This electronic thesis or dissertation has been downloaded from Explore Bristol Research, http://research-information.bristol.ac.uk

Author:
Coombs, Jen

Title:
Structural characterisation of large macromolecular complexes using cryo-TEM

General rights
Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

Take down policy
Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited in Explore Bristol Research. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact collections-metadata@bristol.ac.uk and include the following information in your message:

- Your contact details
- Bibliographic details for the item, including a URL
- An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.

Structural characterisation of large macromolecular complexes using cryo-TEM

Jennifer Coombs

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of PhD in the Faculty of Life Sciences

School of Biochemistry

August 2019

Word Count: 36,231

Abstract

Over the last 20 years cryo transmission electron microscopy (cryo-TEM) has become established as a vital tool used to answer important biological questions structurally and facilitate developments in the fields of medicine and bionanotechnology. In this thesis I discuss two projects in which cryo-TEM has played a pivotal role in our understanding of complex macromolecular systems.

Type II chaperonins (known as thermosomes in archaea) are cage-like structures formed of back to back rings of subunits which close in an ATP dependent manner, sequestering unfolded proteins from the cytosol, allowing them to fold. The ATP binding and hydrolysis behaviour of the thermosome from *T. acidophilum*, along with a panel of mutants, had been previously characterised. Single particle electron microscopy reported in this thesis reveals the structures of the wild type, and a mutant in which the ATP hydrolysis has been blocked in one of the two subunits. The addition of ATP and aluminium fluoride captured the systems in a post-hydrolysis state and allowed comparison of the two systems. Surprisingly, the mutant was found to be in an open conformation, while the wild type was in a closed conformation similar to a previously reported crystal structure, suggesting strong cooperative behaviour between subunits in a ring. We also observed a concerted motion in all three domains of the subunits during closing, analogous to that seen for the homo-oligomeric thermosome from *Methanococcus Maripaludis*.

The second project reported in this thesis is a structural characterisation of a peptide-based nanoparticle, designed for applications such as drug delivery. Previous studies suggested that the modular system self-assembles into a unilamellar spherical particle. I undertook a broad ranging characterisation study which combined cryo-TEM, cryo-correlative light-electron microscopy, electron energy loss spectroscopy and small angle x-ray scattering. These revealed that the particles had a more complex, less ordered structure than previously envisaged. Cryo-TEM demonstrated internal structure which can only be attributed to component peptide. Alternative structures are discussed, and I propose it is most likely a sponge-like structure in which the self-assembling lattice propagates in three dimensions rather than two.

For my grandparents

To Dadcu, I stepped into your electron microscopy shoes without even realising!

To Grandpa, walking around Bristol always reminds me of your shelves full of books about Isambard Kingdom Brunel. I think of you every time I go to the suspension bridge or the train station.

To Granny and Mamgu, because if there’s anything I’ve learned in the last five years it’s that men shouldn’t get all the glory!!

Acknowledgements

It is impossible to imagine that I would have reached this day without the help and support of a huge number of incredible people, I really can't thank you all enough.

Firstly, to my supervisory team: Paul, Danielle, Giulia and Dek - thank you for all of your support, encouragement, patience and advice. Paul, it has been a joy to be a part of your team and I thank you for all the amazing opportunities you have given me. Danielle, you rescued me and taught me what academia can, and should, be. Giulia, for teaching me about parlando del sesso degli angeli! And Dek, although we may not always have seen eye to eye, you pushed me to be the best I can be, and I would not be who I am today without that experience.

Thank you to everyone who has helped with experiments and scientific discussions along the way. Team SAGE, and Harriet in particular, for invaluable advice, support and perseverance! Judith for teaching me everything I know about electron microscopy, you can never be appreciated enough. Andy and everyone at eBic for your help on the microscopes, and to Deb and Sesh for modelling help on both projects! Thanks to Marston for help with data processing and lots of interesting discussions about image analysis.

Thanks to everyone at the BCFN, particularly Annela and Terry, for giving me this incredible opportunity and providing such amazing support. Also to Maddy and Kate, for your endless encouragement, venting, cocktails and 'creativity'!

Thank you to Alastair and everyone at the Victor Chang Institute for being so welcoming, sharing the ice cream joy, and mocking my love of wombats!

Huge thanks to everyone in the office(s). Lorna, Tania and Hugh - you've been there for the highs and the lows and have always been there to pick me back up and make me laugh again. Also to the citizens of Pawnee, Indiana for keeping me company on some long nights!

To my family, old and new, for keeping me grounded and reminding me of what's really important.

To Jessica and Reuben, thank you for absolutely not caring about my PhD and bringing me back to the real world when I needed it most.

And finally to Matt, there are no words! You have been there for me from the very start and taught me to be strong.

This thesis would not exist if it weren't for you.

Author’s Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ....

DATE:..........................

Table of contents

List of Figures

List of Tables

Abbreviations

ADP	Adenosine Diphosphate
AFM	Atomic Force Microscopy
AlF_x	Aluminium Fluoride
ATP	Adenosine Triphosphate
CCD	Charged-Coupled Device
CCT	Chaperonin Containing TCP-1
CLEM	Correlative Light Electron Microscopy
CT	Computed Tomography
CTF	Contrast Transfer Function
DED	Direct Electron Detector
DLS	Dynamic Light Scattering
DMF	Dimethylformamide
DPDS	2,2’-Dipyridyldisulfide
DQE	Detective Quantum Efficiency
eBic	electron Bioimaging centre
EELS	Electron Energy Loss Spectroscopy
EMDB	Electron Microscopy Data Bank
ET	Electron Tomography
FEG	Field Emission Gun
FM	Fluorescence Microscopy
FSC	Fourier Shell Correlation
GFP	Green Fluorescent Protein
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HPLC	High Performance Liquid Chromatography
LaB_6	Lanthanum hexaboride
LMFM	Lateral Molecular Force Microscopy
MBP	Maltose Binding Protein
MDFF	Molecular Dynamics Flexible Fitting
MeCN	Acetonitrile
NMR	Nuclear Magnetic Resonance
pSAGE	protein Self-Assembling peptide Cage
RMSD	Root Mean Squared Deviation
SAGE	Self-Assembling peptide Cage
(E)SAXS	(Extremely) Small Angle X-ray Scattering
SEM	Scanning Electron Microscopy
SNR	Signal to noise ratio
SPA	Single Particle Analysis
ST	Sub-tomogram
TEM	Transmission Electron Microscopy
TEM	Transmission Electron Microscopy
TFA	Trifluoroacetic acid
TFE	Trifluoroethanol
TriC	TCP-1 Ring Complex
UA	Uranyl Acetate

Chapter 1: **Introduction**

This chapter will serve as an introduction to the transmission electron microscope (TEM), with a particular focus on those features and issues which are pertinent to the imaging of biological macromolecules.

Light microscopy is limited by the wavelength of visible light. According to the Abbé diffraction limit, the maximum achievable resolution for a conventional microscope is approximately half the wavelength of the illuminating light. In practise this corresponds to a resolution limit of around 200 nm for standard light microscopes [1]. Super resolution techniques have pushed this to sub 100 nm [2], but what if we wish to image and determine the structure of an object such as a protein or nanoparticle which is on the order of 10-100 nm?

The wavelength of an electron is inversely related to the energy by the de Broglie relationship. For a microscope producing electrons with an energy of 100 keV, the de Broglie wavelength is on the order of picometres, therefore its theoretical resolving power is much smaller than the size of an atom [3]. In practise this is not achieved, due to limitations posed by aberrations of the electromagnetic lenses and detector systems. As developments such as lens correctors and better detectors are beginning to overcome these issues, atomic resolution can now be obtained in a well ordered, strongly scattering object, such as a metal or semiconductor. Less well ordered, soft materials such as biological macromolecules, however, present a different challenge [4].

The transmission electron microscope was first demonstrated in 1931, but it was not until 1986 that Ernst Ruska won the Nobel Prize in physics for its development. During this time, the TEM had become established as “one of the most important inventions of this century” with significance in fields of medicine and biology [5].

Today the TEM is ubiquitous in biological, chemical, and materials science laboratories around the globe, as a vital tool for visualising internal structure and composition. It has also recently received worldwide attention as the 2017 Nobel Prize for chemistry was awarded to Jaques Dubochet, Joachim Frank and Richard Henderson “for developing cryo-electron microscopy for the high resolution structure determination of biomolecules in solution” [6].

1.1 Anatomy of an electron microscope

This section will discuss the main components of a transmission electron microscope (TEM) from electron source to image. An overview of the anatomy of a basic electron microscope is shown in Figure 1-1.

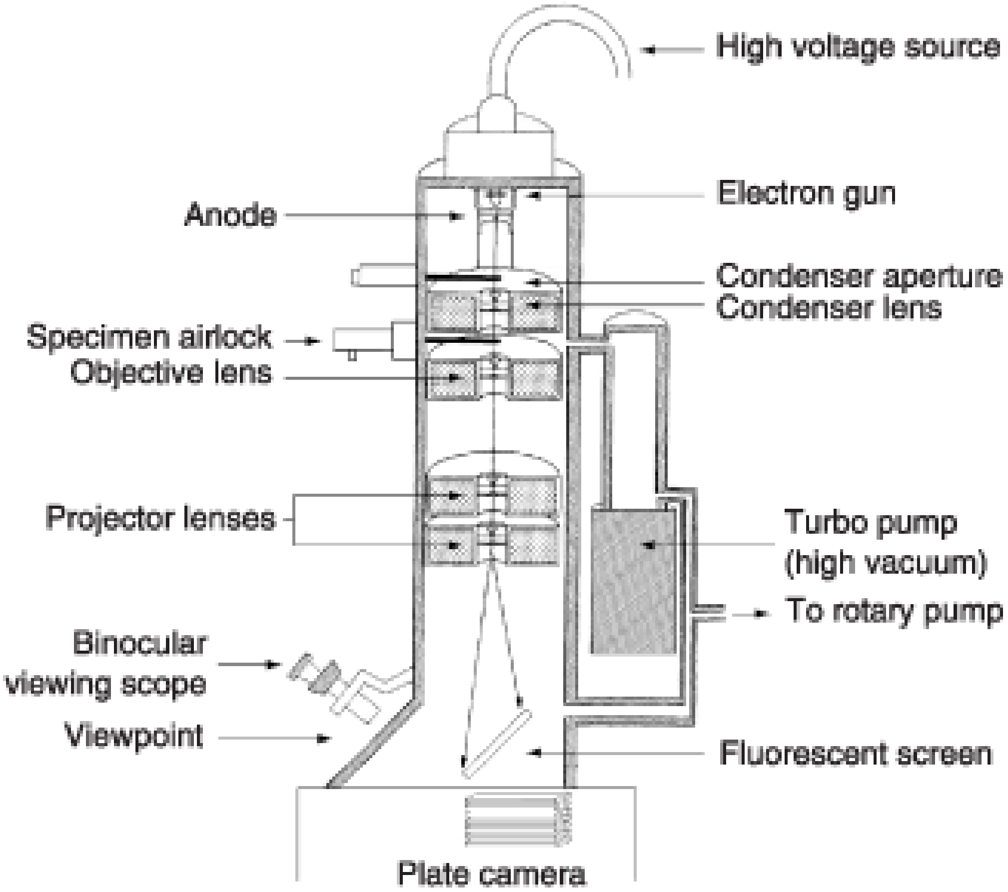


Figure 1-1: Overview of a transmission electron microscope. Reproduced from [7]

The key elements in an electron microscope are the electron source, or gun, the lenses, apertures, and detectors, each of which is discussed in more detail below. The whole system is maintained under vacuum by the high vacuum pumps seen in Figure 1-1. Lead shielding is placed around the electron microscope to protect the user from the high energy electron beam. The viewing scope and fluorescent screen allow the user to see an image in real time. The fluorescent screen is then lifted to allow the electrons to reach the detector which in this representation is a plate camera. Figure 1-1 is a simplified description of the imaging system where the sample itself is inserted in between the condenser and objective lenses. In fact, in a modern microscope the sample is inserted between upper and lower components of a Twin Objective lens. The gap between upper and lower polepieces affects the achievable resolution, with smaller gaps giving better resolutions, however this reduces the available room for sample holders and limits the tilt range in tomography (see Section 1.4.1).

Sample holders may be in a side entry configuration as shown here, or alternatively a cartridge system which holds several grids at once with an internal mechanism to switch between them.

1.1.1 Electron Source

Electrons can be produced by two types of source, known as thermionic or field-emission sources.

Thermionic sources use heat to release electrons from a bulk material. The first example of these is the tungsten filament, which closely resembles the wire in a lightbulb. The high melting point of tungsten means that we can put a high voltage though this filament causing it to emit electrons without melting. The alternative is a Lanthanum hexaboride (LaB_6) crystal, whose low work function means that electrons are emitted at temperatures well below its melting point. In practise, LaB_6 electron sources are now more commonly used than tungsten due to their longer life, higher efficiency, brightness and greater electron coherence (*i.e*. smaller spread of emitted electron energies) [3].

Field-emission electron sources on the other hand function by application of a very high electric field to a sharply pointed emitter. This effectively pulls electrons from the surface at very coherent energies. Field emission guns (FEGs) can be further classified into warm and cold, where surface contamination is minimised either by operating the gun at ambient temperatures in ultra-high vacuum (cold) or by heating the gun and working at lower vacuum (warm) [3].

Field emission guns produce electron beams which are extremely bright, and coherent (particularly cold-FEGs). However, the sharp point required for the source means that the illumination area is small compared to thermionic sources. This is not ideal for cellular EM for example where contextual information is needed. The ideal electron gun for a microscope is therefore not exclusively governed by physical properties but also the applications for which it will most often be used.

As mentioned above, the wavelength of an electron is governed by the accelerating voltage of the microscope: the higher the voltage, the higher the resolving power (assuming a perfect lens system). There is an ultra-high voltage electron microscope which operates at 3000 kV based at Osaka university, in Japan, which can reach resolutions of 0.14 nm [8]. However, such high energies are rarely necessary and would cause damage to typical biological materials. Standard biological TEM is therefore typically performed on microscopes operating between 100 and 300 kV.

1.1.2 Lenses

Instead of the glass lenses used in light microscopes, electron beams are focussed using magnetic lenses. Magnetic lenses consist of large coils of copper wire which carry current. This creates an electric field which bends the electron beam. The strength of the field can be varied in order to change the strength of the lens. This means that in an electron microscope the lenses are in a fixed position rather than having to be moved up and down like optical lenses.

Spherical aberration is a result of the fact that the field within the lens is not uniform, so electrons further from the central axis are bent more than on-axis electrons. This leads to a single point being spread out in the image, limiting the resolution of the microscope.

Magnetic lenses also bend electrons of different energies to different extents. Images are created by scattering electrons off a sample, producing different energies. These scattered electrons have lower energies and are therefore bent more by the lens. This once again leads to spreading of the signal, in this case known as **chromatic aberration**.

An **aperture** is a small hole in a metal disk which is placed in the line of the beam. This limits the off-axis electrons which are worst affected by these aberrations.

Many other aberrations can result from physical defects in the lenses, misalignment of lenses and apertures, contamination *etc*. It is therefore vital that a microscope is well aligned prior to data collection, particularly if high resolutions are required [3].

1.1.3 Detectors

Recent development of the cameras used in transmission electron microscopes has had a huge effect on our ability to collect high resolution images of biological samples.

Until electronic detectors were developed in the 1990s, TEM images were collected on photographic film. These produced high quality images and were not limited by the pixelation inherent to electronic devices. However, developing images was a time-consuming process, limiting the throughput of the technique [9].

The signal to noise ratio of an image (SNR) created by a detector is affected by the **detective quantum efficiency** (DQE) of that detector according to:

$$DQE(\omega) = \frac{(SNR_{out})^2}{(SNR_{in})^2}$$

When working with biological materials, relatively low electron doses (short illumination times) are used in order to minimise damage to the sample. This results in noisy images. Ideally the DQE of a detector would be 1, *i.e.* the detector would not introduce any extra noise to the images. In practise however, this is never the case and all detectors introduce noise to the image. In order to obtain interpretable images for biological materials we must work with maximum possible DQE [10].

The first electronic cameras used in TEM were **charge-coupled devices** (or CCDs). For electron energies higher than 100 keV, the DQE of these devices was lower than that of photographic film due to scattering of electrons and within the sensor. Photographic film was therefore favoured for structural biological imaging despite the increased contrast at low frequencies and instant image production that CCDs provided [9].

The latest cameras being developed for TEM use **direct electron detection**. These read out signal on a rolling basis, meaning that electrons can be detected accurate to a single pixel by reducing the dose per frame and detecting single electron events. This is known as counting mode which can give DQEs as high as 0.8, double the equivalent for photographic film [11].

For the user, another key advantage of the direct electron detector is the ability to collect multi-framed micrographs or movies. This means that the total dose used is separated into several individual micrographs which can then be aligned to correct for sample motion during imaging. Furthermore, it is possible to remove later frames if the cumulative dose has resulted in damage to the sample [12].

1.2 Image formation

TEM images are formed by two contrast components, amplitude and phase. These are caused by the different types of interaction an electron can have with the sample. These interactions, or scattering events, can be either inelastic, in which electrons lose energy to the sample, or elastic, in which no energy is transferred.

Amplitude contrast is the differential loss of amplitude from the beam across the sample. When an electron interacts with a heavy atom such as that of a heavy metal stain (see Section 1.3.2), it results in an inelastic scattering event. In this case electrons are deflected or absorbed by the sample. The transmitted beam is then focused by the lens but the deflected electrons are blocked by the aperture, preventing them from contributing to image formation. This leaves a transmitted wave of unscattered electrons which has lower amplitude at positions corresponding to regions of electron density in the sample, creating an image at the detector

Phase contrast on the other hand arises from the interaction of unscattered and elastically scattered electrons. Biological materials can be described as phase objects as they contain light atoms, and consequently much less inelastic scattering occurs. When the electrons interact with the sample, elastic scattering results in a change in phase but not amplitude. An image is then formed as the unscattered waves interact constructively and destructively with the elastically scattered beam depending on whether they are in or out of phase with each other. The amplitude of the resulting wave at the detector is therefore modulated resulting in a final image.

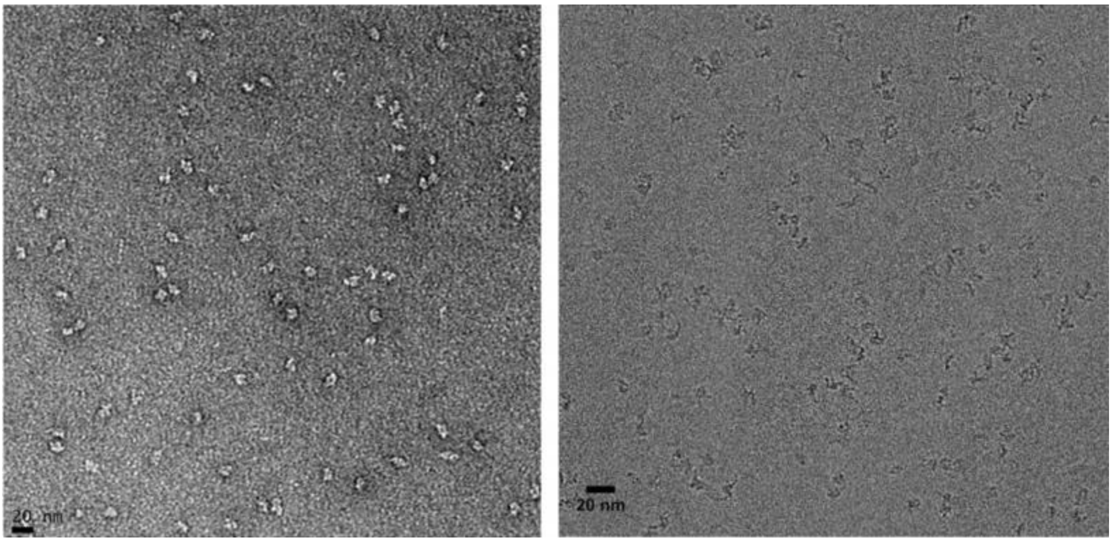


Figure 1-2: Amplitude contrast from negative stain (left) vs phase contrast in cryo-TEM (right) Adapted from [13].

As discussed above (section 1.1.2), magnetic lenses are imperfect which results in spherical aberration. This creates additional phase shifts in electrons which are scattered at a higher angle, i.e. those corresponding to higher resolution features have a higher phase shift relative to the unscattered beam. This creates an oscillating pattern of high and low contrast with increasing spatial frequency as shown by the black line in Figure 1-3 (A). In a Fourier transform of an image, this manifests as concentric light and dark rings known as **Thon rings** [14].

