in fact a clade 7 AAA+ ATPase itself, but will be discussed briefly here in order to understand the mechanism of ORC/Cdc6 action. The eukaryotic helicase loading mechanism is very different to the bacterial one - the initiator first binds to DNA at the origin of replication, and only afterwards recruits the helicase to the DNA strand. The origin recognition complex (ORC) is formed by six Orc1-6 subunits. Orc1-5 are clade 2 AAA+ ATPases, and upon binding of ATP they form a five-membered spiral around double-stranded DNA with Orc6 binding at the periphery (Li et al., 2018). The gap between the first and fifth Orc subunits in this ring (Orc1 and Orc5) allows the spiral to bind to the DNA strand, and the related clade 2 ATPase Cdc6 then binds at the Orc1-Orc5 interface and closes the ring (Yuan et al., 2017).

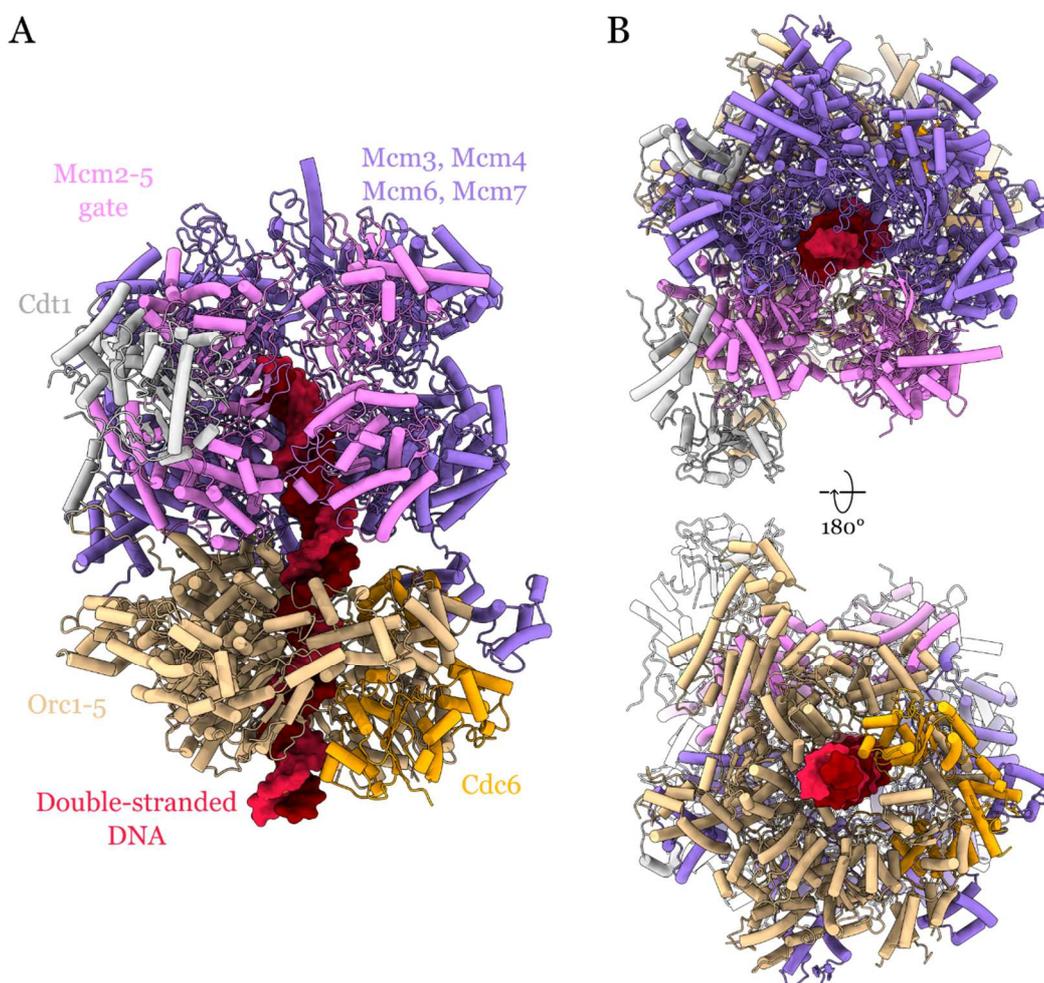


Figure 2.16. **A)** Structure of the *S. cerevisiae* ORC/Cdc6-Cdt1/MCM complex bound to double-stranded DNA (PDB ID: 5V8F) shown as cartoons, viewed from the side. The peripheral Orc6 is not shown. The Mcm2-5 gate which opens to accommodate the DNA double helix is highlighted in pink, other components are coloured as indicated. **B)** The ORC/Cdc6-Cdt1/MCM complex viewed from above (top) focusing on the MCM hexamer, and viewed from below (bottom) focusing on the ORC-Cdc6 ring.

Unlike the bacterial helicase DnaB, which requires the initiator DnaC to adopt an open ring for binding to DNA, the eukaryotic helicase MCM pre-adopts an open spiral conformation which is stabilised by Cdt1 (Frigola et al., 2017). MCM is made up of the six subunits Mcm2-7. The two subunits Mcm2 and Mcm5 form a ‘gate’ at the spiral seam which allows the helicase to bind DNA. The ORC/Cdc6 complex then recruits the Cdt1-stabilised MCM hexamer to the double-stranded DNA (Miller et al., 2019). The structure of the ORC/Cdc6-Cdt1/MCM complex is shown in Figure 2.16. After helicase loading, Cdc6 and Cdt1 are released and MCM hydrolyses ATP to close the helicase ring around the DNA molecule (Miller et al., 2019; Ticau et al., 2017). The overall mechanism is summarised in Figure 2.17 and contrasts with the bacterial system where the initiator DnaC, and not the helicase DnaB, hydrolyses ATP to close the helicase ring.

After closure of the MCM ring, a second ORC/Cdc6 complex is then recruited on the opposite side of the MCM hexamer. A second copy of the MCM helicase is then recruited, which binds facing the opposite direction to the first copy – this allows bidirectional DNA unwinding at the origin (Miller et al., 2019).

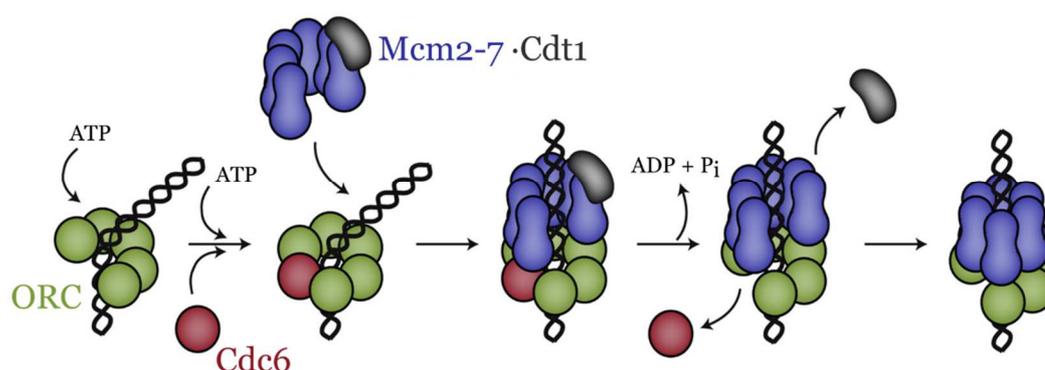


Figure 2.17. Mechanism of MCM loading by ORC/Cdc6. Adapted from Arias-Palomo et al., 2019.

### 2.6.3. Clade 3 – Classical

Clade 3 consists of members from the original AAA family, as established in 1993 (Kunau et al., 1993). Later on, bioinformatic analysis expanded the AAA family to cover many more proteins, and the AAA family was reclassified as a clade within the wider AAA+ superfamily (Neuwald et al., 1999). Members of clade 3 share a small  $\alpha$ -helical insertion after the canonical  $\beta$ 2 strand, which upon hexamerisation is located in the central pore and is involved in substrate recognition (Miller and Enemark, 2016). Unlike the DNA-binding clade 1 and 2 ATPases, clade 3 members act as protein remodellers, threading polypeptides through the central pore (Puchades et al., 2020).

There are two key differences between members of clade 3 and most other clades at the level of the ATP binding site. Classical clade 3 AAA+ proteins lack the sensor-II arginine (see Figure 2.4), which is mutated to an alanine for these members (Ogura et al., 2004). In addition, adjacent to the canonical arginine finger residue lies a second conserved arginine finger, which also interacts with bound ATP and is important for both ATP hydrolysis and oligomerisation (Erzberger and Berger, 2006; Sun et al., 2017). Because of these differences, it has been hypothesised that the ATP hydrolysis mechanism may differ between clade 3 AAA+ proteins and members of other clades (Ye et al., 2015).

Like many AAA+ ATPases, clade 3 members often have inserted N- or C-terminal domains, which are responsible for substrate binding or additional enzymatic activity (Erzberger and Berger, 2006). Because of this, there is huge functional diversity in this family, and consequently only a few examples will be discussed here. Clade 3 ATPases are key to a wide array of cellular processes that involve protein remodelling. For example, the human mitochondrial ATPase YME1 removes proteins from mitochondrial inner membrane and threads it through a C-terminal protease ring for degradation (Puchades et al., 2017). Similarly, the enzymes katanin, fidgetin and spastin are involved with disassembling microtubules in eukaryotes, and possess N-terminal microtubule interacting domains (Zehr et al., 2017).

Clade 3 AAA+ proteins can be categorised as type I or type II, depending on whether they possess a single or double AAA+ ring (Puchades et al., 2020). Type I proteins have only a single AAA+ ring, and include proteins such as katanin and Vps4. Type II proteins have a double AAA+ ring, and include proteins such as p97/Cdc48, VAT, ClpB-NTD/Hsp104-NTD and NSF. Occasionally one of the double rings is inactive and unable to hydrolyse ATP, as is the case for NSF (Brunger and DeLaBarre, 2003). In other cases such as for Hsp104, the C-terminal AAA+ domain possesses structural features of other AAA+ clades rather than clade 3 as discussed in section 2.6.5 (Frickey and Lupas, 2004).

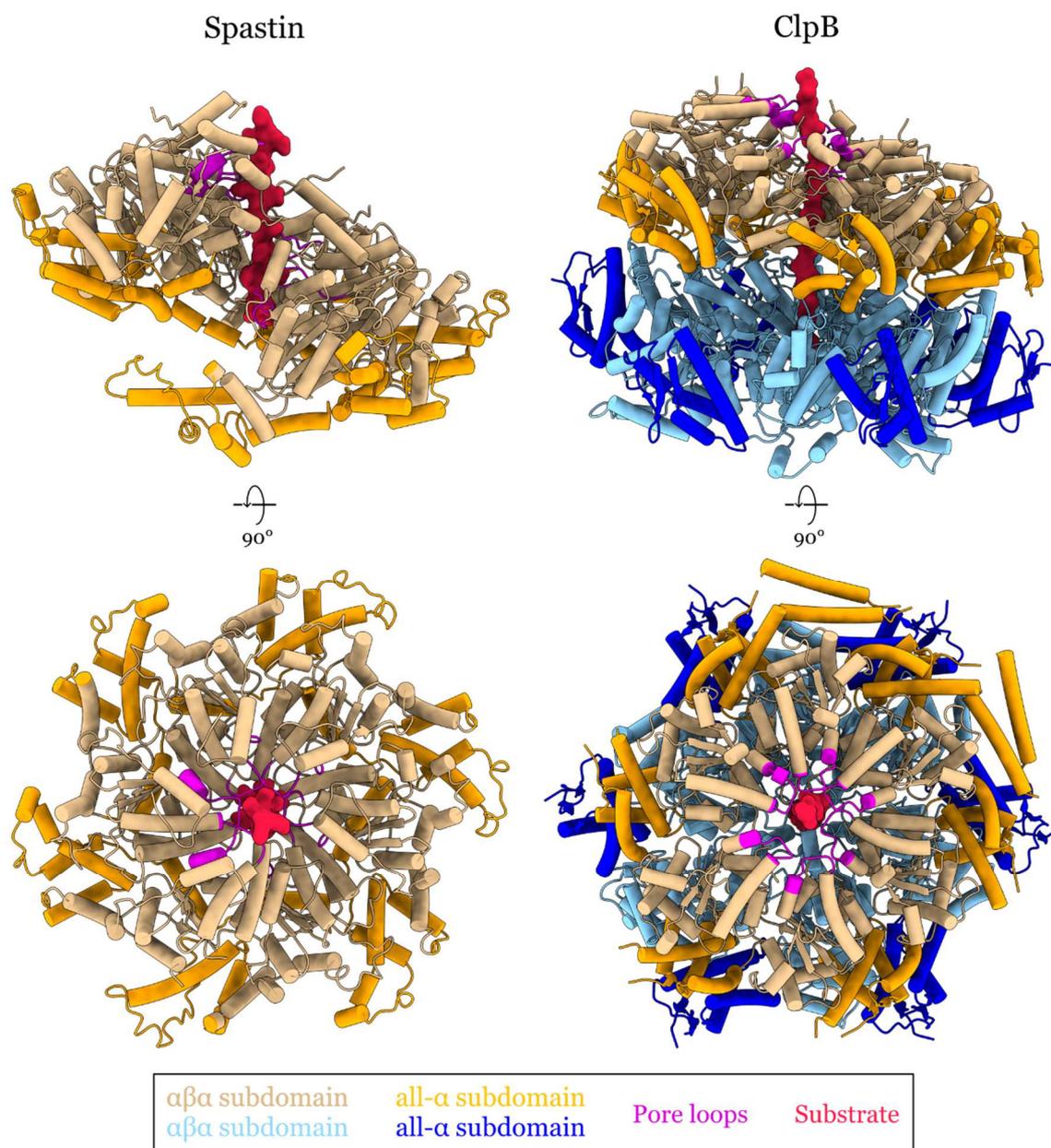


Figure 2.18. Structures of spastin (PDB ID: 6Po7) and ClpB (PDB ID: 6QS6) AAA+ domains bound to substrate (polyglutamate for spastin, casein for ClpB). The αβα subdomain is coloured brown for spastin and ClpB-NTD, and light blue for ClpB-CTD. The all-α subdomain is coloured orange for spastin and ClpB-NTD, and dark blue for ClpB-CTD. Substrate is coloured in red, while pore loops (corresponding to the small clade-specific insertion before α2) are coloured in magenta.

Examples of type I and type II clade 3 ATPases bound to polypeptide substrates are shown in Figure 2.18. Spastin is a type I clade 3 protein with a single AAA+ ring, while ClpB is a type II clade 3 protein with a double AAA+ ring. The N-terminal domain (NTD) is a clade 3 AAA+ ATPase, whereas the C-terminal domain (CTD) is evolutionarily distinct (see section 2.6.5). As discussed in section 2.5, clade 3 is one of the best studied clades of AAA+ ATPases, and most cryo-EM structures of AAA+ ATPase protein translocases with bound substrate are from this clade (Puchades et al., 2020).

### 2.6.4. Clade 4 – SF3 helicases

Clade 4 AAA+ ATPases are only found in DNA and RNA viruses and function as helicases, unwinding the viral genome for transcription by the host cell machinery (Hickman and Dyda, 2005; Iyer et al., 2004b). Examples of clade 4 members include the SV40 large T antigen (LTag), the D5 protein from vaccinia virus and the E1 protein from papillomavirus (Miller and Enemark, 2016). Instead of the canonical all- $\alpha$  C-terminal domain, SF3 helicases have a unique helical bundle made up of both N- and C-terminal helices (Hickman and Dyda, 2005; Kazlauskas et al., 2016; Medagli and Onesti, 2013). As such, clade 4 is the only other clade besides clade 3 to lack a sensor-II arginine residue (Erzberger and Berger, 2006).

SF3 helicases also have inserted N- or C-terminal DNA-binding domains that recognise the origin of replication – therefore, both origin recognition and helicase activity are carried out by the same protein in SF3 helicases, in contrast to helicase loader/helicase pairing of clade 2 AAA+ ATPases (Chang et al., 2013; Hickman and Dyda, 2005; Kazlauskas et al., 2016; Singleton et al., 2007). An example is shown in Figure 2.19 – the domain

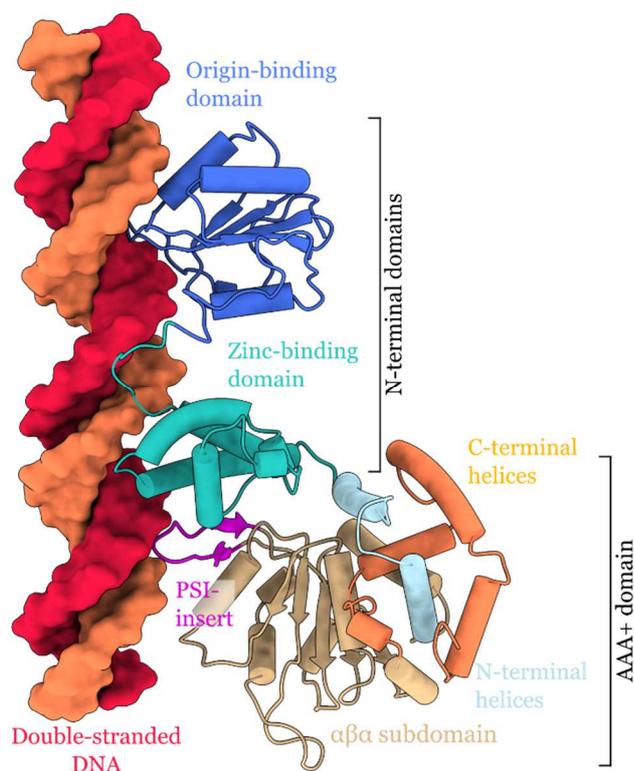


Figure 2.19. Structure of an SV40 helicase LTag monomer bound to double-stranded DNA (PDB ID: 4GDF) shown as cartoons. Domains are coloured as indicated. Instead of the canonical AAA+ all- $\alpha$  C-terminal subdomain, the unique clade 4 helical bundle is composed of helices from both the N- and C-terminal ends of the AAA+ domain.

architecture of the SV40 LTag is shown, indicating the conserved AAA+ core and N-terminally inserted DNA-binding domains.

Like other PSI-insert clades (see Figure 2.9), clade 4 helicases possess a  $\beta$ -hairpin insertion between  $\alpha_3$  and  $\beta_4$ , before the S-I motif (Erzberger and Berger, 2006; Iyer et al., 2004b). This  $\beta$ -hairpin loop is often located in the centre of the hexameric ring and involved with the translocation of substrates through the pore (Miller and Enemark, 2016). In clade 4 ATPases, the loop interacts with double-stranded DNA during origin recognition (as shown in Figure 2.19) and with single-stranded DNA during helicase unwinding (Chang et al., 2013; Enemark and Joshua-Tor, 2006) as shown for the E1 helicase in Figure 2.20. In the hexameric helicase ring, these pore loops are arranged in a spiral staircase-like arrangement, similarly to clade 3 AAA+ ATPases (Enemark and Joshua-Tor, 2006). The ATPase mechanism of SF3 helicases is also hypothesised to occur in an around-the-ring fashion, with the downward movement of PSI-insert loops predicted to pull single-stranded DNA through during helicase unwinding.

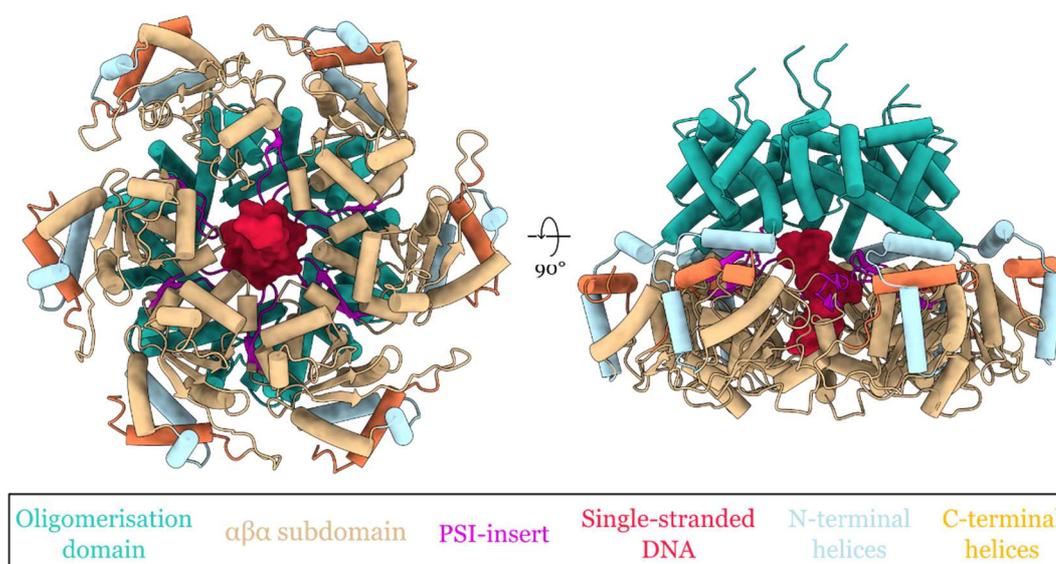


Figure 2.20. Structure of the E1 helicase oligomerisation and AAA+ domains from papillomavirus bound to single-stranded DNA (PDB ID: 2GXA) viewed from the bottom (left) and side (right), with domains coloured as indicated. Similarly to SV40 LTag, E1 also possesses the helical bundle composed of helices from the N- and C-terminal end of the AAA+ domain.

### 2.6.5. Clade 5 – HCLR

The HCLR clade is named after its founding members HslU/ClpX, ClpABC-CTD, Lon, and RuvB. Apart from the PSI-insert, clade 5 AAA+ proteins do not have any other insertions in the ancestral AAA+ core. Similarly to clade 3 proteins, clade 5 proteins are protein translocases, often involved in proteolytic degradation (Miller and Enemark, 2016). However unlike clade 4 helicases, the PSI-insert in clade 5 AAA+ ATPases appears to be involved with substrate recognition rather than substrate translocation (Martin et al., 2008; Miller and Enemark, 2016; Shin et al., 2019).

Two examples of clade 5 AAA+ proteins are Lon and ClpX. The Lon protease consists of a AAA+ domain and a protease domain, and is involved in protein degradation in the bacterial and archaeal cytoplasm, and in the mitochondrial matrix in eukaryotes (Shin et al., 2019). Similarly, the unfoldase ClpX is involved in proteolysis in bacteria, archaea, and mitochondria, but rather than possessing a protease domain it interacts with the dedicated protease ClpP (Hanson and Whiteheart, 2005; Ogura and Wilkinson, 2001). Figure 2.21 shows the structure of ClpXP complex determined by cryo-EM, with the hexameric ClpX sitting above the seven-fold symmetric ClpP protease barrel into which unfolded substrates are threaded for proteolysis (Gatsogiannis et al., 2019).

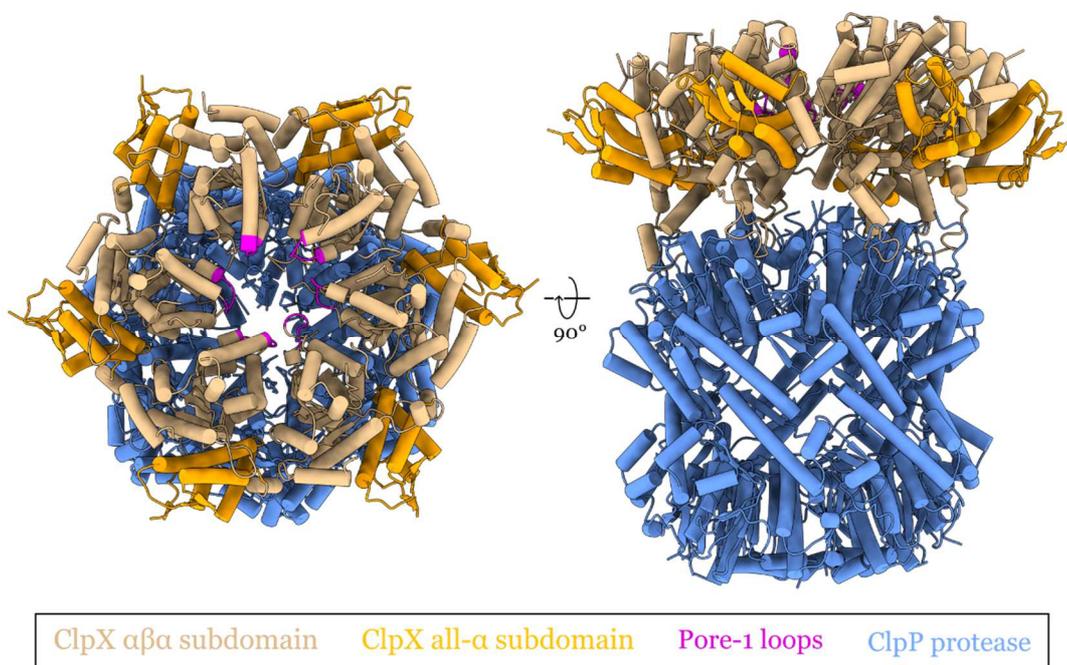


Figure 2.21. Structure of the ClpXP complex (ClpX PDB ID: 6SFW, ClpP PDB ID: 6SFX) viewed from the top (left) and side (right). ClpX subdomains and ClpP are coloured as indicated. The pore-1 loops are coloured magenta, while the RKH loops (corresponding to the PSI-insert) and pore-2 loops are not resolved in the model.

Three separate pore loops are responsible for the interaction between ClpX and its substrate (Ripstein et al., 2020). One of these is the PSI-insert typical of clades 4-7 (termed the ‘RKH loop’ for ClpX), which is crucial for substrate recognition (Martin et al., 2008). Due to its inherent flexibility when not bound to substrate it is not resolved in the PDB model in Figure 2.21. Two recent cryo-EM studies resolve the PSI-inserts and show that they are indeed bound to substrate above the ClpX hexamer, confirming the importance of the PSI-insert for substrate recognition (Fei et al., 2020; Ripstein et al., 2020). However, the corresponding PDB models have not yet been released at the time of writing.

There are currently unresolved questions about whether the general around-the-ring mechanism proposed for clade 3 ATPases is generally applicable to clade 5 ATPases as well (Gates and Martin, 2019). A mechanism for ClpX has been proposed based on biochemical and X-ray crystallography studies in which ATP hydrolysis does not occur sequentially around the hexameric ring but rather occurs stochastically (Glynn et al., 2009; Stinson et al., 2013). One of the recent cryo-EM studies on the ClpXP complex appears to support this mechanism (Fei et al., 2020), but another instead proposes that the clade 3-like rotary mechanism applies to ClpX (Ripstein et al., 2020). In addition, a recent cryo-EM structure of the Lon protease suggests that the rotary mechanism indeed extends to clade 5 ATPases (Shin et al., 2019). Further work on other clade 5 AAA+ proteins (and indeed other clades) is therefore needed to resolve the debate about whether a generally-applicable ATP hydrolysis mechanism exists.

While carrying out structural comparisons between various clade 5 proteins, I noticed that there are some inconsistencies between observed structural features and classifications of some clade 5 proteins. In particular, the C-terminal AAA+ domains of ClpB and Hsp104 (which are homologues from bacteria and yeast respectively) and Lon appear to have an additional insertion in the  $\alpha 2$  helix, which is a structural hallmark of clade 6 and 7 AAA+ ATPases (the ‘helix-2 insert’). A comparison between selected clade 5 proteins, with reference to structures from clade 3, 6, and 7 representatives is shown in Figure 2.22, focussed on the region around the  $\alpha 2$  and  $\alpha 3$  helices. These ambiguities may raise questions about whether some supposed clade 5 members may in fact be better classified as clade 6 proteins.

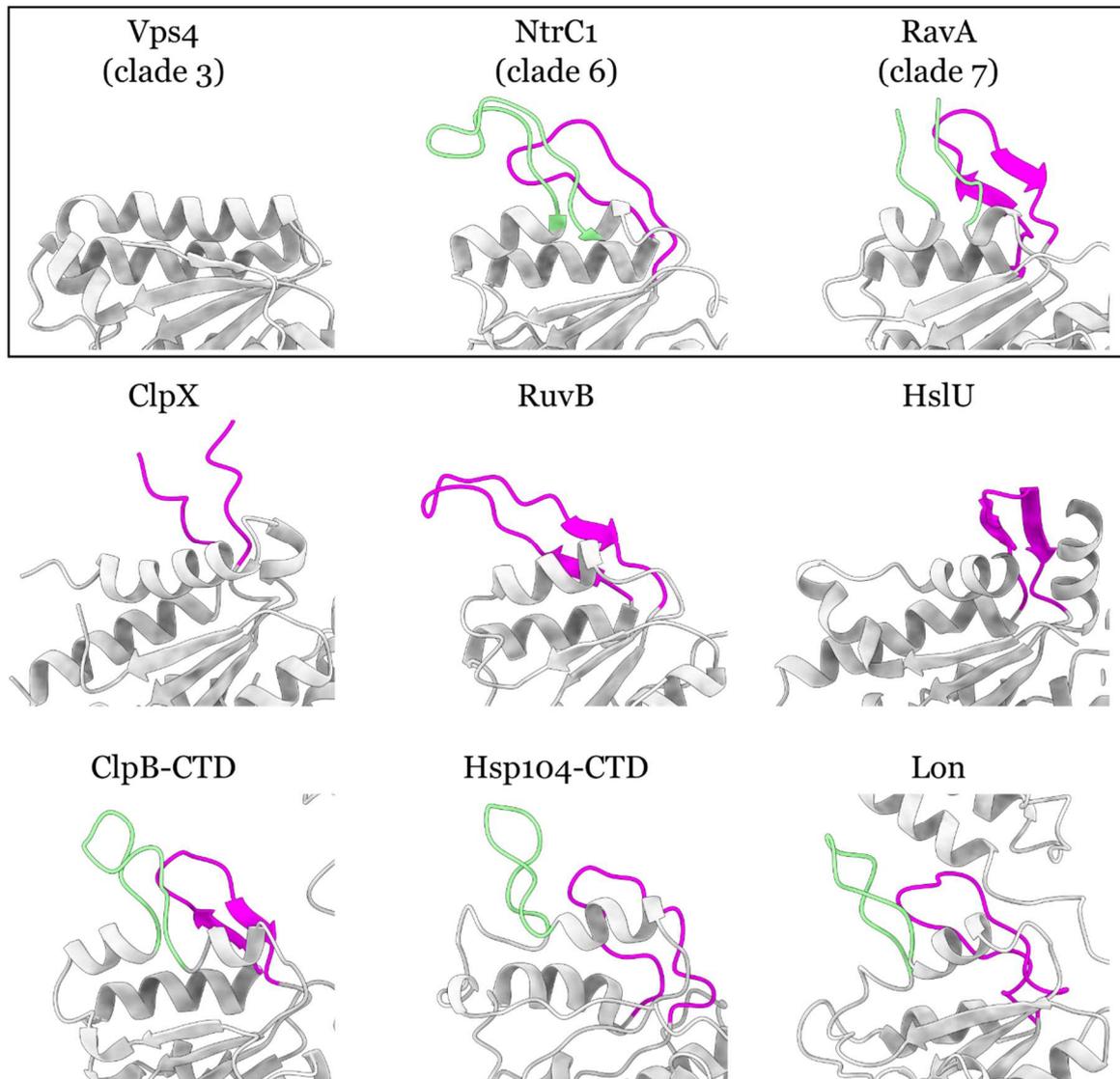


Figure 2.22. Comparison between monomer structures, focussed on  $\alpha$ -helices  $\alpha 2$  and  $\alpha 3$ , from the selected clade 5 proteins ClpX (PDB ID: 3HWS), RuvB (PDB ID: 1HQC), HslU (PDB ID: 1HT2), ClpB-CTD (PDB ID: 6QS6), Hsp104-CTD (PDB ID: 6AHF), and Lon (PDB ID: 6N2I). For comparison, representative structures from each of clade 3 (Vps4, PDB ID: 6AP1), clade 6 (NtrC1, PDB ID: 1NY5), and clade 7 (RavA, PDB ID: 3NBX) are shown in the box at the top. PSI-inserts are present in all proteins except for the clade 3 Vps4, and are coloured in magenta. The clade 6 and 7-specific H2-insert is coloured in light green for NtrC1 and RavA. The clade 5 proteins ClpX, RuvB and HslU do not show any insertion in  $\alpha 2$ , whereas ClpB-CTD, Hsp104-CTD and Lon all possess  $\beta$ -hairpin insertions in  $\alpha 2$ .

### 2.6.6. Clade 6 – H2-insert

As well as the PS-I insert, members of clade 6 possess an additional  $\beta$ -hairpin insertion in the  $\alpha 2$  helix termed the helix-2 insert (H2-insert) as shown in Figure 2.22 above. Clade 6 contains the NtrC and McrB families, although as discussed in the previous section other AAA+ proteins may also appear to possess a  $\beta$ -hairpin insertion in  $\alpha 2$ .

NtrC-family proteins activate transcription of genes under the control of  $\sigma^{54}$  (also known as RpoN) in bacteria, which are involved in diverse processes such as nitrogen metabolism, biofilm formation and flagellum synthesis (Banerjee et al., 2019; Joly et al., 2012).  $\sigma^{54}$  recruits RNA polymerase to specific sequences in gene promoters, and using the energy from ATP hydrolysis, NtrC-family AAA+ proteins remodel  $\sigma^{54}$ -bound RNA polymerase from a closed to an open conformation, allowing transcription to occur. The NtrC family proteins NtrC1, FleQ and PspF all contain the motif GAFTGA in the H2-insert, and this motif is crucial for the interaction with  $\sigma^{54}$  in the  $\sigma^{54}$ /RNA polymerase complex (Bordes et al., 2003; Joly et al., 2012; Rappas et al., 2006). One X-ray crystal structure of NtrC1 shows a planar heptameric oligomeric state (as shown in Figure 2.23), while another instead shows an asymmetric hexamer, with the H2-inserts forming a spiral staircase in the hexameric pore (Sysoeva et al., 2013).

In contrast, the bacterial clade 6 protein McrB uses the energy from nucleotide hydrolysis to power endonuclease activity by McrC (Nirwan et al., 2019). In comparison to most other AAA+ proteins, McrB hydrolyses GTP more efficiently than ATP (Erzberger and Berger, 2006). The N-terminal DNA-binding domain of the McrB hexamer recognises methylated DNA, which is typical to certain bacteriophage genomes, and GTP hydrolysis is likely to induce DNA translocation (Nirwan et al., 2019). Two McrB hexamers are bridged by a McrC dimer, and the H2-insert in McrB (often called ‘Loop L1’) is responsible for the interaction with McrC, which plugs the central pore of McrB. In addition, the H2-insert is hypothesised to interact with the DNA substrate during translocation. GTP hydrolysis is expected to occur in a sequential manner around the ring, feeding methylated DNA to the McrC endonuclease for cleavage. However, further studies are required to confirm this mechanism.

Figure 2.23 shows structures of NtrC1 and McrBC oligomers, with the H2-inserts and PSI-inserts highlighted. Clade 6 proteins are relatively under-studied compared to other clades, and further work is needed to understand if the mechanisms proposed for other ATPases are applicable to clade 6 proteins such as NtrC1 and McrB, and if these proteins also form asymmetric open-ring spirals.

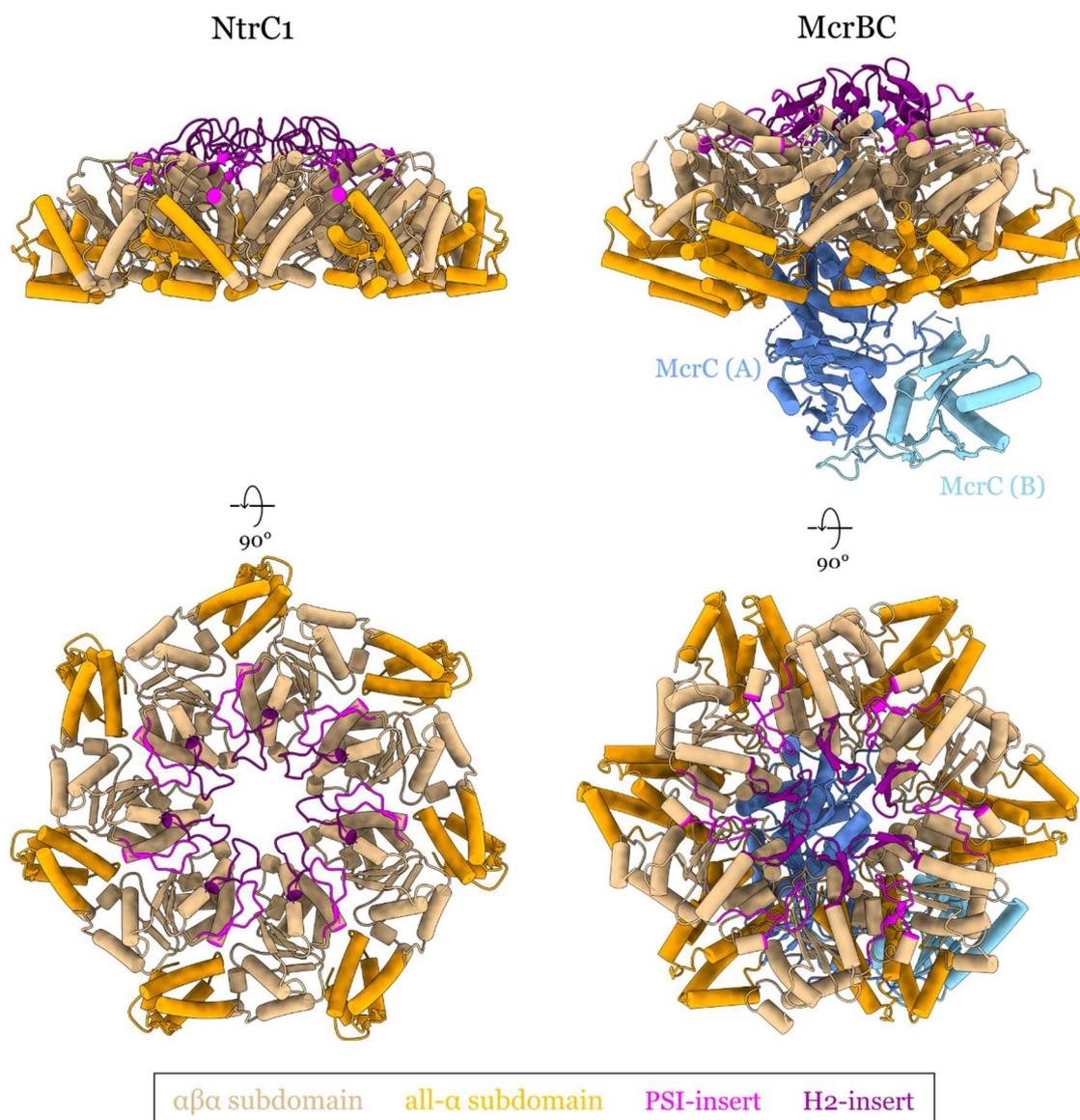


Figure 2.23. Structures of the NtrC1 ATPase domain heptamer (left, PDB ID: 3MoE) and the McrBC complex (right, with only one McrB hexamer displayed, PDB ID: 6HZ4) from the side (above) and top (below). AAA+ subdomains are coloured as indicated, and the PSI-inserts and H2-inserts are coloured in magenta and purple respectively. For the McrBC complex, the two McrC molecules are coloured in different shades of blue.

## 2.6.7. Clade 7 – PSII-insert

RavA is a member of clade 7, which is typified by an unusual arrangement of  $\alpha\beta\alpha$  and all- $\alpha$  subdomains. The defining structural feature of clade 7 AAA+ ATPases is an  $\alpha$ -helical insertion after  $\alpha_5$ , which repositions the entire C-terminal all- $\alpha$  subdomain to the opposite side of the  $\alpha\beta\alpha$  subdomain compared to other clades (Erzberger and Berger, 2006; Miller and Enemark, 2016). This insertion is located before the sensor-II motif and is therefore termed the PSII-insert. The PSII-insert repositions the S-II motif away from the ATP binding site of the  $\alpha\beta\alpha$  domain, such that the S-II motif instead acts *in trans* and contacts the  $\gamma$ -phosphate of the neighbouring monomer (Erzberger and Berger, 2006). Consequently, the nucleotide binding site of clade 7 ATPases is predominantly located between the large and small AAA domains of adjacent monomers, rather than between large and small domains of the same monomer (Wong and Houry, 2012). A comparison between the structures of NtrC1 and RavA monomers (from clade 6 and 7 respectively) is shown in Figure 2.24, showing the rearrangement of the all- $\alpha$  subdomain after the PSII-insert. A comparison of the ATP binding site differences between clade 7 and other clades is presented in Figure 5 of the manuscript presented in Chapter 4.

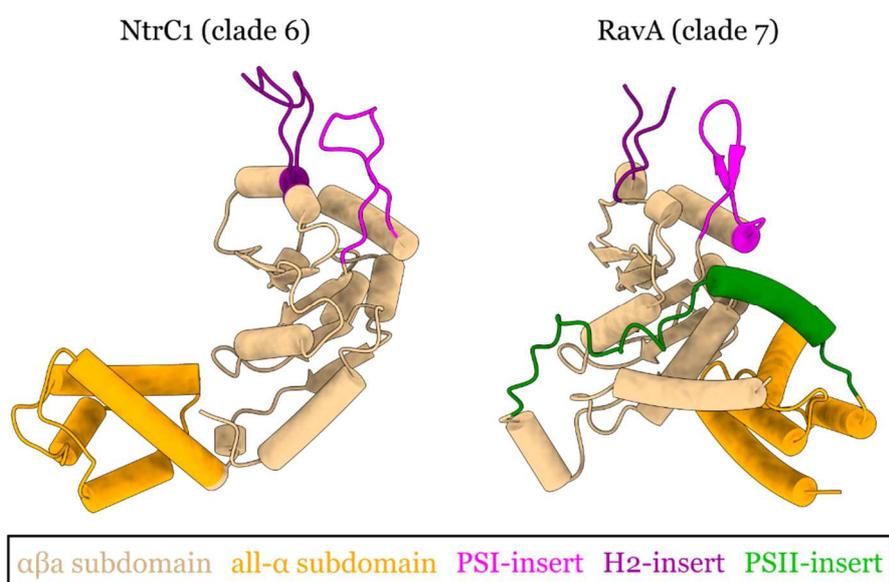


Figure 2.24. Monomer structures of NtrC1 (PDB ID: 1NY5) and the RavA AAA+ domain (PDB ID: 3NBX) showing the rearrangement of the all- $\alpha$  subdomain following the PSII-insert typical to clade 7 AAA+ ATPases. Subdomains and structural features are coloured as indicated.

The PSII-insert clade is a diverse clade which was established containing DNA-interacting helicases, the eukaryotic motor protein dynein, the ribosome biogenesis protein midasin (also known as Rea1), and the MoxR family of which RavA is a member, amongst others (Erzberger and Berger, 2006). Like most other AAA+ proteins, clade 7 members form hexamers, although in both dynein and midasin the six monomers are encoded by a

single polypeptide chain, with additional insertions each chain conferring additional functionality (Carter, 2013; Kon et al., 2012; Sosnowski et al., 2018). Figure 2.25 shows the structure of dynein, an example of a clade 7 motor protein, with a striking stalk insertion at the end of the fourth AAA+ domain that is responsible for binding to microtubules (Schmidt et al., 2015).

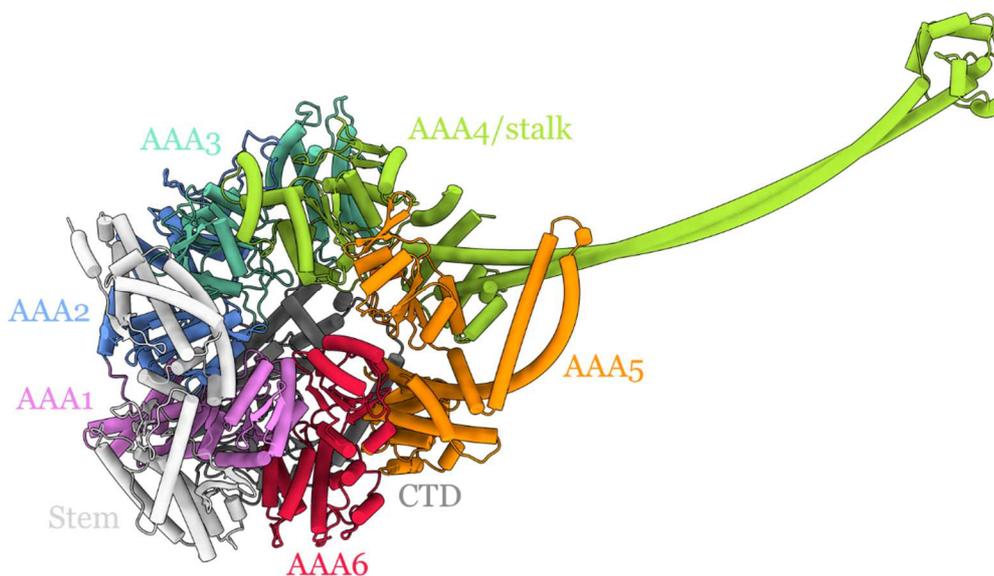


Figure 2.25. Structure of human dynein (PDB ID: 4RH7), with individual AAA+ domains in the fused protein chain coloured as indicated. The N-terminal stem and C-terminal domain are coloured light and dark grey respectively.

Like for clade 6 AAA+ proteins, the H2-insert plays a crucial role in the activity of clade 7 proteins. The MCM helicase (introduced in section 2.6.2) interacts with DNA during unwinding through both its PSI-insert and H2-insert loops, which are arranged in a spiral around the DNA molecule and point into the central hexameric pore (Meagher et al., 2019; Yuan et al., 2017). Key residues in the H2-insert are also necessary for the MoxR protein CbbQ to function as a RuBisCO activase, and in the CbbQ hexamer the H2-inserts protrude into the central hexameric pore (Tsai et al., 2019).

The H2-insert also plays important functional roles in dynein and midasin/Rea1. The H2-insert in the second AAA+ domain of dynein (AAA2) is critical for dynein's motor activity, with AAA2 H2-insert mutants still able to hydrolyse ATP but not able to perform the power stroke associated with motor function (Kon et al., 2012). In contrast, the H2-insert of the Rea1 AAA2 is extended by an  $\alpha$ -helical bundle, which sits in the centre of the hexameric ring as a plug and inhibits ATPase action of the hexamer when not bound to its substrate (Sosnowski et al., 2018).

The AAA+ clades were defined based on sequence analysis coupled with the comparison of protein structures available at the time (Ammelburg et al., 2006; Erzberger

and Berger, 2006; Iyer et al., 2004b). When clade 7 was established (Erzberger and Berger, 2006) the only clade 7 protein with a structure in the PDB was the magnesium chelatase BchI (Fodje et al., 2001). Since then, there have been several structures published of clade 7 proteins, including of dynein, midasin/Rea1, MCM, and the MoxR-family proteins RavA, CHU\_0153 and CbbQ. During the course of preparing the manuscript on RavA presented in Chapter 4, I compared the available structures of different clade 7 AAA+ ATPases. While several of the structures possess the PSII-insert that repositions the all- $\alpha$  domain, dynein and CbbQ notably lack this insert, as shown in Supplementary Figure 8 of the manuscript. This would suggest that, at least structurally, some ATPases that have been classified into clade 7 may be more similar to clade 6 proteins.

In the past decade there have been a large number of structures of AAA+ proteins published, with many of them being obtained using cryo-EM. In addition to the inconsistencies with CbbQ and dynein being classified as clade 7 proteins despite lacking a PSII-insert, several clade 5 proteins possess the H2-insert typical of clade 6 and 7 proteins (as discussed in section 2.6.5). Furthermore, some AAA+ proteins such as TRIP13/PCH2 do not appear to fall into any single clade, but possess features of several clades and therefore are ambiguously placed in the AAA+ superfamily (Ye et al., 2015). Perhaps with the recent wealth of structural information available on AAA+ ATPases, it is worth another look at how the different clades are defined in order to understand how the mechanism of ATP hydrolysis may differ amongst AAA+ proteins, and whether a generally-applicable model for AAA+ ATPase activity indeed exists.

## Chapter 3. The MoxR AAA+ ATPase RavA

### 3.1. The MoxR family

RavA is a MoxR-family AAA+ ATPase, which is a member of the AAA+ clade 7. The MoxR family is widespread across both bacteria and archaea, and is named after the MoxR protein from *Paracoccus denitrificans* which facilitates methanol oxidation by inserting a Ca<sup>2+</sup> cofactor into the methanol dehydrogenase complex (Van Spanning et al., 1991; Toyama et al., 1998). MoxR-family proteins are generally proposed to act as molecular chaperones using the energy of ATP hydrolysis to mature or remodel protein complexes, often through the insertion of metal cofactors (Iyer et al., 2004b; Snider and Houry, 2006; Snider et al., 2006).

For example, the *P. denitrificans* MoxR protein NorQ is involved in the insertion of an Fe<sub>B</sub> cofactor into nitric oxide reductase (Kahle et al., 2018). Similarly, the *Acidithiobacillus ferrooxidans* MoxR protein CbbQ is crucial for the activation of the carbon-fixing enzyme RuBisCO (Sutter et al., 2015; Tsai et al., 2019, 2015). Despite the ubiquity of MoxR proteins in bacteria and archaea the family is relatively poorly-studied, and there is little structural or functional information on MoxR-family proteins compared to other AAA+ ATPase families (Snider and Houry, 2006; Wong and Houry, 2012).

One common feature of MoxR proteins is that they are often found in an operon together with genes coding for von Willebrand factor A (VWA) domain-containing proteins (Snider and Houry, 2006). VWA proteins are involved in the mediation of protein-protein interactions, and are dependent on bound metal ions for their function (Whittaker and Hynes, 2002). For some MoxR subfamilies, at least 70% of genes are found in close proximity to genes coding for VWA proteins (Snider and Houry, 2006). This is particularly the case for the RavA subfamily, with 33 of the 36 identified RavA genes occurring next to a VWA protein. These VWA proteins often act in tandem with the MoxR protein to facilitate their activity – for example, CbbQ and NorQ rely on the VWA proteins CbbO and NorD respectively for their chaperone activity (Kahle et al., 2018; Tsai et al., 2015).

### 3.2. The *ravAviaA* operon

In an effort to identify new and uncharacterised bacterial chaperones, our collaborator Walid Houry's group began investigating MoxR proteins in the early 2000s. During this search, they identified the MoxR protein RavA (regulatory ATPase variant A), the founding member of the RavA MoxR subfamily, which is coded for by an operon that also encodes the VWA protein ViaA (VWA interacting with AAA+ ATPase) in *E. coli*. In collaboration with the Houry group, Irina Gutsche's team first characterised the *E. coli* RavA in 2006 (Snider et al., 2006). The *ravAviaA* operon is controlled by a  $\sigma^S$  promoter, suggesting that RavA and ViaA might play a chaperone role under stress conditions in the cell. Bioinformatic analysis revealed that RavA is a 56 kDa protein consisting of an N-terminal AAA+ domain and a C-terminal domain with little homology to previously-identified protein domains. RavA forms hexamers in the presence of nucleotides, similarly to other AAA+ ATPases.

### 3.3. RavA interacts with the inducible lysine decarboxylase LdcI

In order to determine the function of RavA, pull-down studies were carried out using TAP-tagged RavA to find RavA binding partners (Snider et al., 2006). These studies identified the inducible lysine decarboxylase LdcI as a potential binding partner. LdcI is involved in the acid stress response, and is a key metabolic enzyme of *E. coli* that has been studied for decades, because of its importance to *E. coli* survival in low pH environments (Gale, 1946; Gale and Epps, 1942; Meng and Bennett, 1992; Sabo et al., 1974). LdcI is a member of the LAOdc (Lysine-Arginine-Ornithine decarboxylase) family of PLP-dependant basic amino acid decarboxylases. LdcI catalyses the decarboxylation of lysine to the polyamine cadaverine (Figure 3.1), in a reaction which consumes  $H^+$  and produces  $CO_2$ . This buffers both the intracellular medium and extracellular environment, mitigating acid stress.

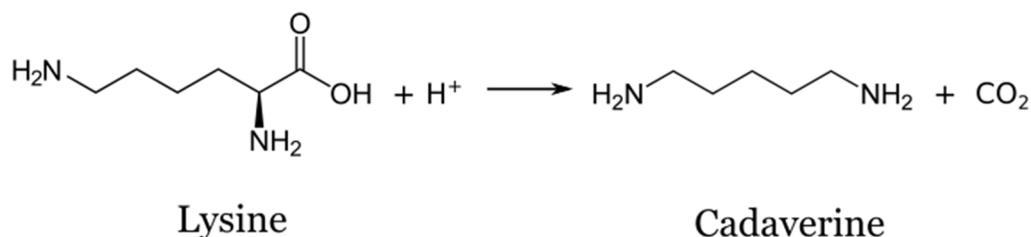


Figure 3.1. LdcI catalyses the reaction of lysine to cadaverine, consuming a proton and producing  $CO_2$ .

*E. coli* possesses several other LAOdcS that share a common evolutionary ancestor with LdcI (Kanjee et al., 2011a). However, RavA only interacts with LdcI and not with other LAOdcS, including the closely-related biosynthetic lysine decarboxylase LdcC (Snider et al., 2006). Irina Gutsche's group have been extensively involved in the characterisation of these decarboxylases in both *E. coli* and other bacteria, which differ in their specificity or active pH (Carriel et al., 2018; Kandiah et al., 2016, 2019; Kanjee et al., 2011a; Felix et al., in preparation).

At pH values below 7.0 LdcI forms decamers, and at higher pH these decamers dissociate into dimers (Sabo et al., 1974; Snider et al., 2006). The group characterised LdcI by negative stain EM in 2006 and showed that the LdcI decamers are D<sub>5</sub>-symmetric and are formed of a tight ring of dimers, as shown in Figure 3.2.

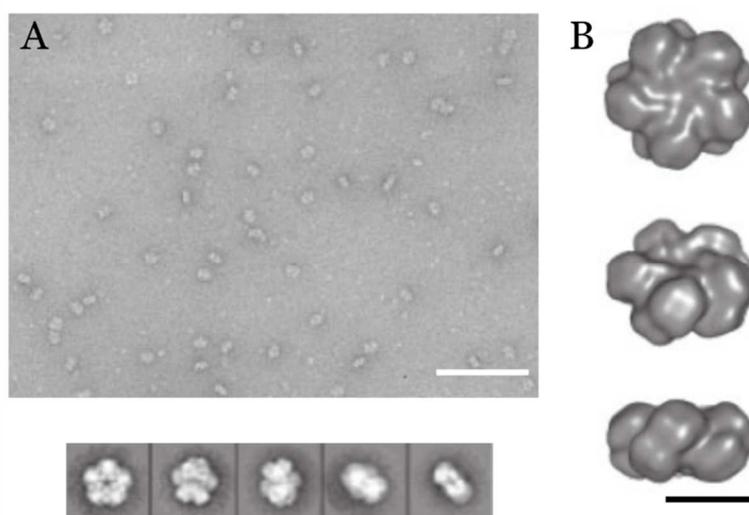


Figure 3.2. Negative stain EM characterisation of LdcI. **A**) Micrograph of LdcI complex showing compact donut-shaped particles (above, scale bar = 100 nm) and five 2D class averages showing different views (below). **B**) 3D reconstruction of the LdcI decamer complex with applied D<sub>5</sub> symmetry, viewed from top (above) tilted (middle) and side (below) views. Five dimers are arranged in the decameric ring, with a pronounced tilt to each dimer. Scale bar = 100 Å, adapted from Snider et al., 2006.

Unexpectedly, rather than being a substrate for RavA it was shown that LdcI formed a large complex with RavA. The first characterisation by negative stain EM of the LdcI-RavA complex was presented in 2006 by the group, showing that the complex is composed of two LdcI decamers bridged by five densities attributed to RavA hexamers (shown in Figure 3.3). They hypothesised that LdcI may act as a scaffold for the ATPase function of RavA, and that the complex would perhaps act as a novel chaperone. However, the resolution of the reconstruction (~30 Å) was only sufficient to visualise the overall shape of the complex, and none of its molecular determinants. In order to better understand the RavA-LdcI interaction, the group subsequently sought to determine the high-resolution crystal structures of both RavA and LdcI.

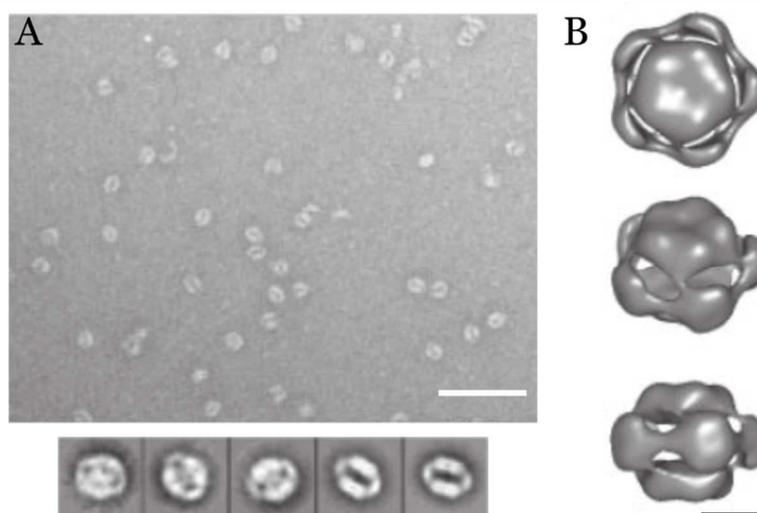


Figure 3.3. Negative stain EM characterisation of the LdcI-RavA complex. **A)** Micrograph of the LdcI-RavA complex showing circular particles with a central cavity (above, scale bar = 100 nm) and five 2D class averages (below). **B)** 3D reconstruction of the LdcI-RavA complex with applied D5 symmetry, viewed from top (above) tilted (middle) and side (below) views. The D5-symmetric LdcI decamer is visible in the centre, surrounded by five other densities corresponding to RavA. Scale bar = 100 Å, adapted from Snider et al., 2006.

### 3.4. RavA and LdcI X-ray crystal structures

The structure of the 56 kDa RavA monomer was solved in 2010 (El Bakkouri et al., 2010), and is shown in Figure 3.4. The N-terminal AAA+ domain of RavA displays the characteristic features of clade 7 ATPases, including the PSII-insert which forms a linker that rearranges the all- $\alpha$  AAA subdomain and repositions the S-II motif (see section 2.6.7). The C-terminal region of RavA, which is poorly-conserved across species and possesses no detectable homology to other proteins, forms two distinct domains in the structure. An extended triple-helical bundle extends below the AAA+ core, and a small globular domain mostly consisting of  $\beta$ -sheets adopts a unique fold and sits at the end of the triple-helical bundle. This globular domain is responsible for the interaction between RavA and LdcI, and removing this domain completely abolishes the LdcI-RavA interaction. Because of this, it is termed the LARA domain (LdcI associating domain of RavA).

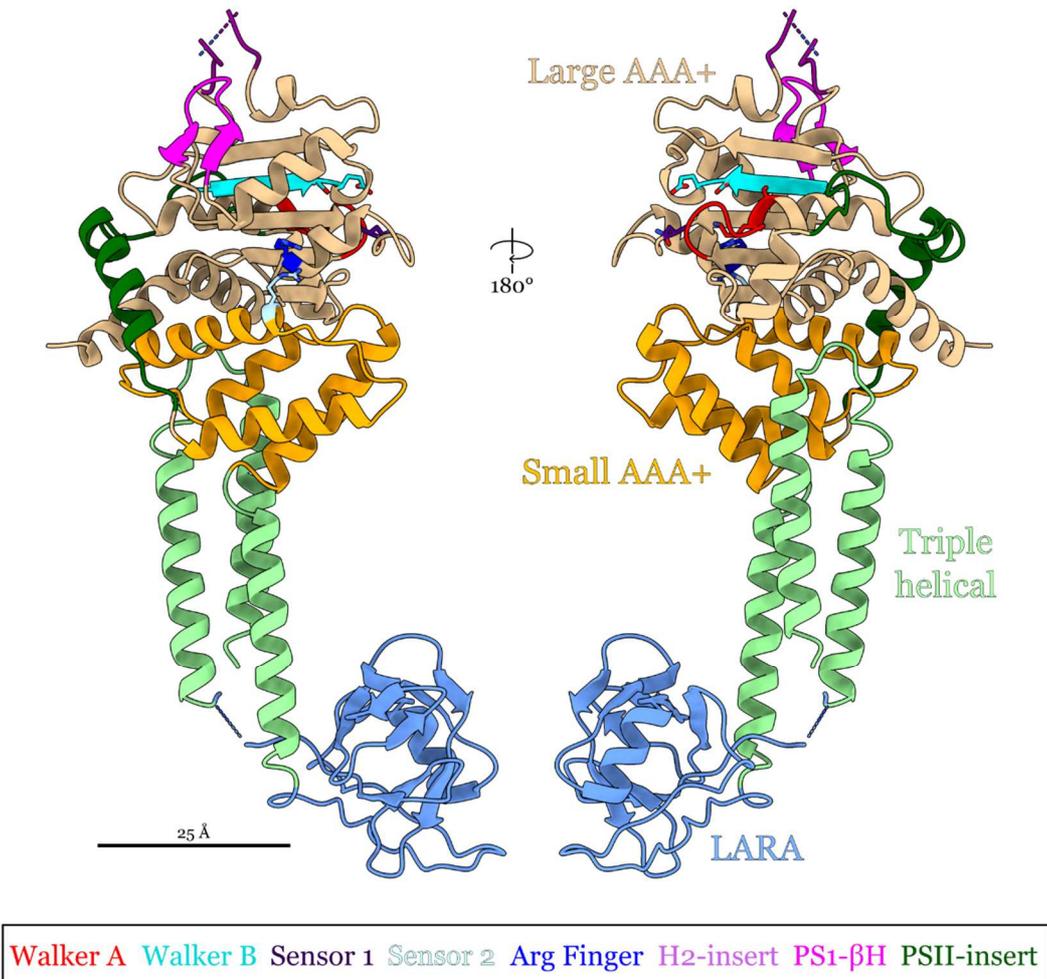


Figure 3.4. X-ray crystal structure of RavA (PDB ID: 3NBX) from two different views. The triple helical domain extends from the core AAA+ domain, with the small LARA domain at the end. The major domains are labelled, and key structural features are coloured as indicated.

The negative stain EM structure of the RavA hexamer in the presence of ADP was presented in the same paper, with applied six-fold symmetry, as shown in Figure 3.5. Fitting the crystal structure of the RavA monomer into the negative stain EM density revealed that the core AAA+ domains are arranged in the centre of the hexamer, with the triple-helical and LARA domains extending outwards. The fitting was guided by comparison with the crystal structure of the HslU hexameric ring, which possesses the canonical arrangement of the  $\alpha\beta\alpha$  and all- $\alpha$  subdomains. In particular, the sensor-II motif of the all- $\alpha$  subdomain of one RavA monomer was positioned using HslU as a reference, so that it contacted the ADP molecule bound to the next monomer in the ring, in order to preserve the geometry of the ATP binding site.

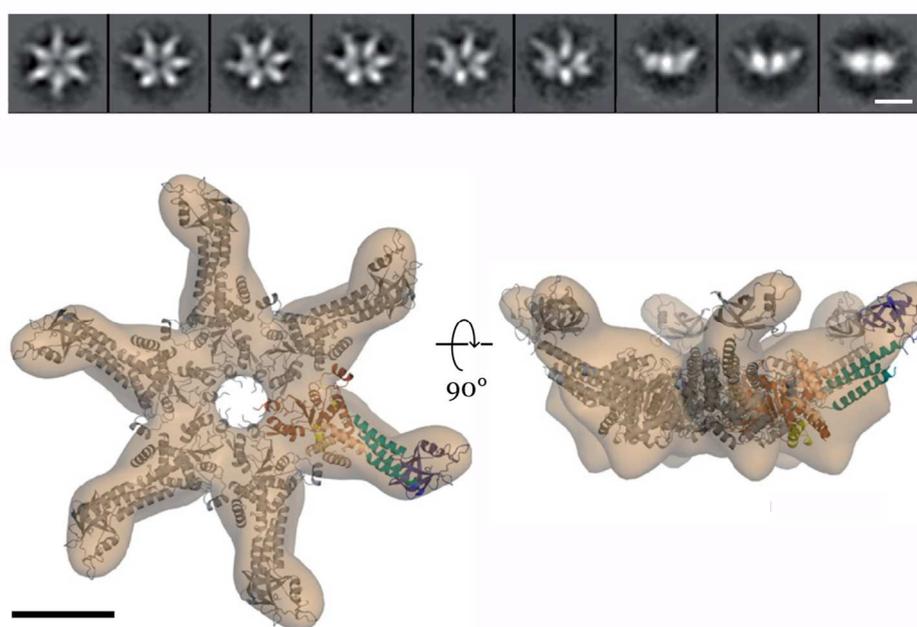


Figure 3.5. Negative stain EM analysis of the RavA hexamer. 2D classes (top, scale bar = 100 nm) show that the RavA hexamer has an approximate six-fold symmetry, with protrusions extending from the centre of the ring. The 3D reconstruction (below, scale bar = 50 Å) fitted with the RavA crystal structure (PDB ID: 3NBX) shows that the protrusions correspond to the triple-helical bundles and LARAs. One monomer is colored in rainbow, the rest are in brown. Adapted from El Bakkouri et al., 2010.

The crystal structure of LdcI was also determined in 2010 by the Houry group (Figure 3.6), confirming the 2006 negative stain EM model (Kanjee et al., 2011b). Unexpectedly, the stringent response alarmone guanosine tetraphosphate (ppGpp) was bound at the dimer interface. ppGpp is produced in response to environmental stress or nutrient starvation to modulate cell metabolism. The Houry group showed that ppGpp binding to LdcI inhibits LdcI activity and proposed that this interaction functions to limit the consumption of lysine, an important metabolic building block, under nutrient stress. LdcI therefore links the acid stress response to the nutrient stress response in *E. coli*. Notably, upon binding of RavA to LdcI the inhibitory effect of ppGpp is removed.

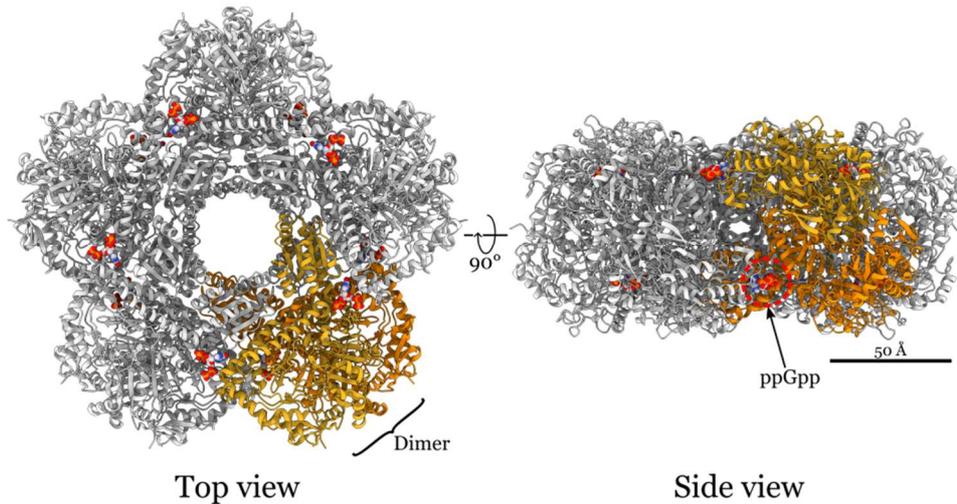


Figure 3.6. Top (left) and side (right) views of the X-ray crystal structure of the LdcI decamer (PDB ID: 3N75). The decamer is built of five dimers, one of which has monomers coloured in gold and orange. There are 10 bound molecules of ppGpp, which bind at the interface between dimers as indicated.

The atomic model of the LdcI decamer and the pseudo-atomic model of the RavA hexamer were combined to give a proposition for how RavA and LdcI would be arranged in the 2006 negative stain EM map, as shown in Figure 3.7. However, the model was speculative and required validation. Therefore, in order to probe the molecular determinants of the LdcI-RavA complex, cryo-EM analysis was carried out.

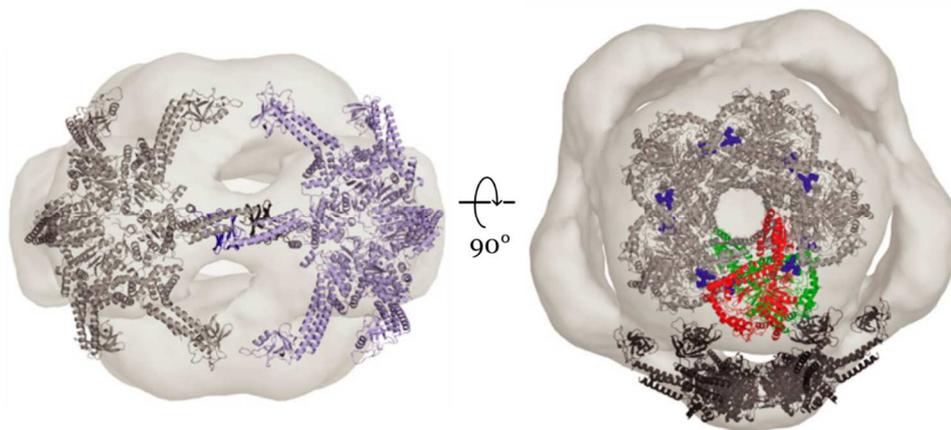


Figure 3.7. Proposed architecture of the LdcI-RavA complex, based upon the crystal structures of LdcI and RavA and the negative stain EM reconstruction of the complex. Adapted from El Bakkouri et al., 2010.

### 3.5. The cryo-EM structure of the LdcI-RavA cage

In 2014, the group published the D<sub>5</sub>-symmetric cryo-EM structure of the LdcI-RavA complex at an overall resolution of 11 Å (shown in Figure 3.8) which revealed the details of complex formation for the first time (Malet et al., 2014). Confirming the negative stain EM structure, the group showed that LdcI and RavA form a 3.3 MDa cage-like complex, composed of two parallel LdcI decamers bridged by five RavA hexamers surrounding a large central cavity. RavA interacts with LdcI through its LARA domains which bind either the top or bottom of the decameric ring. A C-terminal β-strand of LdcI is crucial for the LdcI-RavA interaction, as shown by a 7 Å cryo-EM structure of the isolated LARA domain in complex with LdcI. The sequence differences at this β-strand between LdcI and LdcC explain why RavA only interacts with LdcI (Kandiah et al., 2016; Malet et al., 2014). The complex is unlike any other formed by AAA+ ATPases, with lateral interactions between hexamers mediated by the triple-helical and LARA domains.

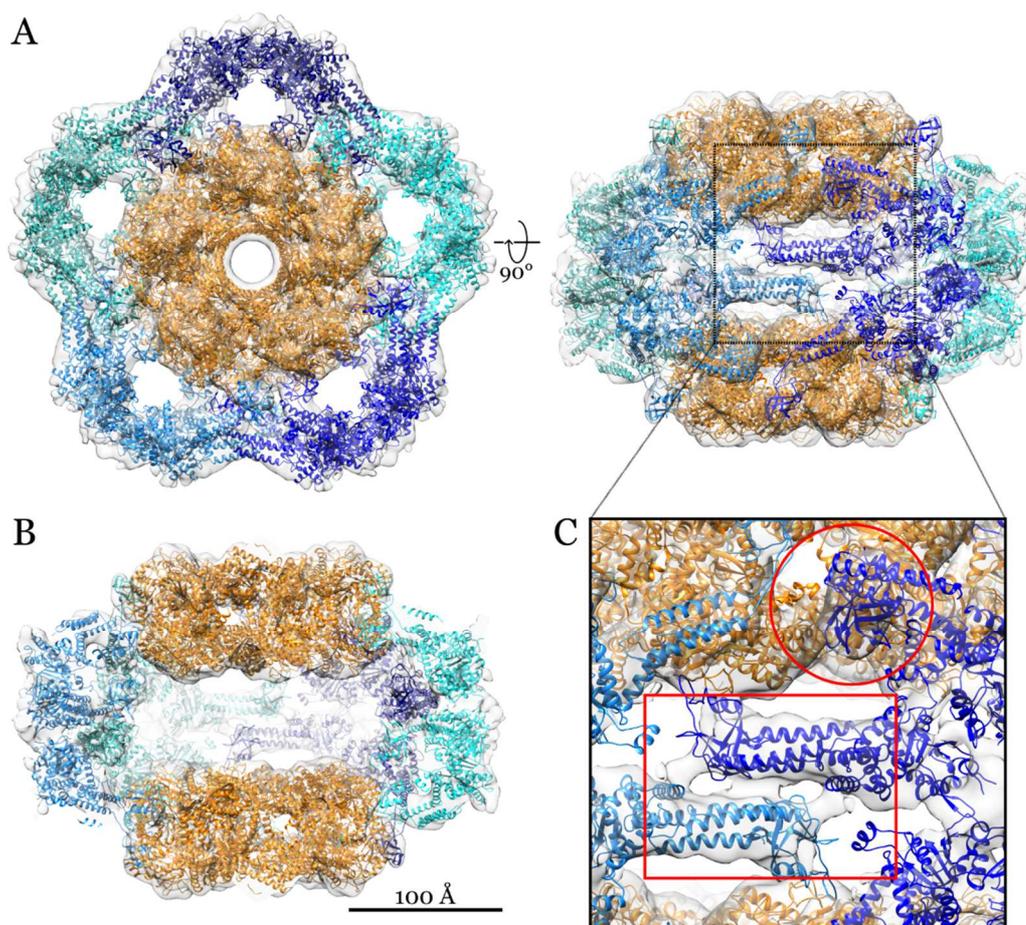


Figure 3.8. Cryo-EM map (EMDB ID: 2679, transparent grey) of the LdcI-RavA cage, with fitted atomic models (PDB ID: 4UPB). LdcI decamers are coloured orange, and individual RavA hexamers are coloured in shades of blue. **A)** Top (left) and side (right) views of the LdcI-RavA cage, displaying the overall arrangement of the complex. **B)** Cutaway side view of the LdcI-RavA cage, showing the large central cavity of the complex. **C)** Close-up view showing the lateral RavA-RavA triple-helical domain interactions (rectangle) and a RavA LARA-LdcI interaction (circle).

The central cavity of the LdcI-RavA cage is 220 Å in diameter and 80 Å in height, with a total volume of  $\sim 3 \times 10^6$  Å<sup>3</sup>. Based on the fact that RavA is a MoxR family AAA+ ATPase, a family which has known chaperone activity, the group proposed that the LdcI-RavA cage might act as a macromolecular chaperone for large substrates (Malet et al., 2014). It was hypothesised that the energy of ATP hydrolysis by RavA may thread substrates into the large central cavity for folding, possibly under acid stress conditions at which LdcI is expressed.

### 3.6. The biological role of the LdcI-RavA cage

In 2014, the same year as the cryo-EM structure of the cage was published, our collaborators from Walid Houry's group proposed potential substrates for RavA based on extensive genetic studies and pull-down experiments. Genome-wide genetic screening was carried out to reveal novel protein interactions in *E. coli*, suggesting that both RavA and ViaA were linked to the iron-sulphur (Fe-S) cluster biogenesis systems Isc and Suf (Babu et al., 2014). These are multicomponent systems involved in assembling Fe-S clusters, which are metal cofactors that are crucial for many fundamental cellular processes, including DNA repair and cellular respiration (Rouault, 2015). This suggested that, similarly to other MoxR-family proteins, RavA may be involved in metal cofactor insertion in conjunction with its VWA partner protein ViaA. Furthermore, immunoprecipitation was carried out and showed that both RavA and ViaA had physical interactions with Isc proteins.

The Houry group then showed using pull-down experiments that both RavA and ViaA had physical interactions with specific subunits of respiratory Complex I, the first complex in the electron transport chain (Wong et al., 2014a). Complex I, like other respiratory complexes, uses Fe-S clusters for shuttling electrons. It was proposed that RavA and ViaA would aid in the maturation of Complex I, both by inserting its Fe-S cofactors and facilitating the assembly of the catalytic domain. Interestingly, knockouts of *ravA**viaA* were also shown to desensitise cells to aminoglycoside antibiotics, which require correctly-matured and active Complex I for uptake (Girgis et al., 2009).

Another group showed that LdcI unexpectedly interacts with a partially-assembled Complex I that lacks its final Fe-S cluster (Erhardt et al., 2012). Therefore, it was hypothesised that the biological function of the LdcI-RavA-ViaA triad would be to facilitate the maturation of respiratory Complex I with the help of the Fe-S cluster biogenesis systems Isc and Suf. The large central cavity of the LdcI-RavA cage, which is more than large enough to fit an entire folded Complex I assembly, would act as a local environment for maturation of the Complex I catalytic domain under acid stress conditions. ViaA would bind to target substrates using its VWA protein interaction domain, and would act as a shuttle to bring substrates to the cage through its N-terminal domain, which was later shown to interact with RavA (Wong et al., 2017).

### 3.7. The goal of the thesis

The original rationale for this thesis was to analyse interactions between LdcI, RavA and ViaA and Fe-S cluster biogenesis proteins, from both a biochemical and structural point of view using cryo-EM. We expected that the LdcI-RavA cage would interact and possibly encapsulate the Fe-S cluster biogenesis system proteins and Complex I assembly intermediates, and could foresee exciting structures between the cage, its substrates and its interaction partners.

In parallel to studying the enterobacterial Complex I assembly triad LdcI-RavA-ViaA, we began a collaboration with Montserrat Soler-López at the European Synchrotron Radiation Facility in Grenoble on an seemingly functionally analogous triad in mitochondria composed of the proteins NDUFAF1, ACAD9 and ECSIT. These proteins are a part of the mitochondrial Complex I assembly (MCIA) complex, and are involved in the maturation of mitochondrial Complex I. I was involved in the structural characterisation of this triad during my PhD, which is presented in Chapters 8 and 9.

For the first year of my PhD, I worked to screen for interactions between RavA, ViaA and LdcI with the Isc and Suf systems using both biophysical and biochemical techniques, as well as by negative stain EM. In conjunction with a post-doc in the group Jan Felix, who was primarily working on validating the interactions between RavA, ViaA and LdcI with subunits of Complex I, we were unable to replicate any of the proposed interactions.

Therefore, my PhD was reoriented away from characterising the interactions with the proposed substrates to structure-functional relationships of the triad itself. The key goal of my thesis became the characterisation of the RavA hexamer by cryo-EM. The original negative stain EM model showed a six-fold symmetric hexamer, as was expected for many AAA+ ATPases at the time. After I started my PhD, an updated structure of the LdcI-RavA cage from Benoît Arragain in the team showed that the RavA hexamer, when bound to LdcI, adopts an asymmetric spiral conformation (as presented in Chapter 4). This came at a time when the asymmetric conformations of AAA+ ATPases were novel, and we therefore decided to use cryo-EM to investigate the conformation of the RavA hexamer alone. Indeed, no previous characterisation of MoxR ATPases had been carried out by cryo-EM, and it was therefore pertinent to investigate whether MoxR ATPases form similar oligomeric structures to AAA+ proteins from other clades, to understand their ATP hydrolysis mechanism.

The proposed functional role of the LdcI-RavA-ViaA triad as a chaperone for Fe-S cluster insertion into respiratory complexes seemed to not be reproducible based on our experiments. In order to gain clues into the true cellular functions of these proteins, we decided to analyse the physiological distribution of the triad in the cell by super-resolution fluorescence microscopy. For this, I purified fluorescently-tagged LdcI and RavA proteins and characterised them by negative stain EM to assess the suitability of constructs for carrying out super-resolution fluorescence microscopy, which was then performed by Clarissa Liesche, a post-doc in the group. The results of this work are presented in the manuscript in Chapter 6.

There were also other experiments that I carried out on proteins of the triad, which are not presented in the context of this thesis. Altogether the results obtained here, in parallel with those obtained by other members of the group, are starting to paint a picture of how this enigmatic triad functions in bacteria, from both a structural and functional point of view.

## II. RESULTS

## Chapter 4. Manuscript on RavA and the LdcI-RavA cage

In 2017, a new cryo-EM structure of the LdcI-RavA cage was solved by Benoît Arragain, a Masters student at the time, without the imposition of D<sub>5</sub> symmetry. In the resulting map, it was clear that RavA formed asymmetric spirals in the context of the cage. These spiral hexamers were constrained to two orientations inside the cage, raising questions about how the ATPase activity of RavA would occur in the LdcI-RavA complex. At the same time, a wealth of new AAA+ ATPase structures were being solved by cryo-EM that also showed asymmetric spiral conformations.

In light of these results, we decided to investigate the structure of the RavA hexamer using cryo-EM, in order to understand whether the conformation seen in the LdcI-RavA cage was also present for unbound RavA. I found that RavA exists in two distinct conformations in solution – an open spiral conformation, similar to the conformation seen in the LdcI-RavA cage, but also a two-fold symmetric closed ring conformation that had not previously been seen for AAA+ ATPases by cryo-EM. The manuscript in Chapter 4 presents both the updated structure of the LdcI-RavA cage and the cryo-EM structures of the RavA hexamer, along with biochemical analysis that helps to investigate the ATPase mechanism of RavA.

ARTICLE

<https://doi.org/10.1038/s42003-020-0772-0>

OPEN

# Structural insights into ATP hydrolysis by the MoxR ATPase RavA and the LdcI-RavA cage-like complex

Matthew Jessop <sup>1,3</sup>, Benoit Arragain <sup>1,3</sup>, Roger Miras<sup>2</sup>, Angélique Fraudeau<sup>1</sup>, Karine Huard<sup>1</sup>, Maria Bacia-Verloop<sup>1</sup>, Patrice Catty<sup>2</sup>, Jan Felix <sup>1\*</sup>, Hélène Malet <sup>1\*</sup> & Irina Gutsche <sup>1\*</sup>

The hexameric MoxR AAA+ ATPase RavA and the decameric lysine decarboxylase LdcI form a 3.3 MDa cage, proposed to assist assembly of specific respiratory complexes in *E. coli*. Here, we show that inside the LdcI-RavA cage, RavA hexamers adopt an asymmetric spiral conformation in which the nucleotide-free seam is constrained to two opposite orientations. Cryo-EM reconstructions of free RavA reveal two co-existing structural states: an asymmetric spiral, and a flat C2-symmetric closed ring characterised by two nucleotide-free seams. The closed ring RavA state bears close structural similarity to the pseudo two-fold symmetric crystal structure of the AAA+ unfoldase ClpX, suggesting a common ATPase mechanism. Based on these structures, and in light of the current knowledge regarding AAA+ ATPases, we propose different scenarios for the ATP hydrolysis cycle of free RavA and the LdcI-RavA cage-like complex, and extend the comparison to other AAA+ ATPases of clade 7.

<sup>1</sup>Institut de Biologie Structurale, Univ. Grenoble Alpes, CEA, CNRS, IBS, 71 Avenue des martyrs, F-38044 Grenoble, France. <sup>2</sup>Laboratoire de Chimie et Biologie des Métaux, Univ. Grenoble Alpes, CEA, CNRS, DRF, IRIG, UMR 5249, 17 rue des Martyrs, F-38054 Grenoble, France. <sup>3</sup>These authors contributed equally: Matthew Jessop, Benoit Arragain. \*email: [jan.felix@ibs.fr](mailto:jan.felix@ibs.fr); [helene.malet@ibs.fr](mailto:helene.malet@ibs.fr); [irina.gutsche@ibs.fr](mailto:irina.gutsche@ibs.fr)

**A**AA+ ATPases of the MoxR family are ubiquitous and found in all major phyla of bacteria and archaea. They are proposed to fulfil chaperone-like functions assisting the maturation or assembly of metabolic protein complexes<sup>1,2</sup>, and are often found in an operon upstream of a gene encoding a von Willebrand factor A (VWA) domain-containing protein. Recent examples include the *P. denitrificans* genes *norQ* and *norD*, which code for a MoxR ATPase and a VWA domain-containing protein facilitating the insertion of the non-heme Fe<sub>B</sub> cofactor into nitric oxide reductase<sup>3</sup>, and the *A. ferrooxidans* MoxR-related protein CbbQ which binds the VWA domain-containing CbbO to activate Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco)<sup>4,5</sup>. The most well-characterised representative of the MoxR family is the *E. coli* ATPase RavA, encoded by the *ravA*/*viaA* operon, together with the VWA domain-containing protein ViaA. These two proteins were proposed to play a role in the maturation of both respiratory Complex I and fumarate reductase<sup>6,7</sup>. In addition, RavA is involved in the *E. coli* acid stress response by binding to the acid stress-inducible lysine decarboxylase LdcI<sup>8,9</sup>. LdcI catalyses the conversion of lysine into cadaverine, thereby consuming a proton and buffering both the intra- and extracellular medium<sup>10,11</sup>. Under conditions of combined acid and nutrient stress, LdcI is inhibited by the stringent response alarmone ppGpp, preventing excessive consumption of lysine<sup>12</sup>. However, binding of RavA to LdcI was shown to alleviate this inhibition<sup>8</sup>. Remarkably, RavA and LdcI together form a 3.3 megadalton cage-like complex, consisting of two D5-symmetric decameric LdcI rings located at the top and bottom of the cage, surrounded by five RavA hexamers<sup>9</sup>.

Combined with crystal structures of the LdcI decamer (PDB ID: 3N75) and the RavA monomer (PDB ID: 3NBX), our first low resolution cryo-electron microscopy (cryo-EM) map of the LdcI–RavA cage, performed imposing the D5 symmetry of the LdcI onto the whole assembly (EMD-2679), provided initial insights into the elements involved in the complex formation<sup>13</sup>. Specifically, rotations of the C-terminal arms of RavA with respect to the N-terminal AAA+ ATPase modules, and accompanying massive reorientation of the tip domains called LARA (LdcI Associating domain of RavA), were shown to mediate RavA binding to either LdcI or adjacent RavA monomers in the cage<sup>13</sup>. The lateral contacts observed between neighbouring RavA hexamers in the LdcI–RavA complex are unique amongst AAA+ ATPases.

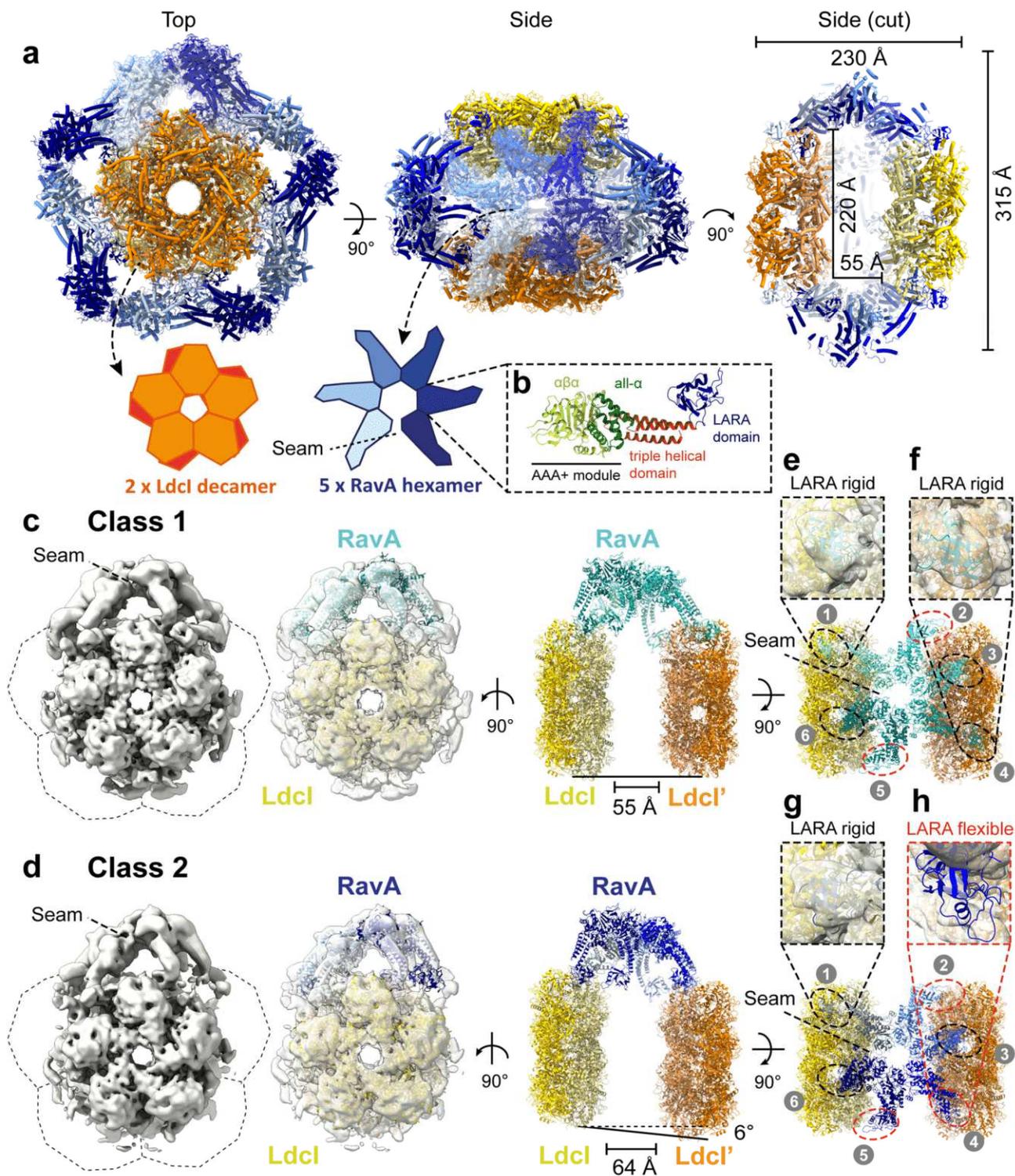
Building further upon these results, we now present a higher resolution cryo-EM structure of the LdcI–RavA cage in the presence of ADP, obtained without symmetry imposition. We show that the complex is built by five RavA hexamers arranged into spirals, with a prominent gap (or “seam”) between two LdcI-binding RavA monomers facing either the top or the bottom LdcI decamer. Spiral conformations have recently been observed for AAA+ ATPases such as katanin, Vps4, Hsp104, ClpB and Lon<sup>14–17</sup>, but have not yet been described for the MoxR family. In addition, cryo-EM analysis of free RavA in the presence of ADP reveals the presence of two distinct conformational states: a RavA spiral containing a single seam, equivalent to the one inside the LdcI–RavA cage, and a planar C2-symmetric ring with two nucleotide-free seams at opposite positions in the RavA hexamer. This second conformation may represent an intermediate state between the “seam up” and “seam down”-oriented RavA spirals inside the LdcI–RavA complex. Moreover, it displays remarkable structural similarity to the approximately two-fold symmetric “dimer of trimers” arrangement of subunits in crystal structures of the extensively studied AAA+ unfoldase ClpX<sup>18,19</sup> and the protein-remodeling AAA+ ATPase PCH2<sup>20</sup>. Consequently, the mechanism of the RavA ATPase cycle may be unexpectedly similar to the meticulously dissected ATP hydrolysis cycle

of ClpX<sup>19,21–23</sup>, although the respective families of these two proteins belong to different clades of AAA+ ATPases<sup>24–26</sup>. Finally, we characterise the LdcI–RavA interaction using bio-layer interferometry (BLI) binding studies and ATPase activity assays. We demonstrate that while the affinity of LdcI for RavA is pH-independent, LdcI-binding results in an increase in RavA ATPase activity at acidic pH, at which this complex should be formed inside the cell. Based on these results, we propose different possible scenarios for the ATP hydrolysis cycle of RavA, both alone and in the context of the LdcI–RavA cage, and discuss their functional implications.

## Results

**The LdcI–RavA cage is formed by spiral RavA hexamers.** Initial attempts to reconstruct the LdcI–RavA complex by imposing C5 symmetry resulted in maps with visible heterogeneity for the five RavA copies (Supplementary Fig. 1). Therefore, we applied a symmetry expansion procedure (Supplementary Fig. 1, Methods), followed by a masked 3D classification without angular search using a soft mask focussing on one RavA hexamer and two LdcI decamers. This resulted in two essentially identical classes, apart from a 180° rotation around the centre of the RavA hexamer. These classes displayed left-handed spiral RavA hexamers containing a seam pointing either to the top (orientation A) or bottom (orientation B) LdcI decamer in the cage (Fig. 1, Supplementary Fig. 1). The particles from orientation A and B were grouped together after applying a 180° rotation to orientation B, and used in a second masked 3D refinement. To account for observed heterogeneity in the LARA domains of RavA, a final round of 3D classification was carried out followed by 3D refinement. The resulting two classes, with an overall resolution of 7.6 Å and 7.8 Å, respectively (Fig. 1, Supplementary Fig. 2, Supplementary Table 1), both display spiral RavA hexamers bound to LdcI, but show clear differences in the presence (Class 1) or absence (Class 2) of density for one specific LARA domain (Fig. 1c and d, panels f and h). The distance between the two LdcI decamers is about 10 Å larger in Class 2, which displays a 6° tilt between opposite LdcI rings in contrast to the parallel position of LdcI copies in Class 1. Both classes show a higher overall resolution for LdcI (5–6 Å), compared to RavA (12 Å), likely originating from the inherent flexibility of the RavA spirals in the cage (Supplementary Fig. 2). The complete LdcI–RavA cage is formed by two parallel LdcI decamers surrounded by five hexameric RavA spirals. Therefore, to illustrate the overall architecture of the LdcI–RavA cage, we constructed a C5-symmetrised map of Class 1 (Methods). Each RavA hexamer harbours six binding interfaces: two lateral interactions with neighbouring RavA monomers mediated by the triple helical domain of RavA (Fig. 1b), and two interactions per LdcI decamer mediated by the LARA domains at the end of the four other RavA monomers (Fig. 1). The resulting map differs dramatically from the previously published one<sup>13</sup> (EMD-2679, PDB ID: 4UPB) that was calculated with a D5 symmetry inherent to LdcI, thereby leading to a distortion of the RavA spirals into a C2-symmetrical assembly.

A pseudo-atomic model of the cage was then created by flexibly fitting crystal structures of LdcI and RavA into the maps of Class 1 and 2 using iMODFIT<sup>27</sup> (Methods). The positions of the LARA domains of RavA contacting LdcI were inferred from a cryo-EM map of the LdcI–LARA complex (EMD-3206)<sup>28</sup>. In contrast to what was anticipated from the LdcI–LARA cryo-EM map, the crystal structure of the LdcI decamers (PDB ID: 3N75)<sup>12</sup> remained virtually unchanged upon fitting into the map of the LdcI–RavA complex, indicating that RavA binding does not affect the LdcI conformation. As for the atomic model of RavA to be used for flexible fitting into the spiral RavA density inside the



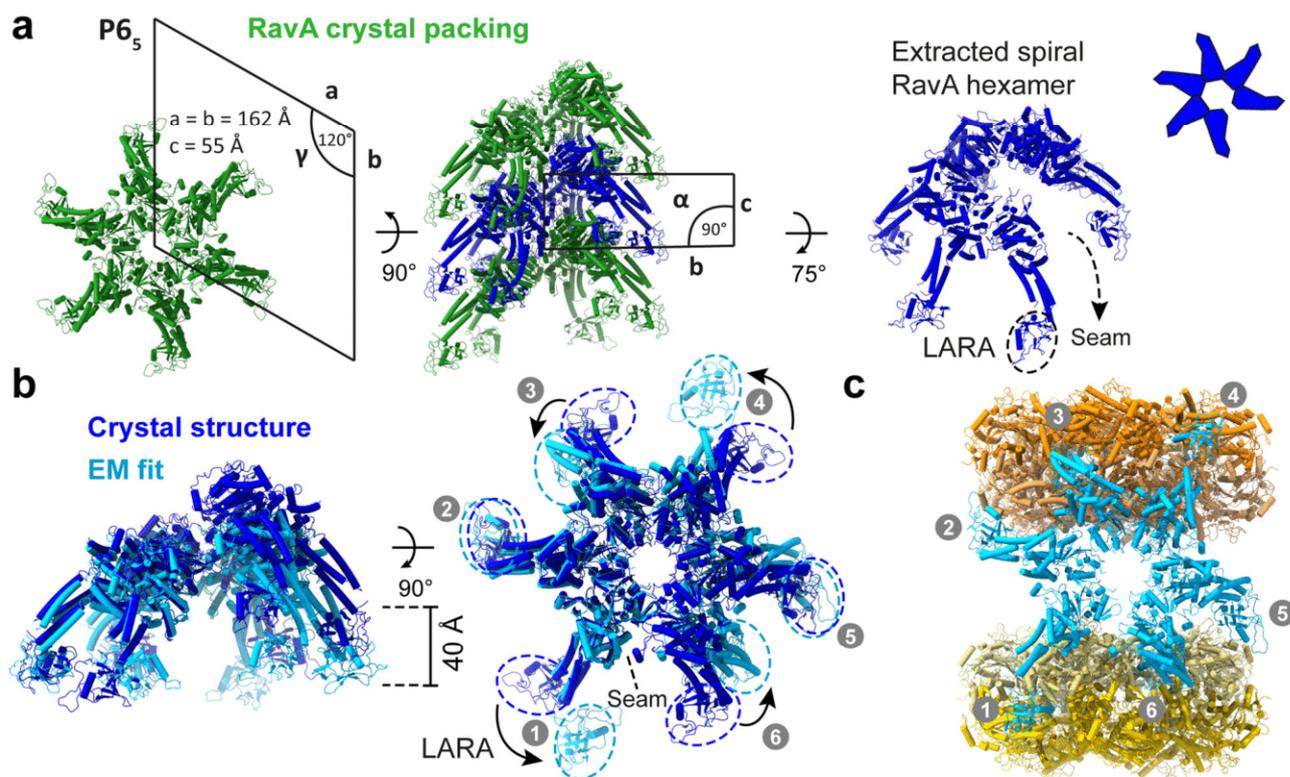
**Fig. 1 Cryo-EM structure of the Ldcl-RavA cage-like complex.** **a** Pseudo-atomic model of the Ldcl-RavA complex, based on flexible fitting of crystal structures of RavA (PDB ID: 3NBX)<sup>8</sup> and Ldcl (PDB ID: 3N75)<sup>12</sup>. The cryo-EM map used for fitting corresponds to the “Class 1” map after 3D classification (containing two Ldcl decamers and one RavA hexamer, see panel **b**) to which C5 symmetry has been applied. Two Ldcl decamers (coloured yellow and orange) and five spiral RavA hexamers (individually coloured light to dark blue) are shown as cartoons. Top and side views are shown, as well as a cut-away side view displaying the inner cavity of the cage. A dashed box (**b**) shows one RavA monomer with annotations for the different domains: AAA+ module (green), triple helical domain (red), and LARA domain (blue). The seam in the spiral RavA hexamer is indicated by a dashed line. **c, d** Classes 1 (**c**) and 2 (**d**) obtained after C5 symmetry expansion followed by a masked 3D classification without angular search in RELION-2.0, resulting in C1 asymmetric maps. For each class, a post-processed cryo-EM map is shown (left) along with a fit of two Ldcl decamers (yellow and orange) and one RavA hexamer (Class 1: cyan, Class 2: dark blue). Dashed lines indicate the positions of the masked-out RavA hexamers during symmetry expansion. On the right, side and top views of the fits are shown, with panels (**e**)–(**h**) (dashed boxes) focusing on specific LARA domains (numbered 1–6, black circles: rigid, red circles: flexible) contacting Ldcl and their corresponding fits in the EM map. Class 2 displays a 6° tilted orientation of the second Ldcl decamer (coloured orange) compared to Class 1.

cage, we decided to reexamine the crystal packing in the RavA structure (PDB ID: 3NBX)<sup>8</sup>. Indeed, while RavA was reported to crystallise as a monomer, analysis of the crystal packing in space group  $P6_5$  reveals a continuous left-handed RavA helix (Fig. 2a) with ADP molecules bound at the intersubunit interface (Supplementary Fig. 3). This interface is essentially equivalent to the one that we originally inferred from a fit of the RavA monomer crystal structure into a low resolution C6-symmetric negative stain EM map constrained by a comparison with other AAA+ ATPases<sup>8</sup>. Specifically, ADP is coordinated by the Walker-A and -B residues A51, K52, S53 and D114 from one RavA monomer and by Sensor 2 and R-finger residues (RavA R251 and R170, respectively) from another RavA monomer<sup>8</sup> (Supplementary Fig. 3). These observations favour the idea that the crystallographic RavA–RavA interface constitutes (or at least closely resembles) the biological interface. Therefore, we opted for the usage of a spiral RavA hexamer generated from the RavA crystal structure as a starting point for flexible fitting in the cryo-EM maps. Thus, despite the overall low local resolution of RavA in the LdcI–RavA cryo-EM map, the combined use of crystal structures of LdcI decamers, RavA spiral hexamers and a cryo-EM map of LdcI–LARA allows us to confidently model the LdcI–RavA complex.

A comparison between a RavA hexamer generated from the crystal structure and RavA fitted in the cryo-EM map of Class 1 is shown in Fig. 2b. While the RavA–RavA interface is retained after fitting, major differences are observed in the pitch of the

RavA spiral, and in the positions of the four LARA domains contacting LdcI. Indeed, these LARA domains (numbered 1, 3, 4 and 6 in Fig. 2b, c) undergo massive rotations when compared to the crystal structure, whereas the LARA domains of the RavA monomers involved in lateral RavA interactions only show minor movements. The apparent flexibility of RavA in our cryo-EM maps is accentuated by the disappearance of density for LARA domain 4 in Class 2, while it is clearly present in Class 1 (Fig. 1c, d, panels e–h).

**RavA seams are oriented up or down in the LdcI–RavA cage.** In the LdcI–RavA complex, the RavA spiral seam is oriented between the two RavA monomers that have LARA domains interacting with LdcI. This results in a seam that always faces either towards the upper (orientation A: “seam up”) or lower (orientation B: “seam down”) LdcI decamer, and never towards adjacent RavA hexamers (Fig. 1, Supplementary Fig. 1). Generally, for AAA+ ATPases forming spiral assemblies, the progressive movement of the seam around the hexameric ring is shown to be functionally important<sup>14,15,29–32</sup>. Therefore, the occurrence of only two opposite seam orientations of RavA spirals inside the LdcI–RavA cage may be explained as follows: (i) the lateral RavA–RavA interactions impose local geometrical constraints causing ATP hydrolysis to occur solely at the active sites formed between RavA monomers 3–4, and 1–6 in the hexamer (Fig. 2c), or (ii) binding to LdcI stalls RavA in an inactive form by preventing ATP hydrolysis from proceeding



**Fig. 2 Comparison between a RavA hexamer generated from the RavA crystal structure and a fit in the cryo-EM map of the LdcI–RavA complex.** **a** Crystal structure of *E. coli* RavA (PDB ID: 3NBX)<sup>12</sup>, displayed as cartoons, showing the helical crystal packing of RavA crystallised in spacegroup  $P6_5$ . A top view along the helical screw axis of the assembly (left, with annotated unit-cell parameters) resembles a RavA hexamer. A side view of the helical assembly is shown as well (middle, with annotated unit-cell parameters), with one spiral RavA hexamer coloured dark blue. On the right, an extracted spiral hexamer (dark blue) is displayed with a slight tilt to allow visualisation of the seam, along with a schematic representation. **b** Comparison of a spiral RavA hexamer extracted from the crystal structure (dark blue) and a fit of RavA in the cryo-EM map of Class 1 (light blue). Side (left) and top views (right) are shown, with dashed circles around the LARA domains (numbered 1–6) to highlight the differences between the crystal structure and EM fit. The position of the seam is indicated by a dashed line. **c** LdcI–RavA complex obtained after fitting of structures of RavA (light blue) and LdcI (yellow and orange) in the cryo-EM map of Class 1, with LARA domains numbered as in (b).

in a progressive fashion. The former hypothesis would imply a switching between the two observed orientations of the RavA spiral seam and would therefore contradict a strictly sequential ATPase cycle for RavA. In addition, there remains a possibility that other seam orientations of RavA spirals in the LdcI–RavA cage also exist but are less stable and thus not resolved in our analysis because of the limited number of particles used. In light of the observed spiral conformation of RavA inside the LdcI–RavA complex, we then revisited the structure of the free RavA hexamer. Indeed, the previous negative stain EM reconstruction of the hexameric RavA was performed with a C6 symmetry imposed<sup>8</sup>, based on planar symmetric hexamers observed in numerous AAA+ ATPase crystal structures available at the time<sup>2,8,33–35</sup>.

### Free RavA has a spiral and a flat C2-symmetric conformation.

Initial processing of the RavA-ADP dataset indicated a strongly preferred top-view orientation of the particles on the cryo-EM grid, resulting in a nonuniform distribution of angular projections. A second dataset was therefore collected at a 30° tilt (see “Methods”). 2D class averages from the combined dataset revealed the presence of two distinct conformational states of the RavA hexamer, containing either one (Fig. 3a, b, blue squares) or two gaps (Fig. 3a, b, red squares). 3D classification and refinement (see “Methods”, Supplementary Figs. 4 and 5, Supplementary Table 1) resulted in two maps (at overall resolution of 7 and 6 Å, respectively), corresponding to an open spiral with a single nucleotide-free seam (Fig. 3c, Supplementary Figs. 4 and 5) and a planar C2-symmetric closed ring characterised by two nucleotide-free seams at opposite positions (Fig. 3d, Supplementary Figs. 4 and 5). These medium resolution maps are sufficient to enable the high-confidence flexible fitting of a RavA hexamer (for the spiral) or a trimer (for the C2-symmetric ring) extracted from the RavA helix generated from the crystal structure (Fig. 4a, b, Supplementary Fig. 6). Noteworthy, while caution in interpreting fine details at the level of the interfaces and loop regions should still be exercised, rigid-body fitting of RavA monomers into the two maps without any prior knowledge leads to an interface virtually identical to the one observed in the RavA crystal structure, further validating the RavA intersubunit interface.

The spiral structure of free RavA is equivalent to its conformation in the LdcI–RavA complex (Supplementary Fig. 6a), except for the lack of a defined density for the LARA domains that are flexible in solution before binding to LdcI. The map contains densities attributed to five ADP molecules bound to the interface between each contacting RavA monomer (Fig. 4a, c–e). In contrast, the C2-symmetric conformation contains only four ADP molecules bound in the active sites between subunits 1–2, 2–3, 4–5 and 5–6 (Fig. 4b, f). The 27 Å gap between subunits 1 and 6 in the spiral conformation is much wider than the seams in the closed-ring conformation, meaning that rearrangements between subunits 1–6 and 3–4 that destroy the nucleotide-binding site are more subtle (Supplementary Fig. 6b). Loss of nucleotide binding mainly results from a rigid-body rotation of RavA monomers 1 and 4, accounted by shifts of helix  $\alpha 3$  and its preceding loop, which contains the Walker-A residues A51, K52 and S53, and helix  $\alpha 7$ , which contains residue M189. All of these residues directly interact with ADP in the intact active site interface (Supplementary Fig. 3, Supplementary Fig. 6c, d).