In addition, the oscillating function is attenuated by an envelope function (green lines) which are a result of the imperfect electron source. Since the electrons produced are neither perfectly coherent, nor perfectly monochromatic, the contrast produced at higher spatial frequencies are damped as noise is introduced. Together the oscillation and envelope are known as the **contrast transfer function** (CTF) (Figure 1-3), which describes the attenuation and modulation of signal mathematically. The shape of the function depends on several parameters; the defocus, the spherical aberration coefficient, the source size, the defocus spread and the accelerating voltage.

At focus, the CTF contains few zeros, however relatively little low resolution contrast is available, making it difficult to find whole particles in the image. Below the focus level (when the image is defocused), the oscillations are compressed, increasing low resolution contrast and allowing particles to be easily recognised. On the other hand the higher resolution information is more complex to interpret as contrast is inverted more frequently. In both situations there also remain spatial frequencies in which no contrast is present and therefore no information at the corresponding resolution is contained within the image.

By fitting a theoretical CTF to the experimentally derived power spectrum of the original image, we can correct for phase reversals and amplitude attenuation. Correcting for phase reversals or ‘phase flipping’ still leaves zero values at particular frequencies. To resolve this, imaging strategies tend use a range of defocus values such that zeros at one defocus value are non-zero at another, thereby sampling the full frequency range [15].

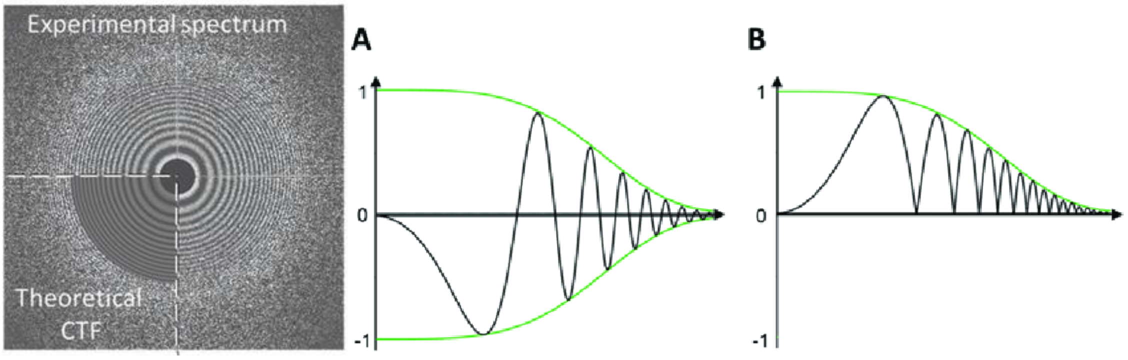


Figure 1-3: CTF Correction. (left) Experimental power spectrum and theoretical CTF. Thon rings match between the two showing good fitting of CTF. (A) shows the oscillation of the CTF within the envelope which results from imperfections of the microscope. (B) Correction of CTF by flipping the phase of the negative regions. Information is lost at the zeros. Adapted from [16]

1.2.1 Cryo-TEM imaging of biological samples

Imaging biological samples present significant challenges. By their nature, they are very beam sensitive but also extremely low contrast phase objects. Visualising a sample at low temperature reduces beam damage, but also offers the opportunity to prepare a sample in a fully hydrated or most natural state. This is what is meant by cryo-TEM. To make this work, requires optimised microscopes and sample holders, sample supports or grids, specialised sample preparation equipment and specialised image processing techniques.

1.3 Sample preparation for electron microscopy based structural analysis

1.3.1 Grid Types

There are many TEM grids available on the market for a wide range of different applications. They are 3 mm diameter metal grids with holes through which a sample suspended on a support is visualised.

The most common grids are made of copper which is cheap, however other materials such as gold or molybdenum may be used for example for studying cells (gold is non-toxic) or working at high temperatures. The size and shape of the holes in the mesh can be varied as well as the coating which supports the sample.

The simplest support film is an amorphous carbon coating which does not diffract the electron beam and is therefore considered transparent. Since negatively stained samples give high contrast, carbon/formvar (a type of plastic) films are often used for added strength. For low contrast samples such as in cryo-TEM we aim to minimise the contrast from support films as there is no added stain. The sample is therefore embedded in vitreous ice which is suspended in holes in a carbon film. These can take the form of regular holes of a specific size and spacing such as Quantifoil or C-flats, but are also available in irregular holes such as lacey or holey films. Sometimes however, it is helpful to add an extra support layer to this holey film. This can take the form of an extremely thin (*e.g*. 4nm) amorphous carbon film, graphene or graphene oxide. The choice of grid is highly sample specific.

1.3.2 Negative Stain

The most commonly used method for creating contrast when imaging materials with light atoms (*e.g*. biological materials) is to use a heavy metal stain such as uranyl acetate, lead citrate, ammonium molybdenate or phosphor tungstic acid. When the stain is added to the grid, it collects around the edges of the sample. The sample then creates a region of negative space, appearing white on a dark background, hence the term negative stain. If the staining is of high quality, the structural details of a macromolecule can be seen. However, the sample has been dried down on the grid, so may well be deformed.

Negative staining can be a quick and easy method, going from grid to microscope in as little as 5 minutes. It creates lots of contrast and does not require a large volume of sample, making it excellent for rapidly screening lots of conditions.

There are several down sides to this method, however. Such stains are prone to artefacts which can be difficult to distinguish from the sample. Stains can interact with the sample or be sensitive to pH. The resolution may also be limited for as the stain cannot fully penetrate into the protein and therefore does not give a true representation of the protein’s internal structure.

1.3.3 Vitrification Methods

In order to reach the highest possible resolutions, the sample must be imaged in as close to native state as possible. We therefore aim to avoid staining, fixing or drying the sample. This can be achieved, and a hydrated state maintained, for solutions of small biological samples (<1 µm) by suspending in ice in a method known as plunge freezing.

In this method, a small film of water is cooled so rapidly that the molecules are not able to rearrange to form crystals before being immobilised. Instead, the water is frozen in a disordered state, known as vitreous (glassy) ice. This type of ice is transparent (non-diffracting) to an electron beam, providing very low background noise in a TEM image. This process, and indeed the full cryo-TEM workflow, is excellently described in Dubochet *et. al*.’s seminal paper published in 1988 [17].

The vitrification process is achieved by plunging a TEM grid with a thin layer of sample in solution into liquid ethane cooled to liquid nitrogen temperatures. It is not possible to do this using liquid nitrogen alone as it does not have sufficient thermal capacity and boils before the sample is fully vitrified.

Initial plunge freezers were generally custom made, guillotine like devices, however these devices tended to produce results which were very difficult to reproduce. The second-generation devices have attempted to make this process more reproducible and applicable to a wide range of samples. Of these, the most popular examples are the Vitrobot from ThermoFisher Scientific, and Leica's EM GP. Both machines offer the same main features. There is a chamber at the top in which temperature and humidity can be controlled. Other settings such as how long the sample is on the grid before blotting, how long the blot lasts, blotting pressure *etc.* are user controlled. The grid is then dropped rapidly into a small pot containing either liquid ethane or a mixture of propane and ethane. This is cooled by a secondary container holding liquid nitrogen. Once the grid is frozen it is transferred into a box for storage or to the microscope to be imaged. Frozen grids must be kept below -140ºC at all times to prevent the formation of crystalline ice [17].

A third generation of plunge freezing devices are currently being developed, which aim to minimise some of the problems seen with standard machines.

One key problem is that proteins in solution migrate towards the air water interface where it is likely that they become denatured. The time frame between blotting and freezing in a standard vitrification device is on the order of a few seconds. In this time a layer of protein can form at the surface, creating additional background noise and leading to preferential orientation of proteins in solution [18].

One method being explored to minimise this is to spray the sample onto the grid in very small droplets while the grid is moving rapidly into the cryogen. In this case the time between sample application and vitrification is ~255ms. To aid this, self-blotting grids were used which have nanowires growing from the grid itself. These make use of the capillary effect between nanowires to spread the small droplets over the surface of the grid [19].

Another alternative to the traditional plunge freezing methods is to 'write' a few nL of solution onto the grid using a microfluidics device. The small volume of liquid applied means that no blotting is required, minimising potential contamination. The issue of sample aggregating at the air water interface is not however, avoided. In conjunction with this work, an automating screening workflow was developed allowing samples and freezing conditions to be rapidly tested. Since at present these must be optimised for each individual plunge freezer and sample this could represent a significant time saving for the experimentalist [20].

The downsides of these new methods at present are that they both produce a relatively small useable area of thin vitreous ice.

Vitrification is a much more time-consuming process that negative stain. It can be very sensitive to buffer conditions, salt, *etc.* and so must be optimised for every sample. All liquid nitrogen must be kept as clean as possible, as any water which freezes in it will tend to stick to surfaces, causing contamination of the grid. Ideally therefore, vitrification and any transfer steps would be performed in a humidity-controlled room. All subsequent grid handling steps create additional opportunity for the introduction of contamination so handling should be kept to a minimum.

Imaging of vitrified grids also poses several difficulties which are not seen in negative stain. The sample must be kept cold during imaging requiring a special sample holder. The ice and sample contained within it are easily damaged by the electron beam. Imaging is therefore performed under low-dose conditions to prevent over-exposure. In general, this means that focusing is performed on an area away from the region being imaged and the beam current is maintained as low as possible while still being enough to produce contrast.

1.4 3-Dimensional TEM

1.4.1 Tomography

The technique of tomography is used to create 3D reconstructions from TEM images, particularly of cellular samples. The method is analogous to Computed Tomography (CT) scanning in medicine which uses X-rays. During a CT scan the camera moves around the person being imaged, but in EM the sample is tilted within the microscope. Images are thus collected at a range of angles, typically from -60 to +60 degrees. The angles between different images in a series are known. The images are then back projected to compute a 3D image. Small particles, called fiducials, with high contrast can be added to facilitate the alignment of images during reconstruction. This can also be done under cryogenic conditions (cryo-ET), where once again it is vital to ensure the focus area is away from the area of interest.

Whilst structural information about the cellular environment and processes can be obtained from cryo-ET there are inherent limitations to the achievable resolution. The reconstructions have low contrast and anisotropic resolution (*i.e*. resolution in z is different to that in x-y). Also, importantly they suffer from a lack of information in certain directions known as the missing wedge due to the data acquisition geometry. A recent and exciting development in the field of cryo-ET is subtomogram averaging (ST) where sub regions containing the volume of interest in the tomogram are extracted, aligned and averaged together. This increases the SNR in the data and has the potential to overcome the effects of the missing wedge if the feature is oriented at a different angle in different tomograms. Reconstructions from ST data have now broken the sub-nanometer barrier and are being applied to a wider range of samples [21].

1.4.2 Single Particle Analysis

Despite the recent advances made in subtomogram averaging, sub-nanometer resolution is not readily obtained. To gain high resolution 3D structural information of individual proteins or macromolecular complexes the image processing technique of single particle analysis (SPA) [7] is becoming an increasingly popular method.

This method relies on a solution of particles (proteins, viruses, nanoparticles *etc.*) which are identical copies of each other. The particles, in solution, are immobilized in ice (see Section 1.3.3) and imaged. Rather than collecting a series of images of the same object at different angles, multiple objects are imaged, providing different views due to their differing orientations. These images are then used to reconstruct a 3D map of the object.

An additional benefit of this is that multiple images of the same view may be averaged together, increasing the signal to noise ratio. This method forms the basis of Chapter 2: of this thesis where a more detailed description of the processing steps may be found. The key steps of a single particle experiment are as follows:

1. Sample is purified and screened by negative stain
2. Grids are prepared by plunge freezing – These two steps may require a lengthy optimisation process
3. Images are collected – typically a multi-day auto-collection after screening grids and selecting good areas
4. Images are pre-processed to correct for drift and CTF
5. Particles are picked from the images – depending on factors such as heterogeneity of the sample, resolution required and symmetry of particles, 10s to 100s of thousands of particles may be needed.
6. 3D models are computationally reconstructed – there are multiple software packages available.

Cryo-TEM allows particles to be imaged in their native, hydrated state. This allows the capture of states which were previously unreachable through crystallography. The number of single particle structures submitted to the Electron Microscopy Data Bank (EMDB) has been growing steadily over the last decade (Figure 1-4). At the time of writing in August 2019, the number of structures deposited this year is almost equal to the number published in 2018.

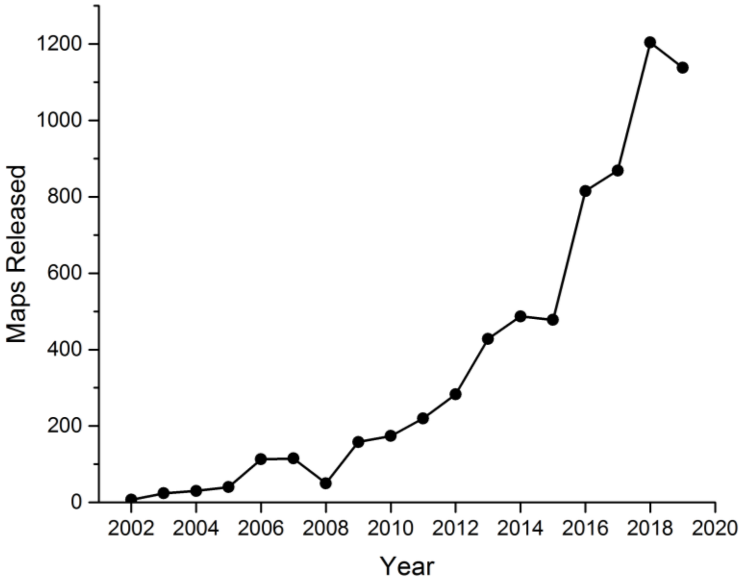


Figure 1-4: Single particle maps updated to EMDB by year. Data taken from EMDB on August 7th 2019.

The resolution of the maps published has also improved year on year as shown in (Figure 1-5). The average resolution has increased from 15-20 Å until 2013 to now being just over 5 Å.

The recent leap in the achievable resolution can be attributed to three main factors: the advent of direct electron detectors, automated data acquisition and the application of maximum likelihood statistical approaches to image processing. These recent advances are discussed in a number of extensive reviews [22-24]

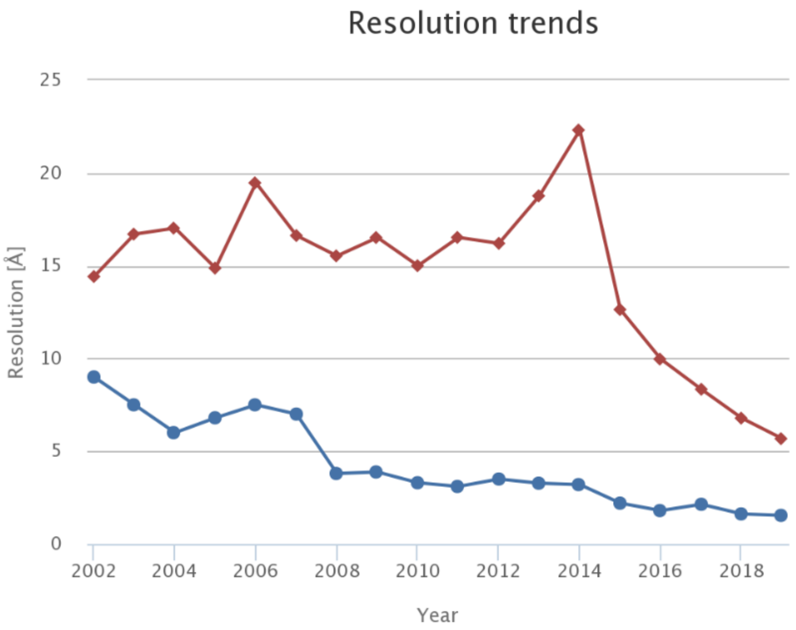


Figure 1-5: Single particle maps by year. Best (blue) and average (red) resolution. Data taken from EMDB on 7th August 2019.

Indeed, new software can improve on maps which have been previously solved. For example, in the release of RELION 3.1, Scheres *et. al*. demonstrated that the new software features could be used to give an improvement of around 0.4 Å on a map which had previously been processed using RELION version 2.1. They also presented a new map of human apo-ferritin with a reported resolution of 1.65 Å which is currently the highest resolution structure (with an associated paper) on the EMDB [25].

With higher resolution maps being produced, the effect of minor aberrations (such as anisotropic magnification) become of increasing concern. Thus, improvements continue to focus on both hardware and software, to continue pushing the abilities of this incredibly powerful technique.

Nevertheless, limitations remain on what may be achieved due to physical limitations of the microscope and poor signal to noise. At present it is extremely challenging to create high resolution maps of small proteins (less than 100 kDa) as the low contrast means small features cannot be aligned.

The development of cryo-TEM has had huge impact in fields of structural biology, and drug design [23, 26, 27]. This thesis will introduce two unique projects to which cryo-TEM has been pivotal in our understanding of a biological system.

References

1. Cremer, C. and B.R. Masters, *Resolution enhancement techniques in microscopy.* European Physical Journal H, 2013. **38**(3): p. 281-344.
2. Vangindertael, J., et al., *An introduction to optical super-resolution microscopy for the adventurous biologist.* Methods and Applications in Fluorescence, 2018. **6**(2).
3. Williams, D.B. and B.C. Carter, *Transmission Electron Microscopy*. 1996, Springer.
4. Zhou, W.Z. and J.M. Thomas, *HRTEM surface profile imaging of solids.* Current Opinion in Solid State & Materials Science, 2001. **5**(1): p. 75-83.
5. NobelPrize.org, *The Nobel Prize in Physics 1986*. 1986.
6. NobelPrize.org, *The Nobel Prize in Chemistry 2017*. 2017.
7. Frank, J., *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State*. 2006: Oxford University Press.
8. *H3000 UHVEM*. 07/08/2019]; Available from: http://www.uhvem.osaka-u.ac.jp/en/what.html.
9. Bammes, B.E., et al., *Direct electron detection yields cryo-EM reconstructions at resolutions beyond 3/4 Nyquist frequency.* Journal of Structural Biology, 2012. **177**(3): p. 589-601.
10. Clough, R.N., G. Moldovan, and A.I. Kirkland, *Direct Detectors for Electron Microscopy.* Electron Microscopy and Analysis Group Conference 2013 (Emag2013), 2014. **522**.
11. McMullan, G., A.R. Faruqi, and R. Henderson, *Direct Electron Detectors.* Methods Enzymol, 2016. **579**: p. 1-17.
12. Faruqi, A.R. and G. McMullan, *Direct imaging detectors for electron microscopy.* Nuclear Instruments & Methods in Physics Research Section

a-Accelerators Spectrometers Detectors and Associated Equipment, 2018. **878**: p. 180-190.

13. Larsen, K.P., et al., *Architecture of an HIV-1 reverse transcriptase initiation complex.* Nature, 2018. **557**(7703): p. 118-122.
14. Thon, F., *Zur Defokussierungsabhangigkeit Des Phasenkontrastes Bei Der Elektronenmikroskopischen Abbildung.* Zeitschrift Fur Naturforschung Part a-Astrophysik Physik Und Physikalische Chemie, 1966. **A 21**(4): p. 476-&.
15. Cong, Y. and S.J. Ludtke, *Single Particle Analysis at High Resolution.* Methods in Enzymology, Vol 482: Cryo-Em, Part B: 3-D Reconstruction, 2010. **482**: p. 211-235.
16. Costa, T.R.D., A. Ignatiou, and E.V. Orlova, *Structural Analysis of Protein Complexes by Cryo Electron Microscopy.* Methods Mol Biol, 2017. **1615**: p. 377-413.
17. Dubochet, J., et al., *Cryo-electron microscopy of vitrified specimens.* Q Rev Biophys, 1988. **21**(2): p. 129-228.
18. D'Imprima, E., et al., *Protein denaturation at the air-water interface and how to prevent it.* Elife, 2019. **8**.
19. Dandey, V.P., et al., *Spotiton: New features and applications.* J Struct Biol, 2018. **202**(2): p. 161-169.
20. Schmidli, C., et al., *Miniaturized Sample Preparation for Transmission Electron Microscopy.* Jove-Journal of Visualized Experiments, 2018(137).
21. Schur, F.K., *Toward high-resolution in situ structural biology with cryo-electron tomography and subtomogram averaging.* Curr Opin Struct Biol, 2019. **58**: p. 1-9.
22. Kuhlbrandt, W., *Structure and Mechanisms of F-Type ATP Synthases.* Annu Rev Biochem, 2019. **88**: p. 515-549.
23. Bradshaw, M. and D.M. Paul, *After the revolution: how is Cryo-EM contributing to muscle research?* J Muscle Res Cell Motil, 2019.
24. Maier, D., et al., *Influence of correlated errors on the estimation of the relaxation time spectrum in dynamic light scattering.* Appl Opt, 1999. **38**(21): p. 4671-80.
25. Zivanov, J., et al., *New tools for automated high-resolution cryo-EM structure determination in RELION-3.* Elife, 2018. **7**.