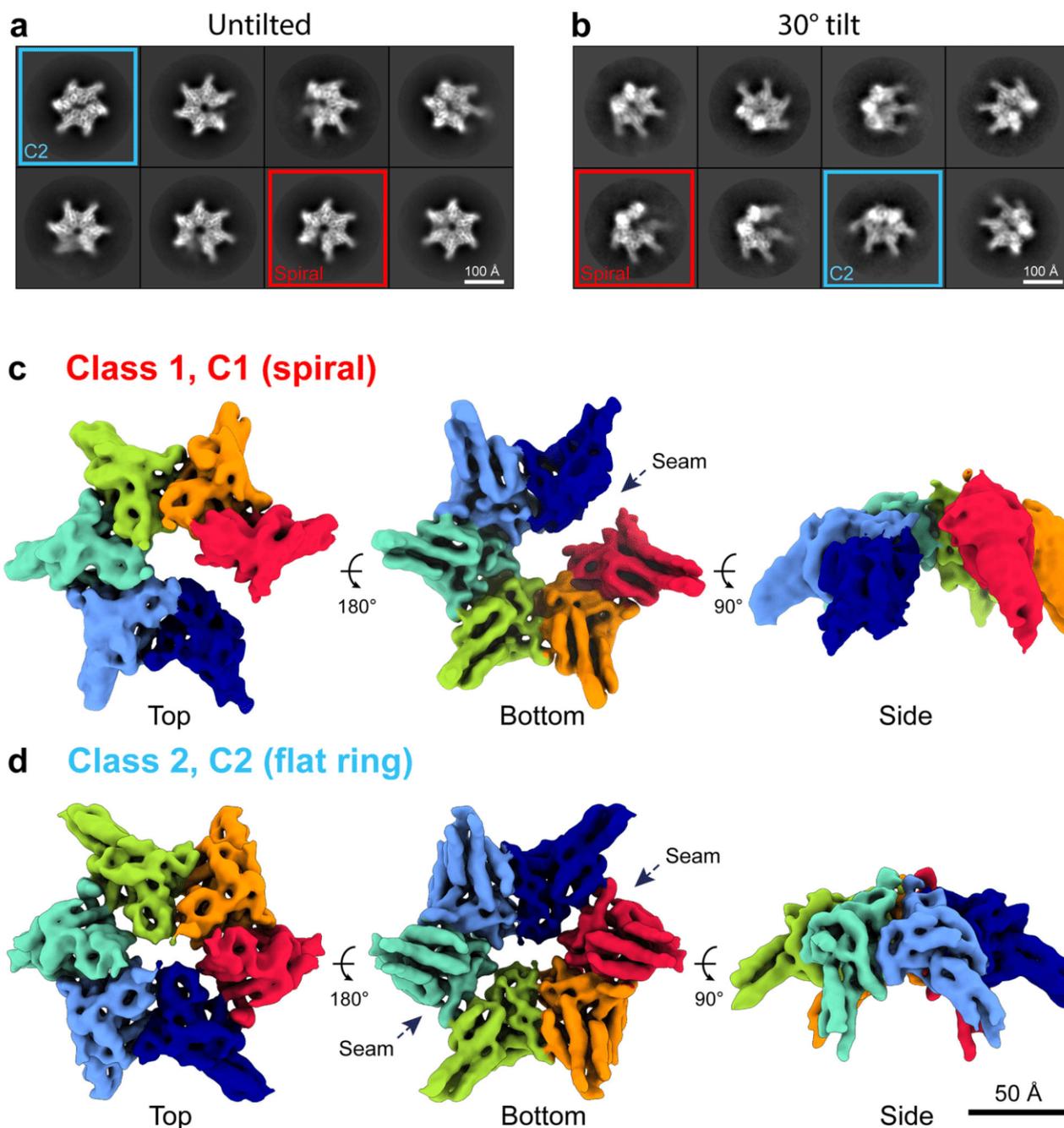
In an attempt to mimic the active ATP-bound state of the RavA hexamer, a cryo-EM dataset was collected on free RavA in the presence of ATP $\gamma$ S, a slowly-hydrolysable ATP analogue often used to stabilise the ATP-bound state of ATPases. However, 2D class averages indicated that RavA-ATP $\gamma$ S displays even more conformational heterogeneity than RavA-ADP (Supplementary

Fig. 7). Both the C2-symmetric and spiral conformations are present, as well as a seemingly C6-symmetric ring and even a C7-symmetric oligomer. The biological relevance of these two additional states is uncertain (see for instance Sysoeva, 2017<sup>36</sup> for review), and this heterogeneity, coupled with a strongly preferential orientation, hampers successful 3D separation. However, a comparison of 2D class averages displaying the asymmetric spiral and C2-symmetric closed ring conformations of RavA-ADP and RavA-ATP $\gamma$ S datasets does not reveal any noticeable differences. Most importantly, this observation suggests that, similarly to ADP, ATP $\gamma$ S binding does not fix RavA in one particular structural state.

**Structural insights into the ATPase cycle of RavA.** The crystal structure of the RavA monomer<sup>8</sup> corroborated the former phylogeny-based classification of the MoxR family as a member of the AAA+ clade 7<sup>24–26,37</sup>. This clade harbours in particular an additional linker (termed the pre-Sensor 2 insertion) that repositions the C-terminal helical lid of the AAA+ module relative to the N-terminal  $\alpha\beta\alpha$  core domain<sup>25,26</sup>. In such a spatial configuration, different from all other AAA+ clades and unique to clade 7, the Sensor 2 motif cannot contribute to ATP binding and hydrolysis in the same monomer (Supplementary Fig. 8). However, based on the crystal structure of the first crystallised clade 7 member, the magnesium chelatase BchI monomer (PDB ID: 1G8P)<sup>38</sup>, aligned onto active hexamers from other clades, AAA+ ATPases of clade 7 were proposed to rely on a trans-acting Sensor 2 contributed by the neighbouring monomer in the hexamer<sup>25</sup>. The first pseudo-atomic model of the RavA hexamer, which was based on a fit of the monomer structure into a negative stain C6-symmetric EM map<sup>2,8</sup> and guided notably by relative positioning of Sensor 2, agreed with this hypothesis and suggested that oligomerisation of MoxR ATPases is required for completion of their ATP binding sites. Specifically, in MoxR-type AAA+ ATPases, the ATP binding site was proposed to be located not between the large ( $\alpha\beta\alpha$ ) and small (all- $\alpha$ ) AAA+ domains of the same monomer, but between the large domain of one monomer in a hexamer and the small domain of its neighbour<sup>2,8</sup>. The cryo-EM maps presented here and the resulting atomic models of the RavA hexamer in spiral and C2-symmetric conformations provide strong experimental support to this model, which is presently extended to all clade 7 members<sup>25,26</sup>.

Viewed from this perspective, the planar double-seam conformation of the RavA hexamer is strikingly reminiscent of the approximately two-fold symmetric “dimer of trimers” arrangement of subunits in hexamers of the AAA+ unfoldase ClpX<sup>18,19</sup>, which belongs to clade 5 AAA+ ATPases and thus lacks the pre-Sensor 2 insertion<sup>25,26</sup>. In crystal structures of ClpX, hexamers are arranged with an approximate two-fold symmetry, and contain four ClpX subunits in a nucleotide loadable (L) and two in unloadable (U) conformation on opposite sides of the hexamer. In the unloadable ClpX subunits, the small and large AAA+ domains are positioned in an “open” conformation which destroys the nucleotide-binding site<sup>18,19</sup>. The resulting 4L-2U arrangement of ClpX contains a characteristic seam which runs along the hexamer centre. A comparison of the C2-symmetric closed ring conformation of RavA with the 4L-2U ClpX crystal structure reveals highly similar assemblies (Fig. 5a, b). While ClpX binds nucleotides in the interface formed between the large and small AAA+ domains within one subunit, the nucleotide-binding interface in RavA hexamers is formed in between adjacent monomers (Fig. 5c, d). Importantly, our structural comparison shows that similar rigid-body like movements between the large and small AAA+ subdomains in a single ClpX subunit, or between the large or small subdomains of adjacent RavA monomers, lie at the basis of

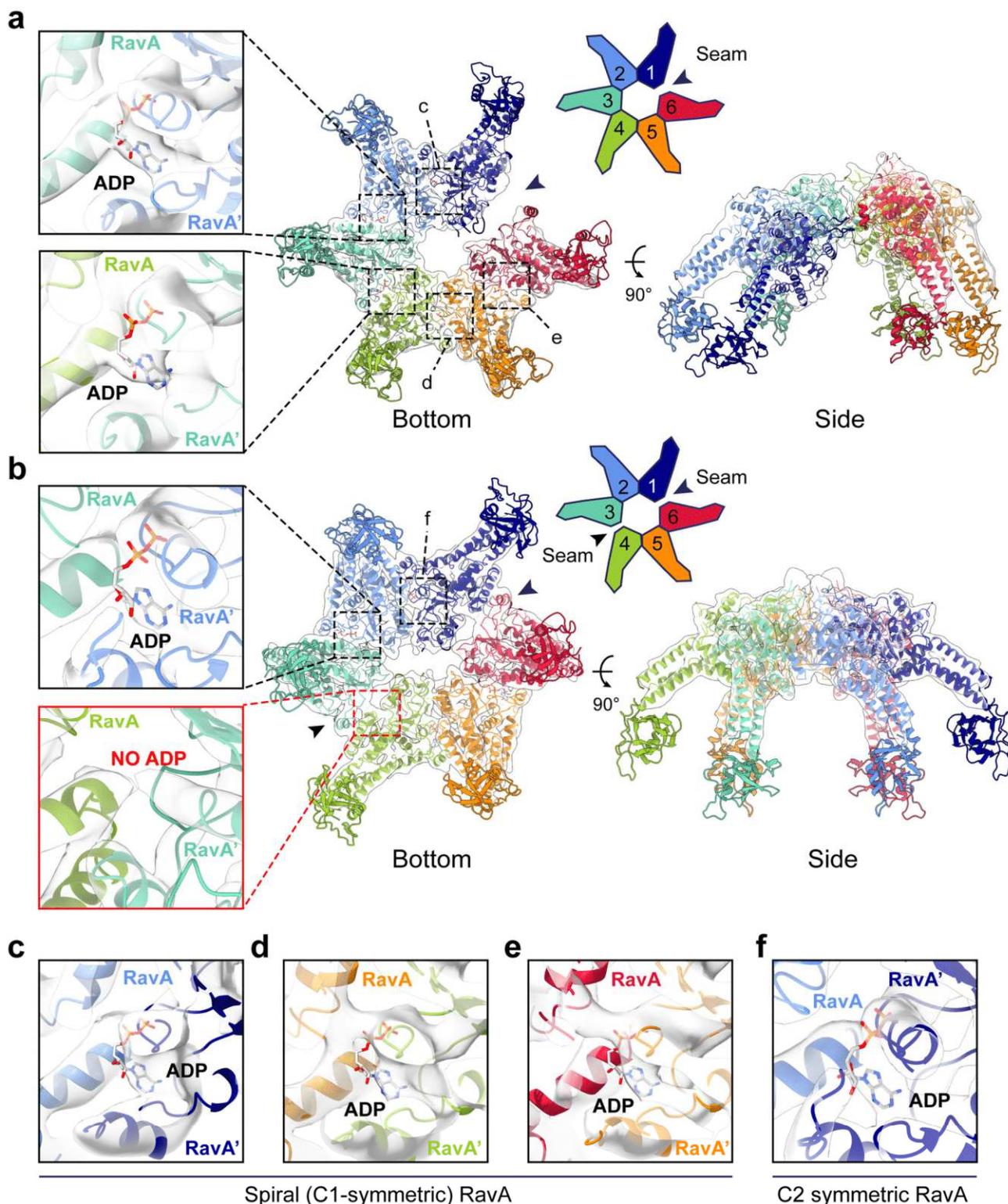
## 2D class averages



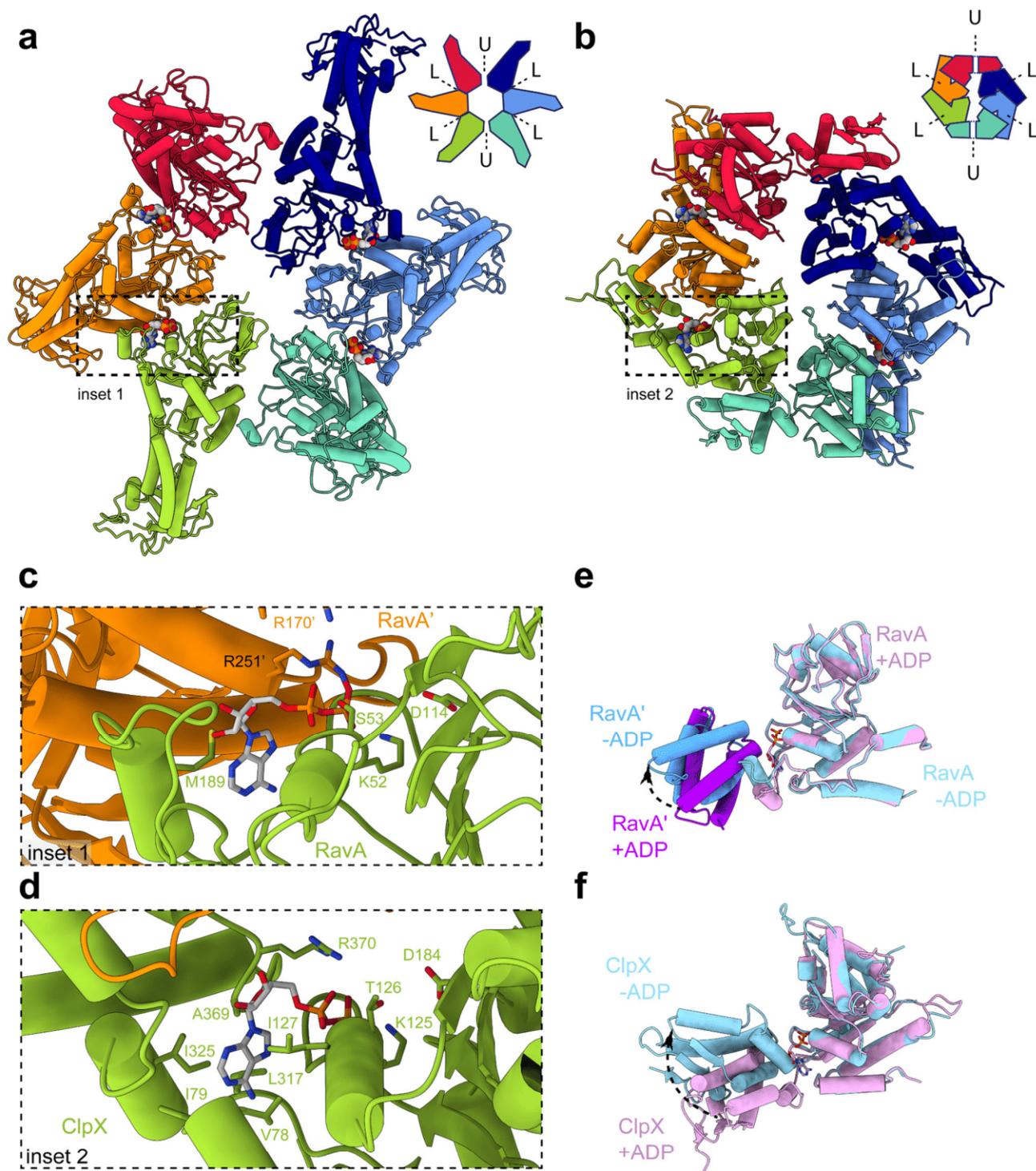
**Fig. 3** Cryo-EM analysis of free RavA in the presence of nucleotide (ADP). **a, b** 2D classes of untitled (**a**) and 30° tilted (**b**) datasets of free RavA in the presence of ADP. Red squares highlight classes belonging to a spiral RavA conformation, while blue squares show classes belonging to a C2-symmetric closed ring conformation of RavA. **c, d** 3D reconstructions of the spiral (**c**) and closed ring (**d**) RavA conformations corresponding to class 1 and 2, respectively. Individual subunits in the maps are coloured according to a rainbow colour scheme. The nucleotide-free seams in the two maps are annotated using dotted arrows.

L to U subunit conversion, resulting in an impaired nucleotide-binding site (Fig. 5e, f). In addition, similar C2-symmetric closed ring hexamer conformations have been observed for the crystal structures of the *T. maritima* metalloprotease FtsH<sup>39</sup>, and the *C. elegans* protein-remodeling AAA+ ATPase PCH2, a TRIP13 ortholog<sup>20</sup>. Interestingly, a recent structure of human TRIP13 solved by cryo-EM in both apo- and substrate-bound states<sup>40</sup> displays a right-handed spiral, but no closed-ring conformation.

Previous studies have shown that conversion between L- and U-states is necessary to couple ATP hydrolysis to ClpX functioning, and provide evidence for a probabilistic model for L to U subunit switching upon ATP hydrolysis. In the proposed model, a 4L:2U ClpX hexamer converts to a 5L:1U hexamer in the presence of nucleotide, followed by subunit switching between L and U states in a non-sequential manner<sup>19</sup>. The cryo-EM analysis of RavA presented here reveals the presence of a mixture of both a spiral (5L:1U) and C2-symmetric closed ring conformation



**Fig. 4 Structural analysis of spiral and C2-symmetric closed ring RavA conformations.** Fit of the spiral (a) and C2-symmetric closed ring (b) conformations of RavA in their respective EM maps, displayed as cartoons. Individual RavA subunits, labelled 1-6 in the accompanying schematic representations, are coloured according to a rainbow colour scheme. Zooms show the presence or absence of ADP in the nucleotide-binding site interface formed between subunits 2-3 and 3-4 in the spiral (a) and C2-symmetric closed ring (b) conformations of RavA. The nucleotide-free seams in the two maps are annotated using black arrows. **c-f.** Insets showing the nucleotide-binding site interface formed between subunits 1-2 (c), 4-5 (d) and 5-6 (e) of the spiral RavA conformation, and between subunits 1-2 (f) of the C2-symmetric closed ring RavA conformation.



**Fig. 5 Comparison between C2-symmetric closed ring conformations of RavA and ClpX.** Comparison between closed ring conformations of RavA (**a**) and ClpX (**b**), shown as cartoons with accompanying schematic representations. RavA and ClpX subunits in equivalent positions around the hexamer are given identical colours following a rainbow colour scheme. In a RavA hexamer, the active site is formed in between the large and small AAA+ domains of adjacent RavA monomers, while in ClpX the nucleotide-binding site is formed between the large and small AAA+ domains of within a single ClpX subunit. Loadable and unloadable ATP binding sites in RavA and ClpX are annotated with L and U, respectively. **c, d** Zooms of the nucleotide-binding interface between adjacent RavA monomers (**c**, coloured orange and green) and the nucleotide-binding interface within one ClpX subunit (**d**, coloured green). Bound ADP molecules and interacting residues of RavA or ClpX are labelled and shown as sticks. **e** Superposition of the large and small AAA+ domains of adjacent RavA subunits (labelled RavA and RavA') from an interface in the C2-symmetric closed ring conformation with bound ADP ("closed" conformation, coloured pink and purple) and without bound ADP ("open" conformation, coloured light and dark blue). **f** Superposition of the large and small AAA+ domains within one ClpX subunit containing bound ADP ("closed" conformation, coloured pink) or without bound ADP ("open" conformation, coloured light blue). Movement of the small AAA+ domains of RavA and ClpX upon nucleotide binding is shown using black arrows.

(4L:2U) in solution. The presence of 4L:2U and 5L:1U RavA states, and the similarity between the 4L:2U conformations of ClpX and RavA, seems to suggest that RavA would function via a similar ATP hydrolysis cycle.

**LdcI binding increases RavA ATPase activity at low pH.** To investigate whether or not the restricted orientations of RavA in the LdcI–RavA complex would result in RavA being incapable of hydrolysing ATP, we performed ATPase activity measurements of RavA in the absence or presence of a three-fold molar excess of LdcI at pH values between 5.0 and 9.0. While the RavA ATPase activity increases with pH in the explored interval, LdcI increases it further at pH below 7.0, and decreases it slightly at pH 7.5 or higher (Fig. 6a, b). At pH 7.0, LdcI exerts no apparent effect on RavA ATPase activity. The observed bimodal effect of LdcI on RavA is unexpected, and contrasts with previous results which showed that LdcI increases the ATPase activity of RavA at pH 7.5<sup>9</sup>.

Given the pH-dependent effect of LdcI on RavA ATPase activity, we investigated the LdcI–RavA interaction at different pH values using BLI. In vitro biotinylated RavA-AviTag was immobilised on streptavidin-coated biosensors, and subsequently exposed to a concentration series of LdcI at pH between 5.0 and 8.0 (see “Methods”). BLI measurements at different pH values did not show a marked difference in the binding affinity of LdcI for RavA, which ranged from  $K_D=20$  nM to  $K_D=40$  nM. However, we observed an increase in the height of the BLI signal (response in nm) with decreasing pH (Fig. 6c). This indicates that the mass of the bound ligand increases with lower pH, which can be explained by the pH-dependent oligomerisation of LdcI. Indeed, while LdcI is predominantly dimeric at pH 8.0, at pH 6.5 and 5.0 LdcI forms decamers and stacks of decamers, respectively (Fig. 6d)<sup>12</sup>. Taken together, our BLI and ATPase activity studies suggest that dimeric LdcI has a moderate inhibitory effect on RavA, while LdcI decamers and stacks increase RavA ATPase activity. Most importantly, when bound to LdcI, RavA can exert its ATPase activity over a broad pH range. Thus, despite the restricted orientations of the RavA spiral seam, RavA is still able to efficiently hydrolyse ATP in the LdcI–RavA cage.

## Discussion

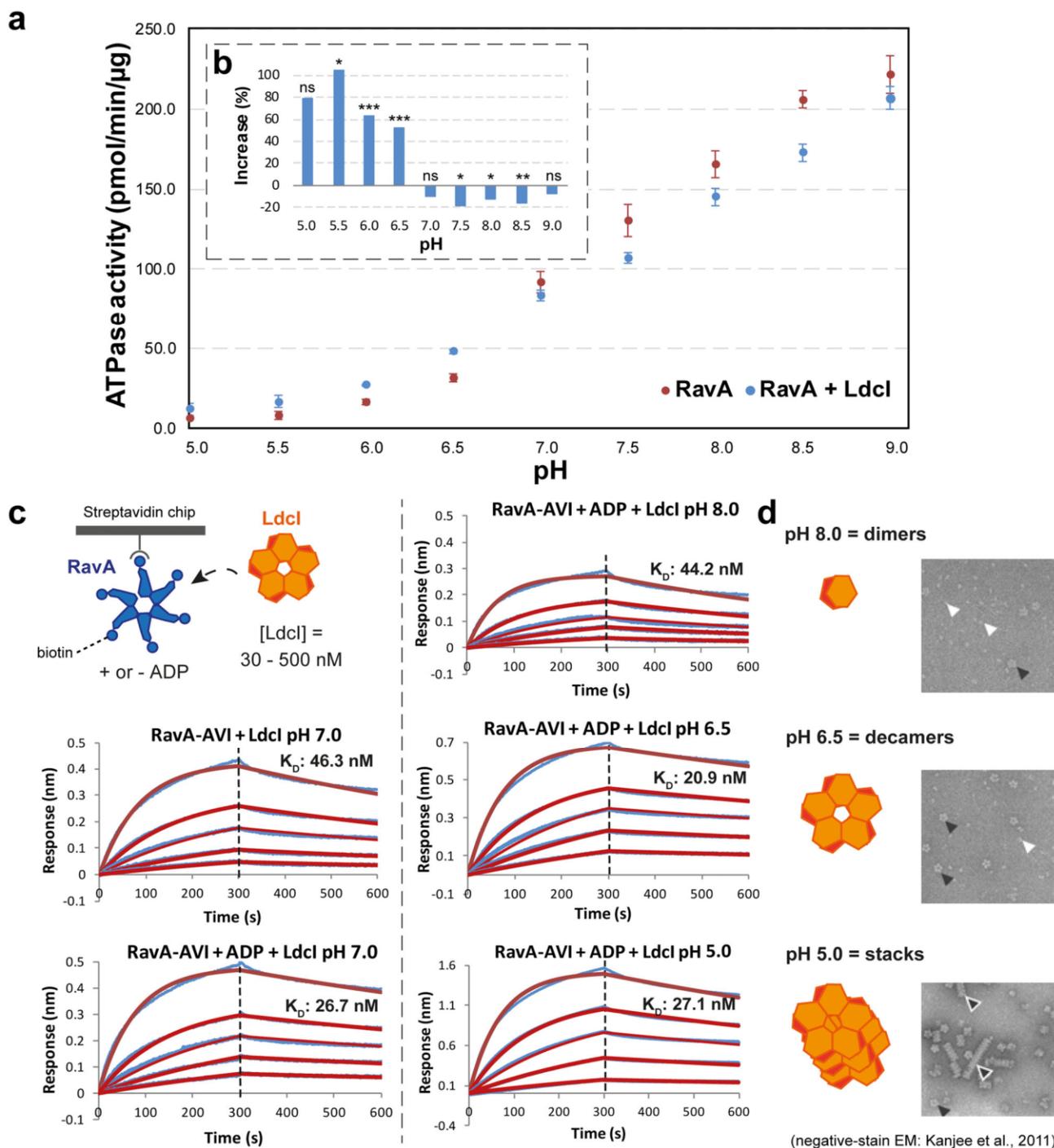
AAA+ ATPases of the MoxR family have been suggested to play a role as chaperones in the assembly of multi-protein complexes, but in general the functions of MoxR family members are not well characterised and their structures are scarce<sup>2</sup>. Functionally, RavA has been implicated in the assembly of *E. coli* respiratory Complex I and modulation of the activity of fumarate reductase<sup>6,7</sup>. The crystal structure of *E. coli* RavA displays monomers that are packed in a left-handed helix, and contains an ATP binding site at the interface between adjacent monomers in the helix<sup>8</sup>. In fact, several other AAA+ ATPases crystallise as apparent monomers in spacegroup P6<sub>5</sub>, thereby forming continuous helices due to crystal packing that resembles hexamers when viewed along the helical screw axis<sup>8,29,30,32,38,41–45</sup>. In some of these cases, the interface between monomers in the helical crystal packing is very similar to the interface elucidated by other structural methods. For instance, the spiral RavA hexamers observed in our cryo-EM reconstructions of the free RavA and the LdcI–RavA complex display an interface which is equivalent to the helical RavA assembly observed in the crystal structure. Likewise, the crystal structure of Spastin forms a helical assembly in which the monomer-monomer interface is compatible with a hexameric model based on docking of monomers in an *ab-initio* small-angle X-ray scattering envelope<sup>42,45</sup>. In contrast, crystal structures of Vps4 show an interface only partially similar to the interface

observed in cryo-EM maps of hexameric Vps4 spirals<sup>30,31,46</sup>. Moreover, the crystal structure of apo-katanin forms a helix with a different handedness than the asymmetric hexamer found in a cryo-EM reconstruction of ATP-bound katanin, and as such does not retain the biologically relevant monomer-monomer interface<sup>29</sup>. In addition, these helices contrast with the planar, symmetric hexamers observed in numerous available AAA+ ATPase crystal structures<sup>5,33–35,47</sup>. Interestingly, a crystal structure of ClpX also shows helically arranged monomers<sup>48</sup>, and the interfaces formed are only slightly shifted compared to the ones observed in hexameric ClpX structures<sup>18</sup>.

Based on negative stain EM data and comparison with related hexameric AAA+ ATPases, RavA was initially modelled as a planar hexameric assembly<sup>2,8</sup>. The cryo-EM analysis of free RavA described here unexpectedly showcases a mixture of two distinct conformational states: a spiral RavA hexamer, also observed in the present cryo-EM structure of the LdcI–RavA cage, and a flat C2-symmetric RavA hexamer characterised by two nucleotide-free seams. Spiral hexameric assemblies with a seam devoid of any bound nucleotide are common among different AAA+ ATPases<sup>14,15,46,49</sup>. For several AAA+ ATPases that form spiral hexamers, a sequential mechanism was proposed whereby ATP hydrolysis causes the seam to move processively around the spiral hexamer via a closed ring intermediate<sup>14,29,49–51</sup>.

Besides a sequential ATP hydrolysis cycle, two other models are put forward to explain how AAA+ ATPases couple ATP hydrolysis to mechanical force to exert their function: the AAA+ lTag is suggested to act via concerted (all-or-none) nucleotide binding and hydrolysis that occurs simultaneously in all subunits<sup>52</sup>, while the AAA+ unfoldases HslU and ClpX are thought to hydrolyse ATP via a probabilistic mechanism where ATP hydrolysis is not strictly sequential around the hexamer<sup>19,21,23,53</sup>. Evidence for probabilistic L to U subunit switching in ClpX stems from assays using individually mutated, disulfide-linked<sup>19</sup> or crosslinked<sup>23</sup> subunits in covalently tethered ClpX pseudohexamers. These studies also demonstrate that ClpX hexamers with one or more L or U locked subunits are able to hydrolyse ATP, but are impaired in substrate binding and degradation. Thus, blocking of L to U switching in a single ClpX subunit uncouples ATP hydrolysis from mechanical work, supporting a probabilistic but coordinated ATP hydrolysis mechanism in which communication between ClpX subunits is obligatory. Similar studies performed on disulfide-crosslinked HslU pseudohexamers<sup>53</sup> show that HslU pseudohexamers with different mixtures of active and inactive subunits can unfold protein substrates and support their degradation by HslV, albeit at a lower rate than wild-type HslUV.

Remarkably, several very recent papers reopen the debate on the exact ATPase mechanism of ClpX by showing high-resolution cryo-EM structures of the ClpXP complex where hexameric ClpX in a spiral 5L:1U state is bound to the tetradecameric ClpP protease<sup>54–56</sup>. Therefore, further studies are required to elucidate whether ClpXP follows a probabilistic<sup>56</sup> or rather processive model<sup>55</sup> of ATPase hydrolysis. Regardless, the existence of both spiral (5L:1U) and a C2-symmetric closed ring (4L:2U) conformations of hexameric RavA, and the resemblance of the latter to 4L:2U ClpX structures, suggests that RavA may follow a similar ATPase mechanism as proposed for ClpX. Consequently, different scenarios for RavA ATPase cycling upon LdcI binding may be envisioned. The current cryo-EM reconstructions of the LdcI–RavA complex in the presence of ADP contain five “seam up” or “seam down” RavA spirals. One possible scenario is that probabilistic L to U subunit switching allows the seam to be transferred to an opposite position in the RavA hexamer. The absence of any alternative observable seam positions other than “seam up” or “seam down” could be the result of geometrical



**Fig. 6** Characterisation of the Ldcl-RavA interaction by ATPase assays and BLI binding studies. **a** Effect of Ldcl on the ATPase activity of RavA. ATPase activity was measured at various pH values (ranging from 5 to 9) as described in “Methods”, either for RavA alone (red) or a mix containing RavA and a three-fold molar excess of Ldcl (blue). Each data point represents the average of three independent measurements. Error bars correspond to the standard deviation. The dashed inset (**b**) shows the percent change of RavA activity when comparing the ATPase activity of RavA alone and RavA plus a three-fold molar excess of Ldcl. The statistical significance is calculated using a 2-sided T-test (2-sample unequal variance).  $P \geq 0.1$ : not significant (ns),  $P \leq 0.1$ : \* $P \leq 0.05$ : \*\* $P \leq 0.001$ : \*\*\*. **c** BLI measurements of the Ldcl-RavA interaction at different pH values ranging from 5 to 8 (blue curves: experimental data, red curves: calculated fit using a 1:1 interaction model). For each experiment, biotinylated RavA-AVITAG was immobilised on a streptavidin-coated BLI biosensor (with or without prior incubation with 1 mM ADP) followed by binding measurements using different concentrations of Ldcl (500 nM, 250 nM, 125 nM, 62.5 nM and 31.25 nM). Average values for the RavA-ADP:Ldcl interaction measured at four different pH values (8, 7, 6.5, 5) are:  $K_D = 29.7 \text{ nM} \pm 10.0$ ,  $k_{on} = 2.68 \times 10^4 \text{ 1/Ms} \pm 2.13 \times 10^3$ , and  $k_{dis} = 8.11 \times 10^{-4} \text{ 1/s} \pm 3.41 \times 10^{-4}$ . The  $K_D$ ,  $k_{on}$  and  $k_{dis}$  values of the individual experiments can be found in Supplementary Table 2. **d** Negative-stain EM micrographs of Ldcl incubated at different pH values (reproduced from Kanjee et al.<sup>12</sup>, with permission of the EMBO Journal)<sup>12</sup>. At lower pH, Ldcl mainly forms decamers (pH 6.5) or stacks of decamers (pH 5), while at higher pH (pH 8) Ldcl is predominantly dimeric.

constraints imposed by lateral contacts between the triple helical domains of neighbouring RavA monomers in the LdcI–RavA cage, thereby restricting the rotations between subunits needed for release of ADP by RavA monomers making these lateral contacts (schematically shown in Supplementary Fig. 9). If RavA indeed follows a similar ATP hydrolysis mechanism to ClpX, this would imply that conformational locking of RavA subunits in the LdcI–RavA cage, as described above, would lead to RavA hexamers that can hydrolyse ATP, but are functionally inactive. However, based on our data we cannot rule out the possibility that other seam orientations of RavA hexamers in the LdcI–RavA cage can occur, but are not observed in our cryo-EM analysis for example because of their transient nature. In addition, alongside the observed alternation of RavA seam states, a comparison of the cryo-EM maps representing the two classes of the LdcI–RavA complex reveals a near 10 Å difference in the distance between opposite LdcI decamers forming the central cavity of the cage (Fig. 1). If, while bound to LdcI, RavA is able to exert its potential protein-remodeling capacity, then the existence of these two classes tends to suggest that ATP hydrolysis by RavA would cause a breathing motion of the LdcI–RavA cage, thereby transferring mechanical force to remodel substrates inside the complex.

Taken together, our synergistic approach, which combines data from complementary structural techniques such as cryo-EM, X-ray crystallography, modelling, as well as biochemical characterisation, has provided insights into nucleotide-dependent conformational changes of RavA during ATP hydrolysis, and the possible ATPase mechanism of RavA in the LdcI–RavA complex. Validation of the proposed interaction partners and characterisation of interaction with substrates are necessary future steps in the elucidation of the structure-function relationships of the LdcI/RavA/ViaA triad, and in uncovering of the mechanism of its action in sensitization of *E. coli* to aminoglycosides<sup>6</sup>. Our work adds to the growing number of AAA+ ATPase structures corresponding to snapshots of the ATP hydrolysis cycle. Clade 7 AAA+ ATPases encompass very divergent families, including MoxR, Chelatase/YifB, the minichromosome maintenance protein MCM built by six different ATPase subunits, and even the eukaryotic Dynein/Midasin where the six AAA+ subunits are all covalently linked. It is tempting to suggest that the observations and hypotheses based on the RavA cryo-EM structures described here may be extended to all clade 7 AAA+ ATPases that share a spatial arrangement of  $\alpha\beta\alpha$  and all- $\alpha$  subdomain resulting in an active site formed between adjacent monomers (Supplementary Fig. 8).

## Methods

**LdcI–RavA complex formation.** RavA and LdcI proteins were expressed and purified as previously described<sup>8,12,13</sup>, with the sole exception that LdcI was expressed in a different ppGpp<sup>-/-</sup> *E. coli* strain (MG1655  $\Delta$ relA  $\Delta$ spoT), generously provided by Dr. Emmanuelle Bouveret. To promote optimal LdcI–RavA cage formation in vitro, purified LdcI and RavA were initially separately diluted to a respective concentration of 0.76 mg ml<sup>-1</sup> and 1.2 mg ml<sup>-1</sup> in a buffer containing 20 mM Tris pH 7.9, 300 mM NaCl, 2 mM ADP, 10 mM MgCl<sub>2</sub>, 0.1 mM PLP and 1 mM DTT. After 10 min at room temperature (RT), equal volumes of both proteins were mixed and incubated 10 min at RT. In the final mix, concentrations of LdcI and RavA were 0.38 mg ml<sup>-1</sup> (4.67  $\mu$ M) and 0.6 mg ml<sup>-1</sup> (10.64  $\mu$ M), respectively, resulting in a RavA monomer:LdcI monomer ratio of 10.64  $\mu$ M:4.67  $\mu$ M = 2.278, or approximately 4.5:2.

**Cryo-electron microscopy on the LdcI–RavA complex.** The quality of complex formation was checked by negative-stain electron microscopy (EM) using 5 times diluted 4.5:2 RavA:LdcI mix (see above). 4  $\mu$ l of sample was applied to the clear side of carbon on a carbon-mica interface and stained with 1% (w/v) uranyl acetate. Images were recorded under low-dose conditions with an FEI T12 microscope operated at 120 kV or FEI F20 microscope operated at 200 kV, at nominal magnifications ranging from 13,000x to 19,000x.

For cryo-EM grid preparation, 4  $\mu$ l of 4.5:2 LdcI–RavA mix was applied onto a glow-discharged quantifoil 400 mesh 1.2/1.3 grid (Quantifoil Micro Tools GmbH,

Germany), the excess solution was blotted for 3 s with a Vitrobot (FEI) using blot force 1, and the grid plunge-frozen in liquid ethane. Data collection was performed on an FEI Polara microscope operated at 300 kV. Movies of 40 frames were collected with a total exposure time of 8 s and a total dose of 40 e<sup>-</sup> Å<sup>-2</sup> on a K2 summit direct electron detector (Gatan) at a magnification of 41,270x, corresponding to 1.21 Å pixel<sup>-1</sup> at the specimen level. Specimen motion during data collection was evaluated and corrected with MotionCor2<sup>57,58</sup>. Frames 3–40 of each movie were dose-weighted, summed and kept for further processing. The contrast transfer function (CTF) of each micrograph was determined with GCTF<sup>59</sup>. 1819 best micrographs were selected based on visual quality control and CTF inspection. As previously noticed<sup>13</sup>, and despite the high affinity of RavA for the LdcI<sup>12</sup>, the LdcI–RavA cage is extremely sensitive to the cryo-EM grid preparation process, which results in a very low amount of intact particles per image. This difficulty in sample preparation limits the number of particles available for further analysis, complicates the particle selection process and hinders obtaining of a high-resolution structure. Eventually, 18,902 particles were manually picked using EMAN2 e2boxer<sup>60</sup> and subjected to two rounds of 2D classification with RELION-2.0<sup>61</sup> to yield a cleaned dataset containing 15,771 particles. For 2D classification and all further steps, CTF-amplitude correction was performed starting from the first peak of the CTF. Visual analysis of 2D class averages immediately revealed considerable heterogeneity in RavA conformations/positions while LdcI appeared more rigid. The initial 3D model based on 2D class averages was calculated with svxviper (SPARX)<sup>62</sup> imposing D5 symmetry and appeared similar to our previously published map<sup>13</sup>. Subsequent 3D classification in RELION-2.0 was performed with C1 symmetry in order to remove the remaining incomplete cages containing 3 or 4 RavA hexamers. This led to a clean dataset containing 11,866 particles. 3D refinement with Relion auto-refine, using the initial svxviper 3D model low-pass filtered to 40 Å as a reference and imposing C5 symmetry, led to a 7.3 Å resolution map.

Albeit already much better than our previous map<sup>13</sup>, the resulting map showed a lower resolution of RavA in comparison to LdcI, which again pointed to structural heterogeneity of RavA particles inside the LdcI–RavA complex. Thus, a soft mask was created from a fit of one hexameric RavA<sup>8</sup> and two LdcI decamers<sup>12</sup> into the map, and the dataset was expanded by replicating each particle from the C5 consensus refinement and adding  $n^{\circ}72^{\circ}$  with  $n = 1, \dots, 5$  to its first Euler angle. A masked 3D classification was then conducted with RELION-2.0 without angular search. This procedure enabled a reliable separation of the dataset into two classes, containing 47 and 53% of the data, respectively. Unexpectedly, both classes showed left-handed RavA spirals, with a clearly defined seam facing either the upper (orientation A) or the lower (orientation B) LdcI decamer and related exactly by a 180° rotation (Supplementary Fig. 1). In order to be able to combine images of both orientations, we applied a  $-\phi, 180^{\circ}$ -theta,  $180^{\circ}$ +psi transformation to the Euler angles of orientation B to bring it into orientation A. Images corresponding to orientation A and rotated orientation B were then subjected to a masked local 3D auto-refinement (RELION-2.0), with the mask that again included two LdcI decamers and one RavA hexamer. This masked reconstruction had a global resolution of 7.3 Å, with local resolution ranging from 5 to 14 Å. The B-factor sharpening, using a B-factor value of  $-270 \text{ \AA}^2$ , was performed in RELION-2.0 as described<sup>63</sup>.

To further address eventual conformational variability of the RavA spiral inside the LdcI–RavA complex, we undertook a final 3D classification using the same mask as before, including two LdcI decamers and one RavA hexamer. This classification allowed separation of two defined states (Class 1 and 2) containing 32 and 28% of particles, respectively, and a third state corresponding to more mobile/flexible RavA containing 40% of particles (Class 3). Refinement of Class 1 and Class 2 gave respective resolutions of 7.6 and 7.8 Å, with local resolution ranging from 5 to 14 Å for both classes. In Class 1, the densities corresponding to all RavA domains contacting LdcI (i.e. LARA domains) are well resolved. In contrast, in Class 2 as well as in the preceding 7.3 Å map from the masked refinement containing all particles, the density of one LARA domain is missing.

**Post-processing of LdcI–RavA Cryo-EM maps and fitting of structures.** Local resolution estimation and subsequent filtering of maps were performed in RELION-3.0, using B-factors of  $-200, -250$  and  $-300 \text{ \AA}^2$  for the masked 3D refinement containing all particles, or particles from Class 1 and 2, respectively. For fitting of atomic models in the resulting filtered maps, we used the previously-determined crystal structures of LdcI (PDB ID: 3N75)<sup>12</sup> and RavA (PDB ID: 3NBX)<sup>8</sup>. Careful analysis of the RavA crystal packing revealed that RavA was crystallised as a continuous helix. In each map, two decameric LdcI molecules extracted from PDB 3N75 and one spiral RavA hexamer extracted from a continuous RavA helix generated from PDB 3NBX were first manually placed using Chimera<sup>64</sup>, and then fitted separately using iMODFIT<sup>27</sup>, followed by a single round of B-factor (ADP) refinement in Phenix<sup>65,66</sup>.

**Cryo-electron microscopy on free RavA.** Purified RavA (see above) was diluted to a final concentration of 0.1 mg ml<sup>-1</sup> in the presence of 1 mM ADP and incubated at room temperature for 10 min. 3  $\mu$ l of RavA:ADP was applied to glow-discharged (20 mA, 45 s) R2/1 400 mesh holey carbon copper grids (Quantifoil Micro Tools GmbH). Grids were plunge-frozen in liquid ethane with a Vitrobot

Mark IV (FEI) operated at 100% humidity using blot force 1 and a blot time of 2 s. Data collection was performed on an FEI Polara microscope operated at 300 kV.

A total of 2944 movies comprising 40 frames were recorded at a tilt angle of 0° on a K2 summit direct electron detector (Gatan Inc) operated in counting mode. Movies were collected with a total exposure time of 6 s and a total dose of  $40 \text{ e}^- \text{ \AA}^{-2}$ . Preliminary processing suggested that RavA adopted a strongly preferred orientation on the grid. To overcome preferred orientations of RavA, a further 1083 micrographs were recorded with a 30° tilt, with a total exposure time of 6 s and a total dose of  $44 \text{ e}^- \text{ \AA}^{-2}$ . All movies were recorded at a magnification of 41,270x, corresponding to a pixel size of  $1.21 \text{ \AA pixel}^{-1}$  at the specimen level, with a target defocus range of 1.8–3.8  $\mu\text{m}$ .

RavA:ATP $\gamma$ S grids were prepared as for RavA:ADP, except with a 10 min incubation with ATP $\gamma$ S instead of ADP. Data collection was performed on a Glacios microscope (Thermo Scientific) operated at 200 kV. A total of 2809 movies (1224 of which were tilted to 30°) comprising 29 frames were recorded on a Falcon II direct electron detector (Thermo Scientific) at a magnification of 116,086x, corresponding to a pixel size of  $1.206 \text{ \AA pixel}^{-1}$  at the specimen level. Movies were collected with a total exposure time of 6 s and a total dose of  $41 \text{ e}^- \text{ \AA}^{-2}$ , with a target defocus range of 1.5–3.5  $\mu\text{m}$ .

**Image processing and 3D reconstruction of free RavA.** Motion correction on both RavA:ADP datasets was carried out using MotionCor2<sup>57</sup>. After discarding the first two frames, the remaining frames were aligned, dose-weighted and summed. CTF parameters were determined on aligned dose-weighted sums using CTFIND4<sup>67</sup>, and micrographs with an estimated resolution by CTFIND4 of better than 8  $\text{\AA}$  were kept for further processing. Because of the observation that RavA was present as a spiral in the LdcI–RavA cage, particles were picked from all micrographs using the particle-picking software FPM<sup>68</sup> using a spiral hexamer extracted from the RavA crystal structure (PDB ID: 3NBX)<sup>8</sup> filtered to a resolution of 20  $\text{\AA}$  as a reference. Per-particle CTF estimation was then carried out on selected particles using GCTF<sup>59</sup> to account for variations in defocus across the tilted micrographs. A total of 924,000 particles were picked, and particles were extracted with a box size of  $256 \times 256$  pixels. Particles from untitled and tilted micrographs were separately subjected to several rounds of 2D classification in RELION-2.0, then combined prior to 3D classification resulting in a cleaned dataset of 562,000 particles. 3D classification without imposed symmetry was subsequently carried out with four classes, using an asymmetric initial model generated in RELION-2.1 filtered to 40  $\text{\AA}$  as a reference. This resulted in three classes displaying asymmetric spirals (comprising 416,000 particles) and unexpectedly, one class showing a 2-fold symmetric closed ring (corresponding to 146,000 particles).

Particles from the three classes displaying asymmetric spirals were grouped together and subjected to a second round of 3D classification into two classes, resulting in one junk class and one good class (comprising 216,000 particles). However, density for the sixth monomer in the spiral was weak, most likely due to partial occupancy or flexibility. To resolve this monomer, a final round of 3D classification was carried out into 5 classes. Particles from the best class were then subjected to 3D refinement, resulting in a map with a final resolution of 6.94  $\text{\AA}$  after post-processing and sharpening with a B-factor of  $-400 \text{ \AA}^2$ .

Due to the fact that the micrographs were originally picked using a spiral hexamer as a reference, it was possible that side views corresponding to the 2-fold symmetric closed ring were missed during the picking process. To overcome this, micrographs were re-picked using the closed-ring map filtered to a resolution of 15  $\text{\AA}$ . 1,072,943 picked particles were subjected to per-particle CTF correction followed by several rounds of 2D classification. The resulting 721,000 particles were imported into CryoSPARC<sup>69</sup>, and divided randomly into four subsets, each containing  $\sim 180,000$  particles. For each subset, particles were subject to ab-initio 3D classification (using the Ab-initio Reconstruction algorithm) into five classes with no imposed symmetry. 257,000 particles which classified into closed-ring classes were combined, and subject to a further round of asymmetric ab-initio 3D classification into two classes, resulting in one volume with visibly less stretching in the z-direction, corresponding to 72,000 particles. Particles from this class underwent a homogeneous refinement against the resulting volume, resulting in a map with a resolution of 5.96  $\text{\AA}$  after post-processing in RELION- 2.1, which was sharpened with a B-factor of  $-350 \text{ \AA}^2$ .

For the RavA:ATP $\gamma$ S dataset, micrographs were motion corrected using MotionCor2, after discarding the first two frames. CTF estimation was carried out using GCTF. Particles from the best 1044 micrographs after manual screening were picked using Gautomatch (<http://www.mrc-lmb.cam.ac.uk/kzhang/>), using a.mrcs stack containing projections of both the C2-symmetric and spiral RavA hexamers as a reference. The resulting  $\sim 477,000$  particles were subject to per-particle CTF estimation using GCTF. Particles were imported into CryoSPARC and 2D classification was then carried out. Classes showed significant heterogeneity and a strongly preferred orientation, even more so than for the RavA:ADP dataset (see Supplementary Fig. 6).

**Fitting of structures and refinement.** Local resolutions of 3D reconstructions were calculated in RELION-3.0<sup>70</sup>. All resolution estimates are calculated using the 0.143 gold-standard Fourier shell correlation (FSC) criterion<sup>63</sup>. For fitting of atomic models, a RavA hexamer (for the spiral RavA conformation) and two RavA trimers (for the C2-symmetric RavA conformation) were extracted from a

continuous RavA helix generated from the crystal structure (PDB ID: 3NBX)<sup>8</sup>, and were fitted into the corresponding maps using iMODFIT<sup>27</sup>. The two resulting models were then subjected to a single round of ADP refinement in Phenix, followed by geometry minimisation<sup>65,66</sup>. Considering the medium resolution of the cryo-EM maps for both LdcI–RavA and RavA alone, we took particular caution not to interpret the models at atomic level. Rather, we focus on large-scale conformational changes such as the orientation of the RavA seam, the distance between two LdcI decamers forming the LdcI–RavA complex, the movement of the LARA domains, the spiral and the C2-symmetric conformation of RavA and the presence or absence of the nucleotide in the RavA intersubunit interface.

**BLI binding studies.** For BLI binding studies, a C-terminal AviTag was added to RavA cloned in the p11 vector (N-terminal cleavable HIS-tag). The AviTag-containing RavA was expressed and purified using the same protocol as described for RavA<sup>8</sup> with the exception that 100  $\mu\text{M}$  of D-biotin was added to the LB medium during expression in *E. coli* BL-21 DE3 cells (overnight expression, 20 °C). Biotinylated RavA-AviTag was purified to homogeneity, concentrated to 9 mg ml<sup>-1</sup>, aliquoted and flash-frozen for later use. BLI experiments were performed in either 1x TBS pH 8 (25 mM Tris, 300 mM NaCl, 10 mM MgCl<sub>2</sub>, 10% glycerol), 1x HBS pH 7 (25 mM HEPES, 300 mM NaCl, 10 mM MgCl<sub>2</sub>, 10% glycerol), 1x MES pH 6.5 (25 mM MES, 300 mM NaCl, 10 mM MgCl<sub>2</sub>, 10% glycerol) or 1x MES pH 5 supplemented with 1x kinetics buffer (0.1% w/v BSA, 0.02% v/v Tween-20), 1 mM ADP, 1 mM DTT and 0.1 mM PLP.

Experiments were performed using an Octet RED96 instrument (FortéBio), operated at 293 K. Before the start of each BLI experiment, RavA-AviTag was incubated with 1 mM ADP for 10 min. Streptavidin-coated Octet biosensors (FortéBio) were functionalised with biotinylated RavA-AviTag, quenched with 10  $\mu\text{g ml}^{-1}$  biocytin, and dipped in wells containing 500, 250, 125, 62.5, 31.25 or 0 nM LdcI. To check for nonspecific binding during the experiments, non-functionalised biosensors were used to measure the signal from the highest ligand concentration as well as running buffer. All data were fitted with the FortéBio Data Analysis 9.0 software using a 1:1 interaction model. Average values and standard deviations for the RavA-ADP:LdcI interaction measured at four different pH values (8, 7, 6.5, 5) are:  $K_D = 29.7 \pm 10.0 \text{ nM}$ ,  $k_{on} = 2.68 \times 10^4 \pm 2.13 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$ , and  $k_{dis} = 8.11 \times 10^{-4} \pm 3.41 \times 10^{-4} \text{ s}^{-1}$  (Supplementary Table 2).

**RavA ATPase activity measurements.** RavA ATPase activity was measured at 25 °C on a Infinite<sup>®</sup> 200 microplate reader (TECAN). The reaction was performed in 100  $\mu\text{l}$  of 50 mM MES/Tris pH 5–9, 10 mM MgCl<sub>2</sub>, 2 mM ATP, 0.02% (v/v) Triton X-100, 1 mM DTT, 17  $\mu\text{g ml}^{-1}$  RavA. Experiments were initiated by the addition of RavA with or without a three-fold molar excess of LdcI, and stopped after 45 min by addition of 50  $\mu\text{l}$  of 3% (w/v) lithium dodecyl sulfate. The amount of inorganic phosphate produced by ATP hydrolysis was assessed using the colorimetric Fiske and Subbarow method<sup>71</sup>. Inorganic phosphate standards were used at each of the experimental conditions to calculate the specific activity of RavA. All measurements were performed in triplicate (technical replicates).

**Statistics and reproducibility.** For BLI binding studies, the reported average value and standard deviation for the RavA:LdcI interaction were calculated using four measurements at pH 8, 7, 6.5 and 5.

For RavA ATPase activity measurements, the reported average values and standard deviations were calculated using three technical replicates per measurement.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Cryo-EM maps of the two classes of the LdcI–RavA complex in ADP-bound state, and the spiral and C2-symmetric closed ring conformations of free RavA in the presence of ADP, together with their corresponding fitted atomic structures have been submitted to the EMDB and PDB with accession codes EMD-4469 and PDB-6Q7L for LdcI–RavA Class 1, EMD-4470 and PDB-6Q7M for LdcI–RavA Class 2, EMD-10351 and PDB-6SZA for the C2-symmetric closed ring RavA conformation, and EMD-10352 and PDB-6SZB for the spiral RavA conformation. Source Data for Fig. 6a–c can be found in Supplementary Data 1.

Received: 22 November 2019; Accepted: 11 January 2020;

Published online: 28 January 2020

## References

1. Snider, J. & Houry, W. A. MoxR AAA+ ATPases: a novel family of molecular chaperones? *J. Struct. Biol.* **156**, 200–209 (2006).

2. Wong, K. S. & Houry, W. A. Novel structural and functional insights into the MoxR family of AAA+ ATPases. *J. Struct. Biol.* **179**, 211–221 (2012).
3. Kahle, M., ter Beek, J., Hosler, J. P. & Ädelroth, P. The insertion of the non-heme FeBcofactor into nitric oxide reductase from *P. denitrificans* depends on NorQ and NorD accessory proteins. *Biochim. Biophys. Acta—Bioenerg.* **1859**, 1051–1058 (2018).
4. Tsai, Y. C. C., Lapina, M. C., Bhushan, S. & Mueller-Cajar, O. Identification and characterization of multiple rubisco activases in chemoautotrophic bacteria. *Nat. Commun.* **6**, 8883 (2015).
5. Sutter, M. et al. Structural characterization of a newly identified component of  $\alpha$ -carboxysomes: the AAA+ domain protein CsoCbbQ. *Sci. Rep.* **5**, 16243 (2015).
6. Wong, K. S. et al. The MoxR ATPase RavA and its cofactor ViaA interact with the NADH: ubiquinone oxidoreductase I in *Escherichia coli*. *PLoS ONE* **9**, e85529 (2014).
7. Wong, K. S., Bhandari, V., Janga, S. C. & Houry, W. A. The RavA-ViaA chaperone-like system interacts with and modulates the activity of the fumarate reductase respiratory complex. *J. Mol. Biol.* **429**, 324–344 (2017).
8. El Bakkouri, M. et al. Structure of RavA MoxR AAA+ protein reveals the design principles of a molecular cage modulating the inducible lysine decarboxylase activity. *Proc. Natl Acad. Sci.* **107**, 22499–22504 (2010).
9. Snider, J. et al. Formation of a distinctive complex between the inducible bacterial lysine decarboxylase and a novel AAA+ ATPase. *J. Biol. Chem.* **281**, 1532–1546 (2006).
10. Lin, J. et al. Mechanisms of acid resistance in enterohemorrhagic *Escherichia coli*. *Appl. Environ. Microbiol.* **62**, 3094–3100 (1996).
11. Sabo, D. L., Boeker, E. A., Byers, B., Waron, H. & Fischer, E. H. Purification and physical properties of inducible *Escherichia coli* lysine decarboxylase. *Biochemistry* **13**, 662–670 (1974).
12. Kanjee, U. et al. Linkage between the bacterial acid stress and stringent responses: the structure of the inducible lysine decarboxylase. *EMBO J.* **30**, 931–944 (2011).
13. Malet, H. et al. Assembly principles of a unique cage formed by hexameric and dimeric *E. coli* proteins. *Elife* **3**, e03653 (2014).
14. Gates, S. N. et al. Ratchet-like polypeptide translocation mechanism of the AAA+ disaggregase Hsp104. *Science* **357**, 273–279 (2017).
15. Yokom, A. L. et al. Spiral architecture of the Hsp104 disaggregase reveals the basis for polypeptide translocation. *Nat. Struct. Mol. Biol.* **23**, 830–837 (2016).
16. Deville, C., Franke, K., Mogk, A., Bukau, B. & Saibil, H. R. Two-step activation mechanism of the ClpB disaggregase for sequential substrate threading by the main ATPase motor. *Cell Rep.* **27**, 3433–3446.e4 (2019).
17. Shin, M. et al. Distinct structural features of the lon protease drive conserved hand-over-hand substrate translocation. *bioRxiv* <https://doi.org/10.1101/617159> (2019).
18. Glynn, S. E., Martin, A., Nager, A. R., Baker, T. A. & Sauer, R. T. Structures of asymmetric ClpX hexamers reveal nucleotide-dependent motions in a AAA+ protein-unfolding machine. *Cell* **139**, 744–756 (2009).
19. Stinson, B. M. et al. Nucleotide binding and conformational switching in the hexameric ring of a AAA+ machine. *Cell* **153**, 628–639 (2013).
20. Ye, Q. et al. TRIP13 is a protein-remodeling AAA+ ATPase that catalyzes MAD2 conformation switching. *Elife* **2015**, 1–44 (2015).
21. Cordova, J. C. et al. Stochastic but highly coordinated protein unfolding and translocation by the ClpXP proteolytic machine. *Cell* **158**, 647–658 (2014).
22. Martin, A., Baker, T. A. & Sauer, R. T. Rebuilt AAA+ motors reveal operating principles for ATP-fuelled machines. *Nature* **437**, 1115–1120 (2005).
23. Stinson, B. M., Baytshtok, V., Schmitz, K. R., Baker, T. A. & Sauer, R. T. Subunit asymmetry and roles of conformational switching in the hexameric AAA+ ring of ClpX. *Nat. Struct. Mol. Biol.* **22**, 411–416 (2015).
24. Iyer, L. M., Leipe, D. D., Koonin, E. V. & Aravind, L. Evolutionary history and higher order classification of AAA+ ATPases. in *J. Struct. Biol.* **146**, 11–31 (2004).
25. Erzberger, J. P. & Berger, J. M. Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 93–114 (2006).
26. Miller, J. M. & Enemark, E. J. Fundamental characteristics of AAA+ protein family structure and function. *Archaea* **2016**, 9294307 (2016).
27. Lopéz-Blanco, J. R. & Chacón, P. IMODFIT: efficient and robust flexible fitting based on vibrational analysis in internal coordinates. *J. Struct. Biol.* **184**, 261–270 (2013).
28. Kandiah, E. et al. Structural insights into the *Escherichia coli* lysine decarboxylases and molecular determinants of interaction with the AAA+ ATPase RavA. *Sci. Rep.* **6**, 24601 (2016).
29. Zehr, E. et al. Katanin spiral and ring structures shed light on power stroke for microtubule severing. *Nat. Struct. Mol. Biol.* **24**, 717–725 (2017).
30. Monroe, N. et al. The oligomeric state of the active Vps4 AAA ATPase. *J. Mol. Biol.* **426**, 510–525 (2014).
31. Sun, S. et al. Cryo-EM structures of the ATP-bound Vps4E233Q hexamer and its complex with Vta1 at near-atomic resolution. *Nat. Commun.* **8**, 16064 (2017).
32. Caillat, C. et al. Asymmetric ring structure of Vps4 required for ESCRT-III disassembly. *Nat. Commun.* **6**, 8781 (2015).
33. Bochtler, M. et al. The structures of HslU and the ATP-dependent protease HslU-HslV. *Nature* **403**, 800–805 (2000).
34. Suno, R. et al. Structure of the whole cytosolic region of ATP-dependent protease FtsH. *Mol. Cell* **22**, 575–585 (2006).
35. Zhang, X. et al. Structure of the AAA ATPase p97. *Mol. Cell* **6**, 1473–1484 (2000).
36. Sysoeva, T. A. Assessing heterogeneity in oligomeric AAA+ machines. *Cell. Mol. Life Sci.* **74**, 1001–1018 (2017).
37. Roessle, M. W. et al. Upgrade of the small-angle X-ray scattering beamline X33 at the European Molecular Biology Laboratory, Hamburg. *J. Appl. Crystallogr.* **40**, s190–s194 (2007).
38. Fodje, M. N. et al. Interplay between an AAA module and an integrin I domain may regulate the function of magnesium chelatase. *J. Mol. Biol.* **311**, 111–122 (2001).
39. Bieniossek, C. et al. The molecular architecture of the metalloprotease FtsH. *Proc. Natl Acad. Sci. USA* **103**, 3066–3071 (2006).
40. Alfieri, C., Chang, L. & Barford, D. Mechanism for remodelling of the cell cycle checkpoint protein MAD2 by the ATPase TRIP13. *Nature* **559**, 274–278 (2018).
41. Putnam, C. D. et al. Structure and mechanism of the RuvB holliday junction branch migration motor. *J. Mol. Biol.* **311**, 297–310 (2001).
42. Roll-Mecak, A. & Vale, R. D. Structural basis of microtubule severing by the hereditary spastic paraplegia protein spastin. *Nature* **451**, 363–367 (2008).
43. Scott, A. et al. Structural and mechanistic studies of VPS4 proteins. *EMBO J.* **24**, 3658–3669 (2005).
44. Stotz, M. et al. Structure of green-type Rubisco activase from tobacco. *Nat. Struct. Mol. Biol.* **18**, 1366–1370 (2011).
45. Taylor, J. L., White, S. R., Lauring, B. & Kull, F. J. Crystal structure of the human spastin AAA domain. *J. Struct. Biol.* **179**, 133–137 (2012).
46. Monroe, N., Han, H., Shen, P. S., Sundquist, W. I. & Hill, C. P. Structural basis of protein translocation by the Vps4-Vta1 AAA ATPase. *Elife* **6**, e24487 (2017).
47. Matsuyama, B. Y. et al. Mechanistic insights into c-di-GMP-dependent control of the biofilm regulator FleQ from *Pseudomonas aeruginosa*. *Proc. Natl Acad. Sci.* **113**, E209–E218 (2016).
48. Kim, D. Y. & Kim, K. K. Crystal structure of ClpX molecular chaperone from helicobacter pylori. *J. Biol. Chem.* **278**, 50664–50670 (2003).
49. Su, M. et al. Mechanism of Vps4 hexamer function revealed by cryo-EM. *Sci. Adv.* **3**, e170032 (2017).
50. Puchades, C. et al. Structure of the mitochondrial inner membrane AAA+ protease YME1 gives insight into substrate processing. *Science* **358** (2017).
51. Ripstein, Z. A., Huang, R., Augustyniak, R., Kay, L. E. & Rubinstein, J. L. Structure of a AAA+ unfoldase in the process of unfolding substrate. *Elife* **6**, e25754 (2017).
52. Gai, D., Zhao, R., Li, D., Finkielstein, C. V. & Chen, X. S. Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen. *Cell* **119**, 47–60 (2004).
53. Baytshtok, V. et al. Covalently linked HslU hexamers support a probabilistic mechanism that links ATP hydrolysis to protein unfolding and translocation. *J. Biol. Chem.* **292**, 5695–5704 (2017).
54. Gatsogiannis, C., Balogh, D., Merino, F., Sieber, S. & Raunser, S. Cryo-EM structure of the ClpXP protein degradation machinery. *bioRxiv* 638692, <https://doi.org/10.1101/638692> (2019).
55. Ripstein, Z. A., Vahidi, S., Houry, W. A., Rubinstein, J. L. & Kay, L. E. A processive rotary mechanism couples substrate unfolding and proteolysis in the ClpXP degradation machinery. *bioRxiv* 780494, <https://doi.org/10.1101/780494> (2019).
56. Fei, X. et al. Structures of the ATP-fueled ClpXP proteolytic machine bound to protein substrate. *bioRxiv* 704999, <https://doi.org/10.1101/704999> (2019).
57. Zheng, S. Q. et al. MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
58. Li, X. et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590 (2013).
59. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
60. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
61. Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
62. Hohn, M. et al. SPARX, a new environment for Cryo-EM image processing. *J. Struct. Biol.* **157**, 47–55 (2007).

63. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
64. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
65. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D: Biol. Crystallogr.* **66**, (213–221 (2010).
66. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D: Biol. Crystallogr.* **68**, 352–367 (2012).
67. Rohou, A. & Grigorieff, N. CTFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
68. Estrozi, L. F. & Navaza, J. Fast projection matching for cryo-electron microscopy image reconstruction. *J. Struct. Biol.* **162**, 324–334 (2008).
69. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
70. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, 1–22 (2018).
71. Fiske, C. & Subbarow, Y. The colorimetric determination of phosphorous. *J. Biol. Chem.* **66**, 375–400 (1925).

## Acknowledgements

We thank Guy Schoehn for establishing and managing the cryo-electron microscopy platform and for providing training and support. We are grateful to Aymeric Peuch for help with the usage of the EM computing cluster and to Ambroise Desfosses and Leandro Estrozi for discussions. This work was funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 647784 to I.G. For electron microscopy, this work used the platforms of the Grenoble Instruct-ERIC Center (ISBG: UMS 3518 CNRS-CEA-UGA-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL, a project of the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003). IBS acknowledges integration into the Interdisciplinary Research Institute of Grenoble (IRIG, CEA). The IBS electron microscope facility is supported by the Rhône-Alpes Region, the Fondation pour la Recherche Médicale (FRM), the fonds FEDER, the Centre National de la Recherche Scientifique (CNRS), the Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), the University of Grenoble Alpes, EMBL, and the GIS-Infrastructures en Biologie Santé et Agronomie (IBISA). MJ was funded by a CEA IRTELIS PhD fellowship, JF was supported by a long-term EMBO fellowship (ALTF441-2017) and a Marie Skłodowska-Curie actions Individual Fellowship (789385, RespVIRALI).

## Author contributions

M.J., B.A., A.F., K.H., J.F. and H.M. purified proteins. M.J., B.A., M.B.V., J.F., H.M. and I.G. made cryo-EM grids. M.J., B.A., M.B.V., H.M. and I.G. collected E.M. data. M.J., B.A., J.F., H.M. and I.G. analysed E.M. data. R.M. and P.C. performed ATPase activity measurements. M.J. and J.F. performed B.L.I. measurements. M.J., J.F. and I.G. analysed structures and mechanisms. H.M. supervised B.A. and contributed to the initial design of the project together with I.G. I.G. designed, directed and funded the overall study. J.F. and I.G. wrote the paper with significant input of M.J. and H.M.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s42003-020-0772-0>.

**Correspondence** and requests for materials should be addressed to J.F., H.M. or I.G.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

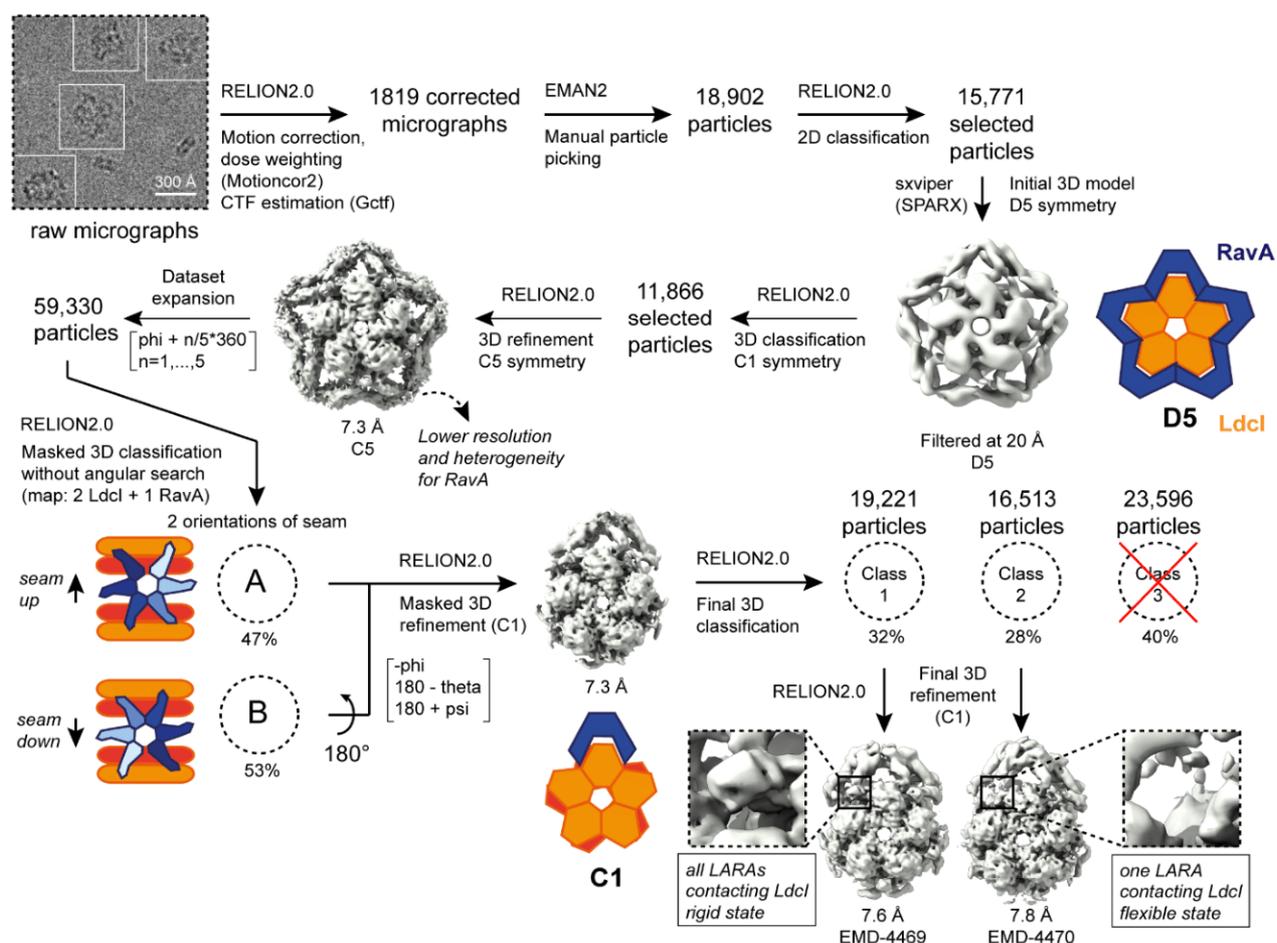
**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

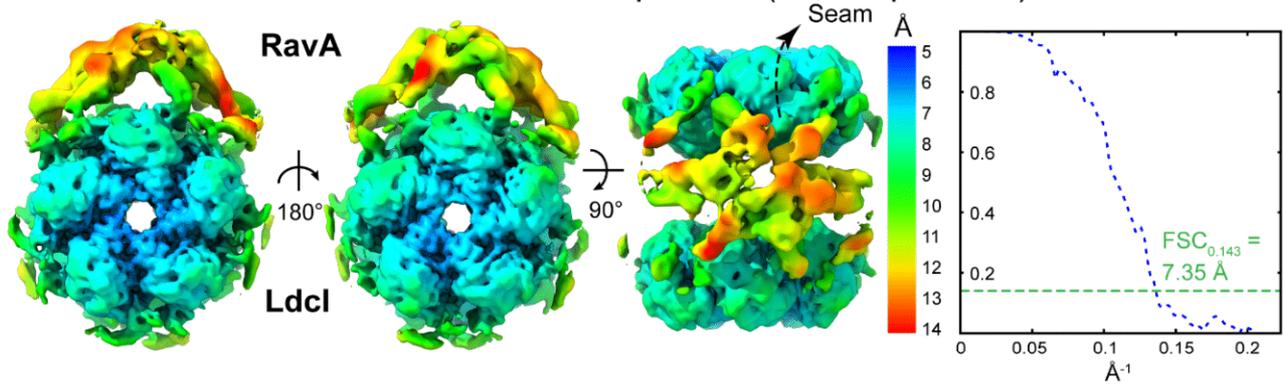
© The Author(s) 2020

## Supplementary information

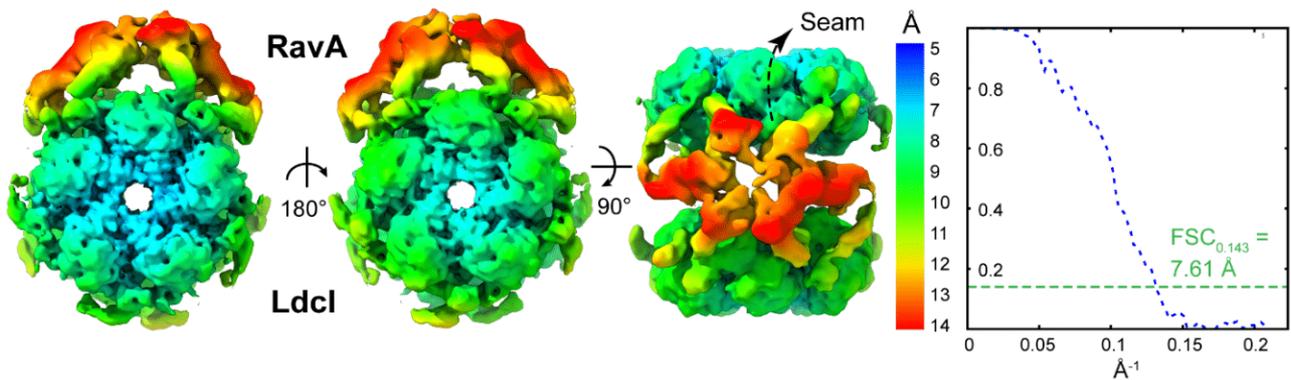


**Supplementary Figure 1.** Image processing pipeline of cryo-EM data of the LdcI-RavA complex supplemented with ADP. A representative micrograph is shown with picked LdcI-RavA particles, followed by subsequent classification and refinement procedures performed in this study. Software packages used at each step are indicated as well as details regarding the procedure. Schematic representations either show a top view of the LdcI-RavA cage centered on one orange-coloured LdcI decamer and five surrounding dark-blue coloured RavA hexamers, or a side view centered on one light to dark blue-coloured RavA hexamer, and two LdcI decamers with the two rings per decamer coloured light and dark orange. A summary of cryo-EM data collection, refinement and validation statistics can be found in Supplementary Table 1.

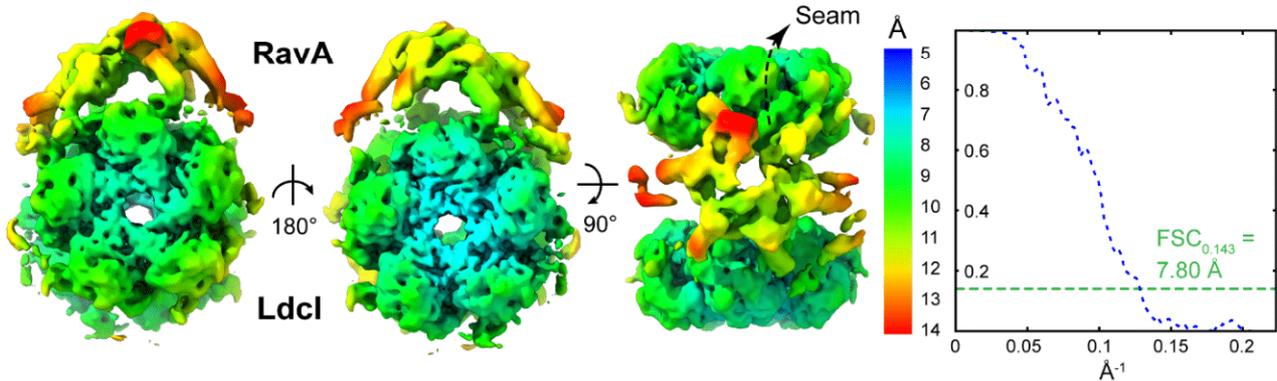
Masked 3D refinement after dataset expansion (59,330 particles)



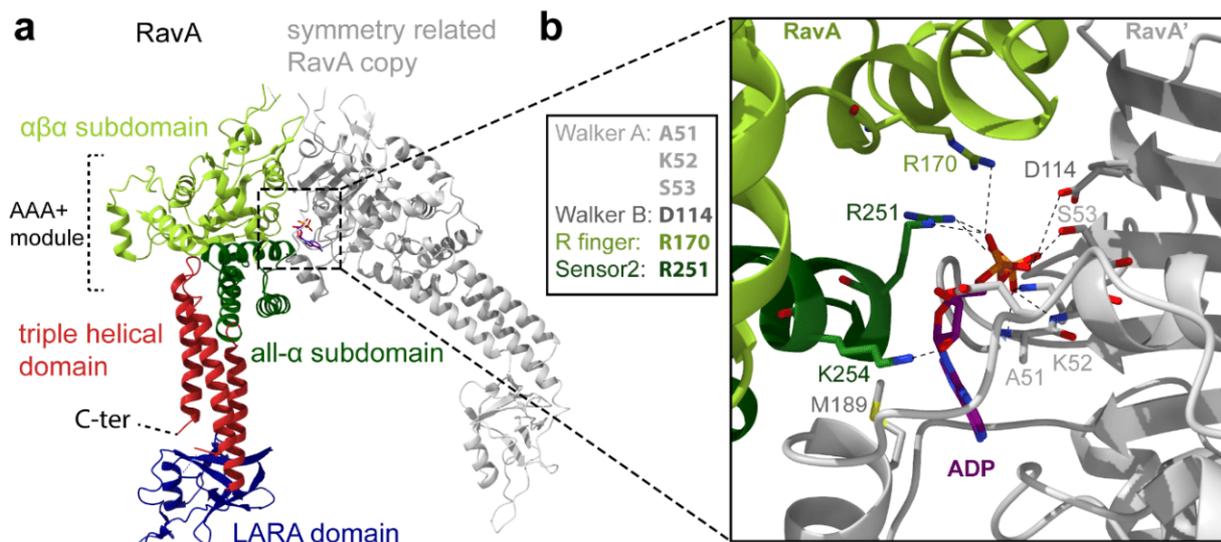
3D classification and refinement, Class 1 (19,221 particles)



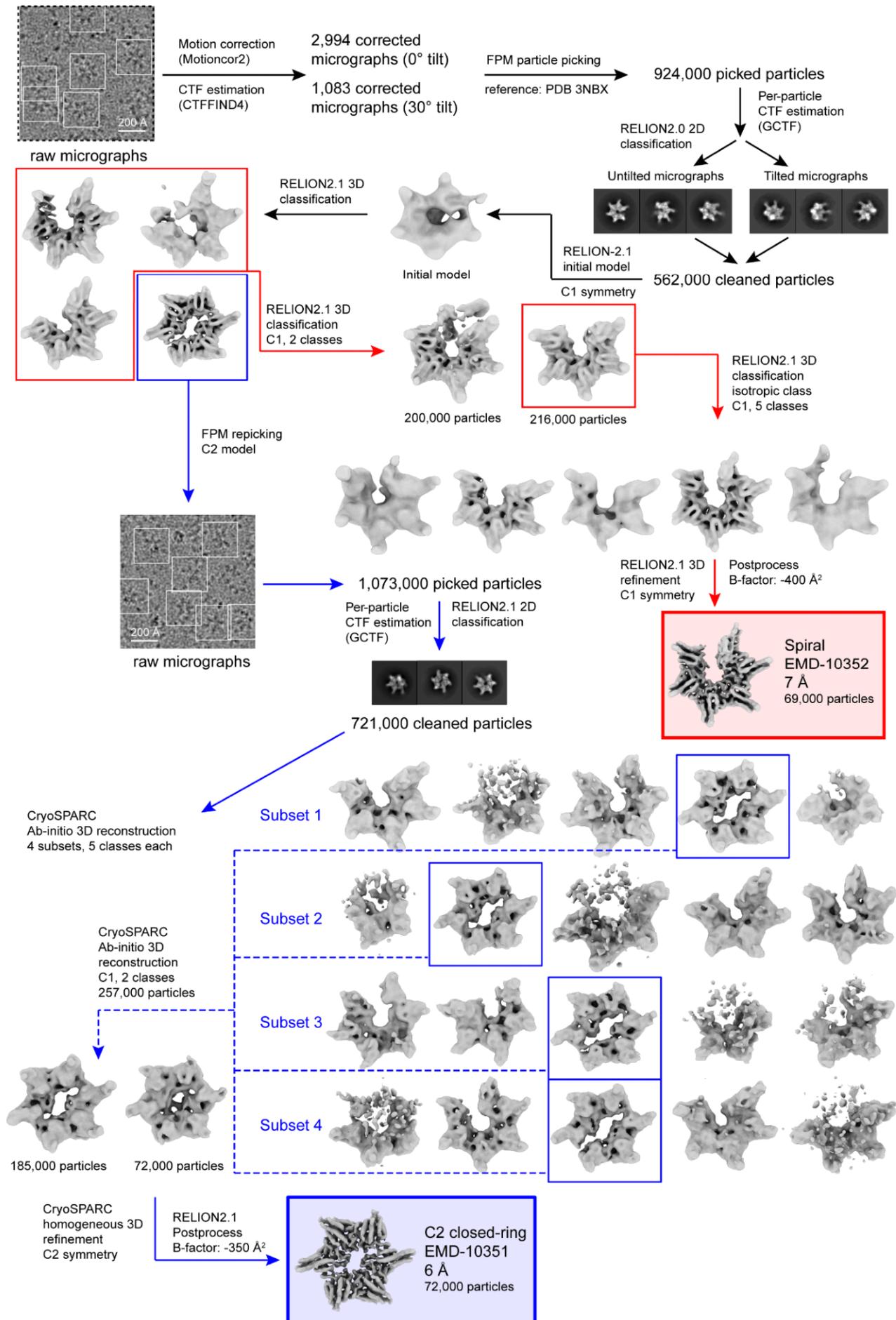
3D classification and refinement, Class 2 (16,513 particles)



**Supplementary Figure 2.** Local resolution estimation in RELION2.0 for a masked 3D refinement containing all LdcI-RavA particles (upper panels), particles from LdcI-RavA Class 1 (middle panels) or particles from LdcI-RavA Class 2 (lower panels). Maps are coloured according to the local resolution. Panels on the right display gold-standard FSC curves with estimated resolution at  $FSC = 0.143$ .

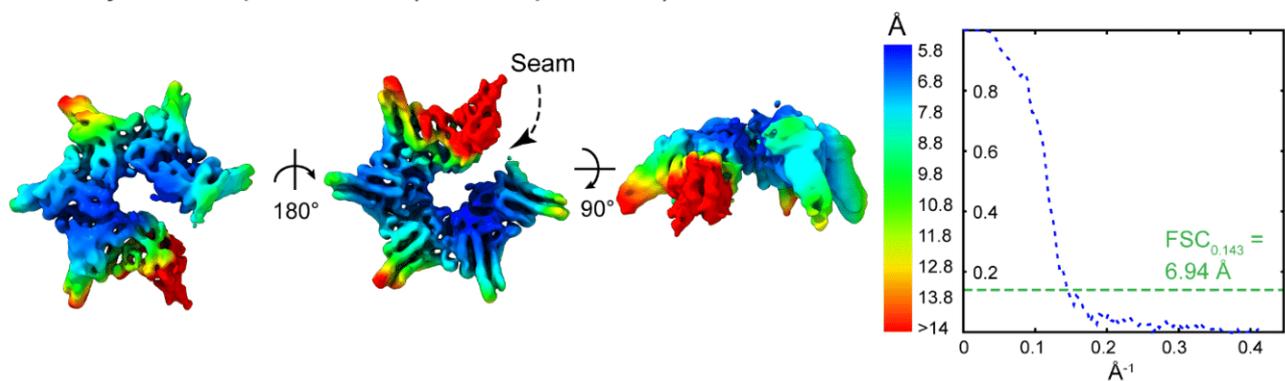


**Supplementary Figure 3.** (a): The crystal structure of *E. coli* RavA (PDB ID: 3NBX) (El Bakkouri *et al.*, 2010) is shown as a cartoon. Different subdomains in the RavA monomer are annotated and coloured accordingly: AAA+ module  $\alpha\beta\alpha$  subdomain in light green, AAA+ module all- $\alpha$  subdomain in dark green, triple helical domain in red and LARA domain in dark blue. One symmetry related RavA copy is shown in light grey. The interface formed between the two monomers containing bound ADP is marked with a dashed square. (b): Zoom displaying bound ADP at the active site formed between two RavA monomers. Interacting residues are annotated and shown as sticks, polar interactions are shown as dashed lines.

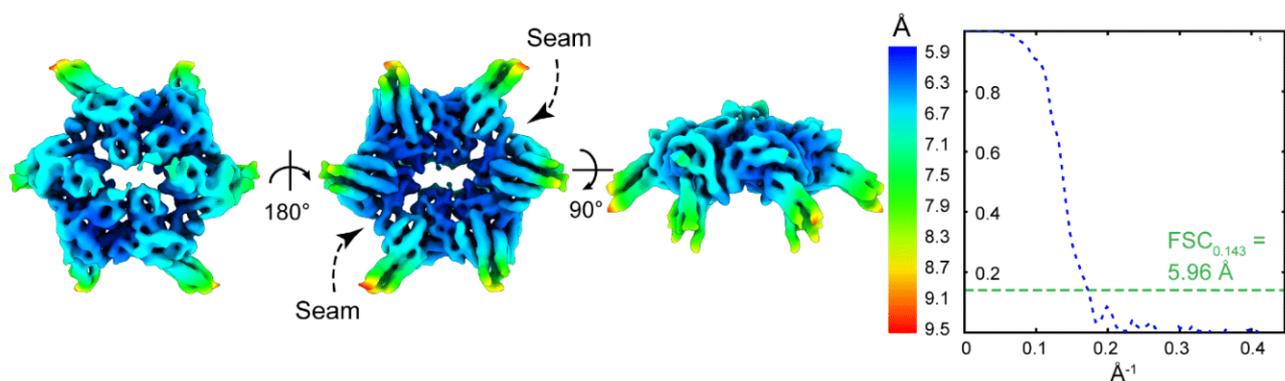


**Supplementary Figure 4.** Image processing pipeline of cryo-EM data of unbound RavA supplemented with ADP. A representative micrograph is shown with picked RavA particles, followed by subsequent classification and refinement procedures performed in this study. Software packages used at each step are indicated as well as details regarding the procedure. A summary of cryo-EM data collection, refinement and validation statistics can be found in Supplementary Table 1.

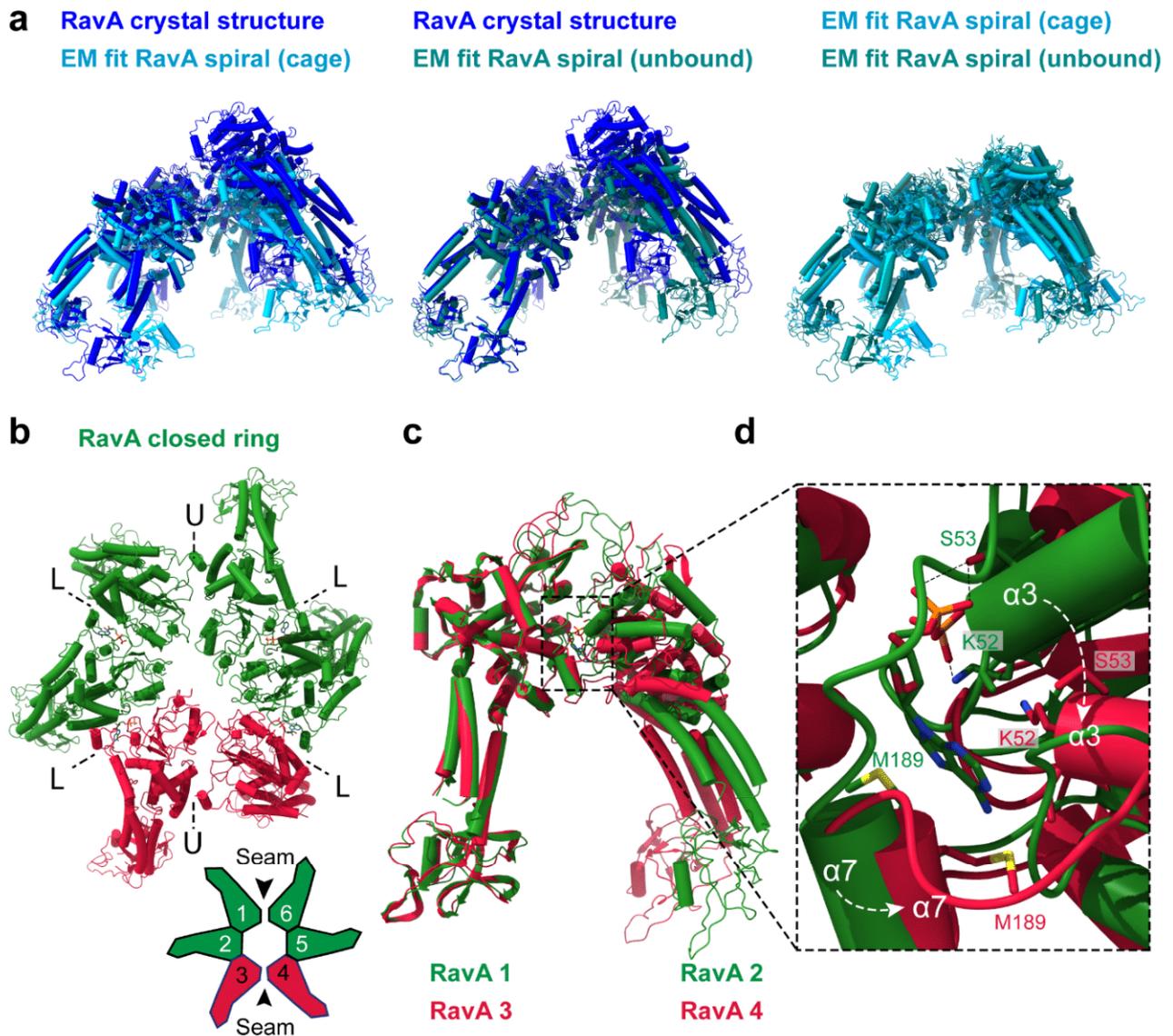
Assymmetric spiral RavA (69,237 particles)



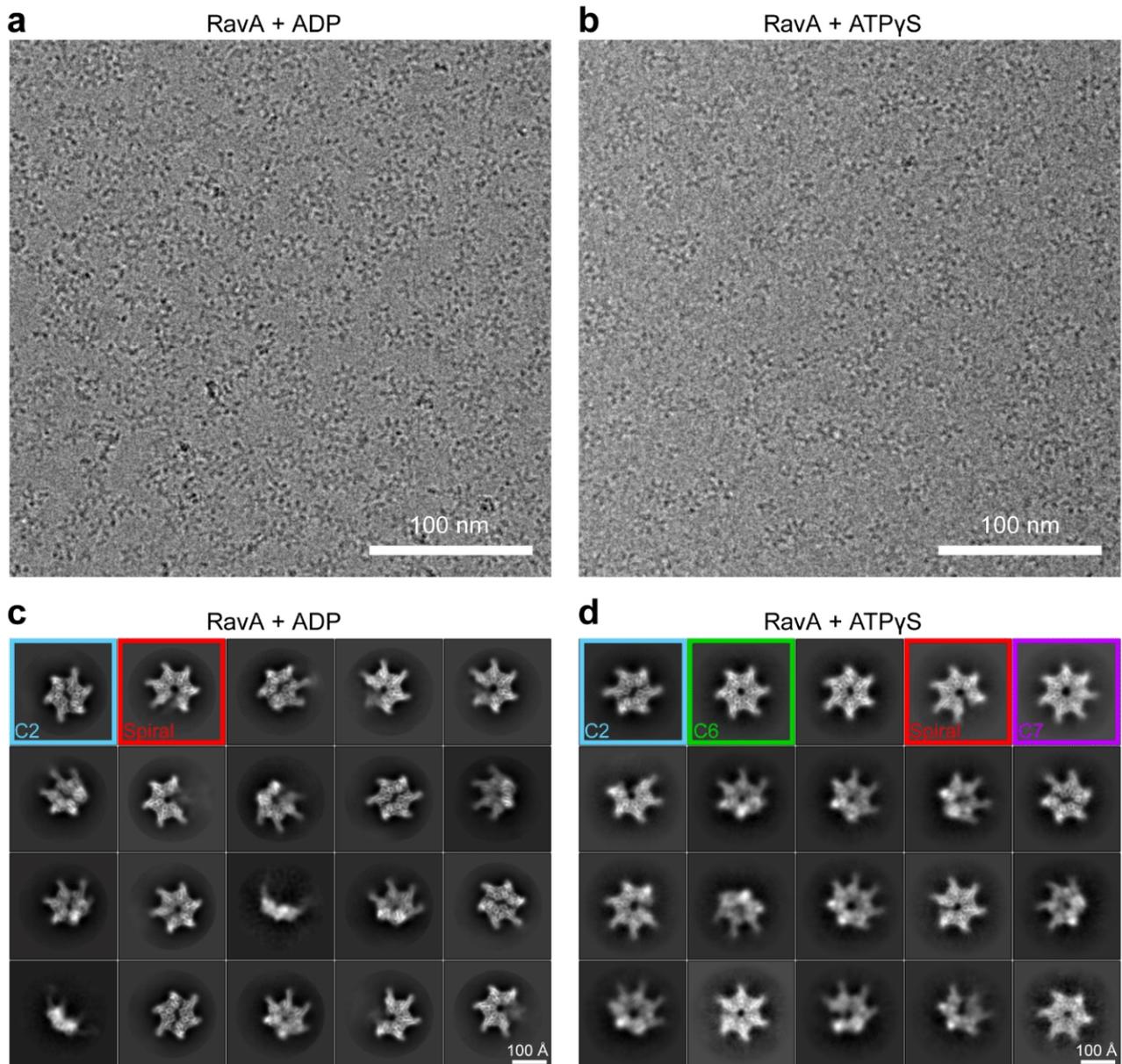
C2-symmetric closed ring RavA (72,175 particles)



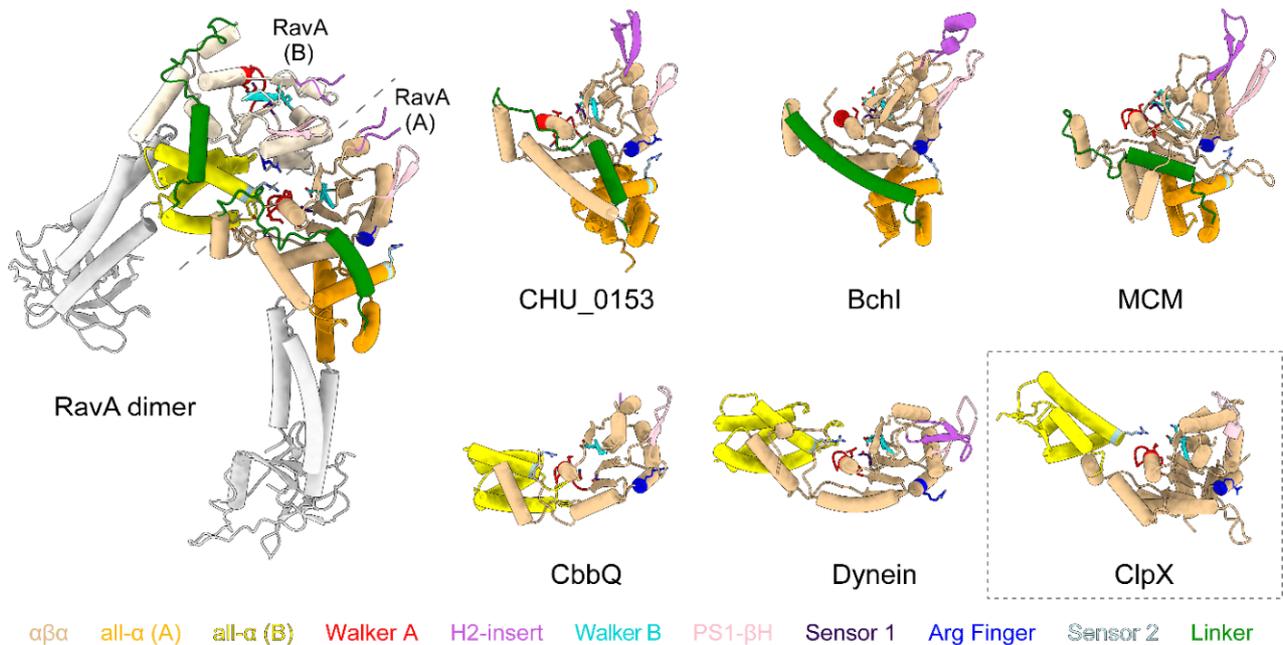
**Supplementary Figure 5.** Local resolution estimation in RELION3.0 for 3D refinements of spiral and C2-symmetric closed ring conformations of RavA. Maps are coloured according to the local resolution. Panels on the right display gold-standard FSC curves with estimated resolution at FSC = 0.143.



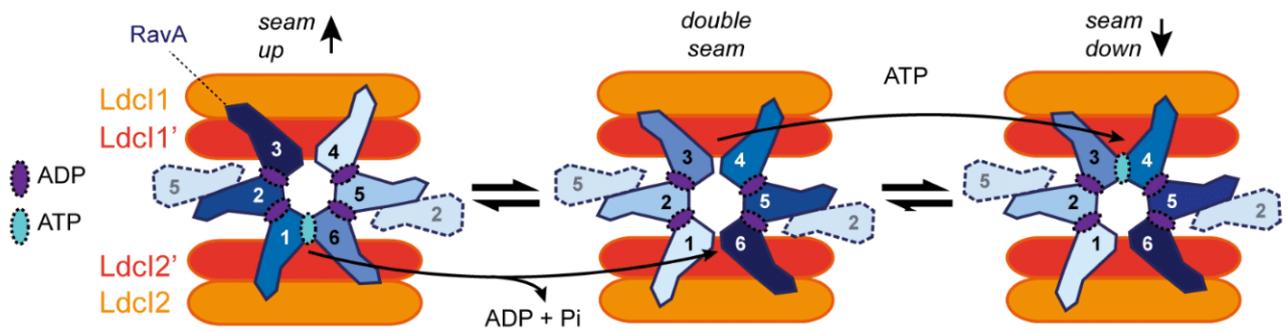
**Supplementary Figure 6.** (A) Side by side comparisons of spiral RavA hexamers extracted from the crystal structure (dark blue) and fits of RavA in the cryo-EM maps of the LdcI-RavA cage (light blue) or the spiral conformation of unbound RavA (dark cyan). Structures are displayed as cartoons. (B) Cartoon representation of the C2-symmetric closed ring conformation of unbound RavA. Loadable and unloadable ATP binding sites are annotated with L and U respectively. An accompanying schematic representation is shown with numbered RavA monomers. The seam positions between monomers 1-6 and 3-4 are annotated using black arrows. (C) Alignment of neighbouring RavA monomers 1 and 2 (green), and 3 and 4 (crimson red) extracted from the C2-symmetric closed ring conformation of unbound RavA. (D) Zoom of the active-site interface formed between adjacent RavA monomers 1 and 2 (green, with bound ADP) and monomers 3 and 4 (crimson red, without present nucleotide). Shifts in helices  $\alpha 3$  and  $\alpha 7$  are accentuated using dotted white arrows. Relevant residues are labelled and shown as sticks.



**Supplementary Figure 7.** (a) & (b): Micrographs of RavA supplemented with ADP (a) or ATP $\gamma$ S (b). (c) & (d): Representative 2D class averages displaying spiral (red squares) and C2 symmetric closed-ring (blue squares) conformations found in both the ADP (c) and ATP $\gamma$ S (d) datasets, while additional C6 symmetric (green square) and C7 symmetric (purple square) 2D class averages are solely found in the ATP $\gamma$ S dataset.



**Supplementary Figure 8.** Cartoon representation of ATPase domains of the Clade 7 AAA+ ATPases CHU\_0153 (PDB ID 2R44), BchI (PDB ID 1G8P), MCM (PDB ID 4R7Y), CbbQ (PDB ID 5C3C), Dynein (PDB ID 4AKG, AAA3 domain) aligned to a RavA monomer (A) (PDB ID 6SZA), with ClpX (PDB ID 3HWS) shown for reference. Structural features of models are coloured as follows:  $\alpha\beta\alpha$  = light orange, all- $\alpha$  = orange/yellow, Walker A = red, H2-insert = purple, Walker B = cyan, PS1- $\beta$ H = pink, Sensor 1 = indigo, Arginine Finger = blue, Linker = green, Sensor 2 = light blue. Where the  $\alpha\beta\alpha$  and all- $\alpha$  subdomains superimpose with the corresponding domains in a single RavA monomer (A), the all- $\alpha$  subdomain is coloured orange. Where the all- $\alpha$  subdomain superimposes with the all- $\alpha$  subdomain of the adjacent RavA monomer (B) in the RavA dimer, it is coloured yellow. RavA, CHU\_0153, BchI and MCM share the reorientation of the all- $\alpha$  subdomain following a linker which is distinctive to clade 7 AAA+ ATPases, whereas CbbQ and Dynein align more closely to the Clade 5 ClpX, and lack the linker region. Relevant residues from the different structural features are shown as sticks.



**Supplementary Figure 9.** Schematic representation of a possible scenario for the ATPase mechanism of LdcI-bound RavA. The two pentameric rings per LdcI decamer are coloured light and dark orange respectively, while RavA hexamers are coloured light to dark blue. ATP hydrolysis only occurs at the active sites formed between monomers 1 and 6 or 3 and 4 in the spiral RavA hexamer. RavA legs from adjacent hexamers (dashed lines) constrain the movement of RavA monomers 2 and 5, resulting in inactive ATPase sites between monomers 1-2, 2-3, 4-5 and 5-6. As a result, the spiral RavA hexamer cycles between two equivalent states with the seam oriented towards either the upper ('seam up') or lower ('seam down') LdcI decamer during ATP hydrolysis via an intermediate 'double seam' state.

<b>Data collection and processing</b>	LdcI-RavA complex + ADP		RavA + ADP	
	Class 1	Class 2	C2-symmetric closed ring	Spiral open ring
	EMD-4469 PDB ID 6Q7L	EMD-4470 PDB ID 6Q7M	EMD-10351 PDB ID 6SZA	EMD-10352 PDB ID 6SZB
Magnification	41,270	41,270	41,270	41,270
Voltage (kV)	300	300	300	300
Electron exposure (e <sup>-</sup> /Å <sup>2</sup> )	40	40	40	40
Defocus range (μm)	1.0 - 3.5 μm	1.0 - 3.5 μm	1.8 - 3.8 μm	1.8 - 3.8 μm
Pixel size (Å)	1.21	1.21	1.21	1.21
Symmetry imposed	C1	C1	C2	C1
Initial particle images (no.)	59,330	59,330	~1,072,000	~924,000
Final particle images (no.)	19,221	16,513	72,175	69,237
Map resolution (Å)	7.61	7.80	5.96	6.94
FSC threshold	0.143	0.143	0.143	0.143
Map resolution range (Å)	6.5 - 14.9	6.9 - 14.8	5.9-10.0	5.8 - 23.0
<b>Fitting/Refinement</b>				
Used Software	iMODFIT, Phenix	iMODFIT, Phenix	iMODFIT, Phenix	iMODFIT, Phenix
Initial model used (PDB code)	3NBX 3N75	3NBX 3N75	3NBX	3NBX
Model resolution (Å)	8.8	9.7	7.1	8.0
FSC threshold	0.5	0.5	0.5	0.5
Model resolution range (Å)	∞ - 8.8	∞ - 9.7	∞ - 7.1	∞ - 8.0
Map sharpening B factor (Å <sup>2</sup> )	-250	-300	-350	-400
Model composition				
Non-hydrogen atoms	137,337	137,351	23,190	23,217
Protein residues	17,102	17,104	2886	2886
Ligands (ADP)	5	5	4	5
B factors (Å <sup>2</sup> )				
Protein	371.76	288.39	425.58	551.42
Ligand	740.63	401.93	224.85	344.86
R.m.s. deviations				
Bond lengths (Å)	0.003	0.003	0.018	0.003
Bond angles (°)	0.462	0.466	0.91	0.870
Validation				
MolProbity score	1.90	1.91	1.76	1.66
Clashscore	6.17	6.13	5.86	4.42
Poor rotamers (%)	0.17	0.16	0.92	0.88
Ramachandran plot				
Favored (%)	89.41	89.27	93.19	93.19
Allowed (%)	9.92	9.99	6.14	6.14
Disallowed (%)	0.67	0.74	0.67	0.67

**Supplementary Table 1.** Cryo-EM data collection, refinement and validation statistics

	<b>K<sub>D</sub> (M)</b>	<b>k<sub>on</sub> (1/Ms)</b>	<b>k<sub>dis</sub> (1/s)</b>
<b>RavA-AVI + LdcI, pH 7.0</b>	4.63E-08 ± 4.28E-10	2.13E+04 ± 1.27E+02	9.87E-04 ± 6.98E-06
<b>RavA-AVI + ADP + LdcI, pH 7.0</b>	2.67E-08 ± 2.97E-10	2.47E+04 ± 1.33E+02	6.59E-04 ± 6.40E-06
<b>RavA-AVI + ADP + LdcI, pH 8.0</b>	4.42E-08 ± 4.72E-10	2.96E+04 ± 2.24E+02	1.31E-03 ± 9.86E-06
<b>RavA-AVI + ADP + LdcI, pH 6.5</b>	2.09E-08 ± 2.73E-10	2.57E+04 ± 1.42E+02	5.38E-04 ± 6.38E-06
<b>RavA-AVI + ADP + LdcI, pH 5.0</b>	2.71E-08 ± 2.42E-10	2.73E+04 ± 1.27E+02	7.40E-04 ± 5.63E-06

**Supplementary Table 2.** Summary of BLI measurements of the LdcI-RavA interaction at different pH values.

# Chapter 5. Towards the cryo-EM characterisation of RavA

In the pursuit of a cryo-EM reconstruction of RavA, work was carried out that is not included in the final manuscript presented in the previous chapter. Chapter 5 will focus on additional steps carried out in the pursuit of characterising RavA by cryo-EM. Other RavA constructs were tried in an attempt to improve particle homogeneity and oligomeric stability, including BC2-tagged RavA and two ATPase-deficient mutants. I will also present a more detailed summary of the image processing steps that were taken for both the RavA + ADP and RavA + ATP $\gamma$ S datasets presented in Chapter 4, in order to resolve the two-fold symmetric closed ring conformation (hereafter referred to as the C2 conformation) and the spiral open ring conformation (hereafter referred to as the spiral conformation). Finally, I will present some initial work carried out to characterise RavA bound to the model substrate casein, in an attempt to characterise the RavA-substrate interaction and the active substrate-bound state of RavA.