26. Razi, A., R.A. Britton, and J. Ortega, *The impact of recent improvements in cryo-electron microscopy technology on the understanding of bacterial ribosome assembly.* Nucleic Acids Research, 2017. **45**(3): p. 1027-1040.
27. Morris, E.P. and P.C.A. da Fonseca, *High-resolution cryo-EM proteasome structures in drug development.* Acta Crystallographica Section D-Structural Biology, 2017. **73**: p. 522-533.

Chapter 2: **Cryo-TEM of the *T.acidophilum* Thermosome**

2.1 Introduction

In this chapter I discuss the application of cryo-TEM single particle methods to the development of two high resolution 3D maps. The aim of this work was to interrogate the molecular function of the thermosome, an archaeal type II chaperonin, through structural analyses of the wild type and a site-directed mutant. A brief background and contextualisation are given. I will then discuss some of the differences in grid preparation and detail the image processing workflow that was carried out. Two maps at sub 4 Å resolution are presented into which models are fitted in the subsequent chapter.

2.2 Background

2.2.1 Chaperonins

Chaperonins are a class of ATP-dependent heat shock proteins which help newly synthesised or misfolded proteins to fold by encapsulating and isolating them from the overcrowded cellular environment. They are large, barrel shaped oligomeric complexes of 0.8-1 MDa formed of two back to back rings of subunits enclosing a central cavity. Substrate proteins are sequestered inside this cavity which protects them from other misfolded proteins in the cytosol, thus preventing aggregation and allowing them to refold [1].

Chaperonins are found in all kingdoms of life and are classified into two groups based on their evolutionary lineage. Type I chaperonins, found in bacteria and eukaryotic organelles such as mitochondria and chloroplasts, are the simplest as they are homomeric [2]. Type II on the other hand, found in archaea and the eukaryotic cytosol, are mostly hetero-oligomeric and can contain up to 8 different subunits [3, 4].

All chaperonin subunits can be subdivided into three distinct domains. The equatorial, in which the majority of intra- and inter-subunit contacts and the nucleotide binding pocket are found, the apical domain which undergoes large conformational rearrangement during the opening and closing process, and the intermediate domain which links the two. This allosteric mechanism is driven by the binding and hydrolysis of ATP, which cause rearrangements in all three domains in order to close off the internal cavity. The mechanisms by which the two types of chaperonin function have important commonalities as well as differences [1].

2.2.2 Type I Chaperonins

Found in bacteria and eukaryotic organelles, type I chaperonins consist of two rings composed of 7 identical subunits stacked back to back. Unlike the Type II chaperonins, they require a co-chaperonin, which acts as a lid, to form a closed chamber for substrate folding. The prototypical Type I system is the bacterial GroEL with its co-chaperonin GroES, the first high resolution structure of which was solved by x-ray crystallography in 1994 [5].

The high molecular mass and relative simplicity of the GroEL complex make it excellent for cryo-TEM studies, with cryo-TEM papers as early as 1994 showing substrate bound [6]. At the time of writing, a search of the EMDB database revealed over 60 cryo-TEM structures for GroEL complexes published since 2002. 2017 saw the release of a 3.7 Å GroEL structure in which subunits were classified individually, leading to the discovery of 3 distinct conformations within single GroEL oligomers [7].

2.2.3 Type II Chaperonins

As mentioned above, type II chaperonins are found in archaea (thermosomes) and in the eukaryotic cytosol (CCT/TriC). Unlike Type I they do not require a co-chaperonin to enclose their internal cavity, but instead contain an additional helical protrusion at the tip of each apical domain which acts as an inbuilt lid that opens and closes in an ATP-dependent manner along the folding cycle.

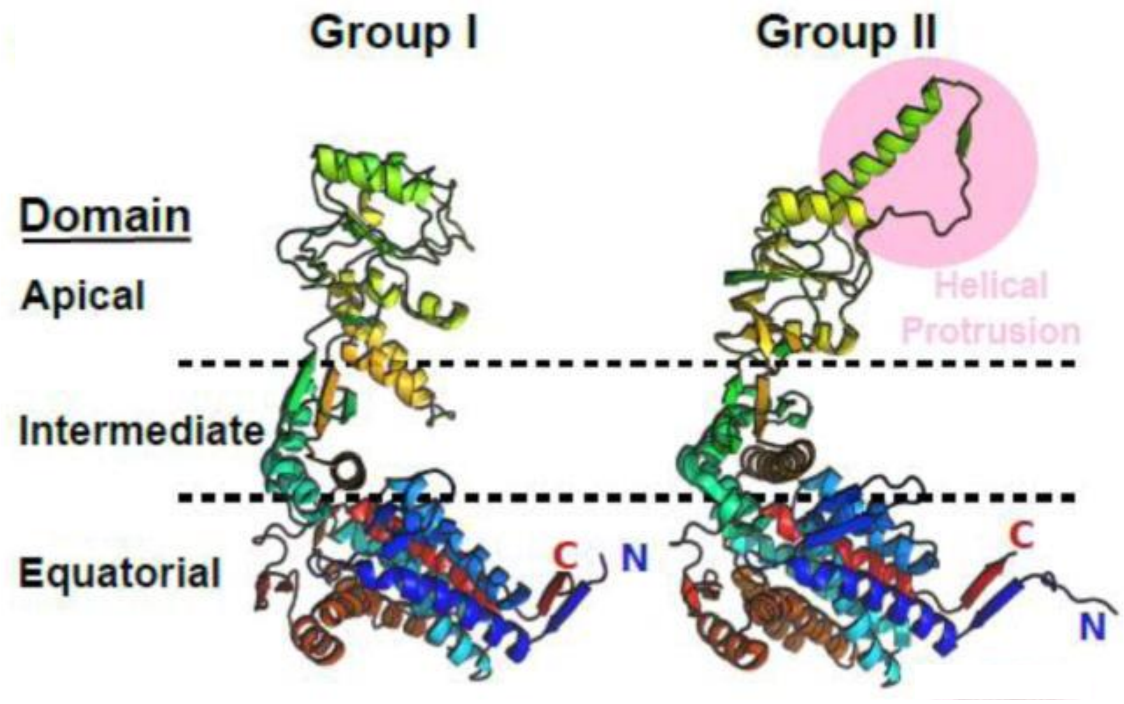


Figure 2-1: Type I vs type II chaperonin structures. Adapted from [3].

Thermosomes can have between 1 and 3 different subunits, arranged in rings of 8 or 9, whereas CCT is formed by two 8-membered rings of different subunits. This increasing level of complexity does make class II chaperonins much more complex to study than GroEL/ES, although in the past 10 years progress in a wide range of structural techniques has been fundamental to our actual understanding of these folding machines [8]. The thermosome therefore represents a good compromise, in that they present all the features of class II chaperonins with a level of complexity, in terms of number of different subunits, lower than their eukaryotic counterparts. In fact, the first high resolution crystallographic structure solved for a group II chaperonin is that of the closed (nucleotide bound) thermosome from the archaeon *Thermoplasma acidophilum* [9].

Excellent reviews comparing the two types of chaperonin can be found at [1, 10].

Some key structural features discussed in these papers which are particularly relevant to this study are summarised below:

1. As mentioned previously, the helical protrusion which acts as an inbuilt lid in type II chaperonins is not present in type I which instead requires a co-chaperonin.

2. Type I rings are offset, with each subunit in contact with two others on the opposite ring. Type II on the other hand have direct contact with only 1 opposite subunit.
3. Type I tend to only be closed in one ring at a time. Steric hindrance means that it can function more efficiently by closing one ring at a time. This is not the case for type II in which this steric hindrance is not seen and has been regularly seen in a fully closed state.
4. The equatorial region of GroEL shows little global motion during opening and closing, although an internal rotation between subdomains is seen [11]. Thermosome equatorial domains act as a rigid body which undergoes a tilt between open and closed conformations [12].

2.2.4 Archaeal Chaperonins: Thermosomes

The variety of Thermosome subtypes that can occur was recently shown in *Sulfolobus shibatae,* for which a series of different subunit combinations were found depending on environmental characteristics and cellular stress [13].

A key difficulty in developing our understanding of the mechanism of thermosomes is capturing them in different states. Cryo-TEM has enabled this, leading to the first sub-nanometer structures of open thermosomes being released in 2010 [14]. This, and two subsequent studies by the same group [15, 16] on the homo-oligomeric thermosome from *Methanococcus maripaludis* (Mm-cpn) are summarised as follows.

Structures were generated for three distinct states of the thermosome: a 4.3 Å closed structure, an 8 Å nucleotide-free open structure [14] and an 8 Å ATP-bound mutant in which hydrolysis is blocked and has an open conformation [16]. In these studies, the open conformations have been depleted of the tips of the apical domains so have no lid region, as it was found that removing this improved the orientation of the particles in ice.

The apical domains of the two open structures were rotated by 45º relative to each other while the other domains remained relatively still. This suggests that the binding of ATP causes an initial motion in the apical domains. Full closure then occurs upon ATP hydrolysis, where comparison with the closed (hydrolysed) structure shows a large movement which encompasses all domains.

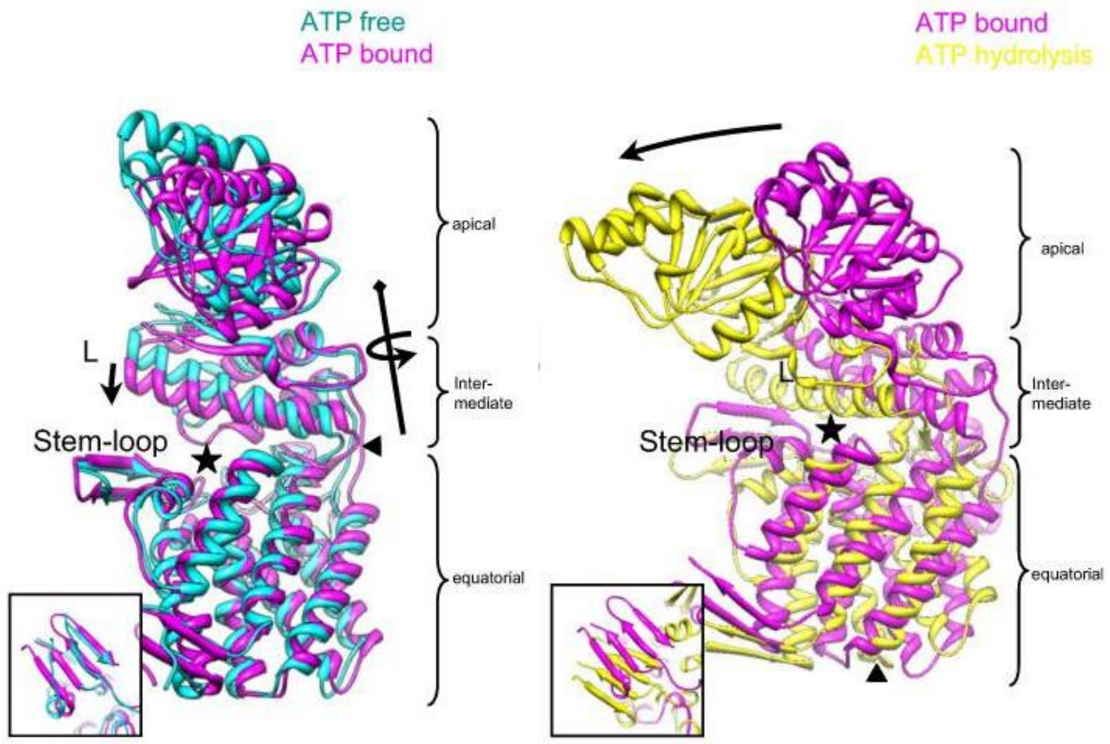


Figure 2-2: Comparison of structure in different states of Mm-cpn. Showing small rotation between the ATP free and ATP bound states (left) and a large tilt between ATP bound and ATP hydrolysis states (right). Adapted from [16]

They identified key movements of residues linked to ATP hydrolysis (in helix L and stem loop labelled in Figure 2-2) and interactions between N and C termini with the stem loop of the adjacent subunit.

2.2.5 Thermosome of *Thermoplasma acidophilum*

Unlike Mm-cpn, the thermosome from *T. acidophilum*, has two distinct subunits alternating in two 8 membered rings. Since it was the first thermosome to be crystallised, with a structure solved at high resolution published in 1998 [9], it became a model system for studies into class II chaperonins. It may be used as a foundation to understand the role of heterogeneity between subunits on the allosteric behaviour and folding mechanism as a whole.

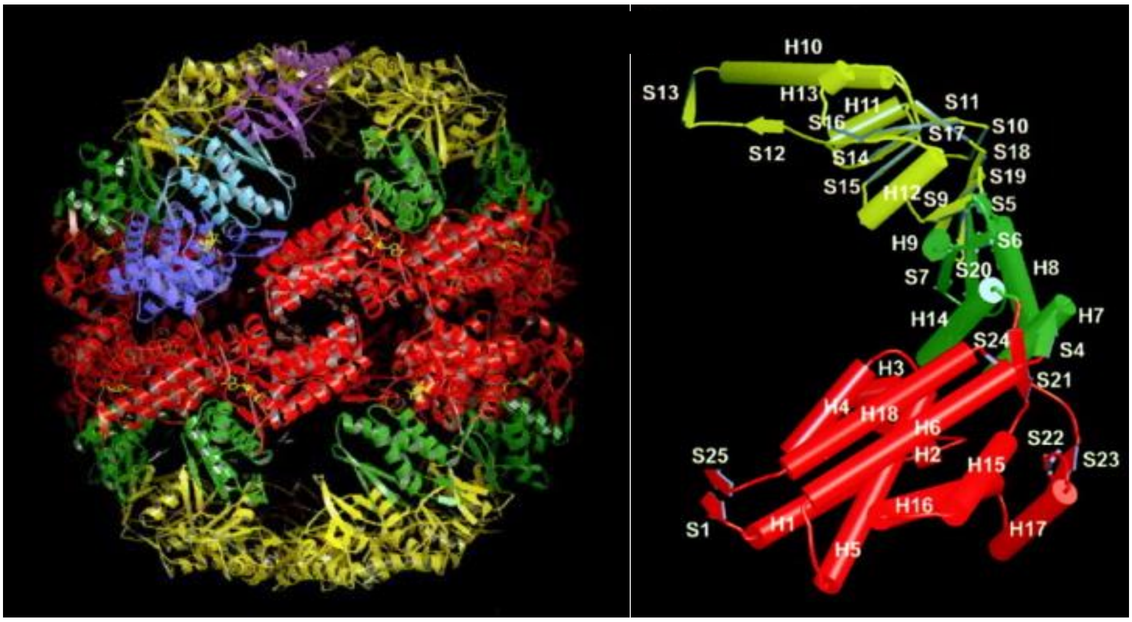


Figure 2-3: Crystal structure of the thermosome of T. acidophilum. Full closed structure (left) and labelled secondary structure of the alpha subunit (right). Adapted from [9]

Previous attempts to visualise the *T. acidophilum* thermosome in an open conformation failed, due to the formation of filaments in alpha-beta mixtures. It was found, however, that a solution of only alpha subunits was able to assemble into cylindrical oligomers. Low resolution (~30 Å) maps were successfully created for these all-alpha thermosomes in a mixture of open and closed conformations (Figure 2-4) [17]. The top panels of Figure 2-4 show a 3D representation of the open structure. On the right are several experimental views and reprojections calculated from the map. From left to right are a top view, several intermediates and a full side view. The lower panels show the same for the closed conformation, note that all such views look very similar while there is a clear difference between the different views for the open. It is further noted, as in previous studies, that thermosomes in an open conformation tend to orient themselves preferentially as top views within the vitreous ice. This is suggested to be a result of hydrophobicity of the lid region. In this study, detergent was added to reduce this effect while the Mm-cpn work discussed above resorted to removing this region entirely [14, 17].

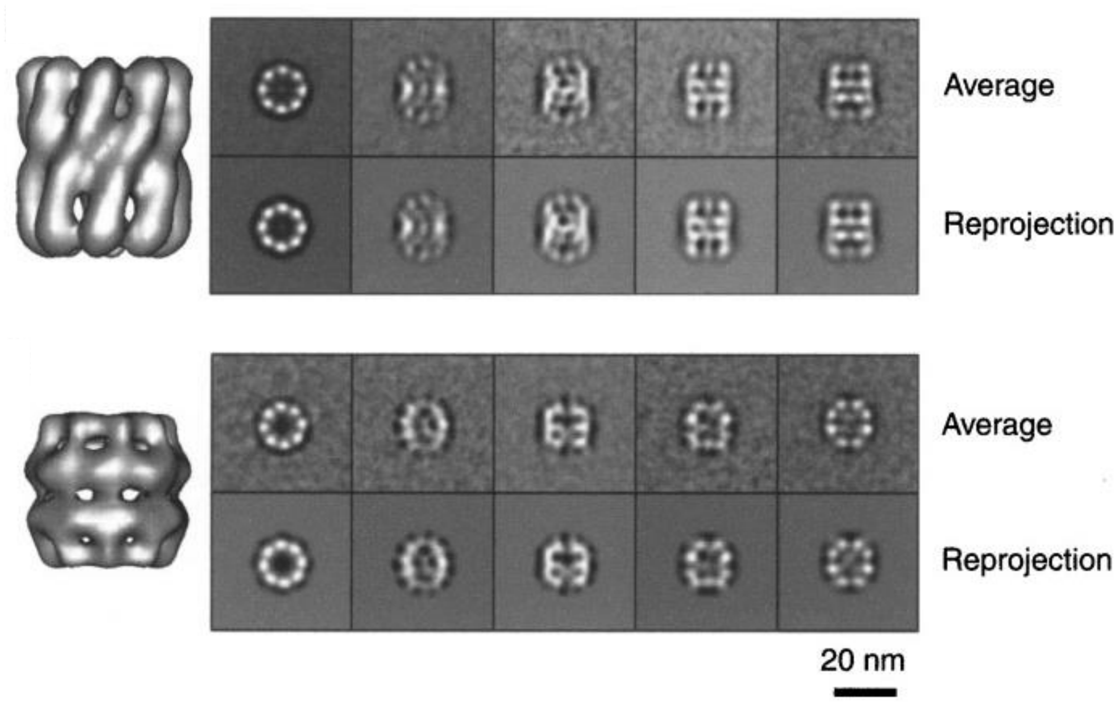


Figure 2-4: Low resolution 3D models of all alpha thermosomes in open (top) and closed (bottom) states. Also shown are class averages, and reprojections of different views over a 90º rotation from full top (left) to full side (right). Adapted from [17].

To our knowledge, at the time of writing no further high-resolution 3D structures have been published of this system. There are however several studies focussing on its ATP-ase and folding activities [18-20]. In 2016 it was shown that the yield of thermosomes purified from *E.coli* could be increased by 4 times by addition of a His-tag within an external loop of the alpha subunit [21]. This facilitated purification and also provides potential for use in distinguishing between the different subunits in 3D structures.

Using this system, a 2018 report [22] demonstrated a panel of mutations in which either ATP binding or ATP hydrolysis were blocked in one of the two subunits at a time. The position of one such mutation in the alpha subunit is shown in Figure 2-5 (top). It was shown that even when hydrolysis was blocked in one subunit, it was still able to occur in the other (bottom). This differing behaviour of subunits towards ATP draws an important analogy with eukaryotic CCT, in which varying strengths of ATP activity are found around the different subunits in a ring.