## 5.1. BC2-tagged RavA

While pursuing the goal of localising the LdcI-RavA cage in the cell by fluorescence microscopy (see Chapters 6 and 7 for further details), C-terminally BC2-tagged RavA and LdcI constructs were developed by Clarissa Liesche in the team. Very briefly, the BC2 tag is a short peptide of 12 amino acids that is recognised by the BC2 nanobody – labelling the BC2 nanobody with a fluorescent dye allows the visualisation of BC2-tagged constructs by immunofluorescence (Braun et al., 2016). In order to characterise the effect of the BC2 tag on the structure of the LdcI and RavA oligomers, and to assess whether the BC2-tag interfered with the binding of RavA to LdcI, both LdcI-BC2 and RavA-BC2 were expressed and purified.

For the RavA-BC2 purification (performed by Karine Huard), cells were lysed by sonication and the lysate was subsequently centrifuged at 20,000 rpm for 45 minutes. The supernatant was filtered with a 0.20  $\mu\text{m}$  pore filter and loaded onto an IMAC (immobilised metal affinity chromatography) column. After washing the column with a buffer containing 20 mM imidazole, protein was eluted using a linear gradient from 20 mM to 250 mM imidazole. Fractions containing RavA-BC2 were dialysed overnight to remove imidazole and concentrated to  $\sim 500 \mu\text{L}$ , before being loaded onto a Superose 6 gel filtration column. After gel filtration, samples from selected fractions were analysed by SDS-PAGE. The gel filtration profile was identical to that of wild-type RavA and analysis by SDS-PAGE confirmed that the protein was pure.

While the BC2-tagged constructs were primarily designed for eventual fluorescence microscopy applications, negative stain EM analysis of RavA-BC2 (after incubation with 1 mM ADP) gave the impression that hexamers were more homogeneous and of a higher quality than wild-type RavA + ADP on the grid (Figure 5.1). I collected a small negative stain EM dataset of RavA-BC2 for some preliminary analysis and to decide whether the sample would be suitable for cryo-EM. Grids were stained using uranyl acetate, and 25 micrographs were collected on an F20 microscope equipped with a Ceta camera and operated at 200 kV. Images were collected at a nominal magnification of 40,000 x, corresponding to a pixel size 2.73  $\text{\AA}$ /pixel.

For image processing, CTF estimation was first carried out on micrographs using CTFFIND4 (Rohou and Grigorieff, 2015). Semi-automatic particle picking was carried out using the Boxer program in EMAN (Ludtke et al., 1999), resulting in 16,934 picked particles. Further steps were carried out in RELION-2.1 (Kimanius et al., 2016). Particles were extracted with a box size of 112 x 112 pixels and subjected to several rounds of 2D classification. 2D classes showed that the particles were relatively homogeneous and

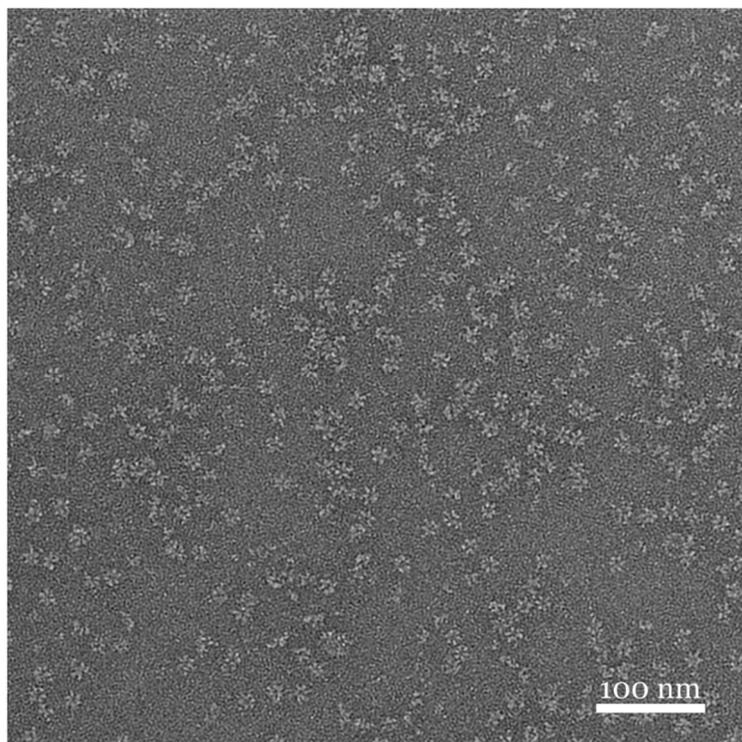


Figure 5.1. Crop of a negative stain EM micrograph of RavA-BC2. Scale bar = 100 nm.

mostly hexameric, but with some classes showing broken particles (Figure 5.2). Several classes corresponding to tilted views or side views were seen as well. With hindsight, 2D classes that correspond to the two-fold symmetric state seen in cryo-EM (presented in Chapter 4) can even be seen by negative stain EM.

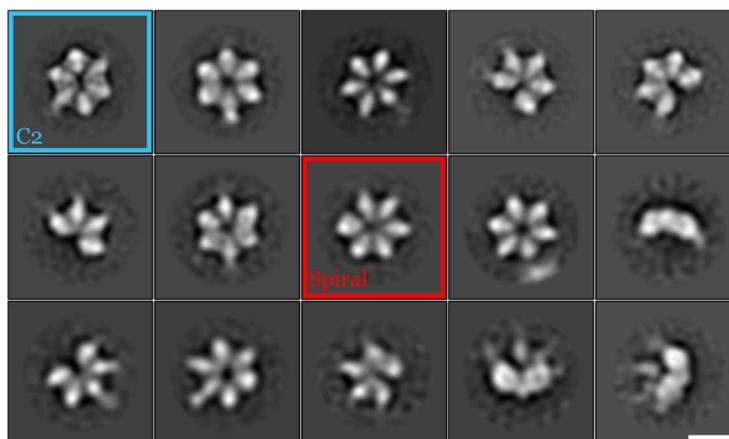


Figure 5.2. Gallery of 2D class averages. The two conformations seen by cryo-EM presented in Figure 3 of Chapter 4 can also be seen, as indicated by blue (C2) and red (spiral) squares. Scale bar = 100 Å.

Based on the quality of the micrographs and the distribution of top and side views seen by 2D classification, we decided to progress to cryo-EM on the RavA-BC2 sample. RavA-BC2 was first incubated with 1 mM ADP for 10 minutes at room temperature. Sample was then applied to cryo-EM grids which were then frozen with a number of freezing conditions using a Vitrobot Mark IV. Two concentrations of protein were frozen on

Quantifoil R 1.2/1.3 Cu grids, both with and without a 2 nm continuous carbon film, after grids were glow discharged for 45 s with a current of 20 mA. Grids were then screened on a Polara microscope operated at 300 kV. Hexamers of RavA were seen but appeared to be heterogeneous and partially aggregated, and the ice quality was often either too thick or too thin. Grids were subsequently frozen and screened three more times with different freezing conditions, including the pumping of grids under vacuum first, and with different protein concentrations. Grids were also frozen with the addition of 0.25% glutaraldehyde to attempt to stabilise the RavA hexamer. Disappointingly, none of the grids showed promising images even after extensive screening. Therefore, we decided to instead try to stabilise the RavA hexamer by introducing mutations that abolished the ATPase activity to stabilise the hexamer, as discussed in the following section.

## 5.2. Stabilising the RavA hexamer – ATPase mutants

RavA often appears to be very heterogeneous in EM images, with mixtures of oligomeric states. Frequently, RavA appears to form ‘broken’ oligomers with only five or even four monomers present, even in the presence of nucleotide, as shown above in Figure 5.2. Therefore, we attempted to stabilise the RavA hexamer by cloning and purifying mutants of RavA that were unable to hydrolyse ATP, which is commonly done for other AAA+ ATPases (Sandate et al., 2019). The two mutants K52Q and E115Q, which affect key residues in the Walker A and Walker B motifs respectively, were cloned into the p11 plasmid (with help from Jan Felix and Angélique Fraudeau). Both mutants were expressed overnight at 37°C and then purified based on the purification protocol described in the previous section, to see how the mutations affected the oligomeric stability.

RavA-K52Q was highly expressed, and although there was some precipitation during overnight dialysis the gel filtration profile looked similar to the wild-type RavA gel filtration profile (Figure 5.3A). There was also an additional peak at ~12 mL that contained no protein (as shown by SDS-PAGE in Figure 5.3B) which was likely due to DNA contamination based on the high  $A_{255}$  absorption.

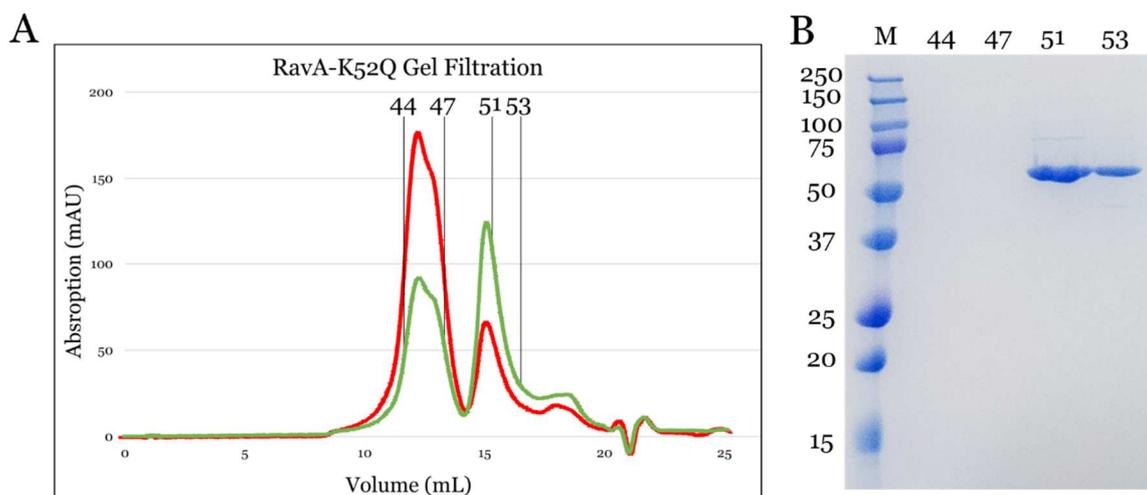


Figure 5.3. **A**) Gel filtration profile of RavA-K52Q showing absorption at  $A_{280}$  (green) and  $A_{255}$  (red). **B**) SDS-PAGE gel of fractions after gel filtration as indicated in the profile in A). The two fractions taken from the first peak do not contain protein and likely correspond to DNA contamination.

RavA-E115Q (purified by Karine Huard) was also highly expressed but significantly less soluble than wild-type RavA and RavA-K52Q. The gel filtration peak appeared at approximately the same elution volume as wild-type RavA and RavA-K52Q, but was approximately 30 times smaller.

Samples from the top of the RavA gel filtration peak were taken for both mutants, and after incubation with 1 mM ADP for 10 minutes at room temperature, samples were analysed by negative stain EM at a concentration of  $\sim 0.05$  mg/mL. RavA-K52Q was imaged on an F20 microscope operated at 200 kV and equipped with a Ceta camera, whereas RavA-E115Q was analysed on a T12 microscope equipped with an Orius 1000 camera and operated at 120 kV.

Hexamers of both mutants were seen in micrographs, as shown in Figure 5.4A. However, these did not appear to be any more homogeneous than wild-type RavA. The K52Q mutation appeared to give more consistent hexamers than the E115Q mutation and had a higher yield, but still were not completely homogeneous. Compared to the RavA-BC2 micrographs (shown in Figure 5.1), the images of the Walker A and Walker B mutants were of lower quality.

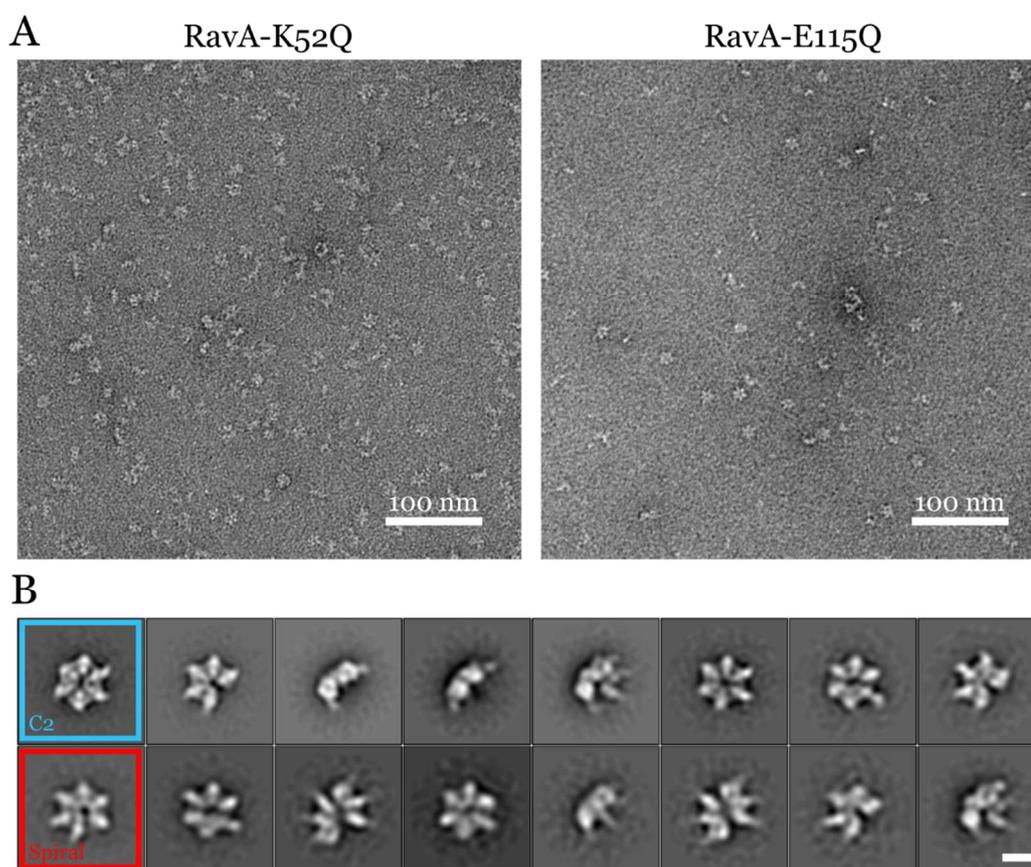


Figure 5.4. **A**) Negative stain EM micrographs of RavA-K52Q (left) and RavA-E115Q (right), scale bar = 100 nm. **B**) 2D class averages of RavA-K52Q. The C2 and spiral conformations are visible in the classes, as indicated by the blue (C2) and red (spiral) boxes. Scale bar = 100 Å.

To better assess the quality of the RavA-K52Q preparation, I collected 88 micrographs and processed them as described for RavA-BC2 in the previous section. Despite the worse appearance of the micrographs, the distinction between the C2 and spiral conformations were even more clear in 2D class averages than for RavA-BC2 as shown in

Figure 5.4B, although at the time we had not yet collected the wild-type RavA cryo-EM dataset and did not recognise the features for what they were. This is likely due to the larger dataset collected. While the classes look promising, they were no better than those I had previously obtained with wild-type RavA (shown for comparison in Figure 5.5) both in terms of visible features and oligomeric homogeneity. We therefore decided not to pursue with the ATPase mutants for the structural characterisation of RavA and instead focus for the time being on obtaining an optimal preparation with wild-type RavA, which was eventually achieved to collect the data presented in Chapter 4.

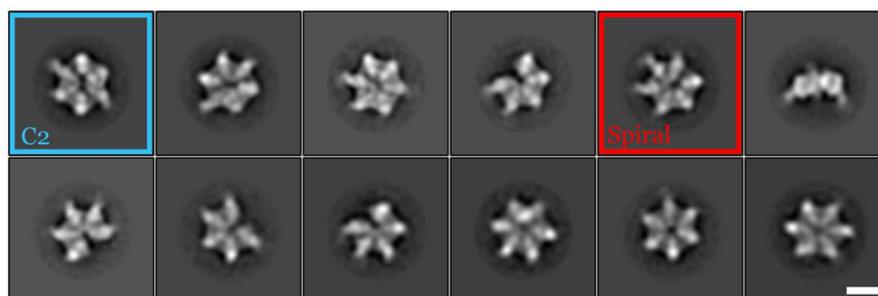


Figure 5.5. 2D classes of wild-type RavA display similar features and oligomeric heterogeneity to classes of RavA-K52Q. The two conformations of RavA are clearly seen, as indicated by the coloured boxes as for Figure 5.2 and Figure 5.4. Scale bar = 100 Å.

### 5.3. RavA + ADP – additional image processing steps

The image processing steps presented in Chapter 4 are the steps that gave the final best maps for the open spiral and C2 state. Here, I will elaborate on a few image processing steps that weren't included in the final manuscript.

The first RavA + ADP dataset was collected without tilting the grid, and resulted in almost exclusively top-views in picked particles. The resulting reconstruction from these ~240,000 particles was extremely stretched in the z-direction because of the lack of side views. Before collecting tilted images, I manually discarded 70,000 particles from 2D classes corresponding to top views, in order to reduce the ratio of top views to side views. This step dramatically improved the resulting map and allowed visualisation of secondary structural features for the first time (Figure 5.6). The map resembled the spiral conformation of RavA seen in the LdcI-RavA complex, although density for the sixth monomer in the spiral was very poor. In reality however, this map contained a mixture of particles corresponding to the C2 state and the spiral state. The nominal resolution of this map as calculated by RELION was 6.07 Å – this in fact was approximately the same as the resolution reported for the final RavA C2 conformation map (5.94 Å). However, upon inspection of the map it is clear that the map is much worse than the 5.94 Å map of the C2 conformation presented in Chapter 4, despite the almost identical reported resolution. This highlights the fact that the nominal resolution is only one measure of a map quality, but often doesn't represent the full story. One very recent paper presents a method for quantifying the directional resolution of cryo-EM maps, which would likely give a more accurate assessment (Vilas et al., 2020). There is a need to develop such alternative metrics in the future, especially for anisotropic datasets such as the RavA datasets presented here.

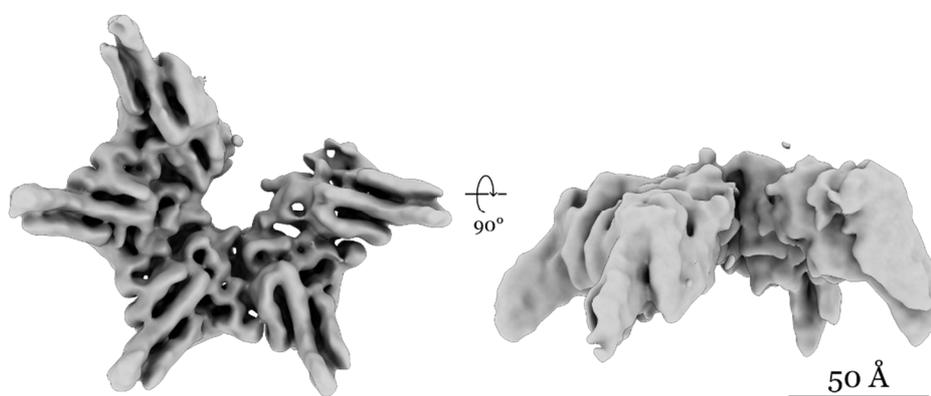


Figure 5.6. 3D reconstruction of RavA + ADP from untilted micrographs after the removal of 70,000 top views. At this threshold, no density is visible for the sixth monomer in the spiral ring. Scale bar = 50 Å.

The second RavA + ADP dataset was collected with a  $30^\circ$  tilt and dramatically improved the resulting maps. Before the tilted dataset was collected, it was not possible to separate out the C2 and spiral states in 3D – it was only after both datasets were combined that 3D classification in RELION yielded the C2 closed-ring conformation. 3D refinement of the spiral conformation gave a map with a nominal resolution of  $6.3 \text{ \AA}$ , which was almost identical to the previous spiral map. Refinement of the C2 conformation (with applied C2 symmetry) yielded a map with a nominal resolution of  $5.0 \text{ \AA}$ . After repicking with the C2 map (as discussed in Chapter 4) and reprocessing, the C2 conformation map improved to a nominal resolution of  $4.36 \text{ \AA}$ . This map is shown in Figure 5.7 – the secondary structural features are significantly improved in the x and y direction but there is still significant anisotropy in the z direction. Therefore, further rounds of classification were carried out to find a subset of particles that would yield improved reconstructions for the spiral and C2 conformations.

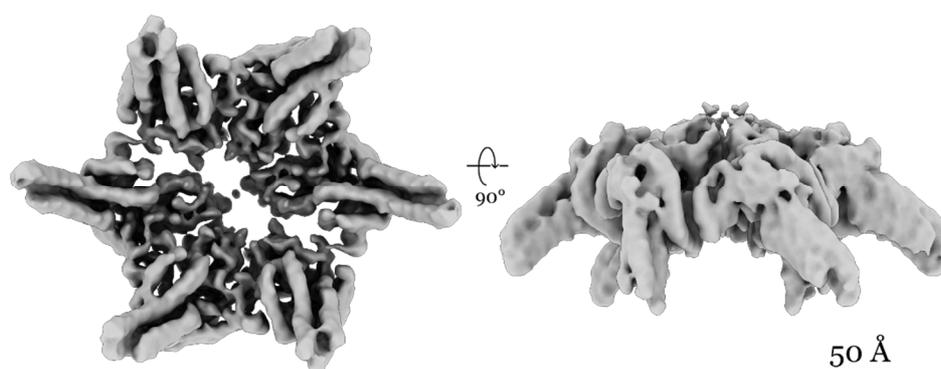


Figure 5.7. 3D refinement of the C2 conformation at a nominal resolution of  $4.36 \text{ \AA}$ . Scale bar =  $50 \text{ \AA}$ .

Many trials were carried out in both RELION and CryoSPARC, such as 2D classification followed by manual selection of particle subsets, 3D classification into different numbers of classes, and *ab initio* 3D reconstruction into multiple classes (in CryoSPARC only). Focussed classification without alignment was also attempted for improving the sixth monomer in the spiral conformation without success. In the end, 3D classification into five classes in RELION followed by 3D refinement of the best class yielded the best spiral map, which is in the final manuscript in Chapter 4.

For the C2 conformation, the  $\sim 257,000$  particles from the previous best refinement in RELION (shown in Figure 5.7) were imported into CryoSPARC. A first homogeneous 3D refinement of all imported particles was carried out against the previous refinement (with applied C2 symmetry), yielding a map with a nominal resolution of  $4.58 \text{ \AA}$  which is shown in Figure 5.8A. *Ab initio* 3D reconstruction of these particles was then carried out into two

classes with no applied symmetry, resulting in two maps that were similar but with one being less stretched in the z-direction (corresponding to ~72,000 particles) than the other (corresponding to ~185,000 particles). Particles from the class with reduced z-stretching were subjected to homogeneous 3D refinement with applied C2 symmetry, resulting in a map with a nominal resolution of 5.94 Å. This map, which is shown in Figure 5.8B and corresponds to the final C2 conformation map presented in Chapter 4, was significantly better than the refinement with all particles, despite having a lower nominal resolution. This again highlights the inadequacy of a single nominal resolution number in accurately conveying the quality of a reconstruction.

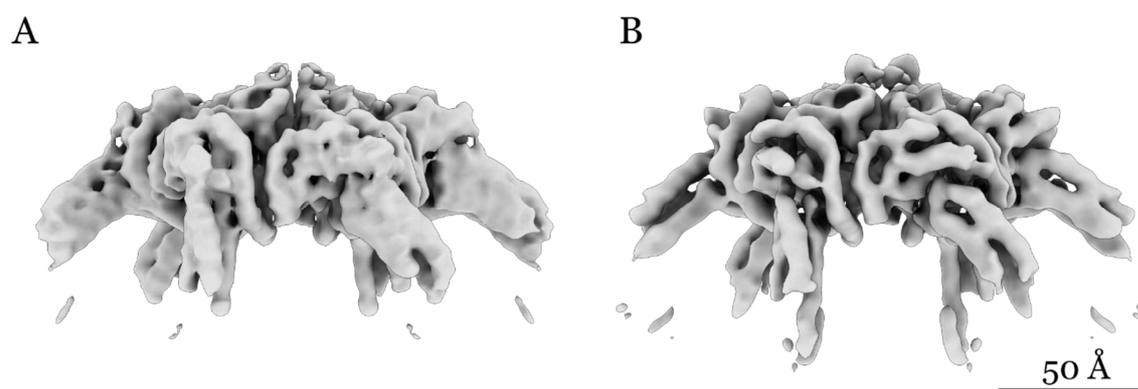


Figure 5.8. **A)** Consensus refinement of all ~257,000 particles in CryoSPARC. **B)** After *ab initio* 3D reconstruction into two classes, a subset of ~72,000 particles refined to a lower nominal resolution but exhibited much better definition of secondary structural features in the z-direction. Scale bar = 50 Å.

Even after extensive classification, significant anisotropy persisted in the spiral map, as well as the C2 map (though to a lesser extent). However, we took the decision to stop at this point because of the availability of the high-resolution RavA crystal structure. The cryo-EM maps presented in the manuscript in Chapter 4 are good enough to allow an unambiguous fit of the crystal structures, which allowed us to make confident assessment of the structural rearrangements of the RavA hexamer between the C2 and spiral conformations. In principle we could have spent significantly more time trialling different conditions for grid freezing with different strategies for overcoming preferential orientation, as discussed further in Chapter 10. However, we decided that the new insights this would have brought would not have justified the added time investment, and pursued other goals instead.

## 5.4. RavA + ATP $\gamma$ S dataset

As described in the manuscript presented in Chapter 4, a dataset with RavA bound to ATP $\gamma$ S, a slowly-hydrolysable ATP analogue, was collected in an attempt to stabilise the active state of RavA. This was done in order to shed further light on the RavA ATPase mechanism, and to understand which of the spiral or C2 conformation would correspond to the active substrate-bound state. However, the conformational heterogeneity was even higher than that of RavA + ADP, with significant populations of seven-fold symmetric and six-fold symmetric closed rings in addition to the C2 and open spiral states. Coupled with the strong preferential orientation, the datasets were not optimal for high-resolution structural analysis, as discussed further in Chapter 10.

In the manuscript in Chapter 4, we stop at the level of presenting 2D class averages. I also carried out 3D analysis of the RavA + ATP $\gamma$ S dataset, but this was not included in the final manuscript. The reconstructions did not reach higher resolutions than the RavA + ADP dataset and were overall of lower quality. Importantly, the density corresponding to bound nucleotides seen in the RavA + ADP dataset was not as clear for the RavA + ATP $\gamma$ S dataset, which prevented the analysis of the nucleotide occupancy of individual ATP binding sites in the hexamer. The anisotropy was also worse for the ATP $\gamma$ S dataset than the RavA + ADP dataset, complicating analysis in 3D.

However, the reconstructions definitively and independently confirm the existence of the two states in equilibrium, in addition to other possible states which may or may not be biologically relevant. After 2D classification in CryoSPARC (as described in the Chapter 4 manuscript), *ab initio* 3D reconstruction was carried out into multiple classes, reproducibly producing C2 and spiral states as well as apparently six-fold symmetric and seven-fold symmetric closed rings. The seven-fold state could not be reconstructed in 3D due to a lack of side views, but the other three states were refined to nominal resolutions of  $\sim 7.3$  Å for the C2 state (with applied two-fold symmetry),  $\sim 8.0$  Å for the open spiral state (with no applied symmetry) and  $\sim 7.4$  Å for the six-fold symmetric closed ring state (with applied six-fold symmetry) as shown in Figure 5.9. The anisotropy in the resulting 3D reconstructions was significantly worse than for the RavA + ADP dataset however, as discussed in Chapter 10. The biological relevance of the C6 state is uncertain, and based on the map it is not possible to determine the number of bound nucleotides. However, it is interesting to note that a recent publication on the Abo1 AAA+ ATPase characterises a six-fold symmetric conformation which is proposed to correspond to a defined state in the ATPase cycle (Cho et al., 2019).

One interesting observation is that for the C2 state, density for the LARA domains appeared at much higher thresholds for the RavA + ATP $\gamma$ S dataset than for the RavA + ADP dataset. For the RavA + ADP dataset the LARA domains only become visible at very low thresholds. For the RavA + ATP $\gamma$ S dataset the LARA domains are much more visible in the C2 map, but there is a rotational smearing of density due to the fact that the LARA domains have high flexibility. Whether or not the ATP $\gamma$ S-bound state has reduced flexibility of the LARAs is unclear. However, because the LARA domain is primarily composed of  $\beta$ -sheets, the resolution of the map prevented any further meaningful analysis.

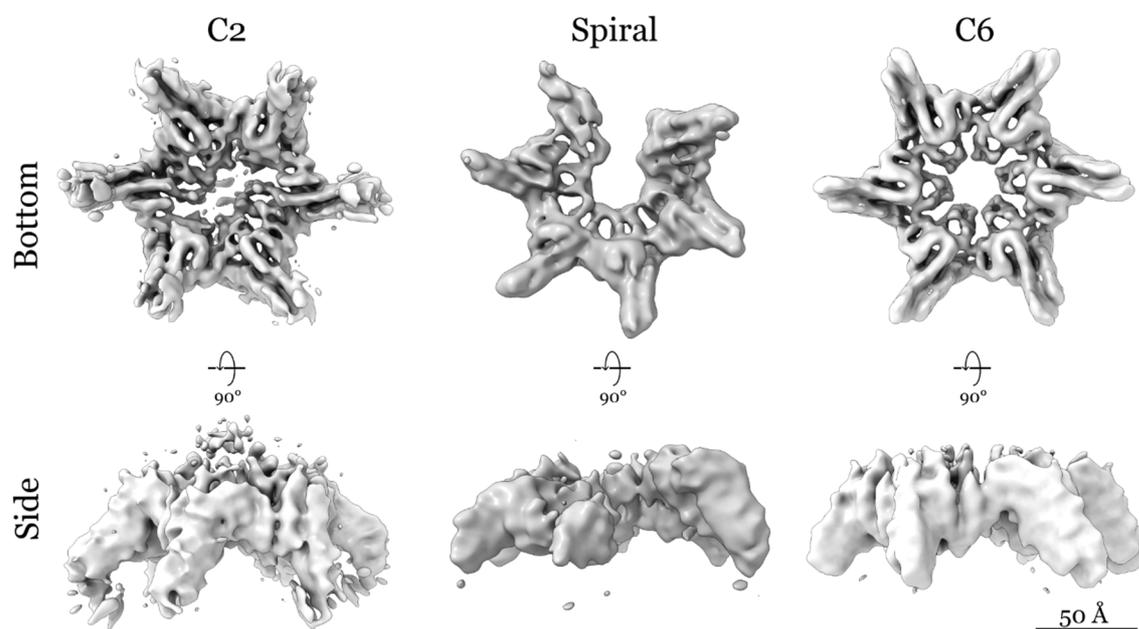


Figure 5.9. 3D reconstructions of the C2, spiral and C6 states of RavA + ATP $\gamma$ S viewed from below and from the side. Scale bar = 50 Å.

## 5.5. Stabilising the substrate-bound conformation

In order to better understand the functional significance of the two RavA states seen (open spiral and closed two-fold symmetric), we decided to investigate the structure of RavA when bound to fluorescein-labelled (FITC) casein, an unfolded protein often used as a model substrate for AAA+ ATPases (Deville et al., 2019; Gates and Martin, 2019; Gates et al., 2017; Rizo et al., 2019). Several cryo-EM structures of AAA+ ATPases have been solved with bound casein that shed light on interactions between pore loops and substrate, and have helped to define the mechanism of AAA+ action (Gates and Martin, 2019; Puchades et al., 2020). We reasoned that having substrate bound may favour the active translocating state of RavA and yield information about how RavA would thread substrates through its central pore.

Therefore, I carried out cryo-EM analysis on RavA with casein added. RavA at a concentration of 1.33  $\mu\text{M}$  was incubated with 2 mM ATP $\gamma$ S for 10 minutes at room temperature, along with 1.33  $\mu\text{M}$  of FITC-casein to give a 1:1 molar ratio of RavA to casein as per Rizo et al., 2019. 3  $\mu\text{L}$  of the mixture was added to UltrAuFoil R 2/2 200 mesh gold grids after glow discharging (20 mA, 45 s). Grids were plunge frozen using a Vitrobot Mark IV operated at 100% humidity at room temperature with a blot force of 1 and blot times of either 2.5 or 3.0 s.

Grids were screened on a Glacios microscope operated under identical conditions to the RavA + ATP $\gamma$ S dataset, as described in Chapter 4. 1,269 movies of 29 frames were collected with an electron dose of 38  $e^-/\text{\AA}^2$ , at a magnification of 116,086 x corresponding to a pixel size of 1.206  $\text{\AA}/\text{pixel}$  at the specimen level. The images were of a lower quality than the RavA + ATP $\gamma$ S dataset – the ice was thicker and there was a higher background in the images, probably corresponding to unbound casein, and as shown in Figure 5.10A. Because of this, only untilted images were collected, as the further loss of contrast from tilting would have prevented processing of the images.

Motion correction, CTF estimation, particle picking and extraction were carried out as described for the RavA + ATP $\gamma$ S dataset in Chapter 4. Particles were imported into CryoSPARC and several rounds of 2D classification were carried out. 2D classes showed that the addition of 1  $\mu\text{M}$  casein did not favour one conformational state of RavA, but rather that the mixture of open spirals and closed rings was retained. One class showed density in the central pore that could possibly correspond to bound casein, as shown in Figure 5.10B. Interestingly, there were relatively more side views in the RavA + ATP $\gamma$ S + casein dataset than for both RavA + ADP and RavA + ATP $\gamma$ S, despite not collecting tilted images. One

hypothesis for this is that casein, which is a hydrophobic unstructured protein, may in fact adsorb to the air-water interface and slightly improve the particle orientation distribution.

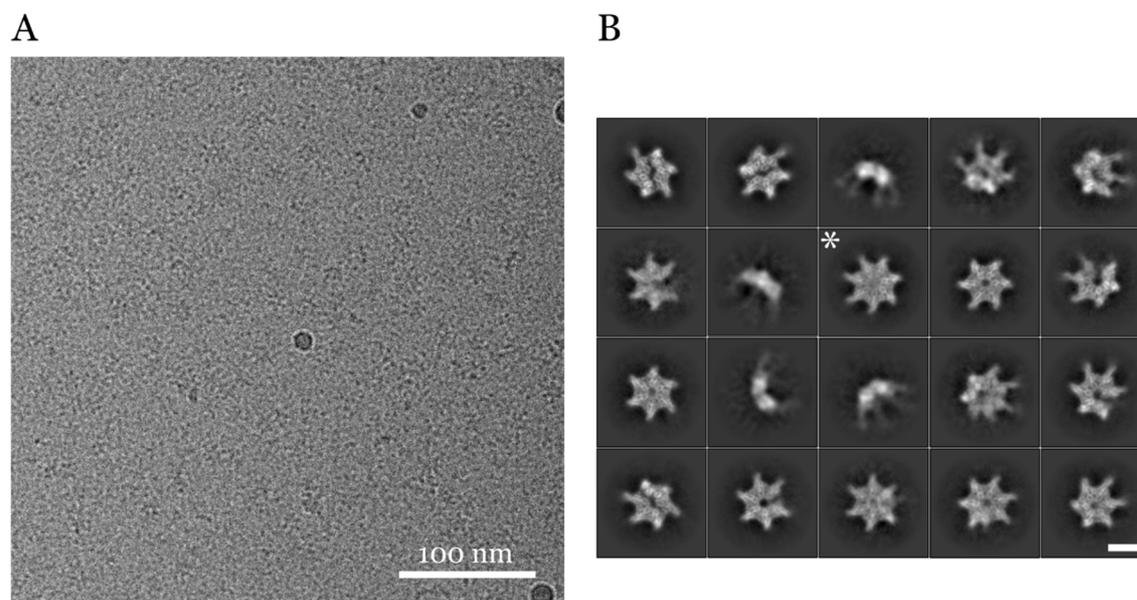


Figure 5.10. **A)** Micrograph of RavA + ATP $\gamma$ S + casein, scale bar = 100 nm. **B)** 2D class averages showing a better distribution of top and side views than for RavA + ATP $\gamma$ S. One class (indicated with an asterisk), corresponding to a seven-fold symmetric oligomer displays density in the central pore, which could possibly be attributed to bound casein. Scale bar = 100 Å.

Both the C2 and spiral conformations could be resolved by *ab initio* 3D reconstruction, and after further processing (several additional rounds of *ab initio* 3D reconstruction followed by homogeneous 3D refinement) the maps were of a similar quality and appearance to those for RavA + ATP $\gamma$ S presented in the previous section. Despite the lack of tilted images, the stretching in the z-direction was similar for both maps,

However, for both the C2 and spiral maps there was no density corresponding to bound casein in the central pore. Therefore, casein was probably bound in substoichiometric amounts. Given more time, additional steps such as running gel filtration with RavA and casein in the presence of ATP $\gamma$ S could have been carried out. Further steps are currently being taken by other members of the group to better characterise the RavA + casein binding for both wild-type RavA and the Walker A mutant K52Q, which should stabilise the substrate-bound active state. This includes the use of fluorescence anisotropy, which indicates an interaction with a low micromolar affinity, amongst other techniques. This will hopefully allow the characterisation of the RavA-substrate interaction in the future and allow the characterisation of the active conformation of RavA.

## Chapter 6. Manuscript on LdcI

In order to investigate the biological function of the LdcI-RavA cage in *E. coli*, we decided to use super-resolution fluorescence microscopy to probe the cellular distribution of LdcI and RavA, in particular under acid stress conditions. To achieve this goal, we created fluorescent protein fusions to both LdcI and RavA. In order to assess the effect of the fluorescent tags on both the structure and function of LdcI and RavA, I expressed and purified fluorescently-tagged LdcI and RavA and used negative stain EM to characterise the fusions. We saw that for LdcI, the addition of a fluorescent protein tag dramatically affected both the structure and function of LdcI.

Therefore, we turned to immunofluorescence studies for characterising the LdcI cellular distribution. To this end we produced anti-LdcI nanobodies, and I characterised the binding of the nanobody to LdcI using negative stain EM. This then enabled STORM imaging to be carried out by Clarissa Liesche on endogenous LdcI, which showed the presence of distinct clusters under acid stress. We then used cryo-EM to solve the structure of stacks of LdcI, which form under acidic conditions, and identify key residues involved in stack formation. The manuscript in this chapter, which is currently under review, presents these results and provides a basis for future studies on the role of LdcI under acid stress conditions.

1 **Title**

2  
3 **Supramolecular assembly of the *E. coli* Ldcl upon acid stress**

4  
5 **Authors**

6 Matthew Jessop\*, Clarissa Liesche\*, Jan Felix\*, Ambroise Desfosses, Megghane Baulard, Virgile  
7 Adam, Angélique Fraudeau, Karine Huard, Grégory Effantin, Jean-Philippe Kleman, Maria Bacia-  
8 Verloop, Dominique Bourgeois, Irina Gutsche#

9  
10 <sup>1</sup>Institut de Biologie Structurale, Univ Grenoble Alpes, CEA, CNRS, IBS, 71 Avenue des martyrs, F-  
11 38044 Grenoble, France

12  
13 \*: equal contribution, #: corresponding author.

14  
15 **Abstract**

16 Pathogenic and commensal bacteria often have to resist the harsh acidity of the host stomach. The  
17 inducible lysine decarboxylase Ldcl buffers the cytosol and the local extracellular environment to  
18 ensure enterobacterial survival at low pH. Here, we investigate the acid-stress response regulation  
19 of *E. coli* Ldcl by combining biochemical and biophysical characterisation with negative stain and  
20 cryo-electron microscopy, and wide-field and super-resolution fluorescence imaging. Due to  
21 deleterious effects of fluorescent protein fusions on native Ldcl decamers, we opt for three-  
22 dimensional localisation of endogenous wild-type Ldcl in acid-stressed *E. coli* cells, and show that it  
23 organises into patches following an apparent long-range pseudo-helical order. Consistent with  
24 recent hypotheses that *in vivo* clustering of metabolic enzymes often reflects their polymerisation as  
25 a means of stimulus-induced regulation, we show that Ldcl assembles into filaments *in vitro* at low  
26 pH. We solve the structures of these filaments and of the Ldcl decamer formed at neutral pH by  
27 cryo-electron microscopy, and reveal the molecular determinants of Ldcl polymerisation, confirmed  
28 by mutational analysis. Finally, we propose a model for Ldcl function inside the enterobacterial cell,  
29 providing a structural and mechanistic basis for further investigation of the role of its supramolecular  
30 organisation in the acid stress response.

31  
32 **Introduction**

33 Cell survival requires the adaptation of metabolism to changing environmental demands.  
34 Biochemical regulation of metabolic enzymes by cellular metabolites has been intensely studied for  
35 many decades. In addition, a growing body of recent light microscopy observations highlights the  
36 spatial regulation of enzymes by stimulus-induced phase separation into distinct loci – liquid droplets,  
37 amyloid-like aggregates or ordered polymers<sup>1</sup>. In eukaryotes, and particularly in yeast, the observed  
38 condensation of fluorescently-labelled metabolic enzymes is triggered by stress, including medium  
39 acidification, hypoxia and nutrient limitation. Enterobacteria such as *Escherichia coli*, *Salmonella* and  
40 *Vibrio* encounter these types of conditions in the host gastrointestinal tract<sup>2</sup>. One of the key  
41 enterobacterial proteins expressed during the acid stress response, upon oxygen limitation and  
42 regulated by the nutrient stress alarmone guanosine tetraphosphate (ppGpp), is the acid stress-

43 inducible lysine decarboxylase LdcI<sup>3-6</sup>. LdcI has been scrutinised since the early 1940s because of  
44 direct links between the efficiency of the acid stress response and pathogenicity<sup>7-9</sup>. This enzyme  
45 transforms lysine into cadaverine while consuming protons and producing CO<sub>2</sub>, thereby contributing  
46 to buffering of the bacterial cytosol and the extracellular medium under acid stress conditions to  
47 promote bacterial survival. While both the structure and the function of LdcI have been thoroughly  
48 studied<sup>6,10</sup>, nothing is known about its localisation inside the bacterial cell.

49 Whereas the overwhelming majority of super-resolution fluorescence imaging is focused on  
50 eukaryotes, optical studies of bacterial systems are nearly exclusively centred on large  
51 macromolecular complexes with obvious superstructure such as cytoskeletal, cell division,  
52 chromosome partitioning, RNA degradation and secretion machineries<sup>11-13</sup>. However, the relevance  
53 of the documented patchy or long-range helical localisation of some of these assemblies is now  
54 questioned and requires re-evaluation. Indeed, the vast majority of these studies were based on  
55 labelling with either fluorescent proteins or epitope tags shown to be able to induce artefactual  
56 associations and localisations<sup>12-15</sup>.

57 A handful of examples of regulation of bacterial metabolic enzymes by phase separation  
58 through stimulus-triggered polymerisation concern well-conserved oligomeric proteins involved in  
59 nucleotide and amino acid metabolism, such as CTP synthase<sup>16,17</sup> and glutamine synthetase<sup>18</sup>.  
60 Interestingly, these enzymes are also able to polymerise into filaments *in vitro*, and their *in vivo*  
61 condensates, detected both in bacteria and in eukaryotes, have been suggested to correspond to  
62 the polymerised state of the enzymes. Other examples of bacterial metabolic enzymes purified as  
63 polymers from bacterial extracts or forming polymers *in vitro* are the aldehyde-alcohol  
64 dehydrogenase AdhE<sup>19</sup> and the hydrogen-dependent CO<sub>2</sub> reductase HDCR<sup>20</sup>.

65 Specific to bacteria, LdcI is a decamer composed of five dimers tightly arranged into  
66 pentameric double-rings (PDB ID: 3N75)<sup>6</sup>. Interestingly, *in vitro*, at pH below 6, these double-rings  
67 were observed by negative stain electron microscopy (ns-EM) to stack on top of one another into  
68 filament-like structures<sup>6,8</sup>. Here, we address the spatial localisation of the *E. coli* LdcI. To critically  
69 evaluate labelling artefacts and define the optimum constructs for subsequent chromosomal  
70 manipulation, we start by overexpressing and purifying different fluorescent protein (FP) fusions of  
71 LdcI, performing their structural and biochemical characterisation *in vitro*, and observing the  
72 distribution of the overexpressed constructs *in vivo*. This methodological section enables us to  
73 propose a workflow that brings together examinations of *in vivo* protein localisations with *in vitro*  
74 biochemical and ns-EM characterisations of the purified FP fusions in order to ensure artefact-free  
75 optical imaging investigations. This analysis is followed by unveiling a supramolecular organisation  
76 of endogenous LdcI inside the *E. coli* cell upon acid stress, coupled to determination of the *in vitro*  
77 structure of the LdcI polymers by cryo-electron microscopy (cryo-EM). For comparison, we also  
78 determined the cryo-EM structure of the LdcI decamer at neutral pH. In addition, mutational analysis  
79 of the LdcI stack-forming interfaces allowed identification of critical residues involved in stack  
80 formation. Finally, we discuss the observed LdcI localisation pattern in the light of the wealth of  
81 available functional and imaging data, and offer a structural and mechanistic basis of supramolecular  
82 LdcI assembly, which will aid in the design of future experiments linking LdcI stack formation to *E.*  
83 *coli* acid stress fitness.

84

85 **Results**

86 **Fluorescent protein fusions affect Ldcl structure without modifying localisation of the**  
87 **overexpressed fusion constructs**

88 Because of the small size of the *E. coli* cell, we opted for super-resolution microscopy imaging  
89 <sup>>21</sup> and set out to localise Ldcl inside the cell upon acid stress by either Photoactivation  
90 Localisation Microscopy (PALM) or Stochastic Optical Reconstruction Microscopy (STORM)<sup>22–24</sup>. *A*  
91 *priori*, PALM seemed more relevant because this technique relies on genetically-encoded FPs fused  
92 to the protein of interest, and is therefore unmatched in terms of labelling specificity and efficiency.  
93 In addition, PALM does not require delivery of fluorescent molecules across the cell wall, and  
94 enables live cell imaging and single molecule tracking.

95 Considering that Ldcl is an acid stress response enzyme, we opted for either mGeosM<sup>25</sup> or  
96 Dendra2<sub>T69A</sub><sup>26</sup> as FP markers because of their relatively low pKa values, high monomericity and high  
97 fluorescent quantum yields. In the Ldcl decamer structure, the N-termini are oriented inwards and  
98 towards the central pore of the double ring, while the C-termini point to the ring periphery and are  
99 readily accessible from the outside. Therefore, on the one side, one could assume that an N-terminal  
100 labelling with an FP would be likely to interrupt the Ldcl tertiary structure. On the other side, the only  
101 well characterised binding partner of Ldcl, the AAA+ ATPase RavA, is known to interact precisely  
102 with the C-terminal  $\beta$ -sheet of Ldcl<sup>27</sup>. Assembly of two copies of double-pentameric rings of Ldcl and  
103 five copies of hexameric RavA spirals results in a huge 3.3 MDa macromolecular cage of intricate  
104 architecture and largely unknown function<sup>27–29</sup>. The Ldcl-RavA cage is proposed to assist assembly  
105 of specific respiratory complexes in *E. coli* and to counteract acid stress under starvation conditions  
106 by alleviating ppGpp-dependent inhibition of Ldcl<sup>30–32</sup>. Thus, although these functions still require  
107 further investigation, preservation of the RavA-binding propensity should be one of the criteria for  
108 assessing the suitability of an Ldcl-FP fusion. Therefore, both N- and C-terminal fusion constructs  
109 with either mGeosM or Dendra2<sub>T69A</sub> attached to Ldcl via an appropriate linker were cloned into  
110 dedicated plasmids and overexpressed at conditions optimised for Ldcl overproduction (see  
111 Methods for details). Expression of fusion proteins was immediately detected by wide-field  
112 fluorescence imaging that showed a similar distribution for the four fusions (Figure 1A). Each  
113 construct was then purified in order to assess its structural integrity and RavA binding capacity *in*  
114 *vitro*, with a goal of defining the most suitable construct for the subsequent creation of a  
115 corresponding chromosomal fusion.

116 The N-terminal Dendra2<sub>T69A</sub>-Ldcl fusion formed exclusively dimers, confirming the structure-  
117 based prediction that the fluorescent tag at this position would perturb the dimer-dimer interaction  
118 (Figure 1B, Supplementary Figure 1). Admittedly, native Ldcl was shown to dissociate into dimers *in*  
119 *vitro* at pH above approximately 7.5, but in this pH range Ldcl is not supposed to be expressed in  
120 the cell and therefore this dissociation may be irrelevant. Surprisingly, in contrast to Dendra2<sub>T69A</sub>-  
121 Ldcl, mGeosM-Ldcl assembled into regular symmetric non-native higher-order oligomers, with a  
122 dramatically altered quaternary structure (Figure 1B, Supplementary Figure 1). These oligomers  
123 were built of three Ldcl tetramers, bridged together by additional densities. Noteworthy, mEos2, the  
124 fluorescent protein from which mGeosM was derived, crystallises as a tetramer (PDB ID: 3S05) that  
125 can be straightforwardly fitted into the Ldcl-bridging densities (Figure 1B). This illustrates that despite  
126 the fact that mGeosM was designed as a monomeric FP (with the first “m” explicitly standing for

127 monomeric), some residual oligomeric tendency is still maintained. This propensity of mGeosM to  
128 oligomerise when bound to Ldcl may be driven by avidity effects – as Ldcl dimers begin to assemble  
129 into a decamer, the local concentration of mGeosM increases to such a point that oligomerisation  
130 becomes energetically favourable, despite the apparent monomer behaviour of mGeosM in gel  
131 filtration<sup>25</sup>. This is also in line with the known propensity of mEos2, from which mGeosM is derived,  
132 to form tetramers at high concentration<sup>33</sup>. To conclude, both N-terminal fusions induced non-native  
133 assembly of Ldcl, and therefore neither were appropriate for determining the native localisation of  
134 this enzyme in the cell.

135 As expected, C-terminal Ldcl fluorescent fusions formed decamers with protruding densities  
136 that can be attributed to flexibly attached FPs (Figure 1B, Supplementary Figure 1). Nevertheless,  
137 despite the long flexible linker between Ldcl and the FP, these constructs were unable to interact  
138 with RavA as shown by Bio-Layer Interferometry (BLI) binding studies (see Methods and  
139 Supplementary Figure 1 for details). This means that the functionality of these fusions cannot be  
140 considered as entirely retained. Thus, none of the four fusions were suitable for an in-depth imaging  
141 analysis under conditions of native Ldcl expression upon acid stress that we planned to undertake.  
142 Altogether, we demonstrated that structural integrity and unaltered interaction with known partners  
143 are useful read-outs for functional preservation. Based on this result we propose that, when feasible,  
144 purification and structure-function analysis of FP fusions should be performed prior to interpretation  
145 of the protein localisation inferred from observation of FP fusions by optical methods. This workflow  
146 can be used for example in cases where chromosomal manipulation for assessment of intact function  
147 is difficult and/or the phenotype is condition-dependent or unclear, whereas basic *in vitro* biochemical  
148 and ns-EM characterisation are efficiently set up.

149

## 150 **Immunofluorescence of Ldcl in *E. coli* under acid stress reveals its supramolecular** 151 **organisation**

152 Because none of the FP fusions possessed the properties of native Ldcl, we decided to turn  
153 to STORM imaging of the endogenous enzyme. One of the drawbacks of STORM is the requirement  
154 for exogenous labelling and therefore the fixation and permeabilisation of cells. These procedures  
155 are notoriously known to potentially affect cell morphology, and therefore, when possible, the usage  
156 of live cell imaging with FP fusions (as for example in PALM), or observation of unfixed samples  
157 under cryogenic conditions would be ideal. Moreover, fixation, permeabilisation and specific  
158 exogenous labelling in bacteria present unique challenges because of the complex, often multi-  
159 layered cell wall<sup>34,35</sup>. A significant advantage of STORM however is the possibility of direct imaging  
160 of the wild-type (WT) protein, which should circumvent the dangers associated with FP fusions  
161 described above. The three prerequisites for imaging of WT systems by immuno-based labelling in  
162 general and of Ldcl in particular are (i) availability of an antibody or nanobody (antibody fragment  
163 derived from heavy-chain-only camelid antibody) coupled to an organic fluorescent dye and directed  
164 towards the native protein, (ii) precise knowledge of endogenous expression conditions, (iii)  
165 validation of an efficient permeabilisation and immunolabelling technique that enables the  
166 antibody/nanobody to enter the cell and specifically target the protein of interest without altering the  
167 native organization of the cell ultrastructures.

168 As a first step towards STORM imaging of endogenous LdcI, we probed an anti-LdcI  
169 nanobody, hereafter called anti-LdcI-Nb. The complex between purified *E. coli* LdcI and the  
170 nanobody was purified by size exclusion chromatography (see Methods), and imaged by ns-EM.  
171 The resulting 3D map of the LdcI/anti-LdcI-Nb complex demonstrates binding of anti-LdcI-Nb to each  
172 LdcI monomer in the decamer and reveals the location of the interaction site which is clearly distinct  
173 from the RavA binding site (Figure 2A, Supplementary Figure 2). This nanobody was thus identified  
174 as a suitable labelling agent for LdcI imaging in *E. coli* cells, and labelled with the fluorescent dye  
175 AlexaFluor647 (AF647) or AlexaFluor488 (AF488) (see Methods).

176 Consistently with published data, LdcI expression could be induced by a pH shift experiment  
177 (i.e. transfer of bacteria from pH 7.0 into a pH 4.6 growth medium) in the presence of lysine under  
178 oxygen-limiting conditions (Supplementary Figure 3A). While the WT strain grew well under these  
179 conditions and efficiently buffered the extracellular medium up to pH 6.2 in approximately 1.5 to two  
180 hours (Supplementary Figure 3A) concomitantly with the increase in the level of LdcI expression  
181 (Supplementary Figure 3B, C), the growth and the acid stress response of the  $\Delta/dcl$  mutant strain  
182 were severely impaired (Supplementary Figure 3A). Since the peak of LdcI expression by the wild  
183 type cells was achieved between one and two hours after exposure to acid stress (Supplementary  
184 Figure 3B, C), a time point of 90 minutes was chosen for the subsequent labelling and imaging  
185 experiments. The specificity and performance of anti-LdcI-Nb in immunofluorescence labelling of  
186 permeabilised *E. coli* cells were characterised by flow cytometry and wide-field fluorescence imaging  
187 (Supplementary Figure 3D, E). Both techniques demonstrated that in the absence of LdcI expression  
188 no specific fluorescence is seen, confirming the suitability of anti-LdcI-Nb for immunofluorescence  
189 studies. Thus, the above-mentioned prerequisites for immuno-based imaging of cellular LdcI have  
190 been fulfilled. Noteworthy, the LdcI expression profile highlighted a considerable asset of the usage  
191 of STORM instead of PALM for visualisation of the endogenously expressed LdcI: indeed, the  
192 transient nature of the expression and the necessity of work under oxygen-limiting conditions may  
193 have created difficulties due to the longer maturation time of the FPs under these conditions.

194 Initial characterisation of the cellular distribution of LdcI 90 minutes after exposure of *E. coli*  
195 cells to acid stress was carried out by wide-field fluorescence imaging. Based on these images, it  
196 appeared that natively expressed LdcI did not display a homogeneous cytoplasmic distribution but  
197 rather showed a patchy localisation pattern (Figure 2B). 3D STORM imaging subsequently provided  
198 a more detailed view of this patchy distribution. As shown in Figure 3, Supplementary Figure 4, and  
199 Supplementary Movies 1-5, the labelling density was lower in the centre of the bacterium. This  
200 indicates a propensity for LdcI to cluster near the cell periphery, towards the inner membrane and  
201 the cell poles, rather than being distributed homogeneously through the volume. Recent  
202 investigations argue that, in most cases, *in vivo* clustering of metabolic enzymes corresponds to their  
203 polymerised states, and represents an efficient means of regulation of enzymatic activity and  
204 metabolic homeostasis in response to a stimulus<sup>1</sup>. Thus, structure determination of these polymers  
205 is a crucial step towards the understanding of regulation mechanisms. Using ns-EM, we previously  
206 documented that at low pH and high concentration, LdcI decamers tend to stack<sup>6</sup>. Therefore, the  
207 high local concentration of LdcI, clustered in patches in the *E. coli* cell under acid stress conditions  
208 boosting LdcI expression, is likely to induce enzyme polymerisation via stack formation. Remarkably,  
209 the images convey a visual impression that individual patches tend to be further arranged into bands

210 or helix-like turns, showing a circumferential distribution with a sort of a long-range stripy or pseudo-  
211 helical organisation (Figure 3, Supplementary Figure 4, Supplementary Movies 1-5).

### 212 213 **Structural determinants of Ldcl stack formation revealed by cryo-EM**

214 In order to understand the molecular determinants of Ldcl polymerisation at low pH and to  
215 provide a framework for the analysis of Ldcl function under acid stress, we solved the 3D structures  
216 of the Ldcl decamer at pH 7.0 and of Ldcl stacks at pH 5.7 by cryo-EM (See Methods, Figure 4,  
217 Supplementary Figures 5-6, Supplementary Table 1). The 2.8 Å resolution structure of the Ldcl  
218 decamer at pH 7.0 (Figure 4A) is extremely similar (Supplementary Table 2 & 3) to the Ldcl crystal  
219 structure solved at pH 8.5 in an inhibited ppGpp-bound state<sup>6</sup>. However, contrary to pH 7.0 and even  
220 pH 6.2 where Ldcl is still predominantly decameric<sup>10</sup>, Ldcl forms straight rigid filaments on the cryo-  
221 EM grid at pH 5.7, which corresponds to the pH of maximum Ldcl enzymatic activity (Figure 4B).  
222 The structure of a three-decamer stack was solved to a resolution of 3.3 Å, revealing the structural  
223 details of acid stress-induced Ldcl polymerisation (Figure 4C). Ldcl decamers stack tightly on top of  
224 one other, with negligible rotation between decamers along the stack. Each dimer fits snugly in the  
225 inter-dimer groove of decamers above and below.

226 A comparison of Ldcl decamer structures taken from the Ldcl stack cryo-EM map (at pH 5.7),  
227 the Ldcl decamer cryo-EM map (at pH 7.0) and the crystal structure of decameric Ldcl crystallised  
228 with bound ppGpp at pH 8.5 (PDB ID: 3N75) reveals some remarkable differences between the stack  
229 structure and the two decamer structures (Figure 4D, E, Supplementary Table 2 & 3). While the three  
230 structures do not show any major differences at the monomer level, a structural alignment of an Ldcl  
231 dimer extracted from the Ldcl decamer structures with a dimer extracted from the Ldcl stack structure  
232 uncovers a rigid body-like rotation between monomers around a hinge region located at the  
233 monomer-to-monomer interface (Supplementary Table 2 & 3, Figure 4DE). This rotation results in a  
234 5° tilt when comparing the N-terminal wing-domains in Ldcl dimers, and an overall slightly decreased  
235 diameter of the central cavity inside the stacked Ldcl rings (Figure 4D, E), which may contribute to  
236 the tight packing of each dimer into the grooves of an opposing decamer in the stack.

237 A careful examination of the Ldcl stack structure shows that two major inter-decamer  
238 interfaces situated at a two-fold symmetry axis perpendicular to the stack direction contribute to stack  
239 formation (Figure 5A). In particular, the first interface (Figure 5B) is formed between residues K422,  
240 D460, R468, D470, and E482 situated in the ppGpp-binding domain (amino acids 418-564), and  
241 residues N314, D316, and G352 from the PLP-binding domain (amino acids 184-417). D460 from  
242 one decamer makes an electrostatic interaction with K422' of a neighbouring decamer in the stack.  
243 R468 is sandwiched between D316' and E344' from a neighbouring decamer, and makes  
244 electrostatic interactions through its  $\eta$ 1 and  $\eta$ 2 nitrogen atoms with D316'. In addition, D470 interacts  
245 with the backbone of G352', and E482 forms hydrogen bonds with N314'. The second interface  
246 (Figure 5B) is formed between residue N94 of the wing domain (amino acids 1-130) of one set of  
247 opposing dimers, and a stretch of four residues in the ppGpp-binding domain – T444, E445, S446  
248 and D447 - at the end of helix  $\alpha$ 16 from a second set of opposing dimers. The wing domain residue  
249 N94 makes hydrogen bonds with E445' of an opposing dimer. A second charged residue, D447,  
250 interacts with the backbone of T444', and is held in place by R97 from the wing domain of a  
251 neighbouring dimer in the same decamer.

252 Considering that the Ldcl polymerisation is induced by acid stress, we wondered which  
253 residues in the interface would be sensitive to pH changes. Surprisingly, most of the side chains  
254 involved in the inter-decamer interface are charged arginine (pKa 13), aspartate and glutamate  
255 residues (pKa of 4 and 3 respectively), which do not change protonation state in the pH window  
256 relevant for Ldcl activity (pH 5-7)<sup>36</sup>. Nonetheless, other residues, situated outside the interaction  
257 interfaces, may drive stack formation through pH-dependent interactions that would in turn lead to  
258 the observed inter-monomer rotation and the associated constriction of the Ldcl central cavity,  
259 coupled to the alignment of complementary contacts at interfaces one and two (Figure 4D, E). We  
260 note for example that H694 should be protonated in the Ldcl stack structure at pH 5.7 but  
261 deprotonated in the two ring structures at pH 7.0 and pH 8.5, and that an electrostatic interaction  
262 between H694 and D192 situated in the linker region is present in the stack structure only (Figure  
263 5E).

264 To validate the observed interactions at the inter-decamer interface, and to assess the  
265 individual importance of key residues involved in Ldcl stack formation, we constructed four Ldcl point  
266 mutants (R468E, R97E, H694A and H694N), two double mutants (R97E/R468E and E445A/D447A)  
267 and one triple mutant (E445A/D447A/R468E). The mutants were purified following the protocol for  
268 wild-type Ldcl (see Methods), diluted into a buffer at pH 5.7 and observed by ns-EM (Figure 6,  
269 Supplementary Figure 7). Although the grid preparation procedure for ns-EM yields stacks that are  
270 shorter and more curved and distorted when compared to the cryo-EM data (Figure 4B, Figure 6),  
271 our previous observations of the five paralogous *E. coli* amino acid decarboxylases justify the validity  
272 of this approach for a qualitative comparative analysis<sup>6,36</sup>. Ns-EM images make immediately  
273 apparent that the E445A/D447A double mutant does not have a significantly altered capability of  
274 stack formation at pH 5.7 when compared to WT Ldcl, whereas the single R468E mutation is  
275 sufficient to completely abolish stack formation. The R97E mutant has a moderate destabilising  
276 effect and displays fewer and smaller stacks than the WT Ldcl. Consistently with the major role of  
277 R468 in the Ldcl stack formation, the R97E/R468E double mutant and E445A/D447A/R468E triple  
278 mutant exclusively occur as decamers at low pH. Altogether, these results reveal that R468 is one  
279 of the key determinants of Ldcl stack formation. Finally, similarly to the R97E mutant, a modest  
280 destabilisation of stack formation is observed for the two histidine mutants, H694A and H694N,  
281 favouring our hypothesis that H694 may have an influence on the propensity of Ldcl polymerisation  
282 at low pH (Figure 6, Supplementary Figure 7).

283

## 284 Discussion

285 Our synergistic approach, combining several *in vitro* techniques including biochemical  
286 characterisation of purified fluorescent protein fusions, ns-EM observation of mutants, low resolution  
287 ns-EM reconstruction and high resolution cryo-EM analysis, with *in vivo* flow cytometry, wide-field  
288 and 3D-STORM imaging, provides insights into the supramolecular Ldcl assembly upon acid stress.  
289 This work adds to the very few known examples of regular polymerisation as means of regulation of  
290 enzymes involved in amino acid metabolism in bacteria. The cryo-EM structure of the Ldcl stacks  
291 presented here offers a molecular framework for future investigation of the role of the Ldcl  
292 polymerisation in the acid stress response.

293 The high resolution cryo-EM structures of the stack at pH 5.7 and of the decamer at pH 7.0  
294 complement the previously solved crystal structure of the ppGpp-bound LdcI decamer. Previous  
295 serendipitous co-crystallisation of LdcI with the nutrient stress response alarmone ppGpp led to  
296 assessment of the effect of ppGpp on LdcI activity, and to a proposal that ppGpp would act as an  
297 LdcI inhibitor that prevents excessive lysine consumption upon nutrient limitation during acid stress<sup>6</sup>.  
298 In addition, a ppGpp-dependent disassembly of the LdcI stacks had been previously observed but  
299 could not be structurally explained since the ppGpp binding site is situated between two neighbouring  
300 dimers inside the LdcI ring<sup>6</sup>. Our cryo-EM structures show that one of the residues involved both in  
301 ppGpp binding and in the stack formation is arginine 97 (R97). In the crystal structure of ppGpp-  
302 bound LdcI, R97 makes a stacking interaction with the guanosine imidazole ring of ppGpp, while in  
303 LdcI stacks R97 is involved in a key interaction at the second interface (Figure 5D), where it locks  
304 D447 in a conformation allowing interactions between helices  $\alpha$ 16 from opposing LdcI decamers.  
305 Binding of ppGpp to LdcI interferes with the R97-D447 interaction, thereby most likely prohibiting  
306 correct positioning of D447 at the tip of helix  $\alpha$ 16, and resulting in a disruption of the second stack  
307 interface, leading to a moderate yet notable stack destabilisation (Figure 6, Supplementary Figure  
308 7). Furthermore, our cryo-EM structure of ppGpp-free LdcI decamers at neutral pH enables to rule  
309 out the effect of ppGpp on the differences observed between the LdcI stack structure at pH 5.7 and  
310 the ppGpp-LdcI crystal structure at pH 8.5. Indeed, despite the absence of ppGpp, R97 is still  
311 oriented towards the ppGpp binding site and away from the inter-decamer interface in the pH 7.0  
312 decamer cryo-EM map (Figure 5D). This suggests that the conformational changes in LdcI driving  
313 stack formation are mostly driven by low pH and not by the absence of ppGpp, although the D192  
314 and H694 hinge residues are similarly far apart in the pH 8.5 and pH 7.0 structures (Figure 5E). Our  
315 current work provides a solid experimental and structural basis for a future closer evaluation of the  
316 hypothesised role of ppGpp in LdcI regulation *in vitro* and *in vivo*<sup>6</sup>

317 The clusters of endogenous LdcI at the bacterial periphery observed by optical imaging would  
318 presumably correspond to LdcI assembled into stacks. Possible reasons for such assembly may be  
319 to provide an efficient way to locally increase the LdcI concentration and to enhance its activity. But  
320 why would LdcI, an apparently highly soluble protein, be driven towards the inner membrane? What  
321 would be the advantage for acid-stressed *E. coli* cells to increase the concentration of LdcI via stack  
322 formation in these particular peripheral locations? Localisation of proteins to specific sites in the  
323 bacterial membrane was shown to be generally driven by chemical factors such as the phospholipid  
324 composition of the lipid microdomains, and by physical factors such as the degree of local curvature  
325 or the electric potential of the membrane<sup>37</sup>. An attractive hypothesis would be that as an acid stress  
326 response protein performing a proton-consuming enzymatic reaction, LdcI may be attracted to proton  
327 sinks formed by anionic phospholipids which compartmentalise oxidative phosphorylation  
328 (OXPHOS) complexes for efficient functioning in bacterial respiration and adaptation to  
329 environmental changes. Indeed, OXPHOS complexes were often described to be unevenly  
330 distributed in the membrane in the form of mobile patches<sup>38-42</sup>, providing evidence for highly dynamic  
331 and spatially organised bioenergetic membranes in *E. coli* cells<sup>42</sup>. In addition, certain bacterial  
332 flotillins, which are essential scaffold proteins of the functional membrane microdomains, equivalent  
333 to the lipid rafts of eukaryotic cells, also show a patchy distribution and were shown to interact with  
334 specific OXPHOS complexes<sup>42,43</sup>. In this regard, two different lines of evidence would be interesting

335 to note. First, Ldcl was described to co-purify with a partially assembled Complex I<sup>44</sup>, whereas the  
336 Ldcl-binding partner RavA, as well as ViaA, the second protein expressed from the *ravAviaA* operon  
337 and which also interacts with RavA, were shown to interact with both Complex I and fumarate  
338 reductase<sup>31,32</sup>. Second, the other *E. coli* PLP-dependent lysine decarboxylase LdcC, exercising the  
339 role of cadaverine biosynthesis irrespectively of acid stress<sup>45</sup>, neither binds RavA nor forms  
340 stacks<sup>28,36</sup>, in spite of its 69% of identity with Ldcl. It is interesting to note that in LdcC the key stack-  
341 forming residue R468 has been substituted for an alanine while the interacting D316 has been  
342 preserved.

343 The reason for the arrangement of the Ldcl clusters into stripes or pseudo-helical patterns is  
344 also intriguing. While similar distributions have been documented for bacterial cytoskeletal, cell  
345 division, chromosome partitioning, RNA degradation and secretion machineries, an eventual impact  
346 of labelling on these distributions, demonstrated specifically for the YFP-MreB<sup>46</sup>, warrants caution in  
347 their interpretation<sup>12,14</sup>. Here however, we observed endogenous wild-type Ldcl in cells fixed prior to  
348 their permeabilisation and for Ldcl labelled with anti-Ldcl-Nb, which means that the resulting pattern  
349 is likely real and not an artefact. In addition, examination of some published images of OXPPOS  
350 patches (for example Llorente-Garcia et al., 2014) also seem to hint to a possible pseudo-helical  
351 organisation. Excitingly, anionic phospholipid-specific dyes and fluorescently-labelled antibiotics  
352 specific for nascent peptidoglycan synthesis upon cell elongation were also described to be  
353 distributed on helical or stripe patterns<sup>37,47-50</sup>. It is therefore tempting to imagine that inside the  
354 bacterial cell, Ldcl has a tendency to follow a general path upon polymerisation governed by the  
355 underlying patterns in the cell envelope, which are proposed to follow a global right-handed chiral  
356 order<sup>51</sup>.

357 Finally, from the methodological view, our work convincingly illustrates that different FP fusion  
358 constructs can share the same cellular distribution in spite of a completely different structure,  
359 necessitating caution when inferring intact function from the preservation of the protein localisation  
360 inside the cell. Our findings emphasize the importance of characterising FP fusions using both  
361 biochemical and structural techniques, such as ns-EM, to ensure that the FP tag disrupts neither  
362 structure nor function of the target protein.

363

## 364 **Methods**

### 365 **Expression constructs**

366 For fluorescence studies, several FP fusion constructs were generated starting from an  
367 available plasmid containing the coding sequence of Ldcl (Uniprot entry P0A9H3), cloned in the  
368 pET22b(+) vector with a C-terminal 6xHis-tag<sup>6</sup>. All constructs were generated using the Gibson  
369 cloning strategy and verified by sequencing analysis. The Gibson assembly was performed using  
370 0,4U T5 exonuclease (NEB M0363S), 2,5U Phusion polymerase (NEB M0530S) and 400U Taq  
371 ligase (NEB M0208L) in 1X ISO buffer (100mM Tris-HCl pH 7,5, 10 mM MgCl<sub>2</sub>, 0.8 mM dNTP mix,  
372 10mM DTT, 50 mg PEG-8000, 1 mM NAD). 7.5  $\mu$ L of the GIBSON Master Mix was mixed with 2.5  
373  $\mu$ L DNA, containing circa 100 ng of vector. The mix was incubated for 60 min at 50°C.  
374 Transformations were performed in Top10 competent bacteria (One Shot™ TOP10 Chemically  
375 Competent *E. coli*, Invitrogen C404003) and selected using 100  $\mu$ g/mL ampicillin or 50  $\mu$ g/mL  
376 kanamycin sulphate from Euromedex (Ampicillin ref EU0400-D, Kanamycin ref EU0010-D). Agarose

377 Gel purification and DNA plasmid extraction kits were purchased from Macherey-Nagel (NucleoSpin  
378 Plasmid ref 740588-250 and NucleoSpin Gel and PCR Clean-up ref 740609-250).

379 Dendra2<sub>T69A</sub>-Ldcl and mGeosM-Ldcl were both cloned in the pET-TEV vector containing an  
380 N-terminal 6xHis-tag, and a TEV cleavage site between Dendra2<sub>T69A</sub> or mGeosM and the Ldcl gene.  
381 Ldcl-Dendra2<sub>T69A</sub> and Ldcl-mGeosM were both cloned in the pET22b(+) vector with Ldcl followed by  
382 either Dendra2<sub>T69A</sub> or mGeosM containing an uncleavable C-terminal 6XHis-tag.

383 The sequence of a anti-Ldcl-Nb was cloned in the pHEN6 vector containing the pelB leader  
384 sequence from *Erwinia carotovora* for secretion into the periplasm, and a C-terminal 6xHis-tag  
385 (kindly provided by Dr. Aline Desmyte, AFMB Marseille).

386 Plasmids, primers and cloning strategy are summarised in Supplementary Table 4.

387

### 388 **Protein purification**

389 Ldcl-FP fusions and Ldcl mutants were expressed in BL21(DE3) cells grown in LB medium  
390 supplemented with 100 µg/mL ampicillin or 50 µg/mL kanamycin sulphate. Protein expression was  
391 induced using 40 µM IPTG (Euromedex EU008-C) and carried out overnight at 18°C. The Ldcl-FP  
392 fusions and Ldcl mutants were purified as previously described for wild-type Ldcl<sup>27,28,36</sup>, in a final  
393 buffer containing 25 mM Tris (pH 7.5), 0.3 M NaCl, 5% glycerol, 1 mM DTT and 0.1 mM PLP.

394 The anti-Ldcl-Nb was expressed in *E. coli* WK6 cells following the protocol described by<sup>52</sup>,  
395 and purified with immobilised metal affinity chromatography (IMAC, using a Ni-NTA column) followed  
396 by Size exclusion chromatography (SEC) using a a superdex 75 Increase 10/300GL column (GE-  
397 Healthcare, ref: 29-1487-21) equilibrated with a buffer containing 25 mM Tris pH 7.4 and 0.3 M NaCl.

398 In order to characterise the Ldcl complex with anti-Ldcl-Nb, the two proteins were mixed at  
399 a 1:5 molar ratio and submitted to size exclusion chromatography as carried out for Ldcl alone but  
400 without DTT in the buffer. The top of the peak was taken for subsequent ns-EM analysis.

401

### 402 **Ns-EM on Ldcl-FP fusions, Ldcl/anti-Ldcl-Nb and Ldcl mutants**

403 FP fusion samples after gel filtration were diluted to a concentration of approximately 0.025  
404 mg/mL. 3 µL was applied to the clean side of carbon on a carbon–mica interface and stained with  
405 2% uranyl acetate (mGeosM-Ldcl, Ldcl-mGeosM, Ldcl-Dendra2<sub>T69A</sub>, Ldcl/anti-Ldcl-Nb, Ldcl  
406 mutants) or 2% sodium silicotungstate (Dendra2<sub>T69A</sub>-Ldcl). Images were collected on a 120 kV  
407 Tecnai T12 microscope with an Orius 1000 camera (Gatan) or on a 200 kV Tecnai F20 electron  
408 microscope with either a OneView camera (Gatan) or a Ceta camera (Thermo Scientific). All images  
409 were collected with a defocus range of approximately -1.0 µm to -2.0 µm and with pixel sizes between  
410 2.29 Å/pixel and 3.42 Å/pixel.

411

### 412 **Image processing – Ldcl-FP and Ldcl/anti-Ldcl-Nb**

413 36 micrographs of Dendra2<sub>T69A</sub>-Ldcl with a pixel size of 2.73 Å/pixel, 368 micrographs of  
414 mGeosM-Ldcl with a pixel size of 2.29 Å/pixel, 23 micrographs of Ldcl-Dendra2<sub>T69A</sub>, 92 micrographs  
415 of Ldcl-mGeosM with a pixel size of 3.42 Å/pixel, and 124 micrographs of Ldcl/anti-Ldcl-Nb with a  
416 pixel size of 2.82 Å/pixel were used for image analysis.

417 CTF estimation was performed with CTFFIND3<sup>53</sup>. Semi-automatic particle selection was  
418 carried out with BOXER<sup>54</sup>, with box sizes of 98 x 98 pixels for Dendra2<sub>T69A</sub>-Ldcl, 180 x 180 pixels for

419 mGeosM-Ldcl, 128 x 128 pixels for Ldcl-Dendra2<sub>T69A</sub> and Ldcl-mGeosM, and 112 x 112 pixels for  
420 Ldcl/anti-Ldcl-Nb respectively. Particle extraction followed by several rounds of cleaning by 2D  
421 classification in RELION-1.4<sup>55</sup>, resulted in the following number of particles for each dataset:  
422 Dendra2<sub>T69A</sub>-Ldcl = 7140, mGeosM-Ldcl = 5514, Ldcl-Dendra2<sub>T69A</sub> = 832, Ldcl-mGeosM = 12,211  
423 and Ldcl/anti-Ldcl-Nb = 14,075.

424 For Dendra2<sub>T69A</sub>-Ldcl, initial model generation was carried out in RELION-2.1<sup>56</sup> without any  
425 symmetry applied. For mGeosM-Ldcl, initial model generation was carried out in RELION-2.1 with  
426 either C1, C3 or D3 symmetry applied. The results of all three calculations being very similar, the  
427 model with applied D3 symmetry was selected. For Ldcl-Dendra2<sub>T69A</sub>, Ldcl-mGeosM and Ldcl/anti-  
428 Ldcl-Nb, the previously-determined Ldcl decamer structure (PDB ID: 3N75) was filtered to 60 Å and  
429 used as an initial model for 3D refinement.

430 3D refinement was carried out for each dataset with applied C2 symmetry for Dendra2<sub>T69A</sub>-  
431 Ldcl, D3 symmetry for mGeosM-Ldcl, and D5 symmetry for Ldcl-Dendra2<sub>T69A</sub>, Ldcl-mGeosM and  
432 Ldcl/anti-Ldcl-Nb. Rigid body fitting of Ldcl (PDB ID: 3N75), mEos2 (PDB ID: 3S05) and Dendra2  
433 (PDB ID: 2VZX) crystal structures was then carried out in Chimera<sup>57</sup> for the five datasets.

434

### 435 **Cryo-EM on Ldcl stacks (pH 5.7)**

436 Wild-type Ldcl was purified as previously described<sup>10</sup> from an *E. coli* strain impaired in the  
437 production of ppGpp (MG1655  $\Delta$ relA $\Delta$ spoT) in order to avoid any serendipitous ppGpp binding.  
438 Purified Ldcl was diluted to a final concentration of approximately 0.25 mg/mL in a buffer containing  
439 25 mM MES (pH 5.7), 0.3 M NaCl, 5% glycerol, 1 mM DTT and 0.1 mM PLP. 3  $\mu$ L of the sample  
440 was applied to a glow-discharged R2/1 300 mesh holey carbon copper grid (Quantifoil Micro Tools  
441 GmbH) and plunge-frozen in liquid ethane using a Vitrobot Mark IV (FEI) operated at 100% humidity.  
442 Datasets were recorded at the European Synchrotron Radiation Facility (ESRF) in Grenoble,  
443 France<sup>58</sup>, on a Titan Krios microscope (Thermo Scientific) equipped with a BioQuantum LS/967  
444 energy filter (Gatan) and a K2 summit direct electron detector (Gatan) operated in counting mode. A  
445 total of 2564 movies of 30 frames were collected with a total exposure of 6 s, total dose of 29.3 e<sup>-</sup>/Å<sup>2</sup>  
446 and a slit width of 20 eV for the energy filter. All movies were collected at a magnification of 130,000x,  
447 corresponding to a pixel size of 1.052 Å/pixel at the specimen level. A summary of cryo-EM data  
448 collection parameters can be found in Supplementary Table 1.

449

### 450 **Image Processing – Ldcl stacks (pH 5.7)**

451 Motion correction and dose-weighting of the recorded movies were performed using  
452 MotionCor2<sup>59</sup>. CTF parameters were determined on the aligned and dose-weighted sums using  
453 CTFFIND4<sup>60</sup>. After manual inspection of the dose-weighted sums, the best 558 (21.7%) micrographs  
454 were selected for further processing. Ldcl stacks were manually picked using e2heliboxer in  
455 EMAN2<sup>61</sup>. A total of 15,165 Ldcl-stack particles were extracted in RELION-3.0<sup>62</sup> with an extract size  
456 of 320 pixels, resulting in boxes containing three Ldcl decamers, and with the --helix option with an  
457 outer diameter of 160 pixels and a helical rise of 77 pixels. After particle extraction, per-particle CTF  
458 correction was performed using Gctf<sup>63</sup>. Extracted particles were subjected to 2D classification in  
459 RELION-3.0, resulting in a cleaned dataset containing 15,157 particles. Initial 3D refinement with  
460 imposed D5 symmetry was carried out in RELION-3.0, using an initial model generated by manually

stacking three Ldcl decamers (PDB ID: 3N75) in Chimera<sup>57</sup> and low-pass filtering the resulting Ldcl-stack to 40 Å. The resulting 4.3 Å resolution 3D reconstruction, along with the cleaned particle stack, was subsequently imported into CryoSPARC<sup>64</sup>. A final homogeneous 3D refinement in CryoSPARC, using a dynamic mask and imposing D5 symmetry, resulted in a map with a resolution of 3.28 Å based on the 0.143 gold-standard Fourier shell correlation (FSC) criterion<sup>65</sup>. The final map was sharpened using a B-factor of -96 Å<sup>2</sup>. A local resolution estimation of the final 3D reconstruction was calculated in RELION-3.0. A summary of cryo-EM data collection parameters and image processing steps for Ldcl stacks can be found in Supplementary Table 1 and Supplementary Figure 4, with local resolution and FSC curves shown in Supplementary Figure 5.

### **Cryo-EM on Ldcl decamers (pH 7.0)**

Purified Ldcl was dialysed into a buffer at pH 7.0 and diluted to a concentration of 0.25 mg/mL. 3 µL of diluted sample was applied to a glow-discharged (R2/1 300 mesh holey carbon copper grid (Quantifoil Micro Tools GmbH) and plunge-frozen in liquid ethane using a Vitrobot Mark IV (FEI) operated at 100% humidity. Images were recorded on a Glacios microscope (Thermo Scientific) equipped with a Falcon II direct electron detector (Thermo Scientific). A total of 2772 movies of 29 frames were collected with a total exposure of 1.5 s and a total dose of 45 e<sup>-</sup>/Å<sup>2</sup>. All movies were collected at a magnification of 116,086x, corresponding to a pixel size of 1.206 Å/pixel at the specimen level.

### **Image Processing – Ldcl decamers (pH 7.0)**

Motion correction was carried out using patch motion correction in CryoSPARC, discarding the first two frames. Initial CTF estimation was then carried out on summed frames using CTFFIND4. A subset of 600 micrographs were subjected to automatic picking using the blob picker in CryoSPARC, resulting in ~238,000 picked particles. Particles were then extracted with a box size of 256 x 256 and subjected to 2D classification. Particles from the best classes showing clear secondary structural features for Ldcl were selected for homogeneous refinement (with applied D5 symmetry) against EMD-3204 low-pass filtered to 30 Å, resulting in a reconstruction with a resolution of 4.2 Å (FSC = 0.143). This reconstruction was then used to create templates for picking the entire dataset using the template picker in CryoSPARC, after filtering to 12 Å. ~796,000 particles were extracted and subjected to 2D classification, and the best ~428,000 particles were subjected to heterogeneous refinement with applied D5 symmetry against the 4.2 Å map, resulting in one higher-resolution class corresponding to ~229,000 particles. These particles were subjected to homogeneous refinement with applied D5 symmetry, resulting in a map with a resolution of 2.76 Å (FSC = 0.143) which was then sharpened with a B-factor of -173 Å<sup>2</sup>. A summary of cryo-EM data collection parameters and image processing steps for Ldcl decamers can be found in Supplementary Table 1 and Supplementary Figure 4, with local resolution and FSC curves shown in Supplementary Figure 5.

### **Fitting of structures and refinement**

For fitting of atomic models in the 3D reconstructions of Ldcl at pH 5.7 (stack) or pH 7.0 (decamer), two (for the stack) or one copy (for the decamer) of the Ldcl X-ray crystal structure (PDB

503 ID: 3N75) were first rigid-body fitted in the corresponding 3D reconstructions using Chimera.  
504 Refinement was performed using the Phenix software package<sup>66</sup> and was identical for both 3D  
505 reconstructions. A first round of real space refinement was carried out with enabled rigid-body, global  
506 minimization, local grid search and ADP refinement parameters, and imposing rotamer,  
507 Ramachandran, NCS and reference model (PDB ID: 3N75) restraints. A final round of real space  
508 refinement was then performed using the same settings, but without rigid body refinement and  
509 without applying reference restrains, setting the 'nonbonded\_weight' parameter to 4000 and  
510 disabling 'local\_grid\_search'. A summary of refinement and model validation statistics can be found  
511 in Supplementary Table 1.

512

### 513 **pH shift experiment**

514 Stationary phase cultures, which were grown overnight from single colonies in LB medium,  
515 were diluted to OD<sub>600</sub> ~ 0.01 and re-grown at 37°C within approximately 1h 45 min in fresh LB  
516 medium to an OD<sub>600</sub> of 0.1. From this culture, 14 mL were transferred to 15 mL falcon tubes and  
517 pelleted by centrifugation at RT for 5 min. The supernatant was decanted, whereby systematically  
518 around 200 µL LB remained in the falcon tube. The pellet of bacteria was resuspended and  
519 afterwards, LB-4.6 containing 30 mM L-lysine was added to the cells up to 14 mL. To prepare LB-  
520 4.6 medium, LB powder (Sigma-Aldrich) was completely dissolved in distilled water by stirring for 30  
521 min. The pH of 4.6 was then adjusted using HCl. After autoclaving, sterile filtered L-lysine was added  
522 to LB-4.6. For this, L-lysine was dissolved in an aliquot of LB-4.6, sterile filtered and mixed with the  
523 remaining LB-4.6. 30 mM L-lysine were used. In order to grow the culture under oxygen-limiting  
524 conditions, the lid of the falcon tube was closed and the tubes were placed at 37°C on a shaker (150  
525 rpm). After defined time-points, aliquots were taken for OD<sub>600</sub> measurement, pH measurement,  
526 SDS-PAGE or immunofluorescence. For each time-point, a new tube was opened and not reused  
527 further.

528

### 529 **Western Blotting**

530 SDS-PAGE was performed with a Biorad electrophoresis chamber using standard 12% reducing  
531 SDS-PAGE gels. Proteins were transferred to a nitrocellulose membrane (Biorad) using a Trans-  
532 Blot Turbo Transfer System (Biorad). The membrane was blocked for 1h using 5% BSA in TBS  
533 supplemented with 0.1% Tween (TBS-T). Afterwards, the membrane was incubated for 1h with an  
534 anti-Ldcl antibody (Qalam-Antibodies) in BSA/TBS-T (1:5000). The membrane was subsequently  
535 washed 3 x 10 min in TBS-T and incubated for 1 h with HRP-coupled anti-rabbit antibody (1:10000  
536 in BSA/TBS-T). Finally the membrane was washed 3 x 10 min in TBS-T. The membrane was rinsed  
537 once with TBS, prior to detection of antibody-labelled proteins using ECL reagent (GE Healthcare).  
538 Incubation and detection of antibody was performed at 25 °C.

539

### 540 **Nanobody labelling**

541 For the labelling reaction, 50 µL nanobody (i.e. 200 µg) was pipetted into a 1.5 mL tube,  
542 placed on ice and supplemented with 5.5 µL of 1 M bicarbonate buffer at pH 8.3. 100 µg of Alexa-  
543 647 or Alexa-488 NHS ester dye (Life Technologies, A37573 and A20000) were dissolved in 10 µL  
544 DMSO to final concentration of 10 mg/mL. 5 µl (i.e. 40 nmol) dye in DMSO was added to the protein

545 and incubated for 1 h at room temperature in a shaking block, covered with an aluminium foil to  
546 protect the dye from the light. Excess dye was removed by iterative buffer exchange using a 3K spin  
547 column (Amicon-Ultra-4 Centrifugal Filters Ultracell 3K, Millipore UFC800396) to PBS. The degree  
548 of labelling was inferred from measuring the OD.

### 549 **Cell preparation for immunofluorescence staining of *E. coli* cells with nanobodies**

550 For immunofluorescence staining, the OD<sub>600</sub> of the cell culture was measured, and the  
551 volume of cells corresponding to OD = 4, with OD = 1 corresponding to about  $8 \times 10^8$  cells per mL,  
552 were collected by centrifugation. After removal of LB by pipetting, cells were resuspended in 2 mL  
553 of 4 % FA in PBS (made from 16% Formaldehyde solution, Methanol-free from Thermo Scientific).  
554 Falcons were placed on a rotor for constant agitation for 45 min at room temperature. After fixation,  
555 cells were collected by centrifugation and the solution was removed by pipetting. Cells were then  
556 resuspended in 14 mL PBS (Gibco, Thermo Scientific) to remove and dilute the fixative. Cells were  
557 permeabilised for 10 min using 2 mL 0.1 % Triton X-100 in PBS and subsequently washed three  
558 times with 10 mL PBS. Finally, cells were transferred to 1.5 mL tubes, centrifuged and resuspended  
559 in 200  $\mu$ L 1% BSA/PBS (BSA/PBS solution was dissolved for 30 min and sterile-filtered to avoid  
560 clumps). After 30 min of incubation, 0.5  $\mu$ g anti-Ldcl-Nb, labelled with the dedicated dye, was added  
561 to the 200  $\mu$ L bacteria-BSA/PBS suspension. Cells were incubated with the labelled anti-Ldcl-Nb for  
562 16 h at 4°C. The next day, cells were washed three times with 1 mL PBS, centrifuged to remove  
563 antibody solution and resuspended in 250  $\mu$ L PBS. When needed, Hoechst 33342 (Sigma-Aldrich)  
564 was added to a final concentration of about 100 ng/mL.

### 565 **Wide-field imaging and flow cytometry**

566 For epifluorescence imaging, 2  $\mu$ L of cells were placed between a glass slide and a coverslip, which  
567 have been carefully pressed together, and observed using an inverted IX81 microscope, with a  
568 UPLFLN 100X oil immersion objective (N.A. 1.3) (Olympus), using the appropriate specific excitation  
569 and emission filters for AF488 (GFP-3035B set, Semrock) and DAPI (DAPI-5060B set, Semrock).  
570 Acquisitions were performed with Volocity software (QuorumTechnologies™) with a sCMOS  
571 2048x2048 camera (Hamamatsu ORCA Flash 4, 16 bits/pixel) achieving a final magnification of 64  
572 nm per pixel. For flow cytometry, 50  $\mu$ L of cells suspensions were injected in a MACSQuant VYB  
573 flow cytometer (Miltenyi Biotech, Bergish Gladbach, Germany) using the 488 nm excitation and  
574 525(50) nm emission channel (B1). AF488 positive populations were estimated after forward scatter  
575 (FSC) and side scatter (SSC) gating on the cells. Data were further processed with MACSQuantify  
576 software (Miltenyi Biotech).

### 577 **STORM imaging**

578 For Single Molecule Localization Microscopy, cells were transferred to a glucose buffer containing  
579 50 mM NaCl, 150 mM Tris (pH 8.0), 10% Glucose, 100 mM MEA (Mercaptoethylamine) and 1x Glox.  
580 Glox was prepared as a 10x stock and contained 1  $\mu$ M catalase and 2.3  $\mu$ M glucoseoxidase. Menzel  
581 glass slides (Thermo Scientific) and precision coverslips (1.5H from ThorLabs) were cleaned for 30  
582 min using an UV-ozone cleaning device (HELIOS□500, UVOTECH Systems). 2  $\mu$ L of immuno-  
583 labelled cells were placed onto a glass slide and covered with the coverslip, then cells were carefully  
584

587 spread by pressing glass slides firmly together and the sides were sealed with transparent nail polish  
588 to avoid evaporation. Mounted samples were imaged on a homemade SMLM set up based on an  
589 IX81 microscope (Olympus). STORM was performed by focusing a 643 nm excitation laser beam  
590 (Toptica Diode laser) to the back focal plane of an oil immersion UAPON100X (N.A. 1.49) objective.  
591 The intensity of the laser was tuned using an Acousto-Optical Tunable Filter (OATF, Quanta Tech).  
592 Acquisition was obtained with a 16 bits/pixel Evolve 512 EMCCD (Photometrics) using Metamorph  
593 (Molecular Devices), for a final pixel size of 123 nm. 3D-STORM based on point-spread-function  
594 astigmatism<sup>67</sup> was performed using a cylindrical lens (LJ1516L1-A, Thorlabs) placed in the detection  
595 light path. STORM datasets consisting of about 30,000 frames, using a 643 nm laser power density  
596 of 3 kW/cm<sup>2</sup> and a 405 nm laser power density of up to 1 W/cm<sup>2</sup> and a frametime of 50 ms, were  
597 acquired with an EMCCD gain set to 200. 3D point spread function calibration was achieved using  
598 tetraspec beads. Finally, data were processed with the Thunderstorm plugin<sup>68</sup> in ImageJ<sup>69</sup>. 3D-  
599 images were rendered with Visp<sup>70</sup> using a minimum neighbour density threshold of 20 to 28.

600

### 601 **Biolayer interferometry measurements**

602 For BLI binding studies, RavA with a biotinylated C-terminal AviTag was expressed and purified as  
603 previously described<sup>29</sup>. BLI experiments were performed in 1x HBS pH at 7.0 (25 mM HEPES, 300  
604 mM NaCl, 10 mM MgCl<sub>2</sub>, 10% glycerol) supplemented with 1x kinetics buffer (0.1% w/v BSA, 0.02%  
605 v/v Tween-20), 1mM ADP, 1mM DTT and 0.1 mM PLP. Experiments were performed using the BLItz  
606 System instrument (FortéBio), operated at room temperature. Before the start of each BLI  
607 experiment, RavA-AviTag was incubated with 1 mM ADP for 10 min. Streptavidin-coated biosensors  
608 (FortéBio) were functionalised with biotinylated RavA-AviTag, then quenched with 10 µg/mL biocytin.  
609 Experiments with the wild-type Ldcl are from<sup>29</sup>. For C-terminal Ldcl-FP fusions, pins were dipped in  
610 wells containing a range of Ldcl-FP concentrations from 0 to 1000 nM, with no binding signal  
611 recorded at any concentration of Ldcl-FP.

612

### 613 **Data and code availability**

614 Cryo-EM maps, along with the corresponding fitted atomic structures, have been submitted to the  
615 EMDB and PDB with accession codes EMD-10850 and PDB-6YN6 for Ldcl stacks (pH 5.7), and  
616 EMD-10849 and PDB-6YN5 for Ldcl decamers (pH 7.0).

617

### 618 **Acknowledgements**

619 We thank Guy Schoehn for establishing and managing the cryo-electron microscopy platform  
620 and for providing training and support, and Rose-Laure Revel-Goyet and Françoise Lacroix for the  
621 support and access to the M4D Cell imaging Platform. We are grateful to Aymeric Peuch for help  
622 with the usage of the EM computing cluster and Daniel Thédie, Kévin Floc'h and Joel Beaudoin for  
623 discussions. We thank Alain Roussel and Aline Desmyter for the nanobody production, and Julien  
624 Perard for help with the cloning of the Dendra2<sub>T69A</sub>-Ldcl construct. We acknowledge the European  
625 Synchrotron Radiation Facility for provision of beam time on CM01. This work was funded by the  
626 European Union's Horizon 2020 research and innovation programme under grant agreement No  
627 647784 to IG. The nanobody generation platform of the AFMB laboratory (Marseille, France) was  
628 supported by the French Infrastructure for Integrated Structural Biology (FRISBI) ANR-10-INSB-05-

629 01. For electron and fluorescence microscopy studies, this work used the platforms of the Grenoble  
630 Instruct centre (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-  
631 05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology  
632 (PSB). The electron microscope facility (Glacios electron microscope) is supported by the Rhône-  
633 Alpes Region (CIBLE and FEDER), the FRM, the CNRS, the University of Grenoble and the GIS-  
634 IBISA. JF was supported by a long-term EMBO fellowship (ALTF441-2017) and a Marie Skłodowska-  
635 Curie actions Individual Fellowship (789385, RespVIRALI).

636

### 637 **Author Contributions**

638 C.L., M.B., V.A., A.F. and K.H. cloned constructs, M.J., A.F. and K.H. purified proteins. M.J. and I.G.  
639 performed ns-EM imaging and analysis. A.D., G.E. and M.B.-V. performed cryo-EM imaging. J.F.  
640 and A.D. performed cryo-EM analysis. M.J. and J.F. built models resulting from EM maps. M.J., J.F.  
641 and I.G. analysed structures and interpreted data. C.L. and M.B. performed optical imaging of  
642 overexpression constructs with input of V.A., J.-P.K. and D.B. C.L. performed nanobody  
643 characterisation for optical imaging with input from J.-P.K. C.L. analysed endogenous expression  
644 and performed optical imaging of endogenous LdcI with input of V.A., J.-P.K. and D.B. C.L. and D.B.  
645 analysed STORM Images. M.J., C.L., J.F. and I.G. analysed the data and prepared the figures and  
646 tables. C.L. and D.B. contributed to the design of the optical imaging part of the project together with  
647 I.G. I.G. designed, supervised and funded the overall study. M.J., J.F. and I.G. wrote the manuscript  
648 with significant input from C.L. and contributions from all of the authors.

649

### 650 **References**

- 651 1. Prouteau, M. & Loewith, R. Regulation of Cellular Metabolism through Phase Separation of Enzymes.  
652 *Biomolecules* **8**, 160 (2018).
- 653 2. Fang, F. C., Frawley, E. R., Tapscott, T. & Vázquez-Torres, A. Bacterial Stress Responses during Host  
654 Infection. *Cell Host and Microbe* **20**, 133–143 (2016).
- 655 3. Watson, N., Dunnyak, D. S., Rosey, E. L., Slonczewski, J. L. & Olson, E. R. Identification of elements  
656 involved in transcriptional regulation of the Escherichia coli cad operon by external pH. *J. Bacteriol.*  
657 **174**, 530–540 (1992).
- 658 4. Foster, J. W. Escherichia coli acid resistance: Tales of an amateur acidophile. *Nature Reviews*  
659 *Microbiology* **2**, 898–907 (2004).
- 660 5. Moreau, P. L. The lysine decarboxylase CadA protects Escherichia coli starved of phosphate against  
661 fermentation acids. *J. Bacteriol.* **189**, 2249–2261 (2007).
- 662 6. Kanjee, U. *et al.* Linkage between the bacterial acid stress and stringent responses: The structure of  
663 the inducible lysine decarboxylase. *EMBO J.* **30**, 931–944 (2011).
- 664 7. Gale, E. F. The Bacterial Amino Acid Decarboxylases. *Adv. Enzymol. - Relat. Areas Mol. Biol.* **6**, 1–32  
665 (1946).
- 666 8. Sabo, D. L., Boeker, E. A., Byers, B., Waron, H. & Fischer, E. H. Purification and Physical Properties  
667 of Inducible Escherichia coli Lysine Decarboxylase. *Biochemistry* **13**, 662–670 (1974).
- 668 9. Gale, E. F. & Epps, H. M. R. The effect of the pH of the medium during growth on the enzymic activities  
669 of bacteria (Escherichia coli and Micrococcus lysodeikticus) and the biological significance of the  
670 changes produced. *Biochem. J.* **36**, 600–618 (1942).

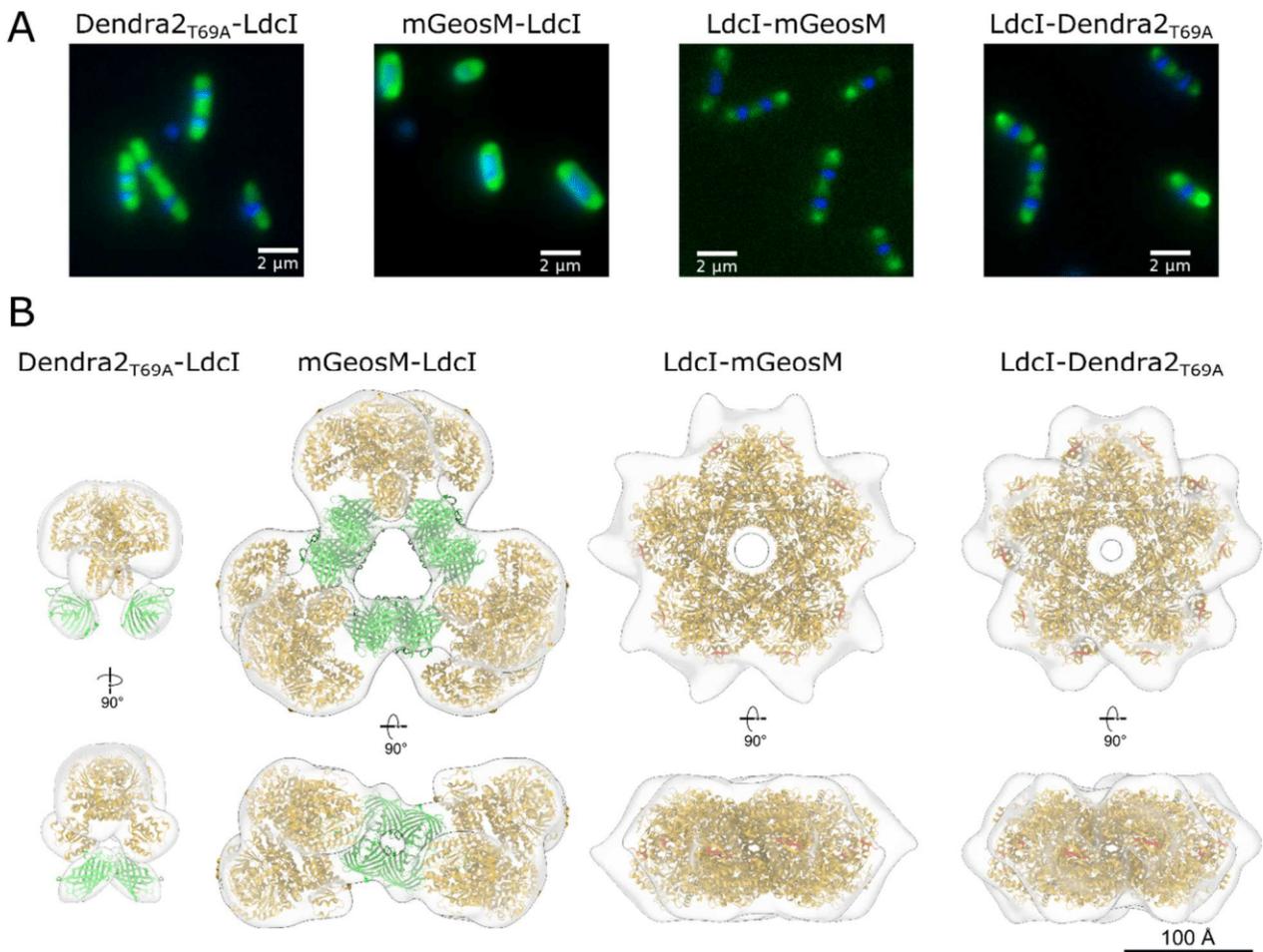
- 671 10. Kandiah, E. *et al.* Structural insights into the Escherichia coli lysine decarboxylases and molecular  
672 determinants of interaction with the AAA+ ATPase RavA. *Sci. Rep.* **6**, (2016).
- 673 11. Surovtsev, I. V. & Jacobs-Wagner, C. Subcellular Organization: A Critical Feature of Bacterial Cell  
674 Replication. *Cell* **172**, 1271–1293 (2018).
- 675 12. Margolin, W. The price of tags in protein localization studies. *J. Bacteriol.* **194**, 6369–6371 (2012).
- 676 13. Gahlmann, A. & Moerner, W. E. Exploring bacterial cell biology with single-molecule tracking and super-  
677 resolution imaging. *Nature Reviews Microbiology* **12**, 9–22 (2014).
- 678 14. Landgraf, D., Okumus, B., Chien, P., Baker, T. A. & Paulsson, J. Segregation of molecules at cell  
679 division reveals native protein localization. *Nat. Methods* **9**, 480–482 (2012).
- 680 15. Costantini, L. M., Fossati, M., Francolini, M. & Snapp, E. L. Assessing the Tendency of Fluorescent  
681 Proteins to Oligomerize Under Physiologic Conditions. *Traffic* **13**, 643–649 (2012).
- 682 16. Ingerson-Mahar, M., Briegel, A., Werner, J. N., Jensen, G. J. & Gitai, Z. The metabolic enzyme CTP  
683 synthase forms cytoskeletal filaments. *Nat. Cell Biol.* **12**, 739–746 (2010).
- 684 17. Lynch, E. M. *et al.* Human CTP synthase filament structure reveals the active enzyme conformation.  
685 *Nat. Struct. Mol. Biol.* **24**, 507–514 (2017).
- 686 18. Frey, T. G., Eisenberg, D. & Eiserling, F. A. Glutamine synthetase forms three and seven stranded  
687 helical cables. *Proc. Natl. Acad. Sci. U. S. A.* **72**, 3402–3406 (1975).
- 688 19. Kim, G. *et al.* Aldehyde-alcohol dehydrogenase forms a high-order spiroosome architecture critical for  
689 its activity. *Nat. Commun.* **10**, (2019).
- 690 20. Schuchmann, K., Vonck, J. & Müller, V. A bacterial hydrogen-dependent CO<sub>2</sub> reductase forms  
691 filamentous structures. *FEBS J.* **283**, 1311–1322 (2016).
- 692 21. Yao, Z. & Carballido-López, R. Fluorescence Imaging for Bacterial Cell Biology: From Localization to  
693 Dynamics, From Ensembles to Single Molecules. *Annu. Rev. Microbiol.* **68**, 459–476 (2014).
- 694 22. Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction  
695 microscopy (STORM). *Nat. Methods* **3**, 793–795 (2006).
- 696 23. Betzig, E. *et al.* Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642–  
697 5 (2006).
- 698 24. Hess, S. T., Girirajan, T. P. K. & Mason, M. D. Ultra-high resolution imaging by fluorescence  
699 photoactivation localization microscopy. *Biophys. J.* **91**, 4258–4272 (2006).
- 700 25. Chang, H. *et al.* A unique series of reversibly switchable fluorescent proteins with beneficial properties  
701 for various applications. *Proc. Natl. Acad. Sci.* **109**, 4455–4460 (2012).
- 702 26. Berardozzi, R., Adam, V., Martins, A. & Bourgeois, D. Arginine 66 Controls Dark-State Formation in  
703 Green-to-Red Photoconvertible Fluorescent Proteins. *J. Am. Chem. Soc.* **138**, 558–565 (2016).
- 704 27. Malet, H. *et al.* Assembly principles of a unique cage formed by hexameric and decameric E. coli  
705 proteins. *Elife* **3**, e03653 (2014).
- 706 28. Snider, J. *et al.* Formation of a distinctive complex between the inducible bacterial lysine decarboxylase  
707 and a novel AAA+ ATPase. *J. Biol. Chem.* **281**, 1532–1546 (2006).
- 708 29. Jessop, M. *et al.* Structural insights into ATP hydrolysis by the MoxR ATPase RavA and the Ldcl-RavA  
709 cage-like complex. *Commun. Biol.* **3**, 46 (2020).
- 710 30. El Bakkouri, M. *et al.* Structure of RavA MoxR AAA+ protein reveals the design principles of a molecular  
711 cage modulating the inducible lysine decarboxylase activity. *Proc. Natl. Acad. Sci.* **107**, 22499–22504  
712 (2010).

- 713 31. Wong, K. S. *et al.* The MoxR ATPase RavA and its cofactor ViaA interact with the NADH: Ubiquinone  
714 oxidoreductase I in Escherichia coli. *PLoS One* **9**, e85529 (2014).
- 715 32. Wong, K. S., Bhandari, V., Janga, S. C. & Houry, W. A. The RavA-ViaA Chaperone-Like System  
716 Interacts with and Modulates the Activity of the Fumarate Reductase Respiratory Complex. *J. Mol. Biol.*  
717 **429**, 324–344 (2017).
- 718 33. Zhang, M. *et al.* Rational design of true monomeric and bright photoactivatable fluorescent proteins.  
719 *Nat. Methods* **9**, (2012).
- 720 34. Coltharp, C. & Xiao, J. Superresolution microscopy for microbiology. *Cell. Microbiol.* **14**, 1808–1818  
721 (2012).
- 722 35. Cramer, K. *et al.* Visualization of Bacterial Protein Complexes Labeled with Fluorescent Proteins and  
723 Nanobody Binders for STED Microscopy. *Int. J. Mol. Sci.* **20**, 3376 (2019).
- 724 36. Kanjee, U., Gutsche, I., Ramachandran, S. & Houry, W. A. The enzymatic activities of the escherichia  
725 coli basic aliphatic amino acid decarboxylases exhibit a pH zone of inhibition. *Biochemistry* **50**, 9388–  
726 9398 (2011).
- 727 37. Barák, I. & Muchová, K. The role of lipid domains in bacterial cell processes. *Int. J. Mol. Sci.* **14**, 4050–  
728 4065 (2013).
- 729 38. Lenn, T., Leake, M. C. & Mullineaux, C. W. Clustering and dynamics of cytochrome bd-I complexes in  
730 the Escherichia coli plasma membrane in vivo. *Mol. Microbiol.* **70**, 1397–1407 (2008).
- 731 39. Erhardt, H. *et al.* Organization of the Escherichia coli aerobic enzyme complexes of oxidative  
732 phosphorylation in dynamic domains within the cytoplasmic membrane. *Microbiologyopen* **3**, 316–326  
733 (2014).
- 734 40. Llorente-Garcia, I. *et al.* Single-molecule in vivo imaging of bacterial respiratory complexes indicates  
735 delocalized oxidative phosphorylation. *Biochim. Biophys. Acta - Bioenerg.* **1837**, 811–824 (2014).
- 736 41. Werner, J. N. *et al.* Quantitative genome-scale analysis of protein localization in an asymmetric  
737 bacterium. *Proc. Natl. Acad. Sci.* **106**, 7858–7863 (2009).
- 738 42. Magalon, A. & Alberge, F. Distribution and dynamics of OXPHOS complexes in the bacterial  
739 cytoplasmic membrane. *Biochimica et Biophysica Acta - Bioenergetics* **1857**, 198–213 (2016).
- 740 43. Bramkamp, M. & Lopez, D. Exploring the Existence of Lipid Rafts in Bacteria. *Microbiol. Mol. Biol. Rev.*  
741 **79**, 81–100 (2015).
- 742 44. Erhardt, H. *et al.* Disruption of individual nuo-genes leads to the formation of partially assembled  
743 NADH:ubiquinone oxidoreductase (complex I) in Escherichia coli. *Biochim. Biophys. Acta* **1817**, 863–  
744 71 (2012).
- 745 45. Yamamoto, Y., Miwa, Y., Miyoshi, K., Furuyama, J. I. & Ohmori, H. The Escherichia coli IdcC gene  
746 encodes another lysine decarboxylase, probably a constitutive enzyme. *Genes Genet. Syst.* **72**, 167–  
747 172 (1997).
- 748 46. Swulius, M. T. & Jensen, G. J. The helical mreB cytoskeleton in Escherichia coli MC1000/pLE7 is an  
749 artifact of the N-terminal yellow fluorescent protein tag. *J. Bacteriol.* **194**, 6382–6386 (2012).
- 750 47. Barák, I., Muchová, K., Wilkinson, A. J., O'Toole, P. J. & Pavlendová, N. Lipid spirals in Bacillus subtilis  
751 and their role in cell division. *Mol. Microbiol.* **68**, 1315–1327 (2008).
- 752 48. Mileykovskaya, E. & Dowhan, W. Cardiolipin membrane domains in prokaryotes and eukaryotes.  
753 *Biochimica et Biophysica Acta - Biomembranes* **1788**, 2084–2091 (2009).
- 754 49. Fishov, I. & Woldringh, C. L. Visualization of membrane domains in Escherichia coli. *Mol. Microbiol.* **32**,

1166–1172 (1999).

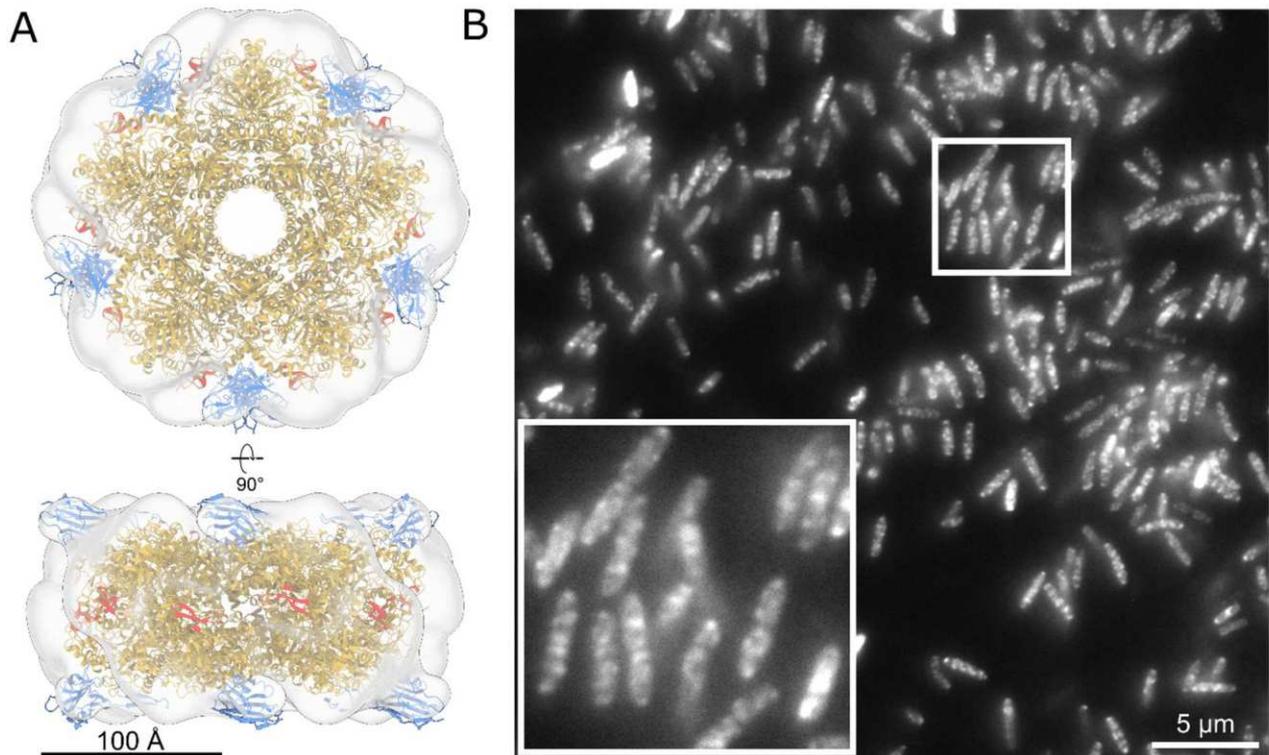
- 755
- 756 50. Daniel, R. A. & Errington, J. Control of cell morphogenesis in bacteria: Two distinct ways to make a  
757 rod-shaped cell. *Cell* **113**, 767–776 (2003).
- 758 51. Wang, S., Furchtgott, L., Huang, K. C. & Shaevitz, J. W. Helical insertion of peptidoglycan produces  
759 chiral ordering of the bacterial cell wall. *Proc. Natl. Acad. Sci. U. S. A.* **109**, (2012).
- 760 52. Pardon, E. *et al.* A general protocol for the generation of Nanobodies for structural biology. *Nat. Protoc.*  
761 **9**, 674–693 (2014).
- 762 53. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron  
763 microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
- 764 54. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: Semiautomated Software for High-Resolution Single-  
765 Particle Reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).
- 766 55. Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure  
767 determination. *J. Struct. Biol.* **180**, 519–530 (2012).
- 768 56. Scheres, S. H. W. Processing of Structurally Heterogeneous Cryo-EM Data in RELION. in *Methods in*  
769 *Enzymology* **579**, 125–157 (2016).
- 770 57. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J.*  
771 *Comput. Chem.* **25**, 1605–12 (2004).
- 772 58. Kandiah, E. *et al.* CM01: A facility for cryo-electron microscopy at the European synchrotron. *Acta*  
773 *Crystallogr. Sect. D Struct. Biol.* **75**, (2019).
- 774 59. Zheng, S. Q. *et al.* MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-  
775 electron microscopy. *Nature Methods* **14**, 331–332 (2017).
- 776 60. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs.  
777 *J. Struct. Biol.* **192**, 216–221 (2015).
- 778 61. Tang, G. *et al.* EMAN2: An extensible image processing suite for electron microscopy. *J. Struct. Biol.*  
779 **157**, 38–46 (2007).
- 780 62. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-  
781 3. *Elife* **7**, 1–22 (2018).
- 782 63. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
- 783 64. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. CryoSPARC: Algorithms for rapid  
784 unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
- 785 65. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and  
786 contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
- 787 66. Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure  
788 solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221 (2010).
- 789 67. Huang, B., Wang, W., Bates, M. & Zhuang, X. Three-dimensional super-resolution imaging by  
790 stochastic optical reconstruction microscopy. *Science (80-. )*. **319**, (2008).
- 791 68. Ovesný, M., Křížek, P., Borkovec, J., Švindrych, Z. & Hagen, G. M. ThunderSTORM: A comprehensive  
792 ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. *Bioinformatics* **30**,  
793 (2014).
- 794 69. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis.  
795 *Nature Methods* **9**, (2012).
- 796 70. El Beheiry, M. & Dahan, M. ViSP: Representing single-particle localizations in three dimensions. *Nature*  
797 *Methods* **10**, (2013).
- 798
- 799

**Figure 1.**



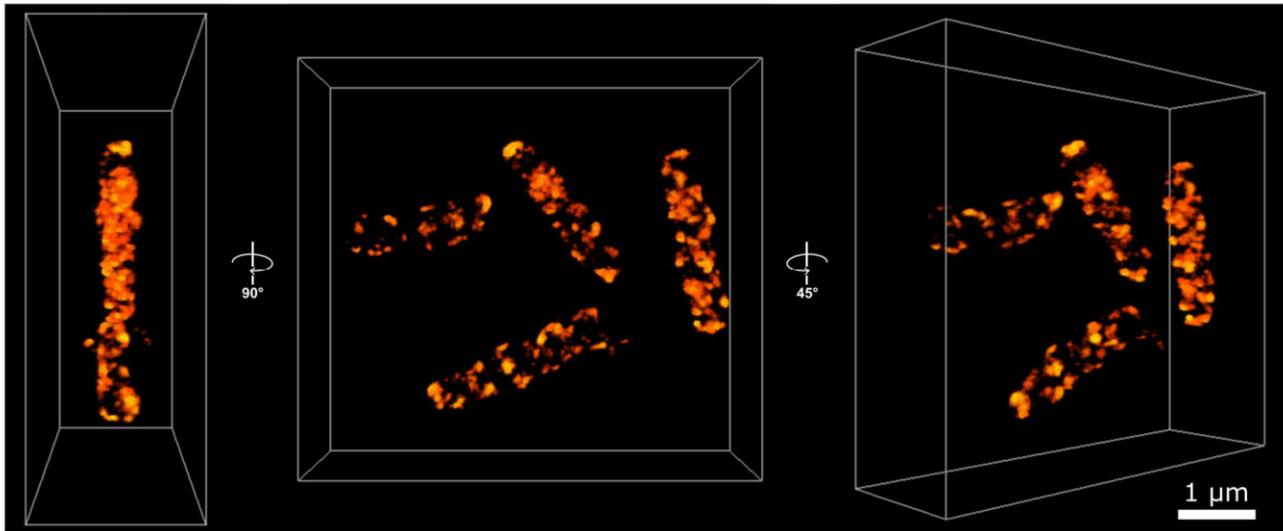
**Figure 1. Fluorescent protein fusions affect LdcI structure without significantly altering cellular localisation of overexpressed constructs. A)** Wide-field fluorescence microscopy images of *E. coli* cells overexpressing the fluorescent fusion proteins in A). Green fluorescence = FP-LdcI, blue fluorescence = DAPI-stained DNA. **B)** ns-EM 3D maps of fluorescently-tagged LdcI, with fitted models. Left – Dendra2<sub>T69A</sub>-LdcI forms dimers, with fluorescent barrels located next to the N-terminus of LdcI as expected. Second from left – mGeosM-LdcI forms large non-native oligomers, composed of three LdcI tetramers bridged by tetramers of mGeosM. Second from right and right – both C-terminal fluorescent fusions LdcI-mGeosM and LdcI-Dendra2<sub>T69A</sub> form decamers, with protrusions at the C-terminus (coloured in red) attributed to flexibly-linked fluorescent proteins. Fitted PDB models are as follows – LdcI: 3N75; Dendra2: 2VZX; mGeosM: 3S05 (mEos2 crystal structure).

Figure 2.



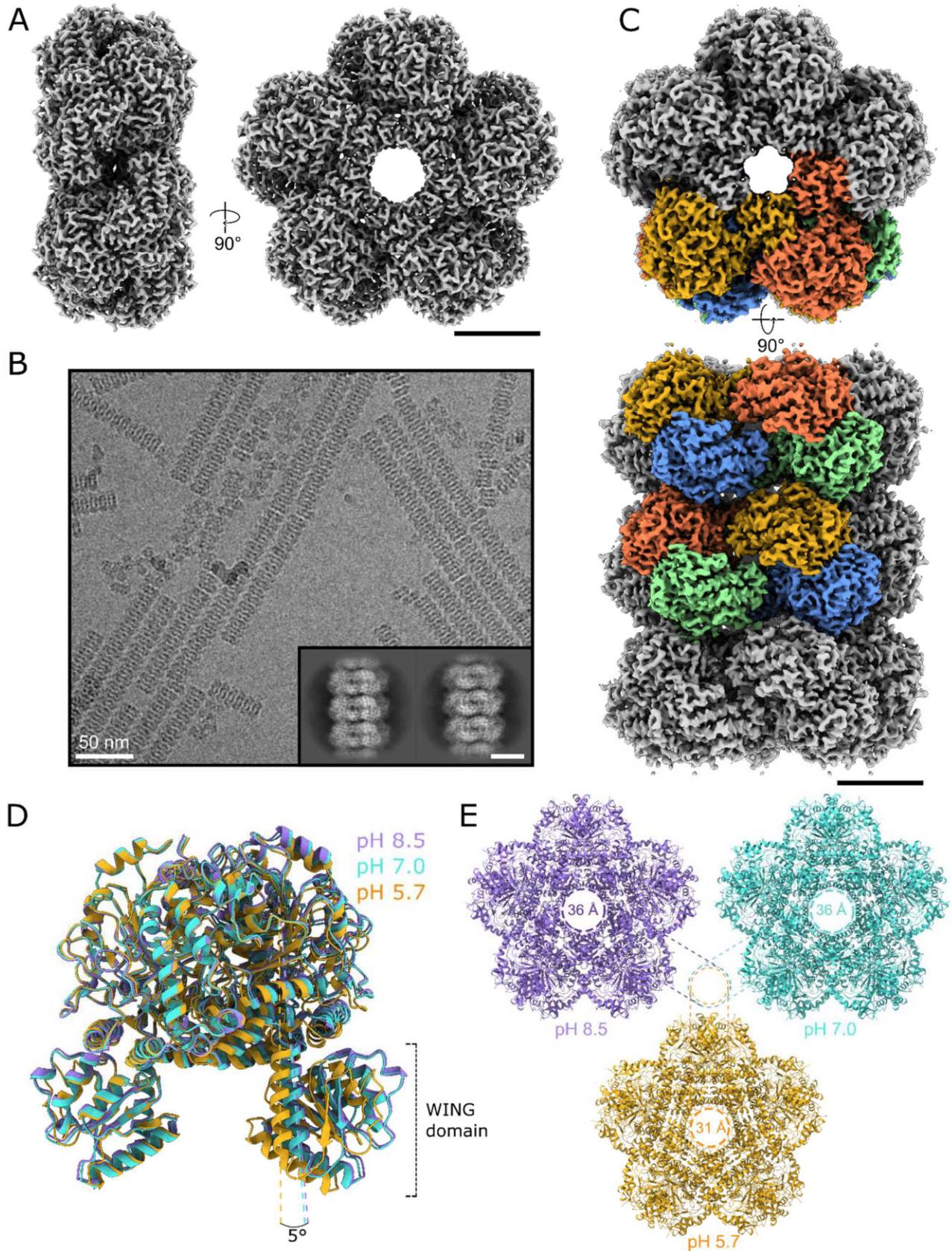
**Figure 2. Anti-Ldcl-Nb is a useful tool to probe cellular localisation of endogenous Ldcl under acid stress conditions.** **A)** Ns-EM 3D reconstruction of the Ldcl decorated by anti-Ldcl-Nb. An Ldcl decamer (gold, PDB ID: 3N75) and 10 nanobodies (blue, PDB ID: 1MEL, anti-lysozyme nanobody) are fitted in the density, with the nanobody binding at the top and bottom of the decameric ring. The C-terminal RavA binding site is indicated in red, and is in a spatially distinct location from the bound nanobodies. Scale bar = 100Å. **B)** Wide-field fluorescence microscopy image of wild-type *E. coli* MG1655 cells grown at pH 4.6 for 90 minutes stained with anti-Ldcl-Nb labelled with AF488. Inset – zoom of the image showing punctuate fluorescence patterns. Scale bar = 5 μm.

Figure 3.



**Figure 3. 3D STORM imaging of *E. coli* cells stained with anti-Ldcl-Nb reveal a patchy pseudo-helically arranged distribution of endogenous Ldcl upon acid stress.** 3D-STORM imaging of wild-type Ldcl with AF647-conjugated anti-Ldcl-Nb. Points are coloured according to localisation density, with brighter points corresponding to higher localisation density. The centre panel shows four cells in the field of view, looking down the z-axis. Left and right panels show side and tilted views respectively. Scale bar = 1  $\mu$ m.

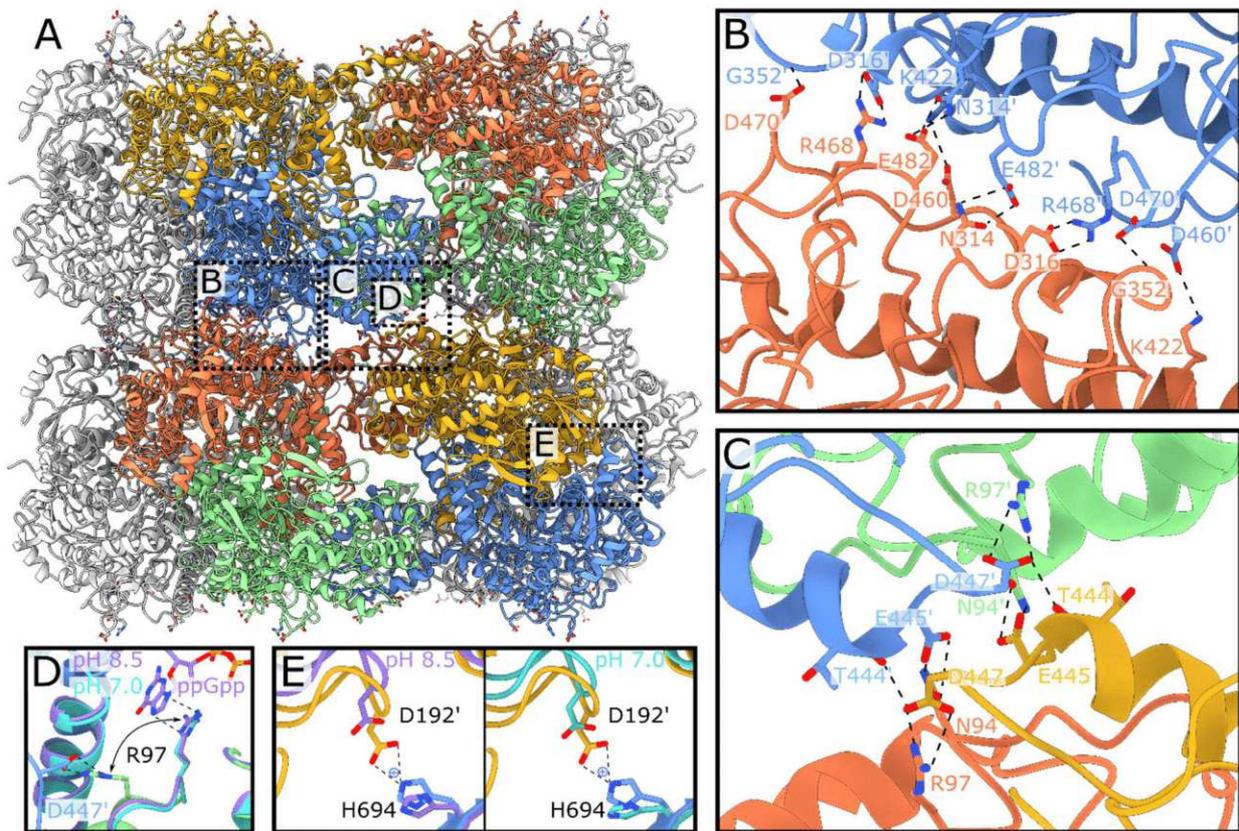
Figure 4.



**Figure 4. Cryo-EM analysis of the Ldcl decamer at neutral pH and Ldcl stacks formed in acid stress conditions. A)** Cryo-EM reconstruction of the Ldcl decamer at pH 7.0 from side (left) and top (right) views. Scale bar = 50 Å. **B)** Cryo-EM micrograph of Ldcl stacks, scale bar = 100 nm. Inset – 2D class averages displaying clear secondary structural features, scale bar = 100 Å. **C)** Top (above) and side (below) views of the cryo-EM reconstruction of a three-decamer Ldcl stack. Four dimers

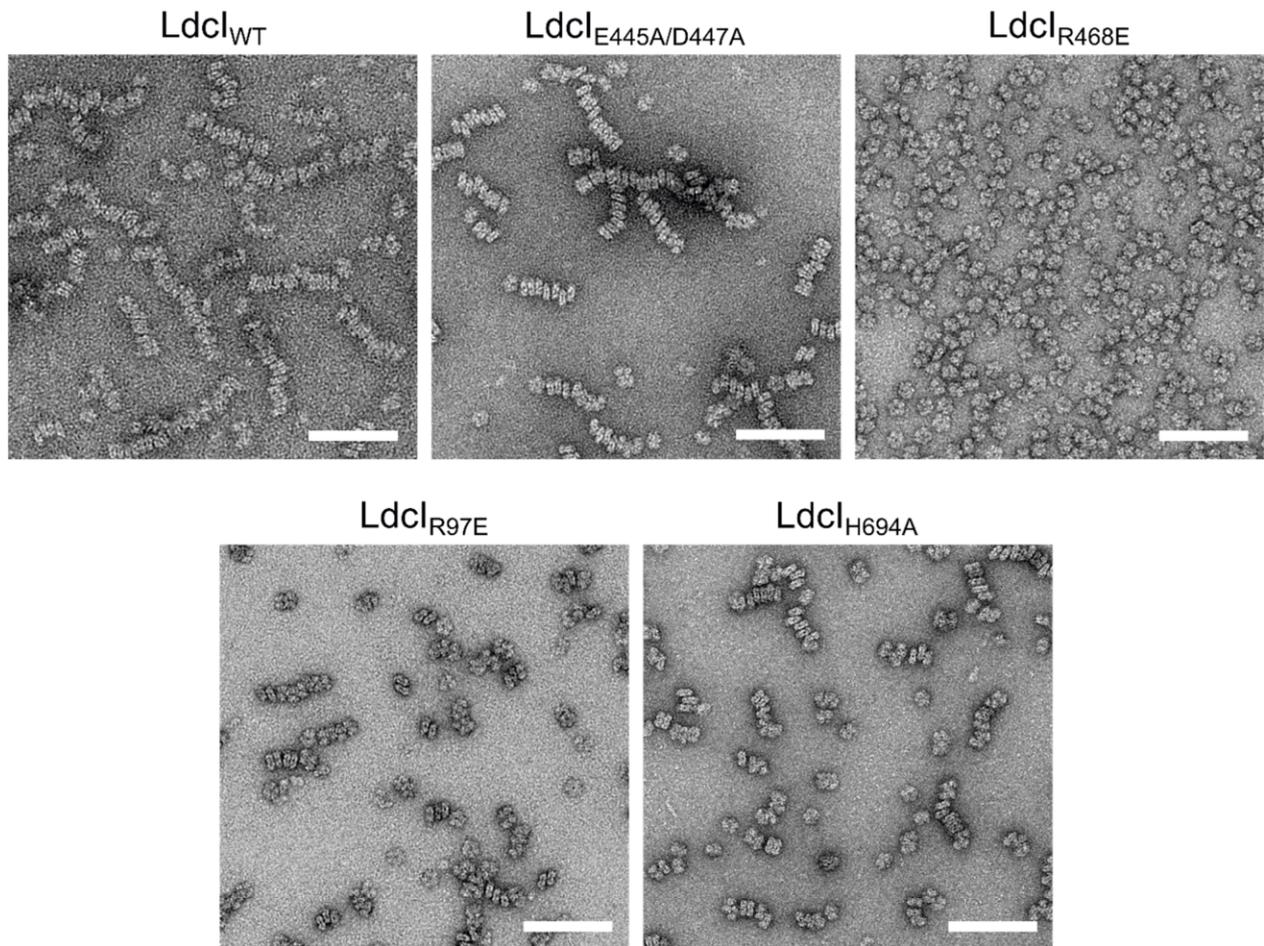
are coloured either blue/gold or green/coral, corresponding to the colouring of the atomic model presented in Figure 5. Scale bar = 50 Å. **D)** Overlay of Ldcl dimers at pH 8.5 (PDB ID: 3N75, shown in lilac), 7.0 (shown in cyan) and 5.7 (shown in gold). Alignment was carried out on a single monomer in the dimer pair. There is a 5° shift in the angle between the wing domains of Ldcl at pH 5.7 and pH 7.0/8.5. **E)** Comparison of the central decamer pore diameter between Ldcl at pH 8.5, 7.0 and 5.7, showing a 5 Å decrease in the pore size upon stack formation at low pH.

Figure 5.



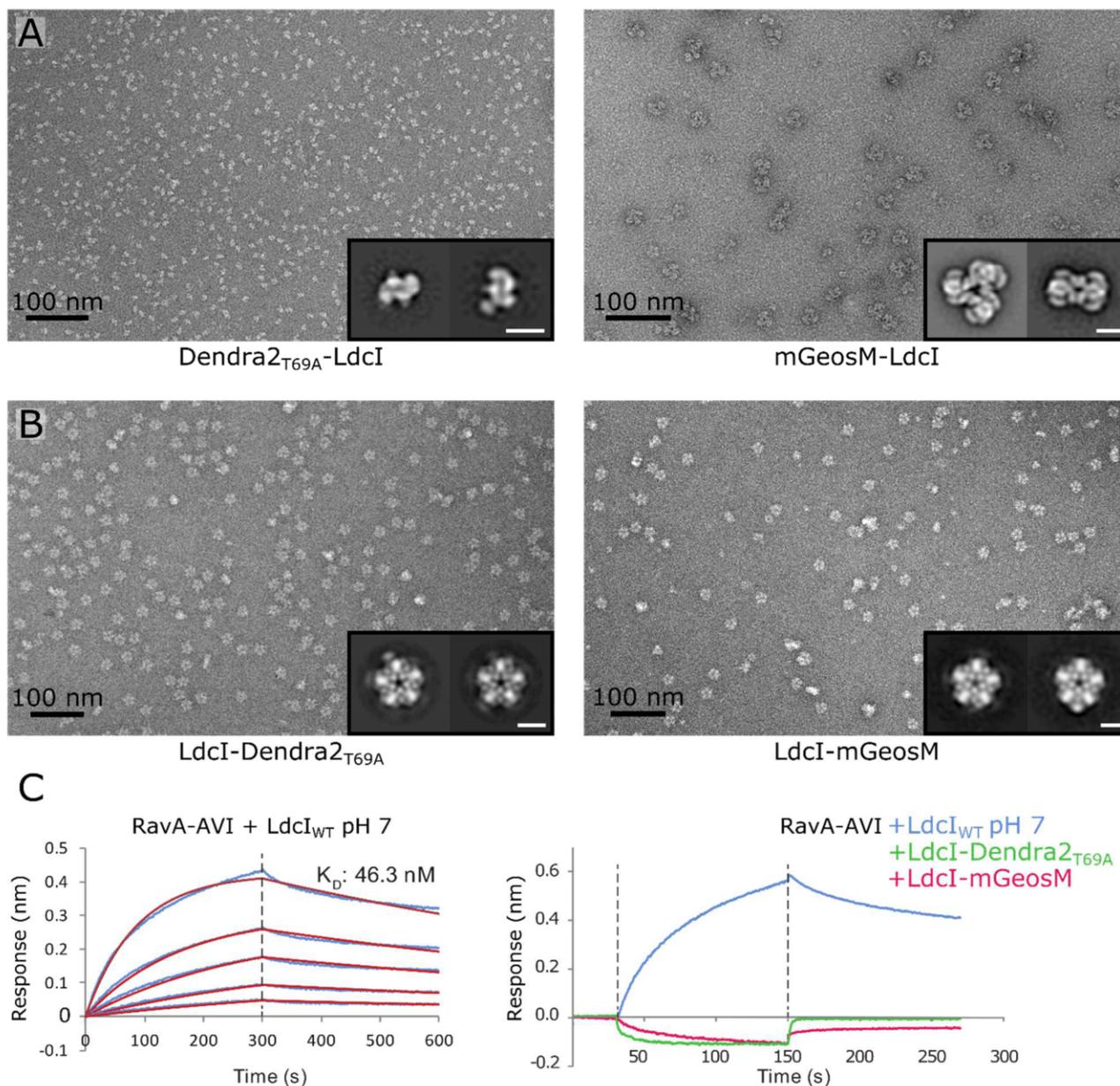
**Figure 5. Structural insights into molecular determinants of the Ldcl polymerisation under acid stress conditions.** **A)** Atomic model of a two-decamer Ldcl stack at pH 5.7. Dimers are coloured as shown in the cryo-EM map in Figure 4. Dotted boxes on the two-decamer stack indicate the locations of the zooms shown in panels B-E. **B)** Close-up of the first decamer-decamer interface, which includes the key stack-forming residue R468. **C)** Close-up of the second decamer-decamer interface. **D)** Overlay of the Ldcl decamer structures at pH 8.5 and 7.0 with the Ldcl stack structure at pH 5.7, focussed on R97. R97 in the Ldcl stack (green) adopts a different conformation compared to the one in the pH 8.5 crystal structure (purple, with ppGpp bound) and the pH 7.0 cryo-EM map (cyan, without ppGpp bound). Despite the absence of ppGpp in the pH 7.0 sample, R97 is still oriented towards the ppGpp binding site. **E)** Comparison between the H694-D192' distance in the Ldcl stack at pH 5.7 (coloured gold and blue), the Ldcl decamer at pH 8.5 (left, coloured purple) and pH 7.0 (right, coloured cyan). Key residues are labelled for all panels, and interactions are shown with dotted lines.

**Figure 6.**



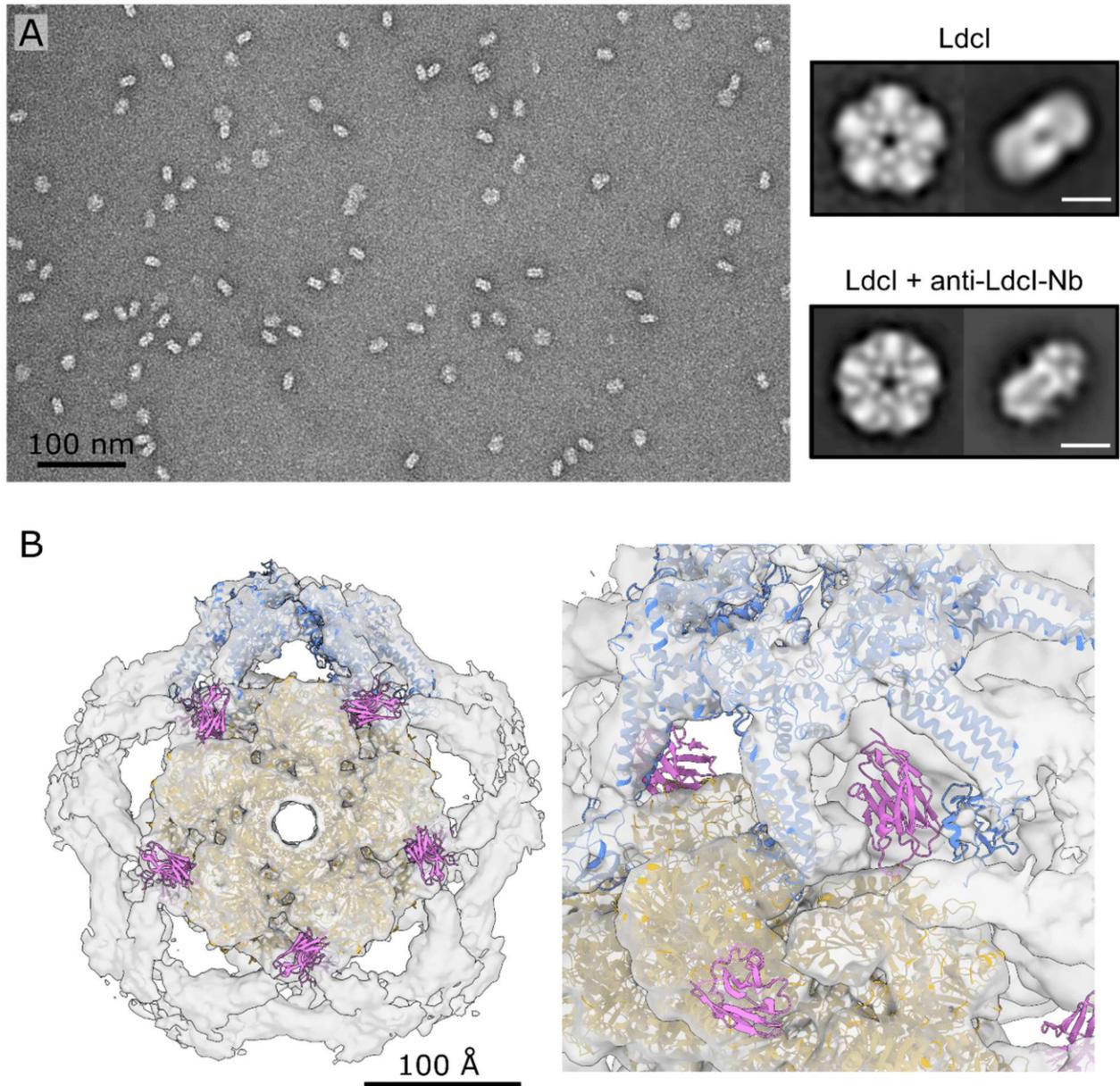
**Figure 6. Mutational analysis of the predicted molecular determinants of the Ldcl polymerisation under acid stress conditions** Cropped negative stain EM micrographs of wild-type and mutant Ldcl at pH 5.7, scale bar = 100 nm. Ldcl<sub>WT</sub> polymerises at pH 5.7, as does the double mutant Ldcl<sub>E445A/D447A</sub>. In contrast, the single mutation R468E abolishes stack formation completely. Both Ldcl<sub>R97E</sub> and Ldcl<sub>H694A</sub> are able to polymerise, however the stacks tend to be shorter than for Ldcl<sub>WT</sub>. Micrographs for Ldcl<sub>R97E/R468E</sub>, Ldcl<sub>E445A/D447A/R468E</sub>, and Ldcl<sub>H694N</sub> were also collected but are not shown here. Ldcl<sub>E445A/D447A/R468E</sub> and Ldcl<sub>R97E/R468E</sub> behaved like Ldcl<sub>R468E</sub> and remained entirely decameric at low pH, while Ldcl<sub>H694N</sub> displayed similar behaviour to Ldcl<sub>H694A</sub>.

## Supplementary Figure 1.



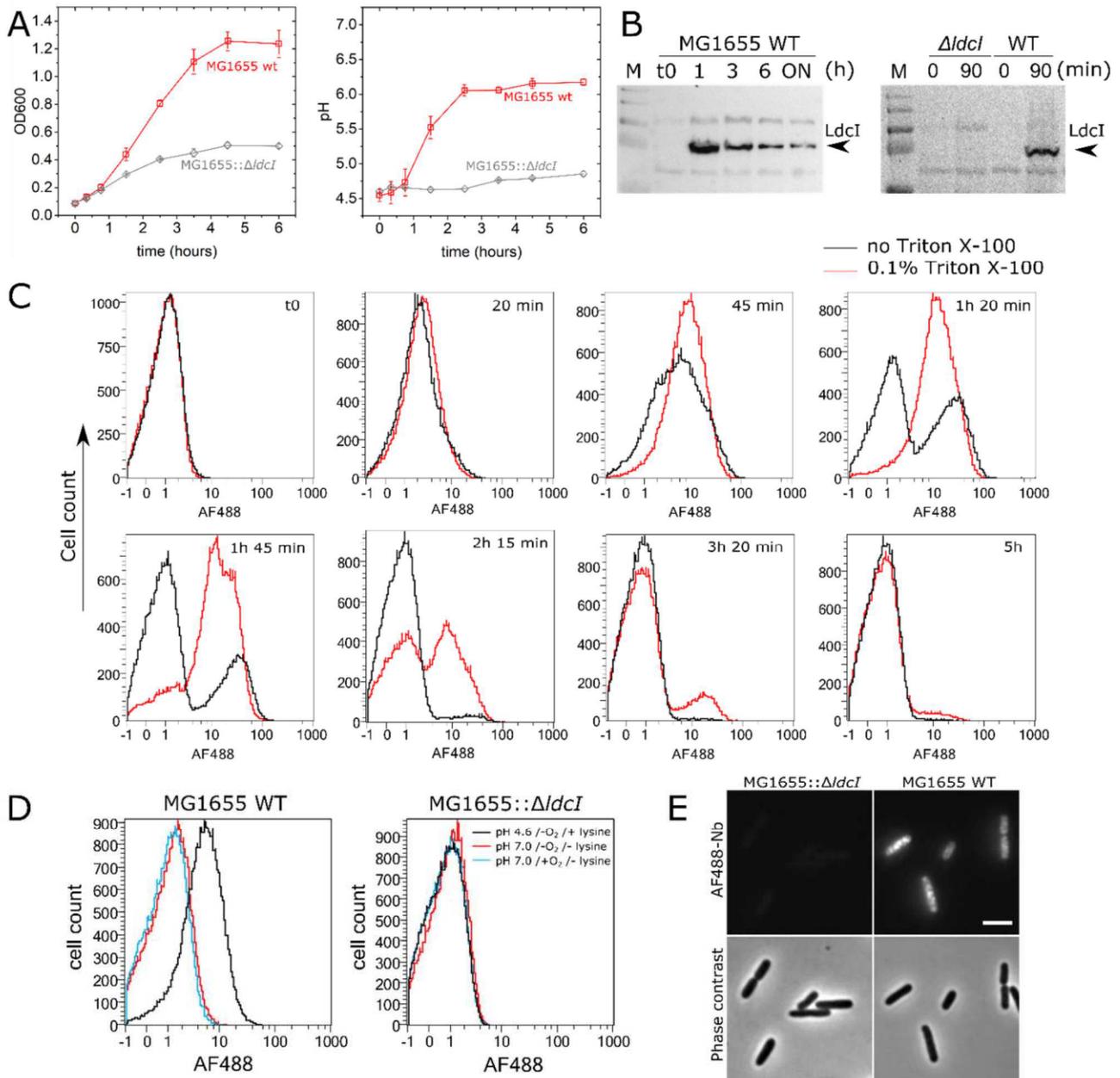
**Supplementary Figure 1.** Characterisation of **Dendra2<sub>T69A</sub>-LdcI** and **mGeosM-LdcI** fusion constructs by negative stain EM and BLI. **A**) Ns-EM micrographs of N-terminal FP-LdcI fusions Dendra2<sub>T69A</sub>-LdcI (left) and mGeosM-LdcI (right), with 2D class averages inset (scale bar = 100 Å). Dendra2<sub>T69A</sub>-LdcI only forms dimers, compared to mGeosM-LdcI which forms large non-native dodecamers. **B**) Ns-EM micrographs of C-terminal LdcI-FP fusions LdcI-Dendra2<sub>T69A</sub> (left) and LdcI-mGeosM (right), with 2D class averages inset (scale bar = 100 Å). Both fusions form decamers, with flexible-linked fluorophores visible as weak densities around the outside of the decameric ring in the 2D class averages. **C**) BLI binding curves of wild-type LdcI (left, [LdcI<sub>WT</sub>] = 500, 250, 125, 62.5, and 31.25 nM) and LdcI<sub>WT</sub> (right, 500 nM), LdcI-Dendra2<sub>T69A</sub>-LdcI (right, 500 nM) and LdcI-mGeosM-LdcI (right, 500 nM) against RavA-AVI. In contrast to wild-type LdcI, which binds RavA-AVI with a 46.3 nM affinity, there is no measurable interaction between either of the C-terminal LdcI-FP fusions and RavA-AVI.

## Supplementary Figure 2.



**Supplementary Figure 2. Ldcl interacts with anti-Ldcl-Nb and RavA at distinct sites.** **A)** Ns-EM micrograph of Ldcl + anti-Ldcl-Nb (left), with 2D class averages (right) of Ldcl alone and Ldcl + anti-Ldcl-Nb (scale bar = 100 Å). **B)** *In silico* model of anti-Ldcl-Nb (represented by the anti-lysozyme nanobody crystal structure, PDB ID: 1MEL) superimposed over the cryo-EM map of the Ldcl-RavA complex (Jessop et al., 2020) (EMDB ID: 4469, PDB ID: 6Q7L). The nanobody (coloured magenta) would occupy a distinct location to the RavA binding site in the Ldcl-RavA cage-like complex (coloured gold and blue).

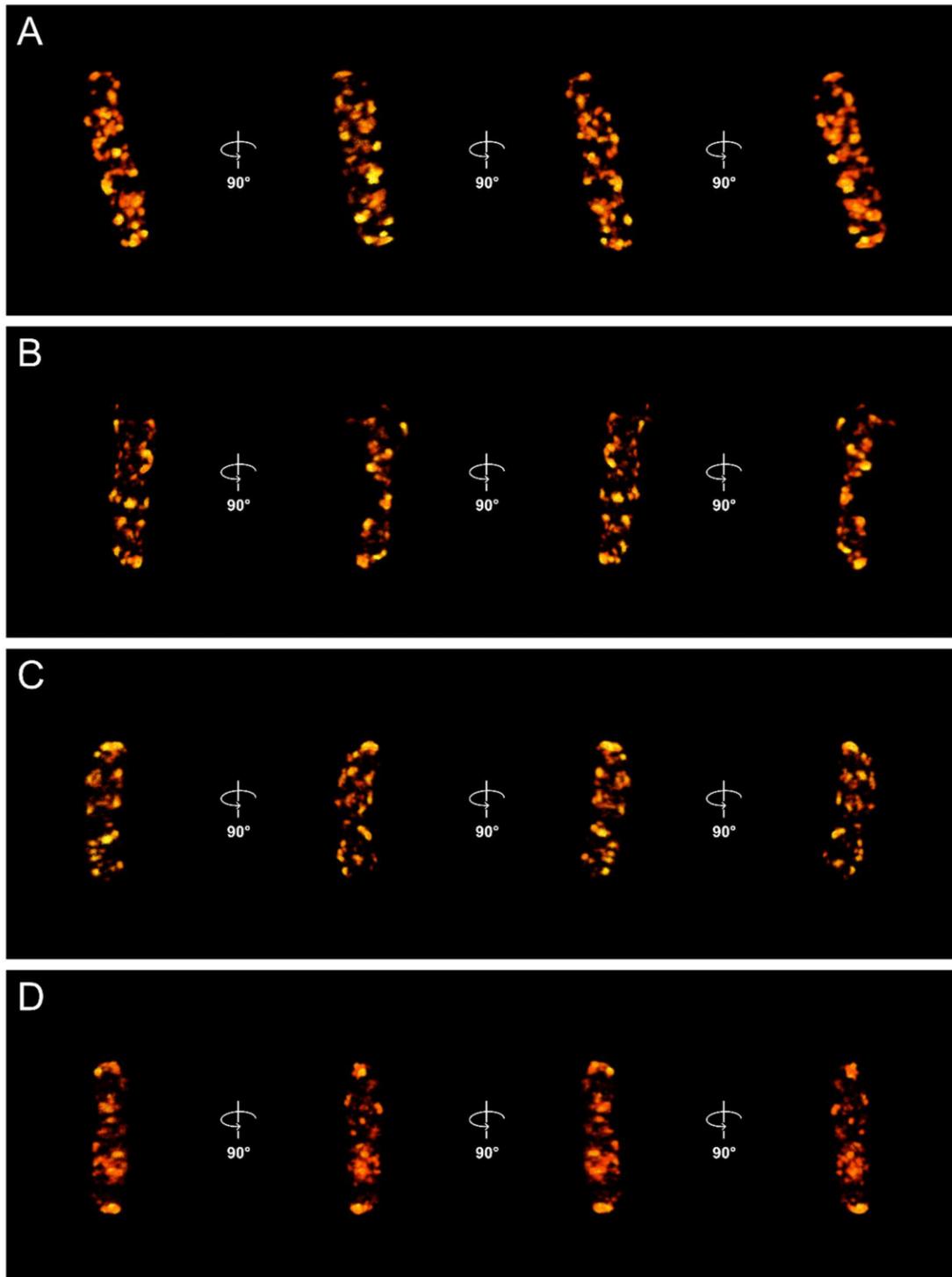
### Supplementary Figure 3.



**Supplementary Figure 3. Immunolabelling of endogenous LdcI in *E. coli* upon acid stress.** Characterisation of LdcI expression using anti-LdcI-Nb. **A**) Left – growth curve of wild-type (red) and LdcI knock-out (grey) cells after exposure to acid stress (pH 4.6), showing compromised growth in LdcI KO cells. Right – measured pH of the growth medium for wild-type and LdcI KO cells after exposure to acid stress (pH 4.6). Wild-type cells are able to buffer the medium under acid stress, whereas LdcI knock-out cells are not. Each data point represents the average of 3 independent measurements taken from 2 biological replicates, error bars correspond to the calculated standard deviation. **B**) Left – western blot showing LdcI expression after exposure to acid stress after 0, 1, 3 and 6 hours, and overnight (ON). LdcI expression is maximal soon after exposure to acid stress, tailing off after three hours. Right – western blot against LdcI in  $\Delta ldcI$  and WT MG1655 cells, 0 minutes and 90 minutes after exposure to acid stress, showing a lack of LdcI expression in  $\Delta ldcI$  cells and a lack of baseline expression before exposure to acid stress in WT cells. M = molecular weight marker. **C**) Flow cytometry measurements carried out after exposure of wild-type cells to acid stress show that LdcI is maximally expressed one to two hours after exposure to acid stress, before tailing off. **D**) Flow cytometry measurements of wild-type and LdcI knock-out cells labelled with anti-LdcI-Nb-AF488 after 90 minutes of growth under three conditions – pH 4.6 in the absence of oxygen and lysine, neutral pH in the absence of oxygen and lysine, and neutral pH in the presence of oxygen

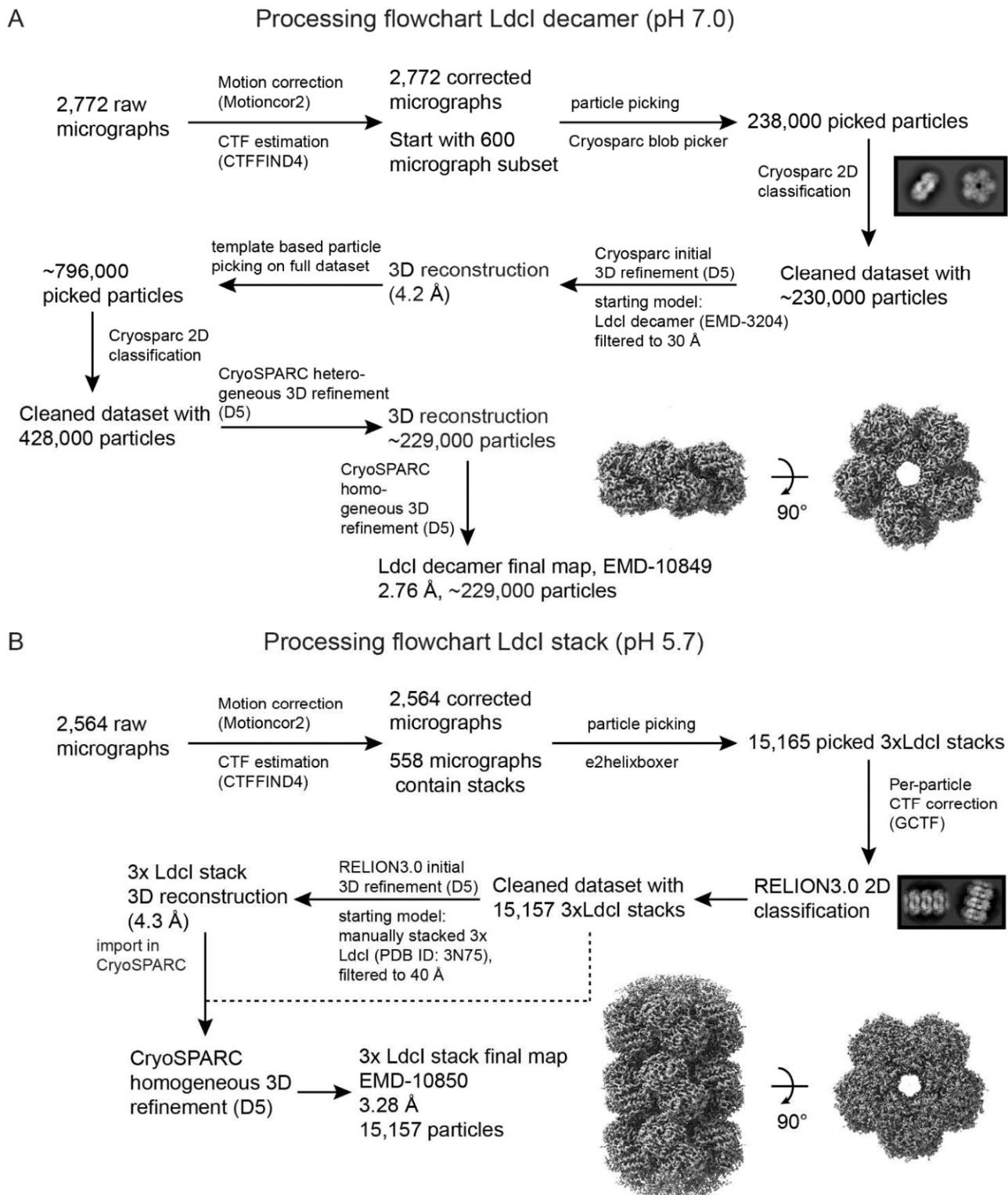
but absence of lysine. No fluorescence is seen above baseline for the LdcI knock-out cells under any condition. For wild-type cells, fluorescence is only seen under acid stress conditions. **E)** Wide-field fluorescence microscopy of LdcI knock-out MG1655 cells (left) and wild-type MG1655 cells (right), 90 minutes after exposure to acid stress. The staining of anti-LdcI-Nb-AF488 is specific, shown by the lack of background in the knock-out strain. Scale bar = 2 $\mu$ m.

Supplementary Figure 4.



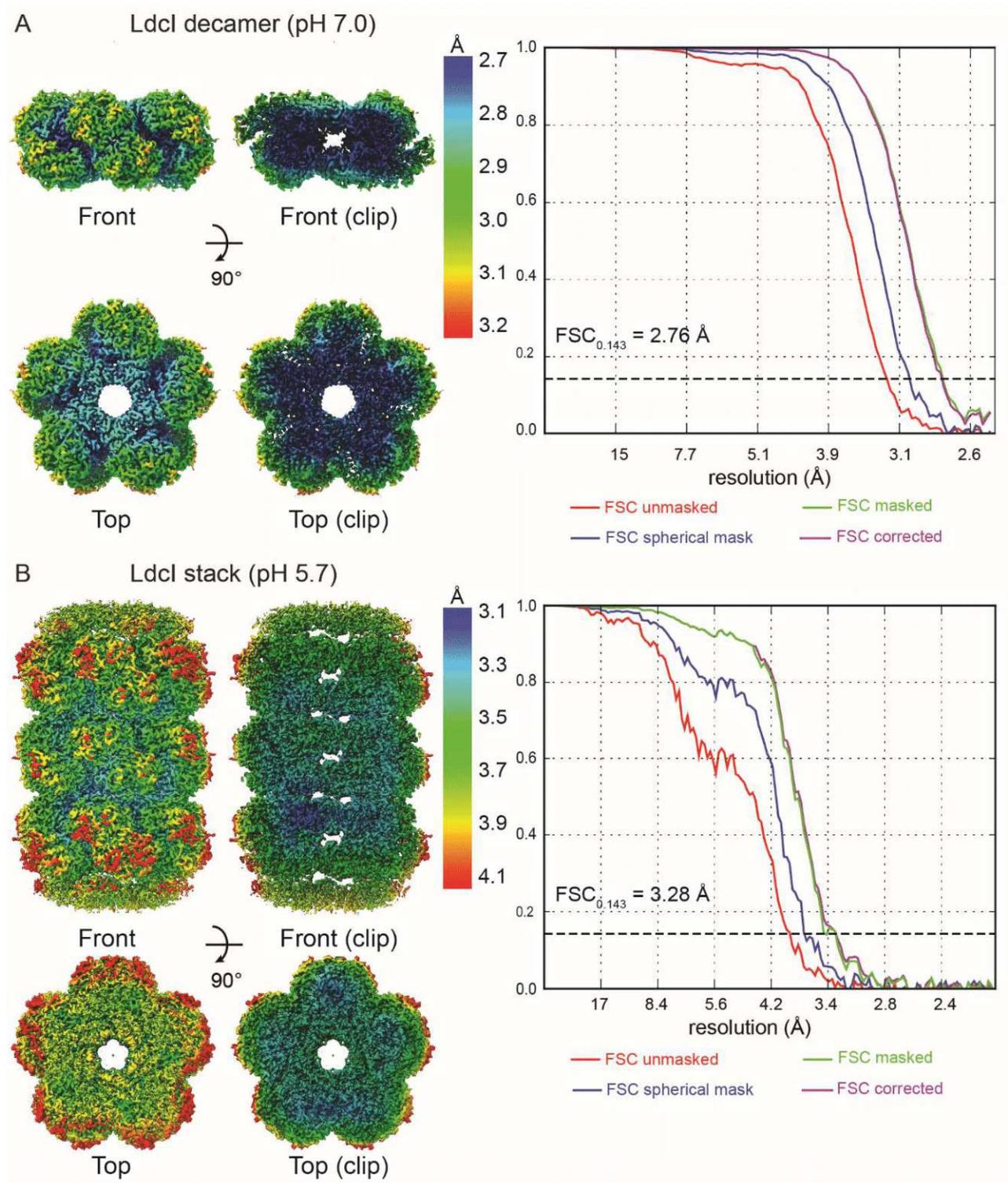
**Supplementary Figure 4. Galleries of 3D-STORM images of the four individual *E. coli* cells shown in Figure 3.** Different orientations of each cell related by  $90^\circ$  rotations around the long axis of the bacterium are shown in order to facilitate visualisation of the patchy distribution of Ldcl and the propensity of the Ldcl patches to assume a stripy or pseudo-helical long-range order.

## Supplementary Figure 5.



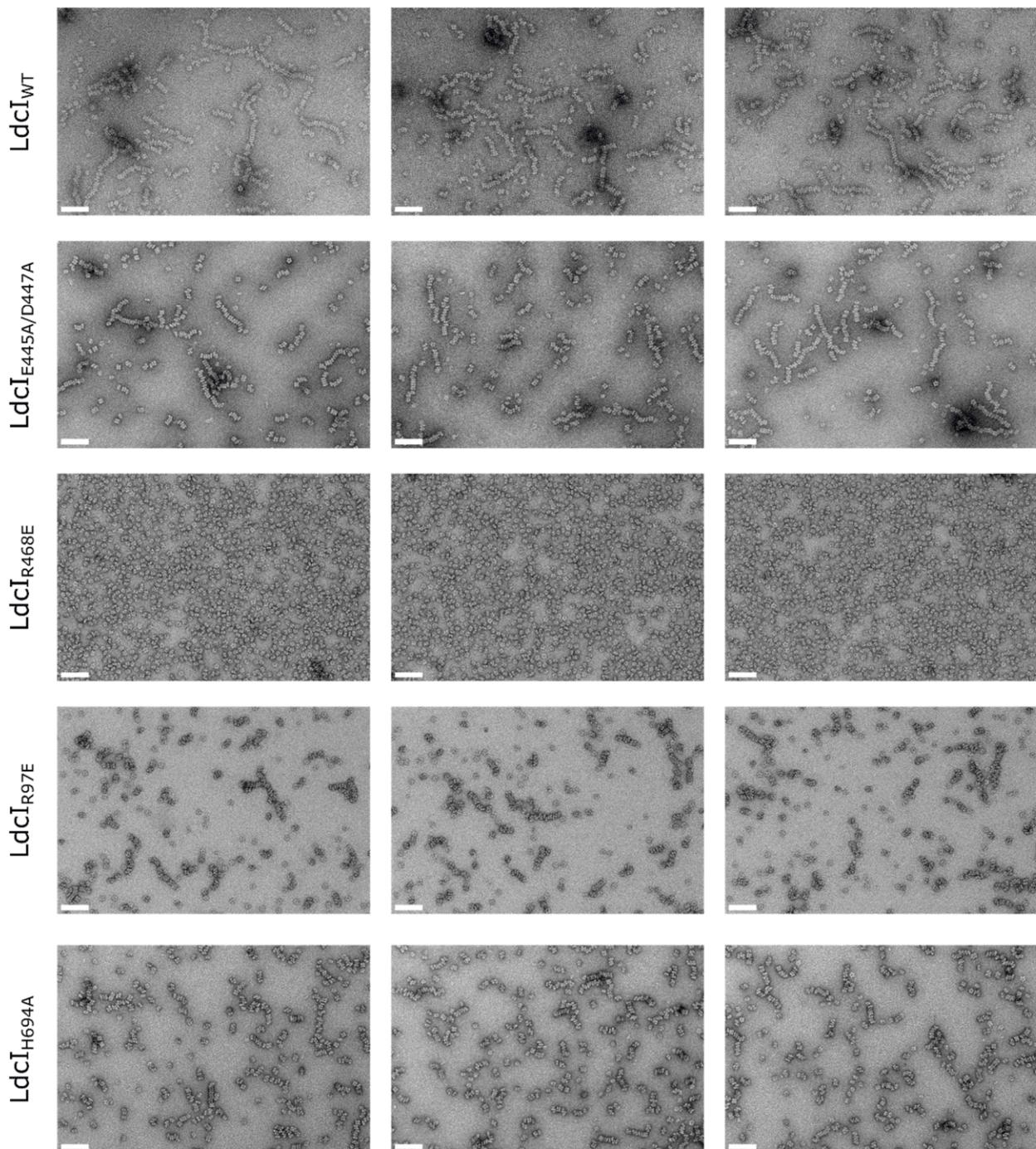
**Supplementary Figure 5. Cryo-EM processing pipelines. A)** Cryo-EM processing pipeline for Ldcl decamers at pH 7.0. Unless otherwise indicated, processing steps were carried out in CryoSPARC. **B)** Cryo-EM processing pipeline for Ldcl stacks at pH 5.7. Software packages used at each step are indicated. For a summary of data collection and processing parameters, and refinement and validation statistics, see Supplementary Table 1.

## Supplementary Figure 6.



**Supplementary Figure 6. Assessment of the resolution of the cryo-EM maps.** 3D reconstructions of **A)** the Ldcl decamer at pH 7.0 **B)** the Ldcl stack at pH 5.7, coloured according to the local resolution. Gold-standard FSC curves with the estimated resolution at FSC = 0.143 are shown on the right.

## Supplementary Figure 7.



**Supplementary Figure 7. Mutational analysis of the predicted molecular determinants of LdcI polymerisation under acid stress conditions.** Full uncropped ns-EM micrographs of wild type and mutant LdcI at pH 5.7, scale bars = 100 nm. Three representative micrographs are shown for each construct. Micrographs for LdcI<sub>R97E/R468E</sub>, LdcI<sub>E445A/D447A/R468E</sub>, and LdcI<sub>H694N</sub> were also collected but are not shown here. LdcI<sub>E445A/D447A/R468E</sub> and LdcI<sub>R97E/R468E</sub> behaved like LdcI<sub>R468E</sub> and remained entirely decameric at low pH, while LdcI<sub>H694N</sub> displayed similar behaviour to LdcI<sub>H694A</sub>.

**Supplementary Table 1. Cryo-EM data collection, refinement and validation statistics**

	Ldcl stacks (pH 5.7)	Ldcl decamers (pH 7.0)
	EMDB 10850 PDB 6YN6	EMDB 10849 PDB 6YN5
<b>Data collection and processing</b>		
Magnification	130,000	116,086
Voltage (kV)	300	200
Electron exposure (e <sup>-</sup> /Å <sup>2</sup> )	29.3	45.0
Defocus range (μm)	1.0 – 2.5	1.0 – 3.0
Pixel size (Å)	1.052	1.206
Symmetry imposed	D5	D5
Initial particle images (no.)	15,165	~796,000
Final particle images (no.)	15,157	~229,000
Map resolution (Å)	3.3	2.8
FSC threshold	0.143	0.143
Map resolution range (Å)	<b>69.0-3.3</b>	<b>51.2-2.8</b>
<b>Refinement</b>		
Initial model used (PDB code)	3N75	3N75
Model resolution (Å) (masked)	3.7	2.9
FSC threshold	0.5	0.5
Model resolution range (Å)	∞-3.7	∞-2.9
Map sharpening <i>B</i> factor (Å <sup>2</sup> )	-96 Å <sup>2</sup>	-173 Å <sup>2</sup>
Model composition		
Non-hydrogen atoms	113,760	57,074
Protein residues	14,220	7,110
Ligands	0	0
<i>B</i> factors (Å <sup>2</sup> )		
Protein	91.22	36.51
Ligand	-	-
R.m.s. deviations		
Bond lengths (Å)	0.009	0.012
Bond angles (°)	1.131	1.268
Validation		
MolProbity score	1.59	1.28
Clashscore	4.78	2.83
Poor rotamers (%)	0.49	0.19
Ramachandran plot		
Favored (%)	95.04	96.74
Allowed (%)	4.96	3.26
Outliers (%)	0.00	0.00

**Supplementary Table 2. Root-mean-square-deviation (RMSD) between extracted Ldcl monomer structures at pH 8.5 (PDB ID: 3N75), 7.0 and 5.7.**

	Ldcl decamer, pH 8.5	Ldcl decamer, pH 7.0	Ldcl decamer, pH 5.7
Ldcl decamer, pH 8.5	0	0.24 Å (670 aligned atoms)	0.61 Å (610 aligned atoms)
Ldcl decamer, pH 7.0	0.24 Å (670 aligned atoms)	0	0.56 Å (604 aligned atoms)
Ldcl decamer, pH 5.7	0.61 Å (610 aligned atoms)	0.56 Å (604 aligned atoms)	0

**Supplementary Table 3. Root-mean-square-deviation (RMSD) between extracted Ldcl dimer structures at pH 8.5 (PDB ID: 3N75), 7.0 and 5.7.**

	Ldcl decamer, pH 8.5	Ldcl decamer, pH 7.0	Ldcl decamer, pH 5.7
Ldcl decamer, pH 8.5	0	0.30 Å (1328 aligned atoms)	1.51 Å (1281 aligned atoms)
Ldcl decamer, pH 7.0	0.30 Å (1328 aligned atoms)	0	1.41 Å (1287 aligned atoms)
Ldcl decamer, pH 5.7	1.51 Å (1281 aligned atoms)	1.41 Å (1287 aligned atoms)	0

## Chapter 7. RavA fluorescent fusions – a cautionary tale

In addition to the fluorescently-tagged LdcI constructs analysed in the manuscript presented in Chapter 6, work was also carried out to characterise two fluorescently-tagged RavA constructs – RavA-mGeos-M, and Dendra2(T69A)-RavA. The rationale behind this was to enable the simultaneous imaging of both LdcI and RavA by two-colour super-resolution fluorescence microscopy, to understand under which conditions the LdcI-RavA cage would form in the cell.

Similarly to tagged LdcI constructs, I expressed and purified fluorescently-tagged RavA to assess the potential influence on the RavA hexamer structure. Both fluorescent fusions were analysed by negative stain-EM, and purified Dendra2(T69A)-RavA was then further analysed by cryo-EM. Structural analysis of the cryo-EM density highlights the need for thorough investigation of unexpected or novel results, and serves as a cautionary tale for the importance of verification for medium-resolution cryo-EM maps.

## 7.1. Expression and purification of fluorescently-tagged RavA

Rosetta 2(DE3) cells were transformed with p11 plasmids containing either His-tagged RavA-mGeos-M or Dendra2(T69A)-RavA and incubated at 37°C overnight. Precultures were grown from a single colony overnight shaking at 37°C, then expression of RavA was induced by addition of 40 mM IPTG and allowed to continue overnight at 18°C.

Cells were lysed using a microfluidizer operated at 15,000 Psi and the lysate was centrifuged at 20,000 rpm for 20 minutes. The supernatant was filtered with a 0.45 µm pore filter, then loaded onto an IMAC column. After washing the column with a buffer containing 20 mM imidazole, protein was eluted using a linear gradient from 20 mM to 250 mM imidazole. The elution peak was earlier than for wild-type RavA, indicating that the His-tag was possibly less accessible on the fluorescent protein than for native RavA. Fractions containing protein were dialysed to remove imidazole and concentrated to 500 µL, then loaded onto a Superose 6 gel filtration column. After gel filtration, samples from selected fractions were analysed by SDS-PAGE. Figure 7.1A and Figure 7.1B show the gel filtration profile and SDS-PAGE analysis respectively for Dendra2(T69A)-RavA; RavA-mGeos-M gave very similar results.

The gel filtration profile shows three main peaks. The earliest peak (fraction 10) contained almost no protein, and most likely contained nucleic acid (indicated by the elevated  $A_{255}$  to  $A_{280}$  ratio). The second peak (fraction 16) eluted near the expected elution volume of wild-type RavA, and exhibited strong fluorescence under a blue light compared to the gel filtration buffer (as shown in Figure 7.1C). This peak contained a large band at approximately 70 kDa, higher than the 56 kDa of RavA but seemingly lower than the expected 87 kDa of the fluorescent fusion, and a smaller band with a slightly higher molecular weight. This seemed to indicate that cleavage was occurring. The final peak (fraction 22) contained several proteins of approximately in the range of ~18 kDa to ~30 kDa. The molecular weight of Dendra2(T69A) alone is approximately 29 kDa – based on this, we expected that there was partial cleavage of the construct, either between the fluorescent protein and RavA or elsewhere, an observation which was also made for the fluorescently-tagged LdcI constructs.

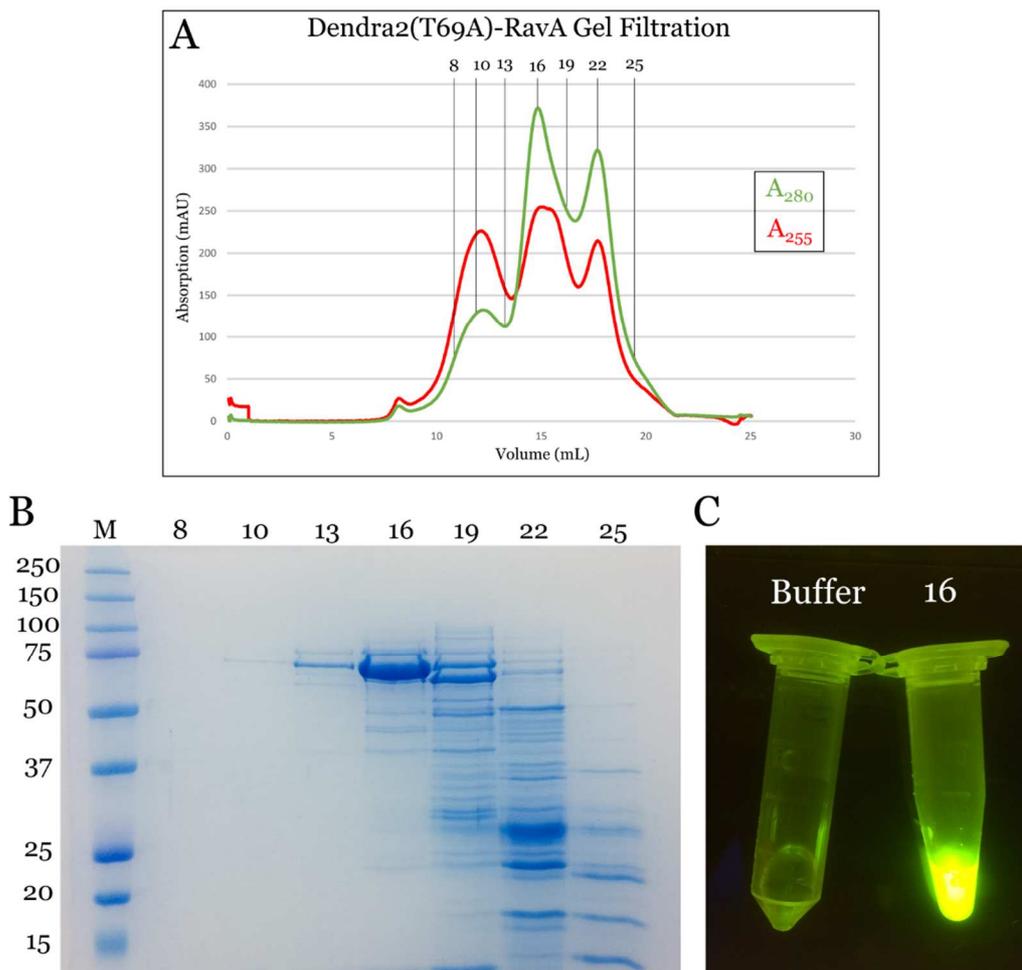


Figure 7.1. Gel filtration of Dendra2(T69A)-RavA. **A)** Gel filtration profile showing absorption at  $A_{280}$  (green) and  $A_{255}$  (red). Fractions analysed in **B)** by SDS-PAGE are labelled. **B)** SDS-PAGE analysis of selected gel filtration fractions in **A)**. M = molecular weight marker, molecular weights are shown in kDa on the left. **C)** Fraction A16 (right) displayed strong fluorescence under a blue light, compared to the gel filtration buffer (left).

## 7.2. Negative stain EM

To analyse the effect of the addition of an N-terminal or C-terminal fluorescent tag on the RavA oligomeric structure, samples were analysed by negative stain EM. Samples from the second gel filtration peak from both RavA-mGeos-M and Dendra2(T69A)-RavA were diluted to  $\sim 0.05$  mg/mL in the presence of 1 mM ADP and incubated for 10 minutes at room temperature. 3  $\mu$ L of sample was applied to the clean side of carbon on a carbon–mica interface and then stained with 2% uranyl acetate. RavA-mGeos-M was observed on a T12 microscope equipped with an Orius 1000 camera and operated at 120 kV, while Dendra2(T69A)-RavA was observed on an F20 microscope equipped with a Ceta camera and operated at 200 kV.

Strikingly, images of both Dendra2(T69A)-RavA and RavA-mGeos-M revealed the presence of regular, compact particles, bearing little resemblance to the wild-type RavA hexamer (Figure 7.2). To understand the nature of these particles, 80 micrographs of Dendra2(T69A)-RavA and 85 micrographs of RavA-mGeos-M were collected for analysis.

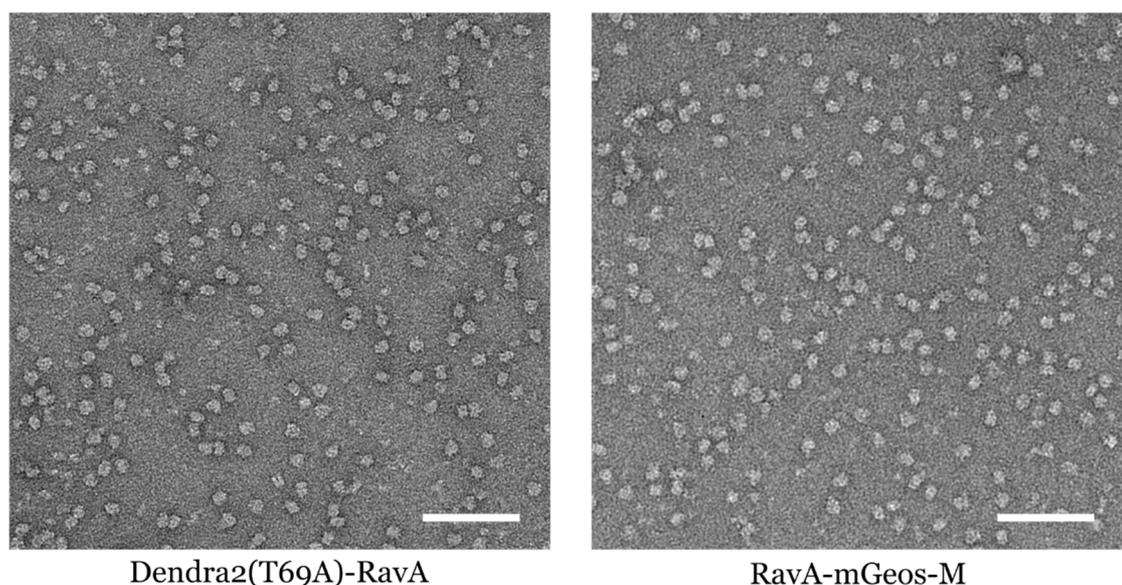


Figure 7.2. Negative stain EM micrographs of Dendra2(T69A)-RavA (left) and RavA-mGeos-M (right). Both samples show particles of a compact nature, which look similar in terms of size and shape. Scale bar = 100 nm.

CTF estimation was carried out on both datasets using CTFFIND4 (Rohou and Grigorieff, 2015). Semi-automatic particle picking was carried out using the Boxer program in EMAN (Ludtke et al., 1999), resulting in 45,094 particles for Dendra2(T69A)-RavA and 25,221 particles for RavA-mGeos-M respectively. Further steps were carried out in RELION-2.1 (Kimanius et al., 2016). Particles were extracted and subjected to several rounds of 2D classification. Samples displayed identical oligomers, which seemed to possess both three-fold and two-fold symmetry (Figure 7.3). The number of particles

retained after 2D classification was 34,285 and 12,546 for Dendra2(T69A)-RavA and RavA-mGeos-M respectively.

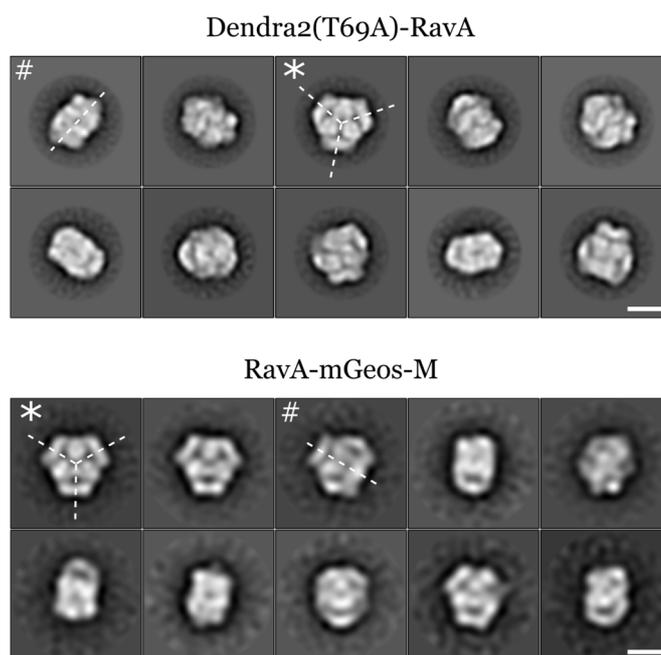


Figure 7.3. Negative stain EM 2D class averages for Dendra2(T69A)-RavA (top) and RavA-mGeos-M (bottom). Example classes with approximate two-fold (#) and three-fold (\*) symmetry are indicated, with symmetry planes shown as dotted lines. Scale bar = 100 Å.

Initial model generation was carried out for both cleaned datasets with no applied symmetry. Both models seemed to possess higher symmetry, with a three-fold rotational axis and a perpendicular two-fold axis, similar to the class averages. Therefore, 3D refinement was carried out with D3 symmetry applied for both datasets, resulting in reconstructions with a resolution of 15 Å and 19 Å for Dendra2(T69A)-RavA and RavA-mGeos-M respectively (using the FSC = 0.143 threshold). The resulting reconstructions are almost identical, although the Dendra2(T69A)-RavA map is better resolved. The core of the map contains six globular lobes around the three-fold axis of symmetry, three on the top and three on the bottom, with two-fold symmetrical rectangular regions at the periphery (Figure 7.4). The density maps do not however resemble the wild-type RavA hexamer, and fitting of the RavA crystal structure did not reveal how the monomers could pack to form such a particle. Considering our previous experience with the D3-symmetric mGeos-M-LdcI (see Chapter 6) we therefore reasoned that the addition of a fluorescent tag at either the N- or C-terminus caused dramatic rearrangement of the RavA oligomer.

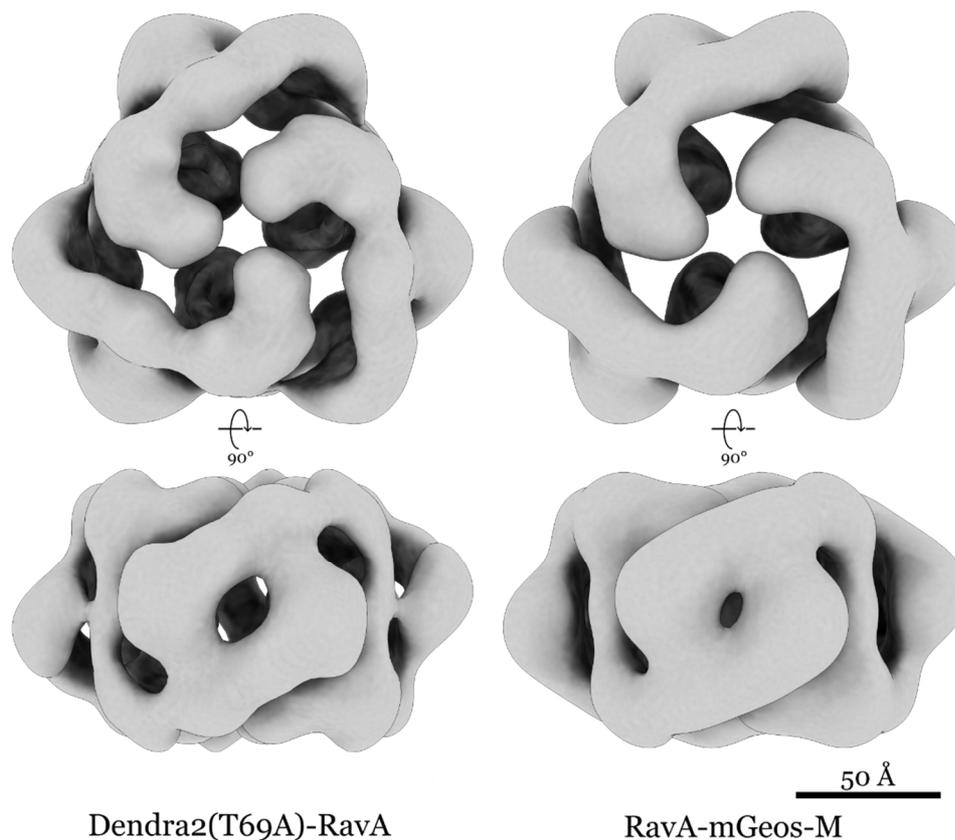


Figure 7.4. Top and side views (above and below) of Dendra2(T69A)-RavA and RavA-mGeos-M (left and right) 3D reconstructions respectively. The Dendra2(T69A)-RavA map possesses slightly more well-resolved features. Scale bar = 50 Å.

### 7.3. MALLS

Multi-angle laser light scattering (MALLS) was carried out on RavA-mGeos-M to estimate the molecular weight. Because of the structural similarity between RavA-mGeos-M and Dendra2(T69A)-RavA, only RavA-mGeos-M was analysed and was considered to be representative of both species. The estimated molecular weight of the single homogeneous peak was calculated to be 445 kDa. This closely corresponds to five RavA-mGeos-M monomers (theoretical molecular weight = 437.5 kDa). However, given the apparent D<sub>3</sub> symmetry of the complex, such an arrangement of was monomers deemed to be unlikely. Therefore, further work was required to characterise these fluorescent protein fusions and to understand the effect of fluorescent tags on the RavA structure.

## 7.4. Cryo-EM

The discrepancies between the theoretical and expected molecular weights, as well as the intriguing D<sub>3</sub>-symmetric complex and the parallels with the D<sub>3</sub>-symmetric mGeos-M-LdcI complex (see Chapter 6 for more details), prompted us to investigate further. Unlike mGeos-M-LdcI, where the fit of crystal structures in the negative stain EM map was obvious and unambiguous, there was no way to unambiguously place the RavA crystal structure in the density. We therefore decided to analyse Dendra2(T69A)-RavA by cryo-EM, to gain insight into how the fluorescent tag induced such large conformational changes.

### 7.4.1. Sample preparation

Fraction 16 from gel filtration (Figure 7.1) was dialysed overnight to remove glycerol, then checked again for sample integrity after dialysis by negative stain EM. 3  $\mu$ L of sample diluted to  $\sim$ 0.25 mg/mL in the presence of ADP was applied to Quantifoil 400 mesh R1.2/1.3 Cu holey carbon grids, after glow discharging at 20 mA for 45 seconds. Grids were frozen with a Vitrobot Mark IV operated at 100% humidity and room temperature, using blot force 1 and varying the blot time between 2-3 seconds.

### 7.4.2. Data collection

Grids were screened for ice thickness and particle distribution on an FEI Polara microscope operated at 300 kV equipped with a K2 detector. 2,774 micrographs were collected using LatitudeS at a nominal magnification of 41,270 x, corresponding to a pixel size of 1.21 Å/pixel at the specimen level, with a target defocus range of -1.5 to -3.5  $\mu$ m. A representative micrograph is shown in Figure 7.5, with several particles indicated. As well as the large three-fold symmetric particles, there were also smaller species present, possibly representing broken particles. The concentration of particles was lower than ideal, but sufficient to pursue with data collection.

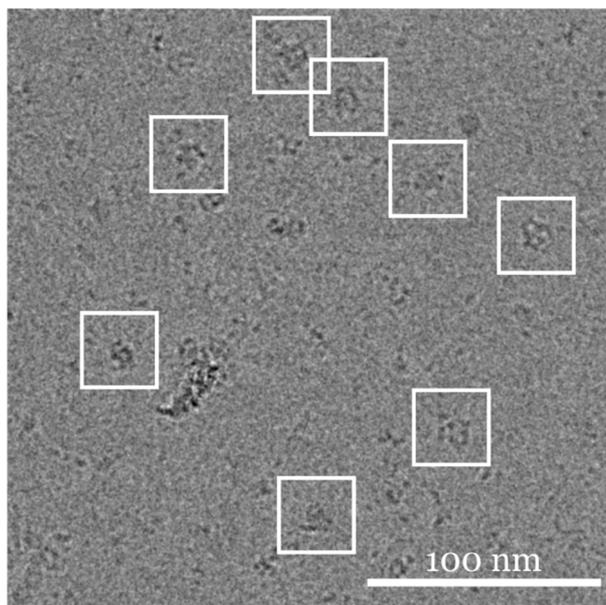


Figure 7.5. Crop of a cryo-EM micrograph of Dendra2(T69A)-RavA. Examples of picked particles are indicated by white boxes, scale bar = 100 nm.

### 7.4.3. Image processing

Motion correction was carried out using MotionCor2 (Zheng et al., 2017), and CTF estimation was carried out using CTFFIND4 (Rohou and Grigorieff, 2015). 2,498 micrographs with an estimated resolution of better than 8 Å were manually screened, and 1,888 micrographs were kept. 79,380 particles were selected from these micrographs with the Fast Projection Matching (FPM) algorithm (Estrozi and Navaza, 2008), using the negative stain model as a reference (see section 7.2). Particles were then extracted with a box size of 220 x 220 pixels and subjected to several rounds of 2D classification in RELION-2.1. 2D classes showed secondary structural features, and 11,006 particles were retained after 2D classification.

Initial model generation was carried out in RELION-2.1, using D3 symmetry. The cleaned particle dataset was then refined against this model with applied D3 symmetry, resulting in a reconstruction with a resolution of 8.5 Å (FSC = 0.143) after post-processing. This model was then used for a further round of particle picking by FPM, after filtering to a resolution of 20 Å. 125,240 re-picked and re-extracted particles were then subjected to several rounds of 2D classification, and the results are shown in Figure 7.6A. Classes clearly displayed the two-fold and three-fold symmetric views seen by negative stain EM, with visible secondary structure.

28,842 selected particles from the best 2D classes were then manually inspected and cleaned, resulting in a dataset of 23,165 particles. These particles were subjected to 3D

refinement with applied D3 symmetry against the previous model filtered to a resolution of 60 Å, resulting in a reconstruction with a resolution of 7.0 Å after post-processing (FSC = 0.143) and sharpening with a B-factor of -460 Å<sup>2</sup> (Figure 7.6B).

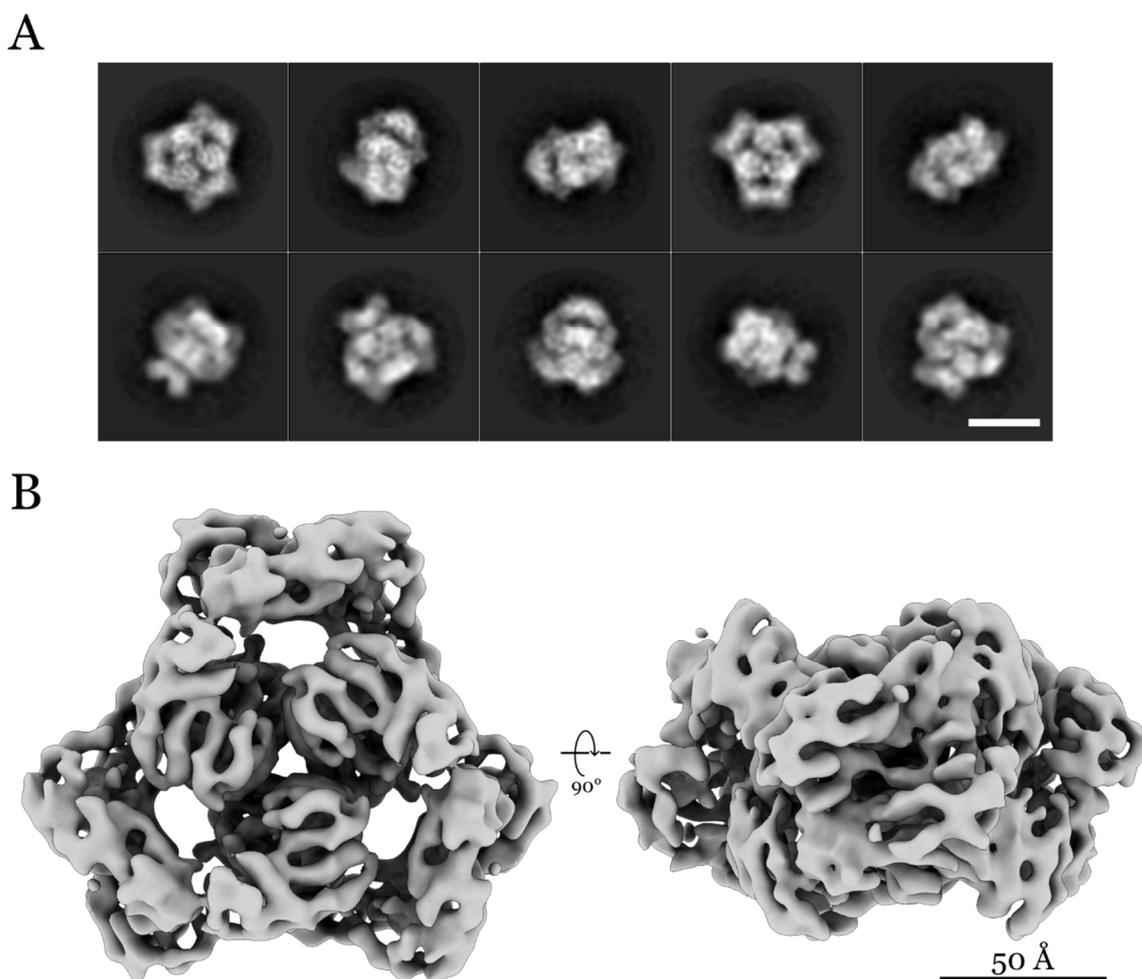


Figure 7.6. **A**) Representative 2D class averages, showing a good distribution of different views and secondary structural features. Scale bar = 100 Å. **B**) Top (left) and side (right) views of the 3D reconstruction. Secondary structural features are well-resolved, and the core of the map shows clear  $\alpha$ -helical density.

#### 7.4.4. Structural analysis

Despite the higher resolution and clear presence of secondary structural features, there was still no way to unambiguously fit the crystal structure of either RavA or the fluorescent protein into the cryo-EM density. Significant time was spent trying to fit either the entire RavA monomer or individual domains of RavA into the map, using exhaustive searches in Chimera (Pettersen et al., 2004). The fit which best explained the density is shown in Figure 7.7. The AAA+ domain of RavA is predominantly  $\alpha$ -helical and was similar in dimensions to the  $\alpha$ -helical core of the cryo-EM density.

Placing RavA monomers with the AAA+ domain in this way provided a plausible fit in terms of size and overall shape of the map. However, the fit was still unsatisfactory – while elements of secondary structure matched, particularly in the  $\alpha$ -helical core, not all of them fit the density. It was also unclear where the fused Dendra2(T69A) was located – we reasoned that this was most likely due to the flexible linker between the fluorophore and RavA.

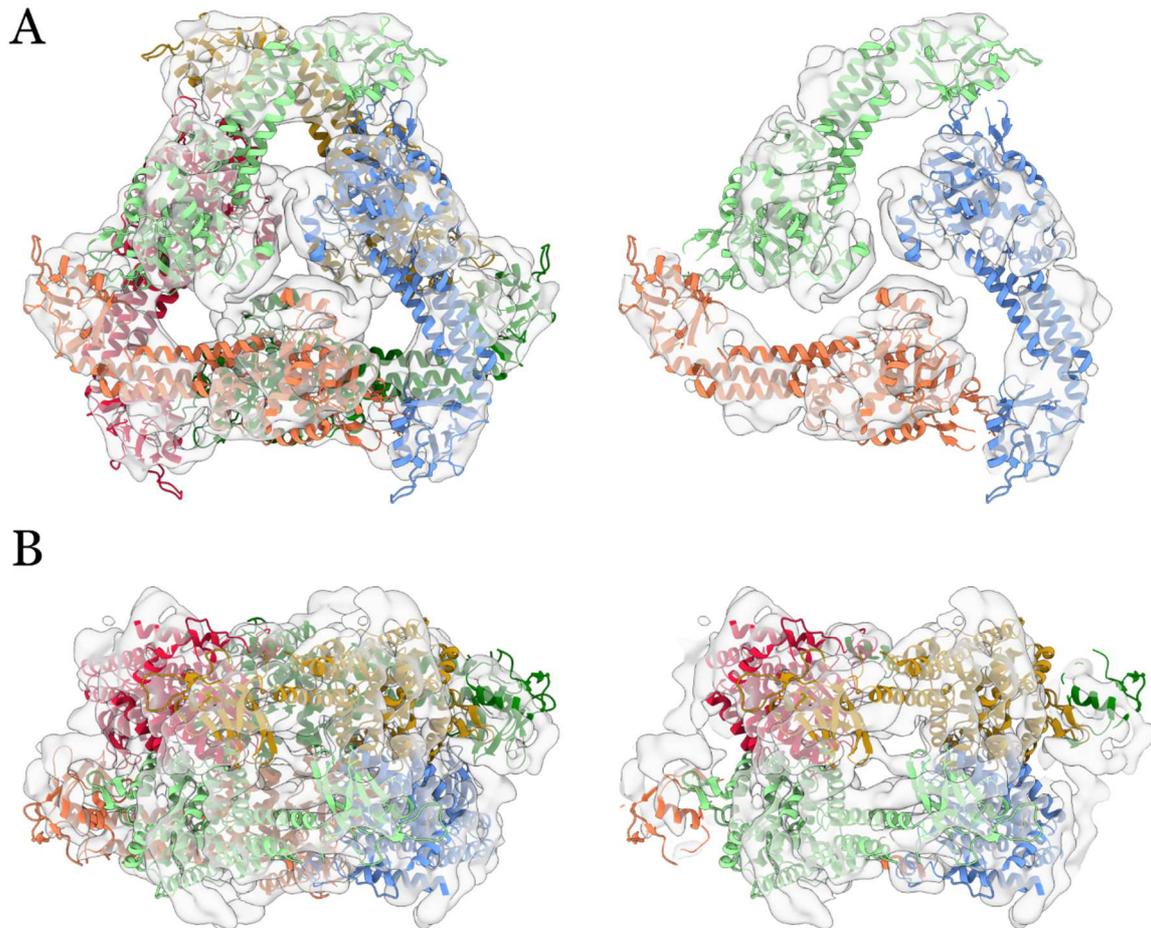


Figure 7.7. RavA monomers (PDB ID: 3NBX, monomers individually coloured) fitted into the cryo-EM density with the AAA+ domains placed in the  $\alpha$ -helical central core. **A)** Top view (left) and clipped top view (right) showing a possible arrangement of RavA monomers in the cryo-EM density. **B)** Corresponding side views, showing the full (left) and clipped (right) cryo-EM density with fitted monomers.

## 7.5. Mass spectrometry and N-terminal sequencing

Because of the apparent cleavage seen during purification and the unsatisfactory fit of the RavA monomer in the cryo-EM map (see Figure 7.1 and Figure 7.7), we decided to carry out mass spectrometry on the sample to understand exactly where the cleavage occurred, to ensure we were fitting the exact fragment of RavA into the cryo-EM density. The band from SDS-PAGE corresponding to the major species of fraction 16 after gel filtration (see Figure 7.1C) sample was subjected to LC-ESI-TOF (liquid chromatography-electrospray ionisation-time of flight) mass spectrometry and the major species in the sample corresponded to a molecular weight of 74.291 kDa. There were eight possible cleavage fragments of Dendra2(T69A)-RavA which could correspond to such a molecular weight, and unambiguous identification of the fragment was not possible.

Therefore, N-terminal sequencing was carried out using Edman degradation to unambiguously determine the sequence of the major species, and the result is shown in Figure 7.8. The N-terminal sequence of the major fragment was revealed to be MKTVV – unexpectedly, and somewhat distressingly, a sequence which did not appear in any part of the Dendra2(T69A)-RavA sequence!

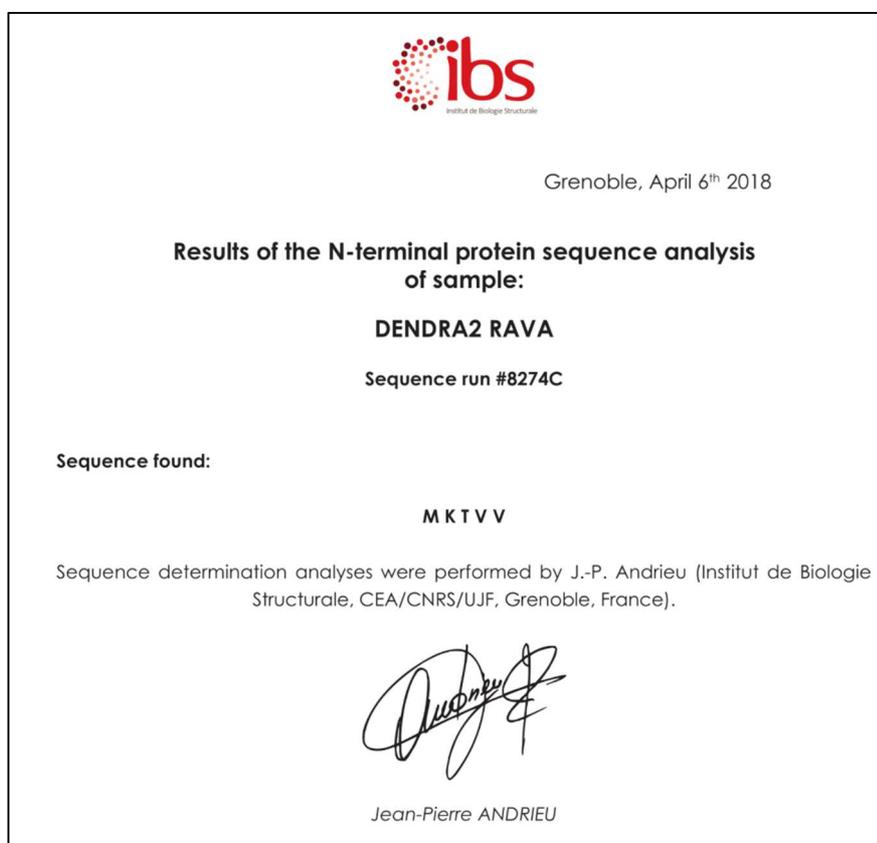


Figure 7.8. N-terminal sequencing of fraction 16 revealed a sequence of MKTVV, which is not found anywhere in Dendra2(T69A)-RavA.

## 7.6. Arn-believable

A search for the peptide fragment MKTVV in the genome of *E. coli* K-12 MG1655 was carried out using the EcoCyc database (Keseler et al., 2017), resulting in two hits. One of these, corresponding to the protein ArnA, contained the sequence at its N-terminus. Searching the Protein Data Bank for ArnA revealed that we had indeed succeeded in solving the structure of ArnA, not Dendra2(T69A)-RavA. Unfortunately, there are already several published high-resolution crystal structures of ArnA dating back to the early 2000s (Gatzeva-Topalova et al., 2004). The ArnA crystal structure fits unambiguously in the cryo-EM density (Figure 7.9). Further investigation revealed that ArnA is in fact a common contaminant during purification of His-tagged proteins using IMAC (Bolanos-Garcia and Davies, 2006).

ArnA is involved in the modification of lipid A, a key lipid present in the outer membrane of gram-negative bacteria (Raetz and Whitfield, 2002). It is a bifunctional enzyme, consisting of an N-terminal formyltransferase domain and a C-terminal decarboxylase domain (Gatzeva-Topalova et al., 2004). These domains catalyse two steps in the formation of 4-amino-4-deoxy-L-arabinose (Ara4N) from UDP-Glucose. Ara4N, a positively-charged sugar, is eventually added to lipid A by ArnT – this reduces the negative charge on the outer membrane, resulting in diminished binding of positively-charged antibiotics such as polymyxin, and therefore conferring resistance to the bacterium (Gatzeva-Topalova et al., 2004).

Compared to the crystal structure (PDB ID: 1Z7E), there is a slight rearrangement between the two domains of ArnA in the cryo-EM map. Therefore, the N- and C-terminal domains were docked separately in the fit shown in Figure 7.9. The C-terminal decarboxylase domain of ArnA is from the short-chain dehydrogenase/reductase family, which possesses a characteristic fold consisting of parallel  $\beta$ -sheets surrounded by  $\alpha$ -helices (Jörnvall et al., 1995). This fold is similar to the  $\alpha\beta\alpha$  fold of the AAA+ domain of RavA (see Figure 7.10), which we had indeed attempted to overlay with the C-terminal region of the ArnA cryo-EM map.

There were several pieces of evidence that led us to believe that we were indeed looking at a novel conformation of RavA induced by the addition of a fluorescent tag. ArnA is also hexameric, with a similar molecular weight to RavA, in gel filtration eluted near the wild-type RavA elution volume. The D<sub>3</sub> symmetry of the complex was reminiscent of the D<sub>3</sub>-symmetric mGeos-M-LdcI – possibly because of this, we expected that a similar artefactual oligomerisation had been induced by the fluorescent tag. In addition, the RavA AAA+ core and the ArnA decarboxylase domain both possess an  $\alpha\beta\alpha$  fold. Because of this,

it was not immediately obvious that the fit of the RavA crystal structure in the cryo-EM map was dramatically wrong, especially as we hypothesised that the addition of a fluorescent tag might have perturbed the RavA structure.

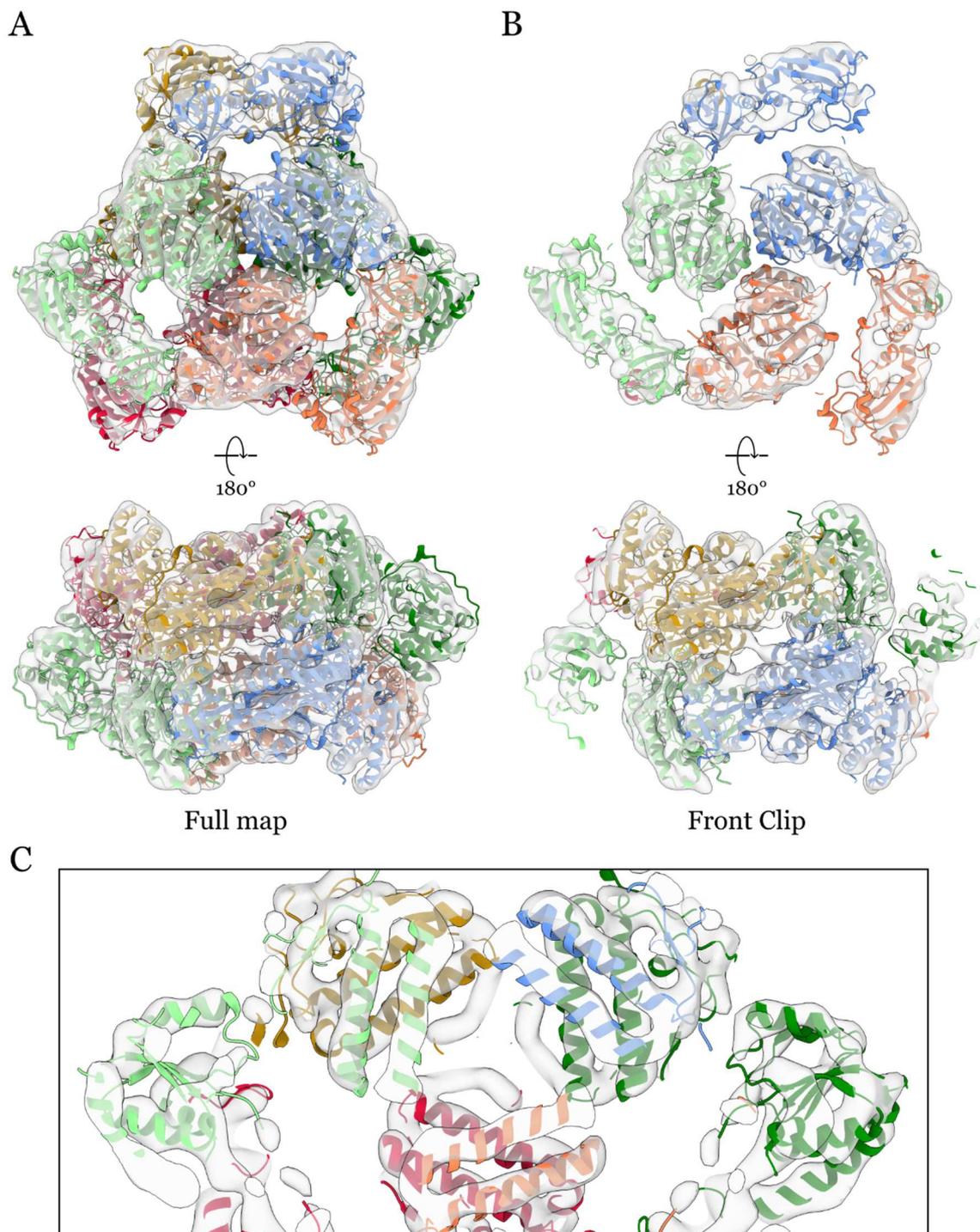


Figure 7.9. Fit of the ArnA crystal structure (PDB ID: 1Z7E) in the cryo-EM density. **A)** Top (above) and side (below) views of the reconstruction with fitted crystal structure. **B)** Top and side views as in A) but clipped to show just the front half of the reconstruction for better visualisation of the fitted crystal structure. The fit is unambiguous, and secondary structural elements fit the density well, as shown in **C)**.