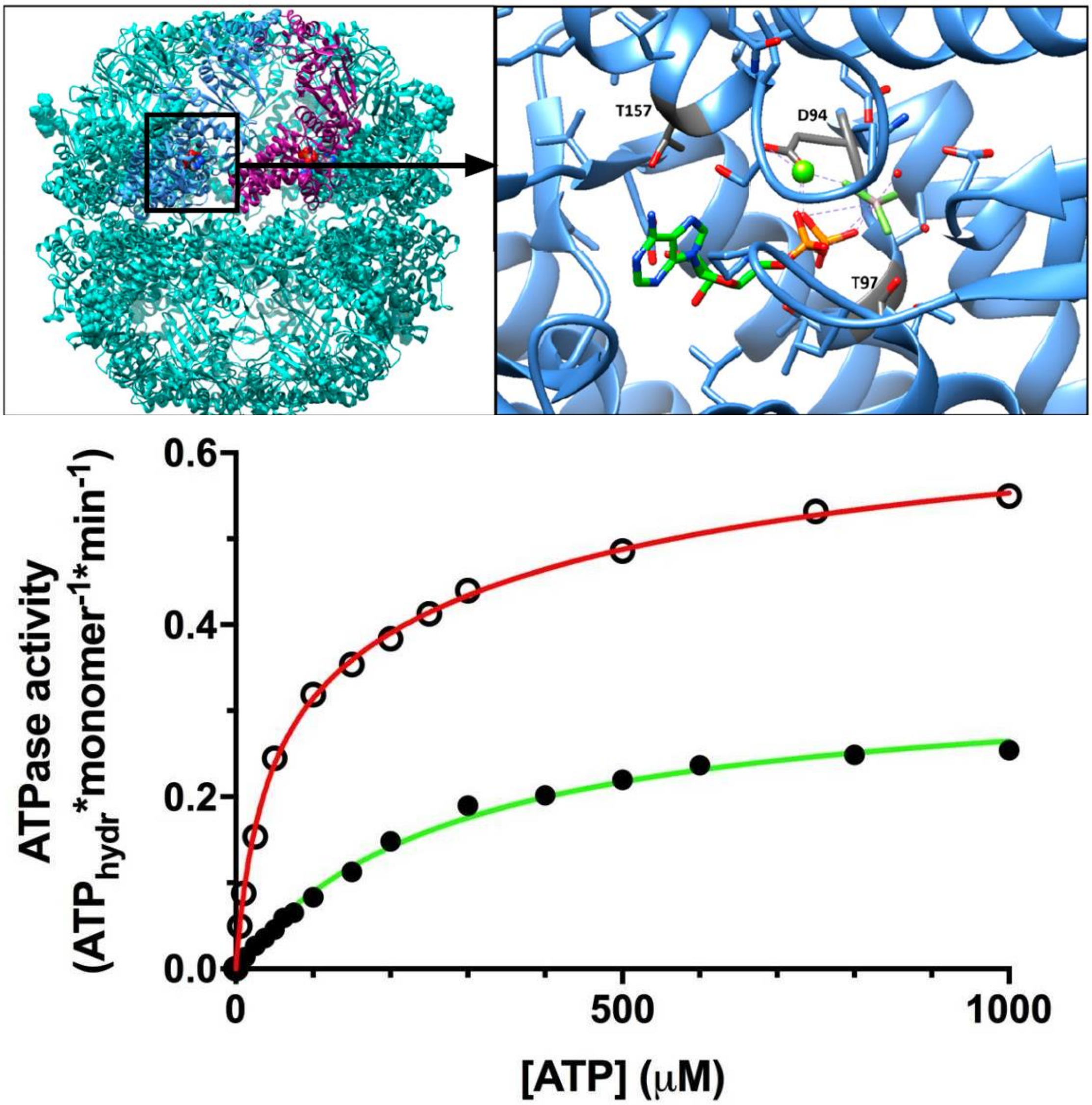


Figure 2-5: Mutation in nucleotide binding pocket of alpha subunit. Full modelled structure on the left shows location of binding pocket and additional His-tags. Right hand image shows location of D94 residue involved in ATP hydrolysis. (Bottom) Steady-state ATP-ase activity of wild type (red) fitted as a two-phase transition and mutant (green) fitted as a one-phase transition. Adapted from [22]

Combining these biochemical results with structural analysis like that reported for homo-oligomeric thermosomes should yield further insight into the role of subunit heterogeneity in type II chaperonins.

2.3 Chapter Aims

The following chapters of this thesis focus on two distinct thermosome systems. The first being the wild type with ATP-AlF_x, representing the hydrolysed state with bound hydrolysis product. The second is a mutant in which a single residue in the nucleotide binding pocket of the beta subunit (Asp 93) has been replaced with a lysine in order to abolish ATP hydrolysis. This is the beta subunit equivalent of the mutation illustrated in Figure 2.5. This residue was chosen based on its structural homologue in GroEL which has been found to be required for ATP-hydrolysis [23]. The equivalent residue in GroEL is responsible for coordinating the Mg^{2+} ion which in turn coordinates the beta and gamma phosphates of ATP. The equivalent D87K mutation in GroEL has been used extensively as a ‘trap’ for unfolded polypeptide as it is able to bind but not refold the substrate. Steady state ATP-ase experiments on the D93K mutant revealed monophasic behaviour with respect to ATP hydrolysis, compared to the biphasic behaviour of the wild type thermosome (Figure 2-5 (bottom) [20]. The latter had been assigned to the presence of two separate sites with different characteristics of ATP binding and hydrolysis (interpreted as either the alpha and beta subunits, or the two facing rings). In the mutant, ATP is only hydrolysed in half of the subunits, while still being permitted in the others. Again, it is unknown whether this is only the alpha subunits or only one of the rings. We may therefore expect to capture a pre-hydrolysis state, analogous to that studied in Mm-cpn [16], in the beta subunit, or a post-hydrolysis state of whichever portion of the thermosome is able to hydrolyse ATP. As with the wild type, ATP-AlF_x is added to capture the hydrolysed state with bound hydrolysis product.

This chapter discusses the collection and processing of cryo-TEM data with the aim of reaching sufficiently high resolution to distinguish between subunits. Models are then fitted to these maps in the chapter that follows, and conclusions are subsequently drawn.

2.4 Materials and Methods

2.4.1 Protein expression and purification

Proteins were expressed and purified by Dr Giulia Bigotti according to the protocol described in [21]. An overview of the purification procedure for the His-tagged thermosome is shown in Figure 2-6. Note that in the figure WT refers to un-tagged thermosome while Ht refers to the His-tagged thermosome, referred to as wild type in the remainder of this thesis. The mutant thermosome was purified in the same way.

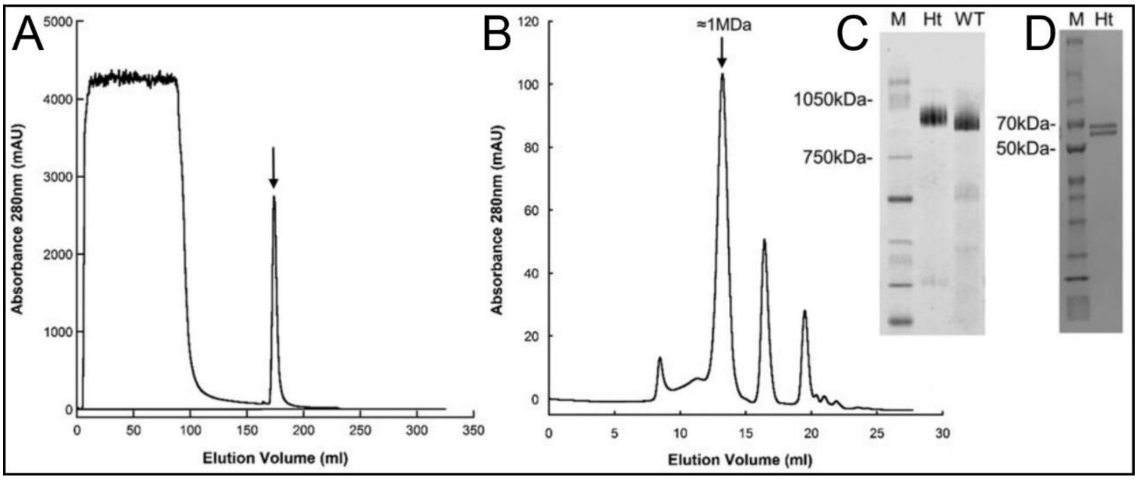


Figure 2-6: Two step purification of thermosomes. (A) chromatogram of nickel affinity column elute showing sharp peak of His-tagged thermosome. (B) Size exclusion chromatography separating 1MDa thermosome complex. (C) His-tagged thermosome shows lowered electrophoretic mobility than un-tagged thermosome. (D) SDS-PAGE separates bands corresponding to alpha and beta subunits in a 1:1 ratio. Adapted from [21]

Prior to negative stain or cryo-TEM grid preparation a solution of AlF_3 was prepared at 20mM and added to the protein solution to a final concentration of 10mM. This was then incubated at 55ºC with 500 µM ATP for 30 minutes to allow for any possible hydrolysis.

2.4.2 Negative Stain

Negative stain TEM grids were prepared by Dr Fabienne Beuron at ICR as described in [21].

Quantifoil grids with a layer of continuous carbon were glow discharged for 20s to make them hydrophilic. The sample was applied and negatively stained with

2% uranyl acetate. Imaging was performed on a FEI Tecnai F20 electron microscope at 200 kV and with a 4K × 4K F416 TVIPS CMOS detector.

2.4.3 Plunge Freezing

Grids were prepared using different plunge freezers and thus conditions were not identical between samples. Vitrobot conditions were based on those in [21]. Wild type conditions were developed to mimic these as closely as possible.

	Wild Type	Mutant
Prepared by	The Author	Dr Fabienne Beuron
Plunge freezer	Leica GP	Vitrobot Mark IV
Chamber conditions	18 ºC 95% humidity	18 ºC 100% humidity
Wait time	2 secs	20 secs
Blot time	1.2s blot	4 secs
Drain time	1 sec	1 sec
Volume	3 µl	2 µl
Grid type	Copper 300 mesh 2/2 Quantifoil	Gold 300 mesh, 1.2/1.3 µm Quantifoil, Continuous carbon

Table 2-1: Plunge freezing conditions for wild type and mutant grids

2.4.4 Data Collection

Details of image collection protocols are described in Table 2-2.

	Wild Type	Mutant
Microscope	Titan Krios II at eBic	Talos Arctica at GW4 Cryo-TEM Facility – Bristol University
Camera Type	K2	K2
Dose	46.6 $e^-/Å^2$	45.4 $e^-/Å^2$
Defocus	-1, -1.43, -1.86, -2.28, -2.7	-1, -1.43, -1.86, -2.28, -2.7
Pixel size	1.048 Å	1.06 Å
Number of frames	38	25
Number of micrographs	3335	1194
Microscope help from	Dr Yuriy Chaban	Dr Andy Richardson

Table 2-2: Data Collection details for wild type and mutant datasets.

2.4.5 Data Analysis

Unless specifically noted in the text, all analysis was performed using RELION version 2.1. Other software accessed through the RELION workflow were Unblur [24], and CTFFind-4.1 [25].

2.4.5.1 Pre-processing

As described in section 1.1.3, recent improvements in cameras have resulted in the recording of movies rather than single micrographs. This allows for the correction of drift, and down-weighting of later frames where the sample has received higher amounts of radiation. Unblur aligns movies by cross correlating an individual frame with an average calculated from the other frames, thus calculating a shift for each frame. The overall trajectory is then smoothed by fitting to a spline or polynomial. This process is repeated iteratively until the shifts fall below a certain predefined value [26].

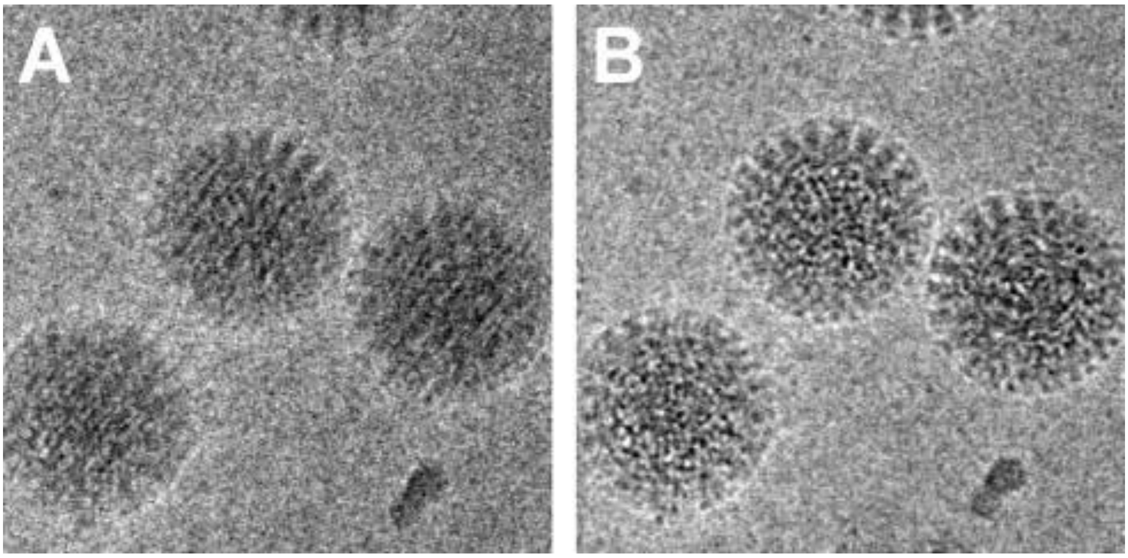


Figure 2-7: Motion correction of movie revealing details which were not visible in the original. (A) Average of 60 unaligned movie frames. (B) The same 60 frames after alignment. Reproduced from [27].

The datasets in this thesis were motion corrected, applying dose weighting according to the total dose and number of frames listed in Table 2-2. CTF correction was performed by CTFFIND-4.1 using the dose-weighted, motion corrected micrographs. See section 1.2 for theoretical background to CTF correction.

2.4.5.2 2D Classification

RELION 2.1 uses a reference-free maximum likelihood method to average and align 2D images. Firstly, the particles are randomly separated into a user defined number of groups and averaged together. Each particle is then compared with each averaged group and given a probability weighting for each one. Finally, new averages are calculated using the weighted averages, *i.e*. a single particle may contribute to multiple classes but with different weights. This process is repeated over many iterations after which particles have generally converged into a single class [28].

In this study, all 2D classifications were performed over 25 iterations and outputs monitored to ensure that convergence had been reached. The number of classes varied between 10 and 100 depending on the number of particles being classified.

2.4.5.3 3D Classification

Similar to 2D classification, 3D classification is performed using a maximum likelihood method. The key difference is in the generation of the initial averages. After splitting the dataset into a user defined number of groups, each particle is refined against the initial model which is low pass filtered to avoid fitting to high resolution noise. This results in multiple differing models forming the basis for the initial classification. Once again, each particle is compared to each reference with user defined angular sampling and given probabilities for each. This is repeated to convergence [29].

All 3D classifications were performed over 25 iterations with 4 classes with the exception of the initial wild type classifications which had 6 classes due to the large dataset size. All references were low pass filtered to 60 Å and sampling performed at RELION 2.1 default values.

2.4.5.4 3D Auto refinement

The dataset is randomly split into two halves and the angles and translations refined according to a reference map provided – usually an output from a 3D classification. The two halves are iteratively improved and compared to ensure that they do not diverge from each other. This gives assurance that the final map is the global minimum rather than a local minimum. A mask may be added to this step (or 3D classification) such that alignment is focussed on the protein density of interest. In the following chapter, all reference maps were taken from a previous round of 3D classification and were filtered to a resolution of 60 Å to minimise reference bias. If necessary, the handedness of the map was flipped in agreement with the crystal structure.

2.4.5.5 Post Processing

In this step a Fourier Shell Correlation (FSC) between the two maps formed during 3D auto refinement is performed, allowing gold standard resolution estimation to be achieved [30]. The loss of high-resolution information is corrected by the calculation of a global temperature, or B-factor, which describes the drop off in contrast with increasing resolution. The calculated global B-factor is then used to sharpen this map to correct for this effect [31].

2.4.5.6 Local Resolution

Local resolution estimations were performed using RELION’s own implementation, in which post processing style calculations are performed in conjunction with a mask. The mask is moved over the whole map, resulting in local estimates of resolution for each region [32]. This outputs a local resolution map which contains a 3D array of local resolution estimates. This is then used to apply surface colour to the filtered map using Chimera.

2.5 Results and Discussion

2.5.1 Negative Stain

To check the suitability of the thermosome preparations for a single particle cryo-TEM experiment, they were first screened by negative stain TEM. In this way appropriate conditions for the protein concentration and buffer were checked rapidly prior to large scale data collection in cryo-TEM. Sample preparation was performed by Dr Giulia Bigotti following the protocol discussed in [21], and as such will not be discussed here. This section will discuss how negative stain data were analysed in order to confirm suitability for structural studies.

2.5.1.1 Wild Type

543 particles were picked manually from 5 micrographs to create a reference for autopicking. These particles were sorted into 10 classes as shown in Figure 2-8 in which the classes have been sorted from highest to lowest number of particles per class. The class highlighted by the red box contained 43.6% of the particles and was the only side view. For this reason, it was selected as the reference for autopicking. While in general one would use multiple references for autopicking, the relatively low resolution of the negative stain images compared to cryo images meant that this reference was sufficient. Furthermore, the closed conformation of this thermosome results in an approximately round side projection which is similar enough to the top views for autopicking all particles.

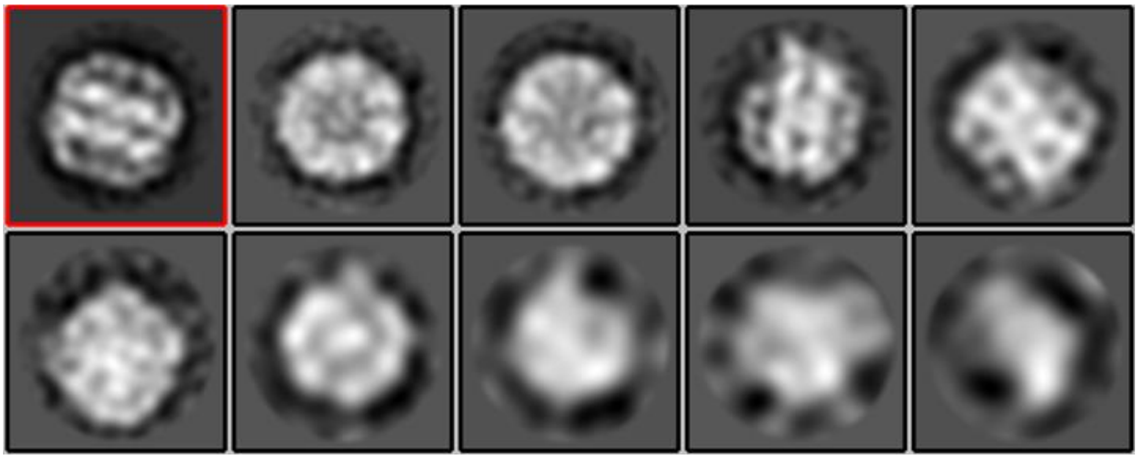


Figure 2-8: Class averages of manually picked, negatively stained, wild type thermosome. Ordered by number of particles, highest (top left) to lowest (bottom right). Red box indicates class used as reference for autopicking.

Autopicking all 133 micrographs resulted in a dataset of approximately 72 000 particles. This was reduced to 60 000 by manually checking the images and removing false positives, such as those sitting on the Quantifoil film or those in aggregates. An example of this process for a single micrograph is shown in Figure 2-9 which shows the raw micrograph (A), autopicked particles highlighted by green circles (B) and finally the remaining particles after manual removal of aggregates and false positives (C).

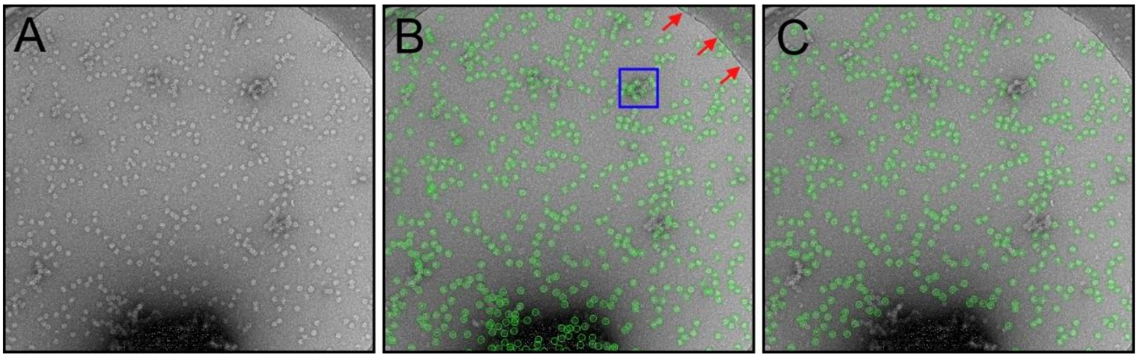


Figure 2-9: Negative stain of wild type Thermosome. (A) raw micrograph, (B) autopicked particles circled in green, aggregate indicated by blue box, red arrows show edge of Quantifoil film, (C) particles after manual removal of false positives.

Following this, the dataset was further cleaned by subjecting it to two rounds of 2D classification as shown in Figure 2-10. These classes represent the full dataset with a variety of views, including multiple different side views. The best classes from the first classification are highlighted in red (approx. 44 000 particles) and were subjected to a second round of classification. In this way the poor particles which were not removed manually, *e.g*. those which are too close to others (blue boxes) and non-particles (yellow boxes) were still removed. Classes which are not highlighted with a coloured box contained particles which did not look like predicted projections based on those shown in Figure 2-4. These classes did not represent a significant number of particles and were also disregarded.