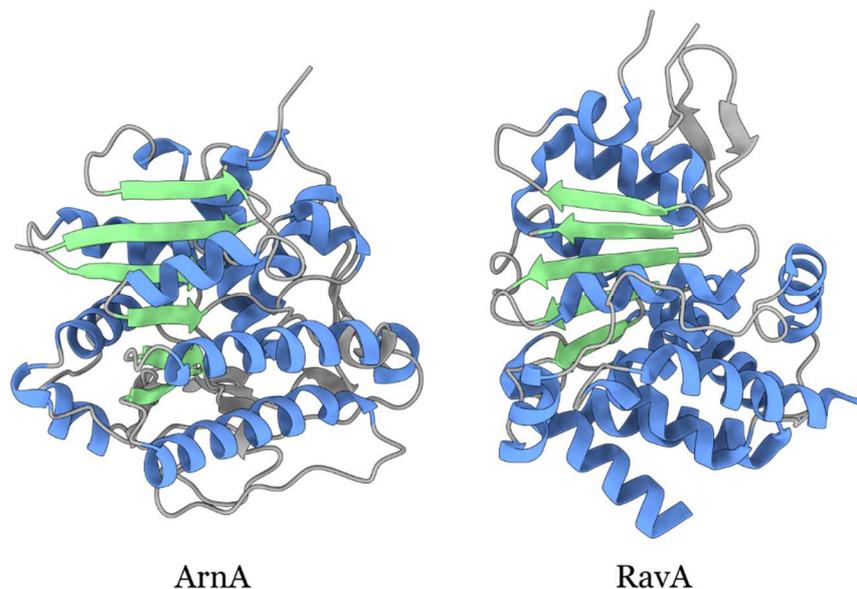


Figure 7.10. Comparison between the C-terminal decarboxylase domain of ArnA (PDB ID: 1Z7E, left) and the AAA+ domain of RavA (PDB ID: 3NBX, right). The core  $\beta$ -strands are coloured in light green, and the surrounding  $\alpha$ -helices are coloured in blue.

In hindsight, there were a few additional steps that could have been taken to verify the identity of the protein at an earlier stage, including carrying out a western blot with anti-RavA antibodies. However, the fact that the sample was also fluorescent (Figure 7.1C) further led us to believe that we were looking at the right protein. Revisiting the SDS-PAGE gel in Figure 7.1B reveals that what we had initially thought was partial cleavage, with a thin band of higher molecular weight and a thicker band of lower molecular weight, in fact corresponded to Dendra2(T69A)-RavA and ArnA respectively. The most likely reason that we only saw ArnA hexamers in electron microscopy images is that the addition of both an N-terminal Dendra2(T69A) and C-terminal mGeos-M completely abolished RavA hexamer formation. The ‘broken particles’ that we saw in the cryo-EM images may possibly correspond to monomeric Dendra2(T69A)-RavA, although the small molecular weight of a Dendra2(T69A)-RavA monomer makes this difficult to verify.

Although this was a disappointing result for me, with several months spent on data collection and analysis, I was still able to take some positives from the experience. As it was the first cryo-EM data collection carried out during my PhD, it gave me the chance to set up a data processing pipeline and learn some of the tools associated with cryo-EM image processing. In addition, it provided a very useful lesson about the importance of verifying seemingly novel or unexpected results through orthogonal methods.

## Chapter 8. Manuscript on the MCIA complex

As I started the PhD, we expected that the LdcI-RavA-ViaA triad would act as a chaperone in the assembly of *E. coli* respiratory Complex I based on genetic and biochemical studies by our collaborators (Babu et al., 2014; Wong et al., 2014a). In mitochondria, the triad ACAD9, ECSIT and NDUFAF1 form the core of the mitochondrial Complex I assembly (MCIA) complex, which we expected to play a similar functional role in mitochondria to the LdcI-RavA-ViaA triad in *E. coli*. Therefore, we began a collaboration with Montserrat Soler-López's group at the European Synchrotron Radiation Facility, who have been working on the MCIA complex for several years, to use electron microscopy to characterise the ACAD9-ECSIT-NDUFAF1 triad.

The manuscript presented in this chapter, which is currently under review, contains biochemical and functional studies on the MCIA complex from the Soler-López group together with my cryo-EM analysis on the ACAD9-ECSIT<sub>CTD</sub> subcomplex.

# ECSIT induces allosteric deflavination of a mitochondrial Complex I assembly factor

Gabriele Giachin<sup>1†</sup>, Matthew Jessop<sup>2†</sup>, Romain Bouverot<sup>1</sup>, Samira Acajjaoui<sup>1</sup>, Melissa Saïdi<sup>1</sup>, Anaïs Chretien<sup>1</sup>, Maria Bacia-Verloop<sup>2</sup>, Luca Signor<sup>2</sup>, Philippe J. Mas<sup>3</sup>, Adrien Favier<sup>2</sup>, Eve Borel Meneroud<sup>4</sup>, Michael Hons<sup>5</sup>, Darren J. Hart<sup>2</sup>, Eeazhisai Kandiah<sup>1</sup>, Elisabetta Boeri Erba<sup>2</sup>, Alain Buisson<sup>4</sup>, Gordon Leonard<sup>1</sup>, Irina Gutsche<sup>2\*</sup> & Montserrat Soler-Lopez<sup>1\*</sup>

<sup>1</sup> Structural Biology Group, European Synchrotron Radiation Facility (ESRF), CS 40220, 38043 Grenoble Cedex 9, France.

<sup>2</sup> Institut de Biologie Structurale (IBS), CNRS, CEA, Université Grenoble Alpes, 71 avenue des Martyrs, 38044, Grenoble, France.

<sup>3</sup> Integrated Structural Biology Grenoble (ISBG) CNRS, CEA, Université Grenoble Alpes, 71 avenue des Martyrs, 38042, Grenoble, France.

<sup>4</sup> Grenoble Institut des Neurosciences (GIN), Centre Inserm U1216, Equipe Neuropathologies et Dysfonctions Synaptiques, Université Grenoble Alpes, BP 170, 38042 Grenoble, France.

<sup>5</sup>European Molecular Biology Laboratory (EMBL), Grenoble Outstation, CS 90181, 38042 Grenoble, Cedex 9, France.

\* Correspondence to: [irina.gutsche@ibs.fr](mailto:irina.gutsche@ibs.fr); [solerlop@esrf.fr](mailto:solerlop@esrf.fr)

† Equal contribution

**Keywords:** ECSIT, ACAD9, Mitochondrial Complex I, Mitochondrial Complex I Assembly Complex, Fatty Acid Oxidation, Amyloid-beta, Alzheimer's disease, Cryo-EM, SAXS

**One Sentence Summary:** Mitochondrial bioenergetics in amyloid toxicity

34 **Abstract**

35 Fatty acid  $\beta$ -oxidation (FAO) and oxidative phosphorylation (OXPHOS) are redox processes  
36 driven by mitochondrial protein complexes to generate ATP. How these pathways are  
37 coordinated and how they are affected by the amyloid-beta toxicity of Alzheimer's disease are  
38 unknown. Here we show that soluble amyloids enhance the activity of the mitochondrial  
39 respiratory Complex I and its association by the assembly factor ECSIT. ECSIT functions as  
40 the central organizing component of the Complex I assembly complex, enabling recruitment of  
41 its partners NDUFAF1 and ACAD9. Remarkably, ECSIT induces the deflavination of  
42 ACAD9, switching ACAD9 from its role as a FAO enzyme to a Complex I assembly factor.  
43 These findings suggest a molecular mechanism for coordinating the regulation of the FAO and  
44 OXPHOS pathways to ensure an efficient energy production. Moreover, they identify ECSIT  
45 as a potential diagnostic biomarker for early Alzheimer's disease stages when mitochondria are  
46 sensitive to low levels of amyloid toxicity.

## 47 **Introduction**

48 Mitochondria coordinate central functions in the cell and have a crucial role in energy  
49 metabolism. Fatty acid  $\beta$ -oxidation (FAO) and oxidative phosphorylation (OXPHOS) are  
50 mitochondrial enzymatic pathways responsible for coupling redox reactions to ATP production  
51 (1). The OXPHOS system, composed of the respiratory chain components embedded in the  
52 mitochondrial inner membrane, includes five multimeric enzymes, the first of which is  
53 Complex I (CI or NADH:ubiquinone oxidoreductase) (2, 3). The redox reactions involved in  
54 ATP production generate reactive oxygen species (ROS), which have important roles in cell  
55 signalling and homeostasis but also cause oxidative damage to DNA, lipids and proteins  
56 including those in the respiratory chain itself. Neurons are extremely energy demanding and  
57 are therefore endowed with a high OXPHOS activity (4). Consequently, the particularly high  
58 ROS levels generated in synaptic terminals are at the root of their utmost vulnerability to  
59 mitochondrial dysfunctions (5). As the first enzyme of the respiratory chain, CI is the main  
60 source and target of ROS (6) and deficiencies in CI activity lead to the most common OXPHOS  
61 disorders in humans, which are often characterized by defects in the CI assembly process (7,  
62 8). Mammalian CI is the largest (~1 MDa) membrane protein complex, composed of 45  
63 different subunits (9). While the atomic structure of CI is known (9, 10), the mechanisms of CI  
64 biogenesis remain elusive (2, 11). Recent evidence suggests that CI biogenesis is assisted by  
65 assembly factors, including the recently described mitochondrial CI assembly (MCIA)  
66 complex, which are believed to dissociate from the final enzyme (12, 13).

67 The MCIA complex is composed of the three core subunits called NDUFAF1, ACAD9  
68 and ECSIT (14-16), knockdown of any of which in cultured cells leads to decreased levels of  
69 the other two partners and results in impaired CI assembly and activity (2, 17-19). However,  
70 the molecular details underlying of the activity of the MCIA complex are unclear, partly  
71 because besides participating in CI maturation, the individual components of the MCIA  
72 complex are also involved in other cellular functions (20, 21). In particular, ECSIT was initially  
73 identified as a cytoplasmic and nuclear signalling protein (22, 23) whereas ACAD9 was  
74 annotated as an acyl-CoA dehydrogenase (ACAD) enzyme due to its sequence homology with  
75 the very long chain acyl-CoA dehydrogenase VLCAD (17), which initiates the FAO pathway  
76 with the concurrent reduction of its FAD cofactor (24). Recent findings suggest physical  
77 interactions between FAO and OXPHOS proteins (25). Molecular insights into the protein  
78 interactions linking these two pathways are thus crucial for understanding the pathogenesis of  
79 diseases entailing OXPHOS and FAO deficiencies. A notable example is Alzheimer's disease  
80 (AD), with considerable evidence indicating a strong association between CI activity and

81 amyloid-beta ( $A\beta$ ) toxicity, a hallmark of AD (24). Indeed, the inhibition of CI seems to  
82 accelerate  $A\beta$  toxicity with a corresponding increase in ROS and may therefore contribute to  
83 age-related AD pathogenesis (26). Noteworthy, ECSIT has been identified as a molecular node  
84 interacting with enzymes producing  $A\beta$  (27, 28). However, the mechanism by which  $A\beta$  affects  
85 CI remains an open question.

86 In this work, we report a novel deflavination activity mediated by ECSIT that switches  
87 ACAD9 from its role as an FAO enzyme to a CI assembly factor, suggesting a possible  
88 mechanism for coordinating the regulation between OXPHOS and FAO protein complexes to  
89 ensure an efficient energy production. Furthermore, we provide molecular insights into the role  
90 of ECSIT as a CI stabiliser under amyloidogenic conditions.

## 91 **Results**

### 92 **$A\beta$ oligomers enhance ECSIT expression and its association with CI in neuronal cells**

93 Intra-neuronal  $A\beta$  can translocate directly from the endoplasmic reticulum into mitochondria,  
94 where it may affect mitochondrial respiration (29). We sought to investigate whether  
95 mitochondrial  $A\beta$  alters the activity of CI in human neuronal cells. To this end, we  
96 overexpressed the amyloid precursor protein (APP) carrying the AD-related Swedish mutation  
97 (KM670/671NL) in neuroglioma cells which enables the investigation of the toxic  $A\beta_{1-42}$   
98 soluble oligomeric form (**Fig. 1A**) as may occur during AD (26, 30). We analysed cytosolic  
99 and mitochondrial fractions from wild-type (WT) and APP overexpressing cells, as well as CI  
100 immunopurified from isolated mitochondria (**Fig. 1, B inset and C**). Interestingly, while we  
101 obtained similar levels of immunopurified CI from both cell types (**Fig. 1B inset**), the APP  
102 cells exhibited a significantly enhanced NADH-dehydrogenase activity (**Fig. 1B**). We then  
103 investigated whether the  $A\beta_{1-42}$  soluble oligomers can affect the expression of the MCIA  
104 factors. While the levels of ACAD9 and NDUFAF1 did not significantly vary, both cytosolic  
105 and mitochondrial levels of ECSIT increased under amyloidogenic conditions (**Fig. 1, C and**  
106 **D and Fig. S1**), in particular the levels of mitochondrial ECSIT associated with CI (**Fig. 1, C**  
107 **and D**).

### 108 **ECSIT is the central organizing component of the MCIA complex**

109 The increased association of ECSIT with CI under amyloidogenesis prompted us to investigate  
110 the role of ECSIT in the formation of the MCIA complex and its interaction with its MCIA  
111 partners ACAD9 and NDUFAF1. We first examined the interactions between the three proteins  
112 *in cellulo* by yeast two-hybrid (Y2H) (**Fig. 2A**) and mammalian cell bimolecular fluorescence  
113 complementation (BiFC) (**Fig. 2, B and C**) assays. While mitochondrial ECSIT, as well as

114 individual N- and C-terminal fragments (ECSIT<sub>NTD</sub> and ECSIT<sub>CTD</sub> respectively, described  
115 below) were found to associate with NDUFAF1 (**Fig. 2A**), only ECSIT<sub>CTD</sub> showed interaction  
116 with ACAD9 (**Fig. 2, B and C**). No direct interaction between NDUFAF1 and ACAD9 was  
117 observed (**Fig. 2A**), confirmed by *in vitro* interaction analysis with recombinant purified  
118 proteins (**Fig. 2D**).

119 That ECSIT appears to carry independent interaction sites for NDUFAF1 and ACAD9  
120 suggests that ECSIT is a central subunit in MCIA complex formation (**Fig. 2E**). We therefore  
121 further examined the interaction of ECSIT with either ACAD9 or NDUFAF1. ECSIT is  
122 predicted to contain an ordered N-terminal domain followed by a disordered C-terminal region  
123 (20). This, and the tendency of mitochondrial ECSIT (residues 49- 431) purified from  
124 recombinant bacteria to form soluble aggregates (**Fig. S2**), motivated us to apply ESPRIT  
125 (Expression of Soluble Proteins by Random Incremental Truncation) technology (31) in order  
126 to identify soluble and monodisperse ECSIT fragments (**Fig. S3**). Whereas all N-terminal  
127 fragments were insoluble, a C-terminal fragment (denoted ECSIT<sub>CTD</sub>, residues 248-431)  
128 showed high expression and solubility (**Fig. S3**). Interestingly, this fragment corresponds to  
129 the isoform ECSIT-3 (22), an alternative splicing variant that has been observed at the  
130 transcript level (23). Analyses of purified ECSIT<sub>CTD</sub> both by size-exclusion chromatography  
131 coupled to multi-angle laser light scattering (SEC-MALLS) (**Fig. S4**) and by small-angle X-  
132 ray scattering (SAXS) (**Fig. S5, A and Table S1**) yielded an approximate molecular mass of  
133 46 kDa, suggesting that ECSIT<sub>CTD</sub> is predominantly a dimer in solution. SAXS and NMR data  
134 (**Fig. S5, A and B**) indicated that these dimers are elongated and partially folded, consistent  
135 with the prediction of an intrinsically disordered domain. In addition, native mass spectrometry  
136 (native MS) showed that those dimers may further associate into tetramers or even higher-order  
137 multimers (**Fig. S5, C**).

138 Purified mature NDUFAF1 (residues 25-327) behaved as a monodisperse monomer in  
139 SEC-MALLS and SAXS experiments (**Fig. S4 and Table S1**) whereas, in agreement with  
140 previous studies (24), recombinant, purified ACAD9 (residues 38-621) was detected as a  
141 monodisperse homodimer by SEC-MALLS (**Fig. S4**). Negative-stain electron microscopy (ns-  
142 EM) (**Fig. S6, A and B**) and SAXS analysis (**Fig. S6, B and Table S1**) of ACAD9 revealed  
143 compact, rectangular envelopes (**Fig. S6, A**) consistent in shape with a previously reported  
144 homology model of the ACAD9 dimer (17). This model is based on the crystal structure of  
145 VLCAD, which shares 47% sequence identity with ACAD9 (**Fig. S6, B to D**).

146 In contrast to the interaction observed by Y2H (**Fig. 2A**), ECSIT<sub>CTD</sub> did not bind  
147 monomeric NDUFAF1 in electrophoretic mobility assays (**Fig. S7, A**), suggesting that the N-

148 terminal domain of ECSIT is required for a high-affinity interaction with NDUFAF1.  
149 Conversely, we were able to reconstitute a complex between ECSIT<sub>CTD</sub> and ACAD9 (ACAD9-  
150 ECSIT<sub>CTD</sub>) *in vitro* (**Fig. 3A** and **Fig. S7, B**). The complex is much more thermostable than  
151 either of its individual components, as observed by differential scanning calorimetry (**Fig. 3B**).  
152 However, SAXS analysis revealed a large volume implying a non-compact supramolecular  
153 assembly (**Fig. 3C**). In addition, NMR spectroscopy showed shift perturbations in an ECSIT<sub>CTD</sub>  
154 random coil region upon the addition of ACAD9 (**Fig. 3D** and **E**), suggesting that ACAD9  
155 interacts with flexible ECSIT<sub>CTD</sub> segments upon complex formation. The specificity of the  
156 ACAD9-ECSIT<sub>CTD</sub> interaction was further confirmed by the protection of ACAD9 from  
157 proteolytic cleavage upon ECSIT<sub>CTD</sub> binding (**Fig. 3F**). Interestingly, NDUFAF1 was unable  
158 to co-elute with the ACAD9-ECSIT<sub>CTD</sub> subcomplex (**Fig. S7, C**), again suggesting that the N-  
159 terminal domain of ECSIT is required for a high-affinity interaction with NDUFAF1.

160 We next investigated the molecular architecture of the ACAD9-ECSIT<sub>CTD</sub> complex by  
161 cryo-electron microscopy (cryo-EM) (**Fig. S8**). Analysis of the 2D class averages (**Fig. 4A**)  
162 showed a strongly preferred orientation and revealed that the complex possesses a defined core  
163 with visible secondary structural features (**Fig. 4B** and **Fig. 5A**). This core clearly corresponds  
164 to density for an ACAD9 dimer, consistent with the ns-EM map (**Fig. S6, B**). Some classes  
165 also show diffuse densities at the periphery of this core (**Fig. 4A**, white arrows). Subsequent  
166 3D classification and refinement (**Fig. S8**) encountered significant difficulties in achieving high  
167 resolution, most likely due to the preferred orientation but also the contribution of the  
168 peripheral densities hindering the accurate assignment of Euler angles to particles (**Fig. S8**).  
169 We therefore chose to filter the final reconstruction to a resolution of 15 Å, in order to first  
170 focus only on the global shape of the ACAD9-ECSIT<sub>CTD</sub> complex with the aim of defining the  
171 ECSIT<sub>CTD</sub> binding site. The extra densities visible at low thresholds most likely correspond to  
172 ordered segments of ECSIT<sub>CTD</sub> that protrude out of the central ACAD9 dimer core (**Fig. 4B**).  
173 The fact that only a small proportion of the bound ECSIT<sub>CTD</sub> is visible in the cryo-EM density  
174 is compatible with the NMR data (**Fig. 3D**), indicating that the disorder of ECSIT<sub>CTD</sub> is mostly  
175 retained upon binding to ACAD9 which thus hampers high-resolution cryo-EM analysis. The  
176 cryo-EM sample was also analysed by native MS (**Fig. S9**). This confirmed the presence of  
177 ECSIT<sub>CTD</sub> in the complex, yielding two main peaks of mass 197.5 kDa and 219.45 kDa  
178 respectively (**Fig. S9, B**). These correspond to an ACAD9 dimer bound to 3 or 4 ECSIT<sub>CTD</sub>  
179 monomers, suggesting that each ACAD9 monomer can bind up to a dimer of ECSIT<sub>CTD</sub>  
180 depending on the conditions used.

## 181 **ECSIT recognises the vestigial dehydrogenase domain of ACAD9**

182 ACAD9 comprises an N-terminal dehydrogenase domain and a C-terminal region, which  
183 includes: (i) a solvent exposed helical stretch (residues 445-482) showing low sequence  
184 conservation with VLCAD; (ii) a vestigial domain (residues 493-587) that appears to have  
185 arisen by duplication of the N-terminal domain but has lost the catalytic activity (17, 32); (iii)  
186 a C-terminal helix that stabilizes the homodimer (residues 588-621) (**Fig. S10, A**). Previous  
187 studies postulated that the solvent exposed helical stretch was responsible for the ability of  
188 ACAD9, but not VLCAD, to participate in Cl assembly (17, 33), raising the possibility that  
189 this helix might interact with ECSIT. Strikingly however, the densities attributed to ECSIT<sub>CTD</sub>  
190 in our cryo-EM map protrude from a distinct surface of ACAD9 (**Fig. 4B**), formed by two  
191 structural elements within the vestigial dehydrogenase domain: a helix-turn-helix (residues  
192 554-560) and a loop (residues 488-496) (17) (**Fig. S10**, regions 1 and 2 respectively).  
193 Interestingly, both regions are highly conserved across ACAD9 orthologues and,  
194 independently, across VLCAD orthologues but are highly different between these two proteins  
195 (**Fig. S10, B**).

196 In order to test whether the helical stretch of ACAD9 (residues 445-482) located nearby  
197 the vestigial domain in the homology model (**Fig. S10, A**, highlighted in grey) is also important  
198 for binding ECSIT (17), we designed and purified a chimeric protein, hereafter named  
199 ACAD9<sub>VLCAD</sub>, by replacing this stretch by the equivalent region of VLCAD (**Fig. S11, A**).  
200 SEC-SAXS showed that the global conformation of ACAD9<sub>VLCAD</sub> is similar to ACAD9 (**Table**  
201 **S1**). However, in contrast to VLCAD (**Fig. S12, A and B**), the chimera is still able to co-elute  
202 with ECSIT<sub>CTD</sub> (**Fig. S12, C and D**), confirming that the altered helical stretch is not required  
203 for the ECSIT interaction. Intriguingly, the dehydrogenase activity of the chimera was heavily  
204 impaired (**Fig. S11, B to D**), suggesting that the integrity of this region is important for catalytic  
205 efficiency.

## 206 **ECSIT induces ACAD9 deflavination**

207 Purified ACAD9 shows acyl-CoA dehydrogenase activity (**Fig. S11, C**), in agreement with  
208 previous reports (34). To verify whether ECSIT<sub>CTD</sub> affects the enzymatic activity of ACAD9,  
209 we analysed the redox state of the FAD cofactor bound to the catalytic pocket of ACAD9,  
210 which typically shows a strong yellow colour when oxidized (**Fig. 5B** and **Fig. S13, A**). Liquid  
211 Chromatography/Electrospray Ionisation Mass Spectrometry (LC/ESI-MS) analysis confirmed  
212 that only the homodimeric ACAD9, but not the monomeric, form of ACAD9 is able to bind  
213 FAD (**Fig. S13, A and B**), corroborating a previous report that ACAD9 homodimerisation is  
214 necessary for cofactor binding (35).

215 Strikingly, the yellow colour is not observed during co-purification of ACAD9 with  
216 ECSIT<sub>CTD</sub> (**Fig. 5B**). Subsequent quantification of FAD content by UV spectroscopy and  
217 LC/ESI-MS measurements revealed the absence of FAD in ACAD9-ECSIT<sub>CTD</sub> (**Fig. 5, B and**  
218 **C**). In addition, complex reconstitution by incubating purified samples of ACAD9 and  
219 ECSIT<sub>CTD</sub> showed that FAD is actively released when adding ECSIT<sub>CTD</sub> to ACAD9 (**Fig. 5D**).  
220 Finally, to obtain structural verification of the presence or absence of FAD in the binding  
221 pocket, we refined the ACAD9-ECSIT<sub>CTD</sub> cryo-EM reconstruction using a tight mask around  
222 the compact ACAD9 dimeric core to mitigate the contribution from the disordered ECSIT<sub>CTD</sub>  
223 (**Fig. S8**). The resulting cryo-EM map displayed clear secondary structural features consistent  
224 with the VLCAD-based homology model, thus validating this procedure (**Fig. 5A**). Moreover,  
225 the resulting map revealed that the FAD binding site is clearly empty in the ECSIT-bound state  
226 (**Fig. 5E**), corroborating the role of ECSIT in promoting ACAD9 deflavination.

## 227 **Discussion**

228 FAO and OXPHOS are key pathways involved in cellular energetics. Despite their functional  
229 relationship, evidence for a physical interaction between the two pathways is sparse.  
230 Understanding how FAO and OXPHOS proteins interact and how defects in these two  
231 metabolic pathways contribute to mitochondrial disease pathogenesis is thus of utmost  
232 importance for the development of new tailored therapeutic strategies.

233 Here, we provide evidence that ECSIT, a protein involved in cytoplasmic and nuclear  
234 signalling pathways, acts as the central organising component of the mitochondrial Complex I  
235 assembly complex (MCIA) required for the Complex I biogenesis at the entry of the  
236 mitochondrial respiratory chain and thereby for the activation of the OXPHOS system (**Fig. 6**).  
237 ECSIT is able to recruit its MCIA partners NDUFAF1 and ACAD9 using independent  
238 interaction interfaces, reflecting the modular nature of ECSIT, which is organised into N- and  
239 C-terminal binding domains. We show that the C-terminal region of ECSIT is highly  
240 disordered but forms a stable dimer and binds ACAD9 in a region of the vestigial domain.  
241 Strikingly, ECSIT acts as a negative allosteric effector by inducing ACAD9 to eject its FAD  
242 cofactor from the catalytic site, triggering the shutdown of dehydrogenase activity. Thus, upon  
243 binding to ACAD9, ECSIT actively antagonizes the FAO pathway and redirects ACAD9 to a  
244 function as a CI assembly factor. Considering the observation of a similar phenomenon for  
245 ETF dehydrogenase holoenzyme, involving the efficient removal of FAD upon binding the  
246 ETF regulatory protein LYRM5 (36), this negative allosteric regulatory mechanism may apply  
247 to many electron transfer flavoproteins.

248 Structural analysis of the ACAD9-ECSIT<sub>CTD</sub> complex reveals that ECSIT stabilizes the  
249 ACAD9 homodimer through the vestigial domain in the absence of FAD, suggesting that this  
250 domain was co-opted for protein interaction and evolved as a mode of complex assembly. In  
251 contrast, the solvent exposed helical region is likely implicated in other functions, possibly the  
252 association of ACAD9 with the inner mitochondrial membrane as previously hypothesized (34,  
253 35).

254 Very interestingly, at low levels of oligomeric A $\beta$  toxicity in neuronal mitochondria,  
255 such as occurs in early AD progression (24), the translocation of ECSIT to mitochondria is  
256 upregulated and ECSIT remains highly associated with CI even after CI assembly is completed,  
257 unlike in normal physiological conditions, where assembly factors are believed to dissociate  
258 from the assembled enzyme (12, 13). The fact that this association is concomitant with an  
259 enhanced CI activity suggests that ECSIT may be a stabiliser of CI under amyloidogenic  
260 conditions. Progressively, this increased activity may lead to an NADH/NAD<sup>+</sup> redox imbalance  
261 that generates oxidative stress and in turn, exacerbates the accumulation of A $\beta$  oligomers in a  
262 vicious cycle, which may, in agreement with previous studies (26), ultimately result in the  
263 inhibition of the CI activity.

264 Our results lead us to propose a new role for ECSIT as a stabiliser molecule, which may  
265 simultaneously direct the MCIA complex to the membrane assembly intermediates of CI by  
266 stabilising ACAD9 in the inner mitochondrial membrane and to the matrix-soluble  
267 intermediates of CI via the interaction with NDUFAF1 (**Fig. 6**). The ECSIT-mediated  
268 regulation of the dual functions of ACAD9 would enable a coordinated crosstalk between the  
269 FAO and OXPHOS pathways to ensure efficient energy production. On the one hand, the  
270 formation of the MCIA complex facilitates the assembly of membrane and matrix  
271 intermediates as part of the CI biogenesis, while on the other, the removal of FAD from the  
272 ACAD9 catalytic site makes ACAD9 quiescent in the initiation of the  $\beta$ -oxidation of fatty acids  
273 (**Fig. 6**). Finally, our findings also suggest that ECSIT acts as a molecular link between CI and  
274 mitochondrial bioenergetics and may be a potential biomarker for mitochondria-targeted  
275 diagnostics aimed at the detection of early AD pre-symptomatic stages when mitochondria are  
276 primarily affected by oligomeric A $\beta$  toxicity.

277

## 278 **Materials and Methods**

### 279 **DNA plasmids**

280 The human DNA sequences coding for mitochondrial ECSIT (ECSIT<sub>Mt</sub>, residues 49-431),  
281 ECSIT N-terminus (ECSIT<sub>NTD</sub>, residues 49-252), ECSIT C-terminus (ECSIT<sub>CTD</sub>, residues 247-  
282 431), mature ACAD9 (residues 38-621) and chimeric ACAD9 (ACAD9<sub>VLCAD</sub>, replacement of  
283 ACAD9 residues 445-482 by VLCAD residues 481-518) were *E. coli* codon-optimized and  
284 synthesized by ShineGene Molecular Biotech. Mature NDUFAF1 (residues 25-327) and  
285 mature VLCAD (residues 75-655) were PCR-amplified from human cDNA clones (GE  
286 Healthcare plasmid #3504355 and Addgene plasmid #38838, respectively). All the other DNA  
287 sequences were inserted into the pET-21d(+) expression vector. To generate the Venus  
288 fragment-fused plasmids for BiFC assays, the DNA sequences of mitochondrial ECSIT, ECSIT  
289 N-terminus, ECSIT C-terminus and mature ACAD9 were inserted into the pBiFC-VC155  
290 (Venus residues 155-238, A206K) or pBiFC-VN173 (Venus residues 1-172) vectors (Addgene  
291 plasmids #22011 and #22010, respectively). To generate the bait and prey plasmids for yeast  
292 two-hybrid assays, the coding DNA sequences for mitochondrial ECSIT, ECSIT N-terminus,  
293 ECSIT C-terminus, mature NDUFAF1, mature ACAD9 and full-length TRAF6 (residues 1-  
294 522) were inserted into both pEXP22 (prey) and pEXP32 (bait) vectors (ThermoFisher  
295 Scientific). All the constructs were obtained using restriction free cloning protocol (37).

### 296 **Expression of soluble proteins by random incremental truncation (ESPRIT) library** 297 **screening**

298 DNA fragments encoding ECSIT residues 1-245, 1-266, 1-431, 1-389, 1-401 and 1-431 were  
299 subcloned into pESPRIT002, downstream of an N-terminal TEV protease-cleavable His<sub>6</sub>-tag  
300 and upstream and in-frame with a C-terminal biotin acceptor peptide (BAP) used as a marker  
301 of protein solubility and stability. Full methods and materials for ESPRIT have been detailed  
302 elsewhere (31). A total of 14025 colonies (9405 for the N-terminal sub-library and 4620 for  
303 the C-terminal sub-library) were screened. Results are summarised in Figure S3.

### 304 **Recombinant protein expression and purification**

305 Mitochondrial ECSIT (residues 49-431), ECSIT<sub>CTD</sub> and NDUFAF1 were expressed in *E. coli*  
306 BL21 Star (DE3) cells (ThermoFisher Scientific) in 6 LB medium at 25 °C (for ECSIT  
307 proteins) and 16 °C (for NDUFAF1) and induced with isopropyl-β-D-galactopyranoside  
308 (IPTG) to a final concentration of 250 and 80 μM (for ECSIT proteins and NDUFAF1,  
309 respectively). ACAD9, VLCAD and chimeric ACAD9<sub>VLCAD</sub> were expressed in *E. coli*  
310 C43(DE3) cells (Lucigen) in 6 L Terrific Broth medium with 8 g/L glycerol at 37 °C and  
311 induced with 500 μM IPTG and harvested 12 hours after induction. Bacterial pellets were

312 resuspended in lysis buffer (200 mM potassium phosphate buffer or PBS, 1 mM DTT, 0.25  
313 mM EDTA, 0.2% Tween-20, pH 7.8) supplemented with protease inhibitor cocktails (Merck)  
314 and *DNase I* (Merck). Bacterial lysis was performed by sonication (QSonica), lysates were  
315 centrifuged for 30 min at 35,000 g (Avanti J-20 XP centrifuge, Beckman Coulter) and the  
316 supernatants were filtered (0.22  $\mu$ m pore size filter) before protein purification.

317 The single proteins were purified by a combination of affinity and size-exclusion  
318 chromatography (SEC) performed on Äkta Purification systems (GE Healthcare). Typically,  
319 0.5 L of the single protein soluble fractions was loaded onto a 5-mL HisTrap column (GE  
320 Healthcare) equilibrated in 200 mM PBS, 1 mM DTT, 0.5 mM EDTA, pH 7.8 and eluted with  
321 a linear imidazole gradient up to 500 mM imidazole in 200 mM PBS, 1 mM DTT, 0.5 mM  
322 EDTA, pH 7.8. Subsequently, the proteins were further purified by a combination of SEC  
323 columns. In particular, ECSIT<sub>CTD</sub> and NDUFAF1 were purified on a HiLoad Superdex 75 pg  
324 16/600 and on a Superdex 75 10/300 GL SEC columns (GE Healthcare) with 25 mM HEPES,  
325 250 mM NaCl, pH 7.5 as running buffer. Mitochondrial ECSIT, ACAD9 and VLCAD were  
326 purified on a HiLoad Superdex 200 pg 16/600 and on a Superdex 200 10/300 GL SEC columns  
327 (GE Healthcare) equilibrated with 25 mM HEPES, 250 mM NaCl, pH 7.5. Single protein  
328 identity was confirmed by electron spray ionization time-of-flight mass spectroscopy (ESI-  
329 TOF/MS). The MCIA subcomplex formed by ACAD9 and ECSIT<sub>CTD</sub> was obtained by co-  
330 purification. The single proteins were expressed as previously described and the bacterial  
331 pellets were co-lysed and centrifuged together. The soluble fraction was loaded overnight onto  
332 a 5-mL His-Trap column equilibrated in 200 mM PBS, 1 mM TCEP, 0.5 mM EDTA, pH 7.8  
333 and eluted with a linear imidazole gradient up to 500 mM imidazole in 200 mM PBS, 1 mM  
334 TCEP, 0.5 mM EDTA, pH 7.8. Subsequently, the proteins were concentrated using 100 kDa  
335 cut-off Amicon Ultra filters and purified by Sephacryl S300 HR 16/600 and Superose 6/10/300  
336 GL SEC columns (GE Healthcare) equilibrated with 25 mM HEPES, 250 mM NaCl, 1 mM  
337 TCEP, pH 7.5. An identical protocol was followed to attempt the co-purification of VLCAD-  
338 ECSIT<sub>CTD</sub> and ACAD9<sup>VLCAD</sup>-ECSIT<sub>CTD</sub> respectively. Single proteins and protein complexes  
339 were stored at 4 °C until use.

#### 340 **Size Exclusion Chromatography coupled to Multi-Angle Laser Light Scattering (SEC-** 341 **MALLS)**

342 The molecular mass of ECSIT<sub>CTD</sub>, NDUFAF1 and ACAD9 in solution was determinate by  
343 SEC coupled to multi-angle laser light scattering (SEC-MALLS) using a Superdex 200 10/300  
344 GL column equilibrated in buffer 25 mM HEPES, 250 mM NaCl, pH 7.8. The measurements  
345 were performed at 20 °C using 50  $\mu$ L of proteins at 5 mg/mL with a flow rate of 0.5 mL/min

346 and eluted proteins were monitored using a DAWN-EOS detector with a laser emitting at 690  
347 nm for online MALLS measurement (Wyatt Technology Corp.) and with a RI2000 detector  
348 for online refractive index measurements (SchambeckSFD). Molecular mass calculations were  
349 performed with the ASTRA software using a refractive index increment (dn/dc) of 0.185 mL/g.  
350 Data were visualized using OriginPro 9.0 (OriginLab) software.

### 351 **Detection of FAD content by UV-vis Spectroscopy**

352 The estimation of FAD content in ACAD9 and in the ACAD9-ECSIT<sub>CTD</sub> subcomplex was  
353 performed using UV-vis spectrophotometer (UV-2401PC, Shimadzu) at 16  $\mu$ M protein  
354 concentration in buffer 25 mM HEPES, 250 mM NaCl, pH 7.5. Absorbance spectra were  
355 recorded with 1 nm steps within a wavelength range from 300 nm to 550 nm using a 10 mm  
356 light path quartz cuvette. To test FAD quenching, decanoyl-CoA (Sigma) was added to the  
357 samples at different molar ratio (i.e. 1:1, 3:1, 10:1 and 20:1) and spectra were recorded within  
358 a wavelength range from 300 nm to 550 nm as before. All the UV-vis measurements were  
359 carried at 20 °C and in triplicate. Data were processed and visualized using OriginPro 9.0.

### 360 **Differential Scanning Calorimetry (DSC)**

361 The melting temperatures of ECSIT<sub>CTD</sub>, ACAD9, VLCAD and the ACAD9-ECSIT<sub>CTD</sub>  
362 subcomplex were determined by differential scanning calorimetry (DSC). Thermal  
363 denaturation experiments were carried out on a micro-DSC III microcalorimeter (Setaram  
364 Instrumentation). Protein solutions at 100  $\mu$ M concentration were degassed under vacuum at  
365 20 °C for calorimetric measurements. The DSC scans were run between 20 and 95 °C at a rate  
366 of 60 °C/hour. Reversibility of the unfolding transition was estimated by rescanning the sample  
367 from 95 to 20°C at the same rate. The denaturation temperature,  $T_m$ , corresponding to the  
368 maximum of the transition peak, was determined from 3 replicate runs and varied not more  
369 than 0.25 °C. The data were processed and analysed using the Thermal Analysis Software  
370 CALISTO (Setaram Instrumentation) and OriginPro 9.0 software.

### 371 **<sup>15</sup>N-Heteronuclear Single Quantum Coherence (HSQC) NMR Spectroscopy**

372 NMR experiments were performed on 0.5 mM <sup>15</sup>N-labelled ECSIT<sub>CTD</sub> and on 0.5 mM <sup>15</sup>N-  
373 ECSIT<sub>CTD</sub>-ACAD9 subcomplex expressed and purified as described above. For isotope  
374 labeling 1 g/L [<sup>15</sup>N] ammonium chloride was added to 2 L of minimal medium. BEST TROSY  
375 experiments (<sup>1</sup>H-<sup>15</sup>N correlation spectra) and DOSY experiments (for measuring the  
376 translational diffusion) were recorded at 300 K on Bruker ADVANCE III HD spectrometers  
377 operating at <sup>1</sup>H frequency of 700 MHz and equipped with a triple resonance pulsed field  
378 gradient cryoprobe (37). Data processing was conducted with either FELIX (Biosym  
379 Technologies) or NMRPipe (38).

### 380 **Small Angle X-ray Scattering (SAXS) data collection and analysis**

381 The experiments were performed at the ESRF BioSAXS beamline BM29, Grenoble, France  
382 (39). ECSIT<sub>CTD</sub> and NDUFAF1 were measured in batch mode at 20 °C immediately after the  
383 protein purification. For these measurements, 45 µL of sample solution at three different  
384 concentrations (5, 2.5 and 1 mg/mL) were used. ACAD9-ECSIT<sub>CTD</sub> subcomplex sample was  
385 measured in batch mode immediately after the SEC elution using 0.8, 1.2 and 2 mg/mL  
386 concentration dilutions. ACAD9 and VLCAD were measured by SEC-SAXS to separate larger  
387 protein aggregates. A volume of 250 µL of protein per each ACAD9 or VLCAD sample at 10  
388 mg/mL was loaded on a Superose 6 10/300 GL column via a high performance liquid  
389 chromatography device (HPLC, Shimadzu, France) attached directly to the sample-inlet valve  
390 of the BM29 sample changer (40). All the samples were measured in buffer 25 mM HEPES,  
391 250 mM NaCl, pH 7.5 at 20 °C. The column was equilibrated with 3 column volumes to obtain  
392 a stable background signal that was confirmed before measurement. A flow rate of 0.5 mL/min  
393 was used for all sample measurements. All parameters for SAXS analysis, sample details and  
394 results are described in **Table S1** according to recent recommended guidelines (41).

### 395 **Electrophoresis Mobility Assay - Native Polyacrylamide Gel Electrophoresis (native** 396 **PAGE)**

397 Purified proteins (ACAD9 and VLCAD at 16.8 µM) were incubated with increasing  
398 concentrations of ECSIT<sub>CTD</sub> (4.5 to 36 µM) for 30 min at room temperature in a buffer  
399 containing 25 mM HEPES pH 7.5, 50 mM NaCl, 1 mM TCEP. Similarly, ECSIT<sub>CTD</sub> at 12.5  
400 µM was titrated with increasing NDUFAF1 concentrations (3.4 to 27 µM). Native  
401 electrophoresis was then carried out using a Mini-PROTEAN® TGX Stain-Free™ Precast  
402 Gels (5% acrylamide) (Bio-Rad) in 1x TBE buffer. Migration of proteins was visualized by  
403 Coomassie Brilliant Blue (Sigma) staining.

### 404 **Liquid Chromatography/Electrospray Ionization Mass Spectrometry (LC/ESI-MS)**

405 LC/ESI-MS was performed on a 6210 TOF mass spectrometer coupled to a HPLC system  
406 (1100 series, Agilent Technologies). All solvents used were HPLC grade; the HPLC mobile  
407 phases were A (H<sub>2</sub>O 95%, ACN 5%, TFA 0.03%) and B (ACN 95%, H<sub>2</sub>O 5%, TFA 0.03%).  
408 Protein samples were desalted on-line on a C8 reversed-phase micro-column (Zorbax 300SB-  
409 C8, 5 µm, 5 x 0.3 mm, Agilent Technologies) for 3 min at 50 µL/min with 100% of mobile  
410 phase A, then eluted at 50 µL/min with 70% of mobile phase B. MS acquisition was carried  
411 out in the positive ion mode in the 300–3000 *m/z* range and the data processed with MassHunter  
412 software (v. B.02.00, Agilent Technologies). The mass spectrometer was calibrated with tuning  
413 mix (ESI-L, Agilent Technologies). The mass spectrometer settings were the following: gas

414 temperature (azote) 300 °C, drying gas (azote) 7 L/min, nebulizer gas (azote): 10 psig, Vcap:  
415 4 kV, fragmentor: 250 V, skimmer: 60 V, Vpp (octopole RF): 250 V.

#### 416 **Native Mass Spectrometry (native MS)**

417 The proteins were analysed by native MS in the concentration range of 3–5  $\mu$ M.  
418 Macromolecular complexes formed by ACAD9 and ECSIT<sub>CTD</sub> were analysed by native MS  
419 (42, 43). Protein buffer was freshly exchanged to 250 mM ammonium acetate. Protein ions  
420 were generated using a nanoflow electrospray (nano-ESI) source. Nanoflow platinum-coated  
421 borosilicate electrospray capillaries were bought from Thermo Electron SAS (Courtaboeuf,  
422 France). MS analyses were carried out on a quadrupole time-of-flight mass spectrometer (Q-  
423 TOF Ultima, Waters Corporation). The instrument was modified for the detection of high  
424 masses (44). The following instrumental parameters were used: capillary voltage = 1.2–1.3 kV,  
425 cone potential = 40 V, RF lens-1 potential = 40 V, RF lens-2 potential = 1 V, aperture-1  
426 potential = 0 V, collision energy = 30–140 V, and microchannel plate (MCP) = 1900 V. All  
427 mass spectra were calibrated externally using a solution of caesium iodide (6 mg/mL in 50%  
428 isopropanol) and were processed with the Masslynx 4.0 software (Waters Corporation,  
429 Manchester, U.K.) and with Massign software package (45).

#### 430 **Negative stain Electron Microscopy (ns-EM)**

431 ACAD9 protein sample at 0.05 mg/mL was applied to the clean side of carbon on mica  
432 (carbon/mica interface). The carbon layer was subsequently floated onto a 2% sodium  
433 silicotungstate solution (pH 7.4), recovered with a 400 mesh copper grid (Agar) and air dried  
434 for 10 min. Micrographs were taken under low-dose conditions (exposing for 1 s at an electron  
435 dose of  $30 \text{ e}^-/\text{\AA}^2$ ) on a Tecnai 12 LaB6 microscope (FEI) operated at 120 kV. Images were  
436 recorded at a nominal magnification of 22000 x with a defocus between 1.2 and 2.5  $\mu$ m on a  
437 GatanOrius 1000 CCD camera, corresponding to a pixel size of 3.24  $\text{\AA}$ /pixel on the specimen.  
438 CTF estimation was carried out using CTFFIND4 (46), and particles were selected using the  
439 LoG picking algorithm implemented in RELION-3 (47). 92,410 particles were extracted with  
440 a box size of 100 x 100 pixels and subjected to several rounds of 2D classification in RELION-  
441 3, resulting in a cleaned particle set of 79,924 particles. From these particles 39,795 particles  
442 from classes with isolated ACAD9 dimers (i.e. with no close neighbouring particles) were used  
443 to generate an initial model with applied C2 symmetry using the initial model generation  
444 algorithm in RELION-3. All 79,924 particles were then refined against this initial model with  
445 applied C2 symmetry, with a loose mask to remove the contribution from neighbouring  
446 ACAD9 dimers, resulting in the negative stain EM envelope shown in Figure S5.

447 **Cryo-Electron Microscopy (cryo-EM), data acquisition, processing and 3D**  
448 **reconstruction**

449 Grids of ACAD9-ECSIT<sub>CTD</sub> were frozen on Quantifoil R 2/1 Cu/Rh 300 mesh holey carbon  
450 grids, after grids were pumped under vacuum for 1 hour and glow discharged at 45 mA for 20  
451 s. Grids were frozen at 100 K with a Vitrobot Mark IV (FEI). 7,510 movies of 40 frames were  
452 collected on a Titan Krios microscope at the CM01 ESRF facility (48). The microscope is  
453 equipped with a K2 Summit camera (Gatan) using EPU (Thermo Fisher) with a total electron  
454 dose of 41 e<sup>-</sup>/Å<sup>2</sup>. Images were collected at a nominal magnification of 131,000 x which  
455 corresponds to a pixel size of 1.067 Å/pixel at the specimen level, with ~50% of images  
456 collected with a 30° tilt to counteract the strong preferential orientation observed while  
457 screening grids. Motion correction was carried out using MotionCor2 (49) and the CTF was  
458 corrected using GCTF (50). Micrographs with an estimated resolution in GCTF of better than  
459 8 Å were retained and manually screened by eye, resulting in 2,999 micrographs used for  
460 further processing. A subset of micrographs was first manually picked using the Boxer program  
461 in EMAN (51) resulting in 2,901 picked particles. Particles were extracted with a box size of  
462 160 x 160 pixels and subjected to 2D classification in RELION-2.1 (52). The best 2D classes  
463 were then used as templates for auto-picking in RELION, on a larger subset of the data. This  
464 resulted in ~490,000 picked particles, which were then extracted and subjected to several  
465 rounds of 2D classification. The best 2D class averages were then used as templates for GPU-  
466 accelerated picking in Gautomatch (<http://www.mrc-lmb.cam.ac.uk/kzhang/>). The coordinates  
467 of the ~700,000 picked particles were then used as input for per-particle CTF correction in  
468 GCTF. Particles were then imported into CryoSPARC (53) for further processing.

469 2D and 3D classification were extensively employed to find a stable subset of particles which  
470 yielded a map consistent with the secondary structural features seen in both 2D class averages  
471 and the VLCAD-based ACAD9 homology model. The contribution of ECSIT<sub>CTD</sub>, coupled with  
472 the strong preferential orientation of the complex, complicated the image analysis process.  
473 Many different trials were taken, and the strategy that yielded reconstructions which were most  
474 consistent with expected secondary structural features are described in the Supplementary  
475 Materials and in Figure S8.

476 **Mammalian cell culturing**

477 Human neuroglioma H4 cells, both wild-type (WT) and stably transfected with human amyloid  
478 precursor protein (APP) carrying the AD-related Swedish mutation (KM670/671NL) (54),  
479 were kindly provided by Dr. Patrick Aloy, Institute for Research in Biomedicine, Barcelona.  
480 The cell lines were maintained in Opti-MEM I reduced serum medium (GIBCO®)

481 supplemented with 10% fetal bovine serum (FBS, GIBCO®) and 100 U/mL penicillin–  
482 streptomycin (Invitrogen) at 37 °C and 5% CO<sub>2</sub> in a humidified incubator. Cells were  
483 trypsinized at 80% confluence after 72 hour incubation on T225 flasks (10<sup>5</sup> cells / flask).  
484 HEK293T cells were kindly provided by Dr. Pascal Fender, Institut de Biologie Structurale,  
485 Grenoble. Cells were cultured in Dulbecco's Modified Eagle Medium (DMEM, Sigma)  
486 supplemented with 10% FBS at 37 °C and 5% CO<sub>2</sub>.

#### 487 **Mitochondria isolation, CI immunopurification, CI activity assay and mitochondrial A $\beta$ <sub>1-42</sub>** 488 **detection**

489 About 25 x 10<sup>6</sup> cells were resuspended in homogenization buffer (10 mM Tris-HCl pH 6.7, 10  
490 mM KCl, 0.15 mM MgCl<sub>2</sub>, 1 mM PMSF, 1 mM DTT) and then transferred to a glass  
491 homogenizer and incubated for 10 min on ice. Cells were lysed using a tight-fitting pestle for  
492 approximately 2 min. The homogenized cellular extract was then gently mixed with 2 M  
493 Sucrose and centrifuged three times at 1,200 × g for 5 min to obtain the supernatant.  
494 Mitochondria were isolated by differential centrifugation steps according to published  
495 protocols (55). Complex I (CI) was immunopurified from H4 cell mitochondria using a  
496 commercial kit (Abcam ab109721) according to the manufacturer instructions. CI activity was  
497 analysed by measuring the absorbance at OD<sub>450 nm</sub> in a kinetic mode at room temperature for  
498 up to 16 hours in a microplate reader Quantamaster QM4CW (Horiba). The human Amyloid  
499 Beta (residues 1-42, A $\beta$ <sub>1-42</sub>) content in isolated mitochondria from WT and APP cells was  
500 measured by enzyme-linked immunosorbent assay (ELISA) according to the manufacturer  
501 instructions (ThermoFisher khb3544).

#### 502 **SDS-PAGE and Western blotting (WB)**

503 SDS-PAGE (10 or 12% acrylamide) gels were performed to verify the identity of the proteins.  
504 Gels were stained with Coomassie brilliant blue (Sigma). For mammalian cell protein  
505 detection, samples were loaded on Mini-PROTEAN® TGX Stain-Free™ Precast Gels (Biorad)  
506 or a NuPAGE 4–12% Bis-Tris gels (ThermoFisher Scientific). Protein samples were transferred  
507 electrophoretically to PVDF membranes (Biorad) using a semidry blotting system. Blots were  
508 incubated for four hours at 4 °C with primary antibodies: NDUFAF1 rabbit monoclonal  
509 (Abcam, ab79826); ECSIT rabbit polyclonal (Abcam, ab21288); ACAD9 mouse polyclonal  
510 (Abcam, ab72655); VDAC1/Porin mitochondrial loading control (Abcam, ab15895); and  
511 subsequently incubated overnight at 4 °C with secondary antibody conjugates: anti-rabbit  
512 (Abcam, ab150077) or anti-mouse (Abcam, ab150113) Alexa Fluor® 488, diluted 1:20,000.  
513 Migration of proteins was visualized using a Typhoon Trio™ fluoroimager (GE Healthcare  
514 LifeSciences), and gel images were processed using ImageJ software (56).

### 515 **Yeast two-hybrid (Y2H) interaction assays**

516 Bait and prey plasmids were pairwise co-transformed into MaV203 yeast strain (kindly  
517 provided by Dr. Ulrich Stelzl, University of Graz, Austria) as previously reported (28), plated  
518 onto selective SD2 (lacking Leu and Trp amino acids) agar media and incubated for 72 hours  
519 at 30°C to detect colony growth. Co-transformant arrays were then replica-cleaned before being  
520 plated onto different selective media for interaction screening. To assay the activation of the  
521 HIS3 reporter gene, SD3 (lacking Leu, Trp, His) agar plates were supplemented with 15 to 50  
522 mM of 3-aminotriazole (3AT, Sigma). Colony growth was inspected after 6-7 days of  
523 incubation.

### 524 **Bimolecular Fluorescence Complementation (BiFC) assays**

525 HEK293 cells were grown in Dulbecco's modified eagle medium containing 10% fetal bovine  
526 serum, 10,000 units/mL penicillin and 10,000 µg/mL streptomycin at 37 °C in a humidified  
527 atmosphere containing 5% CO<sub>2</sub>. The day before transfection, HEK293 cells were plated on 12-  
528 well plates with OPTI-MEM medium (ThermoFisher Scientific). The cells were co-transfected  
529 with 0.2 µg of BiFC probes and the actin filament marker LifeAct-RFP using Lipofectamine  
530 transfection agent (ThermoFisher Scientific). LifeAct-RFP is a cell marker through the  
531 labelling of actin with Red Fluorescent Protein (RFP). The cellular intensities of BiFC  
532 fluorescence were analysed 48h after transfection using the FITC (green spectrum) and  
533 estimated as the ratio of FITC vs colocalised CY3.5 (red spectrum). Each experiment was  
534 performed as triplicate.

### 535 **Limited proteolysis**

536 Limited proteolysis of proteins was carried out by adding trypsin proteolytic enzyme (Sigma).  
537 ACAD9, VLCAD and ACAD9<sup>VLCAD</sup> at 1 mg/mL, respectively, were incubated with trypsin at  
538 1:10000 ratio in absence or presence of ECSIT<sub>CTD</sub> (5 mg/mL) at room temperature for 30  
539 minutes. The incubated mixtures were then mixed with SDS gel loading dye, boiled prior to  
540 their resolution by a SDS-15% PAGE and stained with Coomassie brilliant blue (Sigma).

### 541 **Statistical analysis**

542 For mitochondrial assays, every sample of three independent experiments was presented as a  
543 mean average with the standard deviation. Noted differences amongst the various sample types  
544 within H4 WT and APP cell types; cytosol vs mitochondria subcellular types; and mitochondria  
545 vs CI binding were assessed for significance. Statistically significant differences were  
546 determined by one-way ANOVA followed by Tukey-Kramer post-test to identify pair wise  
547 differences. Differences were considered significant at  $P < 0.05$  (\*),  $P < 0.01$  (\*\*). Statistical  
548 analyses were carried out using GraphPad Prism version 8 (GraphPad Software).

## 549 **Supplementary Materials**

550 Supplementary Materials and Methods

551 Fig. S1. Quantitative analysis of MCIA protein expression levels in individual replicas.

552 Fig. S2. Recombinant, purified mitochondrial ECSIT forms soluble aggregates.

553 Fig. S3. Identification soluble fragments of ECSIT by the ESPRIT approach.

554 Fig. S4. SEC-MALLS characterisation of recombinant ECSIT<sub>CTD</sub>, NDUFAF1 and ACAD9.

555 Fig. S5. ECSIT<sub>CTD</sub> is highly disordered and forms predominantly dimers.

556 Fig. S6. Conformational analysis of the ACAD9 homodimer.

557 Fig. S7. In vitro interaction assays between ACAD9, ECSIT<sub>CTD</sub> and NDUFAF1.

558 Fig. S8. Cryo-EM processing strategy.

559 Fig. S9. Investigation of different complex assemblies by Native MS.

560 Fig. S10. Structural features of the ACAD9 homodimer and potential interaction sites.

561 Fig. S11. Construction of a chimeric ACAD9 and its dehydrogenase activity.

562 Fig. S12. In vitro interaction assays between VLCAD, chimeric ACAD9<sub>VLCAD</sub> and  
563 ECSIT<sub>CTD</sub>.

564 Fig. S13. The ability of ACAD9 homodimer to bind FAD assessed by UV and MS.

565 Table S1. SAXS data of the proteins and protein fragments of this study.

566 **References**

- 567 1. A. Nsiah-Sefaa, M. McKenzie, Combined defects in oxidative phosphorylation and fatty acid  
568 beta-oxidation in mitochondrial disease. *Biosci Rep* **36**, (2016).
- 569 2. L. E. Formosa, M. G. Dibley, D. A. Stroud, M. T. Ryan, Building a complex complex:  
570 Assembly of mitochondrial respiratory chain complex I. *Semin Cell Dev Biol* **76**, 154-162  
571 (2018).
- 572 3. J. A. Letts, L. A. Sazanov, Clarifying the supercomplex: the higher-order organization of the  
573 mitochondrial electron transport chain. *Nat Struct Mol Biol* **24**, 800-808 (2017).
- 574 4. G. Yellen, Fueling thought: Management of glycolysis and oxidative phosphorylation in  
575 neuronal metabolism. *J Cell Biol* **217**, 2235-2246 (2018).
- 576 5. P. H. Reddy *et al.*, Amyloid-beta and mitochondria in aging and Alzheimer's disease:  
577 implications for synaptic damage and cognitive decline. *J Alzheimers Dis* **20 Suppl 2**, S499-  
578 512 (2010).
- 579 6. I. Lopez-Fabuel *et al.*, Complex I assembly into supercomplexes determines differential  
580 mitochondrial ROS production in neurons and astrocytes. *Proc Natl Acad Sci U S A* **113**, 13063-  
581 13068 (2016).
- 582 7. D. Ghezzi, M. Zeviani, Assembly factors of human mitochondrial respiratory chain complexes:  
583 physiology and pathophysiology. *Adv Exp Med Biol* **748**, 65-106 (2012).
- 584 8. O. V. Poole, M. G. Hanna, R. D. Pitceathly, Mitochondrial disorders: disease mechanisms and  
585 therapeutic approaches. *Discov Med* **20**, 325-331 (2015).
- 586 9. K. R. Vinothkumar, J. Zhu, J. Hirst, Architecture of mammalian respiratory complex I. *Nature*  
587 **515**, 80-84 (2014).
- 588 10. K. Fiedorczuk *et al.*, Atomic structure of the entire mammalian mitochondrial complex I.  
589 *Nature* **538**, 406-410 (2016).
- 590 11. D. A. Stroud *et al.*, Accessory subunits are integral for assembly and function of human  
591 mitochondrial complex I. *Nature* **538**, 123-126 (2016).
- 592 12. E. Fernandez-Vizarra, V. Tiranti, M. Zeviani, Assembly of the oxidative phosphorylation  
593 system in humans: what we have learned by studying its defects. *Biochim Biophys Acta* **1793**,  
594 200-211 (2009).
- 595 13. L. Sanchez-Caballero, S. Guerrero-Castillo, L. Nijtmans, Unraveling the complexity of  
596 mitochondrial complex I assembly: A dynamic process. *Biochim Biophys Acta* **1857**, 980-990  
597 (2016).
- 598 14. R. O. Vogel *et al.*, Cytosolic signaling protein Ecsit also localizes to mitochondria where it  
599 interacts with chaperone NDUFAF1 and functions in complex I assembly. *Genes Dev* **21**, 615-  
600 624 (2007).
- 601 15. J. Nouws, H. Te Brinke, L. G. Nijtmans, S. M. Houten, ACAD9, a complex I assembly factor  
602 with a moonlighting function in fatty acid oxidation deficiencies. *Hum Mol Genet* **23**, 1311-  
603 1319 (2014).
- 604 16. H. Heide *et al.*, Complexome profiling identifies TMEM126B as a component of the  
605 mitochondrial complex I assembly complex. *Cell Metab* **16**, 538-549 (2012).
- 606 17. J. Nouws *et al.*, Acyl-CoA dehydrogenase 9 is required for the biogenesis of oxidative  
607 phosphorylation complex I. *Cell Metab* **12**, 283-294 (2010).
- 608 18. B. Andrews, J. Carroll, S. Ding, I. M. Fearnley, J. E. Walker, Assembly factors for the  
609 membrane arm of human complex I. *Proc Natl Acad Sci U S A* **110**, 18934-18939 (2013).
- 610 19. F. R. G. Carneiro, A. Lepelley, J. J. Seeley, M. S. Hayden, S. Ghosh, An Essential Role for  
611 ECSIT in Mitochondrial Complex I Assembly and Mitophagy in Macrophages. *Cell Rep* **22**,  
612 2654-2666 (2018).

- 613 20. G. Giachin, R. Bouverot, S. Acajjaoui, S. Pantalone, M. Soler-Lopez, Dynamics of Human  
614 Mitochondrial Complex I Assembly: Implications for Neurodegenerative Diseases. *Front Mol*  
615 *Biosci* **3**, 43 (2016).
- 616 21. D. C. Fuhrmann *et al.*, Degradation of the mitochondrial complex I assembly factor  
617 TMEM126B under chronic hypoxia. *Cell Mol Life Sci* **75**, 3051-3067 (2018).
- 618 22. E. Kopp *et al.*, ECSIT is an evolutionarily conserved intermediate in the Toll/IL-1 signal  
619 transduction pathway. *Genes Dev* **13**, 2059-2071 (1999).
- 620 23. C. Xiao *et al.*, Ecsit is required for Bmp signaling and mesoderm formation during mouse  
621 embryogenesis. *Genes Dev* **17**, 2933-2949 (2003).
- 622 24. J. Zhang *et al.*, Cloning and functional characterization of ACAD-9, a novel member of human  
623 acyl-CoA dehydrogenase family. *Biochem Biophys Res Commun* **297**, 1033-1042 (2002).
- 624 25. T. Wang, C. Frangou, J. Zhang, Fatty acid oxidation (FAO) metabolic switch: metastasis in  
625 lymph nodes driven by yes-associated protein (YAP) activation. *Biotarget* **3**, (2019).
- 626 26. Y. Joh, W. S. Choi, Mitochondrial Complex I Inhibition Accelerates Amyloid Toxicity. *Dev*  
627 *Reprod* **21**, 417-424 (2017).
- 628 27. M. Soler-Lopez, N. Badiola, A. Zanzoni, P. Aloy, Towards Alzheimer's root cause: ECSIT as  
629 an integrating hub between oxidative stress, inflammation and mitochondrial dysfunction.  
630 Hypothetical role of the adapter protein ECSIT in familial and sporadic Alzheimer's disease  
631 pathogenesis. *Bioessays* **34**, 532-541 (2012).
- 632 28. M. Soler-Lopez, A. Zanzoni, R. Lluís, U. Stelzl, P. Aloy, Interactome mapping suggests new  
633 mechanistic details underlying Alzheimer's disease. *Genome Res* **21**, 364-376 (2011).
- 634 29. C. M. Pinho, P. F. Teixeira, E. Glaser, Mitochondrial import and degradation of amyloid-beta  
635 peptide. *Biochim Biophys Acta* **1837**, 1069-1074 (2014).
- 636 30. R. N. Kalara *et al.*, Abundance of the longer A beta 42 in neocortical and cerebrovascular  
637 amyloid beta deposits in Swedish familial Alzheimer's disease and Down's syndrome.  
638 *Neuroreport* **7**, 1377-1381 (1996).
- 639 31. P. J. Mas, D. J. Hart, ESPRIT: A Method for Defining Soluble Expression Constructs in Poorly  
640 Understood Gene Sequences. *Methods Mol Biol* **1586**, 45-63 (2017).
- 641 32. J. J. Kim, R. Miura, Acyl-CoA dehydrogenases and acyl-CoA oxidases. Structural basis for  
642 mechanistic similarities and differences. *Eur J Biochem* **271**, 483-493 (2004).
- 643 33. R. P. McAndrew *et al.*, Structural basis for substrate fatty acyl chain specificity: crystal  
644 structure of human very-long-chain acyl-CoA dehydrogenase. *J Biol Chem* **283**, 9435-9443  
645 (2008).
- 646 34. R. Ensenauer *et al.*, Human acyl-CoA dehydrogenase-9 plays a novel role in the mitochondrial  
647 beta-oxidation of unsaturated fatty acids. *J Biol Chem* **280**, 32309-32316 (2005).
- 648 35. M. Schiff *et al.*, Complex I assembly function and fatty acid oxidation enzyme activity of  
649 ACAD9 both contribute to disease severity in ACAD9 deficiency. *Hum Mol Genet* **24**, 3238-  
650 3247 (2015).
- 651 36. B. J. Floyd *et al.*, Mitochondrial Protein Interaction Mapping Identifies Regulators of  
652 Respiratory Chain Function. *Mol Cell* **63**, 621-632 (2016).
- 653 37. F. van den Ent, J. Lowe, RF cloning: a restriction-free method for inserting target genes into  
654 plasmids. *J Biochem Biophys Methods* **67**, 67-74 (2006).
- 655 38. F. Delaglio *et al.*, NMRPipe: A multidimensional spectral processing system based on UNIX  
656 pipes. *Journal of Biomolecular NMR* **6**, 277-293 (1995).
- 657 39. P. Pernot *et al.*, Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. *J*  
658 *Synchrotron Radiat* **20**, 660-664 (2013).
- 659 40. M. E. Brennich, A. R. Round, S. Hutin, Online Size-exclusion and Ion-exchange  
660 Chromatography on a SAXS Beamline. *J Vis Exp*, (2017).

- 661 41. J. Trewhella *et al.*, 2017 publication guidelines for structural modelling of small-angle  
662 scattering data from biomolecules in solution: an update. *Acta Crystallogr D Struct Biol* **73**,  
663 710-728 (2017).
- 664 42. E. Boeri Erba, C. Petosa, The emerging role of native mass spectrometry in characterizing the  
665 structure and dynamics of macromolecular complexes. *Protein Sci* **24**, 1176-1192 (2015).
- 666 43. E. Boeri Erba, L. Signor, M. F. Oliva, F. Hans, C. Petosa, Characterizing Intact Macromolecular  
667 Complexes Using Native Mass Spectrometry. *Methods Mol Biol* **1764**, 133-151 (2018).
- 668 44. F. Sobott, H. Hernandez, M. G. McCammon, M. A. Tito, C. V. Robinson, A tandem mass  
669 spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal*  
670 *Chem* **74**, 1402-1407 (2002).
- 671 45. N. Morgner, C. V. Robinson, Massign: an assignment strategy for maximizing information  
672 from the mass spectra of heterogeneous protein assemblies. *Anal Chem* **84**, 2939-2948 (2012).
- 673 46. A. Rohou, N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron  
674 micrographs. *J Struct Biol* **192**, 216-221 (2015).
- 675 47. J. Zivanov *et al.*, New tools for automated high-resolution cryo-EM structure determination in  
676 RELION-3. *Elife* **7**, (2018).
- 677 48. E. Kandiah *et al.*, CM01: a facility for cryo-electron microscopy at the European Synchrotron.  
678 *Acta Crystallogr D Struct Biol* **75**, 528-535 (2019).
- 679 49. H. Peng *et al.*, Automatic tracing of ultra-volumes of neuronal images. *Nat Methods* **14**, 332-  
680 333 (2017).
- 681 50. K. Zhang, Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**, 1-12 (2016).
- 682 51. S. J. Ludtke, P. R. Baldwin, W. Chiu, EMAN: semiautomated software for high-resolution  
683 single-particle reconstructions. *J Struct Biol* **128**, 82-97 (1999).
- 684 52. D. Kimanius, B. O. Forsberg, S. H. Scheres, E. Lindahl, Accelerated cryo-EM structure  
685 determination with parallelisation using GPUs in RELION-2. *Elife* **5**, (2016).
- 686 53. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: algorithms for rapid  
687 unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296 (2017).
- 688 54. M. Mullan *et al.*, A pathogenic mutation for probable Alzheimer's disease in the APP gene at  
689 the N-terminus of beta-amyloid. *Nat Genet* **1**, 345-347 (1992).
- 690 55. M. Spinazzi, A. Casarin, V. Pertegato, L. Salviati, C. Angelini, Assessment of mitochondrial  
691 respiratory chain enzymatic activities on tissues and cultured cells. *Nat Protoc* **7**, 1235-1246  
692 (2012).
- 693 56. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image  
694 analysis. *Nat Methods* **9**, 671-675 (2012).
- 695 57. Y. Min, S. M. Wi, D. Shin, E. Chun, K. Y. Lee, Peroxiredoxin-6 Negatively Regulates  
696 Bactericidal Activity and NF-kappaB Activity by Interrupting TRAF6-ECSIT Complex. *Front*  
697 *Cell Infect Microbiol* **7**, 94 (2017).
- 698 58. A. Favier, B. Brutscher, Recovering lost magnetization: polarization enhancement in  
699 biomolecular NMR. *J Biomol NMR* **49**, 9-15 (2011).
- 700 59. D. Franke *et al.*, ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering  
701 from macromolecular solutions. *Journal of Applied Crystallography* **50**, 1212-1225 (2017).
- 702 60. R. P. Rambo, J. A. Tainer, Accurate assessment of mass, models and resolution by small-angle  
703 scattering. *Nature* **496**, 477-481 (2013).
- 704 61. D. I. Svergun, M. H. Koch, Advances in structure analysis using small-angle scattering in  
705 solution. *Curr Opin Struct Biol* **12**, 654-660 (2002).
- 706 62. J. Trewhella *et al.*, 2017 publication guidelines for structural modelling of small-angle  
707 scattering data from biomolecules in solution: an update. *Acta Crystallogr D Struct Biol* **73**,  
708 710-728 (2017).

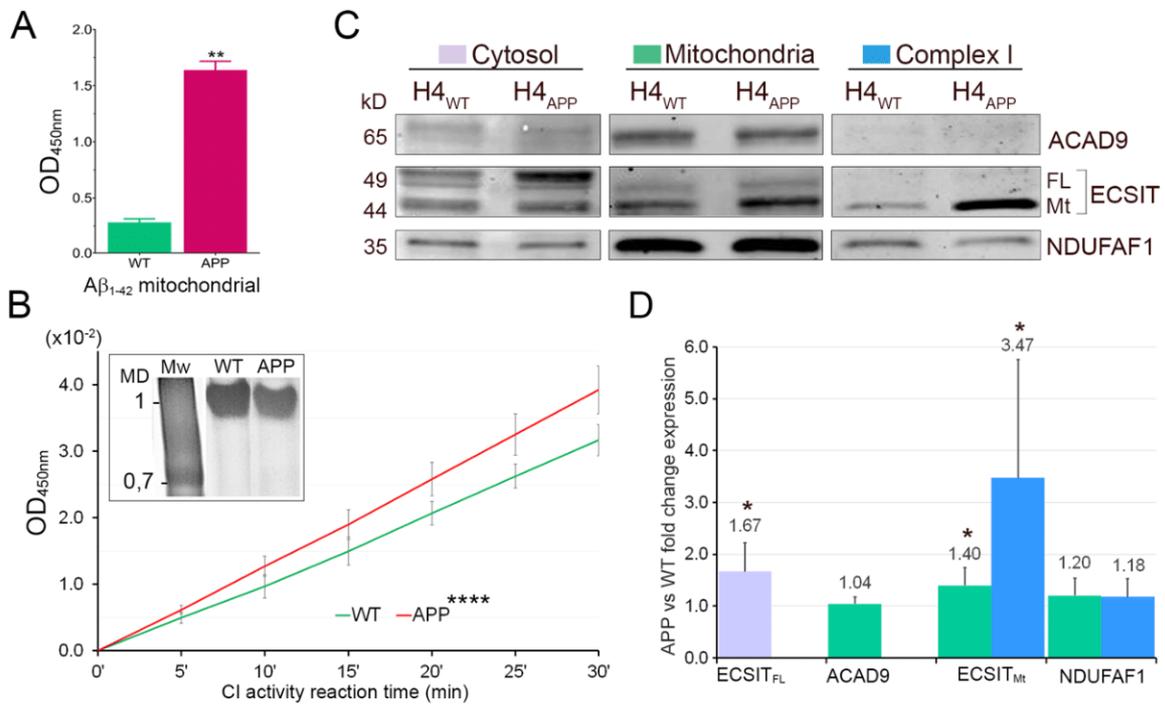
- 709 63. G. Tria, H. D. Mertens, M. Kachala, D. I. Svergun, Advanced ensemble modelling of flexible  
710 macromolecules using X-ray solution scattering. *IUCrJ* **2**, 207-217 (2015).
- 711 64. P. V. Konarev, M. V. Petoukhov, D. I. Svergun, Rapid automated superposition of shapes and  
712 macromolecular models using spherical harmonics. *J Appl Crystallogr* **49**, 953-960 (2016).
- 713 65. E. F. Pettersen *et al.*, UCSF Chimera--a visualization system for exploratory research and  
714 analysis. *J Comput Chem* **25**, 1605-1612 (2004).
- 715 66. R. P. McAndrew *et al.*, Structural basis for substrate fatty acyl chain specificity: crystal  
716 structure of human very-long-chain acyl-CoA dehydrogenase. *J Biol Chem* **283**, 9435-9443  
717 (2008).
- 718 67. M. Souri, T. Aoyama, G. Hoganson, T. Hashimoto, Very-long-chain acyl-CoA dehydrogenase  
719 subunit assembles to the dimer form on mitochondrial inner membrane. *FEBS Lett* **426**, 187-  
720 190 (1998).
- 721
- 722

723 **Acknowledgments:** We thank Prof. S. Zucchelli (University of Piemonte Orientale) for the  
724 TRAF6 plasmid, Dr. U. Stelzl (University of Graz) for the MaV203 yeast strain and Dr. P.  
725 Aloy (IRB Barcelona) and Dr. P. Fender (IBSG Grenoble) for H4 and HEK cell lines  
726 respectively. We are grateful to Dr. C. Mas (IBSG Grenoble) for support with SEC-MALLS  
727 and to Dr. M. Brennich (BM29-BioSAXS beamline, EMBL Grenoble) for support with SAXS  
728 analyses. We thank Dr. C. Petosa and Dr. A. Desfosses (IBS Grenoble) for discussions. We  
729 thank Dr. G. Schoehn for providing training and support, and D. Fenel for preliminary ns-EM  
730 observations. **Funding:** This work used the platforms of Partnership for Soft Condensed Matter  
731 (PSCM) with support from ESRF and ILL; and the platforms of the Grenoble Instruct Center  
732 (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-  
733 02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology  
734 (PSB). The electron microscope facility (Polaris electron microscope) is supported by the  
735 Rhône-Alpes Region (CIBLE and FEDER), the FRM, the CNRS, the University of Grenoble  
736 and the GIS-IBISA. We acknowledge the ESRF for provision of beam time on CM01. The  
737 ESRF in-house Research Program supported this work. The EM work was funded by the  
738 European Union's Horizon 2020 research and innovation programme under grant agreement  
739 No. 647784 to IG. MJ was supported by the CEA PhD Program. RB acknowledges funding  
740 support from the ESRF PhD Program. ESPRIT work has been supported by Instruct-ERIC PID  
741 1710. **Author contributions:** G.G. and M.S-L. conceived the experimental strategy; M.S-L.  
742 supervised the overall project in collaboration with I.G. for the EM analysis; G.G. and M.S-L.  
743 designed all the plasmidic constructs. G.G., R.B. and S.A. performed all biochemical and  
744 spectroscopy studies with contribution of M.S. and A.C. in sample preparation; S.A. and M.S-  
745 L. did the yeast and cell biology assays; E.B-M carried out the BiFC experiments; A.B. was  
746 responsible of the BiFC experiments and image analysis; G.G. and R.B. did the SAXS  
747 experiments; G.G. carried out the SAXS analyses and model reconstructions; L.S. did the  
748 LC/ESI experiments and analysis; A.F. and G.G. did the NMR experiments and analysis;  
749 P.J.M. did the ESPRIT experiments; D.J.H. was responsible of the ESPRIT experiments and  
750 analysis; E. B-E did the native MS experiments and analysis; G.G., M.J. and I.G. performed  
751 ns-EM sample preparation. M.J. and I.G. did ns-EM image collection and analysis; M.J., I.G.  
752 and M.B.-V. did cryo-EM grid preparation; M.J., I.G., M.H. and E.K. carried out cryo-EM data  
753 acquisition; M.J. and I.G. performed the image analysis and 3D reconstruction; G.L. was  
754 involved in all data analysis; G.G. and M.S.L. drafted the manuscript with major contributions  
755 from G.L. and I.G. All the Authors revised the final manuscript. **Competing interests:** Authors  
756 declare no competing interests. **Data and materials availability:** All data are available in the  
757 manuscript or supplementary materials. SAXS data were deposited into SASBDB data bank  
758 (IDs: SASDHU4; SASDHV4; SASDHW4; SASDHX4; SASDHY4 and SASDHZ4). The  
759 cryo-EM maps have been deposited to the EMDB with the following accession codes: XXXX.

760 **Figures**

761

762



763

764

765 **Fig. 1. Expression of MCIA factors in human neuronal mitochondria and under A $\beta$**

766 **oligomeric conditions. (A)** A $\beta$ <sub>1-42</sub> detection in mitochondria isolated from WT and APP cells

767 by ELISA immunoassays (n=4 $\pm$  SD, \*\*P < 0.001). H4 neuroglioma cells overexpress the

768 human amyloid precursor protein (APP) carrying the Alzheimer's-related Swedish mutation

769 (KM670/671NL). **(B)** NADH-dehydrogenase activity assay of immunopurified CI in WT and

770 APP cells (n=3 $\pm$  SD, \*\*\*\*P < 0.0001). In the *inset*, native gel showing purified fully assembled

771 CI (1 MDa) from both cell types and subjected to the activity assays. **(C)** Western blot analysis

772 showing the expression levels in cytosol, mitochondria and immunopurified CI of full-length

773 (FL, residues 1-431, 49 kDa) and mitochondrial (Mt, residues 49-431, 44 kDa) ECSIT, and

774 mature ACAD9 (65 kDa) and NDUFAF1 (35 kDa) respectively. **(D)** Quantitative densitometry

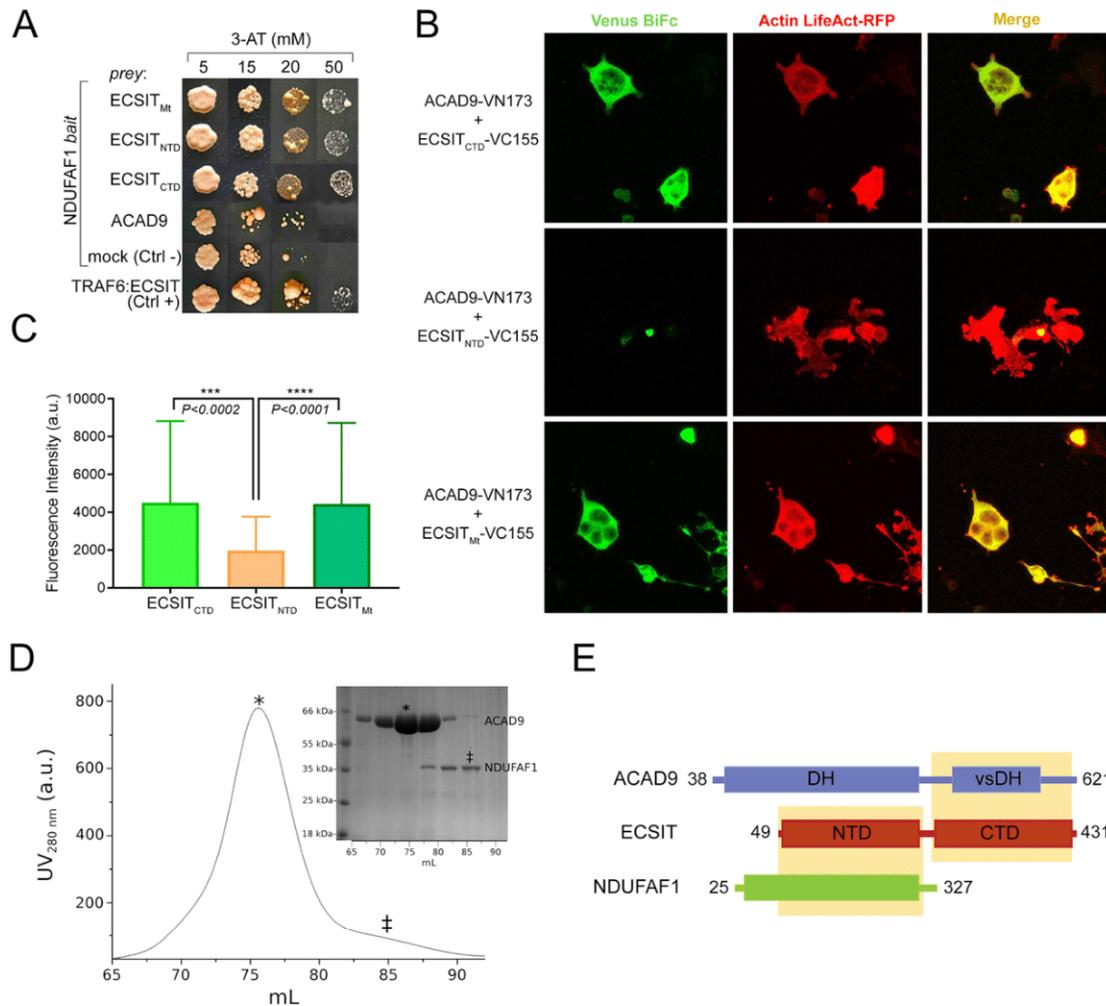
775 analysis of the average APP vs WT fold-change expression levels of ACAD9, NDUFAF1 and

776 ECSIT, based on five independent experiments (see **Fig. S1**). Significant change levels in APP

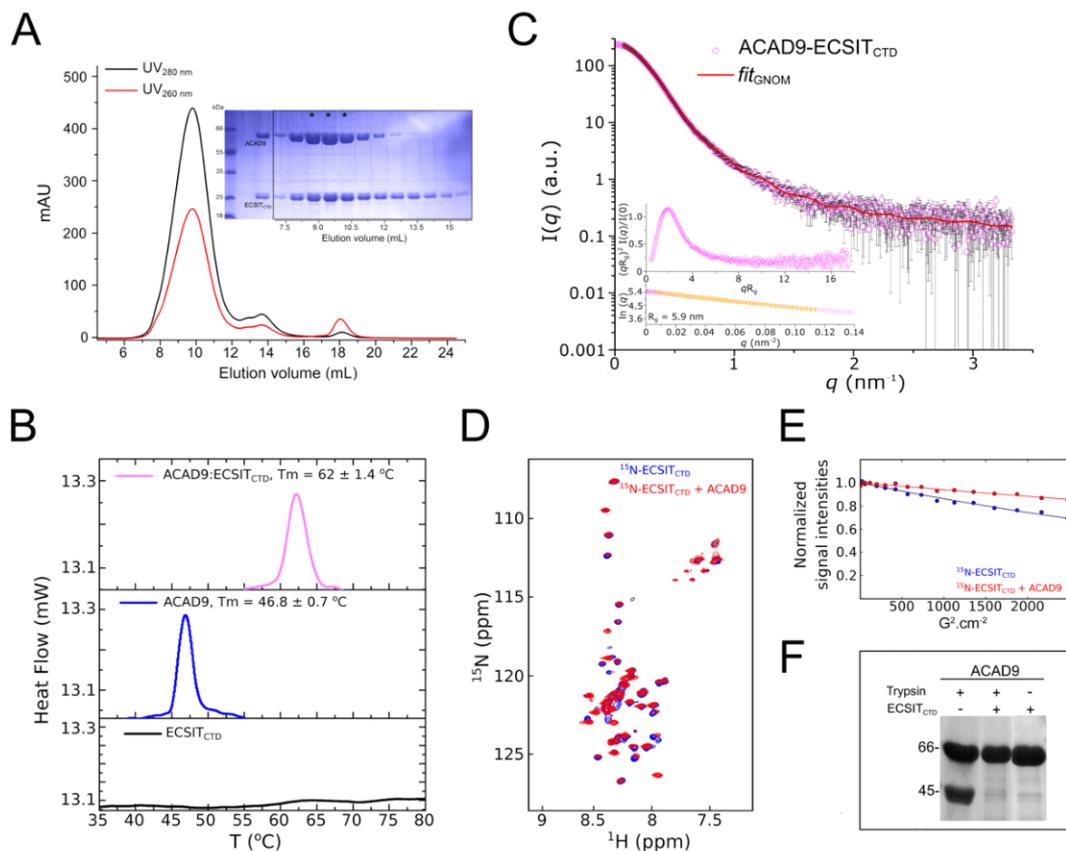
777 conditions are highlighted with \* (P < 0.03).

778

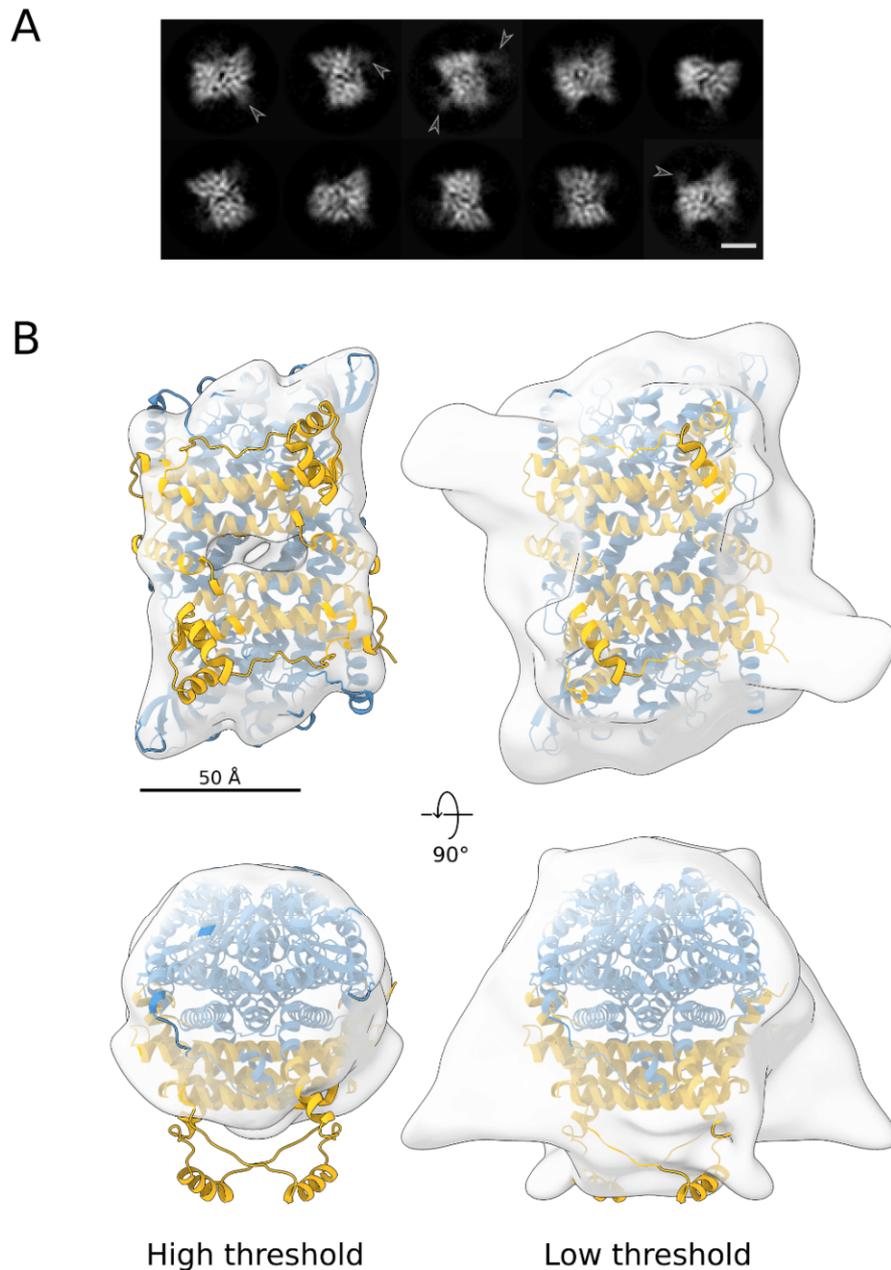
779



777  
778 **Fig. 2. MCIA protein interaction mapping.** (A) Y2H assays (NDUFAF1 as bait, different  
779 ECSIT and ACAD9 constructs as prey) indicate direct interactions between NDUFAF1 and  
780 mitochondrial (*Mt*, residues 49-431), N-terminal (*NTD*, residues 49-252) or C-terminal (*CTD*,  
781 residues 246-431) ECSIT. No interaction is detected between NDUFAF1 and ACAD9. A  
782 previously reported interaction (57) between ECSIT and TRFA6 was used as a positive control.  
783 (B) BiFC assays in HEK293 human cells. Active reassembly of Venus (green signal) indicates  
784 that mitochondrial ECSIT (VC155-ECSIT<sub>Mt</sub>) and the C-terminus (VC155-ECSIT<sub>CTD</sub>) interact  
785 with ACAD9 (ACAD9-VN173). Red immunofluorescence from filamentous actin (LifeAct-  
786 RFP) was used as a cell marker. Overlapping signals (yellow) confirm cells with positive  
787 interacting partners. (C) Quantification of the signal using the FITC (green spectrum) or CY3.5  
788 (red spectrum) filter settings.  $P < 0.002$  (\*\*\*),  $P < 0.0001$  (\*\*\*\*). (D) SEC chromatogram  
789 showing that ACAD9 (\*) elutes as single peak separated from NDUFAF1 (‡), which confirms  
790 the absence of interaction between ACAD9 and NDUFAF1, as observed *in vivo*. (E) Schematic  
791 of the MCIA complex with its potential interaction interfaces highlighted in yellow.  
792

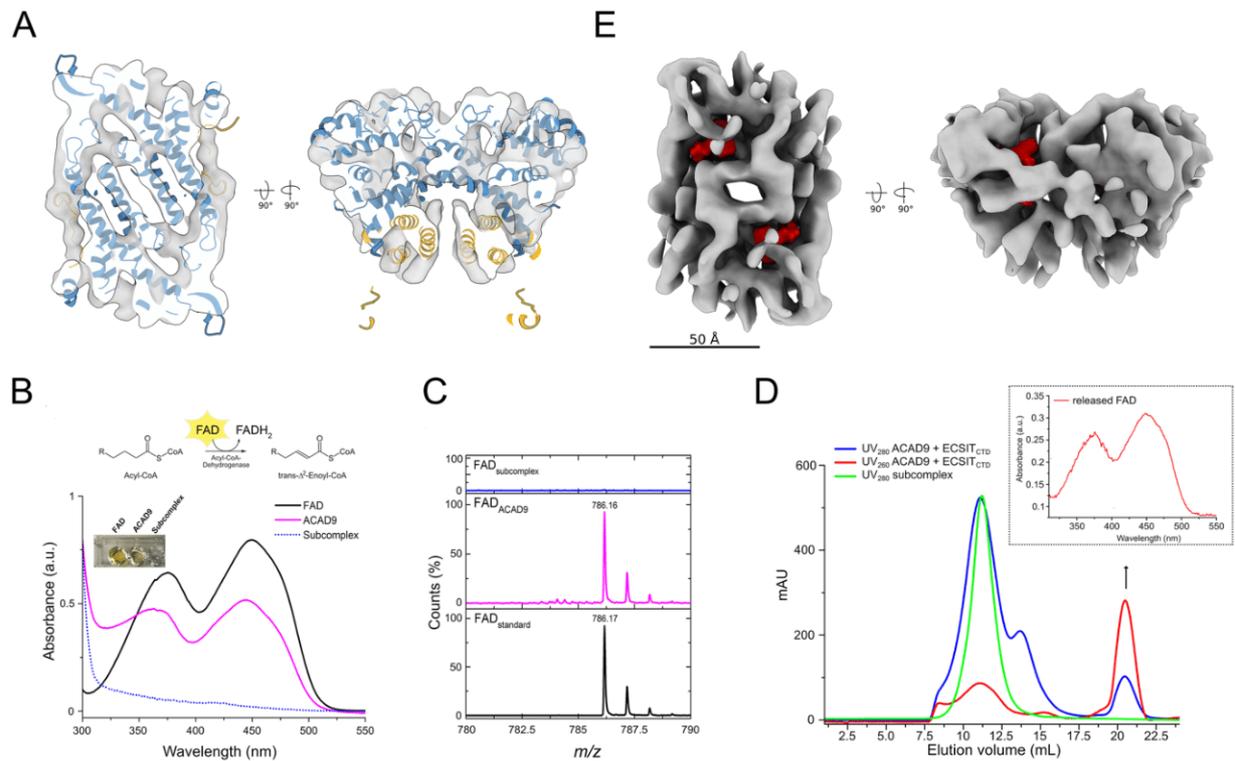


793  
794 **Fig. 3. The ACAD9-ECSIT<sub>CTD</sub> subcomplex.** (A) Chromatographic elution profile of the  
795 subcomplex (\*), showing a small adjacent peak of unbound ECSIT<sub>CTD</sub>. (B) Thermal  
796 denaturation of ECSIT<sub>CTD</sub>, ACAD9 and the subcomplex with the estimated  $T_m$  values. No  
797 unfolding signals were detected for ECSIT<sub>CTD</sub>. (C) SAXS scattering curve for the subcomplex,  
798 with the GNOM fitting curve in red. In the *inset*, normalized Kratky plot and close-up of the  
799 Guinier region. (D) 2D-NMR interaction experiments of <sup>15</sup>N-labelled ECSIT<sub>CTD</sub> and unlabeled  
800 ACAD9. Superimposed <sup>1</sup>H-<sup>15</sup>N correlation spectra of free ECSIT<sub>CTD</sub> and ECSIT<sub>CTD</sub> in complex  
801 with ACAD9 are shown in blue and red, respectively, in the region 8.5-7.5 <sup>1</sup>H ppm in the proton  
802 dimension. This highlights chemical shift perturbations in 11 out of 60 detected peaks in the  
803 amide region, indicating that ACAD9 interacts specifically with a few residues of ECSIT<sub>CTD</sub>.  
804 (E) <sup>15</sup>N-filtered DOSY-NMR measurements on <sup>15</sup>N-ECSIT<sub>CTD</sub> and ACAD9-<sup>15</sup>N-ECSIT<sub>CTD</sub>  
805 complex. The exponential decay curves of ECSIT<sub>CTD</sub>, in the absence and in the presence of  
806 ACAD9 are shown in blue and red respectively. The units on the y-axis are normalized values  
807 of the integrals of the signal measured in the amide proton region. The slower translational  
808 diffusion coefficient of ECSIT<sub>CTD</sub> measured in the presence of ACAD9 is also consistent with  
809 ECSIT<sub>CTD</sub> being part of an object larger than a single dimer (i.e. > 40 kDa in size) (58). (F)  
810 Analysis of ACAD9 stability in presence of ECSIT<sub>CTD</sub>. ACAD9 subjected to trypsin digestion  
811 in absence of ECSIT<sub>CTD</sub> shows a 45 kDa proteolytic fragment. In contrast, ACAD9 is protected  
812 from proteolysis in presence of ECSIT<sub>CTD</sub>, further confirming the formation of a stable  
813 complex between ACAD9 and ECSIT<sub>CTD</sub>. All proteolytic fragments are resolved by SDS-  
814 PAGE.



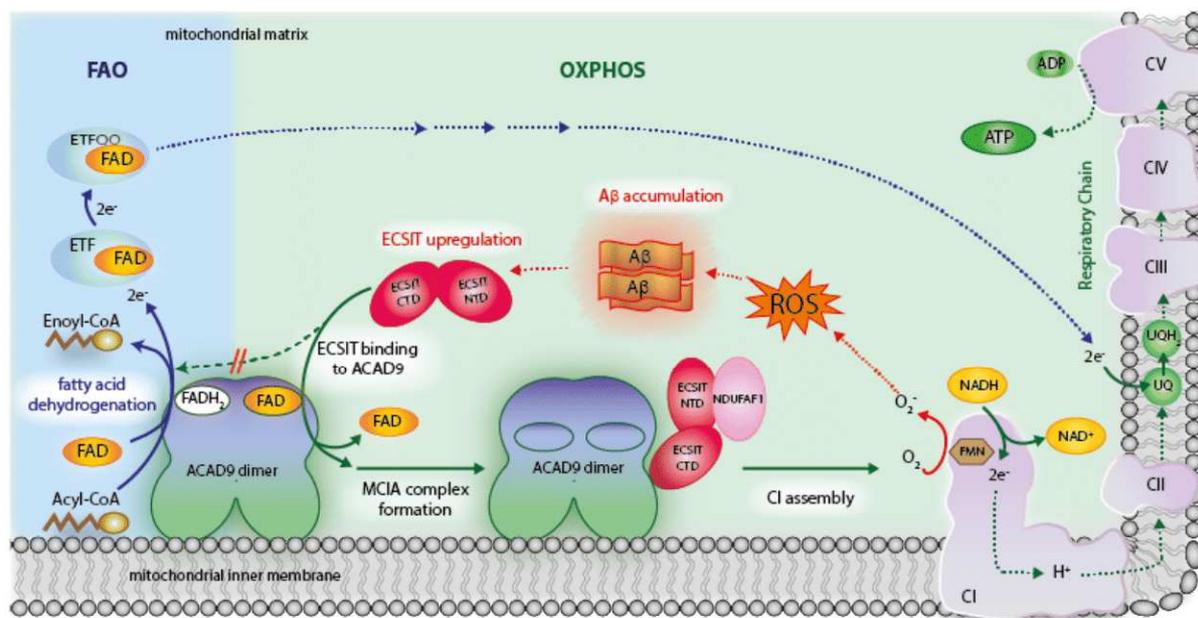
815

816 **Fig. 4. Cryo-EM analysis of ECSIT<sub>CTD</sub>-ACAD9 complex.** (A) Selected 2D class averages of  
 817 the ECSIT<sub>CTD</sub>-ACAD9 complex, showing the presence of a compact rectangular core with  
 818 visible secondary structural features. Diffuse densities at defined locations on the ACAD9 core  
 819 are indicated with arrowheads. Scale bar = 50 Å. (B) The cryo-EM envelope of the complex  
 820 filtered to 15 Å at high (left) and low (right) thresholds, viewed from the bottom (above) and  
 821 end (below). The VLCAD-based homology model is fitted in the density, with dehydrogenase  
 822 domains and the C-terminal vestigial domains shown in blue and gold respectively. The high  
 823 threshold panels highlight the good overall fit of the ACAD9 dimer core into the cryo-EM map.  
 824 The flexible C-terminal region of ACAD9 is only visible at low thresholds, where extra  
 825 protrusions emanating from the vestigial domain also become visible. These protrusions can  
 826 be most likely attributed to the highly disordered ECSIT<sub>CTD</sub>. Scale bar = 50 Å.



827

828 **Fig. 5. ECSI-mediated ACAD9 defflavination.** (A) Slices through the cryo-EM  
 829 reconstruction as viewed in (E) with the VLCAD-based ACAD9 homology model fitted as  
 830 rigid body. The secondary structural features of the homology model are visible in the map,  
 831 reflecting the overall structural conservation between VLCAD and ACAD9. (B) UV visible  
 832 spectra (FAD absorption region only) of free FAD (black), ACAD9 (magenta) and ACAD9-  
 833 ECSIT<sub>CTD</sub> (blue dotted line) in the FAD absorption region. The *inset* shows the colours of the  
 834 different samples. (C) ESI MS spectra recorded in the 780-790 *m/z* region showing free FAD  
 835 standard as control (in black, observed *m/z* 786.17 Da), FAD released from ACAD9 (in  
 836 magenta, observed *m/z* 786.16 Da) and absence of FAD signal in the ACAD9-ECSI<sub>CTD</sub>  
 837 subcomplex (in blue). (D) Release of FAD from the ACAD9 catalytic pocket upon subcomplex  
 838 reconstitution. Chromatographic elution profiles of ACAD9-ECSI<sub>CTD</sub> subcomplex  
 839 reconstituted upon mixing ACAD9 and ECSIT<sub>CTD</sub>. A UV<sub>280nm</sub> peak at 21 mL (blue line, arrow)  
 840 is visible, in contrast to the chromatography profile of the co-purified complex with a single  
 841 elution peak (green). The higher absorbance at UV<sub>260nm</sub> (red line) suggests that it corresponds  
 842 to the FAD released from reconstituted subcomplex. In the *inset*, UV visible spectra of the peak  
 843 contents confirming the presence of FAD. (E) The 7.8 Å resolution cryo-EM map (grey) of the  
 844 ACAD9 dimer core of the ACAD9-ECSI<sub>CTD</sub> complex confirms the lack of bound FAD  
 845 (binding sites in the VLCAD-based homology model are indicated in red). Scale bar = 50 Å.



846

847 **Fig. 6. Proposed model of the deflavination-based mechanism of ACAD9 by ECSIT and**  
 848 **the coordinated regulation between FAO and OXPHOS pathways.** In absence of ECSIT,  
 849 ACAD9 acts as an acyl-CoA dehydrogenase enzyme in the first step of the fatty acid  $\beta$ -  
 850 oxidation (FAO) pathway. Dehydrogenation of the acyl-CoA substrate is concomitant with the  
 851 reduction of the FAD prosthetic group into FADH<sub>2</sub>. Upon ECSIT binding to ACAD9, FAD is  
 852 ejected, shutting down the dehydrogenase activity of ACAD9. This impairs the transfer of  
 853 electrons to the electron transfer flavoprotein (ETF) and the subsequent transfer to the electron  
 854 transfer flavoprotein-ubiquinone oxidoreductase (ETF-QO) until the electrons reach the  
 855 ubiquinone (UQ) pool in the respiratory chain. In this way, the function of ACAD9 in FAO is  
 856 antagonized and ACAD9 becomes committed to a role as a CI assembly factor. Furthermore,  
 857 ECSIT allows the recruitment of NDUFAF1 to the ACAD9-ECSIT subcomplex to form the  
 858 entire MCIAC complex. The latter is required for the assembly of a functional CI and  
 859 subsequently for the activation of the respiratory chain as part of the OXPHOS system. In  
 860 presence of low levels of A $\beta$  toxicity, ECSIT is upregulated and remains associated with CI.  
 861 In addition, A $\beta$  toxicity also generates an increase in CI activity, which may trigger oxidative  
 862 stress (ROS) and therefore further exacerbate the accumulation of A $\beta$  oligomers, leading to an  
 863 inhibition of CI in the long term.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21

Supplementary Materials for

**ECSIT induces allosteric deflavination of a  
mitochondrial Complex I assembly factor**

Gabriele Giachin<sup>†</sup>, Matthew Jessop<sup>†</sup>, Romain Bouverot, Samira Acajjaoui, Melissa Saïdi, Anaïs Chretien, Maria Bacia-Verloop, Luca Signor, Philippe J. Mas, Adrien Favier, Eve Borel Meneroud, Michael Hons, Darren J. Hart, Eaazhisai Kandiah, Elisabetta Boeri Erba, Alain Buisson, Gordon Leonard, Irina Gutsche\* & Montserrat Soler-Lopez\*

Correspondence to: [irina.gutsche@ibs.fr](mailto:irina.gutsche@ibs.fr); [solerlop@esrf.fr](mailto:solerlop@esrf.fr)

<sup>†</sup> Equal contribution

## 22 **Supplementary Materials and Methods**

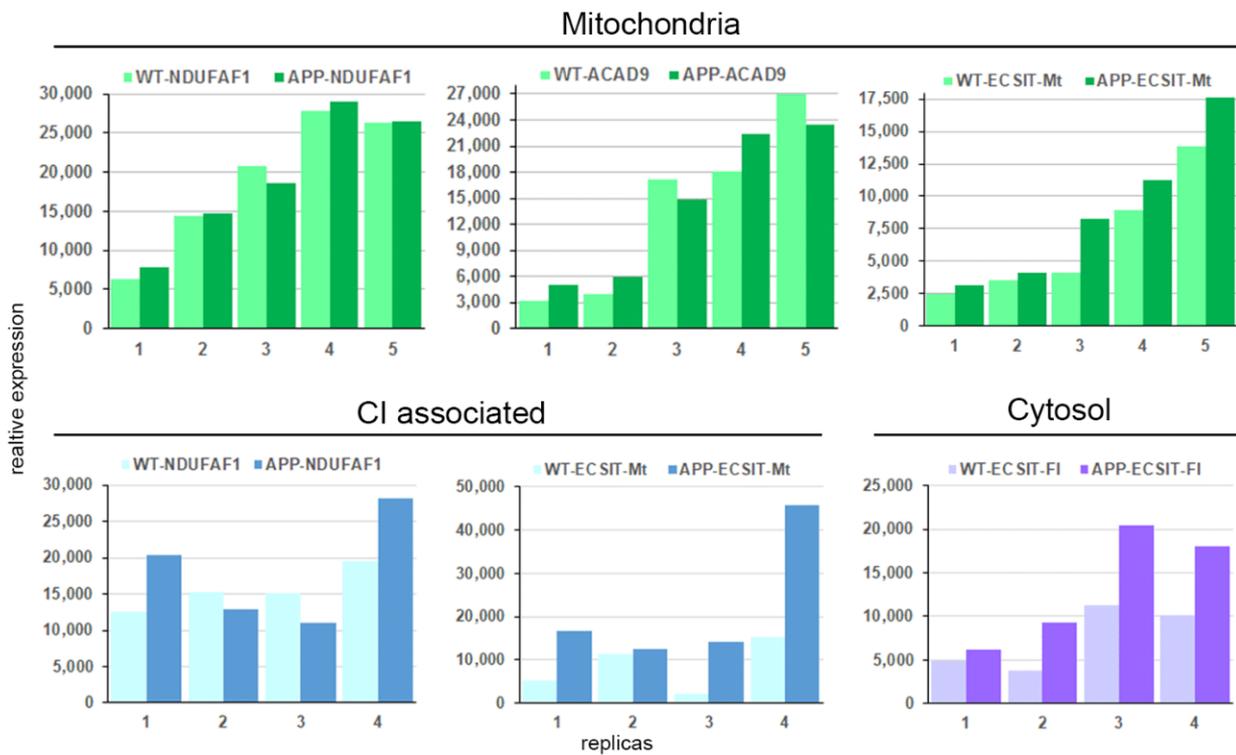
### 23 **Small Angle X-ray Scattering (SAXS) data collection and analysis**

24 SAXS analysis of the overall parameters was carried out by PRIMUS from ATSAS 2.8.4 package  
25 (59) and by ScÅtter 3.0 software (60). The pair distance distribution function,  $P(r)$ , and maximum  
26 diameter of the particle ( $D_{max}$ ) were calculated in GNOM using indirect Fourier transform method  
27 (61). Protein molecular masses were estimated using both Porod volume (61) and scattering mass  
28 contrast (62) methods. For low-resolution structural models of ECSIT<sub>CTD</sub> and NDUFAF1 Ensemble  
29 Optimization Method (EOM) modeling was conducted (63). For ACAD9, VLCAD and the ACAD9-  
30 ECSIT<sub>CTD</sub> subcomplex the pair distance distribution functions were used to calculate *ab initio* models  
31 in C2 symmetry with DAMMIN and GASBOR; the models were averaged, aligned and compared  
32 using DAMAVER (59). The most representative *ab initio* models for ACAD9 and VLCAD were  
33 compared to the ACAD9 homology model and the VLCAD structure, respectively, aligned using  
34 SUPCOMB (64) and visualized using UCSF Chimera software (65). All SAXS data were deposited  
35 into SASBDB data bank.

### 36 **Cryo-Electron Microscopy (cryo-EM), data acquisition, processing and 3D reconstruction**

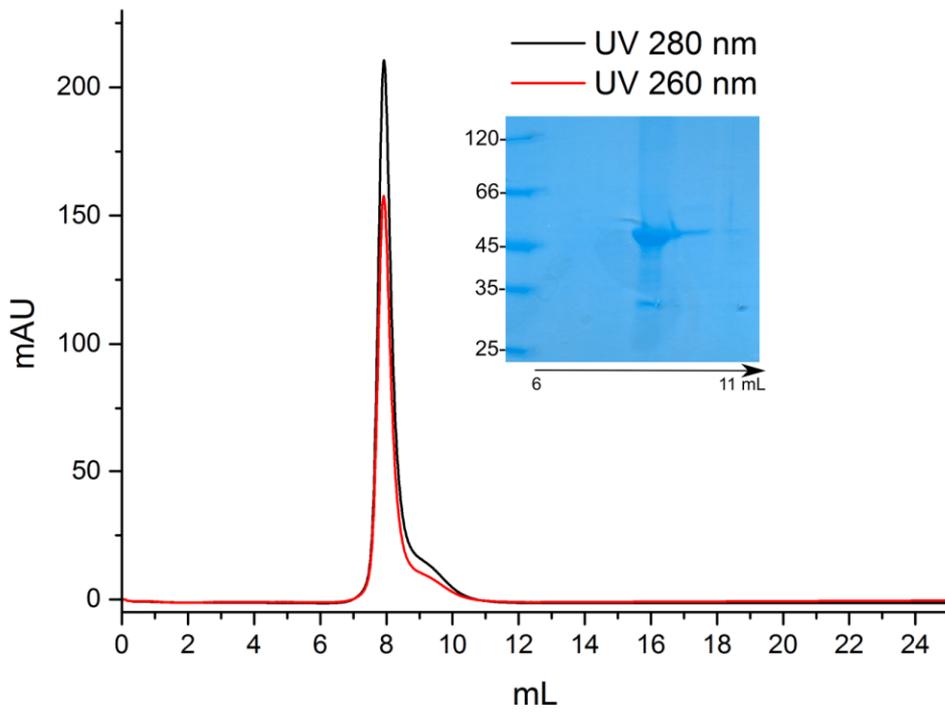
37 We first focussed only on the ACAD9 core, in order to avoid misalignment due to the highly  
38 disordered ECSIT<sub>CTD</sub>. Imported particles were subjected to several rounds of 2D classification with  
39 high initial uncertainty factors. ~16,000 particles from the 2D class averages showing the strongest  
40 secondary structural features of the ACAD9 core were selected and subjected to *ab initio* 3D  
41 reconstruction with applied C2 symmetry. The resulting volume was then used to create a mask using  
42 the Volume Tools utility in CryoSPARC. This mask was applied during homogeneous refinement  
43 with applied C2 symmetry against the *ab initio* volume, resulting in a map with an estimated  
44 resolution of 7.8 Å after sharpening with a B-factor of -500 Å<sup>2</sup>.

45 For the full map of ACAD9-ECSIT<sub>CTD</sub>, particles were split into 6 equal subsets. Each subset was  
46 subjected to *ab initio* reconstruction into 5 classes with no applied symmetry. This consistently  
47 resulted in one class with dimensions agreeing with features of the 2D classes, and four junk classes.  
48 Particles from the best *ab initio* 3D class from each subset were combined, resulting in ~240,000  
49 particles. These particles were then subjected to several rounds of 2D classification. ~36,000 particles  
50 from the 6 best 2D class averages which showed different orientations were subjected to *ab initio*  
51 reconstruction with no applied symmetry into two classes. Particles from the best class were then  
52 subjected to non-uniform refinement with applied C2 symmetry, resulting in a map that possessed an  
53 ACAD9 core that was consistent with the homology model and displayed notable protrusions near  
54 the vestigial dehydrogenase of ACAD9, which are attributed to ECSIT<sub>CTD</sub>. Particles were then refined  
55 against this map in RELION with applied C2 symmetry, resulting in a map with a nominal resolution  
56 of ~8 Å. Considering the better consistency of the masked ACAD9 core map with the VLCAD-based  
57 ACAD9 homology model, we chose to filter the final ACAD9-ECSIT<sub>CTD</sub> reconstruction to a  
58 resolution of 15 Å in order to not over-interpret high-resolution features and to focus on the global  
59 envelope of the ACAD9-ECSIT<sub>CTD</sub> complex with a view to investigate the ECSIT<sub>CTD</sub> binding site.  
60



61  
62  
63  
64  
65  
66  
67  
68  
69

**Fig. S1. Quantitative analysis of MCIA protein expression levels in individual replicas.** The densitometry levels were measured across the subcellular fractions (cytosolic, mitochondrial and CI-immunopurified) in wild type (WT) vs APP-overexpressing H4 cells. The specific band intensity of every Western blot was measured by scanning densitometry and normalized to that of a total protein as detected from the stain-free fluorescence of the transferred protein load in the Western blots. A representative Western blot and the average values as fold-change expression in APP vs WT are shown in **Figure 1**. ECSIT-FI: ECSIT full-length (FI); ECSIT-Mt: mitochondrial ECSIT.



70

71 **Fig. S2. Recombinant, purified mitochondrial ECSIT forms soluble aggregates.**

72 Chromatographic elution profile of the mitochondrial form of ECSIT (residues 49-431, 44kDa)

73 monitored at two wavelengths, showing the elution of the protein main peak in the void volume of

74 the column (8mL, GE Superdex column 200 10/300). In the inset, SDS-PAGE gel showing the eluted

75 fractions in the void volume (from 6 to 11 mL).

76

A

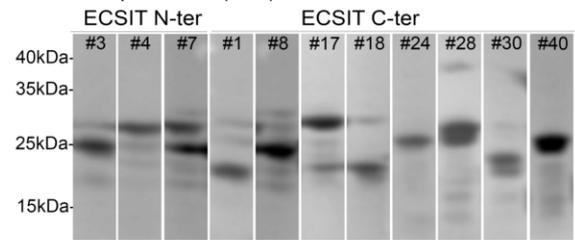
ECSIT library and screening bottleneck	
ECSIT N-ter pool of 3 fixed termini Range of library 950 bp, 9405 clones	ECSIT C-ter pool of 2 fixed termini Range of library 700 bp, 4620 clones
Expression and solubility screening	
7 clones selected	49 clones selected
Large scale expression	
3 clones selected high expression, low solubility: # 3, 4, 7	8 clones selected high expression, low solubility: #1, 8, 18, 24, 30 high solubility: #17, 28, 40

C

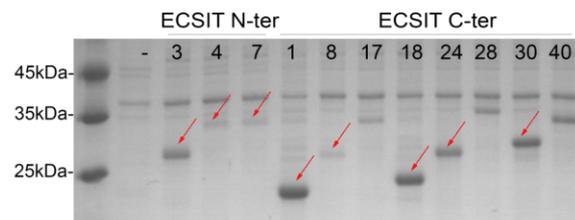
mitochondrial ECSIT (49-431)	#	Mw (kDa)
6xHis 49-431 BAP		
235-431	28	27.4
248-431	40	26.0
252-431	17	25.2
265-431	30	24.2
265-404	18	21.2
251-392	8	21.3
261-392	1	20.4
289-392	24	17.3
50-248	4	28.4
50-247	7	28.3
100-269	3	25.0

B

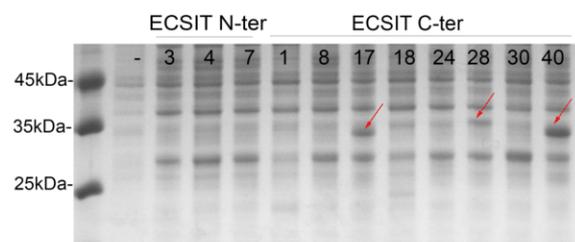
i. Total expression (WB)



ii. Insoluble fractions



iii. Soluble fractions



77

78

79

80

81

82

83

84

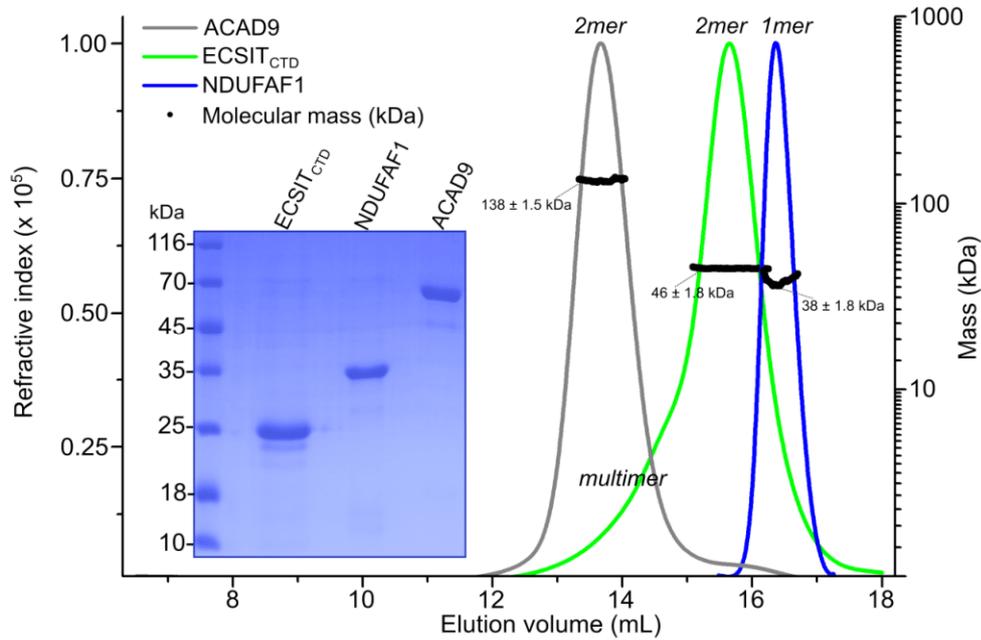
85

86

87

88

**Fig. S3. Identification soluble fragments of ECSIT by the ESPRIT approach.** (A) Schematic representation of the strategy for identifying soluble ECSIT domains: from a high-throughput screen of 9,405 and 4,620 clones corresponding to ECSIT N-terminus and C-terminus pools, respectively, 3 and 8 clones per each pool were selected for the scale-up. (C) From the top, Western streptavidin blot against the C-terminal biotin acceptor peptide (BAP), insoluble and soluble fractions of protein fragments corresponding to N-terminal (clones #3, 4, 7) and C-terminal (clones #17, 18, 24, 28, 30 and 40) ECSIT. Red arrows indicate the proteins expressed as inclusion bodies or as soluble fractions, respectively. (C) Summary of sequenced 3 N-terminal ECSIT and 8 C-terminal ECSIT clones. Construct boundaries are shown with the predicted molecular weights including 5 kDa from the hexahistidine (6xHis) and BAP tags.



90

91 **Fig. S4. SEC-MALLS characterisation of recombinant ECSIT<sub>CTD</sub>, NDUFAF1 and ACAD9.**

92 Elution profiles were monitored by excess refractive index (left ordinate axis). The black dots under

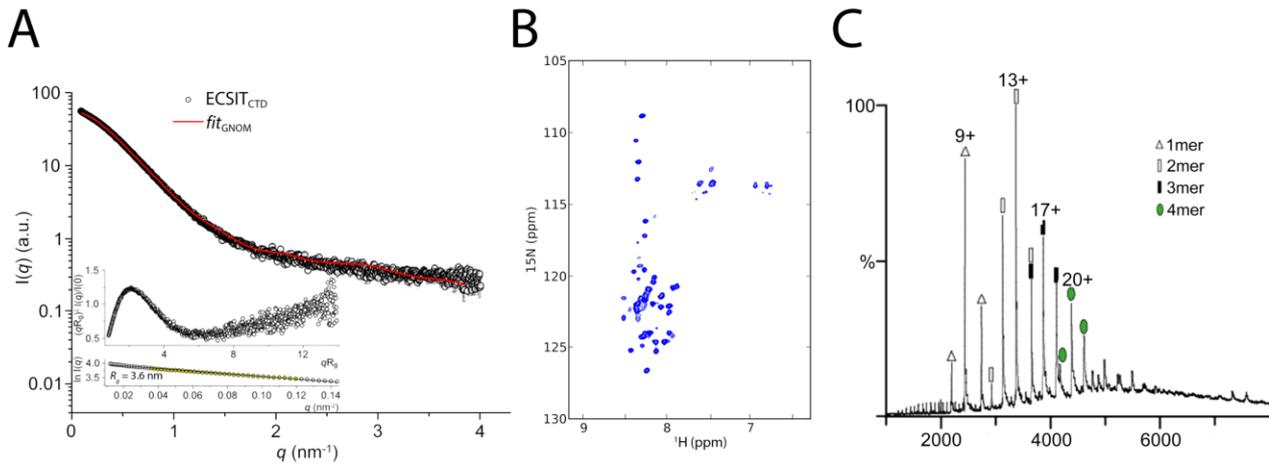
93 each elution peak shows the molecular mass distribution (right ordinate axis). Measured molecular

94 masses show that ACAD9 (grey line) is a  $138 \pm 1.5$  kDa dimer, ECSIT<sub>CTD</sub> (green line) a  $46 \pm 1.8$  kDa95 dimer and NDUFAF1 (blue line) a  $38 \pm 1.8$  kDa monomer. In the inset, SDS-PAGE gel showing the

96 purity and size of monomeric proteins in denaturing conditions.

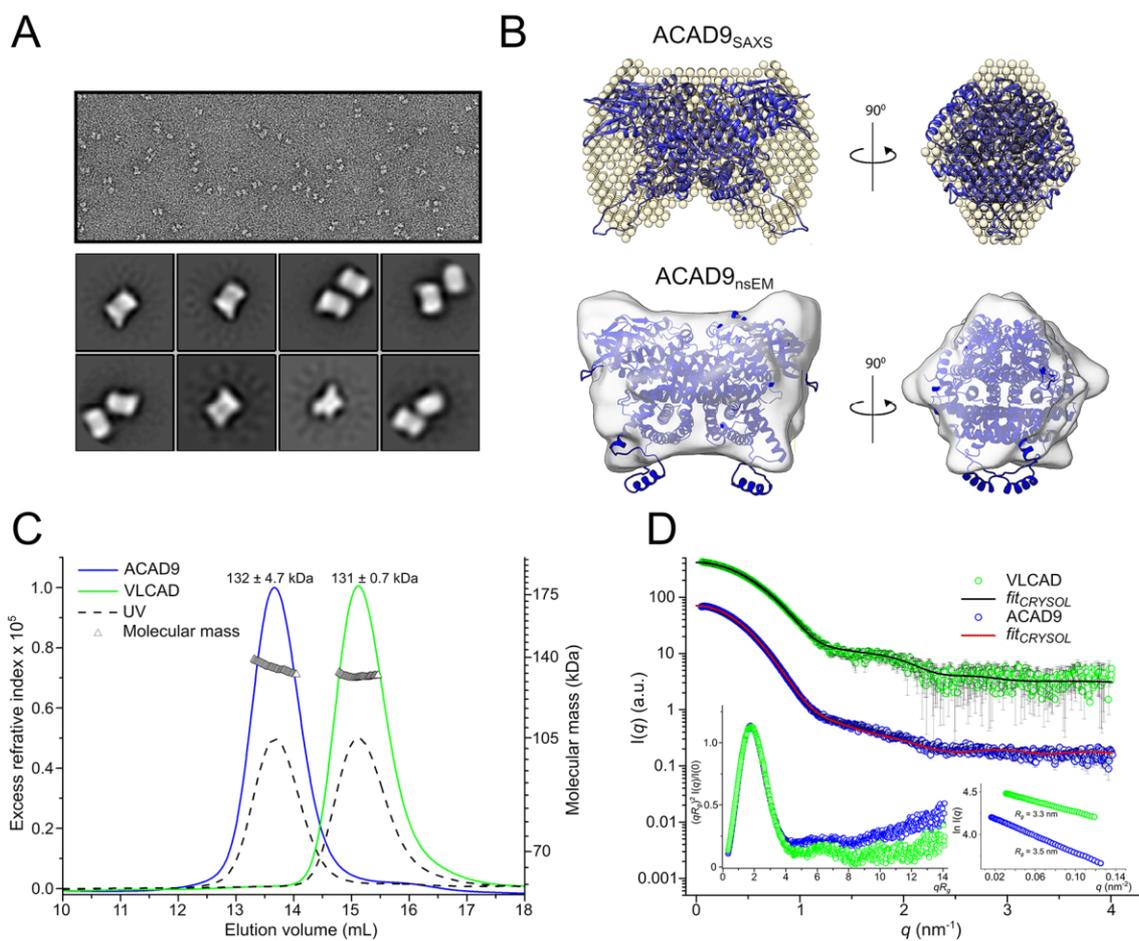
97

98



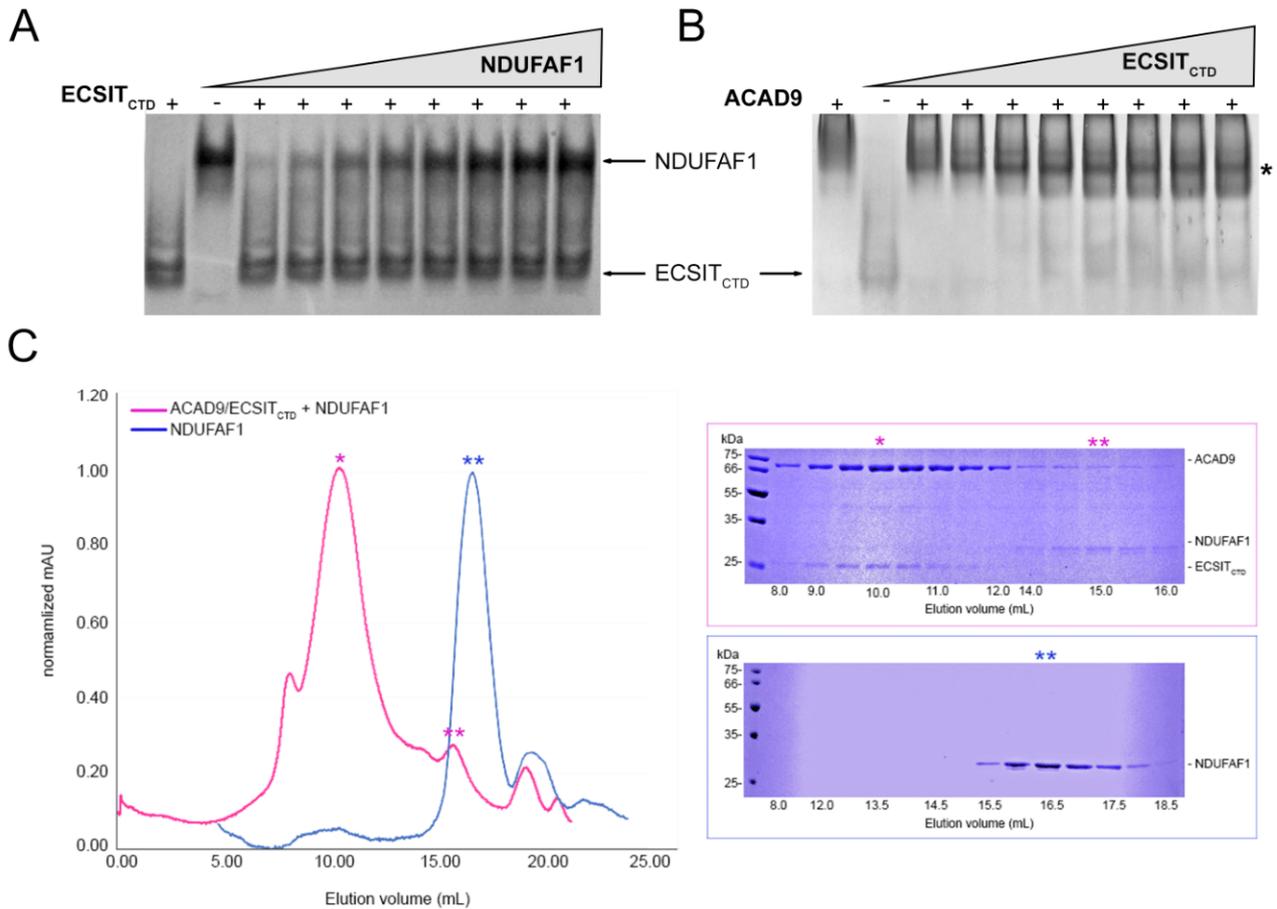
99

100 **Fig. S5. ECSIT<sub>CTD</sub> is highly disordered and forms predominantly dimers.** (A) SAXS  
 101 experimental curve of ECSIT<sub>CTD</sub>. In the inset, dimensionless Kratky plots and Guinier fit analyses,  
 102 with an estimated mass of 46377 Da, corresponding to a dimer. (B) <sup>15</sup>N-HSQC NMR spectrum of  
 103 <sup>15</sup>N-labelled ECSIT<sub>CTD</sub> displays about 60 peaks in the amide region in a narrow <sup>1</sup>H frequency window  
 104 corresponding to the amide random coil region showing that the residues observed by NMR belong  
 105 to one or more unfolded parts of the protein. (C) Native MS spectrum of ECSIT<sub>CTD</sub> with the detection  
 106 of monomers (1mer, 21885 ± 1 Da), dimers (2mer, 43771 ± 2 Da), tetramers (4mer, 87543 ± 4 Da)  
 107 and higher-order multimers.  
 108



109

110 **Fig. S6. Conformational analysis of the ACAD9 homodimer.** (A) Ns-EM raw micrograph of  
 111 ACAD9 particles and 2D class averages. (B) *Top*, SAXS *ab initio* envelop for the ACAD9 dimer  
 112 (bead model representation) matching the ACAD9 homology model (cartoon representation) (17).  
 113 *Bottom*, Ns-EM density map (grey) of ACAD9 generated from averaged class particles and fitted  
 114 with the homology model (cartoon representation). (C) SEC-MALLS analysis of ACAD9 (residues  
 115 38-621, blue line) and VLCAD (residues 75-955, green line) by excess refractive index (left ordinate  
 116 axis) and UV<sub>280 nm</sub> wavelength (dotted lines). The triangular symbols under each elution peak show  
 117 the molecular mass distribution (right ordinate axis), with ACAD9 as a  $131 \pm 0.7$  kDa and VLCAD  
 118 as a  $132 \pm 4.7$  homodimers. (D) X-ray scattering curves for ACAD9 and VLCAD (blue and green  
 119 circles, respectively). *CRYSOLOG* fits using the ACAD9 homology model (17) and the VLCAD dimer  
 120 crystal structure (66) represented as red lines ( $\chi^2$  values 2.5 and 0.92, respectively). In the inset,  
 121 normalized Kratky plots (left) and close-up view of the Guinier region for the two proteins (right).  
 122



123

124

125

126

127

128

129

130

131

132

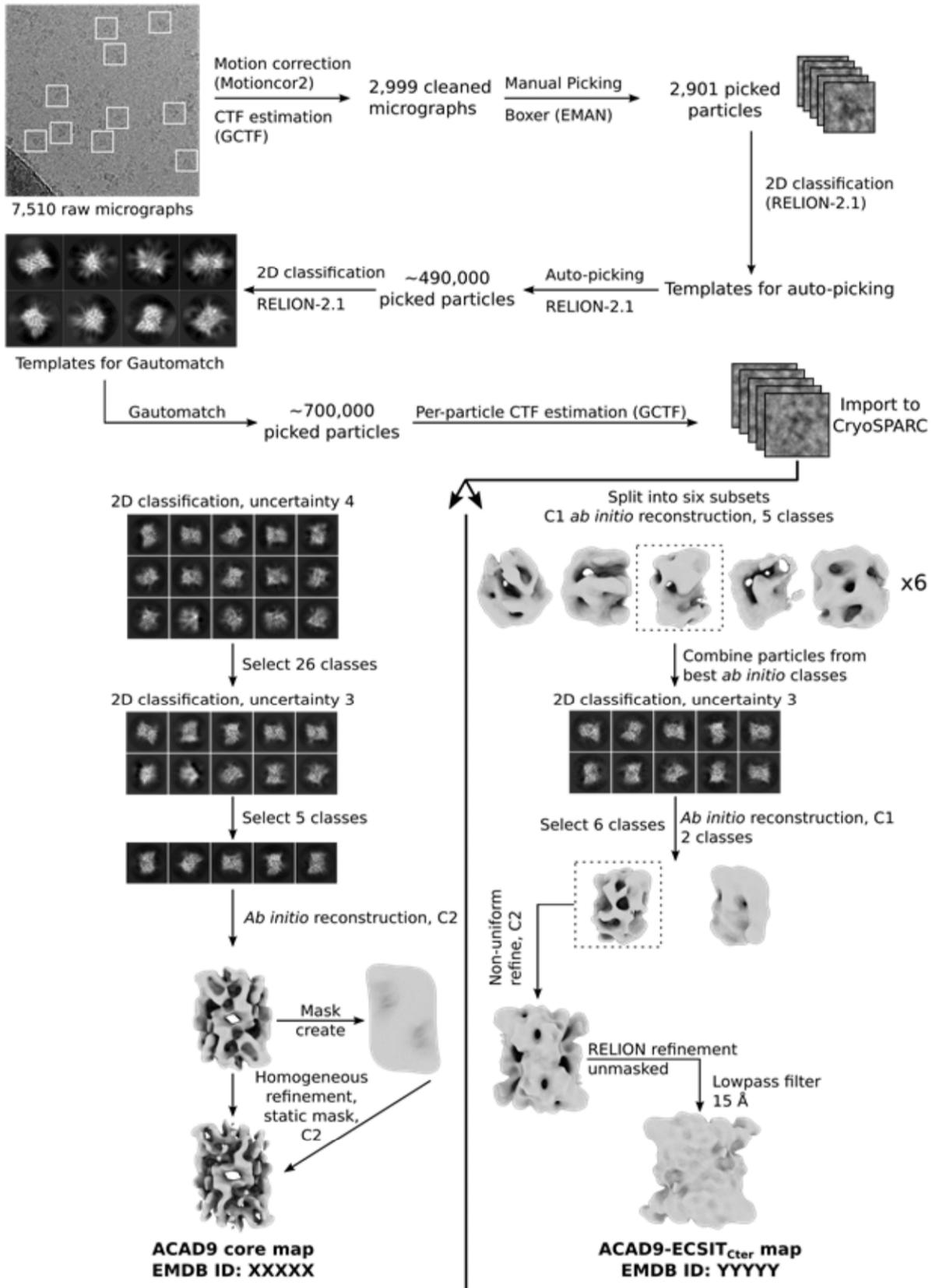
133

134

135

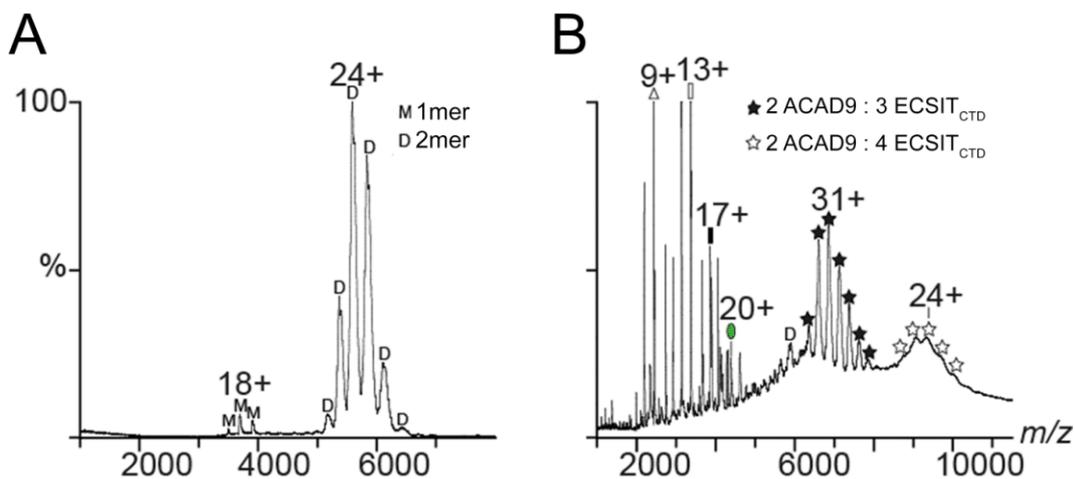
136

**Fig. S7. *In vitro* interaction assays between ACAD9, ECSIT<sub>CTD</sub> and NDUFAF1.** (A) ECSIT<sub>CTD</sub> when titrated with increasing NDUFAF1 concentrations shows no change in mobility. (B) In contrast, ACAD9 titrated with increasing concentrations of ECSIT<sub>CTD</sub> shows the appearance of a new band, concomitant with the disappearance of the free ECSIT<sub>CTD</sub> band, revealing an interaction (\*). (C) On the left, UV<sub>280nm</sub> chromatographic elution profile of NDUFAF1 incubated with ACAD9-ECSIT<sub>CTD</sub> copurified subcomplex (pink line). Comparison with the overlapped elution profile of NDUFAF1 alone (blue line) indicates that the main elution peak of the former profile corresponds to the ACAD9-ECSIT<sub>CTD</sub> subcomplex (\*), consistent with **Figure 3A**, with a small adjacent peak corresponding to unbound NDUFAF1 (\*\*). The absence of a shift in the ACAD9-ECSIT<sub>CTD</sub> subcomplex peak in presence of NDUFAF1 further confirms that NDUFAF1 is unable to form a complex with ACAD9-ECSIT<sub>CTD</sub>. SDS-PAGE gels of the eluted fractions are shown on the right. *Top*, fractions from the gel filtration upon incubation of the three proteins. *Bottom*, fractions corresponding to the gel filtration of NDUFAF1 alone. Fractions corresponding to the peaks are highlighted with \*.



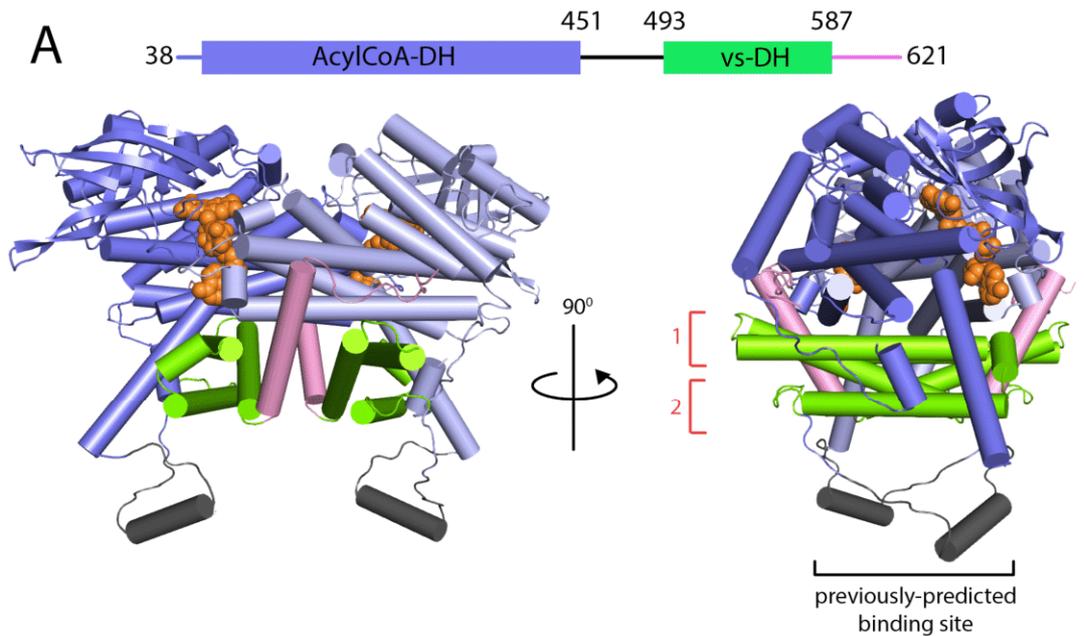
138

139 **Fig. S8. Cryo-EM processing strategy.** Image processing pipeline of cryo-EM data for ACAD9-  
 140 ECSIT<sub>CTD</sub> with software packages and algorithms used at each step indicated, for both the ACAD9  
 141 core and ACA9-ECSIT<sub>CTD</sub> maps.



142

143 **Fig. S9. Investigation of different complex assemblies by Native MS.** (A) Mass spectrometric  
 144 analysis of ACAD9 confirmed its homodimeric form (D 2mer,  $131910 \pm 4$  Da); a very low abundant  
 145 monomer was also detected (M 1mer,  $65957 \pm 3$  Da). (B) The copurified ACAD9-ECSIT<sub>CTD</sub> sample  
 146 used in the cryo-EM experiments. Two major non-covalent species were characterised with a  
 147 stoichiometry of an ACAD9 homodimer and 3 to 4 copies of ECSIT<sub>CTD</sub> (masses of  $197,561 \pm 3$  Da  
 148 and  $219,457 \pm 4$  Da respectively). A series of additional peaks were detected consistent with the  
 149 spectrum of unbound ECSIT<sub>CTD</sub> shown in **Figure S5C**, revealing the tendency of ECSIT<sub>CTD</sub> to form  
 150 higher-order multimeric species in solution.



**B**

	492	502	512	522	532	542	552	562	572
aHuman	GVVHPSLADS	ANKFEENTYC	FGRTVETLLL	RFGKTIMEEQ	LVLKRVANIL	INLYGMTAVL	SRASRSIRIG	LRNHDEVELL	ANTFCVEAYL
aGorilla	GVVHPSLADS	ANKFEENTYC	FGRTVETLLL	RFGKTIMEEQ	LVLKRVANIL	INLYGMTAVL	SRASRSIRIG	LRNHDEVELL	ANTFCVEAYL
aChimpanze	GVVHPSLADS	ANKFEENTYC	FGRTVETLLL	RFGKTIMEEQ	LVLKRVANIL	INLYGMTAVL	SRASRSIRIG	LRNHDEVELL	ANTFCVEAYL
aOrangutan	GVVHPSLADS	ANKFEENTYC	FGRTVETLLL	RFGKTIMEEQ	LVLKRVANIL	INLYGMTAVL	SRASRSIRIG	LRNHDEVELL	ANTFCVEAYL
aHorse	GAVHPTVADS	ANKLEENVYY	FGRTVETLLL	RFGKTIVDEQ	MVLKRVANIL	INLYGMTAVL	SRASRSIRVG	LRNHDEVELL	ANMFCVEAYY
aPolecat	GAVHPSVADS	ANKLEENVYY	FGRTVETLLL	RFGKTIVDEQ	MVLKRVANIL	INLYGMTAVL	SRASRSIRVG	LRNHDEVELL	ANMFCVEAYY
aBeaver	GVVHPSLADS	ANKLEENVYY	FGRTVETLLL	RFGKTIVDEQ	LVLKRVANIL	INLYGMTAVL	SRASRSIRIG	LRNHDEVELL	ANMFCVEAYF
aRat	AVVHPSLADS	ANKLEENVHY	FGRTVETLLL	RFGKTIVDEQ	LVLKRVANIL	INLYGMTAVL	SRASRSIRIG	LRNHDEVELL	ANMFCVEAYF
aMouse	GVVHPSLADS	ANKLEENVHY	FGRTVETLLL	RFGKTIVDEQ	LVLKRVANIL	INLYGMTAVL	SRASRSIRIG	LRNHDEVELL	ANMFCVEAYF
aSalmon	GVVHPSLADS	AKKLEQNVAH	FGSTVENLLY	RYGKTIVDEQ	LILKRVADVIL	INLYAMTAVL	SRSSRSISIG	LRNHDEVELL	ANTFCTDAFF
aZebrafish	GVVHPSLADS	AKMFEQNAAC	FGSTVEGLLY	RYGKTIVDEQ	LLLKRVADVIL	INLYAMTAVL	SRSSRSISIG	LRNHDEVELL	TNTFCKDAH
	. : * : *	. : * : *	. : * : *	. : * : *	. : * : *	. : * : *	. : * : *	. : * : *	. : * : *
vHuman	GLVHPELSRS	GELAVQALEQ	FATVVEAKLI	KHKKGIVNEQ	FLLQRLADGA	IDLYAMVVVL	SRASRSISEG	HPTA <sup>1</sup> QHEKML	CDTW <sup>2</sup> CEIAAA
vGorilla	GLVHPELSRS	GELAVQALEQ	FATVVEAKLI	KHKKGIVNEQ	FLLQRLADGA	IDLYAMVVVL	SRASRSISEG	HPTA <sup>1</sup> QHEKML	CDTW <sup>2</sup> CEIAAA
vChimpanze	GLVHPELSRS	GELAVQALEQ	FATVVEAKLI	KHKKGIVNEQ	FLLQRLADGA	IDLYAMVVVL	SRASRSISEG	HPTA <sup>1</sup> QHEKML	CDTW <sup>2</sup> CEIAAA
vOrangutan	GLVHPELSRS	GELAVQALEQ	FATVVEAKLI	KHKKGIVNEQ	FLLQRLADGA	IDLYAMVVVL	SRASRSISEG	HPTA <sup>1</sup> QHEKML	CDTW <sup>2</sup> CEIAAA
vHorse	GIVHPDLRS	GELAVQALEQ	FATVVEAKLI	KHKKEIVNEQ	FLLQRLADSA	IDLYAMVVVL	SRASRSISEG	HPTA <sup>1</sup> QHEKML	CDSW <sup>2</sup> CEIAAA
vPolecat	GLVHPELNS	GELTVQALEQ	FASVVEAKLI	KHKKGIVNEQ	LVLQRLADSA	IDLYAMVVVL	SRASRSISEG	HSTA <sup>1</sup> QHEKML	CDSW <sup>2</sup> CEIAAA
vBeaver	GVIHPELTRS	GELAVQALEQ	FATVVEAKLI	KHKKGIISEQ	FLLQRLADGA	IDLYAMAVVL	SRASRSISEG	YPTA <sup>1</sup> QHEKML	CDSW <sup>2</sup> CEIAAA
vRat	GIVHPELSRS	GELAVQALEQ	FATVVEAKLM	KHKKGIVNEQ	FLLQRLADGA	IDLYAMVVVL	SRASRSISEG	YPTA <sup>1</sup> QHEKML	CDSW <sup>2</sup> CEIAAT
vMouse	GIVHPELSRS	GELAVQALDQ	FATVVEAKLV	KHKKGIVNEQ	FLLQRLADGA	IDLYAMVVVL	SRASRSISEG	YPTA <sup>1</sup> QHEKML	CDSW <sup>2</sup> CEIAAT
vSalmon	GSIHPELSRS	GDLAVKATEH	FGVTVEMLL	KHGKILHEQ	FVLKRVADSA	IDLYAMVVVL	SRASRSISQG	HASA <sup>1</sup> QHEKLL	CETW <sup>2</sup> CEIAYD
vZebrafish	GTVHPELNHS	GELTVKATEQ	FGAVIEELLL	RHGKIIIDQ	FVLKRVADCA	IDLYAMVVVL	SRASRSINQG	HSSA <sup>1</sup> QHEKML	CETW <sup>2</sup> CEAHE

151

152 **Fig. S10. Structural features of the ACAD9 homodimer and potential interaction sites. (A) Top.**

153 Domain organization of ACAD9, showing the Acyl-Coa (purple) and the vestigial (vs, green)

154 dehydrogenase domains. *Bottom.* Two orientations of the ACAD9 homodimer *in silico* model (17)

155 based on the VLCAD crystal structure (66). The dehydrogenase domain is shown in purple, the

156 vestigial domain in green and the C-terminal segment responsible of homodimerization in pink. The

157 C-terminal stretch of residues, poorly conserved between ECSIT and VLCAD and previously

158 predicted to fold as a helix and be a potential protein-binding site, is shown in grey (17). Bound FAD

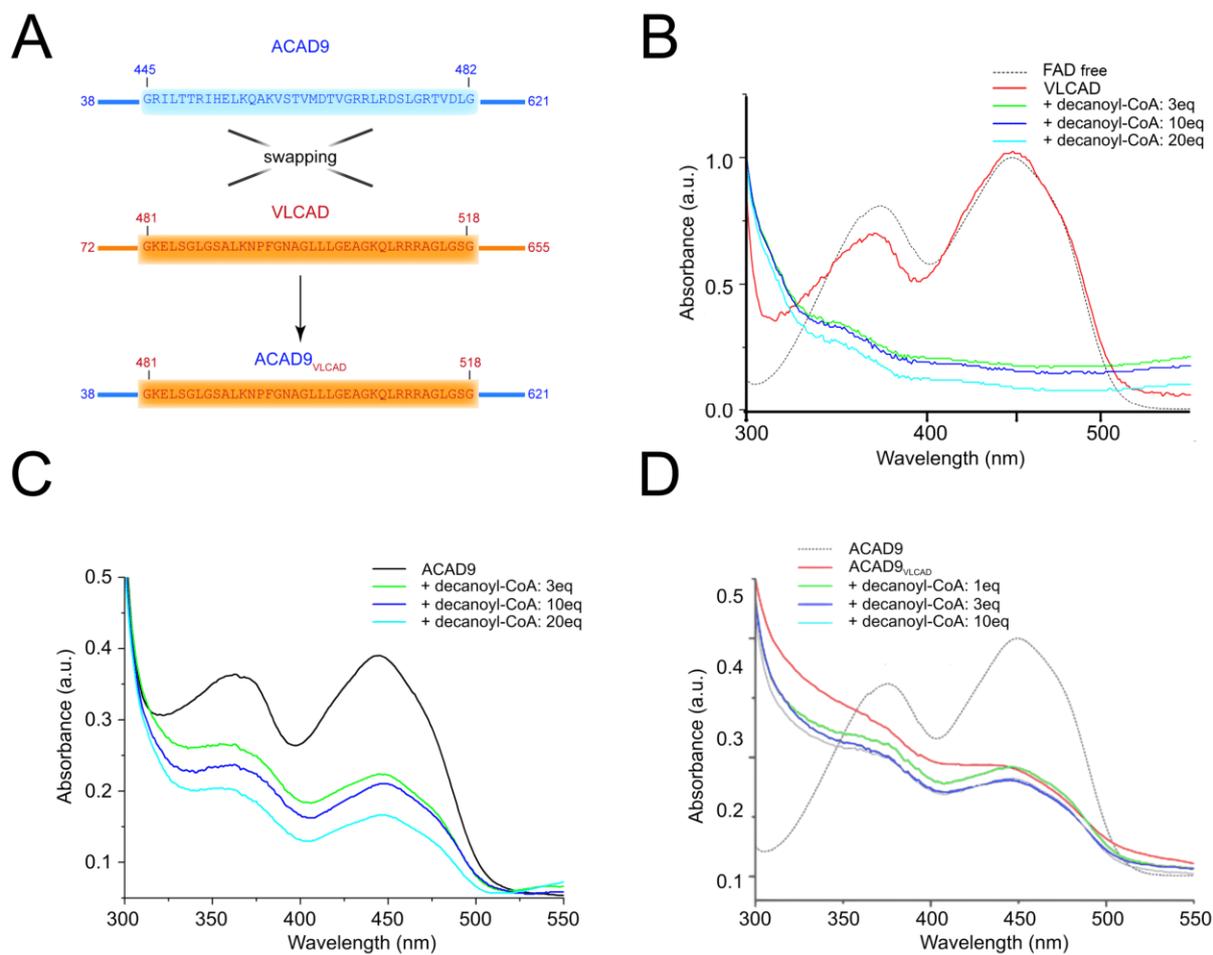
159 is shown as orange spheres. **(B)** Multiple sequence alignment of ACAD9 (a) and VLCAD (v)

160 orthologues. Conserved residues between ACAD9 and VLCAD are highlighted in light blue. The

161 regions corresponding to ACAD9 sequences 488GVVHPSLAD496 and 555RIGLRNH561 are boxed (1

162 and 2), highlighting the conservation of these sequences within ACAD9 and VLCAD orthologues

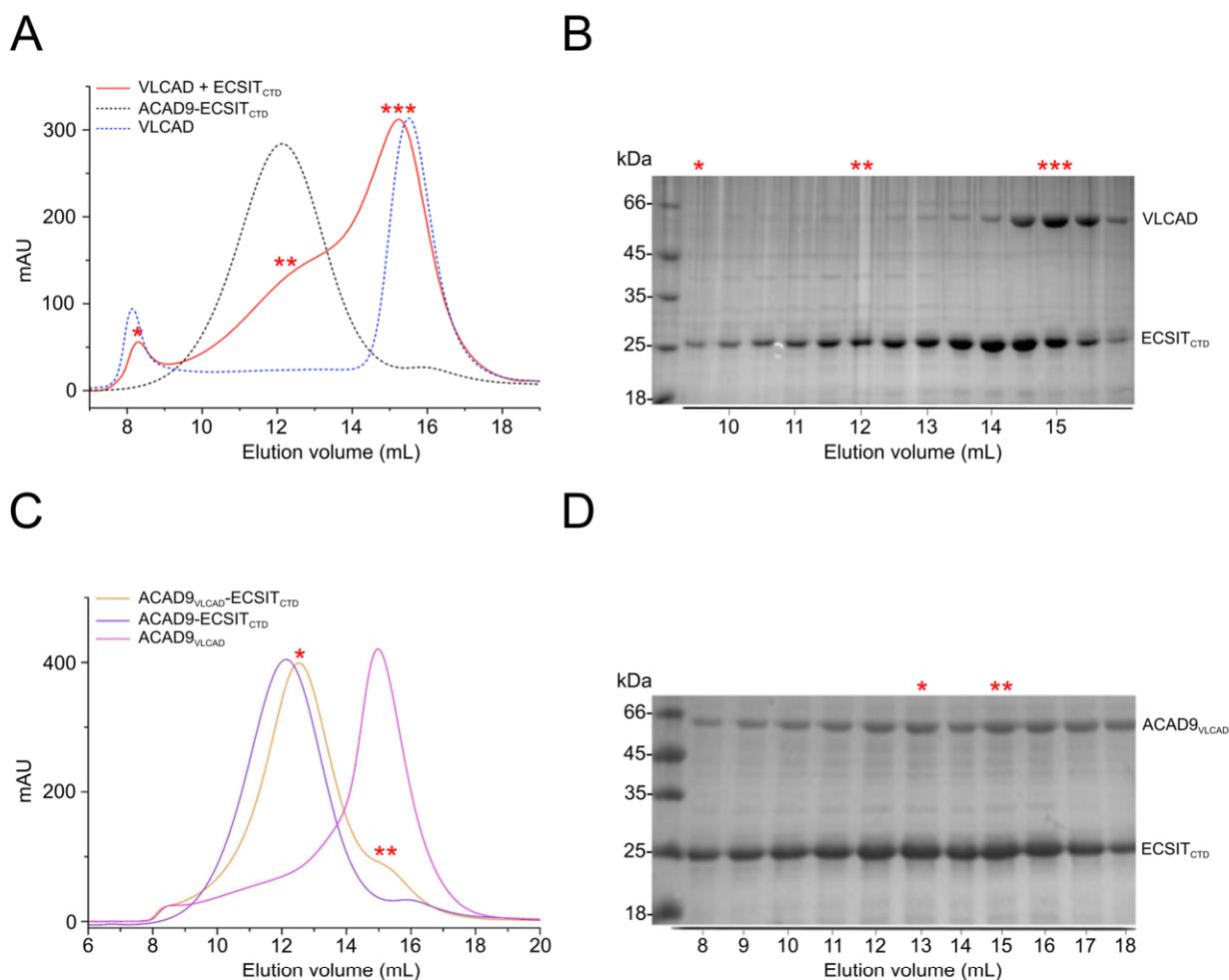
163 and the divergence between the two groups.



164

165 **Fig. S11. Construction of a chimeric ACAD9 and its dehydrogenase activity.** (A) Design of a  
 166 VLCAD-based ACAD9 chimeric mutant by swapping the C-terminal non-conserved 37 amino acid  
 167 (residues 481 to 518) from VLCAD into ACAD9 (replacing residues 445 to 482). (B) UV absorption  
 168 spectra of free FAD (dashed line), showing absorption maxima near 370 and 450 nm, which are  
 169 characteristic of oxidized FAD. The spectra of FAD bound to VLCAD (red line) and in the presence  
 170 of 1, 3, 10, 20 molar equivalent of decanoyl-CoA (blue and green lines) were recorded 2 min after  
 171 the addition of the substrate. Quenching of the absorption peaks results from the reduction of FAD  
 172 into FADH<sub>2</sub> and provides evidence for the formation of a charge transfer complex (67). (C) The same  
 173 experiment after the addition of decanoyl-CoA to ACAD9. (D) ACAD9 chimera and its  
 174 dehydrogenase activity in presence of acyl-CoA substrate. UV visible spectra of native and chimeric  
 175 ACAD9 in the FAD absorption region. The spectra of native ACAD9 (dash line), chimeric  
 176 ACAD9<sub>VLCAD</sub> in the absence (red line) and the presence of 1, 3, 10 molar equivalents of decanoyl-  
 177 CoA (green, purple and cyan lines) were recorded 2 min after the addition of the substrate.

178

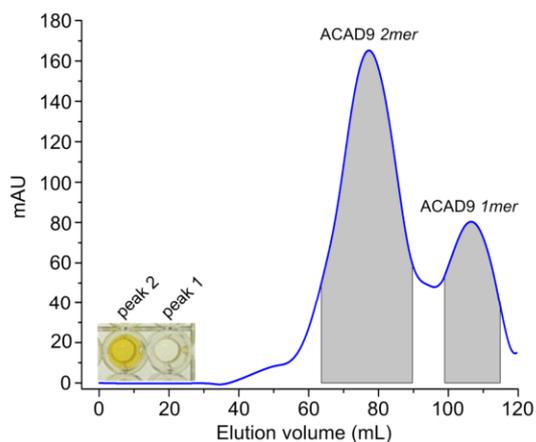
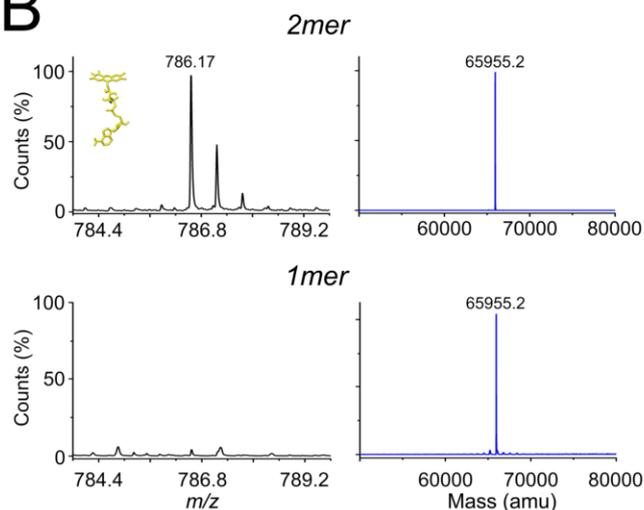


179

180 **Fig. S12. *In vitro* interaction assays between VLCAD, chimeric ACAD9<sub>VLCAD</sub> and ECSIT<sub>CTD</sub>.**

181 **(A)** Elution chromatographic profile monitored at two UV wavelengths (280 and 260 nm) during the  
 182 co-purification of VLCAD and ECSIT<sub>CTD</sub> showing three main elution peaks corresponding to the  
 183 void volume (\*), higher oligomeric species of ECSIT<sub>CTD</sub> (\*\*) and VLCAD with an excess of  
 184 ECSIT<sub>CTD</sub> (\*\*\*). Overlapping of the elution profiles of the ACAD9-ECSIT<sub>CTD</sub> subcomplex (black  
 185 dash line) and VLCAD alone (blue dash line) indicate that there is no complex formation between  
 186 VLCAD and ECSIT<sub>CTD</sub>. **(B)** SDS-PAGE gel of the eluted fractions. **(C)** Elution chromatographic  
 187 profile at 280 nm UV wavelength during the co-purification of the ACAD9<sub>VLCAD</sub>-ECSIT<sub>CTD</sub>  
 188 subcomplex (orange line) showing a main elution peak corresponding to the subcomplex (\*) and a  
 189 small adjacent peak to unbound ECSIT<sub>CTD</sub> (\*\*). Overlapping of the elution profiles of the ACAD9-  
 190 ECSIT<sub>CTD</sub> subcomplex (purple line) and ACAD9<sub>VLCAD</sub> alone (magenta line) clearly indicate that there  
 191 is a shift for the ACAD9<sub>VLCAD</sub>-ECSIT<sub>CTD</sub> co-elution, providing evidence that ACAD9<sub>VLCAD</sub> is still  
 192 able to form a complex with ECSIT<sub>CTD</sub>. **(D)** SDS-PAGE gel of the eluted fractions.

193

**A****B**

194

195 **Fig. S13. The ability of ACAD9 homodimer to bind FAD assessed by UV and MS. (A)** Only  
196 ACAD9 homodimer is able to bind FAD. Elution chromatographic profile of full-length ACAD9  
197 (residues 38-621) monitored at UV<sub>280 nm</sub> during purification with preparative SEC showing two main  
198 elution peaks corresponding to ACAD9 dimer (peak 2, yellowish) and monomer (peak 1, colourless).  
199 **(B)** LC/ESI-MS spectra of ACAD9 acquired after preparative SEC purification. In the upper panels,  
200 signals corresponding to the FAD cofactor (observed  $m/z$  786.17 Da) and to ACAD9 (observed  
201 deconvoluted mass 65955.2 Da) eluted in chromatographic peak 2 (*2mer*, panel A) confirmed that  
202 ACAD9 homodimer binds FAD. In the lower panel, MS data indicated the absence of FAD when  
203 ACAD9 is a monomer.

204

205 **Table S1. SAXS data of the proteins and protein fragments of this study.**

206

207 **(a) Sample details.**

	<b>ECSIT<sub>CTD</sub></b>	<b>NDUFAF1</b>	<b>ACAD9</b>	<b>Chimeric ACAD9<sub>VLCAD</sub></b>	<b>VLCAD</b>	<b>ACAD9- ECSIT<sub>CTD</sub> subcomplex</b>
UniProt sequence ID (residues in construct)	Q9BQ95 (247-431)	Q9Y375 (25-327)	Q9H845 (38-621)	$\Delta$ ACAD9 445- 482/VLCAD 481-518	P49748 (75-655)	Q9BQ95 (247-431) and Q9H845 (38- 621)
Particle contrast from sequence and solvent constituents, $\Delta\rho(\rho_{\text{protein}} - \rho_{\text{solvent}};$ $10^{10}; \text{cm}^{-2})$	2.949 (12.44-9.49)	2.878 (12.37-9.49)	2.744 (12.24-9.49)	2.744 (12.24-9.49)	2.732 (12.23- 9.49)	n.d.
Specific volume from chemical composition ( $v, \text{cm}^3$ $\text{g}^{-1}$ )	0.726	0.732	0.742	0.742	0.742	n.d.
Calculated monomeric mass from sequence (Da)	21887.20	35131.50	65954.82	65456.17	64040.65	n.d.
Total frames	10		2400		10	
Frames used for data analysis	10		60 (range 2280-2340)	100 (range 1900-2000)	100 (range 2080-2180)	10

208

209

**(b) SAXS data collection parameters**

Instrument	ESRF, Grenoble (France), bioSAXS beamline (BM29)		
Wavelength (Å)	0.99		
q-range (Å <sup>-1</sup> )	0.004 – 0.49		
Sample-to-detector distance (m)	2.864		
Exposure time (sec)	0.5/frame	1/frame	0.5/frame
Temperature (° C)	20		
Detector	Pilatus 1M		
Flux (photons/s)	$2 \times 10^{12}$		
Beam size (μm <sup>2</sup> )	700 x 700		
Sample configuration	1.8 mm quartz glass capillary		
Absolute scaling method	Comparison to water in sample capillary		
Normalization	To transmitted intensity by beam-stop counter		
Monitoring for radiation damage	Control of un-subtracted and scaled subtracted data for systematic changes typical for radiation damage		

210

211

**(c) Structural parameters**

	<b>ECSIT<sub>CTD</sub></b>	<b>NDUFAF1</b>	<b>ACAD9</b>	<b>Chimeric ACAD9<sub>VLCAD</sub></b>	<b>VLCAD</b>	<b>ACAD9- ECSIT<sub>CTD</sub> subcomplex</b>
<b>Guinier analysis</b>						
I(0) (cm <sup>-1</sup> )	$0.042 \pm 9.6\text{E-}05$	$0.029 \pm 1.8\text{E-}04$	$0.057 \pm 2.5\text{E-}05$	$0.11 \pm 1.1\text{E-}04$	$0.107 \pm 8\text{E-}06$	$0.190 \pm 2.8\text{E-}05$
$R_g$ (nm)	$3.53 \pm 0.041$	$2.85 \pm 0.023$	$3.54 \pm 0.011$	$3.85 \pm 0.042$	$3.36 \pm 0.024$	$6.09 \pm 0.031$
q range (nm <sup>-1</sup> )	0.0362- 0.1170	0.0173- 0.1550	0.0160- 0.1087	0.0325-0.1008	0.0352- 0.1494	0.0056- 0.0423
Auto $R_g$ fidelity	0.98	0.96	0.99	0.95	0.99	0.96
Mass from I(0) (Da); ratio to predicted	46473.3; 2.1	37700; 1.07	113650; 1.7	131078; 1.98	130875; 2.04	289822; n.d.

<b><i>P(r) analysis</i></b>						
I(0) (cm <sup>-1</sup> )	0.04236	0.02924	0.056736	0.11096	0.10704	0.19024
R <sub>g</sub> (nm)	3.65	2.9	3.51	3.7	3.31	5.97
D <sub>max</sub> (nm)	11.9	10	11.34	11.7	9.8	18.3
q range (nm <sup>-1</sup> )	0.207-3.53	0.282-2.912	0.118-3.827	0.192-3.12	0.213-3.5	0.101-2.952
Porod volume (nm <sup>3</sup> )	116.82	79.29	234	220	233	678
Porod exponent (from Scätter)	1.3	1.1	2.3	2.7	2.9	2.4
Mass estimate (as 0.5 x volume of models, Da); ratio to expected	58410; 2.61	39645; 1.12	117000; 1.8	110000; 1.7	116500; 1.8	339000
Mass from I(0) (Dalton); ratio to predicted	46377; 2.12	31740; 0.9	126800; 1.92	125930; 1.9	125492; 1.95	289737; n.d.
SASBDB Accession ID	SASDHU4	SASDHV4	SASDHW4	SASDHX4	SASDHY4	SASDHZ4

212  
213

**(d) Shape model-fitting results.**

<b><i>DAMMIF</i></b>						
<i>Default parameters, 20 calculation runs</i>						
q range (nm <sup>-1</sup> )	0.067-2.8	0.17-2.94	0.099-4.270	0.064-3	0.184-4.262	0.068-2.149
Symmetry, anisotropy assumptions	P2, none	P1, none	P2, none	P2, none	P2, none	P2, none
Normalized spatial discrepancy (standard deviation)	1.304 (0.174)	0.696 (0.026)	1.377 (0.205)	1.607 (0.228)	1.001 (0.268)	1.803 (0.268)
Resolution (from SASRES) (Å)	39 ± 3	23 ± 2	57 ± 4	57 ± 4	41 ± 3	88 ± 6
χ <sup>2</sup> range of the fitting	1.142-1.169	2.482-2.498	1.852-1.921	3.347-3.542	0.408-0.427	1.137-1.207
<b><i>DAMMIN</i></b>						
<i>Default parameters, fine, smaller beads</i>						
q range (nm <sup>-1</sup> )	0.067-2.8	0.17-2.94	0.099-4.270	0.064-3	0.184-4.262	0.068-2.149
Symmetry, anisotropy assumptions	P2, none	P1, none	P2, none	P2, none	P2, none	P2, none
χ <sup>2</sup> value of the fitting	1.081	4.317	1.743	3.535	0.4217	1.238
<b><i>GASBOR</i></b>						
<i>Default parameters, reciprocal space (fit I(s), slower)</i>						
q range (nm <sup>-1</sup> )	n.d.	n.d.	0.099-4.270	0.064-3	0.184-4.262	0.068-2.149
Symmetry, anisotropy assumptions	n.d.	n.d.	P2, none	P2, none	P2, none	P2, none
χ <sup>2</sup> value of the fitting	n.d.	n.d.	3.95	4.27	0.459	1.205

214  
215

**(e) Atomistic modelling.**

Atomic structures	n.d.	n.d.	Homology model (residues 38-621) <sup>+</sup>	Homology model (residues 38-621) <sup>+</sup>	PDB entry 3B96	n.d.
<b><i>CRY SOL</i></b>						
<i>Default parameters, with constant subtraction</i>						
- χ <sup>2</sup> value of the fitting	n.d.	n.d.	2.55	8.70	0.92	n.d.
- Predicted R <sub>g</sub> (nm)	n.d.	n.d.	3.50	3.50	3.30	n.d.

n.d.: non determined.

216  
217

## Chapter 9. Towards the structural characterisation of the MCIA complex

This chapter presents results that were obtained and steps that were taken during the characterisation of the MCIA complex that were not included in the manuscript presented in Chapter 8. As well as the cryo-EM characterisation of the ACAD9-ECSIT<sub>CTD</sub> subcomplex, I carried out some preliminary work on the full ACAD9-ECSIT-NDUFAF1 triple complex. Initial tests showed that the triple complex was more heterogeneous in comparison to the ACAD9-ECSIT<sub>CTD</sub> subcomplex, and therefore most of the structural characterisation carried out was on the subcomplex. I will also present an extended discussion on the image processing steps that were taken to resolve the ACAD9-ECSIT<sub>CTD</sub> structure by cryo-EM, as well as some initial tests that were carried out to alleviate the preferential orientation difficulties encountered during cryo-EM data collection.

## 9.1. Negative stain EM of the ACAD9-ECSIT-NDUFAF1 triple complex

We first investigated the full ACAD9-ECSIT-NDUFAF1 complex, in order to gain insights into how the MCIA would assemble. The triple complex was co-expressed and purified by Gabriele Giachin in Montserrat Soler-López's group, with several rounds of gel filtration being carried out to achieve a biochemically homogenous complex. We imaged the triple complex by negative stain EM across several sessions from multiple purifications. Each time, several fractions from the gel filtration peak were screened to assess sample homogeneity. 3  $\mu\text{L}$  of sample was applied to the clean side of carbon on a carbon-mica interface and stained with 2% uranyl acetate. Grids were imaged on either a T12 microscope equipped with a Gatan Orius 1000 camera and operated at 120 kV, or an F20 microscope equipped with a Ceta camera and operated at 200 kV.

A representative micrograph is shown in Figure 9.1A. The ACAD9-ECSIT-NDUFAF1 complex appeared as isolated globular particles on the EM grid, but displayed considerable heterogeneity in terms of both size and shape. Some preliminary image processing was carried out on several datasets, and 2D classification confirmed that several different species were present (Figure 9.1B).

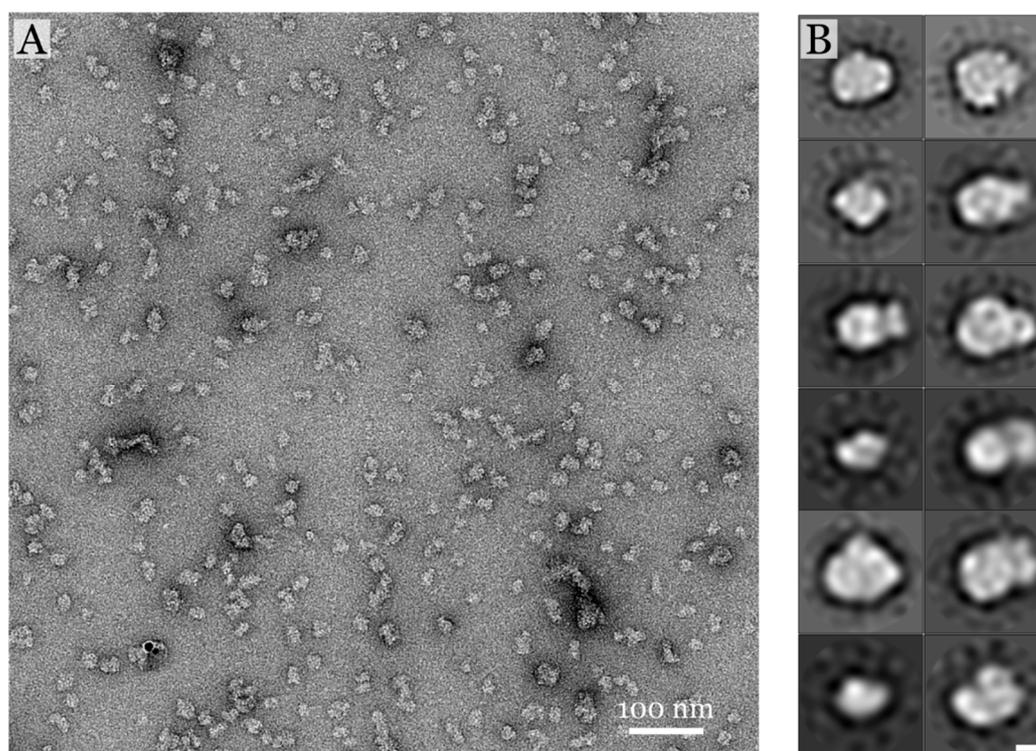


Figure 9.1. **A)** Crop of a negative stain EM micrograph of the ACAD9-ECSIT-NDUFAF1 complex, showing visible heterogeneity. Scale bar = 100 nm. **B)** 2D class averages of particles from the dataset in A). A variety of different species are visible. Scale bar = 100 Å.

Because of the considerable heterogeneity, we decided to use GraFix in an attempt to stabilise a single species. GraFix is a method of preparing samples for electron microscopy which is particularly useful for stabilising and improving the homogeneity of low-affinity complexes (Kastner et al., 2008; Stark, 2010). It involves carrying out ultracentrifugation with both a density gradient and a gradient of a cross-linker at low concentration, as shown in Figure 9.2. In the ultracentrifuge tube, there is a low concentration of glycerol and cross-linker at the top, which increases down the tube. Larger complexes migrate further down the tube during ultracentrifugation and undergo weak intramolecular cross-linking. Fractionation after ultracentrifugation allows the separation of cross-linked species according to size, often improving the sample quality for electron microscopy.

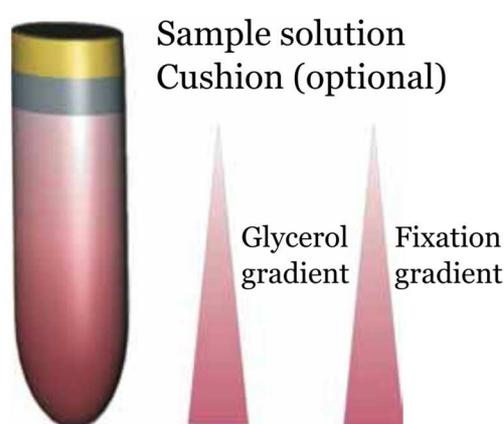


Figure 9.2. Schematic of the gradients used during ultracentrifugation for GraFix. Adapted from (Kastner et al., 2008).

I carried out GraFix several times on the ACAD9-ECSIT-NDUFAF1 complex, trialling different protein concentrations, different glycerol gradients and different concentrations of the cross-linker glutaraldehyde. Glycerol gradients of 10-30% and 10-40% were trialled, with glutaraldehyde gradients ranging from 0-0.025% up to 0-0.1%. The gradients were set up in 800  $\mu$ L ultracentrifugation tubes, and 50  $\mu$ L of protein solution at concentrations of  $\sim$ 0.25 mg/mL or  $\sim$ 0.5 mg/mL was pipetted to the top of the gradient. Ultracentrifugation was carried out for 16 hours at 40,000 rpm at 10°C, and tubes were then fractionated for screening by negative stain EM.

Fractions from the top of the tube consistently contained smaller species compared to fractions further down the tube. Fractions from the bottom of the tube contained mostly larger aggregates, consistent with the higher concentrations of cross-linker in these fractions. We focussed on fractions approximately halfway down the gradient, which contained isolated species most likely corresponding to the full ACAD9-ECSIT-NDUFAF1

complex rather than smaller subcomplexes. Although GraFix improved the overall homogeneity of the sample, there was consistently a mixture of species regardless of the conditions used. Figure 9.3A shows a micrograph from one of the most promising fractions from the GraFix trials. The overall quality of the preparation is better than without GraFix, and the 2D class averages in Figure 9.3B show that the particles have more consistent features and are more homogeneous in both size and shape. However, none of the fractions screened were completely homogeneous, and we therefore did not progress to freezing grids for cryo-EM. Instead, we focussed on the ACAD9-ECSIT<sub>CTD</sub> subcomplex which showed much less heterogeneity in negative stain images.

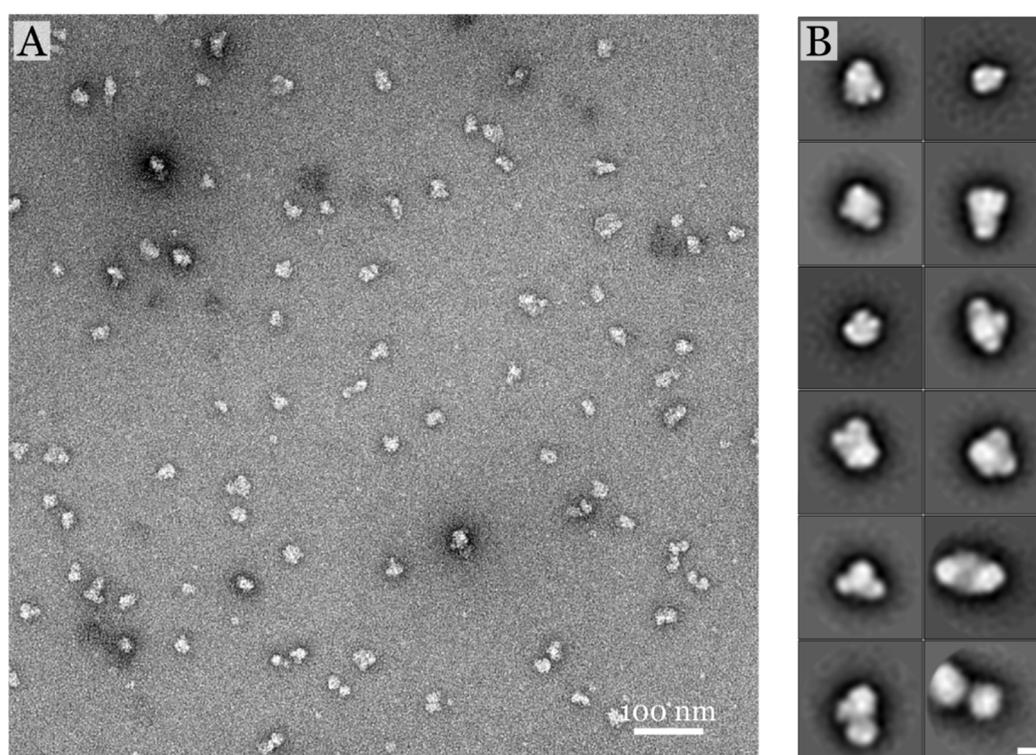


Figure 9.3. **A)** Crop of a negative stain EM micrograph of the ACAD9-ECSIT-NDUFAF1 complex after GraFix. The preparation is more homogeneous than the one shown in Figure 9.1. Scale bar = 100 nm. **B)** 2D class averages of particles from the dataset in A). While the particles are more defined in terms of both size and shape compared to the sample without GraFix, there is still variation in particles size and the sample contains multiple species. Scale bar = 100 Å.

## 9.2. ACAD9-ECSIT<sub>CTD</sub> sample preparation and cryo-EM

In comparison to the full ACAD9-ECSIT-NDUFAF1 complex, the ACAD9-ECSIT<sub>CTD</sub> subcomplex was much more homogeneous in negative stain EM images. However, the key difficulty with characterising the ACAD9-ECSIT<sub>CTD</sub> subcomplex by electron microscopy was overcoming preferential orientation. The extreme preferential orientation of the subcomplex was present in both cryo-EM images and in negative stain EM images. Figure 9.4 shows an example negative stain EM image – while the sample is homogeneous and particles are well dispersed, almost every particle is lying in the same orientation, with the long axis of the particle lying perpendicular to the viewing direction.

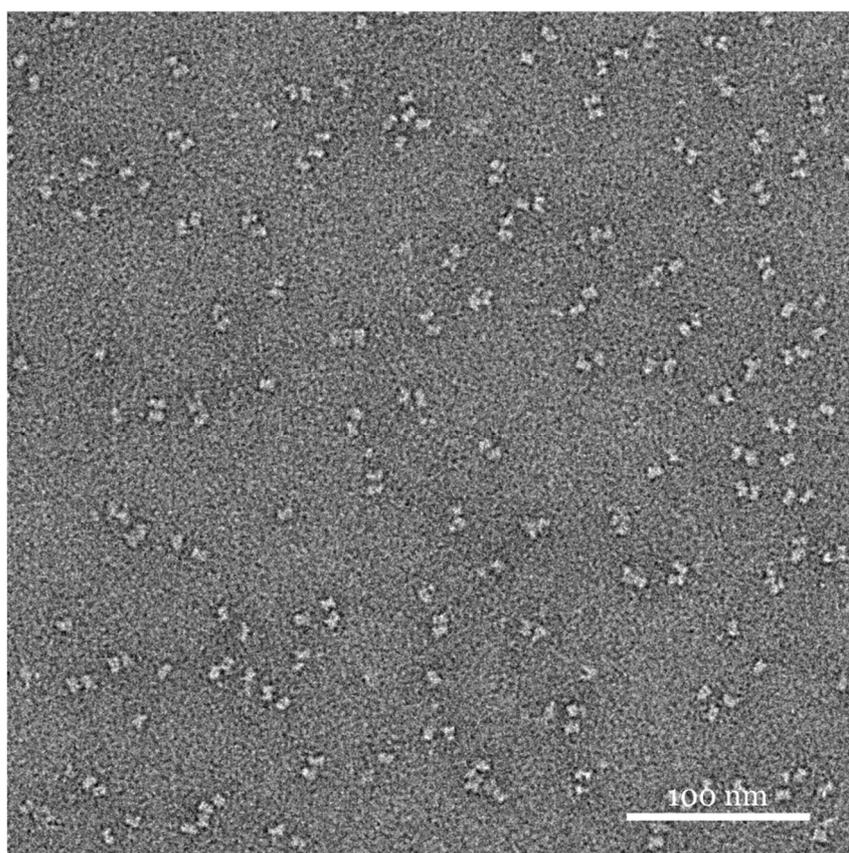


Figure 9.4. Crop of a negative stain EM micrograph of ACAD9-ECSIT<sub>CTD</sub>. The small rectangular particles are much more homogeneous than the full ACAD9-ECSIT-NDUFAF1 triple complex, but show strong preferential orientation. Scale bar = 100 nm.

During grid freezing for cryo-EM, several steps were taken to improve the distribution of particle views in the ice. I carried out glow discharging of cryo-EM grids in the presence of amylamine to modify the surface charge of the grid, which often changes the particle distribution in the ice. Additionally, pumping the cryo-EM grid under vacuum for an hour before freezing was done, which can sometimes also cause the reorientation of particles. However, none of these trials had much effect on the preferential orientation of

particles. Therefore, we decided try tilting the cryo-EM grid for a portion of the data collection because of the success that this strategy for characterising RavA by cryo-EM (see Chapter 4 for details). During two data collections on the ACAD9-ECSIT<sub>CTD</sub> subcomplex, both untilted and 30° tilted images were collected.

For many grids, it was very difficult to find areas containing ice of appropriate thickness for imaging, especially when the grid was tilted. The molecular weight of the ACAD9-ECSIT<sub>CTD</sub> complex is approximately 200 kDa, which is towards the smaller end of the scale for imaging by cryo-EM. Particles could only be visualised in areas containing the thinnest ice, but in these regions the ice was often so thin that it would melt under the electron beam during imaging and cause excessive movement. Therefore, images were not collected in the centre of holes, where the thinnest ice was, but closer to the edge of the holes. Consequently, the areas imaged often had ice thicknesses where the contrast was slightly lower than desired. To try and improve the ice quality, we tried using several different types of grids, with different hole sizes (R 1.2/1.3 and R 2/1) or support materials (copper Quantifoil grids and gold UltrAuFoil grids). After many sessions of freezing and screening grids, the best images for ACAD9-ECSIT<sub>CTD</sub> were obtained on Quantifoil R 2/1 Cu/Rh 300 mesh holey carbon grids after 1 hour of pumping under vacuum, glow discharged at 45 mA for 20 s and frozen at 4°C with a Vitrobot Mark IV.

Data collection was carried out across two sessions on the Titan Krios at ESRF using EPU, resulting in a total of 7,510 micrographs. Images were collected at a nominal magnification of 131,000 x which corresponded to a pixel size of 1.067 Å/pixel at the specimen level. Images consisting of 40 frames were taken with a total electron dose of 41 e<sup>-</sup>/Å<sup>2</sup>. An example image is shown in Figure 9.5.

For image processing, motion correction was carried out using MotionCor2 and the CTF was corrected using GCTF. Micrographs with an estimated resolution in GCTF of better than 8 Å were retained and manually screened by eye, resulting in 2,999 micrographs used for further processing. A subset of micrographs was first manually picked using the Boxer program in EMAN, resulting in 2,901 picked particles. Particles were extracted with a box size of 160 x 160 pixels, and subjected to 2D classification in RELION-2.1. The best 2D classes were then used as templates for auto-picking in RELION, on a larger subset of the data. This resulted in ~490,000 picked particles, which were then extracted and subjected to several rounds of 2D classification.

The best 2D class averages from the RELION auto-picked particles were then used as templates for GPU-accelerated picking in Gautomatch. The coordinates of the ~700,000 picked particles were then used as input for per-particle CTF correction in GCTF. This was

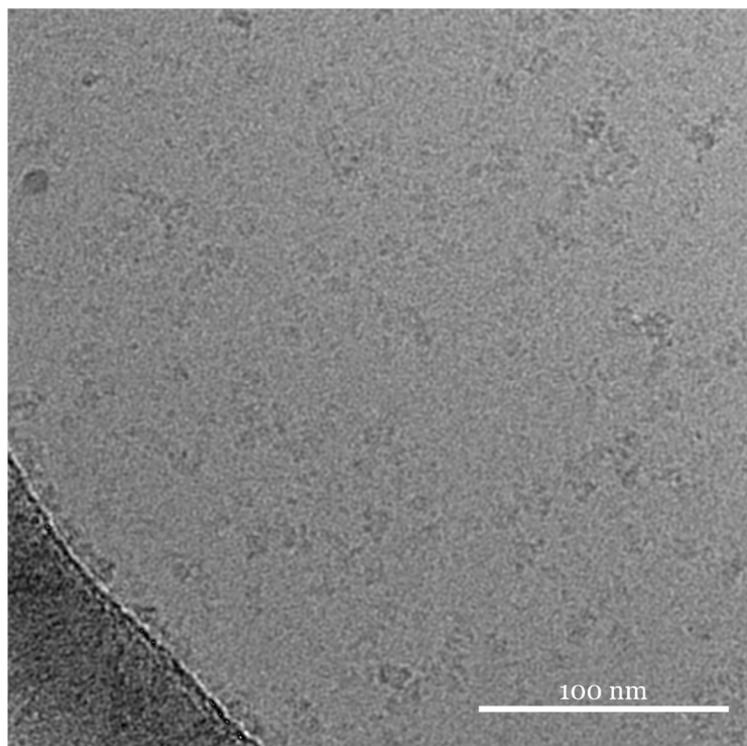


Figure 9.5. Example cryo-EM micrograph of ACAD9-ECSIT<sub>CTD</sub>, scale bar = 100 nm.

carried out to account for particles from tilted micrographs possessing different defocus values. The per-particle CTF-corrected particles were then extracted and subjected to several rounds of 2D classification in RELION. Representative 2D class averages are shown in Figure 9.6. One consistent feature of 2D classes from the two datasets is the presence of a few good classes with clear secondary structural features, which contain the majority of particles. These correspond to the distinctive rectangular view of the complex seen in negative stain EM images.

In the highest occupancy class, shown in the top left in Figure 9.6, a two-fold axis of symmetry is clearly present perpendicular to the image. Therefore, we reasoned that the

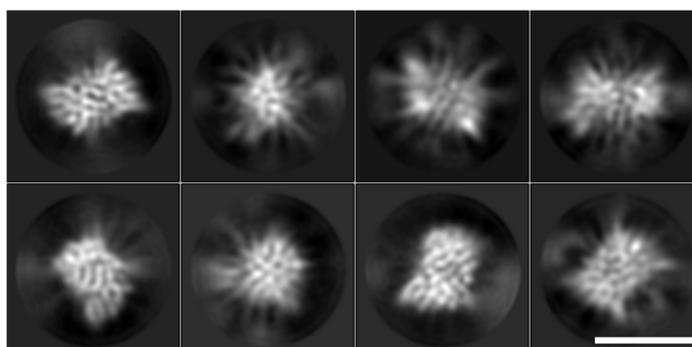


Figure 9.6. Representative 2D class averages from RELION of ACAD9-ECSIT<sub>CTD</sub>, showing the distinctive rectangular particle. Secondary structural features are visible, but the edges of classes show noisy spikes. Scale bar = 100 Å.

core of the class average likely corresponded to the ACAD9 dimer, based on comparisons with the VLCAD-based homology model (see Chapter 8). There were not high-quality 2D class averages corresponding to all views, but classes with clear secondary structural features all seemed to correspond to the same view. Many 2D classes displayed noisy projections at the edge, and it was not clear whether the spiky projections in other classes corresponded to the intrinsically-disordered ECSIT<sub>CTD</sub>, or whether they were artefactual.

Therefore, in order to gain information about the location of ECSIT<sub>CTD</sub>, we moved into 3D analysis. However, this was a difficult and long process – there were significant problems in getting a stable 3D initial model, and 3D refinement usually gave low-quality results, irrespective of whether C2 symmetry was applied or not. I spent many months trying different approaches, including extensive classification in both 2D and 3D to select homogeneous subsets of the data. I trialled different picking algorithms to try and increase the number of side views found, including FPM, Gautomatch, and RELION auto-picking (using templates and template-free approaches). Usually, the resulting 3D maps from refinement after each approach had reported resolutions of ~8-9 Å, but the accuracy of the assignment of Euler angles to particle images was usually estimated to be between 8-10°, which is often an indication that the map may not be correct.

One example of a refinement is shown in Figure 9.7. The overall shape of such volumes were globally consistent with the 2D class averages and the ACAD9 homology model, but apparent high-resolution internal features that appeared in 3D refinements did not match the secondary structural features of the homology model. Because of the high degree of conservation between ACAD9 and VLCAD on which the homology model is based (see Chapter 8), the true secondary structural features of ACAD9 should not be wildly inconsistent with the model.

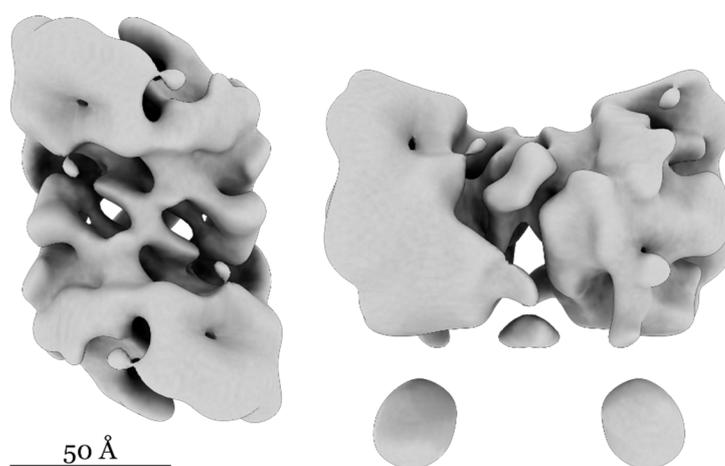


Figure 9.7. One of many 3D refinements from processing in Relion. The global envelope is consistent with the dimensions of the ACAD9 homology model, but internal features are inconsistent. Scale bar = 50 Å.

Because of the extensive difficulties encountered in achieving a stable refinement in RELION, I decided to process the data in CryoSPARC which had given promising results for many other difficult projects. ~580,000 particles after initial cleaning by 2D classification in RELION were imported into CryoSPARC. From here, I tried many different approaches which gave varying results. The two key things which I found to be most useful in CryoSPARC were the ability to finely tune 2D classification parameters, and the *ab initio* 3D reconstruction algorithm, which allows simultaneous generation of multiple initial models and 3D classification between the different models. This allowed me to achieve 3D reconstructions with reasonable secondary structural features, which was not possible using the classification algorithms in RELION. An example of the effects of tuning 2D classification parameters is shown in Figure 9.8. Changing the ‘initial classification uncertainty factor’ changes the amount of clustering of particles into classes. In the RELION class averages shown in Figure 9.6, most of the particles are classified into the top left class. Increasing the uncertainty factor in CryoSPARC increases the diversity of classes, and prevents particles from clustering together as much. This resulted in a larger number of ‘good classes’, and it was possible to identify subtly different views.

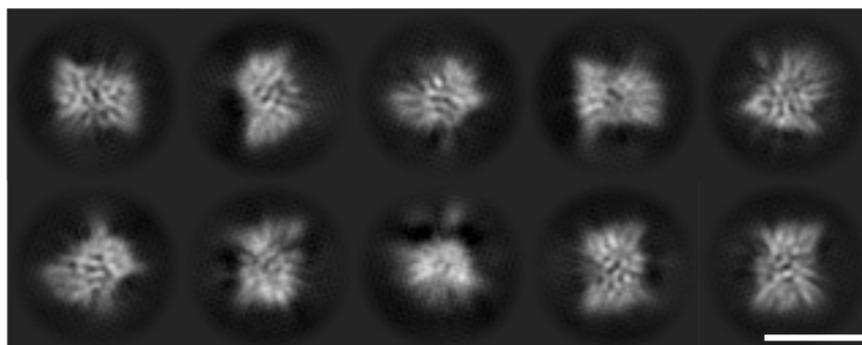


Figure 9.8. Representative 2D class averages from CryoSPARC of ACAD9-ECSIT<sub>CTD</sub>. There is a greater diversity of classes showing secondary structural features, with different (but similar) top views visible. The number of spiky protrusions is decreased compared to the class averages shown in Figure 9.6.

From the hundreds of thousands of particles, it was necessary to find a very small subset of ~10,000-20,000 particles that yielded a map with correct secondary structural features. This is most likely because of two reasons. Firstly, the dominance of top views in the dataset means that most of them had to be discarded to achieve a stable refinement. Secondly, the intrinsic disorder of ECSIT is likely to have caused great difficulty in assigning angles to particle images. The ACAD9 dimer has a molecular weight of 132 kDa, while ECSIT<sub>CTD</sub> has a molecular weight of 22 kDa. As shown in the manuscript presented in Chapter 8, the ACAD9-ECSIT<sub>CTD</sub> complex is composed of an ACAD9 dimer and either three or four copies of ECSIT<sub>CTD</sub>. This means that between 33% and 40% of the total mass of the complex is contributed by ECSIT<sub>CTD</sub>, despite the fact that the final reconstructions

presented in Chapter 8 show features corresponding mostly to the ACAD9 dimer. As shown by NMR in Chapter 8, ECSIT<sub>CTD</sub> only possesses short stretches of secondary structure. Because of the inherent flexibility of the rest of the protein, it is likely that the vast majority averages out in the final cryo-EM density. However, just because the majority of ECSIT<sub>CTD</sub> is not visible in the map does not mean it does not contribute to the misalignment of particles during the 3D reconstruction process.

In the map calculated with a looser mask in Chapter 8, which shows the ACAD9 core and protrusions attributed to the binding of ECSIT<sub>CTD</sub>, the secondary structural features are still not perfectly resolved. Despite this, the quality of the map is sufficient to verify that it is correct, based on the fit of the homology model in the volume and the correct positioning of  $\alpha$ -helices in the density. It is only when masking the ACAD9 core to remove the contribution of ECSIT<sub>CTD</sub> that the map really improved in terms of secondary structural features. The processing steps carried out in CryoSPARC to reach this final map, at a resolution of 7.8 Å, are detailed in Figure 9.9. A first 2D classification was carried out with a high initial uncertainty factor (4, instead of the default 2) to yield many good classes. Because of the low molecular weight of the complex and the low contrast in images, the batch size (i.e. the number of particles used for classification in the initial iterations) was increased from the default 100 to 200. A second round of classification was carried after selecting the best classes, this time with a lower initial uncertainty factor. I then trialled selecting small numbers of 2D classes, showing the best secondary structural features, and carrying out *ab initio* 3D reconstruction on the particles from these classes. I trialled using different numbers of classes in many different combinations, from a single class all the way up to 20 classes. The best results came from selecting five classes showing clean secondary structural features from different views. This corresponded to 16,231 particles, or just ~3% of the initial particle set.

These particles were then used to create an *ab initio* model with applied C2 symmetry. Importantly, changing the starting resolution from the default 35 Å to 12 Å and the final resolution from 12 Å to 7 Å dramatically improved the quality of the initial model, such that it was not significantly worse than the final model. The reason behind this improvement is not certain. One explanation may be that the low-resolution signal which is not taken into account during the initial model generation could in fact correspond to the cloud of flexible ECSIT<sub>CTD</sub> surrounding the ACAD9 core. Lowering the threshold of the initial model revealed large regions of poor quality map which was extremely noisy, possibly corresponding to disordered ECSIT<sub>CTD</sub>. The initial model was therefore used to create a low resolution mask, using a threshold at which only the core of the map was visible. This map was then used during 3D homogeneous refinement, rather than the

default dynamic masking usually used in CryoSPARC. This map showed good consistency with the homology model, as presented in Chapter 8, and allowed the visualisation of the empty FAD binding site.

One of the key benefits to using CryoSPARC for image processing, especially for difficult datasets, is not only the large number of tuneable parameters at each step but the incredible speed at which algorithms run. This simple fact meant that I could really explore

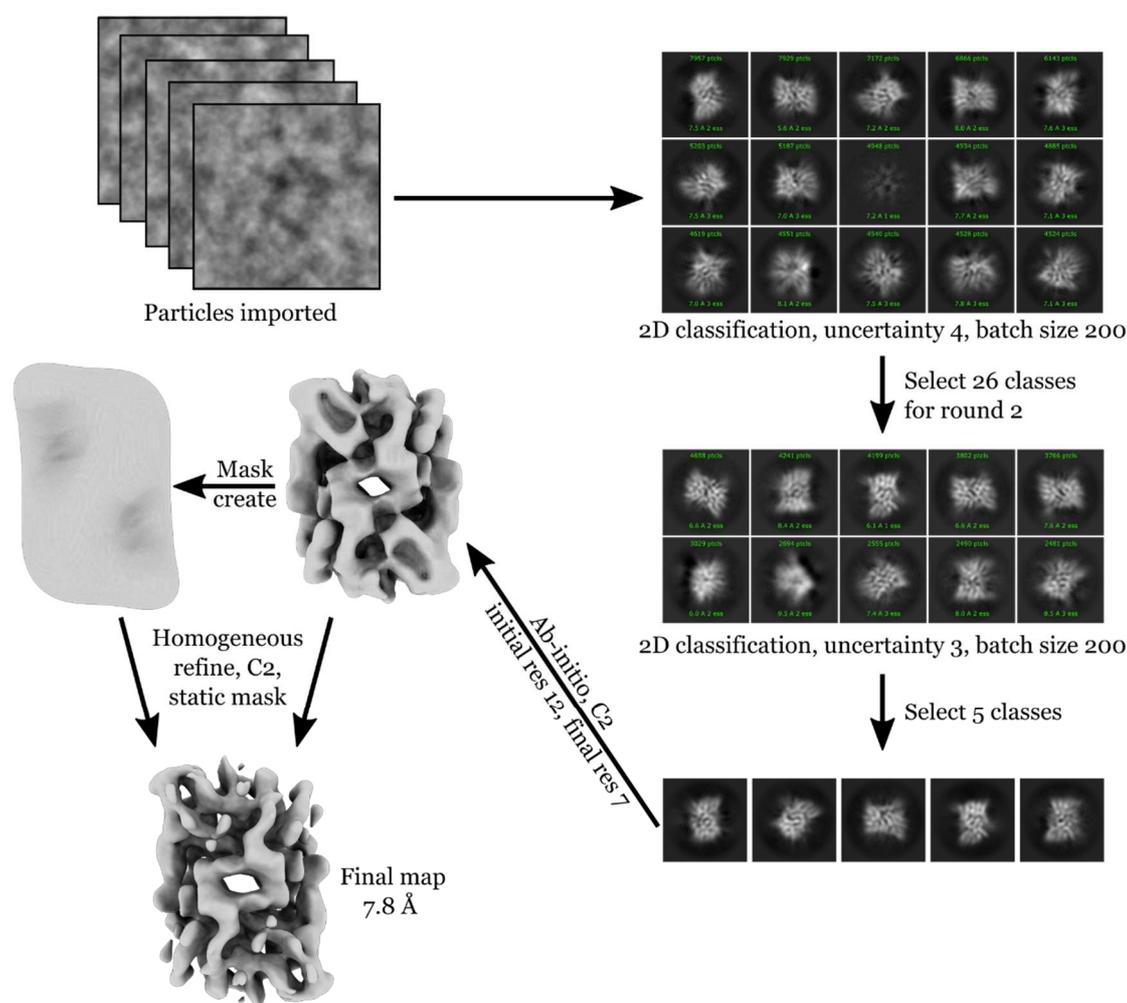


Figure 9.9. Summary of image processing steps, with important changes to default parameters identified, that were carried out to resolve the ACAD9 core map presented in Chapter 8, as described in the text.

a huge number of different combinations of particle subsets and processing tests to find the optimal solution. The processing steps shown in Figure 9.9 represent less than 1% of the things tried for this dataset, and the ability to test such a number of strategies was undoubtedly the reason why I was able to reach the final maps presented in Chapter 8.

### 9.3. Cryo-EM on ACAD9 alone with CHAPSO

In order to overcome the preferential orientation problem, we decided to collect a dataset on ACAD9 alone in the presence of the detergent CHAPSO. As discussed in section 1.6.2, CHAPSO has been used in several cases to improve the distribution of particle orientations for cryo-EM. For simplicity, we decided to try a data collection using CHAPSO for ACAD9 alone first. The addition of detergents can drastically affect the freezing conditions necessary for achieving thin ice in cryo-EM (Chen et al., 2019). In particular the protein concentration required is often much higher than without detergent. We tried freezing grids with three different concentrations, all of them higher than previously used. CHAPSO was added to the ACAD9 sample at a concentration of 8 mM (i.e. at the critical micelle concentration). Sample was frozen on both UltrAuFoil gold grids (R 2/2, 200 mesh) and Quantifoil holey carbon grids (Cu/Rh R 2/1, 300 mesh), with similar parameters for ACAD9-ECSIT<sub>CTD</sub> as described above.

Grids were screened on a Glacios microscope equipped with a Falcon II camera operated at 200 kV. Many grids exhibited the same behaviour as for ACAD9-ECSIT<sub>CTD</sub> – particles were only visible in the very thinnest ice, which was prone to melting under the beam. For most grids, only a subset of the total number of grid squares could be imaged because of this problem. The most promising grid was frozen with the highest concentration of protein on an UltrAuFoil grid. 652 movies were collected at a magnification of 116,086 x, corresponding to a pixel size of 1.206 Å/pixel at the specimen level. Movies were collected with 29 frames and a total electron dose of 45 e<sup>-</sup>/Å<sup>2</sup>, with an exposure time of 1.5 s. An example micrograph is shown in Figure 9.10. While the contrast is good and particles are clearly visible, the majority of them tend to cluster together which may be an effect of CHAPSO addition.

In order to assess whether the preferential orientation was overcome, some initial processing steps were carried out. Motion correction was carried out in MotionCor2, and the CTF was then estimated with GCTF on aligned and dose-weighted sums. Micrographs were then manually screened, with 250 being retained for further processing as the ice thickness in several places was too thick. Manual particle selection was carried out in CryoSPARC on a subset of micrographs, resulting in 1,201 picked particles. Particles were extracted with a box size of 204 x 204 pixels and then subjected to several rounds of 2D classification. The best 2D classes were then used as templates for picking in Gautomatch, resulting in 35,553 picked particles. Particles were extracted in RELION, then imported into CryoSPARC and subjected to 2D classification.

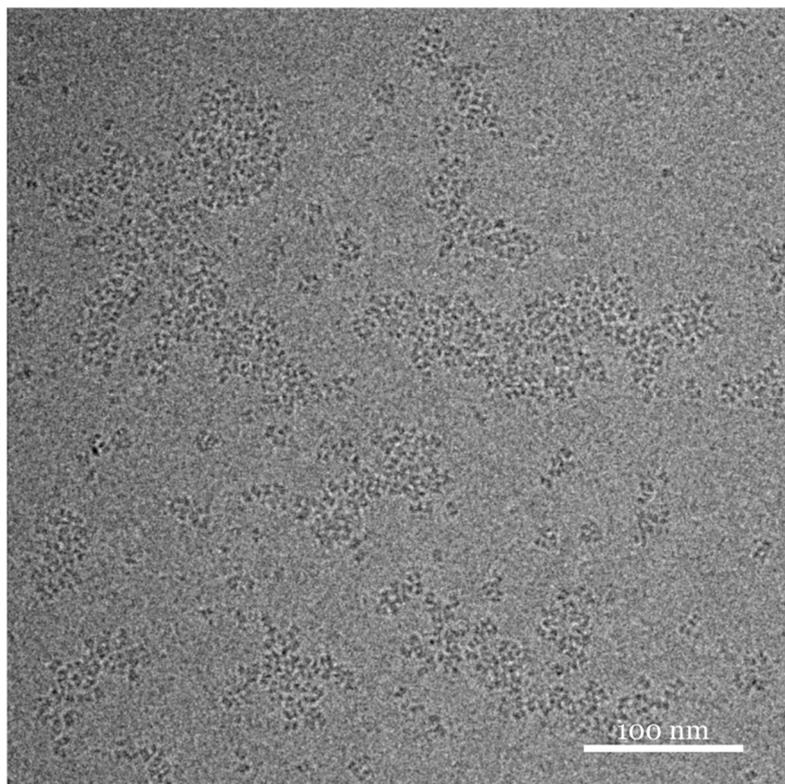


Figure 9.10. Cryo-EM micrograph of ACAD9 in the presence of CHAPSO. The rectangular particles are clearly visible, but have a tendency to cluster together. Scale bar = 100 nm.

Based on the 2D class averages shown in Figure 9.11, it appears that the particle orientation was not significantly affected by the addition of CHAPSO. Despite the improved contrast compared to the ACAD9-ECSIT<sub>CTD</sub> images, the 2D classes are significantly worse, and secondary structural features are barely visible. The likely reason for this is the clustering seen between particles, which is visible in several of the 2D classes and is possibly due to the effects of CHAPSO. Further processing was not carried out on the dataset. Because of time constraints (the data was processed in September 2019), there was not enough time to further optimise conditions and to try different concentrations of detergent. Given more time, a more thorough investigation could have been carried out to improve orientation distribution, as discussed in Chapter 10.

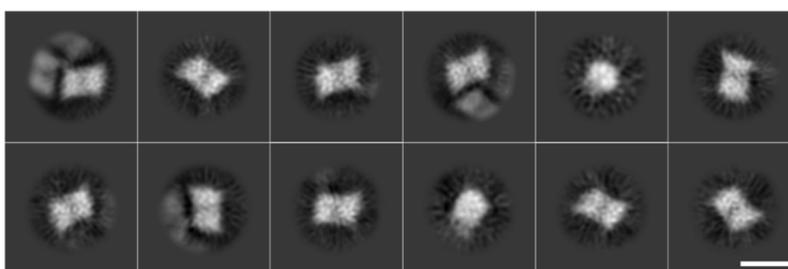


Figure 9.11. 2D class averages of ACAD9 in the presence of CHAPSO. Similarly to the ACAD9-ECSIT<sub>CTD</sub> datasets, the predominant view is along the C2 axis of the ACAD9 dimer with very few end-on views. Scale bar = 100 Å.



# III. DISCUSSION

# Chapter 10. Perspectives and future directions

## 10.1. Improving the RavA reconstruction – overcoming preferential orientation

As discussed in Chapter 5, all three datasets collected (RavA + ADP, RavA + ATP $\gamma$ S, and RavA + ATP $\gamma$ S + casein) suffered from significant preferential orientation that resulted in anisotropy in 3D reconstructions. Tilting the grids during data collection improved the angular distribution of particle views, and the z-stretching could be improved by many rounds of classification during image processing, both in 2D and 3D, but these steps did not remove anisotropy completely. For the RavA + ADP dataset, the maps were good enough to allow confident fitting of the RavA crystal structure and subsequent interpretation of structural features, especially for the C2 map for which anisotropy could be largely eliminated. However, for both the RavA + ATP $\gamma$ S and RavA + ATP $\gamma$ S + casein datasets the separation of different states in 3D was not fully satisfactory. This is likely because of the additional heterogeneity seen in conformational states with ATP $\gamma$ S instead of ADP – while it is easy to classify top views corresponding to the C2, spiral, C6 and C7 states (see Supplementary Figure 7 in Chapter 4), side views corresponding to these states are very similar and limited in number.

Given more time, there are several steps which could have been taken to improve the distribution of particle views, especially for the datasets with ATP $\gamma$ S. Collecting datasets with detergents such as CHAPSO added (as discussed in section 1.6.2) may have improved the distribution of orientations, as well as trying grid supports such as carbon or graphene films. The key limiting factor for trying these approaches was time. Each new condition would have required fresh RavA purifications, and the screening of grid freezing parameters (which may have differed significantly) for each new condition. In addition, the size of the RavA hexamer (~350 kDa) means that it is difficult to screen grids on low-end microscopes without a direct detector, and so the time spent waiting for access to high-end microscopes is also a limitation. The two datasets with ATP $\gamma$ S were collected in June 2019, towards the end of my PhD, and there was not enough time to fully explore conditions to optimise the samples for high-resolution data collection. These avenues can perhaps be further explored in the future, and will help to shed more light on the mechanism of RavA ATPase activity.

## 10.2. The search for the biological function of the LdcI-RavA-ViaA triad

As discussed in Chapter 3, during the first year of my PhD we were unable to confirm any of the proposed substrate interactions with the LdcI-RavA-ViaA triad. Because of this, the majority of my thesis was dedicated to structural characterisation of proteins of triad itself. In addition to the work presented here on RavA and LdcI, I also worked in collaboration with my colleague Jan Felix on the structural characterisation of ViaA by SAXS and negative stain EM. However, there is ongoing work in both the Gutsche lab and by our collaborators to investigate the biological function of LdcI, RavA and ViaA.

The manuscript presented in Chapter 6 presents fresh data on how LdcI may form stacks for mitigating acid stress in the cell, and the identified structural basis of stack dissociation by ppGpp is intriguing. Until now, the only identified function of RavA is to modulate the ppGpp inhibition of LdcI activity (Kanjee et al., 2011b). However, the relationship between RavA and the ppGpp-based inhibition of LdcI is still unclear. It is hard to imagine that this would be the only function of RavA, as this function seems to not rely on the ATPase activity of RavA (El Bakkouri et al., 2010). The interaction between RavA and ViaA, analogously to other MoxR AAA+ ATPases and their respective VWA counterparts, is likely to be functionally important. We have an ongoing collaboration with Frédéric Barras at the Institut Pasteur on the LdcI-RavA-ViaA triad, and further work will soon begin in his lab to identify potential substrates using genetic approaches. In addition, there are exciting ongoing developments in the Gutsche lab which have defined new avenues of investigation for the functional roles of RavA and ViaA.

Another interesting idea would be to analyse the RavA pore loops, and to systematically compare these to pore loops of other AAA+ ATPases. Key conserved residues in the pore loops often directly relate to the nature of substrates. For example, DNA-interacting ATPases often have positive residues at the ends of pore loops for interaction with negatively-charged DNA (Meagher et al., 2019). Similarly, katanin and spastin have positively-charged pore loops for binding to the negatively-charged C-terminal tails of tubulin (White et al., 2007). Many clade 3 AAA+ ATPases have aromatic pore loop residues for intercalating between hydrophobic residues in unfolded substrates (Puchades et al., 2020). The H2-insert is not fully modelled in the RavA crystal structure, but from the cryo-EM density it is clear that they project into the centre of the pore and would be responsible for binding the substrate during translocation. One interesting avenue of investigation would be to model in the missing H2-insert loop residues based on comparison with other AAA+ ATPases to see if this gives clues about the possible nature of the substrate.

### 10.3. Functional implications of LdcI stack formation

As presented in Chapter 6, cryo-EM analysis of the LdcI stack structure allowed the identification of key residues involved in the inter-decamer interaction upon formation of stacks at low pH. Mutating these residues has a direct effect on the ability of LdcI to form stacks, and the R468E mutation completely abolishes stack formation. The next steps for this project are to incorporate these mutations into the *E. coli* chromosome to carry out both functional studies and fluorescence imaging. This should directly reveal whether such mutations have an effect on the acid stress response of *E. coli*, and may confirm that LdcI condensation into patches in the cell is related to its metabolic function.

Another key experiment for the future is carrying out fluorescence microscopy with dyes against both LdcI and cardiolipin, which is present in lipid microdomains in *E. coli*. If the LdcI patches co-localise with lipid microdomains, this would be strong evidence to support the hypothesis that LdcI would be attracted to proton sinks near the membrane for the mitigation of acid stress. In addition, the co-staining of LdcI and respiratory Complex I could be carried out for a similar purpose.

Furthermore, carrying out fluorescence microscopy experiments in cells impaired in the production of ppGpp should be carried out to fully investigate the role of ppGpp-based inhibition of LdcI activity and stack disruption.

Eventually, one potential avenue of investigation may be the characterisation of the LdcI condensates by correlative fluorescence microscopy and cryo-electron tomography. Such an approach may reveal valuable structural and functional insights into the function of LdcI as an acid stress response protein, and add to the growing body of work investigating the spatial regulation of metabolic enzymes.

## 10.4. MCIA complex structural characterisation

Similarly to RavA, the ACAD9-ECSIT<sub>CTD</sub> subcomplex of the MCIA complex suffered from severe preferential orientation on the cryo-EM grid. There are several things which could be attempted to alleviate the problems encountered. As discussed for RavA in section 10.1, further investigation of different grid freezing conditions could be carried out to alleviate this preferential orientation. Although we tried freezing grids on ACAD9 in the presence of CHAPSO, we were not able to carry out extensive trials and further investigation needs to be done.

Another possible solution could be trialling the use of graphene films on cryo-EM grids. As discussed in section 1.6.2, using support films can cause significant reorientation of particles on the grid. Graphene films have been used recently to provide high-resolution structural insights into streptavidin, a 52 kDa protein, using cryo-EM (Fan et al., 2019; Han et al., 2020). For the 132 kDa ACAD9 dimer, following such an approach might lead to improvement in the cryo-EM map.

The intrinsic disorder of ECSIT<sub>CTD</sub> was another factor which complicated the cryo-EM analysis of the ACAD9-ECSIT<sub>CTD</sub> subcomplex. The large amount of disorder in the subcomplex meant that achieving good angular accuracy of particle alignment was difficult, and lowered the final quality of the cryo-EM reconstruction. Despite this, we managed to identify a potential binding site on ACAD9 that ECSIT<sub>CTD</sub> would recognise that fits well with bioinformatic analysis of ACAD9 and its close homologue VLCAD. Further steps, such as cross-linking mass spectrometry, should be taken to validate this binding interface.