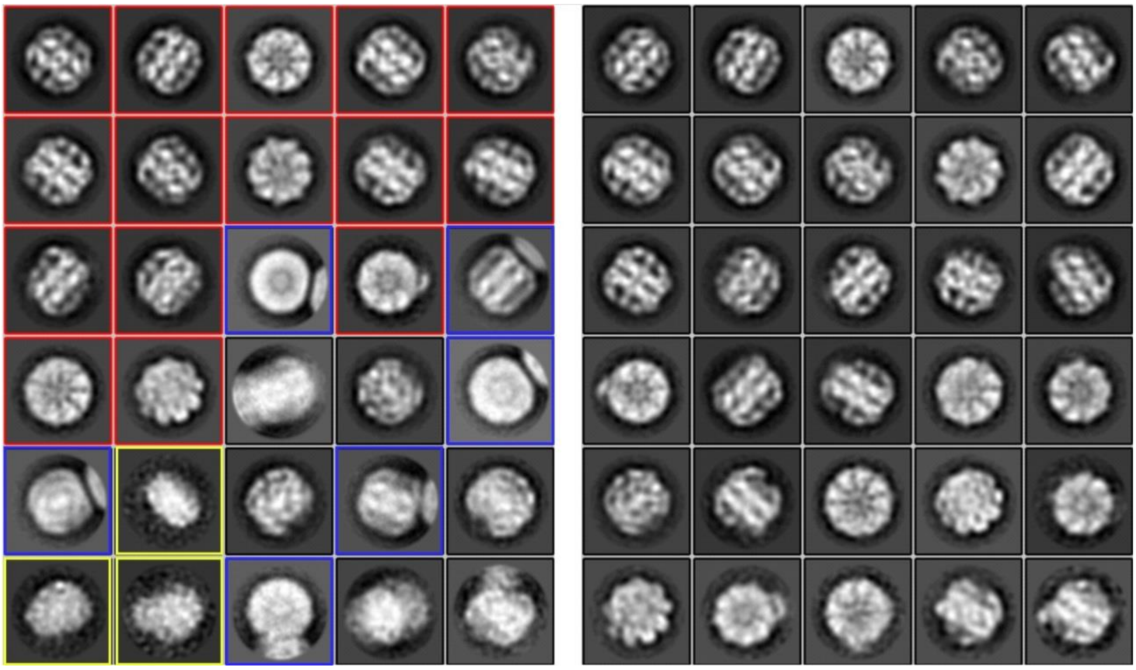


Figure 2-10: 2D classification of full negative stain wild type dataset. Classes sorted in order of particle numbers. Left: 1st classification of full data set, red boxes indicate classes selected for second round of 2D classification (right). Blue boxes show particles which are too close to another, yellow boxes show non-particles.

From the second round of 2D classification it is clear there is a good range of top and side views and that the population is reasonably homogeneous. As these data were collected for sample screening for cryo-TEM, no further analysis was performed. However, if this were a new sample for which no prior structure was known, more extensive 2D classification would be performed to remove further poor particles. An initial 3D model would usually then be created at this stage.

2.5.1.2 Mutant

A full dataset of negative stain mutant thermosome was not collected. A few micrographs were taken in order to ascertain appropriate conditions such as concentration. The negative stain images collected were of a similar quality to those shown for the wild type, so the decision was made to carry on with a full cryo-TEM experiment for this sample.

2.5.2 Plunge Freezing and Data collection

Freezing conditions are detailed in section 2.4.3 above. The mutant sample was frozen and screened at ICR by Dr Fabienne Beuron on a Vitrobot mark IV according to protocol published in [21]. The wild type was prepared in Bristol on a Leica GP. It was necessary to develop a new freezing protocol when using the Leica GP and alternative grid types.

Data collection was also performed on different microscopes for the two samples. Settings such as dose, defocus and pixel size were kept as similar as possible between the two.

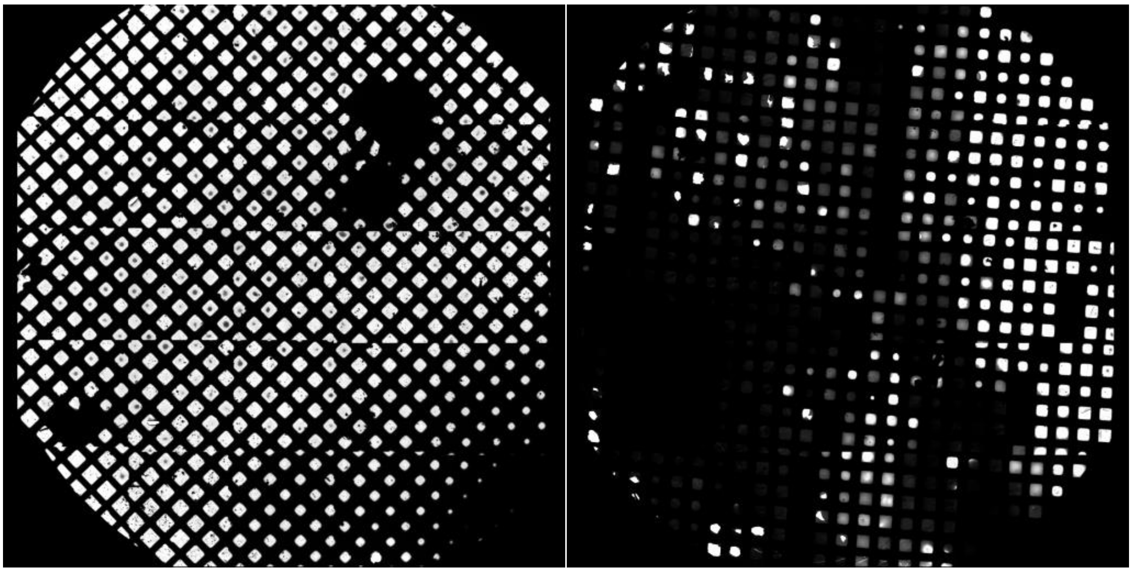


Figure 2-11: Grid atlases for Wild type (left) and mutant (right) grids.

Figure 2-11 shows atlases of example grids for the two samples. The mutant grids (right) showed a gradient of ice thickness to a greater extent than the wild type (left). This is common for grids frozen using the Vitrobot as there is a more pronounced gradient in blotting pressure due to the blotting mechanism.

It was noticed during grid screening that there was a preference for top views on the wild type grid. We attempted tilting the grid by 30º in order to get a wider range of sample orientations but ultimately decided to collect without tilt as it did not appear to provide a significant improvement to the range of views.

2.5.3 Pre-processing

Following data collection, both datasets were submitted to the standard pre-processing steps of motion correction and CTF correction using Unblur and CTFFIND-4.1 respectively. The choice of software was based on efficient use of the available computational power of the CPU cluster.

Figure 2-12 shows some examples of frame trajectories (top) from the wild type dataset along with their motion corrected micrographs (middle) and power spectra showing fitted contrast transfer functions in the upper left quadrants (bottom). The trajectories are plotted to the same scale and are therefore comparable between the three. The starting point for each is indicated by a red dot. In each case, there is a tendency towards large shifts in the first few frames indicating high sample motion when first illuminated. (A) and (B) represent low and high defocus values respectively. Higher contrast is visible in the motion corrected image at higher defocus, and the positions of the Thon rings in the power spectra. The lower defocus images have less contrast, making particle picking more difficult, but the Thon rings can be fitted to higher resolution, indicating a greater information content for these images.

(C) shows a micrograph with poor quality ice, as demonstrated by the dappled quality of the micrograph itself, and the strong ring at ~3.6 Å, also known as a water ring. While this is most pronounced in (C) due to the presence of crystalline ice, the water ring is also present in (A) and (B) as a result of vitreous ice being highly sensitive to beam damage [33].

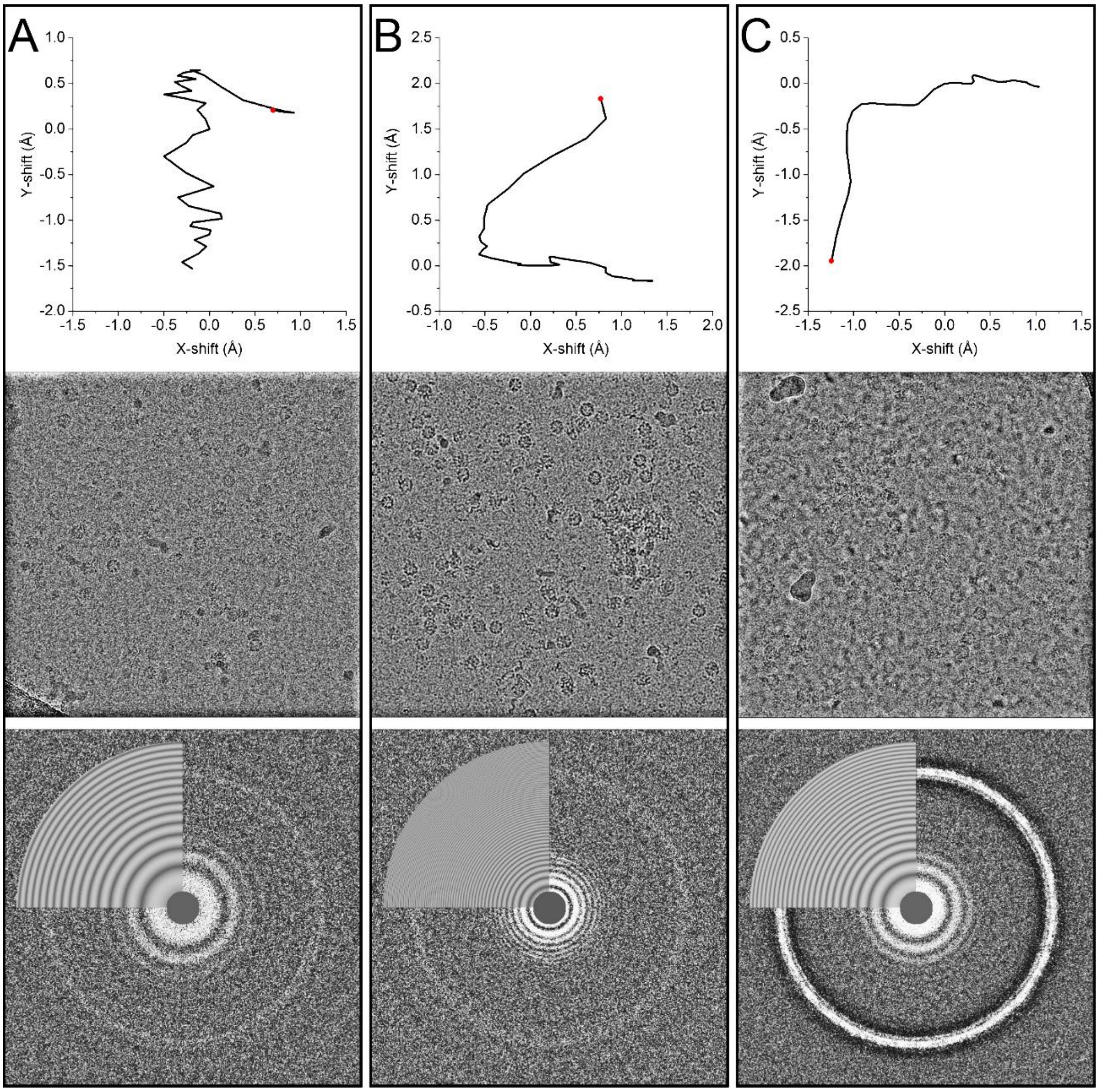


Figure 2-12: Pre-processing of micrographs taken at low defocus (A), high defocus (B) and an area of poor ice (C). Top: Motion of each frame calculated by Unblur. Red dot indicates first frame. Middle: Motion corrected micrographs. Bottom: Power spectra and fitted contrast transfer functions calculated by CTFFIND.

Figure 2-13 shows a plot of estimated defocus vs. the maximum resolution which CTFFind estimates based on how far from the centre the Thon rings can be fitted [34]. The data show clear groupings of defocus values, corresponding to the five defocus values requested while setting up the image acquisition (Table 2-2). There is a slight positive correlation between the values, which is as expected. It should be noted that there was a single outlier which is not shown on this plot for clarity, and the corresponding micrograph was not included in further processing. While the estimated resolution values serve as a useful indication of the quality of data being collected, they should be treated with caution. As discussed in the CTFFind documentation [35], this estimate is prone to error, particularly where

Thon rings drop off quickly but are still present in the water ring which appears to be the case in some of our data (*e.g*. Figure 2-12(A)).

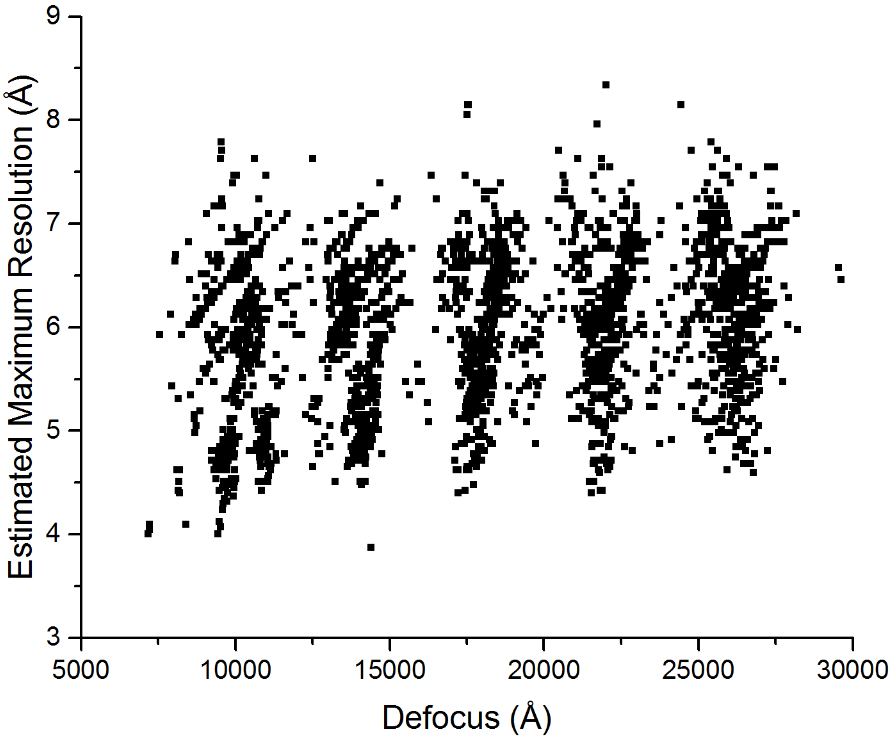


Figure 2-13: Defocus vs Maximum resolution of wild type micrographs as calculated by CTFFind.

At this point, any micrographs with strong water rings were discarded, leaving datasets of 3280 and 1047 micrographs for the wild type and mutant samples respectively.

2.5.4 Particle Picking and 2D Classification

For each dataset, an initial set of particles was manually picked from a subset of micrographs. These were subjected to 2D classification in order to select references for autopicking of the dataset.

	Wild Type	Mutant
# Manually picked	1108	1008
# Autopicked	250 000	35 000
# after 2D classification	173 000	21 000

Table 2-3: Numbers of particles during 2D selection.

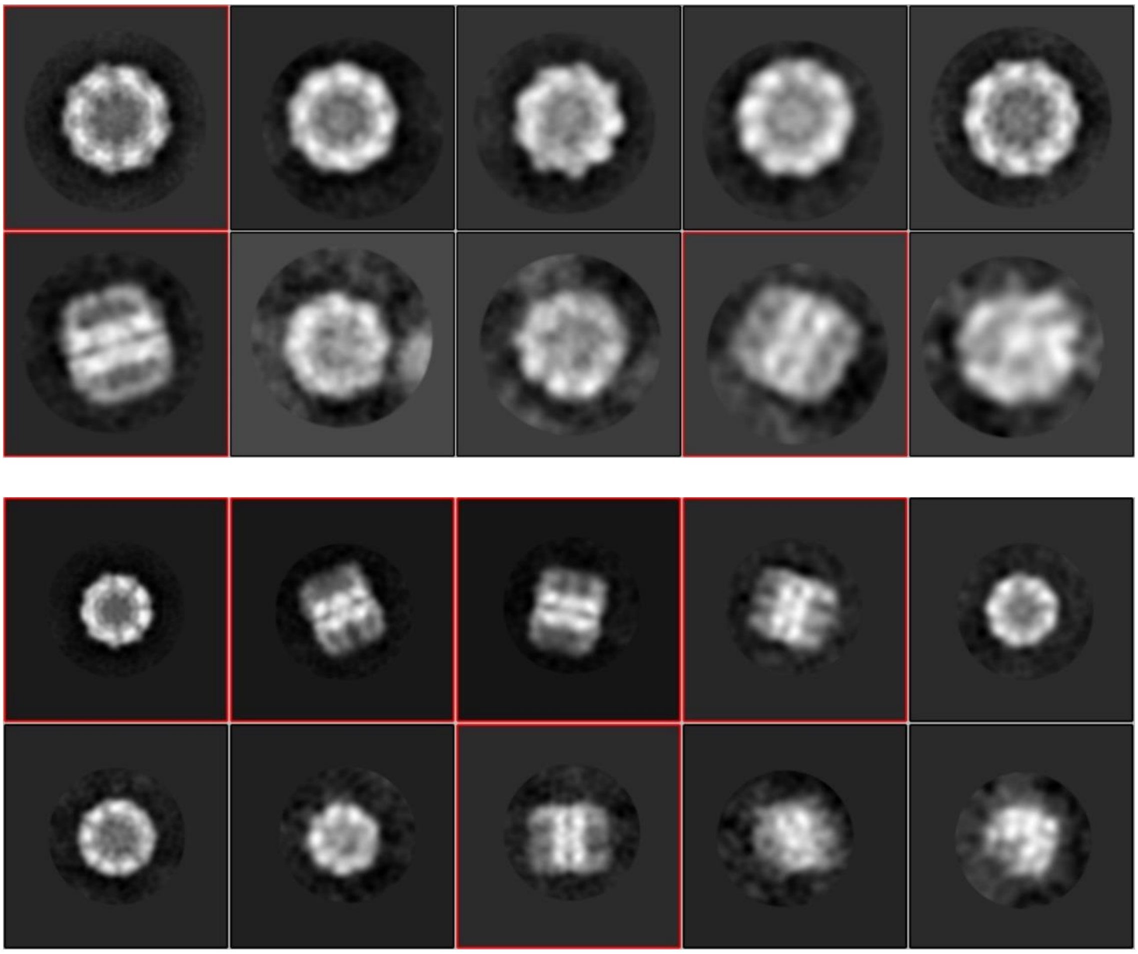


Figure 2-14: 2D classification of manually picked particles from the wild type (top) and mutant (bottom) datasets. Red boxes indicate classes selected as references for autopicking. Box sizes: wild type 300 Å, mutant 450 Å.

The results of classification of the manually picked datasets are shown in Figure 2-14. For the wild type, only two classes showed side views while the mutant gave the same number of side and top views. In both cases, all side views and only the largest top view classes were selected as templates for autopicking. Mutant particles were initially extracted in 450 Å boxes as seen in the figure but this was reduced to 300 Å at the following stage in line with the wild type.

Since the mutant dataset was relatively small, most poor particles were manually removed by checking each micrograph and deselecting any which were false positives. Those which were too close to other particles, neighbouring or sitting on top of the Quantifoil film, ice contamination *etc*. were removed leaving 23 000 particles. The large size of the wild type dataset made this type of cleaning unfeasible so this manual removal step was not performed.

Both datasets were subjected to several rounds of 2D classification to remove any further poor particles and to assess the quality of the data. This left 173 000 wild type and 21 000 mutant particles to be taken forward to 3D processing.

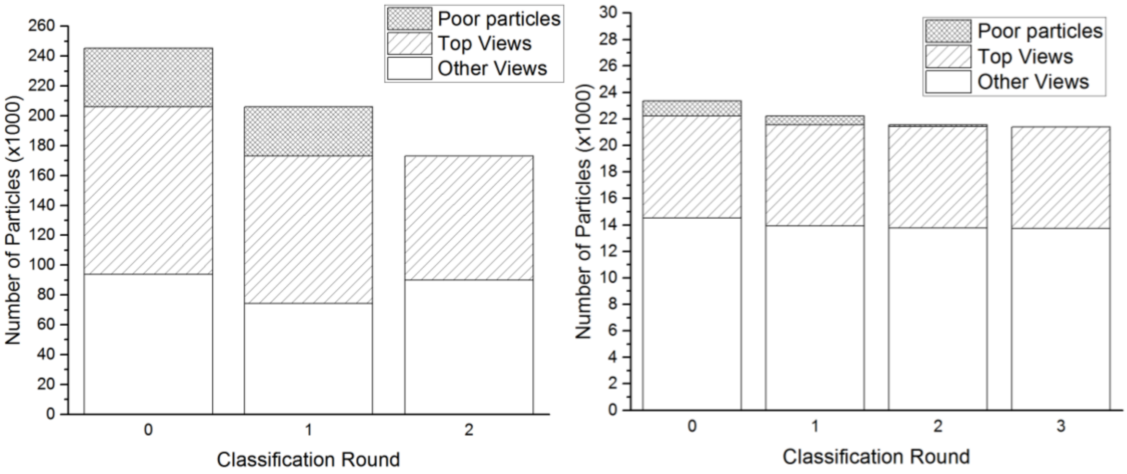


Figure 2-15: Range of particle views in wild type (left) and mutant (right) datasets during 2D classification. Round 0 indicates particles extracted following autopicking (and manual deselection of poor particles for mutant).