One possible way of improving the cryo-EM reconstruction of ACAD9-ECSIT<sub>CTD</sub> is to try and identify the minimal construct of ECSIT<sub>CTD</sub> that would still recognise ACAD9. Carrying out cryo-EM investigation with a smaller ECSIT<sub>CTD</sub> construct would reduce the contribution from disordered regions to the final reconstruction, and should significantly improve the cryo-EM map.

For structural characterisation of the full ACAD9-ECSIT-NDUFAF1 triad, the GraFix trials presented in section 9.1 may be worth revisiting. These experiments were carried out in December 2017 – since then, the ability of cryo-EM image processing packages to deal with both conformational and compositional heterogeneity has improved significantly, as discussed in Chapter 1. While there was still visible heterogeneity in terms of particle size in the best preparation, the most recent classification algorithms may be able to separate out different species and provide insight into the ACAD9-ECSIT-NDUFAF1 complex structure.

In the next few months, a Masters student will jointly work with the Gutsche and Soler-López groups on the structural characterisation of the MCIA. This will begin with the improvement of the cryo-EM reconstruction of ACAD9 itself, by extending the work started here on overcoming the difficulties with preferential orientation. Later, there are plans for a PhD student to continue working on both the ACAD9-ECSIT<sub>CTD</sub> subcomplex and the full ACAD9-ECSIT-NDUFAF1 complex. In combination with the exciting functional work being carried out by the Soler-López group, these future studies should shed light on how the MCIA complex assembles and how this relates to its functional role of Complex I maturation in mitochondria.

## IV. BIBLIOGRAPHY

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., et al. (2010). PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* *66*, 213–221.

Adrian, M., Dubochet, J., Lepault, J., and McDowell, A.W. (1984). Cryo-electron microscopy of viruses. *Nature* *308*, 32–36.

Ahmadian, M.R., Stege, P., Scheffzek, K., and Wittinghofer, A. (1997). Confirmation of the arginine-finger hypothesis for the GAP-stimulated GTP-hydrolysis reaction of Ras. *Nat. Struct. Biol.* *4*, 686–689.

Ammelburg, M., Frickey, T., and Lupas, A.N. (2006). Classification of AAA+ proteins. *J. Struct. Biol.* *156*, 2–11.

Anand, B., Majumdar, S., and Prakash, B. (2013). Structural basis unifying diverse GTP hydrolysis mechanisms. *Biochemistry* *52*, 1122–1130.

Aravind, L., Iyer, L.M., Leipe, D.D., and Koonin, E. V. (2004). A novel family of P-loop NTPases with an unusual phyletic distribution and transmembrane segments inserted within the NTPase domain. *Genome Biol.* *5*.

Arias-Palomo, E., Puri, N., O’Shea Murray, V.L., Yan, Q., and Berger, J.M. (2019). Physical Basis for the Loading of a Bacterial Replicative Helicase onto DNA. *Mol. Cell* *74*, 173–184.e4.

Armstrong, M., Han, B.-G., Gomez, S., Turner, J., Fletcher, D.A., and Glaeser, R.M. (2019). Microscale Fluid Behavior during Cryo-EM Sample Blotting. *Biophys. J.* 791285.

Arnold, S.A., Albiez, S., Bieri, A., Syntychaki, A., Adaixo, R., McLeod, R.A., Goldie, K.N., Stahlberg, H., and Braun, T. (2017). Blotting-free and lossless cryo-electron microscopy grid preparation from nanoliter-sized protein samples and single-cell extracts. *J. Struct. Biol.* *197*, 220–226.

Babu, M., Arnold, R., Bundalovic-Torma, C., Gagarinova, A., Wong, K.S., Kumar, A., Stewart, G., Samanfar, B., Aoki, H., Wagih, O., et al. (2014). Quantitative genome-wide genetic interaction screens reveal global epistatic relationships of protein complexes in *Escherichia coli*. *PLoS Genet.* *10*, e1004120.

Bai, X., McMullan, G., and Scheres, S.H.. (2015a). How cryo-EM is revolutionizing structural biology. *Trends Biochem. Sci.* *40*, 49–57.

Bai, X.C., Rajendra, E., Yang, G., Shi, Y., and Scheres, S.H.W. (2015b). Sampling the conformational space of the catalytic subunit of human  $\gamma$ -secretase. *Elife* *4*, 1–19.

El Bakkouri, M., Gutsche, I., Kanjee, U., Zhao, B., Yu, M., Goret, G., Schoehn, G., Burmeister, W.P., and Houry, W.A. (2010). Structure of RavA MoxR AAA+ protein reveals the design principles of a molecular cage modulating the inducible lysine decarboxylase activity. *Proc. Natl. Acad. Sci.* *107*, 22499–22504.

Baldwin, P.R., Tan, Y.Z., Eng, E.T., Rice, W.J., Noble, A.J., Negro, C.J., Cianfrocco, M.A.,

Potter, C.S., and Carragher, B. (2018). Big data in cryoEM: automated collection, processing and accessibility of EM data. *Curr. Opin. Microbiol.* *43*, 1–8.

Banerjee, P., Chanchal, and Jain, D. (2019). Sensor I Regulated ATPase Activity of FleQ Is Essential for Motility to Biofilm Transition in *Pseudomonas aeruginosa*. *ACS Chem. Biol.* *14*, 1515–1527.

Bartesaghi, A., Aguerrebere, C., Falconieri, V., Banerjee, S., Earl, L.A., Zhu, X., Grigorieff, N., Milne, J.L.S., Sapiro, G., Wu, X., et al. (2018). Atomic Resolution Cryo-EM Structure of  $\beta$ -Galactosidase. *Structure* *26*, 848–856.e3.

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

Bepler, T., Noble, A.J., and Berger, B. (2019b). Topaz-Denoise: general deep denoising models for cryoEM. *bioRxiv* 838920.

Berry, J.L., and Pelicic, V. (2015). Exceptionally widespread nanomachines composed of type IV pilins: The prokaryotic Swiss Army knives. *FEMS Microbiol. Rev.* *39*, 134–154.

Blesa, A., Quintans, N.G., Baquedano, I., Mata, C.P., Castón, J.R., and Berenguer, J. (2017). Role of archaeal HerA protein in the biology of the bacterium *thermus thermophilus*. *Genes (Basel)* *8*.

Blok, N.B., Tan, D., Wang, R.Y.-R., Penczek, P.A., Baker, D., DiMaio, F., Rapoport, T.A., and Walz, T. (2015). Unique double-ring structure of the peroxisomal Pex1/Pex6 ATPase complex revealed by cryo-electron microscopy. *Proc. Natl. Acad. Sci.* *112*, E4017–E4025.

Bolanos-Garcia, V.M., and Davies, O.R. (2006). Structural analysis and classification of native proteins from *E. coli* commonly co-purified by immobilised metal affinity chromatography. *Biochim. Biophys. Acta - Gen. Subj.* *1760*, 1304–1313.

Bordes, P., Wigneshweraraj, S.R., Schumacher, J., Zhang, X., Chaney, M., and Buck, M. (2003). The ATP hydrolyzing transcription activator phage shock protein F of *Escherichia coli*: Identifying a surface that binds  $\sigma_{54}$ . *Proc. Natl. Acad. Sci.* *100*, 2278–2283.

Bos, J.L., Rehmann, H., and Wittinghofer, A. (2007). GEFs and GAPs: Critical Elements in the Control of Small G Proteins. *Cell* *129*, 865–877.

Bowman, G.D., O'Donnell, M., and Kuriyan, J. (2004). Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex. *Nature* *429*, 724–730.

Bowman, G.D., Goedken, E.R., Kazmirski, S.L., Donnell, M.O.Ö., and Kuriyan, J. (2005). DNA polymerase clamp loaders and DNA recognition. *FEBS Lett.* *579*, 863–867.

Bracale, A., Cesca, F., Neubrand, V.E., Newsome, T.P., Way, M., and Schiavo, G. (2007). Kidins220/ARMS is transported by a kinesin-1-based mechanism likely to be involved in neuronal differentiation. *Mol. Biol. Cell* *18*, 142–152.

Bracewell, R. (1956). Strip Integration in Radio Astronomy. *Aust. J. Phys.* *9*, 198.

Braun, M.B., Traenkle, B., Koch, P.A., Emele, F., Weiss, F., Poetz, O., Stehle, T., and Rothbauer, U. (2016). Peptides in headlock - A novel high-affinity and versatile peptide-binding nanobody for proteomics and microscopy. *Sci. Rep.* *6*.

Brenner, S., and Horne, R.W. (1959). A negative staining method for high resolution electron microscopy of viruses. *Biochim. Biophys. Acta* *34*, 103–110.

Brilot, A.F., Chen, J.Z., Cheng, A., Pan, J., Harrison, S.C., Potter, C.S., Carragher, B., Henderson, R., and Grigorieff, N. (2012). Beam-induced motion of vitrified specimen on holey carbon film. *J. Struct. Biol.* *177*, 630–637.

Brunger, A.T., and DeLaBarre, B. (2003). NSF and p97/VCP: Similar at first, different at last. In *FEBS Letters*, pp. 126–133.

Buchholz, T.-O., Jordan, M., Pigino, G., and Jug, F. (2019). Cryo-CARE: Content-Aware Image Restoration for Cryo-Transmission Electron Microscopy Data. In 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), (IEEE), pp. 502–506.

Cai, S., Cai, J., Jiang, W.G., and Ye, L. (2017). Kidins220 and tumour development: Insights into a complexity of cross-talk among signalling pathways (Review). *Int. J. Mol. Med.* *40*, 965–971.

Carragher, B., Kisseberth, N., Kriegman, D., Milligan, R.A., Potter, C.S., Pulokas, J., and Reilein, A. (2000). Leginon: An automated system for acquisition of images from vitreous ice specimens. *J. Struct. Biol.* *132*, 33–45.

Carriel, D., Garcia, P.S., Castelli, F., Lamourette, P., Fenaille, F., Brochier-Armanet, C., Elsen, S., and Gutsche, I. (2018). A novel subfamily of bacterial AAT-fold basic amino acid decarboxylases and functional characterization of its first representative: *Pseudomonas aeruginosa* LdcA. *Genome Biol. Evol.* *10*, 3058–3075.

Carter, A.P. (2013). Crystal clear insights into how the dynein motor moves. *J. Cell Sci.* *126*, 705–713.

Chang, C.-W., Lee, S., and Tsai, F.T.F. (2017). Structural Elements Regulating AAA+ Protein Quality Control Machines. *Front. Mol. Biosci.* *4*, 1–5.

Chang, Y.P., Xu, M., Machado, A.C.D., Yu, X.J., Rohs, R., and Chen, X.S. (2013). Mechanism of Origin DNA Recognition and Assembly of an Initiator-Helicase Complex by SV40 Large Tumor Antigen. *Cell Rep.* *3*, 1117–1127.

Chen, J., Noble, A.J., Kang, J.Y., and Darst, S.A. (2019). Eliminating effects of particle adsorption to the air/water interface in single-particle cryo-electron microscopy: Bacterial RNA polymerase and CHAPSO. *J. Struct. Biol.* *X 1*, 100005.

Cheng, Y. (2018). Single-particle cryo-EM—How did it get here and where will it go. *Science (80-. )*. *361*, 876–880.

Cheng, A., Eng, E.T., Alink, L., Rice, W.J., Jordan, K.D., Kim, L.Y., Potter, C.S., and Carragher, B. (2018). High resolution single particle cryo-electron microscopy using beam-image shift. *J. Struct. Biol.* *204*, 270–275.

Cheng, Y., Grigorieff, N., Penczek, P.A., and Walz, T. (2015). A Primer to Single-Particle Cryo-Electron Microscopy. *Cell* *161*, 438–449.

Cho, C., Jang, J., Kang, Y., Watanabe, H., Uchihashi, T., Kim, S.J., Kato, K., Lee, J.Y., and Song, J.J. (2019). Structural basis of nucleosome assembly by the Abo1 AAA+ ATPase histone chaperone. *Nat. Commun.* *10*, 1–13.

Cox, M.M. (2007). Motoring along with the bacterial RecA protein. *Nat. Rev. Mol. Cell Biol.* *8*, 127–138.

Crowther, R.A. (1971). Procedures for Three-Dimensional Reconstruction of Spherical Viruses by Fourier Synthesis from Electron Micrographs. *Philos. Trans. R. Soc. B Biol. Sci.* *261*, 221–230.

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*, 421–425.

D’Imprima, E., Floris, D., Joppe, M., Sánchez, R., Grininger, M., and Kühlbrandt, W. (2019). Protein denaturation at the air-water interface and how to prevent it. *Elife* *8*, e42747.

Dandey, V.P., Wei, H., Zhang, Z., Tan, Y.Z., Acharya, P., Eng, E.T., Rice, W.J., Kahn, P.A., Potter, C.S., and Carragher, B. (2018). Spotiton: New features and applications. *J. Struct. Biol.* *202*, 161–169.

Danev, R. (2019). Electrons see the light. *Nat. Methods* *16*, 966–967.

Danev, R., and Baumeister, W. (2017). Expanding the boundaries of cryo-EM with phase plates. *Curr. Opin. Struct. Biol.* *46*, 87–94.

Danev, R., Buijsse, B., Khoshouei, M., Plitzko, J.M., and Baumeister, W. (2014). Volta potential phase plate for in-focus phase contrast transmission electron microscopy. *Proc. Natl. Acad. Sci.* *111*, 15635–15640.

Danev, R., Yanagisawa, H., and Kikkawa, M. (2019). Cryo-Electron Microscopy Methodology: Current Aspects and Future Directions. *Trends Biochem. Sci.* *44*, 837–848.

Daumke, O., Weyand, M., Chakrabarti, P.P., Vetter, I.R., and Wittinghofer, A. (2004). The GTPase-activating protein Rap1GAP uses a catalytic asparagine. *Nature* *429*, 197–201.

Davidson, A.L., Dassa, E., Orelle, C., and Chen, J. (2008). Structure, Function, and Evolution of Bacterial ATP-Binding Cassette Systems. *Microbiol. Mol. Biol. Rev.* *72*, 317–364.

Dean, M., and Annilo, T. (2005). Evolution of the Atp-Binding Cassette (Abc) Transporter Superfamily in Vertebrates. *Annu. Rev. Genomics Hum. Genet.* *6*, 123–142.

Denise, R., Abby, S.S., and Rocha, E.P.C. (2019). Diversification of the type IV filament superfamily into machines for adhesion, protein secretion, DNA uptake, and motility. *PLOS Biol.*

Desfosses, A., Venugopal, H., Joshi, T., Felix, J., Jessop, M., Jeong, H., Hyun, J., Heymann, J.B., Hurst, M.R.H., Gutsche, I., et al. (2019). Atomic structures of an entire contractile injection system in both the extended and contracted states. *Nat. Microbiol.* *4*, 1885–1894.

Deville, C., Franke, K., Mogk, A., Bukau, B., and Saibil, H.R. (2019). Two-Step Activation Mechanism of the ClpB Disaggregase for Sequential Substrate Threading by the Main ATPase Motor. *Cell Rep.* *27*, 3433–3446.

Dobro, M.J., Melanson, L.A., Jensen, G.J., and McDowell, A.W. (2010). Plunge freezing for electron cryomicroscopy (Elsevier Inc.).

Drulyte, I., Johnson, R.M., Hesketh, E.L., Hurdiss, D.L., Scarff, C.A., Porav, S.A., Ranson, N.A., Muench, S.P., and Thompson, R.F. (2018). Approaches to altering particle distributions in cryo-electron microscopy sample preparation. *Acta Crystallogr. Sect. D Struct. Biol.* *74*, 560–571.

Dubochet, J., and McDowell, A.W. (1981). VITRIFICATION OF PURE WATER FOR ELECTRON MICROSCOPY. *J. Microsc.* *124*, 3–4.

Dubochet, J., Lepault, J., Freeman, R., Berriman, J.A., and Homo, J.-C. (1982). Electron microscopy of frozen water and aqueous solutions. *J. Microsc.* *128*, 219–237.

Dubochet, J., Adrian, M., Lepault, J., and McDowell, A.W. (1985). Emerging techniques: Cryo-electron microscopy of vitrified biological specimens. *Trends Biochem. Sci.* *10*, 143–146.

Dubochet, J., Adrian, M., Chang, J.-J., Homo, J.-C., Lepault, J., McDowell, A.W., and Schultz, P. (1988). Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophys.* *21*, 129–228.

Duderstadt, K.E., Chuang, K., and Berger, J.M. (2011). DNA stretching by bacterial initiators promotes replication origin opening. *Nature* *478*, 209–213.

Eisenstein, M. (2015). The field that came in from the cold. *Nat. Methods* *13*, 19–22.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.*

Enemark, E.J., and Joshua-Tor, L. (2006). Mechanism of DNA translocation in a replicative hexameric helicase. *Nature* *442*, 270–275.

Erhardt, H., Steimle, S., Muders, V., Pohl, T., Walter, J., and Friedrich, T. (2012). Disruption of individual nuo-genes leads to the formation of partially assembled NADH:ubiquinone oxidoreductase (complex I) in Escherichia coli. *Biochim. Biophys. Acta* *1817*, 863–871.

Erickson, H.P., and Klug, A. (1970). The Fourier transform of an electron micrograph: effects of defocussing and aberrations, and implications for the use of underfocus contrast enhancement. *Berichte Der Bunsengesellschaft Für Phys. Chemie* *74*, 1129–1137.

Erzberger, J.P., and Berger, J.M. (2006). Evolutionary Relationships and Structural Mechanisms of AAA+ Proteins. *Annu. Rev. Biophys. Biomol. Struct.* *35*, 93–114.

- Estrozi, L.F., and Navaza, J. (2008). Fast projection matching for cryo-electron microscopy image reconstruction. *J. Struct. Biol.* *162*, 324–334.
- Fairman-Williams, M.E., Guenther, U.-P., and Jankowsky, E. (2010). SF1 and SF2 : family matters. *Curr. Opin. Struct. Biol.* *20*, 313–324.
- Falcon, B., Zivanov, J., Zhang, W., Murzin, A.G., Garringer, H.J., Vidal, R., Crowther, R.A., Newell, K.L., Ghetti, B., Goedert, M., et al. (2019). Novel tau filament fold in chronic traumatic encephalopathy encloses hydrophobic molecules. *Nature* *568*, 420–423.
- Fan, X., Wang, J., Zhang, X., Yang, Z., Zhang, J.C., Zhao, L., Peng, H.L., Lei, J., and Wang, H.W. (2019). Single particle cryo-EM reconstruction of 52 kDa streptavidin at 3.2 Angstrom resolution. *Nat. Commun.* *10*, 1–11.
- Fei, X., Bell, T.A., Jenni, S., Stinson, B.M., Baker, T.A., Harrison, S.C., and Sauer, R.T. (2020). Structures of the ATP-fueled ClpXP proteolytic machine bound to protein substrate. *Elife* *9*, 1–22.
- Feng, X., Fu, Z., Kaledhonkar, S., Jia, Y., Shah, B., Jin, A., Liu, Z., Sun, M., Chen, B., Grassucci, R.A., et al. (2017). A Fast and Effective Microfluidic Spraying-Plunging Method for High-Resolution Single-Particle Cryo-EM. *Structure* *25*, 663–670.e3.
- Fernandez-Leiro, R., and Scheres, S.H.W. (2017). A pipeline approach to single-particle processing in RELION. In *Acta Crystallographica Section D: Structural Biology*, pp. 496–502.
- Fischer, H., and Glockshuber, R. (1994). A point mutation within the ATP-binding site inactivates both catalytic functions of the ATP-dependent protease La (Lon) from *Escherichia coli*. *FEBS Lett.* *356*, 101–103.
- Fodje, M.N., Hansson, A., Hansson, M., Olsen, J.G., Gough, S., Willows, R.D., and Al-Karadaghi, S. (2001). Interplay between an AAA module and an integrin I domain may regulate the function of magnesium chelatase. *J. Mol. Biol.* *311*, 111–122.
- Fontana, J., Jurado, K.A., Cheng, N., Ly, N.L., Fuchs, J.R., Gorelick, R.J., Engelman, A.N., and Steven, A.C. (2015). Distribution and Redistribution of HIV-1 Nucleocapsid Protein in Immature, Mature, and Integrase-Inhibited Virions: a Role for Integrase in Maturation. *J. Virol.* *89*, 9765–9780.
- Frank, J. (1975). Averaging of low exposure electron micrographs of non-periodic objects. *Ultramicroscopy* *1*, 159–162.
- Frickey, T., and Lupas, A.N. (2004). Phylogenetic analysis of AAA proteins. *J. Struct. Biol.* *146*, 2–10.
- Frigola, J., He, J., Kinkelin, K., Pye, V.E., Renault, L., Douglas, M.E., Remus, D., Cherepanov, P., Costa, A., and Diffley, J.F.X. (2017). Cdt1 stabilizes an open MCM ring for helicase loading. *Nat. Commun.* *8*, 15720.
- Fu, Z., Indrisiunaite, G., Kaledhonkar, S., Shah, B., Sun, M., Chen, B., Grassucci, R.A., Ehrenberg, M., and Frank, J. (2019). The structural basis for release-factor activation during translation termination revealed by time-resolved cryogenic electron microscopy. *Nat. Commun.* *10*, 1–7.
- Gadsby, D.C., Vergani, P., and Csanády, L. (2006). The ABC protein turned chloride channel whose failure causes cystic fibrosis. *Nature* *440*, 477–483.
- Gale, E.F. (1946). The Bacterial Amino Acid Decarboxylases. *Adv. Enzymol. - Relat. Areas Mol. Biol.* *6*, 1–32.
- Gale, E.F., and Epps, H.M.R. (1942). The effect of the pH of the medium during growth on the enzymic activities of bacteria (*Escherichia coli* and *Micrococcus lysodeikticus*) and the biological significance of the changes produced. *Biochem. J.* *36*, 600–618.
- Gates, S.N., and Martin, A. (2019). Stairway to translocation: AAA+ motor structures reveal the mechanisms of ATP-dependent substrate translocation. *Protein Sci.* 1–13.

Gates, S.N., Yokom, A.L., Lin, J., Jackrel, M.E., Rizo, A.N., Kendsersky, N.M., Buell, C.E., Sweeny, E.A., Mack, K.L., Chuang, E., et al. (2017). Ratchet-like polypeptide translocation mechanism of the AAA + disaggregase Hsp104. *Science* (80-. ). *1052*, 1–13.

Gatsogiannis, C., Balogh, D., Merino, F., Sieber, S.A., and Raunser, S. (2019). Cryo-EM structure of the ClpXP protein degradation machinery. *Nat. Struct. Mol. Biol.* *26*, 946–954.

Gateva-Topalova, P.Z., May, A.P., and Sousa, M.C. (2004). Crystal Structure of Escherichia coli ArnA (PmrI) Decarboxylase Domain. A Key Enzyme for Lipid A Modification with 4-Amino-4-deoxy-1-arabinose and Polymyxin Resistance<sup>†</sup>, <sup>‡</sup>. *Biochemistry* *43*, 13370–13379.

Ge, P., Scholl, D., Leiman, P.G., Yu, X., Miller, J.F., and Zhou, Z.H. (2015). Atomic structures of a bactericidal contractile nanotube in its pre- and postcontraction states. *Nat. Struct. Mol. Biol.* *22*, 377–382.

Girgis, H.S., Hottes, A.K., and Tavazoie, S. (2009). Genetic architecture of intrinsic antibiotic susceptibility. *PLoS One*.

Glaeser, R.M. (1971). Limitations to significant information in biological electron microscopy as a result of radiation damage. *J. Ultrastructure Res.*

Glaeser, R.M. (2013). Invited Review Article: Methods for imaging weak-phase objects in electron microscopy. *Rev. Sci. Instrum.* *84*, 1–17.

Glaeser, R.M. (2016). Specimen Behavior in the Electron Beam. *Methods Enzymol.* *579*, 19–50.

Glaeser, R.M., and Han, B.-G. (2017). Opinion: hazards faced by macromolecules when confined to thin aqueous films. *Biophys. Reports* *3*, 1–7.

Glaeser, R.M., Typke, D., Tiemeijer, P.C., Pulokas, J., and Cheng, A. (2011). Precise beam-tilt alignment and collimation are required to minimize the phase error associated with coma in high-resolution cryo-EM. *J. Struct. Biol.* *174*, 1–10.

Glynn, S.E., Martin, A., Nager, A.R., Baker, T.A., and Sauer, R.T. (2009). Structures of Asymmetric ClpX Hexamers Reveal Nucleotide-Dependent Motions in a AAA+ Protein-Unfolding Machine. *Cell* *139*, 744–756.

Grant, T., Rohou, A., and Grigorieff, N. (2018). CisTEM, user-friendly software for single-particle image processing. *Elife* *7*.

Grigorieff, N. (2007). FREALIGN: High-resolution refinement of single particle structures. *J. Struct. Biol.* *157*, 117–125.

Guo, F., and Jiang, W. (2014). Single particle cryo-electron microscopy and 3-D reconstruction of viruses. *Methods Mol. Biol.* *1117*, 401–443.

Gutsche, I., Desfosses, A., Effantin, G., Ling, W.L., Haupt, M., Ruigrok, R.W.H., Sachse, C., and Schoehn, G. (2015). Near-atomic cryo-EM structure of the helical measles virus nucleocapsid. *Science* (80-. ). *348*, 704–707.

Hall, C.E., Jakus, M.A., and Schmitt, F.O. (1945). The structure of certain muscle fibrils as revealed by the use of electron stains. *J. Appl. Phys.* *16*, 459–465.

Han, B.G., Watson, Z., Kang, H., Pulk, A., Downing, K.H., Cate, J., and Glaeser, R.M. (2016). Long shelf-life streptavidin support-films suitable for electron microscopy of biological macromolecules. *J. Struct. Biol.* *195*, 238–244.

Han, Y., Fan, X., Wang, H., Zhao, F., Tully, C.G., Kong, J., Yao, N., and Yan, N. (2020). High-yield monolayer graphene grids for near-atomic resolution cryoelectron microscopy. *Proc. Natl. Acad. Sci. U. S. A.* *117*, 1009–1014.

Hanson, P.I., and Whiteheart, S.W. (2005). AAA+ proteins: have engine, will work. *Nat. Rev. Mol. Cell Biol.* *6*, 519–529.

Hänzelmann, P., and Schindelin, H. (2016). Structural Basis of ATP Hydrolysis and

Intersubunit Signaling in the AAA+ ATPase p97. *Structure* *24*, 127–139.

Harastani, M., Sorzano, C.O.S., and Jonić, S. (2020). Hybrid Electron Microscopy Normal Mode Analysis with Scipion. *Protein Sci.* *29*, 223–236.

Hasemann, C.A., Istvan, E.S., Uyeda, K., and Deisenhofer, J. (1996). The crystal structure of the bifunctional enzyme 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase reveals distinct domain homologies. *Structure* *4*, 1017–1029.

Hattne, J., Shi, D., Glynn, C., Zee, C. Te, Gallagher-Jones, M., Martynowycz, M.W., Rodriguez, J.A., and Gonen, T. (2018). Analysis of Global and Site-Specific Radiation Damage in Cryo-EM. *Structure* *26*, 759–766.e4.

Van Heel, M. (1987). Angular reconstitution: A posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy* *21*, 111–123.

Van Heel, M. (2013). Finding trimeric HIV-1 envelope glycoproteins in random noise. *Proc. Natl. Acad. Sci. U. S. A.* *110*, 4175–4177.

Van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R., and Schatz, M. (1996). A new generation of the IMAGIC image processing system. *J. Struct. Biol.* *116*, 17–24.

Henderson, R. (1995). The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q. Rev. Biophys.* *28*, 171–193.

Henderson, R. (2013). Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proc. Natl. Acad. Sci. U. S. A.* *110*, 18037–18041.

Henderson, R. (2015). Overview and future of single particle electron cryomicroscopy. *Arch. Biochem. Biophys.* *581*, 19–24.

Henning, S., and Adhikari, R. (2017). Scanning Electron Microscopy, ESEM, and X-ray Microanalysis (Elsevier Inc.).

Herzik, M.A., Wu, M., and Lander, G.C. (2019). High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM. *Nat. Commun.* *10*, 1032.

Hickman, A.B., and Dyda, F. (2005). Binding and unwinding: SF3 viral helicases. *Curr. Opin. Struct. Biol.* *15*, 77–85.

Ho, C.-M., Li, X., Lai, M., Terwilliger, T.C., Beck, J.R., Wohlschlegel, J., Goldberg, D.E., Fitzpatrick, A.W.P., and Zhou, Z.H. (2019). Bottom-up structural proteomics: cryoEM of protein complexes enriched from the cellular milieu. *Nat. Methods*.

Hohn, M., Tang, G., Goodyear, G., Baldwin, P.R., Huang, Z., Penczek, P.A., Yang, C., Glaeser, R.M., Adams, P.D., and Ludtke, S.J. (2007). SPARX, a new environment for Cryo-EM image processing. *J. Struct. Biol.* *157*, 47–55.

Horne, R.W. (1973). Contrast and resolution from biological objects examined in the electron microscope with particular reference to negatively stained specimens. *J. Microsc.* *98*, 286–298.

Huang, R., Ripstein, Z.A., Augustyniak, R., Lazniewski, M., Ginalski, K., Kay, L.E., and Rubinstein, J.L. (2016). Unfolding the mechanism of the AAA+ unfoldase VAT by a combined cryo-EM, solution NMR study. *Proc. Natl. Acad. Sci.* *113*, E4190–E4199.

Itsathitphaisarn, O., Wing, R.A., Eliason, W.K., Wang, J., and Steitz, T.A. (2012). The hexameric helicase DnaB adopts a nonplanar conformation during translocation. *Cell* *151*, 267–277.

Iyer, L.M., Makarova, K.S., Koonin, E. V., and Aravind, L. (2004a). Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: Implications for the origins of chromosome segregation, cell division and viral capsid packaging. *Nucleic Acids Res.* *32*, 5260–5279.

Iyer, L.M., Leipe, D.D., Koonin, E. V., and Aravind, L. (2004b). Evolutionary history and higher order classification of AAA+ ATPases. *J. Struct. Biol.* *146*, 11–31.

Jain, T., Sheehan, P., Crum, J., Carragher, B., and Potter, C.S. (2012). Spotiton: A prototype

- for an integrated inkjet dispense and vitrification system for cryo-TEM. *J. Struct. Biol.* *179*, 68–75.
- Jakus, M.A., Hall, C.E., and Schmitt, F.O. (1944). ELECTRON MICROSCOPE OBSERVATIONS OF CLAM MUSCLE FIBRILS. *J. Am. Chem. Soc.* *66*, 313–314.
- Jean, N.L., Rutherford, T.J., and Löwe, J. (2019). FtsK in motion reveals its mechanism for double-stranded DNA translocation. *bioRxiv* 828319.
- Jensen, G.J., and Kornberg, R.D. (2000). Defocus-gradient corrected back-projection. *Ultramicroscopy* *84*, 57–64.
- Jeruzalmi, D., O'Donnell, M., and Kuriyan, J. (2001). Crystal structure of the processivity clamp loader gamma ( $\gamma$ ) complex of *E. coli* DNA polymerase III. *Cell* *106*, 429–441.
- Jin, Q., Sorzano, C.O.S., de la Rosa-Trevín, J.M., Bilbao-Castro, J.R., Núñez-Ramírez, R., Llorca, O., Tama, F., and Jonić, S. (2014). Iterative Elastic 3D-to-2D Alignment Method Using Normal Modes for Studying Structural Dynamics of Large Macromolecular Complexes. *Structure* *22*, 496–506.
- Joly, N., Zhang, N., Buck, M., and Zhang, X. (2012). Coupling AAA protein function to regulated gene expression. *Biochim. Biophys. Acta - Mol. Cell Res.* *1823*, 108–116.
- Jonić, S. (2017). Computational methods for analyzing conformational variability of macromolecular complexes from cryo-electron microscopy images. *Curr. Opin. Struct. Biol.* *43*, 114–121.
- Jörnvall, H., Persson, B., Krook, M., Atrian, S., Gonzalez-Duarte, R., Jeffery, J., and Ghosh, D. (1995). Short-chain dehydrogenases/reductases (SDR). *Biochemistry* *34*, 6003–6013.
- Kaguni, J.M. (2014). DnaA, DnaB, and DnaC. In *Molecular Life Sciences*, E. Bell, ed. (New York, NY: Springer New York), pp. 1–14.
- Kahle, M., ter Beek, J., Hosler, J.P., and Ädelroth, P. (2018). The insertion of the non-heme Fe B cofactor into nitric oxide reductase from *P. denitrificans* depends on NorQ and NorD accessory proteins. *Biochim. Biophys. Acta - Bioenerg.* *1859*, 1051–1058.
- Kandiah, E., Carriel, D., Perard, J., Malet, H., Bacia, M., Liu, K., Chan, S.W.S., Houry, W. a., Ollagnier de Choudens, S., Elsen, S., et al. (2016). Structural insights into the *Escherichia coli* lysine decarboxylases and molecular determinants of interaction with the AAA+ ATPase RavA. *Sci. Rep.* *6*, 24601.
- Kandiah, E., Carriel, D., Garcia, P.S., Felix, J., Banzhaf, M., Kritikos, G., Bacia-Verloop, M., Brochier-Armanet, C., Elsen, S., and Gutsche, I. (2019). Structure, Function, and Evolution of the *Pseudomonas aeruginosa* Lysine Decarboxylase LdcA. *Structure* 1–13.
- Kanjee, U., Gutsche, I., Ramachandran, S., and Houry, W.A. (2011a). The enzymatic activities of the *Escherichia coli* basic aliphatic amino acid decarboxylases exhibit a pH zone of inhibition. *Biochemistry* *50*, 9388–9398.
- Kanjee, U., Gutsche, I., Alexopoulos, E., Zhao, B., El Bakkouri, M., Thibault, G., Liu, K., Ramachandran, S., Snider, J., Pai, E.F., et al. (2011b). Linkage between the bacterial acid stress and stringent responses: the structure of the inducible lysine decarboxylase. *EMBO J.* *30*, 931–944.
- Kastner, B., Fischer, N., Golas, M.M., Sander, B., Dube, P., Boehringer, D., Hartmuth, K., Deckert, J., Hauer, F., Wolf, E., et al. (2008). GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat. Methods* *5*, 53–55.
- Kazlauskas, D., Krupovic, M., and Venclovas, C. (2016). The logic of DNA replication in double-stranded DNA viruses: insights from global analysis of viral genomes. *Nucleic Acids Res.* *44*, 4551–4564.
- Kelch, B.A. (2016). Review: The lord of the rings: Structure and mechanism of the sliding clamp loader. *Biopolymers* *105*, 532–546.
- Kelch, B.A., Makino, D.L., O'Donnell, M., and Kuriyan, J. (2011). How a DNA Polymerase Clamp Loader Opens a Sliding Clamp. *Science* (80-. ). *334*, 1675–1680.

Keseler, I.M., Mackie, A., Santos-Zavaleta, A., Billington, R., Bonavides-Martínez, C., Caspi, R., Fulcher, C., Gama-Castro, S., Kothari, A., Krummenacker, M., et al. (2017). The EcoCyc database: Reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Res.* *45*, D543–D550.

Khoshouei, M., Radjainia, M., Phillips, A.J., Gerrard, J.A., Mitra, A.K., Plitzko, J.M., Baumeister, W., and Danev, R. (2016). Volta phase plate cryo-EM of the small protein complex Prx3. *Nat. Commun.* *7*, 10534.

Khoshouei, M., Radjainia, M., Baumeister, W., and Danev, R. (2017). Cryo-EM structure of haemoglobin at 3.2 Å determined with the Volta phase plate. *Nat. Commun.* *8*.

Kimanius, D., Forsberg, B.O., Scheres, S.H.W., and Lindahl, E. (2016). Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *Elife* *5*.

Klug, A., and Finch, J.T. (1965). Structure of viruses of the papilloma-polyoma type. *J. Mol. Biol.* *11*, 403–423.

Knapke, E., and Dubochet, J. (1980). Beam damage to organic material is considerably reduced in cryo-electron microscopy. *J. Mol. Biol.* *141*, 147–161.

Knowles, J.R. (1980). Enzyme-Catalyzed Phosphoryl Transfer Reactions. *Annu. Rev. Biochem.* *49*, 877–919.

Kon, T., Oyama, T., Shimo-Kon, R., Imamula, K., Shima, T., Sutoh, K., and Kurisu, G. (2012). The 2.8 Å crystal structure of the dynein motor domain. *Nature* *484*, 345–350.

Konings, S., Kuijper, M., Keizer, J., Grollios, F., Spanjer, T., and Tiemeijer, P. (2019). Advances in Single Particle Analysis Data Acquisition. *Microsc. Microanal.* *25*, 1012–1013.

Koonin, E. V. (1993). A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPases including MCM proteins involved in the initiation of eukaryotic DNA replication. *Nucleic Acids Res.* *21*, 2541–2547.

Koonin, E. V., Wolf, Y.I., and Aravind, L. (2000). Protein fold recognition using sequence profiles and its application in structural genomics. *Adv. Protein Chem.* *54*, 245–275.

Kötting, C., Kallenbach, A., Suveyzdis, Y., Wittinghofer, A., and Gerwert, K. (2008). The GAP arginine finger movement into the catalytic site of Ras increases the activation entropy. *Proc. Natl. Acad. Sci. U. S. A.* *105*, 6260–6265.

Kunau, W.H., Beyer, A., Franken, T., Götte, K., Marzioch, M., Saidowsky, J., Skaletz-Rorowski, A., and Wiebel, F.F. (1993). Two complementary approaches to study peroxisome biogenesis in *Saccharomyces cerevisiae*: Forward and reversed genetics. *Biochimie* *75*, 209–224.

Kyrpides, N., Overbeek, R., and Ouzounis, C. (1999). Universal protein families and the functional content of the last universal common ancestor. *J. Mol. Evol.* *49*, 413–423.

Lander, G.C., Estrin, E., Matyskiela, M.E., Bashore, C., Nogales, E., and Martin, A. (2012). Complete subunit architecture of the proteasome regulatory particle. *Nature* *482*, 186–191.

Lane, N. (2015). The unseen World: Reflections on Leeuwenhoek (1677) “Concerning little animals.” *Philos. Trans. R. Soc. B Biol. Sci.* *370*.

van Leeuwenhoek, A. (1677). Observations, communicated to the publisher by Mr. Antony van Leewenhoek, in a dutch letter of the 9th Octob. 1676. here English’d: concerning little animals by him observed in rain-well-sea- and snow water; as also in water wherein pepper had lain infus. *Philos. Trans. R. Soc. London* *12*, 821–831.

van Leeuwenhoek, A. (1682). An account of several very curious discoveries about the internal texture of the flesh of muscles, of strange motions in the finns, and the manner of the production of the shells of oysters, &c. made by Mr. Anthony Leuwenhoek fellow of the Royal Society; . *Philos. Collect. R. Soc. London* *5*, 152–160.

van Leeuwenhoek, A. (1684). An abstract of a letter from Mr. Anthony Leevvenhoek at Delft, dated Sep. 17. 1683. Containing some microscopical observations, about animals in the scurf of the teeth, the substance call’d worms in the nose, the cuticula consisting of scales. *Philos. Trans.*

R. Soc. London *14*, 568–574.

Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., and Aila, T. (2018). Noise2Noise: Learning image restoration without clean data. *35th Int. Conf. Mach. Learn. ICML 2018 7*, 4620–4631.

Leipe, D.D., Wolf, Y.I., Koonin, E. V., and Aravind, L. (2002). Classification and evolution of P-loop GTPases and related ATPases. *J. Mol. Biol.* *317*, 41–72.

Leipe, D.D., Koonin, E. V., and Aravind, L. (2003). Evolution and classification of P-loop kinases and related proteins. *J. Mol. Biol.* *333*, 781–815.

Leipe, D.D., Koonin, E. V., and Aravind, L. (2004). STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: Multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. *J. Mol. Biol.* *343*, 1–28.

Lepault, J., and Pitt, T. (1984). Projected structure of unstained, frozen-hydrated T-layer of *Bacillus brevis*. *EMBO J.* *3*, 101–105.

Li, N., Lam, W.H., Zhai, Y., Cheng, J., Cheng, E., Zhao, Y., Gao, N., and Tye, B.K. (2018). Structure of the origin recognition complex bound to DNA replication origin. *Nature* *559*, 217–222.

Lin, J.H., and Yamazaki, M. (2003). Role of P-Glycoprotein in Pharmacokinetics. *Clin. Pharmacokinet.* *42*, 59–98.

Lipmann, F. (1941). Metabolic Generation and Utilization of Phosphate Bond Energy. In *Advances in Enzymology - and Related Areas of Molecular Biology*, F.F. Nord, and C.H. Werkman, eds. pp. 99–162.

von Loeffelholz, O., Natchiar, S.K., Djabeur, N., Myasnikov, A.G., Kratzat, H., Ménétret, J.F., Hazemann, I., and Klaholz, B.P. (2017). Focused classification and refinement in high-resolution cryo-EM structural analysis of ribosome complexes. *Curr. Opin. Struct. Biol.* *46*, 140–148.

Lopéz-Blanco, J.R., and Chacón, P. (2013). IMODFIT: Efficient and robust flexible fitting based on vibrational analysis in internal coordinates. *J. Struct. Biol.* *184*, 261–270.

Ludtke, S.J., Baldwin, P.R., and Chiu, W. (1999). EMAN: Semiautomated Software for High-Resolution Single-Particle Reconstructions. *J. Struct. Biol.* *128*, 82–97.

Ludtke, S.J., Chen, D.H., Song, J.L., Chuang, D.T., and Chiu, W. (2004). Seeing GroEL at 6 Å resolution by single particle electron cryomicroscopy. *Structure* *12*, 1129–1136.

Lupas, A., and Martin, J. (2002). AAA proteins. *Curr. Opin. Struct. Biol.* *12*, 746–753.

Lyumkis, D., Vinterbo, S., Potter, C.S., and Carragher, B. (2013). Optimod - An automated approach for constructing and optimizing initial models for single-particle electron microscopy. *J. Struct. Biol.* *184*, 417–426.

Mackintosh, S.G., and Raney, K.D. (2006). DNA unwinding and protein displacement by superfamily 1 and superfamily 2 helicases. *Nucleic Acids Res.* *34*, 4154–4159.

Maddhuri Venkata Subramaniya, S.R., Terashi, G., and Kihara, D. (2019). Protein secondary structure detection in intermediate-resolution cryo-EM maps using deep learning. *Nat. Methods* *16*, 911–917.

Malet, H., Liu, K., El Bakkouri, M., Chan, S.W.S., Effantin, G., Bacia, M., Houry, W.A., and Gutsche, I. (2014). Assembly principles of a unique cage formed by hexameric and decameric *E. coli* proteins. *Elife* *3*, e03653.

Marquenet, E., and Richet, E. (2007). How Integration of Positive and Negative Regulatory Signals by a STAND Signaling Protein Depends on ATP Hydrolysis. *Mol. Cell* *28*, 187–199.

Marquenet, E., and Richet, E. (2010). Conserved motifs involved in ATP hydrolysis by MalT, a signal transduction ATPase with numerous domains from *Escherichia coli*. *J. Bacteriol.* *192*, 5181–

5191.

Martin, A., Baker, T.A., and Sauer, R.T. (2008). Diverse Pore Loops of the AAA+ ClpX Machine Mediate Unassisted and Adaptor-Dependent Recognition of *ssrA*-Tagged Substrates. *Mol. Cell* 29, 441–450.

Marton, L. (1934). Electron Microscopy of Biological Objects. *Nature* 133, 911–911.

Marzahn, M.R., Hayner, J.N., Finkelstein, J., O'Donnell, M., and Bloom, L.B. (2014). The ATP Sites of AAA+ Clamp Loaders Work Together as a Switch to Assemble Clamps on DNA. *J. Biol. Chem.* 289, 5537–5548.

Mastrorade, D.N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* 152, 36–51.

McCallum, M., Benlekbir, S., Nguyen, S., Tammam, S., and John, L. (2019). Multiple conformations facilitate PilT function in the type IV pilus. *Nat. Commun.* 1–35.

McMullan, G., Chen, S., Henderson, R., and Faruqi, A.R. (2009). Detective quantum efficiency of electron area detectors in electron microscopy. *Ultramicroscopy* 109, 1126–1143.

McMullan, G., Faruqi, A.R., and Henderson, R. (2016). Chapter One – Direct Electron Detectors. In *Methods in Enzymology*, (Elsevier Inc.), pp. 1–17.

Meagher, M., Epling, L.B., and Enemark, E.J. (2019). DNA translocation mechanism of the MCM complex and implications for replication initiation. *Nat. Commun.* 10.

Medagli, B., and Onesti, S. (2013). Structure and Mechanism of Hexameric Helicases. In *DNA Helicases and DNA Motor Proteins (Advances in Experimental Medicine and Biology)*, M. Spies, ed. (Springer New York), pp. 75–95.

Meents, A., Gutmann, S., Wagner, A., and Schulze-Briese, C. (2010). Origin and temperature dependence of radiation damage in biological samples at cryogenic temperatures. *Proc. Natl. Acad. Sci. U. S. A.* 107, 1094–1099.

Meng, S.Y., and Bennett, G.N. (1992). Nucleotide sequence of the *Escherichia coli* *cad* operon: A system for neutralization of low extracellular pH. *J. Bacteriol.* 174, 2659–2669.

Michalska, K., Zhang, K., March, Z.M., Hatzos-Skintges, C., Pintilie, G., Bigelow, L., Castellano, L.M., Miles, L.J., Jackrel, M.E., Chuang, E., et al. (2018). Structure of *Calcarisporiella thermophila* Hsp104 Disaggregase that Antagonizes Diverse Proteotoxic Misfolding Events. *Structure* 1–15.

Miertzschke, M., Koerner, C., Vetter, I.R., Keilberg, D., Hot, E., Leonardy, S., Søgaard-Andersen, L., and Wittinghofer, A. (2011). Structural analysis of the Ras-like G protein MglA and its cognate GAP MglB and implications for bacterial polarity. *EMBO J.* 30, 4185–4197.

Milburn, M., Tong, L., DeVos, A., Brunger, A., Yamaizumi, Z., Nishimura, S., and Kim, S. (1990). Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins. *Science* (80-. ). 247, 939–945.

Miller, J.M., and Enemark, E.J. (2016). Fundamental Characteristics of AAA+ Protein Family Structure and Function. *Archaea* 2016.

Miller, T.C.R., Locke, J., Greiwe, J.F., Diffley, J.F.X., and Costa, A. (2019). Mechanism of head-to-head MCM double-hexamer formation revealed by cryo-EM. *Nature* 575, 704–710.

Milner-White, E.J., Coggins, J.R., and Anton, I.A. (1991). Evidence for an ancestral core structure in nucleotide-binding proteins with the type A motif. *J. Mol. Biol.* 221, 751–754.

Mindell, J.A., and Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* 142, 334–347.

Monroe, N., Han, H., Shen, P.S., Sundquist, W.I., and Hill, C.P. (2017). Structural basis of protein translocation by the Vps4-Vta1 AAA ATPase. *Elife* 6, 1–22.

Moriya, T., Saur, M., Stabrin, M., Merino, F., Voicu, H., Huang, Z., Penczek, P.A., Raunser,

- S., and Gatsogiannis, C. (2017). High-resolution single particle analysis from electron cryo-microscopy images using SPHIRE. *J. Vis. Exp.* 2017, 1–11.
- Mott, M.L., Erzberger, J.P., Coons, M.M., and Berger, J.M. (2008). Structural Synergy and Molecular Crosstalk between Bacterial Helicase Loaders and Replication Initiators. *Cell* 135, 623–634.
- Nagy, G.N., Suardíaz, R., Lopata, A., Ozohanics, O., Vékely, K., Brooks, B.R., Leveles, I., Tóth, J., Vértessy, B.G., and Rosta, E. (2016). Structural Characterization of Arginine Fingers: Identification of an Arginine Finger for the Pyrophosphatase dUTPases. *J. Am. Chem. Soc.* 138, 15035–15045.
- Nakane, T., Kimanius, D., Lindahl, E., and Scheres, S.H.W. (2018). Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *Elife* 7, 1–18.
- Neuwald, A.F., Aravind, L., Spouge, J.L., and Koonin, E. V. (1999). AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.* 9, 27–43.
- Nirwan, N., Itoh, Y., Singh, P., Bandyopadhyay, S., Vinothkumar, K.R., Amunts, A., and Saikrishnan, K. (2019). Structure-based mechanism for activation of the AAA+ GTPase McrB by the endonuclease McrC. *Nat. Commun.* 10, 1–9.
- Noble, A.J., Dandey, V.P., Wei, H., Brasch, J., Chase, J., Acharya, P., Tan, Y.Z., Zhang, Z., Kim, L.Y., Scapin, G., et al. (2018a). Routine single particle CryoEM sample and grid characterization by tomography. *Elife* 7, 1–42.
- Noble, A.J., Wei, H., Dandey, V.P., Zhang, Z., Tan, Y.Z., Potter, C.S., and Carragher, B. (2018b). Reducing effects of particle adsorption to the air–water interface in cryo-EM. *Nat. Methods* 15, 793–795.
- Nogales, E., and Scheres, S.H.W. (2015). Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity. *Mol. Cell* 58, 677–689.
- Ogura, T., and Wilkinson, A.J. (2001). AAA+ superfamily ATPases: Common structure-diverse function. *Genes to Cells* 6, 575–597.
- Ogura, T., Whiteheart, S.W., and Wilkinson, A.J. (2004). Conserved arginine residues implicated in ATP hydrolysis, nucleotide-sensing, and inter-subunit interactions in AAA and AAA+ ATPases. *J. Struct. Biol.* 146, 106–112.
- Orlova, E. V., and Saibil, H.R. (2011). Structural Analysis of Macromolecular Assemblies by Electron Microscopy. *Chem. Rev.* 111, 7710–7748.
- Pantelic, R.S., Meyer, J.C., Kaiser, U., Baumeister, W., and Plitzko, J.M. (2010). Graphene oxide: A substrate for optimizing preparations of frozen-hydrated samples. *J. Struct. Biol.* 170, 152–156.
- Passmore, L.A., and Russo, C.J. (2016). Chapter Three – Specimen Preparation for High-Resolution Cryo-EM. In *Methods in Enzymology*, (Elsevier Inc.), pp. 51–86.
- Penczek, P. a (2010). Fundamentals of three-dimensional reconstruction from projections. *Methods Enzymol.* 482, 1–28.
- 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. *J. Comput. Chem.* 25, 1605–1612.
- Plattner, H., and Verkhatsky, A. (2016). Inseparable tandem: Evolution chooses ATP and Ca<sup>2+</sup> to control life, death and cellular signalling. *Philos. Trans. R. Soc. B Biol. Sci.* 371.
- Puchades, C., Rampello, A.J., Shin, M., Giuliano, C.J., Wiseman, R.L., Glynn, S.E., and Lander, G.C. (2017). Structure of the mitochondrial inner membrane AAA+ protease YME1 gives insight into substrate processing. *Science* (80-. ). 358.

- Puchades, C., Sandate, C.R., and Lander, G.C. (2020). The molecular principles governing the activity and functional diversity of AAA+ proteins. *Nat. Rev. Mol. Cell Biol.* *21*, 43–58.
- Punjani, A., and Fleet, D.J. (2020). 3D Variability Analysis: Directly resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM images. *bioRxiv* 2020.04.08.032466.
- Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* *14*, 290–296.
- Punjani, A., Zhang, H., and Fleet, D.J. (2019). Non-uniform refinement: Adaptive regularization improves single particle cryo-EM reconstruction. *bioRxiv* 877092.
- Radermacher, M., Wagenknecht, T., Verschoor, A., and Frank, J. (1987). Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of *Escherichia coli*. *J. Microsc.* *146*, 113–136.
- Raetz, C.R.H., and Whitfield, C. (2002). Lipopolysaccharide Endotoxins. *Annu. Rev. Biochem.* *71*, 635–700.
- Ramlaul, K., Palmer, C.M., and Aylett, C.H.S. (2020). Mitigating Local Over-fitting During Single Particle Reconstruction with SIDESPLITTER. *bioRxiv* 2019.12.12.874081.
- Rappas, M., Schumacher, J., Niwa, H., Buck, M., and Zhang, X. (2006). Structural Basis of the Nucleotide Driven Conformational Changes in the AAA+ Domain of Transcription Activator PspF. *J. Mol. Biol.* *357*, 481–492.
- Ravelli, R.B.G., Nijpels, F.J.T., Henderikx, R.J.M., Weissenberger, G., Thewessem, S., Gijssbers, A., Beulen, B.W.A.M.M., López-Iglesias, C., and Peters, P.J. (2019). Automated cryo-EM sample preparation by pin-printing and jet vitrification. *bioRxiv* 651208.
- Rayleigh, J. (1896). XV. On the theory of optical images, with special reference to the microscope. London, Edinburgh, Dublin Philos. Mag. J. Sci. *42*, 167–195.
- Resat, H., Straatsma, T.P., Dixon, D.A., and Miller, J.H. (2001a). The arginine finger of RasGap helps Gln-61 align the nucleophilic water in GAP-stimulated hydrolysis of GTP. *Proc. Natl. Acad. Sci. U. S. A.* *98*, 6033–6038.
- Resat, H., Straatsma, T.P., Dixon, D.A., and Miller, J.H. (2001b). The arginine finger of RasGAP helps Gln-61 align the nucleophilic water in GAP-stimulated hydrolysis of GTP. *Proc. Natl. Acad. Sci.* *98*, 6033–6038.
- Ripstein, Z.A., and Rubinstein, J.L. (2016). Processing of Cryo-EM Movie Data. *Methods Enzymol.* *579*, 103–124.
- Ripstein, Z.A., Huang, R., Augustyniak, R., Kay, L.E., and Rubinstein, J.L. (2017). Structure of a AAA+ unfoldase in the process of unfolding substrate. *Elife* *6*, 1–14.
- Ripstein, Z.A., Vahidi, S., Houry, W.A., Rubinstein, J.L., and Kay, L.E. (2020). A processive rotary mechanism couples substrate unfolding and proteolysis in the ClpXP degradation machinery. *Elife* *9*, 1–50.
- Rivera-Calzada, A., Pal, M., Muñoz-Hernández, H., Luque-Ortega, J.R., Gil-Carton, D., Degliesposti, G., Skehel, J.M., Prodromou, C., Pearl, L.H., and Llorca, O. (2017). The Structure of the R2TP Complex Defines a Platform for Recruiting Diverse Client Proteins to the HSP90 Molecular Chaperone System. *Structure* *25*, 1145–1152.e4.
- Rizo, A.N., Lin, J.B., Gates, S.N., Tse, E., Bart, S.M., Castellano, L.M., DiMaio, F., Shorter, J., and Southworth, D.R. (2019). Structural basis for substrate gripping and translocation by the ClpB AAA+ disaggregase. *Nat. Commun.* *10*, 1–12.
- Rohou, A., and Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* *192*, 216–221.
- Rosenthal, P.B. (2015). From high symmetry to high resolution in biological electron microscopy: a commentary on Crowther (1971) “Procedures for three-dimensional reconstruction

- of spherical viruses by Fourier synthesis from electron micrographs ." *Philos. Trans. R. Soc. B* *370*.
- Rosenthal, P.B., and Henderson, R. (2003). Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. *J. Mol. Biol.* *333*, 721–745.
- De Rosier, D.J., and Klug, A. (1968). Reconstruction of Three Dimensional Structures from Electron Micrographs. *Nature* *217*, 130–134.
- Rouault, T.A. (2015). Mammalian iron-sulphur proteins: novel insights into biogenesis and function. *Nat. Rev. Mol. Cell Biol.* *16*, 45–55.
- Rubinstein, J.L., Guo, H., Ripstein, Z.A., Haydaroglu, A., Au, A., Yip, C.M., Di Trani, J.M., Benlekhir, S., and Kwok, T. (2019). Shake-it-off: a simple ultrasonic cryo-EM specimen-preparation device. *Acta Crystallogr. Sect. D Struct. Biol.* *75*, 1063–1070.
- Ruska, E. (1934). Über ein magnetisches Objektiv für das Elektronenmikroskop. *Zeitschrift Für Phys.* *89*, 90–128.
- Ruska, E. (1987). The Development of the Electron Microscope and of Electron Microscopy (Nobel Lecture). *Angew. Chemie Int. Ed. English* *26*, 595–605.
- Sabo, D.L., Boeker, E.A., Byers, B., Waron, H., and Fischer, E.H. (1974). Purification and Physical Properties of Inducible *Escherichia coli* Lysine Decarboxylase. *Biochemistry* *13*, 662–670.
- Sanchez-garcia, R., Segura, J., Maluenda, D., Carazo, J.M., Unit, B., Diego, S., Diego, S., and Jolla, L. (2019). MicrographCleaner: a python package for cryo-EM micrograph segmentation using deep learning. 1–5.
- Sandate, C.R., Szyk, A., Zehr, E.A., Lander, G.C., and Roll-Mecak, A. (2019). An allosteric network in spastin couples multiple activities required for microtubule severing. *Nat. Struct. Mol. Biol.* *26*, 671–678.
- Saraste, M., Sibbald, P.R., and Wittinghofer, A. (1990). The P-loop - a common motif in ATP- and GTP-binding proteins. *Trends Biochem. Sci.* *15*, 430–434.
- Scheffzek, K., Ahmadian, M.R., Kabsch, W., Wiesmüller, L., Lautwein, A., Schmitz, F., and Wittinghofer, A. (1997). The Ras-RasGAP complex: Structural basis for GTPase activation and its loss in oncogenic ras mutants. *Science (80- )*. *277*, 333–338.
- Schenk, A.D., Cavadini, S., Thomä, N.H., and Genoud, C. (2020). Live Analysis and Reconstruction of Single-Particle Cryo-Electron Microscopy Data with CryoFLARE. *J. Chem. Inf. Model.*
- Scherer, S., Kowal, J., Chami, M., Dandey, V., Arbeit, M., Ringler, P., and Stahlberg, H. (2014). 2dx\_automator: Implementation of a semiautomatic high-throughput high-resolution cryo-electron crystallography pipeline. *J. Struct. Biol.* *186*, 302–307.
- Scheres, S.H.W. (2012a). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* *180*, 519–530.
- Scheres, S.H.W. (2012b). A bayesian view on cryo-EM structure determination. *J. Mol. Biol.* *415*, 406–418.
- Scheres, S.H.W. (2014). Beam-induced motion correction for sub-megadalton cryo-EM particles. *Elife* *3*, e03665.
- Scheres, S.H.W. (2016). Processing of Structurally Heterogeneous Cryo-EM Data in RELION. In *Methods in Enzymology*, (Elsevier Inc.), pp. 125–157.
- Scheres, S.H.W., and Chen, S. (2012). Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* *9*, 853–854.
- Scheres, S.H.W., Valle, M., Nuñez, R., Sorzano, C.O.S., Marabini, R., Herman, G.T., and Carazo, J.M. (2005). Maximum-likelihood multi-reference refinement for electron microscopy images. *J. Mol. Biol.* *348*, 139–149.

- Schilbach, S., Hantsche, M., Tegunov, D., Dienemann, C., Wigge, C., Urlaub, H., and Cramer, P. (2017). Structures of transcription pre-initiation complex with TFIID and Mediator. *Nature* *551*, 204–209.
- Schmidli, C., Albiez, S., Rima, L., Righetto, R., Mohammed, I., Oliva, P., Kovacic, L., Stahlberg, H., and Braun, T. (2019). Microfluidic protein isolation and sample preparation for high-resolution cryo-EM. *Proc. Natl. Acad. Sci.* *116*, 15007–15012.
- Schmidt, H., Zalyte, R., Urnavicius, L., and Carter, A.P. (2015). Structure of human cytoplasmic dynein-2 primed for its power stroke. *Nature* *518*, 435–438.
- Schröder, R.R. (2015). Advances in electron microscopy: A qualitative view of instrumentation development for macromolecular imaging and tomography. *Arch. Biochem. Biophys.* *581*, 25–38.
- Schwartz, O., Axelrod, J.J., Campbell, S.L., Turnbaugh, C., Glaeser, R.M., and Müller, H. (2019). Laser phase plate for transmission electron microscopy. *Nat. Methods* *16*, 1016–1020.
- Scrima, A., Thomas, C., Deaconescu, D., and Wittinghofer, A. (2008). The Rap–RapGAP complex: GTP hydrolysis without catalytic glutamine and arginine residues. *EMBO J.* *27*, 1145–1153.
- Serna, M. (2019). Hands on methods for high resolution cryo-electron microscopy structures of heterogeneous macromolecular complexes. *Front. Mol. Biosci.* *6*, 1–8.
- Sgro, G.G., and Costa, T.R.D. (2018). Cryo-EM Grid Preparation of Membrane Protein Samples for Single Particle Analysis. *Front. Mol. Biosci.* *5*, 1–8.
- Shaikh, T.R., Gao, H., Baxter, W.T., Asturias, F.J., Boisset, N., Leith, A., and Frank, J. (2008). SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nat. Protoc.* *3*, 1941–1974.
- Shin, M., Asmita, A., Puchades, C., Adjei, E., Wiseman, R.L., Karzai, A.W., and Lander, G.C. (2019). Distinct Structural Features of the Lon Protease Drive Conserved Hand-over-Hand Substrate Translocation. *bioRxiv* 617159.
- Singleton, M.R., Dillingham, M.S., and Wigley, D.B. (2007). Structure and Mechanism of Helicases and Nucleic Acid Translocases. *Annu. Rev. Biochem.* *76*, 23–50.
- Smith, D.J. (2008). Ultimate resolution in the electron microscope? *Mater. Today* *11*, 30–38.
- Snider, J., and Houry, W.A. (2006). MoxR AAA+ ATPases: A novel family of molecular chaperones? *J. Struct. Biol.* *156*, 200–209.
- Snider, J., Gutsche, I., Lin, M., Baby, S., Cox, B., Butland, G., Greenblatt, J., Emili, A., and Houry, W.A. (2006). Formation of a Distinctive Complex between the Inducible Bacterial Lysine Decarboxylase and a Novel AAA+ ATPase. *J. Biol. Chem.* *281*, 1532–1546.
- Snider, J., Thibault, G., and Houry, W.A. (2008). The AAA+ superfamily of functionally diverse proteins. *Genome Biol.* *9*, 1–8.
- Sosnowski, P., Urnavicius, L., Boland, A., Fagiewicz, R., Busselez, J., Papai, G., and Schmidt, H. (2018). The CryoEM structure of the *Saccharomyces cerevisiae* ribosome maturation factor Rea1. *Elife* *7*, e39163.
- Van Spanning, R.J.M., Wansell, C.W., De Boer, T., Hazelaar, M.J., Anazawa, H., Harms, N., Oltmann, L.F., and Stouthamer, A.H. (1991). Isolation and characterization of the *moxJ*, *moxG*, *moxI*, and *moxR* genes of *Paracoccus denitrificans*: inactivation of *moxJ*, *moxG*, and *moxR* and the resultant effect on methylotrophic growth. *J. Bacteriol.* *173*, 6948–6961.
- Stark, H. (2010). GraFix: Stabilization of fragile macromolecular complexes for single particle Cryo-EM. *Methods Enzymol.* *481*, 109–126.
- Stinson, B.M., Nager, A.R., Glynn, S.E., Schmitz, K.R., Baker, T.A., and Sauer, R.T. (2013). Nucleotide binding and conformational switching in the hexameric ring of a AAA+ machine. *Cell* *153*, 628–639.

- Stuart, D.I., Subramaniam, S., and Abrescia, N.G.A. (2016). The democratization of cryo-EM. *Nat. Methods* *13*, 607–608.
- Su, M., Guo, E.Z., Ding, X., Li, Y., Tarrasch, J.T., Brooks, C.L., Xu, Z., and Skiniotis, G. (2017). Mechanism of Vps4 hexamer function revealed by cryo-EM. *Sci. Adv.* *3*, e1700325.
- Subramaniam, S. (2013). Structure of trimeric HIV-1 envelope glycoproteins. *Proc. Natl. Acad. Sci. U. S. A.* *110*, 4172–4174.
- Sun, S., Li, L., Yang, F., Wang, X., Fan, F., Yang, M., Chen, C., Li, X., Wang, H.-W., and Sui, S.-F. (2017). Cryo-EM structures of the ATP-bound Vps4E233Q hexamer and its complex with Vta1 at near-atomic resolution. *Nat. Commun.* *8*, 16064.
- Sutter, M., Roberts, E.W., Gonzalez, R.C., Bates, C., Dawoud, S., Landry, K., Cannon, G.C., Heinhorst, S., and Kerfeld, C.A. (2015). Structural Characterization of a Newly Identified Component of  $\alpha$ -Carboxysomes: The AAA+ Domain Protein CsoCbbQ. *Sci. Rep.* *5*, 1–14.
- Sysoeva, T.A. (2017). Assessing heterogeneity in oligomeric AAA+ machines. *Cell. Mol. Life Sci.* *74*, 1001–1018.
- Sysoeva, T.A., Chowdhury, S., Guo, L., and Nixon, B.T. (2013). Nucleotide-induced asymmetry within ATPase activator ring drives 54-RNAP interaction and ATP hydrolysis. *Genes Dev.* *27*, 2500–2511.
- Tan, Y.Z., Cheng, A., Potter, C.S., and Carragher, B. (2016). Automated data collection in single particle electron microscopy. *Microscopy* *65*, 43–56.
- Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., and Ludtke, S.J. (2007). EMAN2: An extensible image processing suite for electron microscopy. *J. Struct. Biol.* *157*, 38–46.
- Taylor, K.A., and Glaeser, R.M. (1976). Electron microscopy of frozen hydrated biological specimens. *J. Ultrastruct. Res.* *55*, 448–456.
- Taylor, K.A., and Glaeser, R.M. (2008). Retrospective on the early development of cryoelectron microscopy of macromolecules and a prospective on opportunities for the future. *J. Struct. Biol.* *163*, 214–223.
- Tegunov, D., and Cramer, P. (2019). Real-time cryo-electron microscopy data preprocessing with Warp. *Nat. Methods* *16*, 1146–1152.
- Thompson, R.F., Iadanza, M.G., Hesketh, E.L., Rawson, S., and Ranson, N.A. (2019). Collection, pre-processing and on-the-fly analysis of data for high-resolution, single-particle cryo-electron microscopy. *Nat. Protoc.* *14*, 100–118.
- Thon, F. (1966). Zur Defokussierungsabhängigkeit des Phasenkontrastes bei der elektronenmikroskopischen Abbildung. *Zeitschrift Für Naturforsch. A* *21*, 476–478.
- Thuman-Commike, P.A., and Chiu, W. (2000). Reconstruction principles of icosahedral virus structure determination using electron cryomicroscopy. *Micron* *31*, 687–711.
- Ticau, S., Friedman, L.J., Champasa, K., Corrêa, I.R., Gelles, J., and Bell, S.P. (2017). Mechanism and timing of Mcm2–7 ring closure during DNA replication origin licensing. *Nat. Struct. Mol. Biol.* *24*, 309–315.
- Topf, M., Lasker, K., Webb, B., Wolfson, H., Chiu, W., and Sali, A. (2008). Protein Structure Fitting and Refinement Guided by Cryo-EM Density. *Structure* *16*, 295–307.
- Toyama, H., Anthony, C., and Lidstrom, M.E. (1998). Construction of insertion and deletion *mx*<sub>a</sub> mutants of *Methylobacterium extorquens* AM1 by electroporation. *FEMS Microbiol. Lett.* *166*, 1–7.
- Tsai, Y.-C.C., Ye, F., Liew, L., Liu, D., Bhushan, S., Gao, Y.-G., and Mueller-Cajar, O. (2019). Insights into the mechanism and regulation of the CbbQO-type Rubisco activase, a MoxR AAA+ ATPase. *Proc. Natl. Acad. Sci.* 201911123.
- Tsai, Y.C.C., Lapina, M.C., Bhushan, S., and Mueller-Cajar, O. (2015). Identification and

- characterization of multiple rubisco activases in chemoautotrophic bacteria. *Nat. Commun.* **6**.
- Unwin, P.N.T. (1971). Phase Contrast and Interference Microscopy with the Electron Microscope. *Philos. Trans. R. Soc. B Biol. Sci.* **261**, 95–104.
- Vetter, I.R., and Wittinghofer, A. (1999). Nucleoside triphosphate-binding proteins: Different scaffolds to achieve phosphoryl transfer. *Q. Rev. Biophys.* **32**, 1–56.
- Vilas, J.L., Tagare, H.D., Vargas, J., Carazo, J.M., and Sorzano, C.O.S. (2020). Measuring local-directional resolution and local anisotropy in cryo-EM maps. *Nat. Commun.* **11**, 55.
- Voss, N.R., Yoshioka, C.K., Radermacher, M., Potter, C.S., and Carragher, B. (2009). DoG Picker and TiltPicker: Software tools to facilitate particle selection in single particle electron microscopy. *J. Struct. Biol.*
- Wagner, T. (2019). MPI-Dortmund/sphire-janni: JANNI 0.0.5.
- Wagner, T., Merino, F., Stabrin, M., Moriya, T., Antoni, C., Apelbaum, A., Hagel, P., Sitsel, O., Raisch, T., Prumbaum, D., et al. (2019). SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, 1–13.
- Walker, J.E., Saraste, M., Runswick, M.J., and Gay, N.J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1**, 945–951.
- Wang, H.W., and Fan, X. (2019). Challenges and opportunities in cryo-EM with phase plate. *Curr. Opin. Struct. Biol.* **58**, 175–182.
- Wang, J., Hu, M., Wang, J., Qi, J., Han, Z., Wang, G., Qi, Y., Wang, H.-W., Zhou, J.-M., and Chai, J. (2019). Reconstitution and structure of a plant NLR resistosome conferring immunity. *Science* (80-. ). **364**, eaav5870.
- Wei, H., Dandey, V.P., Zhang, Z., Raczkowski, A., Rice, W.J., Carragher, B., and Potter, C.S. (2018). Optimizing “self-wicking” nanowire grids. *J. Struct. Biol.* **202**, 170–174.
- Wendler, P., Ciniawsky, S., Kock, M., and Kube, S. (2012). Structure and function of the AAA+ nucleotide binding pocket. *Biochim. Biophys. Acta - Mol. Cell Res.* **1823**, 2–14.
- White, S.R., Evans, K.J., Lary, J., Cole, J.L., and Lauring, B. (2007). Recognition of C-terminal amino acids in tubulin by pore loops in Spastin is important for microtubule severing. *J. Cell Biol.*
- Whittaker, C.A., and Hynes, R.O. (2002). Distribution and Evolution of von Willebrand/Integrin A Domains: Widely Dispersed Domains with Roles in Cell Adhesion and Elsewhere. *Mol. Biol. Cell* **13**, 3369–3387.
- Williams, D.B., and Carter, C.B. (2009). *Transmission Electron Microscopy: A Textbook for Materials Science.*
- Wong, K.S., and Houry, W.A. (2012). Novel structural and functional insights into the MoxR family of AAA+ ATPases. *J. Struct. Biol.* **179**, 211–221.
- Wong, K.S., Snider, J.D., Graham, C., Greenblatt, J.F., Emili, A., Babu, M., and Houry, W.A. (2014a). The MoxR ATPase RavA and its cofactor ViaA interact with the NADH:ubiquinone oxidoreductase I in *Escherichia coli*. *PLoS One* **9**, e85529.
- Wong, K.S., Bhandari, V., Janga, S.C., and Houry, W.A. (2017). The RavA-ViaA Chaperone-Like System Interacts with and Modulates the Activity of the Fumarate Reductase Respiratory Complex. *J. Mol. Biol.* **429**, 324–344.
- Wong, W., Bai, X.C., Brown, A., Fernandez, I.S., Hanssen, E., Condrón, M., Tan, Y.H., Baum, J., and Scheres, S.H.W. (2014b). Cryo-EM structure of the *Plasmodium falciparum* 80S ribosome bound to the anti-protozoan drug emetine. *Elife* **2014**, 1–20.
- Wu, C., Huang, X., Cheng, J., Zhu, D., and Zhang, X. (2019). High-quality, high-throughput cryo-electron microscopy data collection via beam tilt and astigmatism-free beam-image shift. *J.*

Struct. Biol. 208, 107396.

Yao, N.Y., and O'Donnell, M. (2012). The RFC Clamp Loader: Structure and Function. In *The Eukaryotic Replisome: A Guide to Protein Structure and Function*, pp. 259–279.

Ye, Q., Rosenberg, S.C., Moeller, A., Speir, J.A., Su, T.Y., and Corbett, K.D. (2015). TRIP13 is a protein-remodeling AAA+ ATPase that catalyzes MAD2 conformation switching. *Elife* 2015, 1–44.

Yi, X., Verbeke, E.J., Chang, Y., Dickinson, D.J., and Taylor, D.W. (2019). Electron microscopy snapshots of single particles from single cells. *J. Biol. Chem.* 294, 1602–1608.

Yokom, A.L., Gates, S.N., Jackrel, M.E., Mack, K.L., Su, M., Shorter, J., and Southworth, D.R. (2016). Spiral architecture of the Hsp104 disaggregase reveals the basis for polypeptide translocation. *Nat. Struct. Mol. Biol.* 23, 830–837.

Yuan, Z., Riera, A., Bai, L., Sun, J., Nandi, S., Spanos, C., Chen, Z.A., Barbon, M., Rappsilber, J., Stillman, B., et al. (2017). Structural basis of Mcm2–7 replicative helicase loading by ORC–Cdc6 and Cdt1. *Nat. Struct. Mol. Biol.* 24, 316–324.

Zala, D., Schlattner, U., Desvignes, T., Bobe, J., Roux, A., Chavrier, P., and Boissan, M. (2017). The advantage of channeling nucleotides for very processive functions. *F1000Research* 6.

Zehr, E., Szyk, A., Piszczek, G., Szczesna, E., Zuo, X., and Roll-Mecak, A. (2017). Katanin spiral and ring structures shed light on power stroke for microtubule severing. *Nat. Struct. Mol. Biol.* 24, 717–725.

Zhang, K. (2016). Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* 193, 1–12.

Zhang, X., and Zhou, Z. (2011). Limiting factors in atomic resolution cryo electron microscopy: No simple tricks. *J. Struct. Biol.* 175, 253–263.

Zhang, K., Li, S., Kappel, K., Pintilie, G., Su, Z., Mou, T.-C., Schmid, M.F., Das, R., and Chiu, W. (2019). Cryo-EM structure of a 40 kDa SAM-IV riboswitch RNA at 3.7 Å resolution. *Nat. Commun.* 10, 5511.

Zhang, L., Chen, S., Ruan, J., Wu, J., Tong, A.B., Yin, Q., Li, Y., David, L., Lu, A., Wang, W.L., et al. (2015). Cryo-EM structure of the activated NAIP2-NLRC4 inflammasome reveals nucleated polymerization. *Science* (80-. ). 350, 404–409.

Zhao, M., Wu, S., Zhou, Q., Vivona, S., Cipriano, D.J., Cheng, Y., and Brunger, A.T. (2015). Mechanistic insights into the recycling machine of the SNARE complex. *Nature* 518, 61–67.

Zheng, S.Q., Palovcak, E., Armache, J., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* 14, 331–332.

Zhong, E.D., Beppler, T., Davis, J.H., and Berger, B. (2019). Reconstructing continuous distributions of 3D protein structure from cryo-EM images. *arXiv* 1909.05215.

Zhou, M., Li, Y., Hu, Q., Bai, X.C., Huang, W., Yan, C., Scheres, S.H.W., and Shi, Y. (2015a). Atomic structure of the apoptosome: Mechanism of cytochrome c- and dATP-mediated activation of Apaf-1. *Genes Dev.* 29, 2349–2361.

Zhou, Q., Huang, X., Sun, S., Li, X., Wang, H.W., and Sui, S.F. (2015b). Cryo-EM structure of SNAP-SNARE assembly in 20S particle. *Cell Res.* 25, 551–560.

Zi Tan, Y., 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. *Nat. Methods* 14, 793–796.

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, 1–22.

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, 1–13.

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, 253–267.

Zou, J.X., Revenko, A.S., Li, L.B., Gemo, A.T., and Chen, H.-W. (2007). ANCCA, an estrogen-regulated AAA+ ATPase coactivator for ER , is required for coregulator occupancy and chromatin modification. *Proc. Natl. Acad. Sci.* 104, 18067–18072.

## V. APPENDIX

During my PhD, I contributed to a paper published on the structural characterisation of the *Serratia entomophila* anti-feeding prophage (AFP), a bacterial contractile injection system similar to the type VI secretion system and R-pyocin. AFP consists of a central tube surrounded by an external sheath, with an underlying baseplate that attaches to the host cell. During contraction, the baseplate needle penetrates the cell membrane, facilitating the injection of a protein toxin into the cell.

This manuscript was a continuation of work carried out during my BSc (Hons) at the University of Auckland in 2015, where I learned cryo-EM with Ambroise Desfosses under the supervision of Associate Professor Alok K. Mitra. During my BSc (Hons) I used cryo-EM to characterise a mutant form of AFP termed the tube baseplate complex (TBC), which lacks the apical cap protein Afp16 and therefore consists only of the central tube and baseplate. When Ambroise moved to Irina Gutsche's group, I helped to finalise the manuscript on AFP for publication, contributing to the writing, figures, and supplementary materials. The final manuscript was published in Nature Microbiology (Desfosses et al., 2019) and is presented in this appendix.

# Atomic structures of an entire contractile injection system in both the extended and contracted states

Ambroise Desfosses<sup>1,2</sup>, Hariprasad Venugopal<sup>1,3</sup>, Tapan Joshi<sup>1</sup>, Jan Felix<sup>1,2</sup>, Matthew Jessop<sup>1,2</sup>, Hyengseop Jeong<sup>4</sup>, Jaekyung Hyun<sup>4,5</sup>, J. Bernard Heymann<sup>6</sup>, Mark R. H. Hurst<sup>7,8\*</sup>, Irina Gutsche<sup>1,2\*</sup> and Alok K. Mitra<sup>1\*</sup>

**Contractile injection systems are sophisticated multiprotein nanomachines that puncture target cell membranes. Although the number of atomic-resolution insights into contractile bacteriophage tails, bacterial type six secretion systems and R-pyocins is rapidly increasing, structural information on the contraction of bacterial phage-like protein-translocation structures directed towards eukaryotic hosts is scarce. Here, we characterize the antifeeding prophage AFP from *Serratia entomophila* by cryo-electron microscopy. We present the high-resolution structure of the entire AFP particle in the extended state, trace 11 protein chains de novo from the apical cap to the needle tip, describe localization variants and perform specific structural comparisons with related systems. We analyse inter-subunit interactions and highlight their universal conservation within contractile injection systems while revealing the specificities of AFP. Furthermore, we provide the structure of the AFP sheath-baseplate complex in a contracted state. This study reveals atomic details of interaction networks that accompany and define the contraction mechanism of toxin-delivery tailocins, offering a comprehensive framework for understanding their mode of action and for their possible adaptation as biocontrol agents.**

Contractile injection systems (CIS) consist of a long tube with a needle-like tip, a helical sheath surrounding the tube and a baseplate linking external signals with sheath contraction that propels the tube out of the sheath, enabling the needle to perforate the target membrane<sup>1,2</sup>. In addition to contractile bacteriophage tails, CIS include bacterial contractile phage tail-like particles<sup>3</sup>, which are further subdivided into the intracellular type VI secretion system (T6SS)<sup>4–6</sup> and extracellular CIS (eCIS). The latter includes R-pyocins<sup>7–9</sup>, which dissipate the proton motive force of their target bacteria, as well as phage-like protein-translocation structures (PLTS)<sup>10–12</sup>, which inject specific toxic proteins into eukaryotic cells. This major difference with pyocins led to the qualification of PLTS as toxin-delivery tailocins<sup>3,13</sup>, thereby emphasizing their relationship with both pyocins and T6SS<sup>4–6</sup>. These systems, bioinformatically identified in hundreds of bacteria and archaea dwelling in different environments<sup>12</sup>, are exemplified by the *Serratia* antifeeding prophage (AFP)<sup>10,13</sup> and the *Photorhabdus* virulence cassette (PVC)<sup>11,14</sup>—targeting insect larvae and haemocytes, respectively—and are similar to the metamorphosis-associated contractile structures<sup>15</sup> and to the intracellular *Amoebophilis asiaticus* CIS<sup>6,16</sup>.

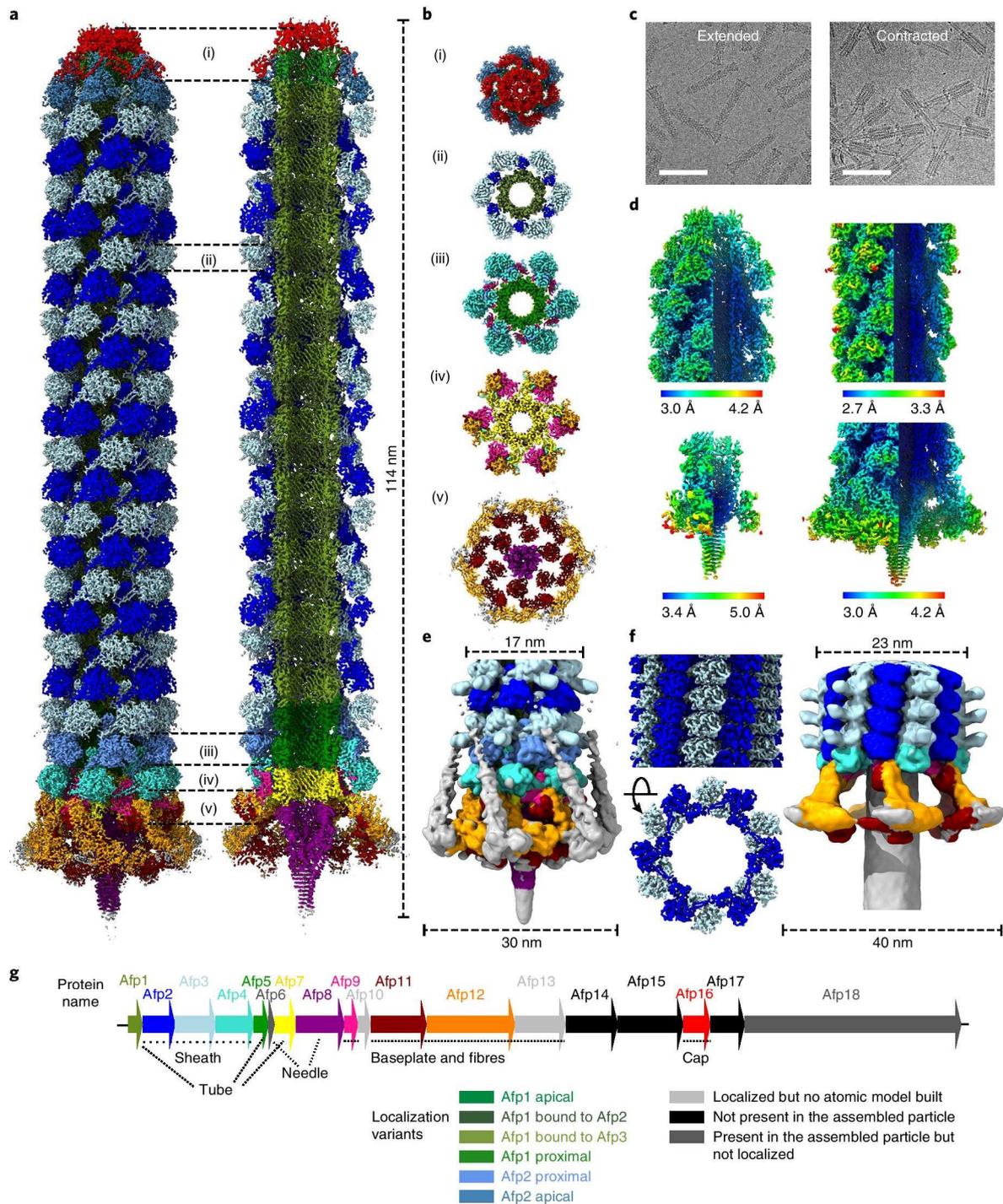
The cryo-electron microscopy (cryo-EM) ‘revolution’ has enabled spectacular progress in the structural characterization of both T6SS and R-pyocin as well as the related myophage T4 (refs. <sup>9,17–20</sup>). However, structural understanding of PLTS is lagging behind, with only a single recent publication detailing the structures of the entire extended PVC particle and the contracted PVC

sheath<sup>14</sup>. Here, we present high-resolution structures of AFP—the archetype PLTS from the soil bacterium *Serratia entomophila*<sup>10,13</sup>—in both the extended and contracted states, including the previously unknown structure of a contracted PLTS baseplate. This analysis enables us to propose a structural mechanism of AFP action as a toxin-delivery syringe and to provide a detailed comparison with other CIS. Conservation of the baseplate components among the PLTS suggests that the inferred contraction mechanism is probably shared by all toxin-delivery tailocins.

## Results

**Overall structure of the AFP particle.** The *S. entomophila* AFP gene cluster<sup>10</sup> comprises 18 open reading frames with translation products denoted as Afp1 to Afp18 (Supplementary Table 1). AFP translocates the insecticidal toxin Afp18 into larvae of the New Zealand pasture pest *Costelytra giveni*, thereby causing cessation of feeding activity<sup>10,21</sup>. The extended AFP is a 110 nm-long bullet-shaped particle built up by a helical trunk with a conical cap at the apical end and a flat base with a mobile fibre network at the proximal end (Fig. 1). To obtain a high-resolution cryo-EM structure of the entire AFP particle, its images were divided into three parts—the trunk (containing the sheath and inner tube), the cap and the baseplate. Each part was reconstructed separately and recombined to give a final composite cryo-EM map (see Methods; Fig. 1a,b,d, Supplementary Figs. 1,2 and Supplementary Table 2). This approach allowed us to: (1) benefit from the helical symmetry

<sup>1</sup>School of Biological Sciences, University of Auckland, Auckland, New Zealand. <sup>2</sup>Institut de Biologie Structurale (IBS), University Grenoble Alpes, CEA, CNRS, Grenoble, France. <sup>3</sup>Ramaciotti Centre for Cryo Electron Microscopy, Monash University, Melbourne, Victoria, Australia. <sup>4</sup>Electron Microscopy Research Center, Korea Basic Science Institute, Cheongju-si, Republic of Korea. <sup>5</sup>Molecular Cryo-Electron Microscopy Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan. <sup>6</sup>Laboratory for Structural Biology Research, National Institute of Arthritis, Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD, USA. <sup>7</sup>Forage Science, AgResearch, Lincoln Research Centre, Christchurch, New Zealand. <sup>8</sup>Bio-Protection Research Centre, Christchurch, New Zealand. \*e-mail: mark.hurst@agresearch.co.nz; irina.gutsche@ibs.fr; a.mitra@auckland.ac.nz



**Fig. 1 | Cryo-EM structures of the complete AFP contractile injection system in the extended and contracted states. a**, Composite cryo-EM map, assembled from individual maps of the AFP cap, trunk, baseplate and needle, showing the overall organization of AFP in the extended state. The map corresponds to the main population of AFP particles terminating in Afp2 below the apical cap. The different subunits are coloured according to **g**. **b**, Cut-out views of the different sections labelled i–v in **a**. **c**, Representative micrographs displaying AFP particles in the extended (left) and contracted (right) state. Scale bars, 100 nm. **d**, Cryo-EM maps of the apical end cap (upper left), sheath and tube (upper right), needle (lower left) and baseplate (lower right) coloured according to the local resolution. **e**, Cryo-EM map of AFP in the extended state, filtered to a resolution of 10 Å, allowing a better visualization of the Afp13 tail fibres and Afp3 protrusions. **f**, Cryo-EM maps of the sheath (left) and baseplate (right) of AFP in the contracted state with the different subunits coloured according to **g**. The lengths and/or widths of the maps shown in **a**, **e**, **f** are annotated using dashed lines. **g**, Schematic representation of the genomic organization of the AFP gene cluster.

of the trunk while applying a six- or threefold rotational symmetry to the baseplate and needle and a sixfold symmetry to the cap, and (2) circumvent long-range distortions and separate two populations

of AFP particles that differ in terms of the trunk length (Fig. 1a and Supplementary Figs. 1,2). The resolution of most of the structure ranges from 2.8 to 3.3 Å, which is overall slightly higher than

that of the recently published PVC structure<sup>14</sup> and falls off towards the peripheral trunk protrusions, baseplate edges and attachment fibres (Fig. 1d,e, Supplementary Fig. 3 and Supplementary Table 2). De novo atomic model building for nearly all proteins composing the AFP particle (Supplementary Table 3) was made possible by the high quality of the maps (Fig. 1 and Supplementary Fig. 3), which reflects the regularity of the assembly and especially of the best-resolved inner tube, which needs to be rigid enough to penetrate the cell membrane following contraction. In the contracted state (Fig. 1f and Supplementary Figs. 2,4,6), the length of the sheath decreases from 850 to 350 Å and the diameter swells from 170 to 230 Å to eject the inner tube through a rearranged baseplate. The post-contraction baseplate (global resolution of 5.4 Å, ranging from 4 Å for the bottom of the sheath to about 12 Å for the outer rim of the baseplate) is less well resolved than the sheath (resolution of 3.8 Å), presumably because of the loss of the needle–baseplate interactions that are required for the tube and needle release.

The final atomic model of the extended AFP contains a total of 11 different proteins (Fig. 1). Three proteins make up the inner tube (Afp1, Afp5 and Afp7), three others build the surrounding sheath (Afp2, Afp3 and Afp4), two comprise the needle (Afp10 and Afp8), three belong to the baseplate (Afp9, Afp11 and Afp12) and one constitutes the apical cap (Afp16). The fibres formed by Afp13 were not well resolved (Fig. 1e and Supplementary Fig. 3), which is consistent with their proposed function in sensing the target-cell surface<sup>10</sup>, requiring mobility and flexibility to orient the AFP particle and position the baseplate for the power stroke<sup>2</sup>. Mass-spectrometry measurements (see Methods; Supplementary Table 1) identified a single copy of the Afp18 toxin, presumably represented by an unassigned density inside the tube, as well as the small and uncharacterized Afp6, which may either be also inside the tube or at the baseplate periphery associated with fibres. These measurements also confirmed the absence of the tape measure protein Afp14 (refs. 22,23), assembly chaperone AAA+ ATPase Afp15 (ref. 22) and ‘inactive pseudotoxin’ Afp17 (refs. 22,24) in the mature particle. Details about AFP proteins and their equivalents in other CIS are summarized in Supplementary Table 1.

**AFP tube.** As in the T4 phage<sup>20</sup>, the first ring of the AFP tube is assembled at the baseplate from Afp7, followed by a ring of Afp5 and subsequent rings of Afp1 that span the interior of the entire particle up to the apical end formed by the tail-tube terminator cap of Afp16 (Figs. 1a and 2). Afp1, Afp5 and Afp7 have a common fold (Fig. 2b), which is also shared with the T4 gp19, the T6SS Hcp and the R-pyocin, phage lambda gpV<sup>25</sup> and T5 pb6 (ref. 26) tube proteins (Supplementary Fig. 5a,b). Similarly to T4 gp19, the inner surface of the tube formed by Afp1 is negatively charged, possibly to allow the efficient release of the Afp18 toxin, which would be negatively charged at pH 11 in the gut of *C. giveni* larvae<sup>27</sup> (Supplementary Fig. 5c). Like Pvc7, Afp7 possesses a LysM-like<sup>28</sup> peptidoglycan-binding domain (Fig. 2e,g,h) and links the C6 symmetry of the tube to the C3 symmetry of the needle protein Afp8 (Fig. 2f and Supplementary Fig. 9c). At the apical end, in the cap-bound conformation, the N-terminal Afp1  $\alpha$ -helix unfolds to follow the interdomain loop of Afp16 (Fig. 2b,c). Each apical Afp1 strongly binds the cap by interacting with three Afp16 subunits, presumably to prevent tube release from the opposite end following contraction (also see below).

The  $\beta$ -hairpins of the adjacent tube subunits constitute a  $\beta$ -sheet barrel with a diameter of 75 nm in the middle of the tube. Each Afp1 monomer interacts with four subunits from the Afp1 layer above and four subunits from the layer below (Fig. 2d); the main interaction occurs through  $\beta$ -sheet intercalation as in other CIS<sup>9,18,29</sup> and a continuous mesh of such  $\beta$ -sheet intercalations spreads through the entire particle, from the needle to the cap (Fig. 2c–f). The free energy of these tube-stabilizing interactions and comparisons between the intra-tube, intra-sheath and tube–sheath interactions in AFP,

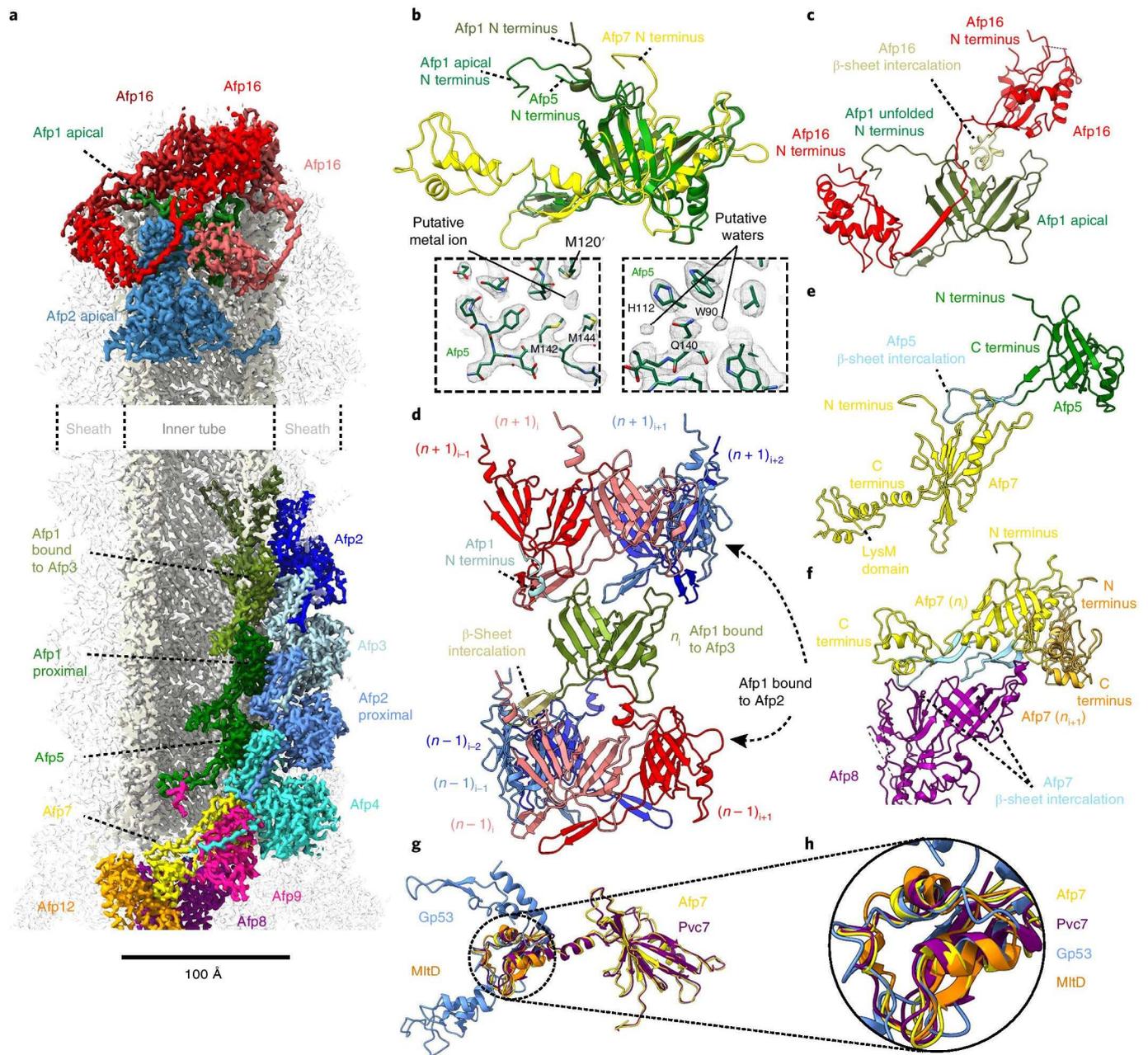
R-pyocin and T6SS are detailed in Supplementary Tables 4 and 5. The observation that the intra-tube interactions, of similar strength in all three assemblies, are much stronger than the consistently weak tube–sheath interactions reflects the need for the rigidity of the tube following perforation of the target cell and the necessity for the tube to slide through the sheath during contraction.

**AFP sheath.** The most striking feature of the AFP sheath is that it is constructed of two alternating hexameric layers, each composed of a different sheath protein<sup>13</sup> (Fig. 3). A comparative genomic study revealed that most PLTS contain one sheath-encoding gene, although some have two and *Enterobacteria* (including *Serratia*) have three<sup>12</sup>. Afp2, Afp3 and Afp4 have a conserved fold (Fig. 3e and Supplementary Fig. 6) integrating elements of domains 1 and 2 of VipA-B/TssB-C<sup>17</sup>. At the proximal end, the sheath starts with a single hexameric ring of Afp4 followed by a ring of Afp2, and then Afp3 and Afp2 rings alternate up to the apical Afp16 cap. 77% of the AFP particles are built by 11 Afp2 and 10 Afp3 layers surrounding 21 layers of the tube protein Afp1 and resulting in a terminal Afp2–Afp16 interaction (Figs. 1a and 2a); 23% contain only 10 Afp2 layers and 10 Afp3 layers surrounding 20 Afp1 layers, and therefore terminate with Afp3–Afp16 (Supplementary Figs. 1,3). The tube–sheath interactions may be considered identical along the length of the trunk, ignoring negligible differences in the contacts between Afp1 and either Afp2 or Afp3 (Supplementary Fig. 6d).

Afp3 is the longest of the three sheath proteins and forms a poorly resolved N-terminal bulge responsible for the characteristic appearance of the AFP trunk<sup>13</sup>, which is carved by six prominent ridges of Afp3 with deep Afp2 grooves in between (Fig. 1e and Supplementary Fig. 3d). In contrast to AFP, PVC has been described<sup>14</sup> to be built of 10 layers of alternating Pvc2 and Pvc3 followed by 12 layers assembled exclusively of Pvc2. However, examination of the entire PVC map (EMDB ID: EMD-9762) suggests that this interpretation may have resulted from processing a mixed population of PVC particles, some of which may have been composed of Pvc2 and Pvc3 that actually alternated beyond the end of the back-folded fibres, further towards the apical cap, as is the case for AFP. Therefore, the N-terminal bulge domain of Afp3/Pvc3 may play another yet unidentified role in addition to the fibre-docking function proposed based on the patterning of the PVC sheath<sup>14</sup>. The shape of this domain, which is partially conserved between Afp3 and Pvc3 (Supplementary Fig. 6f,g), is to some extent reminiscent of the slightly bigger domain 3 of VipA-B/TssB-C. However, the latter is known to be recognized by the cognate AAA+ ATPase ClpV in the contracted T6SS sheath to specifically trigger its disassembly and recycling<sup>17,18,30</sup>, whereas the AAA+ ATPase in the PLTS is proposed to instead be involved in the assembly of the baseplate and the needle<sup>22</sup>.

As in R-pyocins and T6SS, the sheath assembly appears to rely on a  $\beta$ -strand exchange in the innermost layer of the sheath, termed a ‘handshake’ mechanism<sup>9,17</sup> (Fig. 3a,d). As described above for the intra-tube intercalations, the sheath-stabilizing mesh of  $\beta$ -sheet handshakes propagates from the baseplate of the AFP particle to the cap (Fig. 3d). Afp4 forms the first baseplate-proximal sheath ring by offering the corresponding  $\beta$ -strand to the sheath initiator Afp9 while accepting an analogous  $\beta$ -strand of Afp2, which in turn receives a  $\beta$ -strand from Afp3, and this pattern continues up to the top-most sheath layer. Furthermore, in the cap, the tail-terminator Afp16 mimics this handshake interaction by binding into a cleft in the terminal Afp2/3 (Fig. 3d(i); see below).

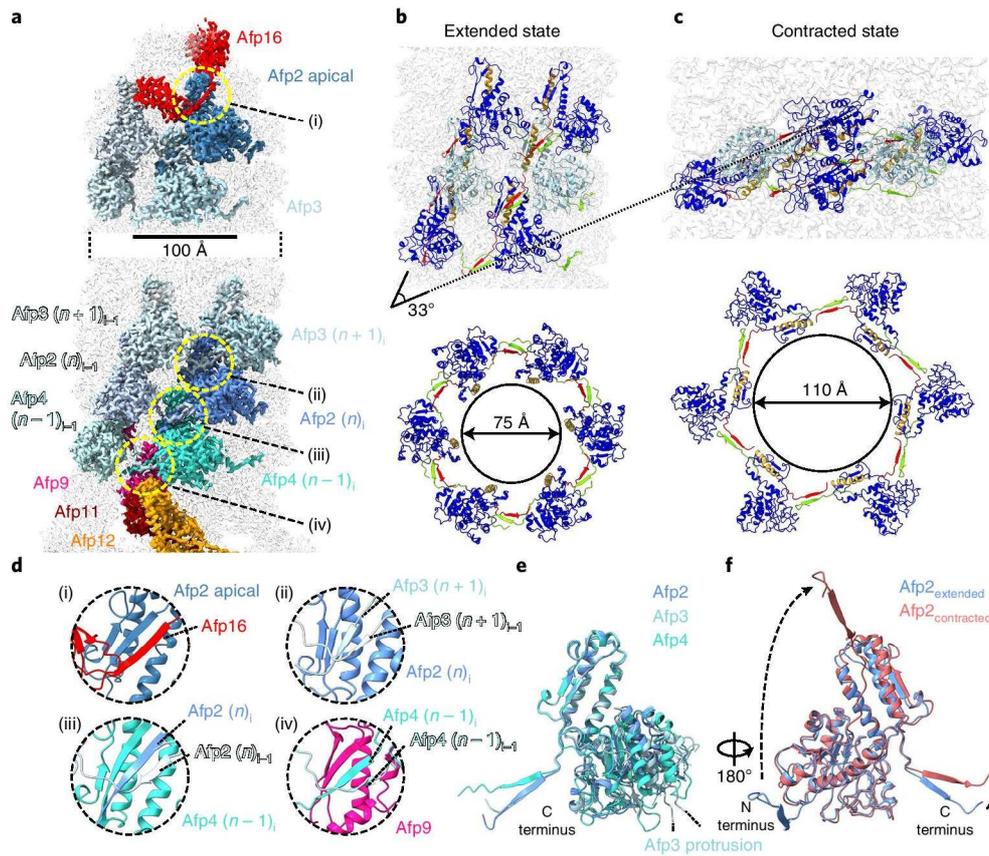
Although contraction was experimentally achieved by briefly subjecting AFP particles to guanidine hydrochloride (see Methods), no loss of protein integrity was detected (Supplementary Fig. 6c). Similarly to R-pyocins<sup>9</sup>, the handshake inter-subunit connectivity does not change between the extended and the contracted state,



**Fig. 2 | Molecular organization of the inner tube of AFP in the extended state. a**, Cryo-EM map of the extended state of AFP displaying select subunits involved in the interactions within the inner tube, between the inner tube and the sheath, and between the inner tube and the apical cap. **b**, Structural alignment between Afp1 in the inner tube (olive green), Afp1 in the apical cap (dark green), Afp5 (light green) and Afp7 (yellow). Note the differences in conformation of the N terminus of Afp1 in the apical cap ( $\alpha$ -helical turn) and inner tube (unfolded). Insets: Afp5 built into the cryo-EM map of the trunk of AFP in the extended state to illustrate that the Afp5 ring is the best-resolved baseplate component (2.9 Å resolution), with the structure featuring a putative metal ion coordinated by three methionines (labelled in black; left) and putative water molecules (right). **c**, Interaction between Afp16 and Afp1 in the apical cap. The loop of Afp16 that intercalates between the central  $\beta$ -sheet of Afp1 is labelled. The N terminus of Afp1 bound to the cap (Afp1 apical) is unfolded and interacts with Afp16. **d**, Interactions between one copy of Afp1 contacting Afp3 (labelled 'Afp1 bound to Afp3'; olive green) in layer  $n$  and four copies of Afp1 contacting Afp2 (labelled 'Afp1 bound to Afp2'; coloured from red to blue) in the layers  $n-1$  (bottom) and  $n+1$  (top) of the inner tube. The N terminus of Afp1 in layer  $n$  forms an  $\alpha$ -helical turn and is indicated in light blue. **e**, Interaction between Afp5 and Afp7. The loop of Afp5 that intercalates between the central  $\beta$ -sheet of Afp7 is labelled. The C-terminal domain of Afp7 (LysM domain) is labelled. **f**, Interaction between two adjacent copies of Afp7, indicated in yellow ( $n_i$ ) and orange ( $n_{i+1}$ ), and Afp8 (purple). The loops of the two Afp7 copies that intercalate between the central  $\beta$ -sheet of Afp8 are shown in two tints of light blue. **g**, Structural alignment of LysM domains of Afp7, Pvc7 (PDB ID: 6J0N), gp53 (PDB ID: 5IV5) and MitD (PDB ID: 1E0G). **h**, Zoom of the LysM domain of the structural alignment shown in **g**.

presumably maintaining the sheath integrity (Fig. 3b,c and Supplementary Fig. 6a,b). Contraction leads to a rigid-body rotation of each sheath subunit by roughly 33° and an outward shift of

40 Å, and the assembly length compresses approximately twofold. The switch from the extended to contracted state results in major displacements of the Afp2 and Afp3 N- and C-termini, with the



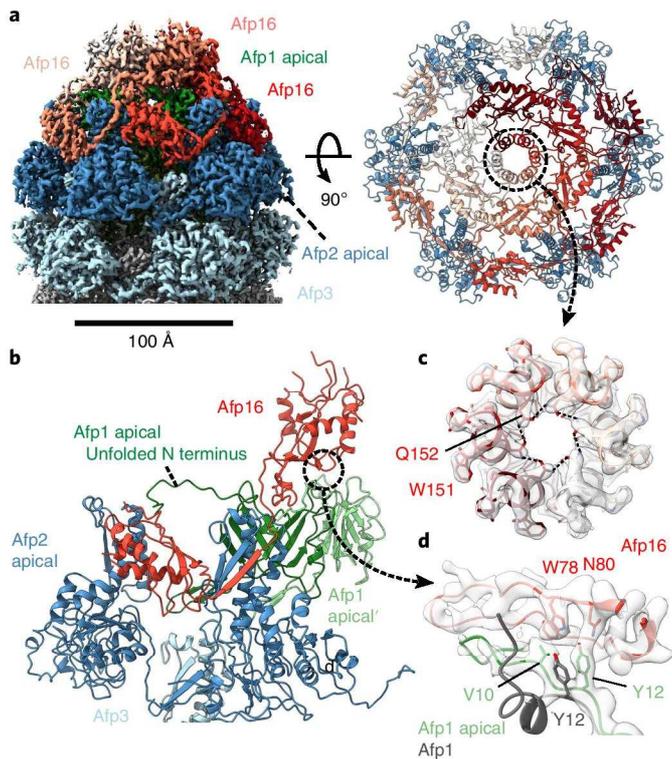
**Fig. 3 | Molecular organization of the AFP sheath in the extended and contracted states.** **a**, Cryo-EM map of AFP in the extended state displaying select subunits involved in the interactions within the sheath, between the sheath and apical cap, and between the sheath and the baseplate. **b,c**, Molecular rearrangement of the Afp2 (dark blue) and Afp3 (light blue) subunits in the sheath following AFP contraction. The N and C termini of both Afp2 and Afp3 are coloured in green and red, respectively; the tube-interacting helix (Afp2 residues 277–297 and Afp3 residues 374–394) is coloured in orange. The angle between alternating Afp2 and Afp3 subunits changes by 33° (between **b**, top; extended state and **c**, top; contracted state) and the diameter of the sheath expands from 75 Å (**b**, bottom; extended state) to 110 Å (**c**, bottom; contracted state). **d**, Zoomed-in views of the dotted yellow circles annotated in **a**. Different instances of a  $\beta$ -strand exchange (handshake mechanism) between Afp2 and Afp16 in the apical cap (i), Afp2 and Afp3 in the sheath (ii), Afp2 and Afp4 in the upper baseplate (iii) and Afp4 and Afp9 in the lower baseplate (iv) are shown. **e**, Structural alignment between Afp2, Afp3 and Afp4. The location of the Afp3 protrusion (not modelled) is indicated with dashed lines. **f**, Structural alignment between Afp2 in the extended and contracted states. Note the large rearrangements of both the N and C termini of Afp2 following AFP contraction.

Afp2 N-terminus rotating more than 120° (Fig. 3f). Following contraction and radial expansion, the subunits of the upper rings insert in the newly formed gaps of the lower rings. Therefore, in spite of the loss of the contact with the tube, each sheath subunit buries circa 900 Å<sup>2</sup> more of its surface (Supplementary Table 6), meaning that a total of circa 730 kcal mol<sup>-1</sup> of energy would be released during a single contraction, resulting in the tube expulsion and drilling into the target cell membrane (Supplementary Tables 5, 6). Interestingly, the difference in free energy between the extended and contracted AFP sheath is much smaller than in T6SS and R-pyocin (Supplementary Table 5), probably reflecting the remarkable stability of extended AFP at a basic pH in the gut of the host.

The sheath–tube interaction appears universally conserved among CIS and mainly relies on complementary surface charges on contacting interfaces (see Supplementary Fig. 6d,e for the conservation of Afp2 and Afp3 residues involved in Afp1 binding.) The fact that the AFP particle inside the cell assembles into a metastable extended state may be explained by our earlier observation that the sheath polymerizes on the already pre-assembled baseplate–tube scaffold<sup>22</sup>, which is corroborated by the recent investigation of PVC mutants<sup>14</sup>. In the proposed directional assembly model, the tube polymerizes on the pre-assembled baseplate until it reaches its

optimal length, controlled by the tape measure protein, and is then apically capped to prevent further growth and to prime the sheath loading<sup>14,22</sup>. The difference between the assembly mechanisms proposed for toxin-delivery tailocins and for T6SS where the tube stays capped during growth<sup>31,32</sup> may be explained by the unique nature of the PLTS cap and its mode of interaction.

**AFP apical-end cap.** The cryo-EM maps of the apex of the major and minor AFP populations are virtually identical, independent of whether the terminal layer under the Afp16 cap is comprised of Afp2 or Afp3 (Fig. 4 and Supplementary Fig. 3c,f). The Afp16 protein is composed of two distinct domains linked by a long  $\beta$ -strand-containing loop and forms a hexameric ring at the apex of the trunk. The N-terminal domain of Afp16, folded similarly to gp15 (T4)<sup>33</sup> and gpU (phage lambda)<sup>34</sup> but differently from TssA (T6SS), constitutes an extra layer of the tube while the PLTS-specific C-terminal domain, slightly shifted inwards in AFP in comparison to PVC, caps the sheath (Fig. 4a–c and Supplementary Fig. 7a,b). The N-terminal Afp16 domains strongly bind to two Afp1 subunits underneath via a  $\beta$ -sheet intercalation, the linker  $\beta$ -strand of Afp16 docks into the C-terminal domain of one Afp2 subunit in a handshake manner (Figs. 3d(i) and 4a,b), while the C-terminal Afp16 domain interacts



**Fig. 4 | Architecture of the apical cap of AFP in the extended state.**

**a**, Cryo-EM map of the apical cap of AFP in the extended state displaying the subunits Afp1, Afp2, Afp3 and Afp16 (left). Top view of fitted Afp2 (dark blue) and Afp16 (white to dark red) subunits (right). Note the central ring formed by a side-by-side stacking of the helix spanning residues 143–153 of the neighbouring Afp16 subunits (annotated by a dotted black circle). **b**, Interactions between Afp1, Afp2, Afp3 and Afp16 in the apical cap. The linker  $\beta$ -strand (residues 184–191) of Afp16 is involved in a handshake with Afp2. **c**, Density map of the central ring in the apical cap with fitted  $\alpha$ -helices containing residues 143–153 of the neighbouring Afp16 subunits. The amino acids Q152 and W151 are labelled and shown as sticks. **d**, Density map displaying the interaction between Afp16 (red) and the N terminus of Afp1 in the apical cap (green). The Afp1 amino acids V10 and Y12, and Afp16 W78 and N80 are labelled and shown as sticks. An alignment of Afp1 in the sheath with apical Afp1 illustrates the rearrangements of the N-terminal residues V10 and Y12 following Afp16 binding.

with a neighbouring Afp2 and even stretches down to an Afp3 subunit one layer below (Fig. 4a,b). The N-terminal domain of Afp16 possesses a short  $\alpha$ -helical insertion which, following Afp16 hexamerization, organizes in a tight ring with a diameter of 9 Å that constricts the extremity of the particle and probably ensures the directionality of the toxin egress through the baseplate and the needle (Fig. 4c and Supplementary Fig. 7a). Noteworthy, no capping of the AFP or PVC sheath was observed in the contracted state.

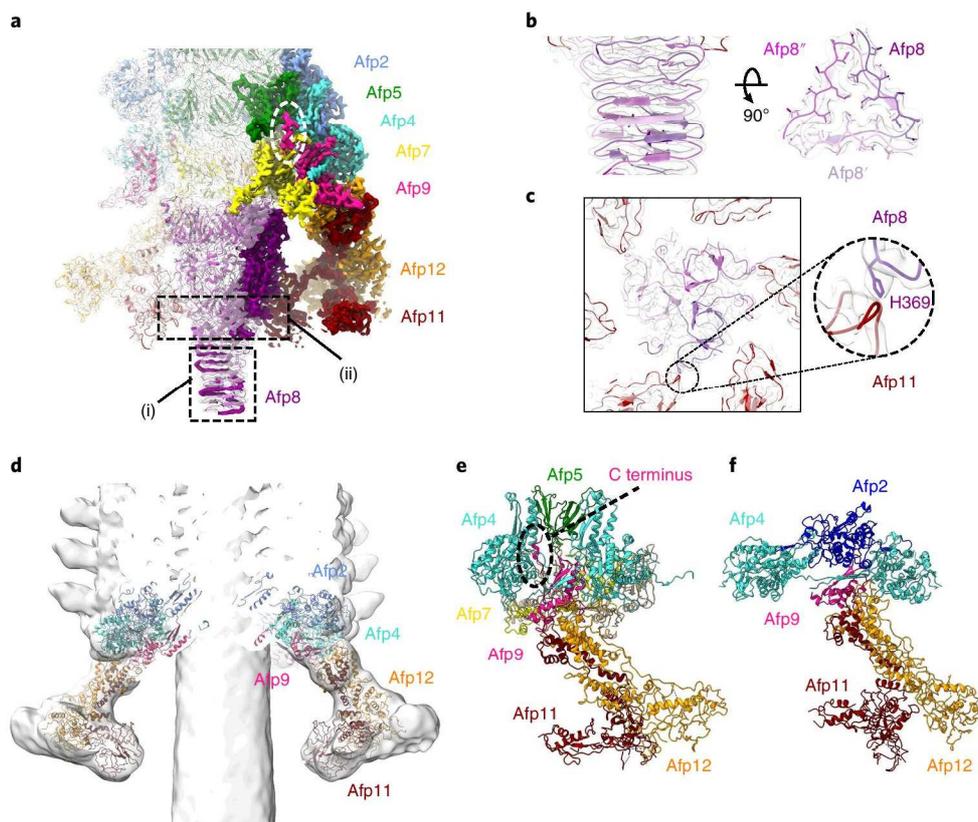
**AFP baseplate and needle.** The overall architecture of the AFP baseplate is similar to PVC and follows the same organization as the conserved inner baseplate of phage T4 (Supplementary Fig. 8). In comparison with the sophisticated full T4 baseplate<sup>20</sup>—which contains a complex tail-fibre network—the AFP, PVC<sup>14</sup> and T6SS<sup>35–37</sup> baseplates are elegantly simple (Supplementary Fig. 8). In agreement with other CIS, the AFP baseplate can be described as a C6-symmetrical assembly of sub-complexes historically called ‘wedges’ that circularize around the C3-symmetrical needle<sup>2</sup>. The structure of the needle-baseplate complex (Fig. 5) was solved to an average resolution of 3.1 Å (see Methods and Supplementary Fig. 1 for details). As in T6SS

VgrG<sup>4</sup>, Afp8 is a fusion of the hub and the needle genes, whereas the tip of the needle is encoded by *afp10* (gp5.4 in T4; VgrG PAAR in T6SS; Supplementary Fig. 9e). Reminiscent of PVC, T4 (ref. <sup>38</sup>) and T6SS<sup>39</sup>, the needle hub is composed of twin  $\beta$ -barrel domains of Afp8 with a fold similar to the core domain of the tube proteins (Fig. 5b,c). This creates a pseudo-hexameric layout that accommodates the transition from the hexameric ring formed by the tube initiator Afp7 (gp48 equivalent) to the C3-symmetry of the rest of the Afp8 needle trimer (Supplementary Fig. 9c). Thus, the top surface of Afp8 packed against the bottom of Afp7 plays the C6-to-C3 symmetry adaptor function attributed to gp27 (ref. <sup>20</sup>). Interestingly, the similarity to gp27 extends even beyond the twin domains (Supplementary Fig. 9e). The bottom of Afp8 forms a tapered needle with an intercalated  $\beta$ -helical wall analogous to the T4 gp5 and T6SS VgrG/VgrG1 but notably shorter; this difference in length (Supplementary Fig. 9e) may be related to the specificity of AFP for eukaryotic targets. The tip of the C3-symmetrical needle is sharpened by a single copy of Afp10. Remarkably, the Afp8-bound region of Afp10 follows a pseudo-C3 symmetry, resulting in three clearly resolved  $\beta$ -sheets characteristic of gp5.4, the T4 homologue of Afp10, and thus enabling fitting of gp5.4 into the density (Supplementary Fig. 9f).

The AFP wedge is comprised of the sheath-initiator protein Afp9, featuring an EPR motif conserved over a variety of CIS, and two other baseplate proteins, Afp11 and Afp12 (Fig. 5e,f). In contrast, the T4 wedge is composed of gp25, gp53 and a module built of two copies of gp6 and one copy of gp7 stabilized by an  $\alpha$ -helical core bundle<sup>20</sup>. Despite these apparent differences, both wedges share a remarkably similar architecture. Furthermore, Afp12 can be regarded as a hybrid between gp7 and one of the gp6 copies, whereas the second gp6 is mimicked by Afp11. Indeed, the alignment of Afp9 and gp25 overlays Afp11 and Afp12 with (gp6)<sub>2</sub>-gp7 (Supplementary Fig. 9d). This structural conservation, which is unexpected from the sequence analysis, highlights the importance of the Afp11–Afp12 module in the assembly and function of the baseplate. Close to the particle axis, Afp11 folds back to loosely bind the Afp8 needle just above the  $\beta$ -helix. Because of the C3-symmetry of the latter, three copies of Afp11 interact with Afp8 more tightly than the other three (Fig. 5c). This fold-back interaction is different from those observed in T4 (ref. <sup>20</sup>) and is mostly polar. The disorder observed in this region of Afp11 may facilitate needle ejection following contraction. Situated at the periphery of the baseplate, Afp12 hosts the proximal end of the Afp13 tail fibre, which can be visualized in the local-resolution-filtered maps and at lower thresholds (Supplementary Fig. 3b,e). As proposed for T4 (ref. <sup>20</sup>), recognition of receptors on the target cell surface may trigger rotation of the wedge and the fibres thereby initiating contraction.

## Discussion

Comparison between the structures of the extended and contracted forms leads us to propose a key role for the sheath initiator protein Afp9. Despite its relatively small 16 kDa size, Afp9 would act as a stabilizer and multivalent connector protein, crucial for the assembly and structural integrity of the whole AFP particle. It is located at the inwards-pointing apex of the baseplate wedge and links the helical trunk to the Afp11–Afp12 bulk of the baseplate. Specifically, in the extended state, while the inner face of the Afp9 ring surrounds the tube initiator Afp7 and reaches to Afp5, its outer top side positions the first sheath ring formed by Afp4 and its lower external loops hold six Afp11–Afp12 modules in place (Fig. 5a,e). Comparison between the structures of the extended and contracted states reveals the crucial role of Afp9 as a power-stroke initiator. Indeed, when the Afp13 fibres transmit the contraction signal through Afp12 to the rest of the wedge, the Afp11–Afp12 module detaches from the needle and swings outwards (Fig. 5d and Supplementary Fig. 9a,b). This momentum is transmitted to Afp9, which, in the contracted state, ensures the main connection between the baseplate wedge and the sheath (Fig. 5d,f). The exact sequence of events and the



**Fig. 5 | Molecular organization of the AFP baseplate in the extended and contracted states.** **a**, Cryo-EM map of the AFP baseplate in the extended state. Select subunits involved in interactions between the tube, sheath and baseplate, and within the baseplate itself are displayed as non-transparent density, whereas the transparent density contains the structures of the subunits Afp2, Afp4, Afp5, Afp7, Afp9, Afp11 and Afp12. **b**, Density map of the AFP needle formed by three neighbouring Afp8 subunits (zoom of (i) in **a**). The hydrophobic residues in the threefold symmetric centre of the needle tip are shown as sticks. **c**, Pseudo-hexameric symmetry of the upper part of the AFP needle (zoom of (ii) in **a**). Note the alternating presence/absence of interactions between the H369 residue of Afp8 and a loop of Afp11 indicating the transition between sixfold to threefold symmetry in the AFP needle. **d**, Cryo-EM map of the AFP baseplate in the contracted state with fitted subunits Afp2, Afp4, Afp9, Afp11 and Afp12. **e, f**, Comparison between the structural organization of the extended (**e**) and contracted (**f**) baseplate. Note the  $\alpha$ -helical conformation of the C terminus of Afp9 in the extended (**e**) AFP state (annotated with a dotted black oval), whereas it is unstructured and thus not present in the contracted (**f**) AFP map.

cause-and-effect relationships require further investigation but examination of the protein interaction networks inside the structures leads us to speculate that the swinging motion of the Afp9 connector with the whole of the wedge may transfer the contraction signal to Afp4 through two routes. The C-terminal  $\alpha$ -helix of Afp9 would act as a hook pulling on the adjacent  $\alpha$ -helix of Afp4, thereby causing rotation of the Afp4 subunit and its detachment from the tube Afp5. In addition, because Afp9 is the sheath initiator and the receiver of the first handshake from Afp4, its reorientation and the resulting movement of Afp4 would activate the transmission belt of handshakes that would spread from Afp4 to Afp2 and further up to the Afp16 cap. This would in turn trigger the sheath contraction, leading to the release of the tube and the needle, perforation of the target membrane and injection of the toxin into its grass grub prey.

In conclusion, this work presents the atomic-resolution structures of an entire contractile injection nanomachine in both the extended and contracted states. It provides a molecular basis for understanding the PLTS mode of biological action and regulation, illustrates the degree of the universal conservation of CIS and offers a framework for their analysis with unprecedented detail. In addition, information offered by the AFP structures may be further used for biomedical design. One such example would be the rational modification of the AFP particle and its attachment fibres towards targeting specific mammalian cell surfaces for the injection of a tailored biocontrol agent.

## Methods

**AFP expression and purification.** Recombinant AFP was obtained as previously described<sup>13</sup>. The purified AFP sample was stored at 4 °C in TM buffer (20 mM Tris-HCl and 20 mM MgCl<sub>2</sub>, pH 7.4). Contracted AFP was generated by incubating purified AFP samples with guanidine hydrochloride at a final concentration of 3 M for 1 min at room temperature and then diluted in a large volume of TM buffer to stop the guanidine-hydrochloride action. The samples were then concentrated by ultracentrifugation for 1.5 h at 150,000g and the pellet was resuspended in TM buffer. The cryo-EM specimens were then prepared immediately (see below).

**Mass spectrometry.** Briefly, protein samples were resolved by SDS-PAGE and bands were excised, destained, reduced with dithiothreitol, alkylated with iodoacetamide and then digested with sequencing-grade porcine trypsin (Promega) in a chilled microwave (CEM) at 45 °C for 1 h at 15 W. A 10  $\mu$ l aliquot of each acidified digest was desalted on a 0.3  $\times$  10 mm trap column packed with 3  $\mu$ m Reprisil C18 media (Dr Maisch). The peptides were then separated on a 0.075  $\times$  200 mm picofrit column (New Objective) packed in-house with 3  $\mu$ m Reprisil C18 media using the following gradient of 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (B) at 300 nl min<sup>-1</sup>: 0 min, 5% B; 0.5 min, 5% B; 47 min, 30% B; 49 min, 98% B; 52 min, 98% B; 53 min, 5% B and 60 min, 5% B. The picofrit spray was directed into a TripleTOF 6600 Quadrupole-Time-of-Flight mass spectrometer (Sciex) operating in positive-ion mode, scanning 350–1,200  $m/z$  for 250 ms, followed by up to forty tandem mass spectrometry scans of 40 ms per cycle (80–1,600  $m/z$ ) on multiply charged species using Dynamic Collision Energy and Dynamic Accumulation.

The resulting tandem-mass-spectrometry data was then searched against an in-house protein sequence database containing selected entries from *S. entomophila* using ProteinPilot v5 (Sciex). The undetected proteins are indicated in Supplementary Table 1. The relative amounts of each detected protein were

estimated using the total intensity of peptides corresponding to each protein divided by the number of peptides expected following trypsin digestion after cleaning the raw results by confidence limit (>80%), *d*<sub>mass</sub> (>0.03) and peptide molecular weight (1,000–3,300 Da). The relative numbers obtained were then normalized by considering the stoichiometry of Afp4 to be exactly six per AFP particle. This semi-quantification yielded relatively precise results, consistent with the cryo-EM map, particularly for proteins with a large number of expected peptides, such as Afp11 (estimated at 6.1 copies per AFP particle) and Afp12 (estimated at 5.09 copies per AFP particle). Only one Afp18 toxin molecule (estimated at 0.97 copies) would be present per AFP particle according to this quantification.

**Cryo-EM sample preparation.** The cryo-EM specimens were prepared using a Vitrobot (Mark IV) device. Typically, 2–3 µl of a purified AFP sample was applied on a R2/2 Quantifoil grid (Pro Sci Tec) freshly glow-discharged in an amyl-amine environment, blotted for 3 s and vitrified using a Vitrobot (ThermoFisher) kept at 4°C and 100% humidity.

**Cryo-EM data collection.** The cryo-EM images of the extended AFP were collected at the EMBL Heidelberg on a Titan Krios, operated at 300 kV by W. Hagen, with a condenser and objective lens-aperture size of 50 µm, a spot size of 8 and a beam diameter of 1.18 µm. A total of 5,153 movies of 20 frames were acquired on a K2 camera in counting mode, with a dose rate of  $2.5 \text{ e}^{-} \text{ \AA}^{-2} \text{ s}^{-1}$ , a total exposure time of 8 s using 0.4 s frames, corresponding to  $1 \text{ e}^{-} \text{ \AA}^{-2}$  per frame and a total dose of  $20 \text{ e}^{-} \text{ \AA}^{-2}$ . The calibrated pixel size at the specimen level was 1.35 Å at a nominal magnification of  $\times 37,000$ .

The cryo-EM images of the contracted AFP were collected by J.H. at the Korea Basic Science Institute Ochang Center on a Titan Krios operated at 300 kV, equipped with a Cs corrector (measured Cs of 500 nm after correction), using condenser and objective lens-aperture sizes of 70 µm and 100 µm, respectively, a spot size of 3 and a beam diameter of 850 nm. A total of 2,847 movies of 30 frames were acquired on a Falcon II camera, with a dose rate of  $35 \text{ e}^{-} \text{ \AA}^{-2} \text{ s}^{-1}$ , a total exposure time of 1.8 s using 0.06 s frames, corresponding to approximately  $2.1 \text{ e}^{-} \text{ \AA}^{-2}$  per frame and a total dose of  $63 \text{ e}^{-} \text{ \AA}^{-2}$ . The calibrated pixel size at the specimen level was 1.397 Å at a nominal magnification of  $\times 47,000$ . A summary of the cryo-EM data collection and processing (see below) parameters can be found in Supplementary Table 2.

**Movie alignment, general processing strategy, particle picking and defocus estimation.** The frame alignment for both datasets was performed with IMOD alignframes<sup>40</sup>, keeping frames 2–20 (total dose of  $20 \text{ e}^{-} \text{ \AA}^{-2}$ ) for the extended AFP and frames 2–13 (total dose of  $27 \text{ e}^{-} \text{ \AA}^{-2}$ ) for the contracted AFP.

Extended AFP particles were divided into three independently processed parts—the baseplate, trunk and cap (Supplementary Fig. 1). To this end, the precise coordinates of the extremities (from the largest baseplate width to the apex of the cap) of 30,858 particles were manually recorded using e2helixboxer.py from EMAN2 (ref. 41), corresponding to a total particle length of 3.19 mm. An additional 16,133 baseplates were manually selected with e2helixboxer.py in order to use baseplates from particles with overlapping trunks or caps, or that extended outside the field of view. This resulted in a total number of 46,991 baseplates (Supplementary Fig. 1).

For the contracted AFP, a three-dimensional (3D) reconstruction without helical symmetry imposition was calculated from one central segment per contracted particle. The helical parameters of this initial map were determined (see below) and a final structure calculated using overlapping segments and taking the helical symmetry into account. For this purpose, the precise coordinates of the extremities of 15,189 particles were manually recorded using e2helixboxer.py, corresponding to a total particle length of 0.63 mm. The details of the optimized segment extraction are summarized in Supplementary Figs. 1 and 2. Helical segmentation for the contracted sheath was performed using SPRING<sup>42</sup>. The contrast transfer function (CTF) estimation was done with either CTFIND3/CTFTILT<sup>43</sup> or CTFIND4 (ref. 44). Phase-flipped segments were extracted for symmetry refinement, whereas convolved segments were used for high-resolution 3D reconstruction. The search in out-of-plane tilt was limited to 12° for all 3D refinements. Figures were prepared using Chimera and ChimeraX<sup>45,46</sup>.

**Extended AFP trunk.** Analogous to all known CIS structures, the trunk of the extended AFP particle can be considered as an assembly of six protofilaments forming a right-handed superhelix (Fig. 2a)<sup>1</sup>. Alternatively, one can adopt a historical description and define the inner tube as tightly stacked hexameric rings or layers that are related to each other by a 20.14° rotation and a translation of 39.32 Å along the particle axis, which is a helical symmetry similar to the other CIS<sup>1</sup>. Due to Afp2/Afp3 alternation, the helical geometry of the sheath is described by a rise of 78.64 Å and a rotation of 40.28°, exactly twice that of the enclosed tube of Afp1. To determine this symmetry, refinement was performed using the previous reconstruction of the AFP trunk with a 20 Å resolution<sup>15</sup> as an initial model, with 20 cycles of IHRSR<sup>47</sup> refinement implemented in Spider<sup>48,49</sup>, wherein the helical rise was varied from 78 to 81 Å in steps of 1.5 Å and the rotation between subunits varied from 40.5 to 42.5° in 1° steps (Supplementary Fig. 1). For each initial symmetry choice, the symmetry parameters were refined with the hsearch<sup>47</sup> program using step sizes of 0.1 Å for the helical rise and 0.1° for the azimuthal rotation for

iterations 2–13. The steps were decreased to 0.05 Å and 0.05°, respectively, for cycles 14–20 and a finer increment was used to project the reference volume. The average final values of 78.57 Å for the rise and 40.15° for the helical rotation were then used for high-resolution refinement in SPRING, which yielded a reconstruction of the trunk at an overall estimated resolution of 2.9 Å at a Fourier shell correlation (FSC) cut-off of 0.143 and with an automatically determined b-factor of  $-105 \text{ \AA}^{-2}$ . A per-particle CTF determination with Gctf<sup>50</sup> and helical auto-refine procedure with symmetry refinement as implemented in RELION-3 (ref. 51) improved the map further, resulting in the final reconstruction at an overall resolution of 2.78 Å (FSC at a 0.143 cut-off) and with an automatically determined b-factor of  $-92 \text{ \AA}^{-2}$ , obtained from 122,721 segments (after the removal of the outplane outliers) corresponding to 736,326 asymmetric units after C6 symmetrization.

**Contracted AFP sheath.** Using a hollow cylinder with an inner diameter of 70 Å and outer diameter of 230 Å as an initial model, a 3D refinement in RELION-2.1 (ref. 52), imposing only C6 symmetry, resulted in a map with a resolution of 4.2 Å after automatic masking (FSC cut-off of 0.143). This map was then used to determine the helical symmetry as follows. Based on a first visual inspection of this C6-symmetrized map, helical rises between 30 and 36 Å (0.2 Å steps) and azimuthal rotations between 2 and 5° (0.1° steps) were tested by imposing each symmetry with the himpose<sup>47</sup> program and recording the variance of each symmetrized reconstruction. A finer search around the parameters corresponding to the highest variance (that is, reinforcement of features) was then performed with the rise ranging from 34.8 to 35.2 Å in 0.02 Å steps and the rotation ranging from 3 to 3.2° in 0.01° steps (Supplementary Fig. 2). The resulting best symmetry of a helical rise of 34.97 Å and azimuthal rotation of 3.14° was then used for further processing in SPRING. High-resolution 3D refinement gave a final map at an overall resolution of 3.8 Å at a FSC cut-off of 0.143 (Supplementary Fig. 4) obtained from 46,022 segments corresponding to 276,132 asymmetric units after helical and C6 symmetrization. This map was filtered to 3.8 Å and sharpened with a b-factor of  $-260 \text{ \AA}^{-2}$  for the model building and figures.

**Contracted AFP baseplate.** The extremities with and without a contracted baseplate were sorted by two-dimensional classification (Supplementary Fig. 2). The C6-symmetrized reconstruction of the entire contracted trunk was shifted and masked to match the position of the extracted extremities, and used as an initial reference for a RELION-2.1 auto-refine procedure using C6 symmetry. The resulting map, with an average resolution of 4.6 Å (FSC at a 0.143 cut-off), showed a poorly resolved baseplate region. Masked 3D classification of this region with no further alignment enabled grouping of the particles by their baseplate conformation (more or less open). A final auto-refine of each group led to three contracted baseplate maps with an overall resolution of 5.4, 5.9 and 7.3 Å. The best map was used for refining the models of the proximal Afp2 and Afp4, and placing Afp9, Afp11 and Afp12 as rigid bodies in the lower-resolution baseplate wedge density.

**Extended AFP baseplate.** The previously published low resolution AFP reconstruction<sup>15</sup> was appropriately masked to enclose the baseplate part and used as initial reference for a RELION-2.1 auto-refine procedure using C6 symmetry. This gave a reconstruction with a resolution of 3.2 Å (FSC at a 0.143 cut-off after automatic masking; 3.6 Å without masking). Alignment and resolution-dependent weighting of the individual particle movie frames (particle polishing) were performed in RELION-2.1 to further improve the resolution, using all the frames and a running average of five frames during the movie expansion. A final refinement using these polished particles resulted in a reconstruction with an overall resolution of 3.1 Å (FSC at a 0.143 cut-off after automatic masking; 3.5 Å without masking) and with a determined overall b-factor of  $-105 \text{ \AA}^{-2}$ .

**Extended AFP needle.** The C6-symmetrized baseplate reconstruction did not allow us to resolve the needle, suggesting that this region may have a threefold symmetry as in other related CIS. However, attempts to reconstruct the baseplate with C3 symmetry imposed did not lead to an improvement. Furthermore, 3D classification of aligned particles (skipping alignment) in two classes was inefficient due to the relatively low mass of the needle region in comparison to the rest of the baseplate. The localized reconstruction approach<sup>53</sup> was therefore used (Supplementary Fig. 1). The contribution of the entire baseplate, except the needle region and close surroundings, was subtracted from the raw particles using `reliion_localized_reconstruction.py` according to the orientations corresponding to the best map of the baseplate before particle polishing. Sub-particles were newly boxed with a size of 180 × 180 pixels around the defined centre of the needle. A 3D reconstruction of the subtracted particles was generated with `reliion_reconstruct` using these orientations to obtain an initial model of the needle. Subsequently, 3D classification without alignment and using C3 symmetry resulted in a separation of the subtracted particles into two classes, each corresponding to a possible orientation of the C3-symmetric needle relative to the C6-symmetric baseplate. Based on this 3D classification, a rotation of 60° around *z* was applied to the particles corresponding to one class by modifying the orientations parameter file. A first correct C3 reconstruction of the needle was computed from the subtracted particles using these modified orientations and a solvent mask surrounding the needle region was created. These were used

as the inputs for further refinement of orientations using only local searches. This workflow led to a final C3-symmetrized map from the subtracted particles with an overall resolution of 3.7 Å (FSC at a 0.143 cut-off). Inspection of this map revealed a poorly resolved upper region of the needle, where significant overlap with the rest of the projection of the baseplate would occur in the side views, suggesting artefacts due to the signal subtraction itself (Supplementary Fig. 1). Orientations of the subtracted images refined with a C3 symmetry were therefore applied to the original non-subtracted baseplate particles and a final local orientation refinement was performed with the C3 symmetry imposed. This gave a threefold symmetric map of the entire baseplate with no artefacts in the upper region of the needle and an overall resolution of 3.3 Å (FSC at a 0.143 cut-off after automatic masking; 3.8 Å without masking), and with an automatically determined overall b-factor of  $-102 \text{ \AA}^{-2}$ .

**Extended AFP cap.** The previously published AFP reconstruction<sup>13</sup> was appropriately masked to enclose the cap region and used as initial reference for a RELION-2.1 auto-refine procedure using C6 symmetry. This gave a reconstruction with an overall resolution of 3.4 Å without masking, but with a poor definition of the cap end, and with sheath protrusions appearing at every sheath layer instead of every two layers (Supplementary Fig. 1). Using the orientations corresponding to this map, 3D classification into two classes (skipping alignment) showed that the terminal sheath ring was built by either Afp2 or Afp3. The particles from both 'Afp2-terminating' and 'Afp3-terminating' classes were further subjected to 3D classification into three classes each, still without alignment. One class of Afp2-terminating particles corresponded to a map shifted by approximately 40 Å (distance between Afp2 and Afp3 rings in the sheath) that ended with Afp3. This can be explained by the fact that the alignment in the first refinement was driven by the helical-sheath region, disregarding the cap region due to its much lower mass. Similarly, two classes from the Afp3-terminating particles were shifted by 40 Å, down and up the helical axis, and corresponded to the actual Afp2-terminating particles. The particles were newly grouped by sheath-terminating protein according to this result and used for 3D refinement with a larger translational search. Two additional rounds of 3D classification and regrouping of particles according to the terminal sheath protein were performed, resulting in the separation of the dataset into 23,797 Afp2-terminating particles (77% of all particles) and 7,061 Afp3-terminating particles (23% of all particles). Finally, a 3D refinement of each dataset with an alignment parameter search and with the final 3D classes as references led to reconstructions of the cap with an overall resolution of 3.2 and 3.4 Å (FSC at a 0.143 cut-off after automatic masking; 3.5 and 3.8 Å without masking), and a determined overall b-factor of  $-102 \text{ \AA}^{-2}$  and  $-82 \text{ \AA}^{-2}$  for the maps with the terminal Afp2–Afp16 and Afp3–Afp16 interactions, respectively. To assess the resolution of only the cap part, a tighter mask enclosing this region was created and combined with the automatic mask calculated by `relion_postprocess`; the cap resolution was estimated to be 3.3 and 3.4 Å for the terminal Afp2–Afp16 and Afp3–Afp16 maps, respectively (FSC at a 0.143 cut-off).

The local resolution was calculated using `relion_postprocess` for each map except the helical contracted sheath, for which `ResMap34` was used.

**Merging maps into one composite map of the extended AFP.** The entire AFP map shown in Fig. 1 was assembled from the following post-processed maps the extended AFP cap terminating with Afp2–Afp16, the helical trunk (SPRING reconstruction), the sixfold symmetric baseplate (needle region removed) and the threefold symmetric baseplate (needle region only). The individual maps were scaled by a factor of two after the appropriate masking to exclude overlaps and gaps, and corresponding translations and rotations were applied. The resulting composite map was subsequently downsampled to the original pixel size and filtered to 3.2 Å.

**Model building and refinement.** Atomic models were built into their respective cryo-EM maps ab initio by first manually tracing each protein sequence into the maps using `Coot35` and then further refined using `phenix.real_space_refine36` implemented in the `rsref` refinement pipeline<sup>37</sup>, using the local resolution of each map as a target. Several iterations of automatic real-space refinement followed by manual adjustments in `Coot` were performed until convergence. Symmetry-related copies and neighbouring subunits were taken into account by including all of the subunits present in each map during the final refinements. `LocScale38` was used to apply local sharpening on the baseplate map for the final steps of the refinements of Afp11 and Afp12, which had regions with densities with poorer resolution. For Afp11 and Afp12, manual building in `Coot` was followed by a single round of molecular-dynamics flexible fitting using `NAMDINATOR39` and by a final round of real-space refinement in `Phenix`. A summary of the model refinement and validation statistics can be found in Supplementary Table 3.

**Interaction energy calculations.** Calculations of the interaction energies between AFP subunits were performed using the PISA server<sup>40</sup>.  $-0.5 \text{ kcal mol}^{-1}$  per hydrogen bond and  $-0.3 \text{ kcal mol}^{-1}$  per salt bridge were then added to the free energy to give the final values (Supplementary Tables 4–6).

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The cryo-EM maps and corresponding atomic coordinates have been deposited to the EMDB and PDB with the following accession codes: full extended AFP composite map, EMD-4783; extended AFP baseplate C6 map, EMD-4782 and 6RAO; extended AFP baseplate C3 map, EMD-4800 and 6RBK; extended AFP cap ending in Afp2–Afp16 EMD-4784 and 6RAP; extended AFP cap ending in Afp3–Afp16, EMD-4801; extended AFP sheath map, EMD-4802 and 6RBN; extended AFP needle map from subtracted images, EMD-4871; contracted AFP sheath map, EMD-4803 and 6RC8; contracted AFP baseplate, EMD-4876 and 6RGL and contracted AFP sheath C6 map, EMD-4859. All other data supporting the findings of this study are available from A.D. and I.G. on request.

Received: 5 April 2019; Accepted: 2 July 2019;  
Published online: 5 August 2019

## References

- Kube, S. & Wendler, P. Structural comparison of contractile nanomachines. *AIMS Biophys.* **2**, 88–115 (2015).
- Taylor, N. M. I., van Raaij, M. J. & Leiman, P. G. Contractile injection systems of bacteriophages and related systems. *Mol. Microbiol.* **108**, 6–15 (2018).
- Patz, S. et al. Phage tail-like particles are versatile bacterial nanomachines—a mini-review. *J. Adv. Res.* **19**, 75–84 (2019).
- Leiman, P. G. et al. Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc. Natl Acad. Sci. USA* **106**, 4154–4159 (2009).
- Brackmann, M., Nazarov, S., Wang, J. & Basler, M. Using force to punch holes: mechanics of contractile nanomachines. *Trends Cell Biol.* **27**, 623–632 (2017).
- Nguyen, V. S. et al. Towards a complete structural deciphering of type VI secretion system. *Curr. Opin. Struct. Biol.* **49**, 77–84 (2018).
- Nakayama, K. et al. The R-type pyocin of *Pseudomonas aeruginosa* is related to P2 phage, and the F-type is related to lambda phage. *Mol. Microbiol.* **38**, 213–231 (2000).
- Michel-Briand, Y. & Baysse, C. The pyocins of *Pseudomonas aeruginosa*. *Biochimie* **84**, 499–510 (2002).
- Ge, P. et al. Atomic structures of a bactericidal contractile nanotube in its pre- and postcontraction states. *Nat. Struct. Mol. Biol.* **22**, 377–382 (2015).
- Hurst, M. R. H., Glare, T. R. & Jackson, T. A. Cloning *Serratia entomophila* antifeeding genes—a putative defective prophage active against the grass grub *Costelytra zealandica*. *J. Bacteriol.* **186**, 5116–5128 (2004).
- Yang, G., Dowling, A. J., Gerike, U., Ffrench-Constant, R. H. & Waterfield, N. R. *Photorhabdus* virulence cassettes confer injectable insecticidal activity against the wax moth. *J. Bacteriol.* **188**, 2254–2261 (2006).
- Sarris, P. E., Ladoukakis, E. D., Panopoulos, N. J. & Scoulica, E. V. A phage tail-derived element worldwide distribution among both prokaryotic domains: a comparative genomic and phylogenetic study. *Genome Biol. Evol.* **6**, 1739–1747 (2014).
- Heymann, J. B. et al. Three-dimensional structure of the toxin-delivery particle antifeeding prophage of *Serratia entomophila*. *J. Biol. Chem.* **288**, 25276–25284 (2013).
- Jiang, F. et al. Cryo-EM structure and assembly of an extracellular contractile injection system. *Cell* **177**, 370–383 (2019).
- Shikuma, N. J. et al. Marine tubeworm metamorphosis induced by arrays of bacterial phage tail-like structures. *Science* **343**, 529–534 (2014).
- Böck, D. et al. In situ architecture, function, and evolution of a contractile injection system. *Science* **357**, 713–717 (2017).
- Kudryashev, M. et al. Structure of the type VI secretion system contractile sheath. *Cell* **160**, 952–962 (2015).
- Wang, J. et al. Cryo-EM structure of the extended type VI secretion system sheath-tube complex. *Nat. Microbiol.* **2**, 1507–1512 (2017).
- Nazarov, S. et al. Cryo-EM reconstruction of type VI secretion system baseplate and sheath distal end. *EMBO J.* **37**, e97103 (2017).
- Taylor, N. M. I. et al. Structure of the T4 baseplate and its function in triggering sheath contraction. *Nature* **533**, 346–352 (2016).
- Hurst, M. R., Glare, T. R., Jackson, T. A. & Ronson, C. W. Plasmid-located pathogenicity determinants of *Serratia entomophila*, the causal agent of amber disease of grass grub, show similarity to the insecticidal toxins of *Photorhabdus luminescens*. *J. Bacteriol.* **182**, 5127–5138 (2000).
- Rybakova, D. et al. Role of antifeeding prophage (Afp) protein Afp16 in terminating the length of the Afp tailocin and stabilizing its sheath. *Mol. Microbiol.* **89**, 702–714 (2013).
- Rybakova, D., Schramm, P., Mitra, A. K. & Hurst, M. R. H. Afp14 is involved in regulating the length of Anti-feeding prophage (Afp). *Mol. Microbiol.* **96**, 815–826 (2015).
- Hurst, M. R. H. et al. *Serratia proteamaculans* strain AGR96X encodes an antifeeding prophage (tailocin) with activity against grass grub (*Costelytra giveni*) and Manuka beetle (*Pyronota* species) larvae. *Appl. Environ. Microbiol.* **84**, e02739-17 (2018).

25. Pell, L. G., Kanelis, V., Donaldson, L. W., Howell, P. L. & Davidson, A. R. The phage lambda major tail protein structure reveals a common evolution for long-tailed phages and the type VI bacterial secretion system. *Proc. Natl Acad. Sci. USA* **106**, 4160–4165 (2009).
26. Arnaud, C.-A. et al. Bacteriophage T5 tail tube structure suggests a trigger mechanism for *Siphoviridae* DNA ejection. *Nat. Commun.* **8**, 1953 (2017).
27. Biggs, D. R. & McGregor, P. G. Gut pH and amylase and protease activity in larvae of the New Zealand grass grub (*Costelytra zealandica*; Coleoptera: Scarabaeidae) as a basis for selection inhibitors. *Insect Biochem. Mol. Biol.* **26**, 69–75 (1996).
28. Bateman, A. & Bycroft, M. The structure of a LysM domain from *E. coli* membrane-bound lytic murein transglycosylase D (MltD). *J. Mol. Biol.* **299**, 1113–1119 (2000).
29. Zheng, W. et al. Refined cryo-EM structure of the T4 tail tube: exploring the lowest dose limit. *Structure* **25**, 1436–1441 (2017).
30. Bönemann, G., Pietrosiuk, A., Diemand, A., Zentgraf, H. & Mogk, A. Remodelling of VipA/VipB tubules by ClpV-mediated threading is crucial for type VI protein secretion. *EMBO J.* **28**, 315–325 (2009).
31. Zoued, A. et al. Priming and polymerization of a bacterial contractile tail structure. *Nature* **531**, 59–63 (2016).
32. Zoued, A. et al. TssA: the cap protein of the type VI secretion system tail. *Bioessays* **39**, 1600262 (2017).
33. Fokine, A. et al. The molecular architecture of the bacteriophage T4 neck. *J. Mol. Biol.* **425**, 1731–1744 (2013).
34. Pell, L. G. et al. The X-ray crystal structure of the phage lambda tail terminator protein reveals the biologically relevant hexameric ring structure and demonstrates a conserved mechanism of tail termination among diverse long-tailed phages. *J. Mol. Biol.* **389**, 938–951 (2009).
35. Cherrak, Y. et al. Biogenesis and structure of a type VI secretion baseplate. *Nat. Microbiol.* **3**, 1404–1416 (2018).
36. Park, Y.-J. et al. Structure of the type VI secretion system TssK–TssF–TssG baseplate subcomplex revealed by cryo-electron microscopy. *Nat. Commun.* **9**, 5385 (2018).
37. Schneider, M. M. et al. PAAR-repeat proteins sharpen and diversify the type VI secretion system spike. *Nature* **500**, 350–353 (2013).
38. Kanamaru, S., Ishiwata, Y., Suzuki, T., Rossmann, M. G. & Arisaka, F. Control of bacteriophage T4 tail lysozyme activity during the infection process. *J. Mol. Biol.* **346**, 1013–1020 (2005).
39. Quentin, D. et al. Mechanism of loading and translocation of type VI secretion system effector Tse6. *Nat. Microbiol.* **3**, 1142–1152 (2018).
40. Kremer, J. R., Mastrorade, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).
41. Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
42. Desfosses, A., Ciuffa, R., Gutsche, I. & Sachse, C. SPRING - an image processing package for single-particle based helical reconstruction from electron cryomicrographs. *J. Struct. Biol.* **185**, 15–26 (2014).
43. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
44. Rohou, A. & Grigorieff, N. CTFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
45. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
46. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
47. Egelman, E. H. A robust algorithm for the reconstruction of helical filaments using single-particle methods. *Ultramicroscopy* **85**, 225–234 (2000).
48. Frank, J. et al. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).
49. Shaikh, T. R. et al. SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs. *Nat. Protoc.* **3**, 1941–1974 (2008).
50. Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
51. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
52. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
53. Ilca, S. L. et al. Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes. *Nat. Commun.* **6**, 8843 (2015).
54. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
55. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
56. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
57. Hoffmann, N. A., Jakobi, A. J., Vorländer, M. K., Sachse, C. & Müller, C. W. Transcribing RNA polymerase III observed by electron cryomicroscopy. *FEBS J.* **283**, 2811–2819 (2016).
58. Jakobi, A. J., Wilmanns, M. & Sachse, C. Model-based local density sharpening of cryo-EM maps. *eLife* **6**, e27131 (2017).
59. Kidmose, R. T., Nissen, P., Boesen, T., Karlsen, J. L. & Pedersen, B. P. Namdinator—automatic molecular dynamics flexible fitting of structural models into cryo-EM and crystallography experimental maps. *IUCr* **6**, 526–531 (2019).
60. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).

## Acknowledgements

This work was supported by a RSNZ Marsden grant to A.K.M. and M.R.H.H. and by the Bio-protection Research Centre. The I.G. lab was funded by a European Union's Horizon 2020 research and innovation programme under grant agreement no. 647784. A.D. was further supported by the Fondation Recherche Médicale (grant no. ARF20160936266) and Labex GRAL (grant no. ANR-10-LABX-49-01). J.F. was supported by a long-term EMBO fellowship (grant no. ALTF441-2017). M.J. was supported by a doctoral grant from the Commissariat à l'Énergie Atomique et aux Énergies Alternatives (CEA). The authors wish to acknowledge the New Zealand eScience Infrastructure (NeSI) high performance computing facilities—particularly B. Roberts and P. Maxwell who set up the cryo-EM software on the cluster and provided precious support—for the image processing. The national facilities of New Zealand are provided by the NeSI and are jointly funded by collaborator institutions of the NeSI and through the Research Infrastructure programme of the Ministry of Business, Innovation and Employment (<https://www.nesi.org.nz>). We also acknowledge A. Peuch and G. Schoehn for support and access to the joint IBS/EMBL EM computing cluster, which was used as part of the platforms of the Grenoble Instruct-ERIC center (ISBG; UMS 3518 CNRS-CEA-UGA-EMBL) within the Grenoble Partnership for Structural Biology. Platform access was supported by FRISBI (ANR-10-INBS-05-02) and GRAL, a project of the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003). IBS acknowledges integration into the Interdisciplinary Research Institute of Grenoble (IRIG, CEA). We thank A. Jakobi for advice and help with the atomic model refinements and local amplitude scaling of the maps, and for providing early versions of LocScale. We thank A. Turner and the Imaging Centre of the University of Auckland for support and help with the use of the electron microscopes. We thank C. Sachse for help and advice in image processing and in setting SPRING on the NeSI and the IBS/EMBL clusters. We thank M. Middleditch for the mass-spectrometry experiments and help with the data analysis. The EMBL Cryo-Electron Microscopy Service Platform is acknowledged for support in image acquisition and analysis. We are grateful to W. Hagen for the acquisition of high-quality cryo-EM images of extended AFP in 2016. Data acquisition has been supported by iNEXT (grant no. 653706) funded by the EU Horizon 2020 programme. We acknowledge P. Leiman for critical comments of the manuscript.

## Author contributions

A.K.M. and M.R.H.H. designed and funded the project. A.D., H.V. and M.J. purified and initially characterized samples by negative staining and cryo-EM. A.D. analysed the mass-spectrometry data. A.D. performed the image analyses with significant input from H.V. A.D. performed the atomic model building with significant input from H.V. and J.E. J.H. and H.J. collected the high-resolution cryo-EM datasets of the contracted AFP. A.D., H.V., T.J., J.F., M.J., J.B.H., M.R.H.H., I.G. and A.K.M. analysed the data. A.D., J.F., M.J. and I.G. prepared the figures and tables. A.K.M., M.R.H.H. and I.G. provided support and supervised the project at various stages. I.G. and A.K.M. wrote the manuscript with significant input from M.R.H.H., J.B.H., A.D., J.F. and M.J., and contributions from all of the authors.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41564-019-0530-6>.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Correspondence and requests for materials** should be addressed to M.R.H.H., I.G. or A.K.M.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

## **SUPPLEMENTARY INFORMATION**

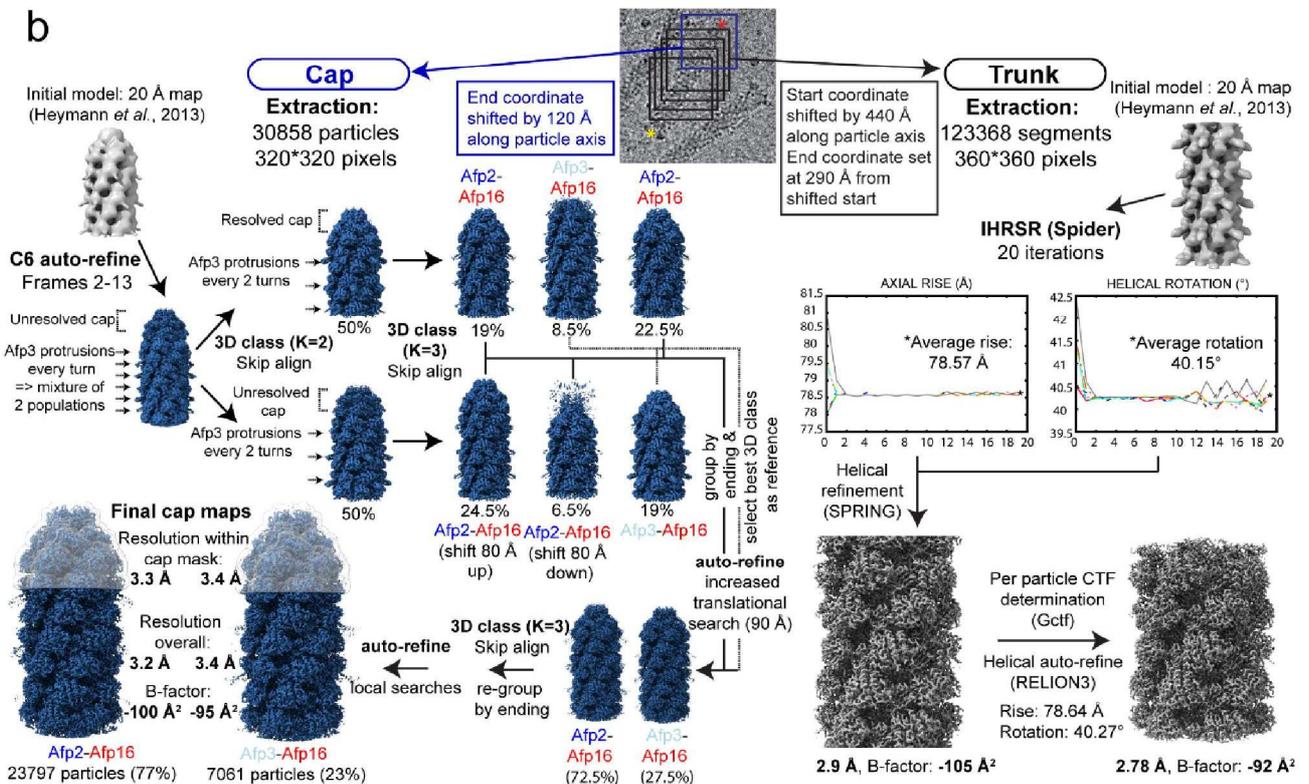
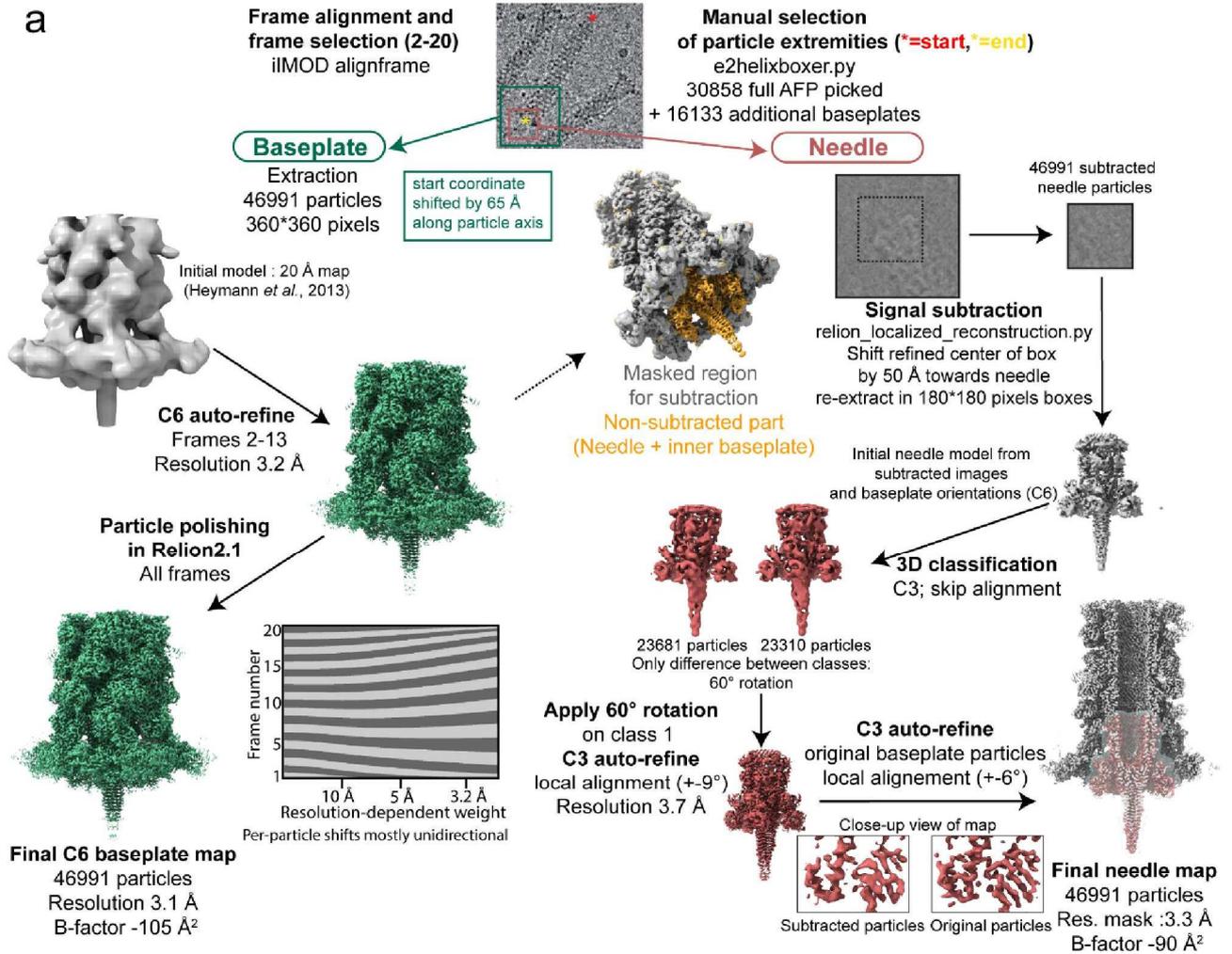
### **Atomic structures of an entire contractile injection system in both the extended and contracted states**

Ambroise Desfosses, Hariprasad Venugopal, Tapan Joshi, Jan Felix, Matthew Jessop,  
Hyengseop Jeong, Jaekyung Hyun, J. Bernard Heymann, Mark R. H. Hurst \*,  
Irina Gutsche \*, Alok K. Mitra \*

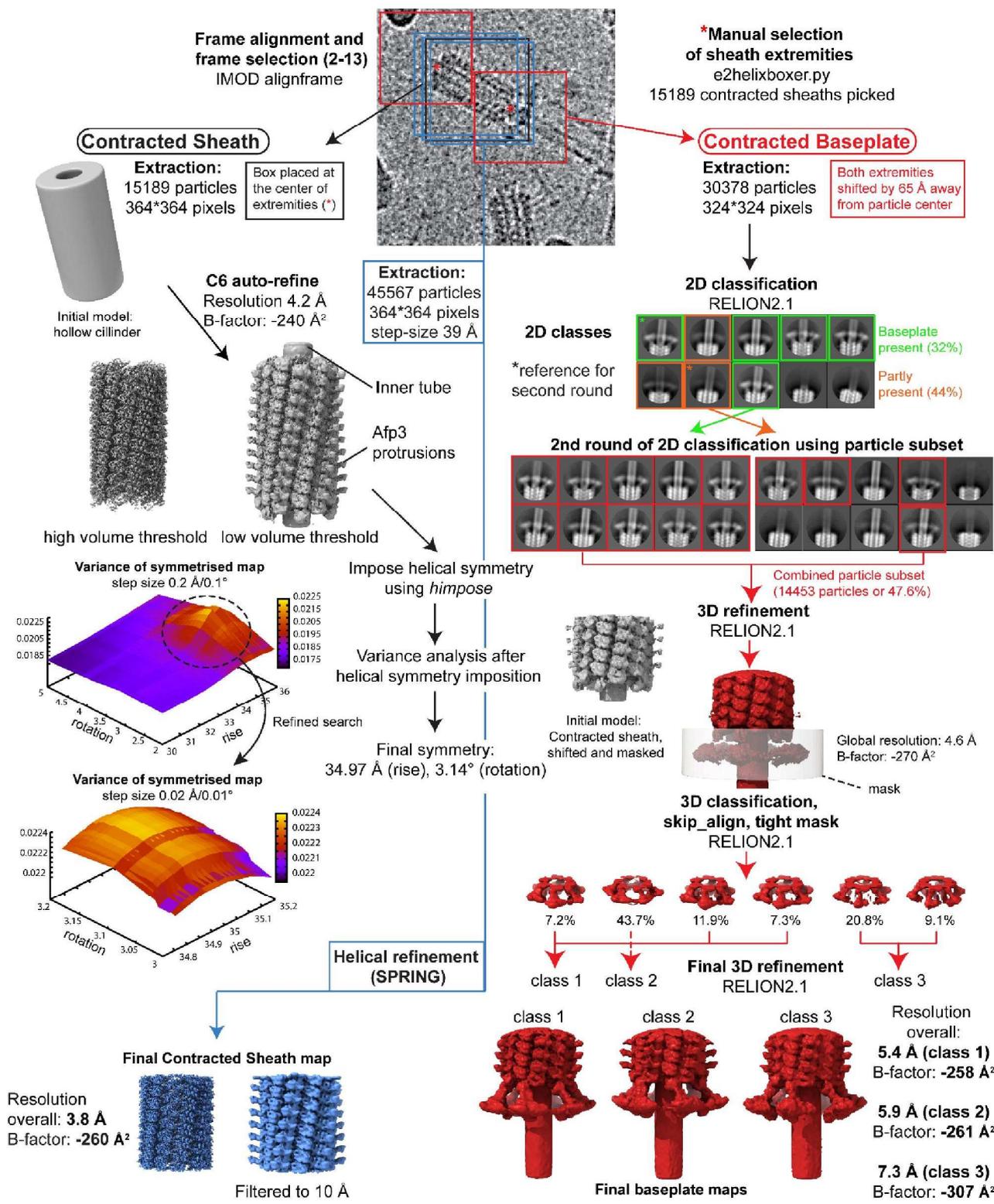
\*: corresponding author

**Supplementary Figures 1 to 9**

**Supplementary Tables 1 to 6**

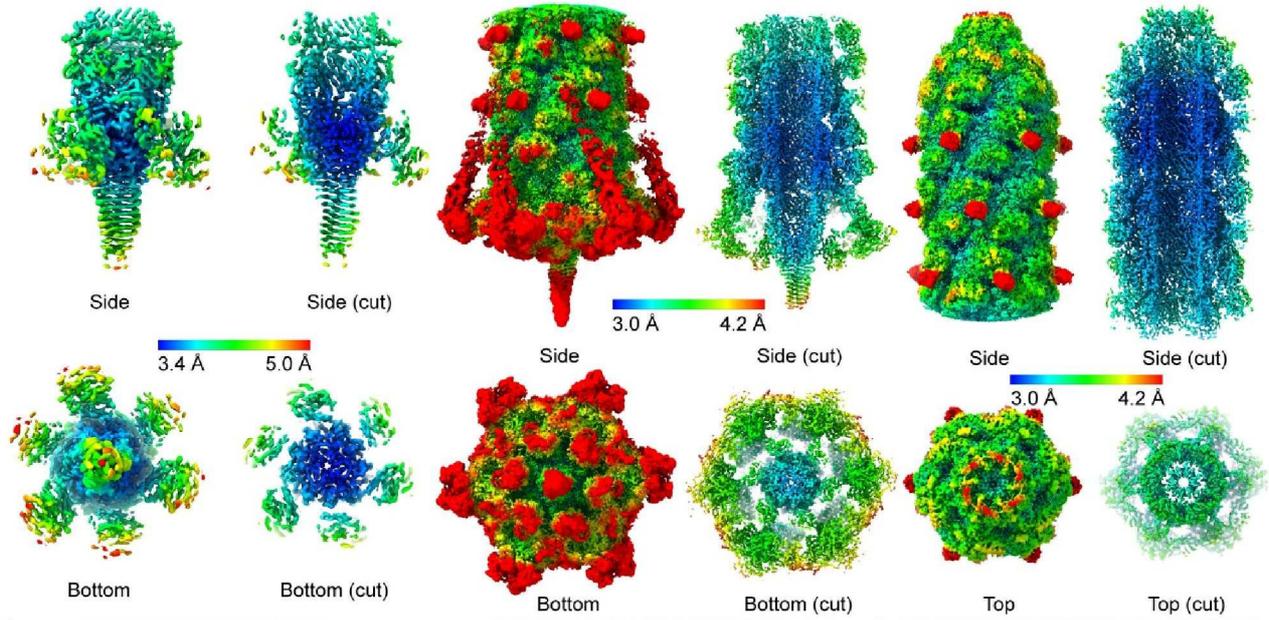


**Supplementary Figure 1: Flowcharts of the image processing pipeline for the AFP baseplate and needle (a), and apical cap and trunk (b) in the extended state.** In the flowcharts, a representative micrograph is shown with boxed picked particles for the Baseplate (a, green), Needle (a, red), Cap (b, blue) and Trunk (b, black), followed by subsequent particle extraction, masking, signal subtraction, classification and refinement procedures performed in this study. At the bottom of each flowchart, final maps are shown with annotated overall resolution and b-factor. Used software packages are indicated where needed. For a full description of the image processing pipeline, see 'Methods'.

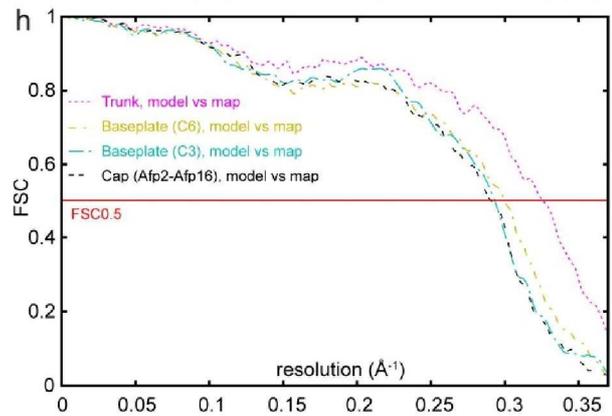
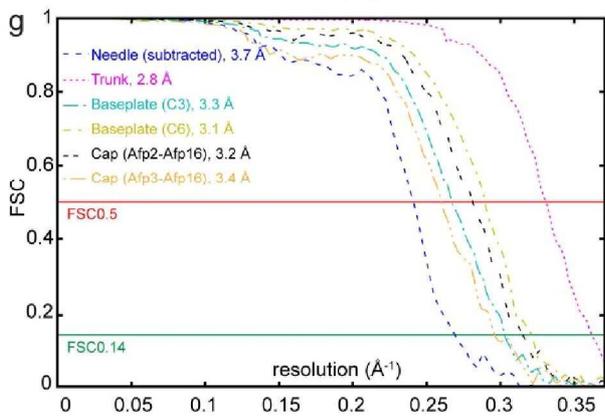
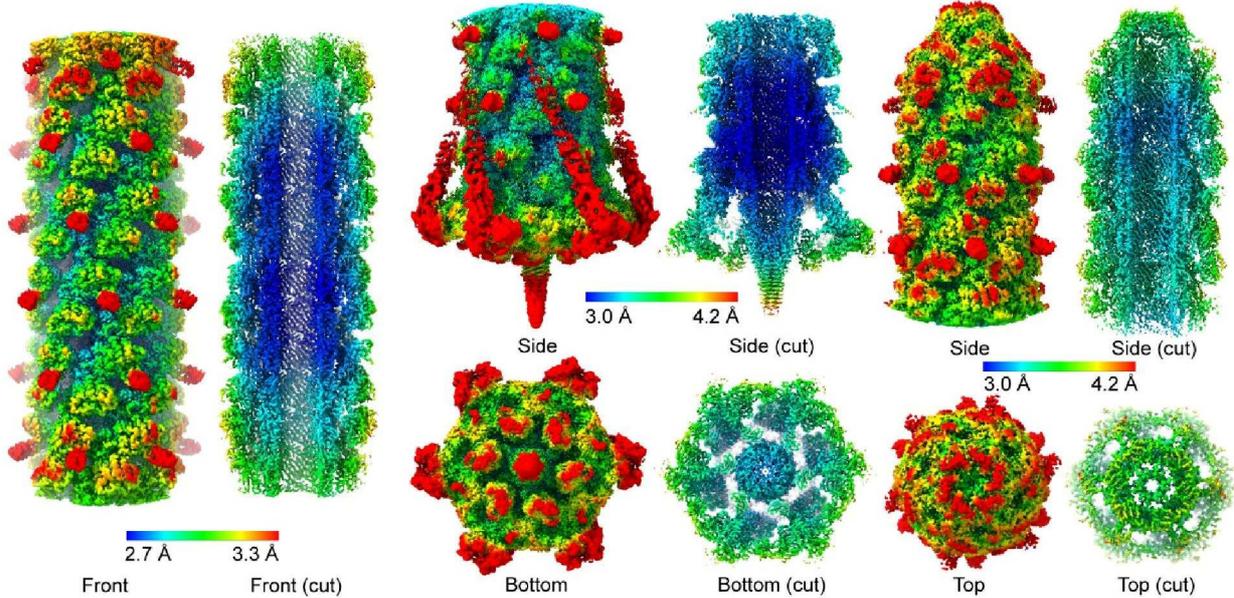


**Supplementary Figure 2: Flowchart of the image processing pipeline for the AFP sheath and baseplate in the contracted state.** A representative micrograph is shown with boxed picked particles (black: Contracted Sheath start, blue: Contracted Sheath final, red: Contracted Baseplate), followed by subsequent particle extraction, masking, classification and refinement procedures performed in this study. At the bottom of the flowchart, final maps are shown with annotated overall resolution and b-factor. Used software packages are indicated where needed. For a full description of the image processing pipeline, see 'Methods'.

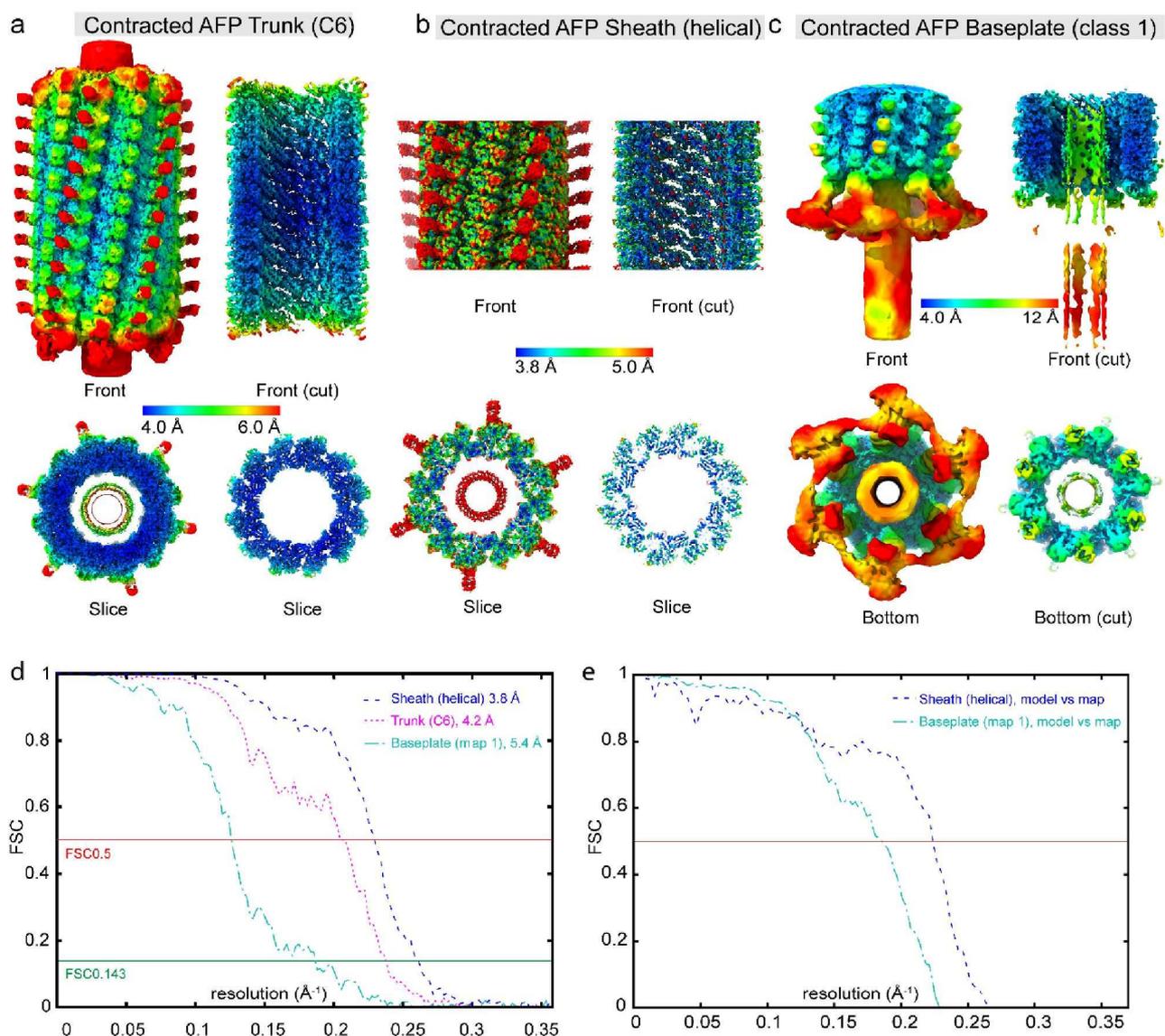
**a** Extended AFP Needle (subtracted)    **b** Extended AFP Baseplate (C3)    **c** Extended AFP Cap (Afp2-Afp16)



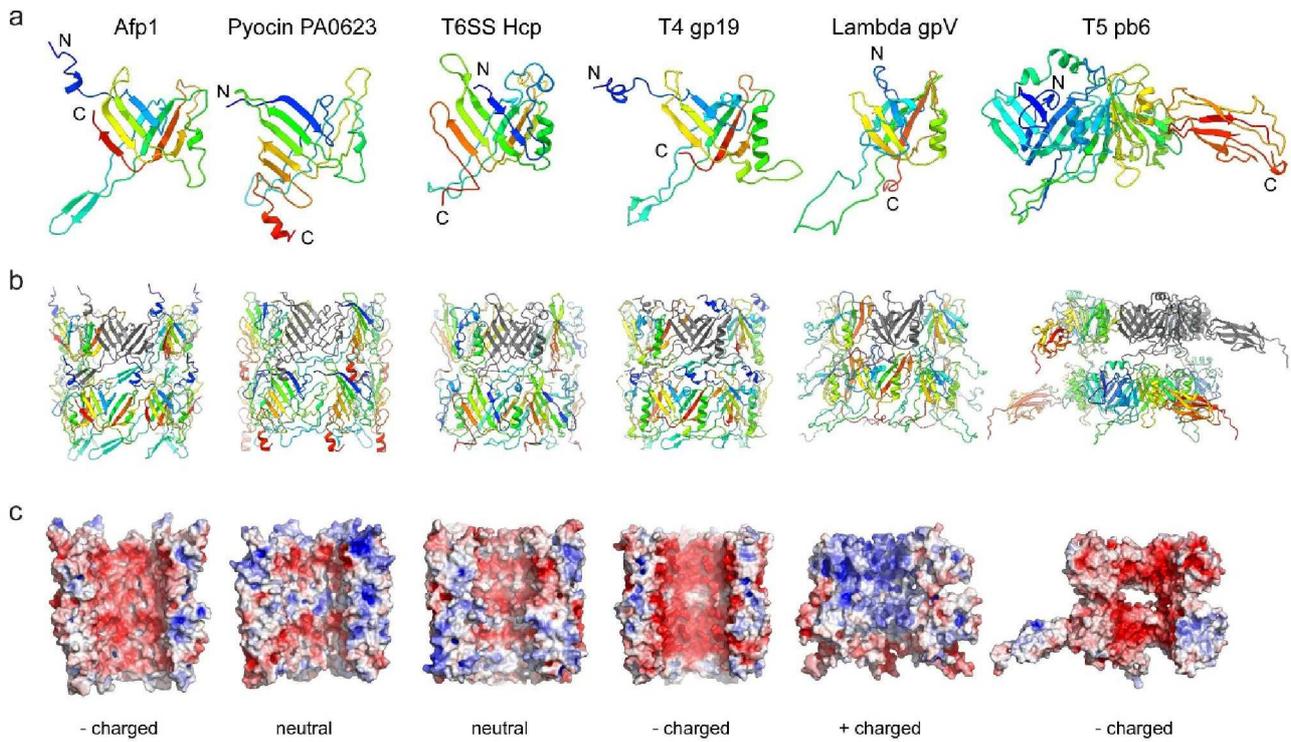
**d** Extended AFP Trunk    **e** Extended AFP Baseplate (C6)    **f** Extended AFP Cap (Afp3-Afp16)



**Supplementary Figure 3: Gallery of unmasked cryo-EM maps, filtered and coloured by local resolution, and accompanying FSC curves for the extended AFP apical cap, trunk, baseplate and needle.** Unmasked cryo-EM maps for the extended AFP needle (a), baseplate processed with C3 symmetry (b), cap terminating with Afp2-Afp16 (c), trunk (d), baseplate processed with C6 symmetry (e), and cap terminating with Afp3-Afp16 (f) filtered and coloured by local resolution. For panels b, c, d, e and f, contour levels ( $\sigma$ -values) used in Chimera to generate full and cut-out views are: full = 0.015 and cut-out = 0.07 (b), full = 0.02 and cut-out = 0.09 (c), full = 0.025 and cut-out = 0.08 (d), full = 0.015 and cut-out = 0.07 (e), full = 0.02 and cut-out = 0.09 (f). (g) Accompanying FSC curves for the maps shown in (a-f). In the legend, the resolution at the FSC cut-off of 0.5 (red) and 0.143 (green) is indicated. (f) Accompanying model vs map FCS curves for the extended AFP trunk, baseplate (C6), baseplate (C3) and cap. The resolution at the FSC cut-off of 0.5 is indicated in red.



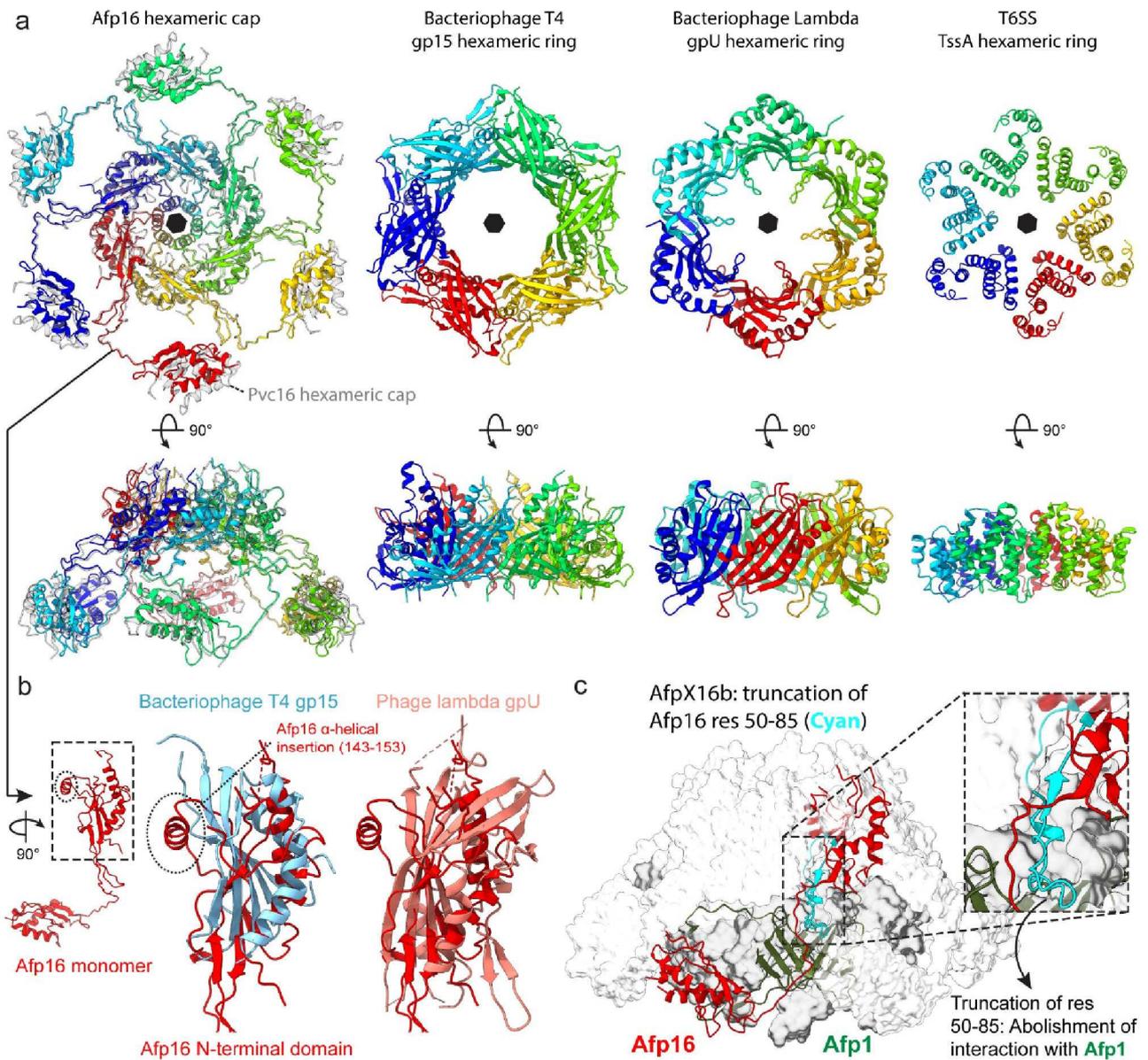
**Supplementary Figure 4: Gallery of unmasked cryo-EM maps, filtered and coloured by local resolution, and accompanying FSC curves for the contracted AFP apical sheath and baseplate.** Unmasked cryo-EM maps for the entire contracted AFP trunk processed with C6 symmetry (a), sheath processed with helical symmetry (b), and baseplate map 1 (c) filtered and coloured by local resolution. For panels a, b and c, the contour levels ( $\sigma$ -values) used in Chimera to generate full and cut-out views are: full = 0.035 and cut-out = 0.1 (a), full = 0.9 and cut-out = 2.5 (b), full = 0.03 and cut-out = 0.065 (c). (d) Accompanying FSC curves for the maps shown in a-c. In the legend, the resolution at the FSC cut-off of 0.5 (red) and 0.143 (green) is indicated. (e) Accompanying model vs map FSC curves for the contracted AFP helical sheath and baseplate. The resolution at the FSC cut-off of 0.5 is indicated in red.



**Supplementary Figure 5: Comparison of tube proteins for AFP (Afp1), Pyocin (PA0623), T6SS (Hcp), phage T4 (gp19), phage Lambda (gpV), and phage T5 (pb6).** (a) Comparison of extracted monomers of Pyocin PA0623 (PDB ID: 5W5E) T6SS Hcp (PDB ID: 5OjQ), T4 gp19 (PDB ID: 5Iv5), phage Lambda gpV (PDB ID: 2K4Q) and phage T5 pb6 (PDB ID: 5NGJ). Structures are shown as cartoons and coloured using a rainbow gradient with annotated N- and C-termini. (b) Double hexameric ring assemblies, corresponding to a 2-layer tube, formed by the tube proteins shown in (a). One monomer per tube is coloured dark grey. (c) Cut-out view of the hexameric ring assemblies shown in (b), coloured according to an electrostatic potential map calculated using APBS (Adaptive Poisson-Boltzmann Solver), where red and blue correspond to negative and positive charges respectively.

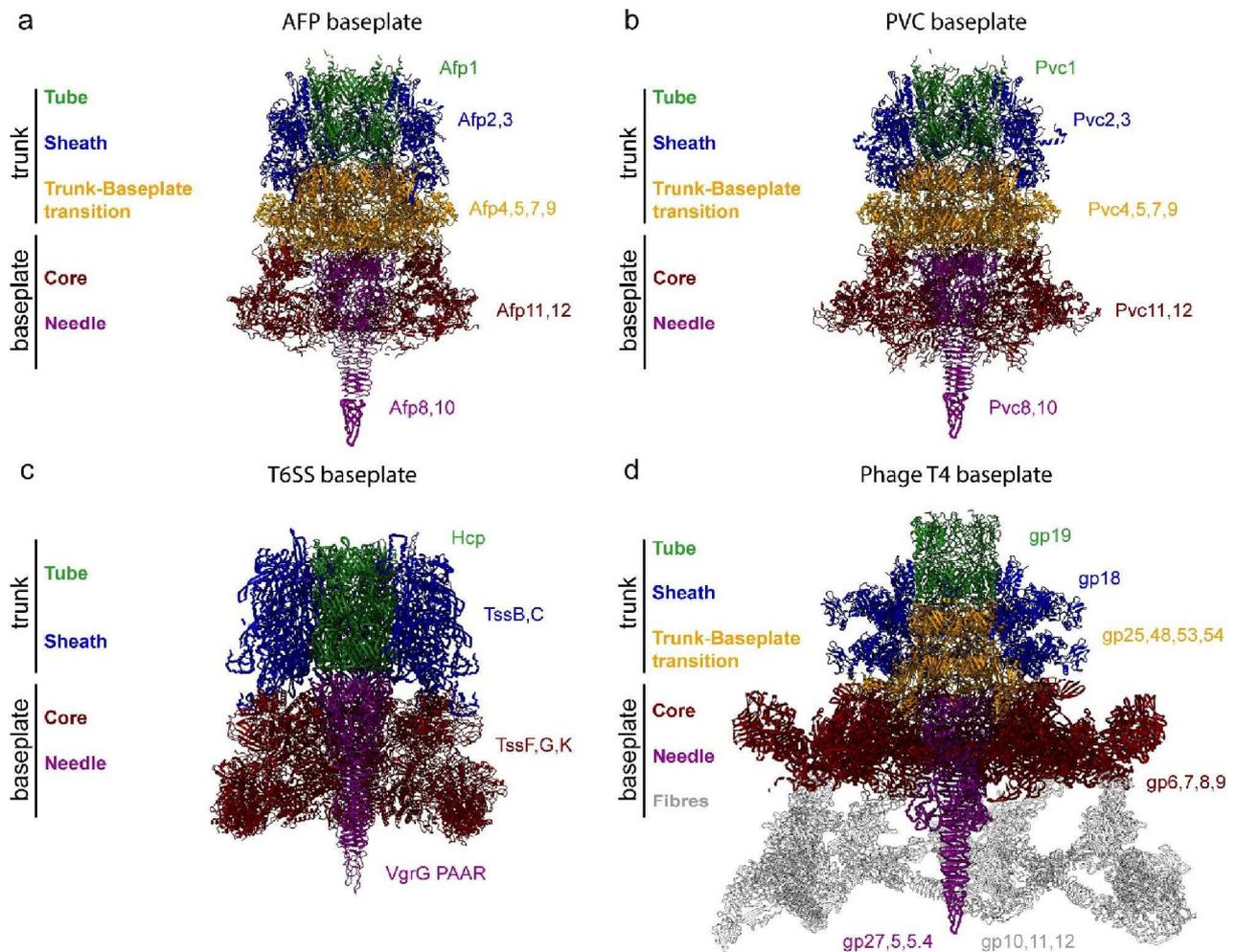


state) addition of guanidinium chloride (GuHCL) is shown. Note the slightly disappearing density for residues Arginine 282 (R282), Lysine 284 (K284), Valine 287 (V287) and Leucine 294 (L294) after addition of GuHCL. (d) Interaction between two copies of Afp1 (light and dark green), Afp2 (dark blue) and Afp3 (light blue) in the AFP extended state. Inset 1 and 2 show a zoom of the interactions between Afp1 (light green) and Afp3 (light blue) and Afp1' (dark green) and Afp3 (dark blue) respectively. Structures are shown as cartoons while selected interacting residues are shown as sticks. H-bonds and electrostatic interactions are depicted using dotted black lines. (e) Sequence alignment between the  $\alpha$ -helices spanning residues 263-322 of Afp2 and residues 360-419 of Afp3. Residues of Afp2/Afp3 interacting with Afp1 are annotated using black boxes. (f) Comparison between the sheath proteins from AFP (Afp2 and 3), PVC (Pvc2, PDB ID: 6J0B), Pyocin (PA0622, PDB ID: 3J9Q), T6SS (VipA and B, PDB ID: 3J9G) and phage T4 (gp18, PDB ID: 3J2M). Structures are shown as cartoons and coloured using a rainbow gradient, except for Pvc2 which is coloured grey, with annotated N- and C-termini. For Afp3, density is shown for the protrusion made up by two separate loops containing residues 65-105 and 215-265 (not modelled) in which the corresponding protrusion of Pvc3 is fitted (PDB ID: 6J0N). Conserved residues between Afp3 and Pvc3 in the protrusion domain are shown in red. (g) Sequence alignment between the protrusion domain of Afp3 and Pvc3. Conserved residues between Afp3 and Pvc3 are labelled in black, while conserved residues between Afp3 and Pvc3 which are present in the Pvc3 structure are labelled red. The split Afp3/Pvc3 protrusion domain is indicated by two black boxes.

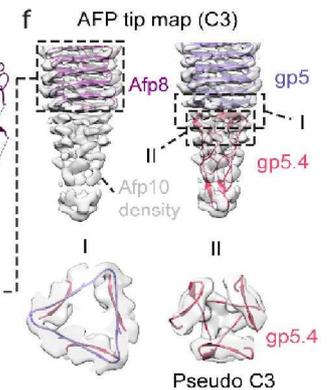
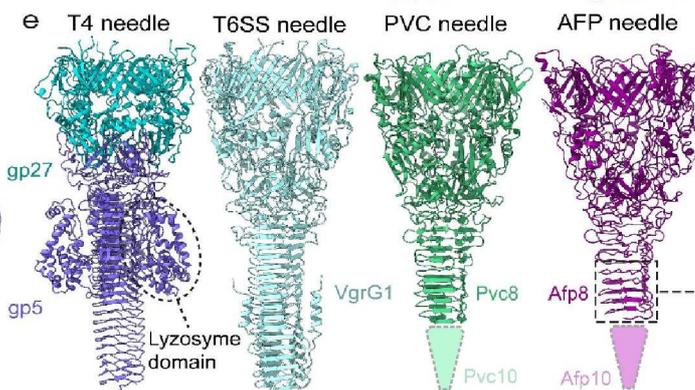
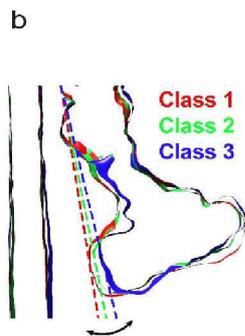
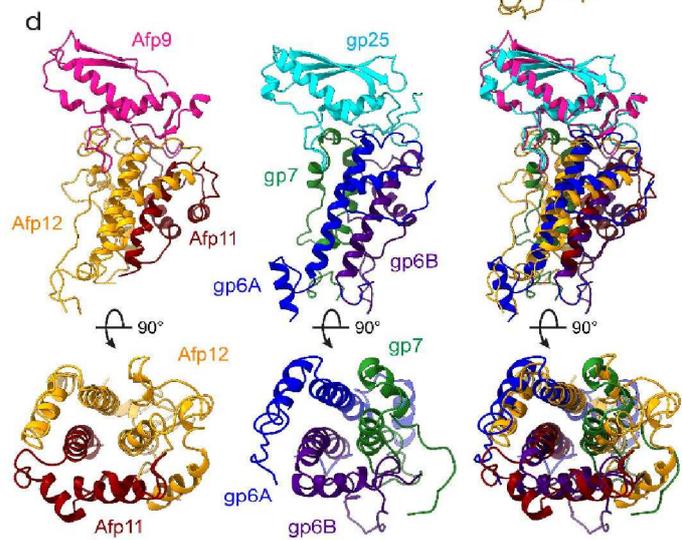
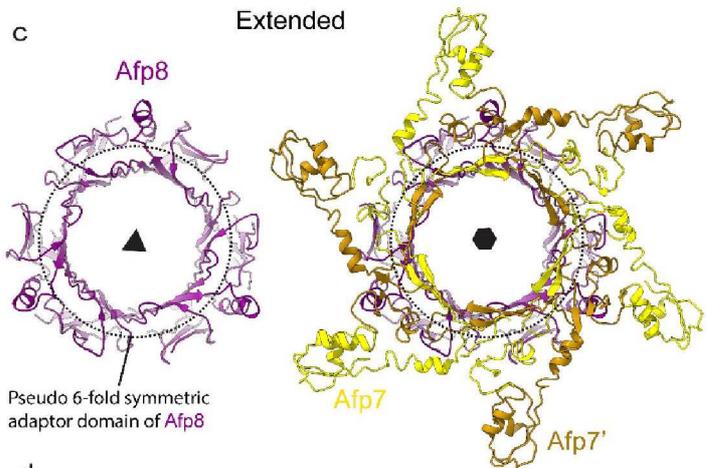
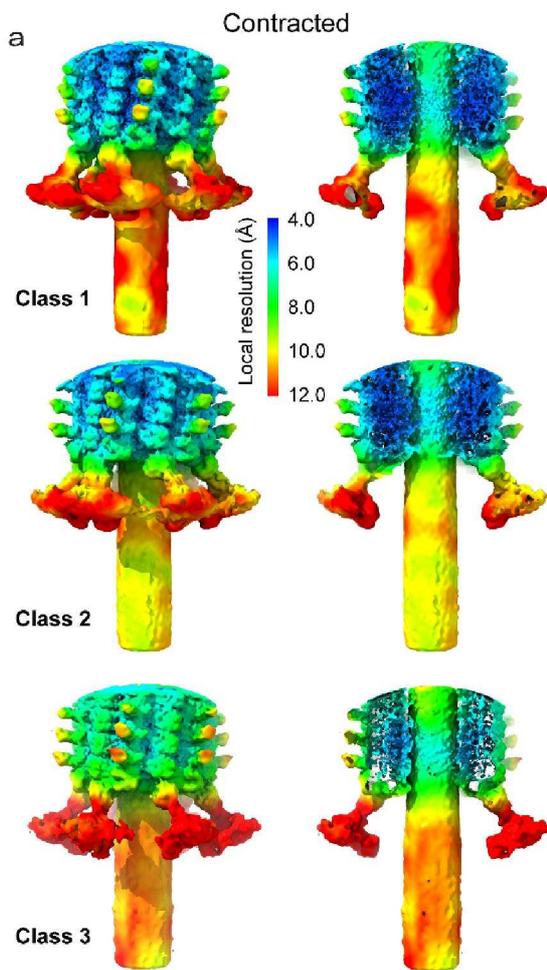


**Supplementary Figure 7: Structural comparison between the AFP apical cap protein Afp16, the PVC apical cap protein Pvc16, and the tail terminator proteins gp15 from phage T4, gpU from phage Lambda and TssA from T6SS.** (a) Top- (upper panels) and side-views (lower panels) of the hexameric cap formed by Afp16 aligned with the hexameric cap formed by Pvc16 (PDB ID: 6J0F), and hexameric rings formed by gp15 (PDB ID: 4HUH), gpU (PDB ID: 3FZB) and TssA (PDB ID: 4YO5). 6-fold rotation axes are indicated using black hexagons. (b) Structural alignment between the Afp16 N-terminal domain (red) and extracted monomers of phage T4 gp15 (light blue) and phage Lambda gpU (salmon). The  $\alpha$ -helical insertion of Afp16 spanning residues 143-153 is annotated using a dotted black oval. (c) Position of Afp1 and Afp16 in the AFP extended state apical cap. One copy of Afp1 and Afp16 are shown as cartoons and coloured green and red respectively, while

remaining Afp1 and Afp16 copies are shown as a transparent white surface. Afp16 residues 50-85 (coloured light cyan) are absent in AfpX16b, the cap protein in the AFP variant AFP-X in *S. proeamaculans*, thus leading to an abolished Afp1 interaction (see inset). This diminishes the capping capability of AfpX16b, resulting in greater variability in the lengths of AFP-X.



**Supplementary Figure 8: Comparison between the overall architecture of AFP, PVC, T6SS and phage T4 baseplate structures.** Cut-out front views of AFP (a), PVC (b), T6SS (c) and phage T4 (d) baseplate structures. For each structure, the first layers of the trunk are displayed. The individual proteins are coloured based on their position in the trunk (Tube: green, Sheath: blue, Trunk-Baseplate transition: orange) or baseplate (Core: red, Needle: purple and Fibres: transparent grey). Baseplate structures were generated as follows: For AFP and PVC, deposited structures of the respective baseplate (AFP: 6RAO, PVC: 6J0N) and needle (AFP: 6RBK, PVC: 6J0M) were combined with gp5.4 from phage T4, fitted in the AFP and PVC baseplate maps (AFP: EMD-4800, PVC: EMD-9764). For T4, the structure of the baseplate (PDB ID: 5IV5) and the structure of the sheath/tube (PDB ID: 3J2M), were fitted together in the cryo-EM reconstruction of the extended T4 tail (EMD-1126) to generate the full assembly. For T6SS, each subcomplex structure (Tube and sheath: PDB ID 5OJQ, Needle: PDB ID 4MTK and 4JIV, Core: PDB ID 6N38) was fitted into the cryo-EM reconstruction of the *Vibrio Cholerae* T6SS baseplate/needle (EMD-3879).



**g**

Afp10	1	MSNLVVIDGDALTFNPQFGAITITPPQPRISGSGEASIGGKKICIVGDEKQVSFTVDYIKPPFVAS----	PG	69
Pvc10	1	MSEAIVVDGDLVLOFDNFGNROVTVPSPGKISGTGHAOVSGKKVCILGDEKQVRVSATYITTT-HTT----	PG	68
gp5.4	1	MSG-----LSYDKCVTAGHEAWPTVVNATQSKVFTGGIAVLVAGDPITE---HTEIKKPYETHGGVTPR		63
	**	* : : * : : * : : * : : * : : * : : *		
Afp10	70	KGTLTIKALASEQRAEFATAPAPMIVVGSQFTTQFQPTTPAQDPQGKPD-D-LSAVTGVGTFINQSQSFVTAG		140
Pvc10	69	TGTTISALDAGQALQCTSGAALI IKGQFTAMFTPELPAMNNTVTPPQPDVITPSSGKGRFITQQNFATVN		141
gp5.4	64	TSKVYVT---GKKA VQMA DP---ISCGD-----TVAQASSKVFIK-----		97
	...	: : * : * * : : * : : * : : * : : *		

Residues conserved between Afp10 and Pvc10

Residues conserved between Afp10, Pvc10 and gp5.4

**Supplementary Figure 9: Structural description of the AFP baseplate and needle.** (a) Cryo-EM maps of three classes of the contracted AFP baseplate after 3D classification, coloured according to local resolution. Overall resolution estimates are 5.4 Å, 5.9 Å and 7.3 Å for Class 1, 2 and 3 respectively (see also Extended Figure 3). (b) Cut-out views of maps for Class 1, 2 and 3 of the contracted AFP baseplate illustrate the variability of the position of the wedge formed by Afp9, Afp11 and Afp12 (c) Depiction of the switch between pseudo 6-fold (3-fold) symmetry found in the adaptor domain of Afp8 (purple) in the extended AFP needle, and the 6-fold symmetry of the preceding Afp7 ring, with alternating Afp7 monomers coloured yellow/orange. 3-fold and 6-fold rotation axes are annotated with black triangles and hexagons respectively. (d) Structural comparison (left and middle) and alignment (right) between the complex formed by Afp9 (violet), Afp11 (dark red) and Afp12 (orange), extracted from the extended AFP baseplate, and the complex formed by phage T4 proteins gp6A (dark blue), gp6B (purple), gp7 (green) and gp25 (light blue) extracted from the phage T4 pre-attachment baseplate-tail tube complex (PDB ID: 5IV5). (e) Structural comparison between the needle from phage T4 (formed by gp5, coloured dark blue, and gp27, coloured cyan), T6SS (formed by VgrG1, coloured light blue), PVC (formed by Pvc8 and Pvc10) and AFP (formed by Afp8 and Afp10, coloured purple). The lysozyme domain in gp5 is annotated by a black dotted oval. T4, T6SS and PVC needle structures are extracted from PDB ID 1WTH, 6H3L and 6J0M respectively. (f) Zoom showing the cryo-EM map for the tip end of AFP with fitted Afp8 (left) or gp5 and gp5.4 (right) from phage T4. Inset I: gp5 is fitted in the upper part of the AFP tip map, which displays C3 symmetry. Inset II: The lower part of the AFP tip map displays pseudo-C3 symmetry and contains fitted gp5.4 (PDB ID: 4KU0). (g) Sequence alignment between Afp10, Pvc10 and gp5.4. Residues conserved between Afp10, Pvc10 and gp5.4 are labelled red, while residues conserved between Afp10 and Pvc10 are labelled black.

## Supplementary Tables

Supplementary Table 1. AFP proteins and their homologues

Afp Protein	MW (no. of amino acids)	Number of copies in extended AFP <sup>1</sup>	Location, function	Phage T4 homologue	T6SS homologue	R-type pyocin homologue	PVC homologue
Afp1	16.4 kDa (149)	126	Inner tube	gp19	Hcp	PA0623	Pvc1
Afp2	38.8 kDa (354)	66	Sheath	gp18	TssB/C, VipA/B	PA0622	Pvc2
Afp3	48.7 kDa (451)	60	Sheath	gp18	TssB/C, VipA/B	PA0622	Pvc3
Afp4	45.5 kDa (417)	6	Sheath	gp18	TssB/C, VipA/B	PA0622	Pvc4
Afp5	17.0 kDa (149)	6	Second layer of inner tube, tube initiator	gp54	Hcp	PA0623	Pvc5
Afp6	6.7 kDa (55)	Unknown <sup>2</sup>	Unknown	-	-	-	Pvc6
Afp7	25.2 kDa (229)	6	First layer of inner tube, tube initiator	gp48/gp53 LysM	Hcp	PA0623/PA0627 LysM	Pvc7
Afp8	58.0 kDa (529)	3	Baseplate, needle	gp5/gp27	VgrG	PA0616/PA0628	Pvc8
Afp9	15.7 kDa (140)	6	Baseplate, sheath initiator	gp25	TssE	PA0617	Pvc9
Afp10	14.5 kDa (140)	1	Baseplate, needle tip	gp5.4	VgrG PAAR	PA0616	Pvc10
Afp11	67.3 kDa (607)	6	Baseplate	gp6	TssF	PA0618	Pvc11
Afp12	106.8 kDa (963)	6	Baseplate	gp6/gp7	TssF/G	PA0619	Pvc12
Afp13	60.7 kDa (636)	6	Tail fibre protein	gp12	-	PA0620	Pvc13
Afp14	62.5 kDa (554)	Not present <sup>3</sup>	Tape measure protein	gp29	-	PA0625	Pvc14
Afp15	78.9 kDa (696)	Not present <sup>3</sup>	Chaperone/ATPase (predicted)	-	CipV	-	Pvc15
Afp16	32.2 kDa (295)	6	Cap	gp15	TssA	PA0626	Pvc16
Afp17	40.5 kDa (358)	Not present <sup>3</sup>	Remnant toxin	-	-	-	-
Afp18	263.4 kDa (2366)	1 <sup>1,2</sup>	Toxin	-	-	-	-

<sup>1</sup>Number of proteins fitted in the cryo-EM map of the major population of extended AFP (terminated by Afp2-Afp16), except for Afp18, which was estimated to be present as a single copy by mass spectrometry

<sup>2</sup>Detected by mass spectrometry but not localised in the cryo-EM map

<sup>3</sup>Not detected by mass spectrometry

**Supplementary Table 2. Cryo-EM data collection, refinement and validation statistics**

	Extended trunk (EMDB-4802) (PDB 6RBN)	Extended cap ending Afp2-Afp16 (EMDB-4784) (PDB 6RAP)	Extended cap ending Afp3-Afp16 (EMDB-4801)	Extended baseplate C6 (EMDB-4782) (PDB 6RAO)	Extended baseplate C3 (EMDB-4800) (PDB 6RBK)	Extended needle from subtracted images (EMDB-4781)	Contracted sheath C6/helical (EMDB-4803) (PDB 6RC8)	Contracted sheath C6 (EMDB-4859)	Contracted baseplate (EMDB-6RGL)
<b>Data collection and processing</b>									
Magnification	37,000x	37,000x	37,000x	37,000x	37,000x	37,000x	47,000x	47,000x	47,000x
Voltage (kV)	300	300	300	300	300	300	300	300	300
Electron exposure (e-/Å <sup>2</sup> )	20	20	20	20	20	20	27	27	27
Defocus range (µm)	1.0-3.5	1.0-3.5	1.0-3.5	1.0-3.5	1.0-3.5	1.0-3.5	1.0-3.5	1.0-3.5	1.0-3.5
Pixel size (Å)	1.35	1.35	1.35	1.35	1.35	1.35	1.40	1.40	1.40
Symmetry imposed	C6, helical	C6	C6	C6	C3	C3	C6, helical	C6	C6
Initial particle images (no.)	30,858	30,858	30,858	46,991	46,991	46,991	15,189	15,189	30,378
Final particle images (no.)	30,858	23,797	7061	46,991	46,991	46,991	15,189	15,189	14,453
Map resolution (Å)	2.78 Å	3.2 Å	3.4 Å	3.1 Å	3.3 Å	3.7 Å	3.8 Å	4.2 Å	4.6 Å
FSC threshold	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143
Map resolution range (Å)	2.7-3.3 Å	3.0-4.2 Å	3.0-4.2 Å	2.9-4.2 Å	3.1-4.2 Å	3.4-5.0 Å	3.8-5.0 Å	4.0-6.0 Å	4-12 Å
<b>Refinement</b>									
Initial model used (PDB code)	<i>de novo</i>	<i>de novo</i>	-	<i>de novo</i>	<i>de novo</i>	-	<i>de novo</i>	-	-
Model resolution (Å)	3.1	3.4	-	3.3	3.4	-	4.4	-	-
FSC threshold	0.5	0.5	-	0.5	0.5	-	0.5	-	-
Model resolution range (Å)	∞ - 3.1	∞ - 3.4	-	∞ - 3.3	∞ - 3.4	-	∞ - 4.4	-	-
Map sharpening B factor (Å <sup>2</sup> )	-92	-100	-95	-105	-102	-166	-260	-240	-270
<b>Model composition</b>									
Non-hydrogen atoms	7745	9879	-	22,409	7378	-	5423	-	-
Protein residues	1004	1398	-	3608	987	-	805	-	-
Ligands	0	0	0	0	0	0	0	-	-
<b>B factors (Å<sup>2</sup>)</b>									
Protein	77	55	-	51	46	-	91	-	-
Ligand	-	-	-	-	-	-	-	-	-
<b>R.m.s. deviations</b>									
Bond lengths (Å)	0.0068	0.0084	-	0.0086	0.0090	-	0.0036	-	-
Bond angles (°)	1.27	1.40	-	1.44	1.57	-	1.07	-	-
<b>Validation</b>									
MolProbity score	1.56	1.82	-	2.31	2.17	-	1.91	-	-
Clashscore	3	5	-	10	5	-	5	-	-
Poor rotamers (%)	0.8	1.5	-	2.0	1.7	-	1.1	-	-
<b>Ramachandran plot</b>									
Favored (%)	92.9	90.1	-	87.6	84.5	-	87.6	-	-
Allowed (%)	6.8	9.7	-	11.8	15.0	-	12.4	-	-
Disallowed (%)	0.3	0.2	-	0.6	0.5	-	0.0	-	-



**Supplementary Table 4.** Interaction energy and buried surface area between extended AFP needle, tube and cap subunits

Interaction	$\Delta G$ , kcal/mol	Buried Surface Area, $\text{\AA}^2$	Description
Afp8-Afp7	-19.6	1568.8	Value for 1x Afp8 + 4x Afp7
Afp7-Afp5	-12.0	859.9	Value for 1x Afp5 + 2x Afp7
Afp5-Afp1	-17.2	1087.9	Value for 1x Afp5 + 4x Afp1
Afp1-Afp1	-22.7	1325.8	Value for 1x Afp1 in one layer with 4x Afp1 from adjacent layers
Afp1-Afp16	-20.7	1594.0	Value for 1x Afp16 + 3x Afp1

**Supplementary Table 5.** Comparison of AFP, R-pyocin and T6SS state tube-tube, sheath-sheath and tube-sheath interaction energies

Interaction	AFP $\Delta G$ , kcal/mol	R-Pyocin $\Delta G$ , kcal/mol	T6SS $\Delta G$ , kcal/mol
Tube-sheath	-4.2	-1.4	-3.3
Tube-tube	-51.9	-49.0	-49.1
Sheath-sheath (extended)	-105.2	-57.5	-129.4
Sheath-sheath (contracted)	-111.0	-68.0	-161.2
<b>Contracted and extended sheath-sheath difference</b>	<b>-5.8</b>	<b>-10.5</b>	<b>-31.8</b>

**Supplementary Table 6.** Interaction energies and buried surface areas between Afp2 and Afp3 subunits for extended and contracted AFP sheath

Interaction	Extended state		Contracted state	
	$\Delta G$ , kcal/mol	Contribution to total	Buried surface area, $\text{\AA}^2$	Buried surface area, $\text{\AA}^2$
Afp2 C-ter-Afp3 C-ter (layer n-1)	-27.4	26%	1417.4	1529.5
Afp2 C-ter-Afp3 C-ter (layer n+1)	-30.0	29%	1613.1	1559.6
Afp2 C-ter-Afp3 N-ter (layer n+1)	-10.9	10%	691.5	531.4
Afp2 N-ter-Afp2 C-ter (layer n)	-12.4	12%	617.0	554.7
Afp2 N-ter-Afp3 C-ter (layer n-1)	-10.6	10%	649.6	640.1
Afp3 N-ter-Afp3 C-ter (layer n)	-13.9	13%	657.4	569.2
Afp2-Afp2 (layers n+2/n-2)	-	-	-	567.7
Afp3-Afp3 (layers n+2/n-2)	-	-	-	587.7
<b>TOTAL</b>	<b>-105.2</b>	<b>100%</b>	<b>5646.0</b>	<b>6539.9</b>