Particle counts for each step are shown in Figure 2-15, separated into top views, other views which were kept for the next round of classification, and discarded particles. The final wild type dataset had approximately equal numbers of top and non-top views, while the mutant contained around twice as many side views as top views.

It appears from this analysis that the number of 'other' views has increased between rounds 1 and 2 for the wild type. This is likely due to miss-classification in the earlier rounds where side or intermediate views may have been incorrectly placed in top-view classes.

The most populous 15 classes of each dataset are shown in Figure 2-16 for comparison. Close inspection of the side views reveals that the mutant particles are in a more open conformation filling a larger proportion of the unmasked area. They appear to be wider at the top than the centre while the wild type is narrower at the ends than the centre.

This indicates that the mutant is likely to be in a more open state than the wild type. This is quantified in (C-D) in which similar side views are shown with measurements of height and widths at the equatorial and apical regions. There is a 20-25% increase in both the height and apical width for the mutant compared to the wild type, suggesting that it is indeed in a more open conformation. There is also a modest increase in equatorial diameter of 7%, suggesting that the motion is not limited to the apical regions.

It is clear from the 2D classifications of the wild type dataset that the thermosome has a preferential orientation within the ice. This effect was mitigated by the presence of a continuous carbon film on the mutant grid, which resulted in a much higher proportion of side views in the mutant. A further result of the additional carbon film is that intermediate views are no longer seen for the mutant while they remain present in the wild type dataset. Such classes are highlighted by white circles in Figure 2-16.

Having a good range of side views is vital for successful 3D reconstruction. However, since the wild type data set is considerably larger than the mutant, the lower proportion of side views available in this dataset does not pose a problem in terms of reaching comparable resolutions for the two datasets.

A final observation is that both datasets contained classes which appear to have 9-fold rather than 8-fold symmetry (not shown as these were smaller classes). While it is generally easy to distinguish these in the top views, it is more difficult for the side views and intermediate angles. It was therefore decided that any such classes would be taken through to the 3D classification stage and removed at that point. See section 2.5.6 for this analysis.

It would be possible to continue removing particles by 2D classification by repeating this process many more times. However, this yields diminishing returns as, in general, fewer particles are removed at each iteration. It was therefore decided to continue to the 3D classification stage and remove any remaining poor particles at that stage.

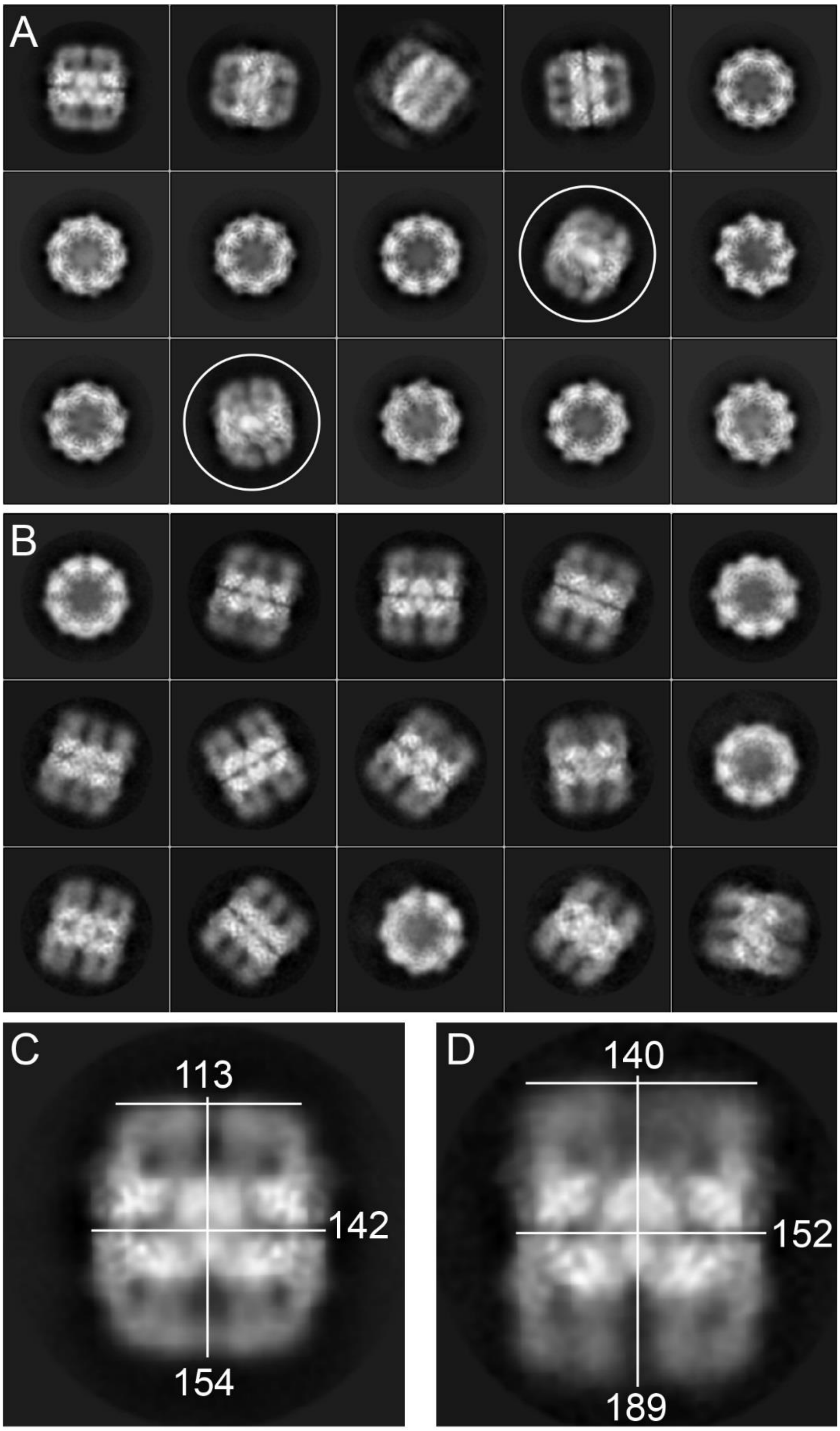


Figure 2-16: Comparison of 2D class averages. 15 largest classes of wild type (A) and mutant (B) particles. Measurements of side view classes of wild type (C) and mutant (D) (in Å). White circles indicate intermediate views.

2.5.5 Initial 3D map

Initial 3D maps were created using a subset of particles from 2D classes representing a range of different views of the particle. In order to minimise bias, no external reference was used, nor was any symmetry applied. Initial maps for each of the datasets are shown in Figure 2-17. In both cases a rough barrel-shaped structure is formed with protrusions corresponding to the apical domains. Comparison with the previously solved crystal structure [9] gives assurance that these maps are appropriate for use as references for 3D classification following filtering to avoid fitting to high resolution features.

The wild type map is in a closed conformation similar to that of the crystal structure while, interestingly, even from these initial 3D maps the mutant appears to be in a more open state. Both maps display a somewhat 'confused' chirality, with the helicity being inconsistent between some subunits. The overall handedness of a particle cannot be determined by TEM without knowledge of relative angles, *e.g*. by collecting images in tilt-pairs. Therefore, throughout the following analysis, maps are shown with both possible chiralities which are only corrected at the final stages.

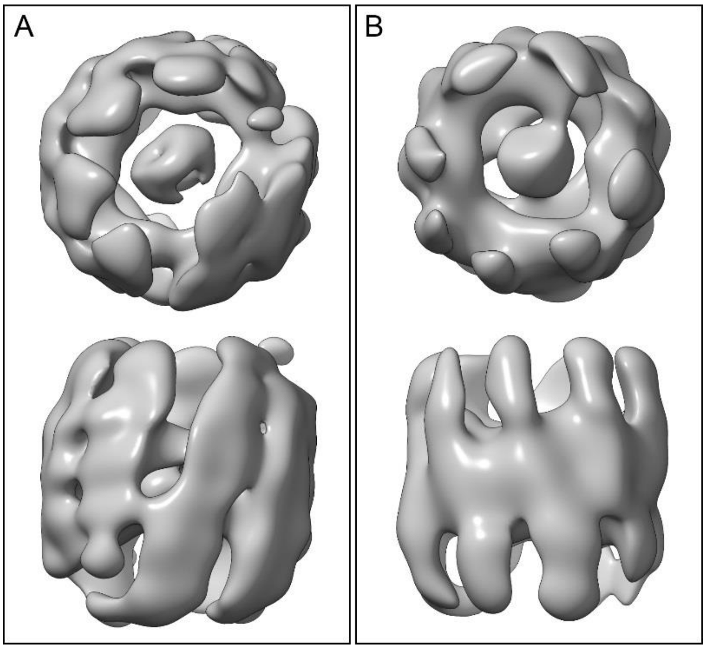


Figure 2-17: Top and side views of initial 3D maps: (A) wild type, (B) mutant.

2.5.6 3D Classification – Identification of 16mers vs. 18mers

Once an initial map had been calculated, 3D classes were generated and low pass filtered to 60 Å to use as references. For the smaller mutant dataset, the particles were classified into 4 classes while for the larger wild type dataset 6 classes were initially used. In each case, there was at least one map which was a clear 16mer but also one which was a clear 18mer. These nine-fold symmetrical particles have been previously reported in recombinantly produced alpha-only thermosomes, and although it is possible that they are physiologically significant, investigating this was beyond the scope of this study [40]. It was therefore necessary to remove these 18 subunit particles from the dataset in order to continue with the structural analysis of the 16mer.

2.5.6.1 Wild type

Since the wild type dataset had not been as well 'cleaned' during the picking and 2D classification stages, 3D classification was also used to remove a significant number of poor particles. The initial round of 3D classification (Figure 2-18) revealed many particles which were very close to the edge of the hole in the Quantifoil film. This can be seen in the 1st and 3rd classes which clearly show a flat disc to the side of the particle. Following their removal, the subsequent round of classification gave only one good 16mer class (indicated by a red box) so only the particles in this class were taken forward to a 3rd round. The 23 000 particles contributing to a good 18mer class (5th) were also discarded at this stage. Finally, a third round again gave only one good class containing the majority of particles (red box), and several very poor ones. This single class was taken forward for further processing giving a final 40 237 particles for further processing.

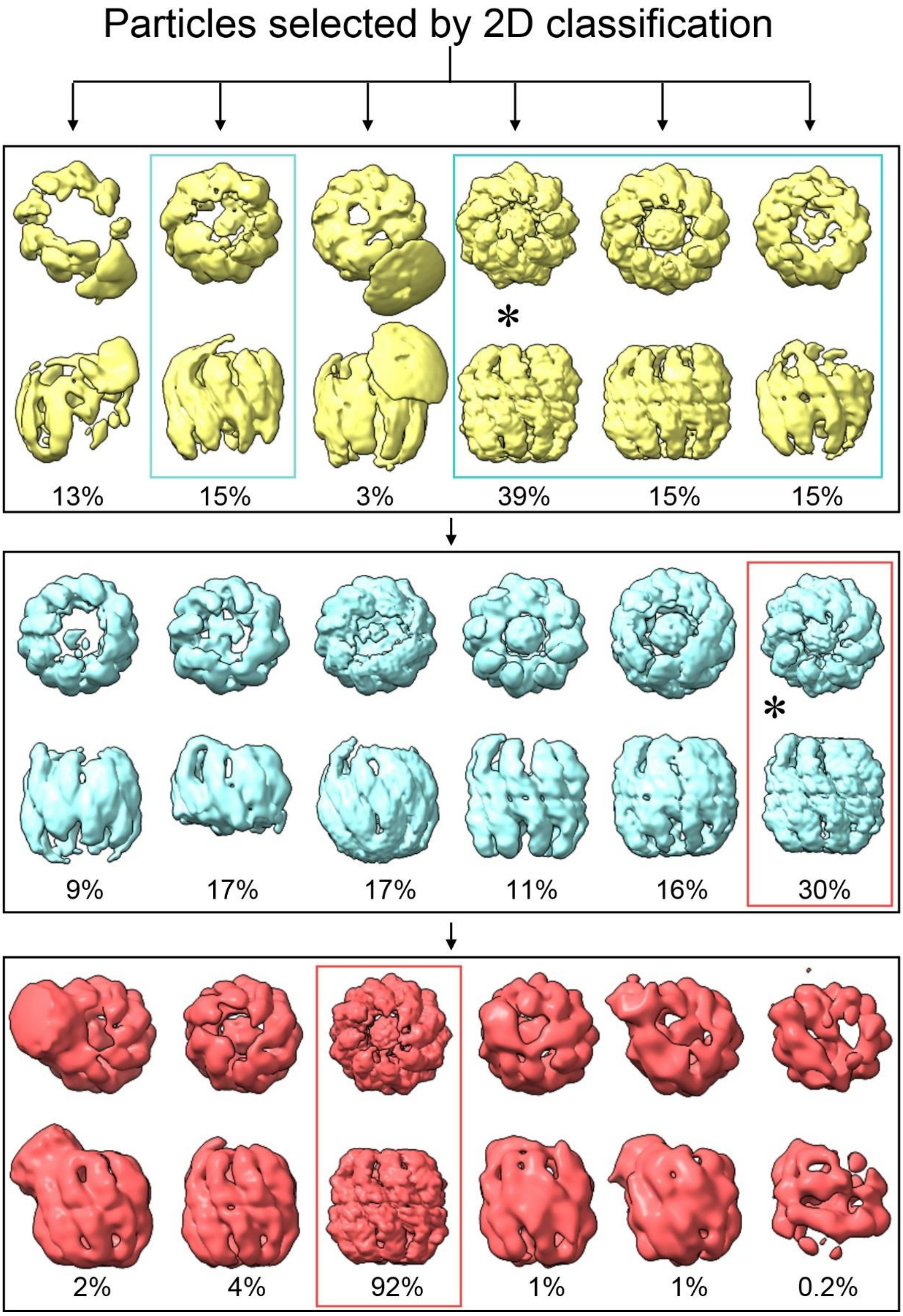


Figure 2-18: WT dataset cleaning by sequential 3D classifications. Black asterisks indicate the map used as a reference for the subsequent round of classification. Boxes indicate particles which were taken forward to the subsequent round. Numbers below maps indicate proportion of particles included in each class.

2.5.6.2 Mutant

Initial 3D classification of the dataset is shown in Figure 2-19(A). Each of the 4 initial classes (left) was reclassified using the best 16mer as a reference map (centre) and again using the best 18mer as a reference (A-right). This verified that any particles originally classified as 16mers were indeed reclassified as such and similarly for 18mers. At the end of this process particles indicated by red boxes were taken forward as good 16mers while blue boxes indicate good 18mers. Since these selections came from different initial classifications, it is not possible that any particles were selected as both 16mers and 18mers.

Any particles contributing to classes which did not have a clear structure were discarded and not used for any further analysis. The choice of 16mer classes was relatively forgiving at this stage due to the small number of particles in some of the classes. Thus, any with visible 8-fold rotational symmetry were selected.

Finally, one further round of 3D classification was performed to more strictly remove any remaining poor particles contained within the selected classes (B). This resulted in a final dataset of 10 307 16mer particles which were used for all subsequent processing. A similar final classification of the 18mer particles is included for comparison(C) and showed a final count of 2074 good 18mers.

2.5.6.3 Conclusions of 3D dataset cleaning

The maps produced continue to show the mutant in a more open conformation than the wild type as suggested by the 2D classification. Following this dataset cleaning, the overall number of 16mers taken forward was 43 536 and 10 307 for the wild type and mutant respectively, while the number of 18mers was 22 895 and 2 074. This corresponds to 34.5% of the total population (of good particles) for the wild type and 16.8% for the mutant. For further discussion of the origin of these 18mers, see Appendix I.

All maps thus far have included a region of density in the centre of the particle. The origin of this is unclear, though it may be due to protein or DNA bound to the particle which was not removed during the purification. In future steps of this study this density is masked in order to ensure that classification and refinement focuses on the thermosome density.

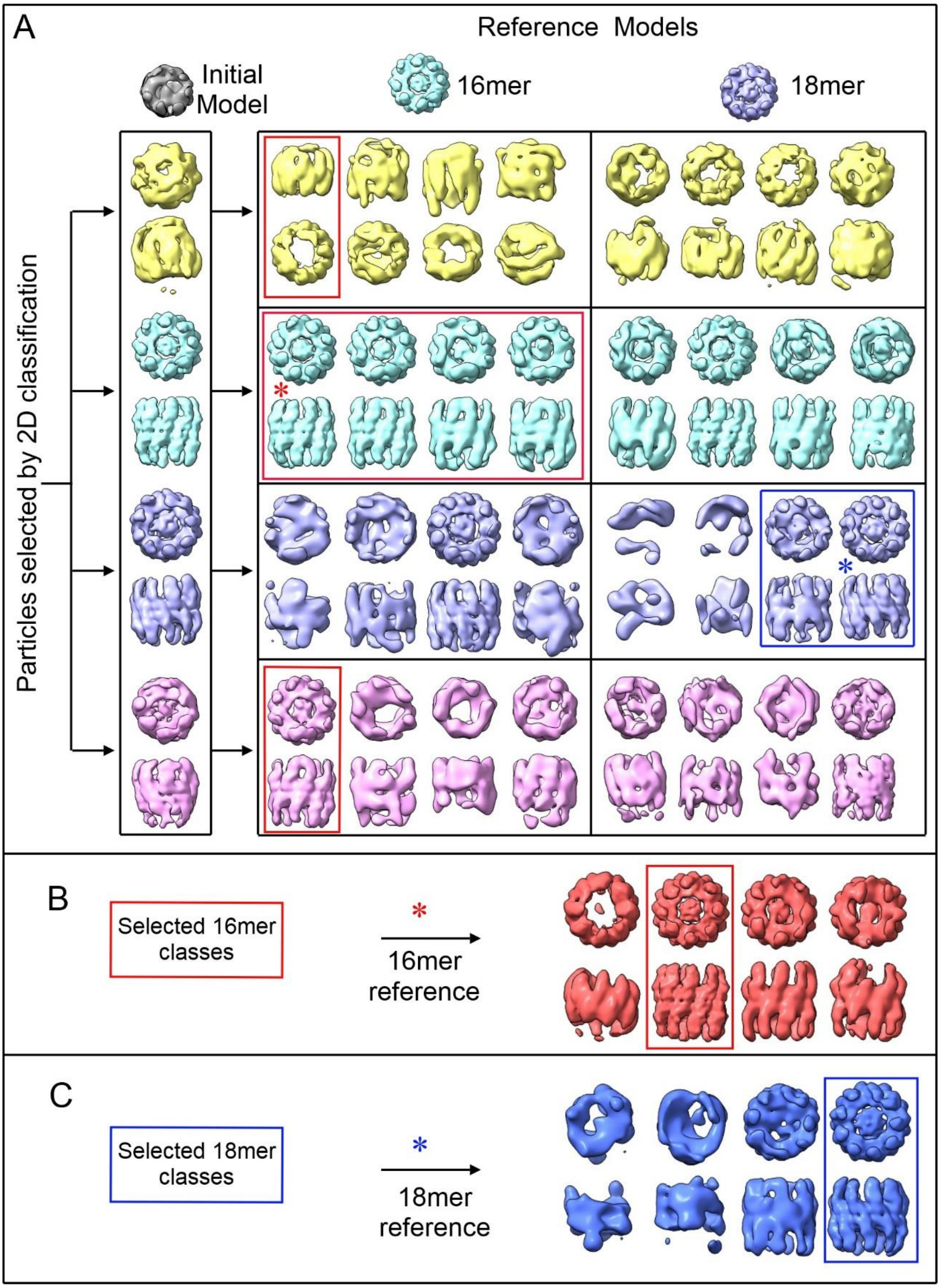


Figure 2-19: Choosing 16mers – Mutant. (A) Initial 3D classification with initial 3D map as a reference (left). Each of these classes was then reclassified with a 16mer reference map - blue (centre) and an 18mer reference map – purple (right). Red boxes indicate 16mer classes selected for final classification with reference map marked by red asterisk. Blue boxes indicate 18mer classes selected for final classification with reference map marked by blue asterisk. Results of final classification of 16mers (B) and 18mers (C).

2.5.7 Map symmetries

Having cleaned up the dataset sufficiently, it was now possible to begin to investigate the symmetry of the particles. Whilst there is evidence for D4 symmetry from the crystal structure, the aim was to make no assumptions as far as possible, particularly for the mutant which has not been seen before and contains a mutation in only one of the two subunits which may alter the symmetry compared to the wild type.

Applying symmetry to a map during reconstruction should increase the resolution if it is a valid symmetry operation for the data. Applying symmetry effectively increases the amount of data available by allowing each particle to contribute multiple times to a reconstruction. For example, applying 4-fold rotational symmetry means that each particle contributes once for each rotational position, thus four times. The higher the order of symmetry, the higher the effective number of particles contributing to the reconstruction. However, if a symmetry operation is used which does not accurately describe the data, the ‘extra’ data will add to noise in the image and therefore prevent the resolution from improving for a finite dataset.

From the literature [9, 36] we know that the thermosome has 16 subunits arranged in two back to back rings of 8. We can therefore propose that it may have either cyclic (C) or dihedral (D) symmetry, with Cn being simply n-fold rotation around the central axis (z), and Dn being n-fold rotation around the central axis (z) and n 2-fold rotational axes in the x-y plane (Figure 2-20).

Put simply, dihedral symmetry should be the same as cyclic symmetry for the thermosome, with the added constraint that the two hemispheres are related to each other. In this way it should be possible to differentiate between structures with a bullet-like shape (one hemisphere open while the other is closed [21, 37]), and a symmetrical shape where both hemispheres are in the same state.

For an 8 membered ring, we may have a maximum of 1, 2, 4 or 8-fold rotational symmetry. The reported crystal structure [9] shows the two subunits arranged in an alternating pattern, which would lead us to expect 4-fold rotational symmetry and indeed the crystal structure was solved to D4 symmetry. The crystal structures of the two subunits are extremely similar however, so we may expect to see pseudo-8-fold symmetry rather than 4-fold if the resolution is not sufficiently high [9].

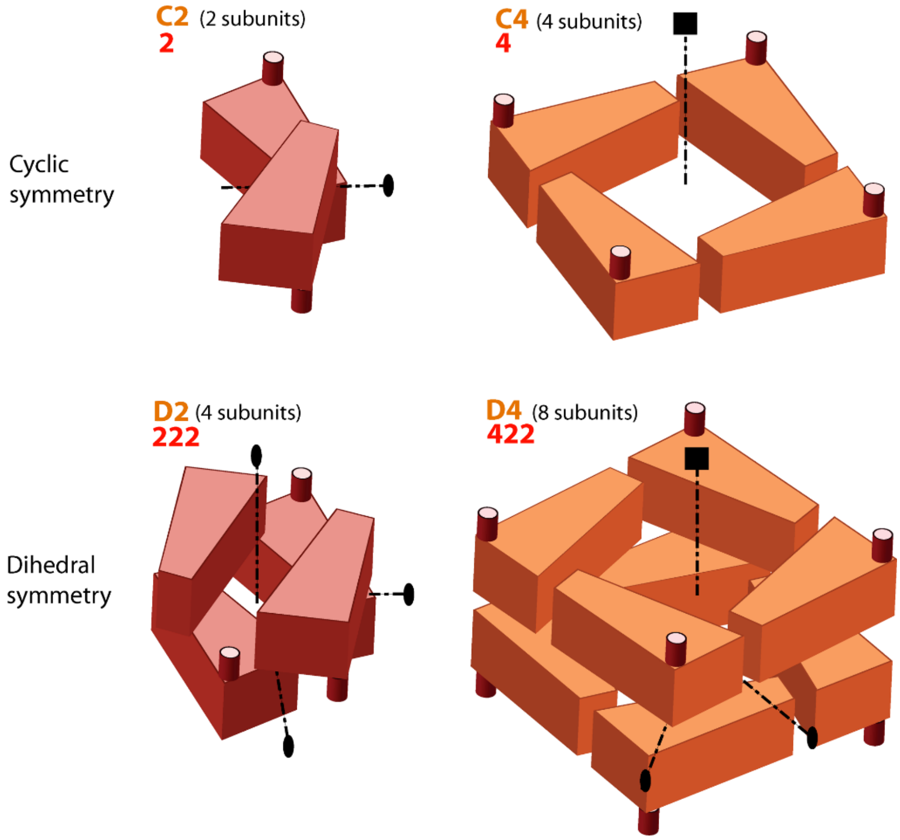


Figure 2-20: Cyclic vs Dihedral symmetries. 2-fold symmetry axes are indicated by an oval. 4-fold axes indicated by square. Reproduced from [38].

The 16mer particles identified in the previous section were subjected to further 3D classifications to identify which symmetry type best describes them. In each case the best 16mer class from the previous section was filtered and used as a reference map. These maps are indicated by red boxes in Figure 2-18 and Figure 2-19. Additionally, in order to encourage the formation of structures with appropriate shape, a loose mask was applied to the classifications. These masks were created by low pass filtering the best 16mer maps, extending by 3 pixels and applying a smooth edge of 3 pixels using RELION’s MaskCreate function. As mentioned above, at this stage the central density was also masked out, in order to focus only on the known thermosome density.

The results for the symmetrised classifications of the wild type and mutant datasets are shown in Figure 2-21 and Figure 2-22 respectively. Top and side views of each of the four classes are displayed along with the estimated resolutions for the best classes in each group.

D8 symmetry yielded the highest resolution maps at 6.7 Å and 8.0 Å. In both cases the trend in estimated resolution values follows that expected for a highly symmetrical particle, *i.e.* a particle which is well modelled by a high symmetry also contains the lower order symmetries.

Given the similarity of the subunits, we would not expect to see any differences between them in the wild type at these resolutions. For the mutant however we may have expected to see a conformational change between subunits due to the mutation only being present in the beta subunit. We hypothesised that the beta would be more open than the alpha (see Chapter 4:), but this does not appear to be the case and we can say that both the wild type and mutant thermosomes have pseudo D8 symmetry.

In general, classifications with all symmetries yielded good models with very little variation indicating a high quality, homogeneous dataset at least to the resolutions achieved thus far.

Dihedral symmetry appears to better describe the data than cyclic for equivalent rotational orders, indicating that there is no significant difference between the two hemispheres at the current resolution values. Furthermore, the fact that C8 and D4 (with equal effective number of particles) yield the same resolution in each case, suggests that the resolution is currently limited by the number of particles rather than heterogeneity of the datasets.

From the sequence and crystal structure, we expect that the overall layout is of alternating alpha and beta subunits. It was therefore decided to refine with D4 symmetry rather than D8 symmetry. Since one of the key aims of the study was to discover differences between the two subunits, particularly in the mutated thermosome, it did not make sense to continue with D8 symmetry in which any differences would be effectively blurred out in a final reconstruction.

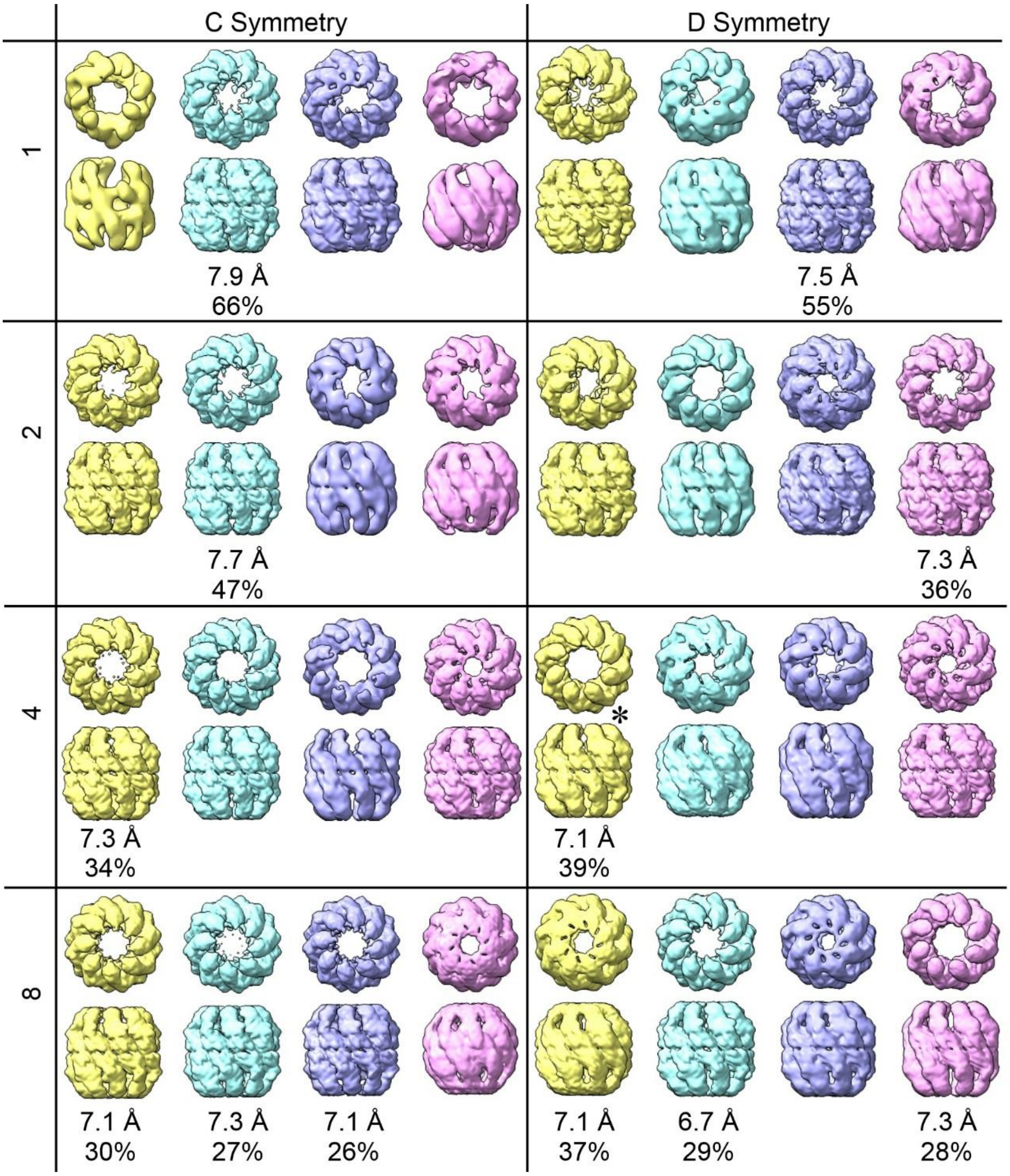


Figure 2-21: Symmetrised classification of wild type 16mers. Number on the left indicates the order of rotational symmetry of the primary axis. Numbers below indicate the estimated resolution and proportion of particles for the highest resolution and most populous classes.

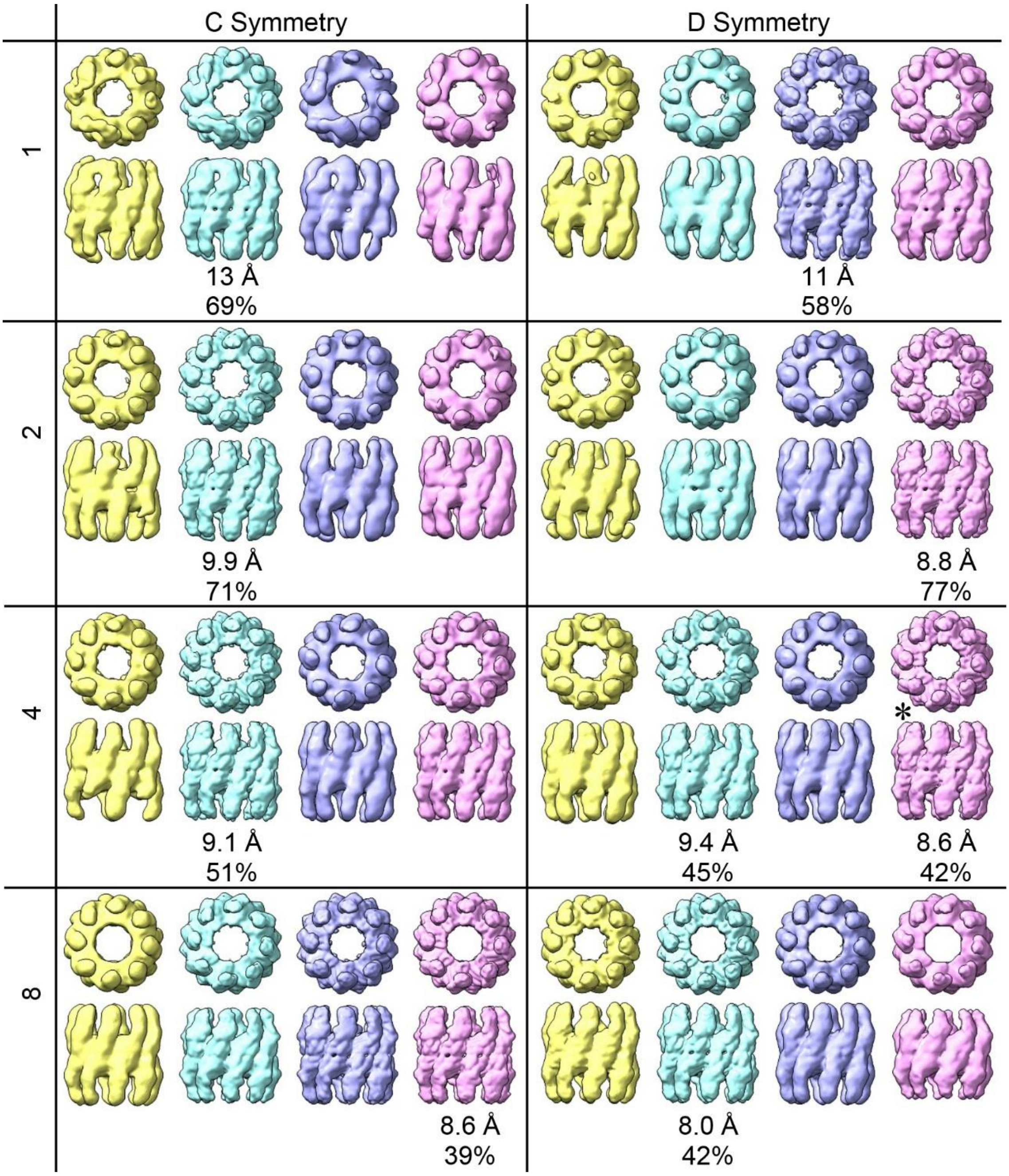


Figure 2-22: Symmetrised classification of mutant 16mers. Number on the left indicates the order of rotational symmetry of the primary axis. Numbers below indicate the estimated resolution and proportion of particles for the highest resolution and most populous classes.

2.5.8 D4 symmetry axis locations

As discussed above, the D4 symmetry group has a 4-fold rotation around a central axis (z), and four further 2-fold axes of rotation perpendicular to this (x and y axes and in between).

For a thermosome with D4 symmetry there are two possibilities for where these 2-fold axes can be located. If we take a thermosome with 16 subunits and unroll it, we can place the 2-fold axes either on the face between two subunits or through the vertex where four subunits meet (Figure 2-23).

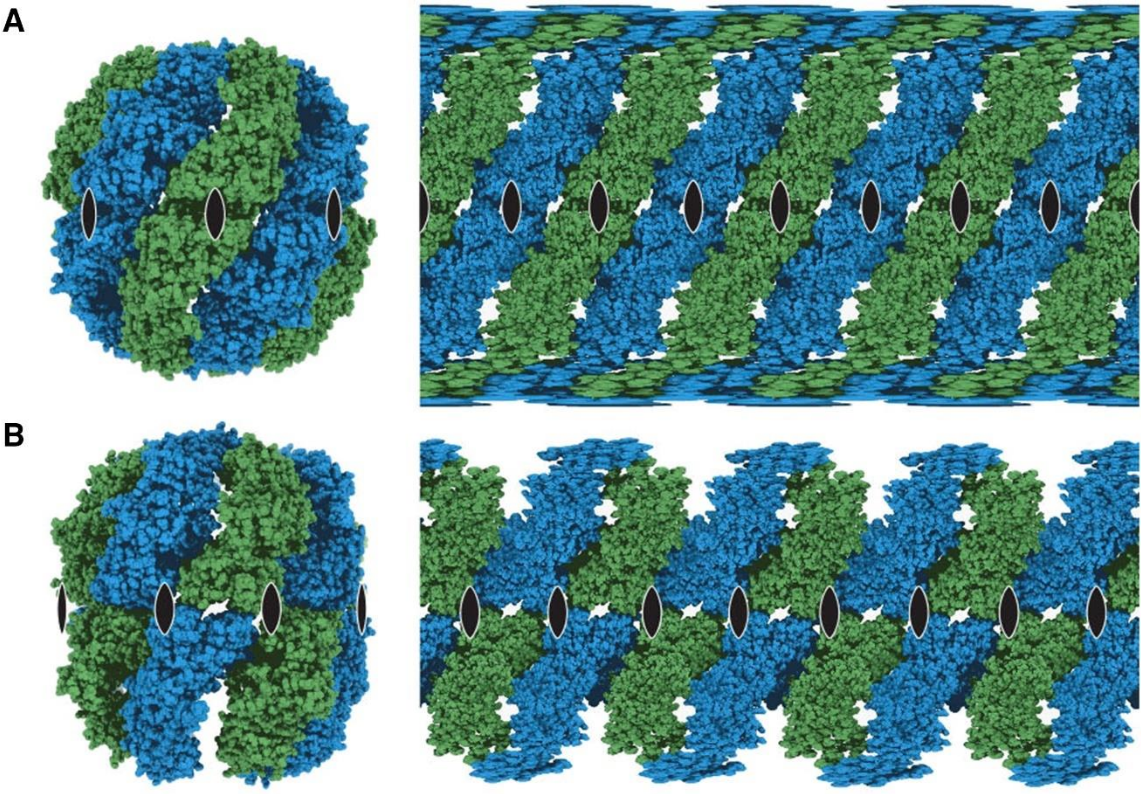


Figure 2-23: D4 symmetry axis locations. Mercator projections (A) D4A symmetry (B) D4B symmetry. Reproduced from [13]

The location of this axis affects the overall conformation. While each ring will remain alternating α β α β α β α β, in the case of the axis going through the joining face (A), α will map onto another α directly below. However, placing the axis at the junction between four subunits (B) will cause an α to be mapped onto an α which is shifted around by one subunit as illustrated by the colours in the figure. This means that the junction across the equator is heterogeneous rather than homogeneous. For the remainder of this thesis I will use the notation D4A to indicate a homogeneous junction (symmetry axis through centre of junction – Figure 2-23(A)) and D4B to indicate a heterogeneous junction (symmetry axis in between junctions - Figure 2-23 (B)).

The reported crystal structure for the thermosome of interest was solved with D4A symmetry, therefore we may expect the TEM data to reflect this. Other type II chaperonins however have been solved with D4B symmetry, *e.g*. the TF55 complex of *Sulfolobus shibatae* [13].

In order to resolve such differences, high resolution reconstructions were required. We therefore made use of RELION's 3D Autorefinement, Post Processing and Local Resolution protocols which are described in section 2.4.5.

2.5.8.1 Wild Type

The best three classes from the D8 classification were taken forward for refinement using D4 symmetry and the best map as a reference. (See Appendix II). The resulting map had a resolution of 3.9 Å and D4A symmetry.

In order to create a D4B reference and mask, the map was rotated by 22.5° and resampled at the same pixel size onto the original grid using Chimera. The resulting map was then inserted back into the RELION workflow. D4A and D4B masks were created using the original and rotated maps. All refinement, post processing, mask creation and local resolution settings were kept the same for both symmetry operations and the resulting maps, coloured by local resolution, are shown in Figure 2-24.

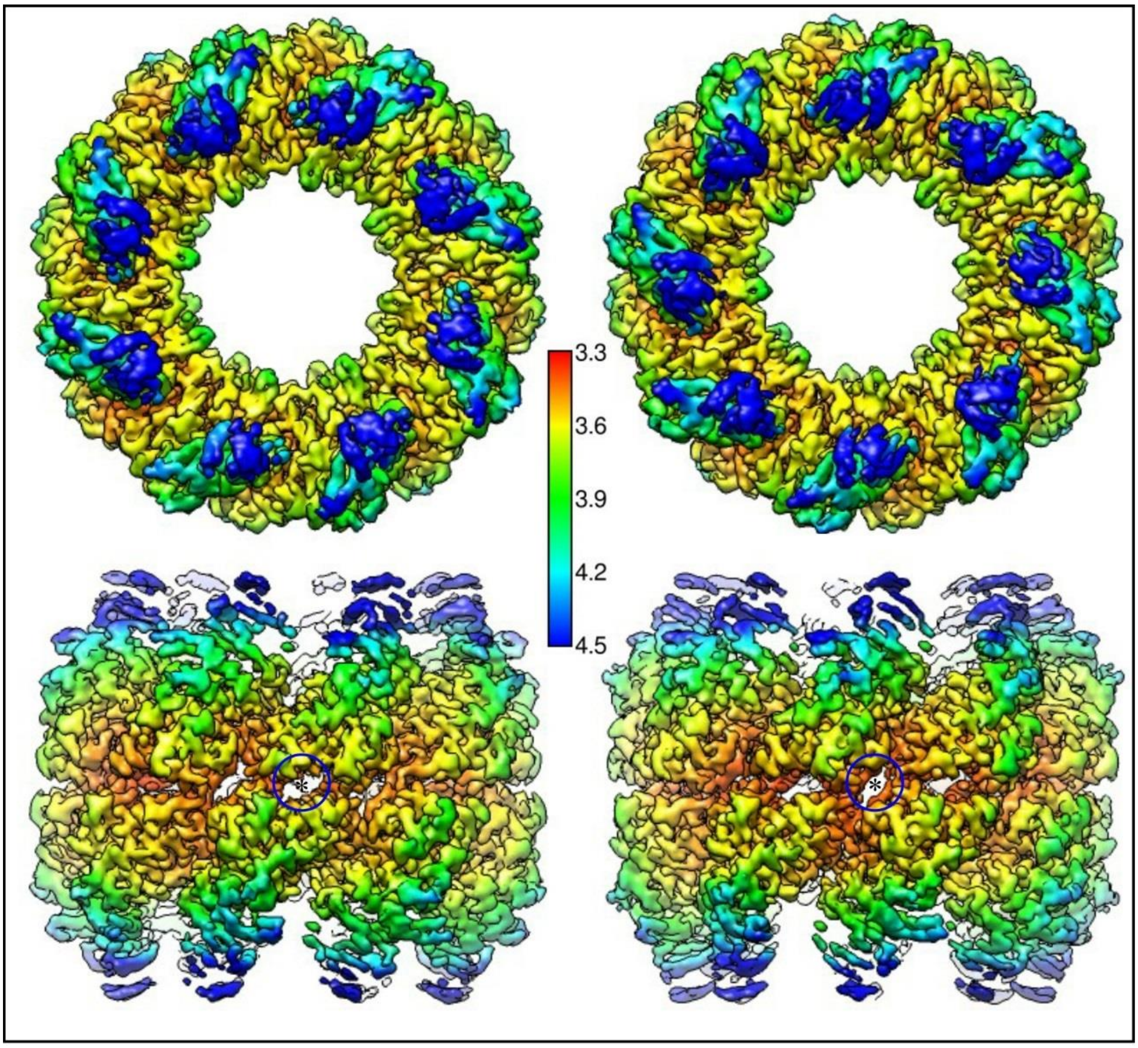


Figure 2-24: Wild type D4A (left) vs D4B (right) symmetrised maps. Circled stars in the lower map indicate the location of one of the 2-fold rotation axes.

	Refinement	Post-process	B-factor
D4A	3.9 Å	3.5 Å	-115
D4B	3.9 Å	3.5 Å	-114

Table 2-4: D4A vs D4B resolution values for wild type

The resolutions calculated at each stage were identical as shown in Table 2-4, and fitting the maps together resulted in a correlation of 0.96 implying that they are morphologically almost identical.

The two maps are indistinguishable based on both the global and local resolution calculations for this data.

2.5.8.2 Mutant

Rather than refining to create a reference, the mutant dataset was classified one final time using the map marked with an asterisk in Figure 2-22 as a reference. The best resulting class was then used to create masks for both D4A and D4B symmetrisation as described above. Finally, the data were refined, post processed and local resolutions determined, the results of which are shown in Table 2-5 and Figure 2-25.

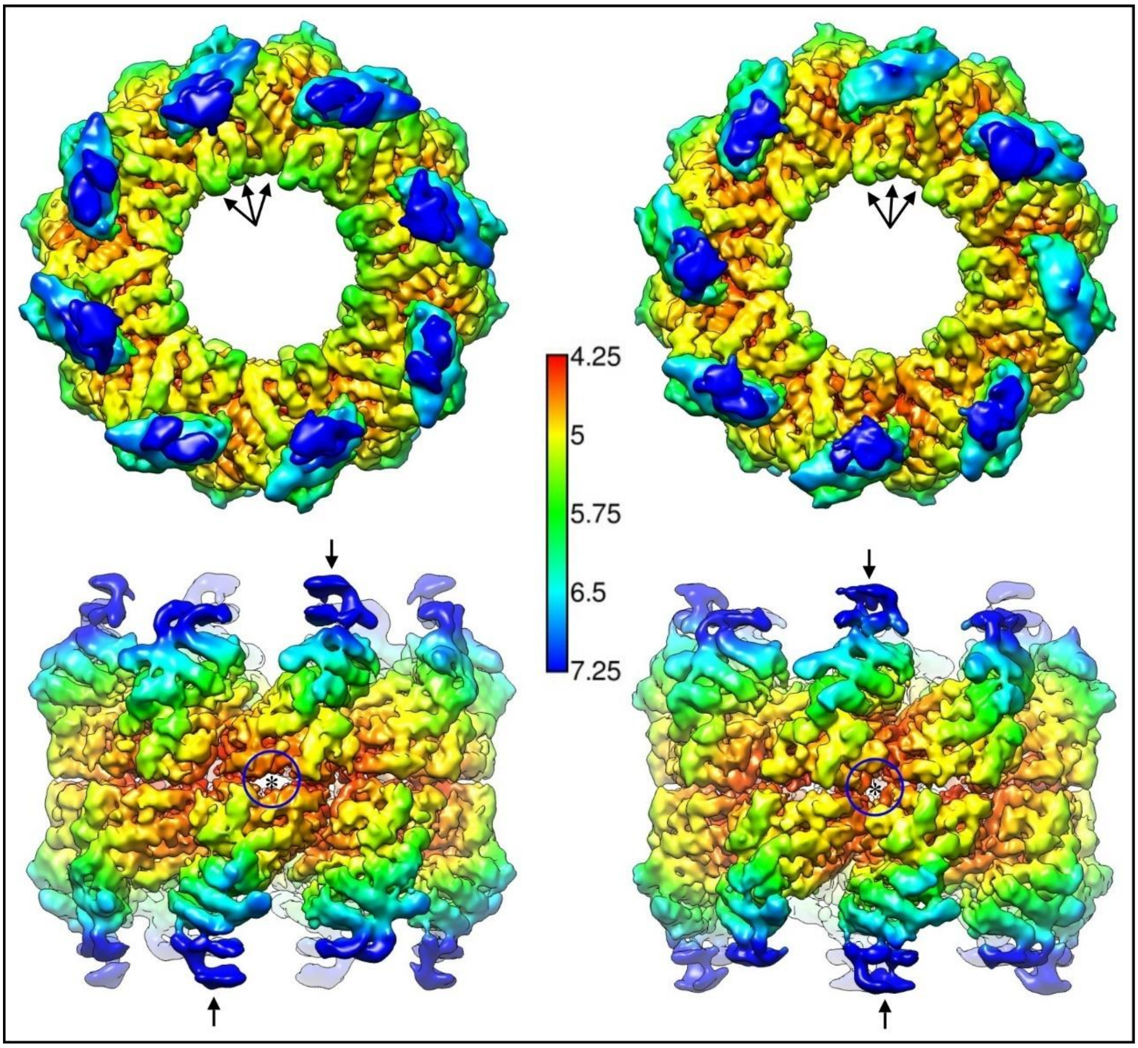


Figure 2-25: Mutant D4A (left) vs D4B (right) symmetrised maps. Circled stars in the lower map indicate the location of one of the 2-fold rotation axes. Arrows in the top views indicate alpha helices in equatorial domain where different local resolutions may be seen. Arrows in the side views indicate the apical domains of symmetry related subunits.

	Refinement	Post-process	B-factor
D4A	6.8 Å	4.7 Å	-168
D4B	6.4 Å	4.5 Å	-171

Table 2-5: D4A vs D4B resolution values for mutant

Overall the D4B map yields slightly higher resolution at each stage of processing. Comparing the local resolution maps, which have the same threshold value, we can see that the best-defined regions are the equatorial domains, where resolutions reach 4.3 Å, while further from the centre, and further up the flexible apical domains, the resolution falls to 6-8 Å and worse.

The differences between D4A and D4B symmetries can be clearly seen by comparing the noise of the apical regions in the side views. In the D4A we can see that one subunit protrudes further (arrows), with a matching protrusion on its opposite subunit. However, in D4B, the matching protrusions are offset between hemispheres.

We would expect to resolve secondary structure for resolutions below around 7 Å, which is indeed the case in the equatorial and intermediate domains. A clear example is the equatorial alpha helices indicated by arrows in the top views.

2.5.9 Summary thus far

Two high resolution maps have now been produced for both the wild type and a mutant variant of the *T. acidophilum* thermosome. To our knowledge, at the time of writing these constitute the first cryo-TEM maps of this thermosome, and the first map created by any method in an open conformation.

The key findings thus far are as follows:

1. The single point mutation in the beta subunit nucleotide binding domain has caused a significant change in morphology of the thermosome. While the wild type is in a closed state similar, to that of the published crystal structure, the mutant is in a previously unseen open conformation.
2. All local resolution maps show significantly lower resolution at the apical domains due to their increased flexibility compared to the equatorial domains.
3. In the equatorial regions, the wild type yielded up to 0.8 Å increase in resolution over the mutant. This is a significant difference as at around 4.5 Å we may expect to resolve larger side chains which would be required to differentiate between subunits with greater confidence [14].
4. It was not possible to distinguish between D4A and D4B symmetries for the wild type data.
5. While estimated local resolutions in the equatorial regions were slightly better for D4B symmetry map in the mutant, these differences are only very slight.
6. No definitive conclusions on whether these particles have D4A or D4B symmetry may be drawn at this point.

2.5.10 Combining Mutant Datasets

As discussed previously, it is likely that the achievable resolution of the mutant dataset was limited by the available number of particles. In order to distinguish between the alpha and beta subunits it is likely that fitting of side chains will be necessary requiring a slight improvement in the current resolution achieved. It is possible to estimate the number of particles required to reach a certain resolution by solving structures from particle subsets and extrapolating to predict the behaviour of a larger dataset [39].

Random subsets of mutant particles were created from the full dataset and refined in the same way as above. The resulting resolutions were then plotted as shown in Figure 2-26.

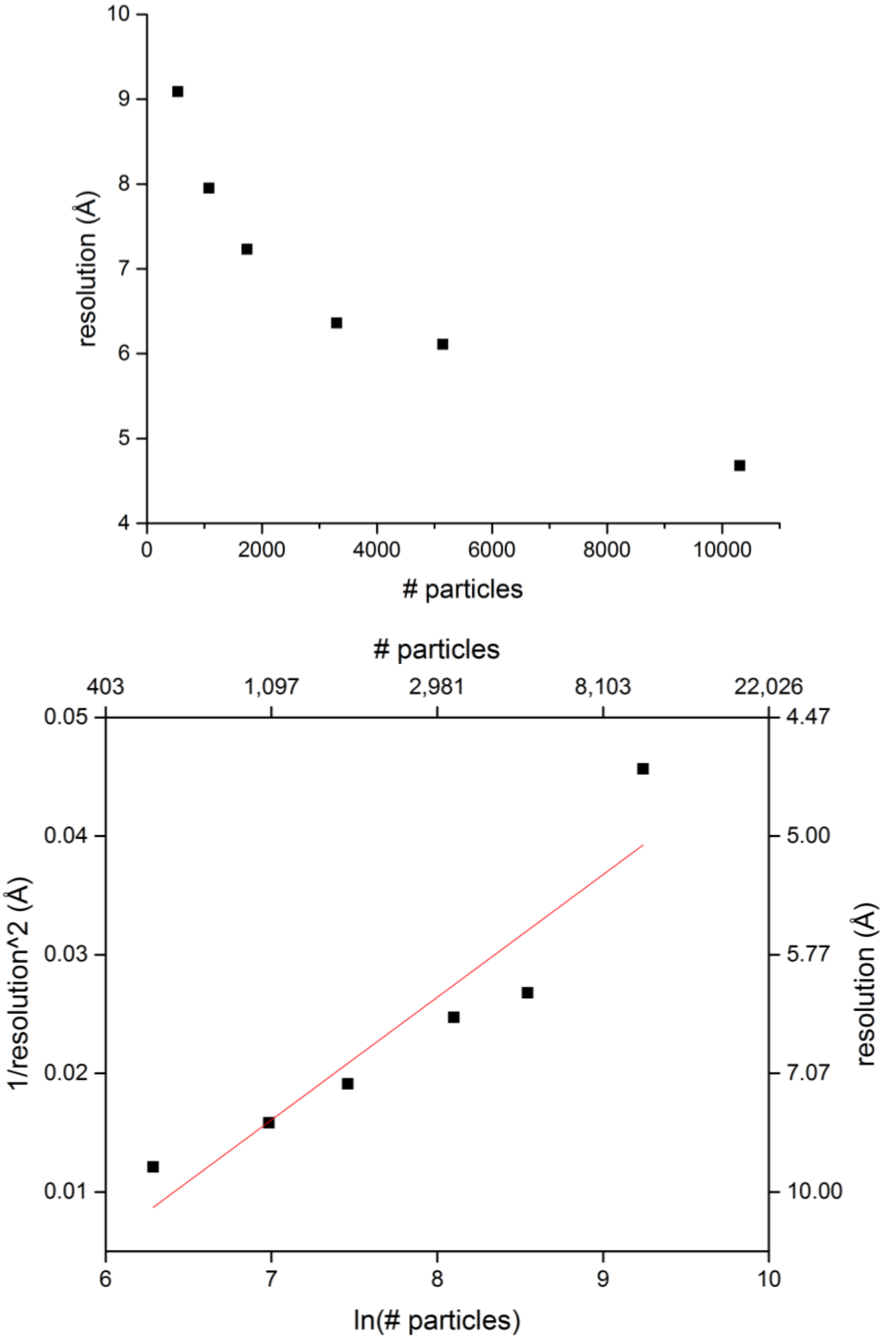


Figure 2-26: B-factor plots calculated from mutant data subsets. Top – number of particles vs resolution achieved by autorefinement. Bottom – the same data linearised.

The first plot shows the number of particles in the subset against achieved resolution. This drops off exponentially, demonstrating that increasing the dataset size indefinitely will not result in ever-improving resolution. This is because high resolution contrast is lost during imaging (*e.g*. due to sample motion, radiation damage, charging) [39]. This relationship can be linearised by plotting the natural log of the number of particles against 1/resolution^2. Fitting a straight line to this plot then allows extrapolation of predicted resolution for a given number of particles and vice versa, examples of which are presented in Table 2-6.

# particles	**Predicted resolution**	**Resolution**	**Predicted # particles**
20 000	4.7 Å	4.0 Å	97 600
40 000	4.3 Å	3.5 Å	621 000
60 000	4.2 Å	3.0 Å	10 700 000

Table 2-6: B-factor plot extrapolation.

We can clearly see from these results that we are unlikely to ever create a 3 Å map with these data. However, increasing the dataset to 40 000 may yield sufficient improvement to be able to distinguish side chain features which will aid in model fitting (see Chapter 3:).

It was felt that more data were required, and this experimental prediction agreed. A second complementary set was therefore collected using grids prepared in the same batch as above. This dataset was processed in parallel by Marston Bradshaw. Specific details of each analysis were not shared during the process (other than extraction box size) in order to eliminate bias and to confirm the validity of later combination of datasets. The full image processing workflow of the second dataset analysis will not be shown here however the key methods and results are noted in Table 2-7.

	Dataset 1 (described above)	Dataset 2
Number of micrographs after pre-processing	1047	1524
Number of particles picked	250 000	141 000
Dataset cleaning in 2D	Manual removal after autopicking Removed poor 2D classes – 3x	Removed poor 2D classes – 6x Measured widths of good classes and removed larger as 18mers.
Dataset cleaning in 3D	Several rounds of classification. 18mers removed by 3D classification	One round of classification
Number of particles in final map	10 307	28 238
Point at which symmetry was first applied in processing	3D classification	3D classification
FSC resolution	4.7 Å	3.9 Å

Table 2-7: Comparison of processing of two mutant datasets

2.5.10.1 Comparison of maps from each individual dataset

The maps generated from the two datasets are shown side by side in Figure 2-27. Both have been refined using D4A symmetry and show the same overall morphology. The most striking difference between these maps is the difference in resolution to which each has been refined, with an improvement of almost 0.8 Å for the second. This can clearly be seen in the equatorial domain where the whole region has reached the same or better resolution than even the best regions in the first. This is likely since the second map is formed from almost 3 times the number of particles. Extrapolating from the plot above (Figure 2-26) we would have predicted 30 000 particles to give a resolution of 4.5 Å. This represents a significant improvement on that, strengthening the reasoning for adding more data.

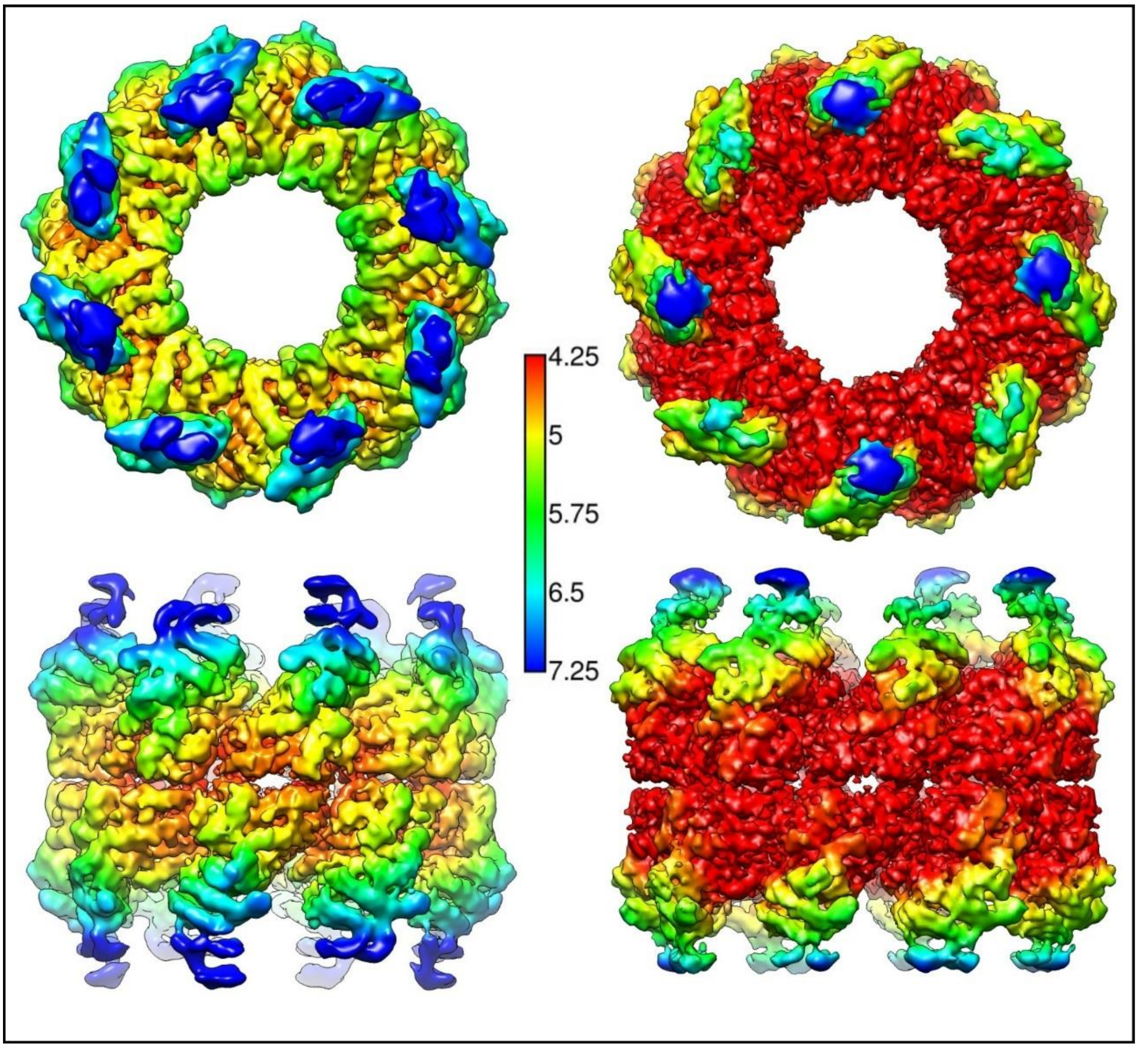


Figure 2-27: Final maps produced from separate mutant datasets processed by the author (left) and Marston Bradshaw (right). Threshold chosen to show approximately the same features for comparison.

2.5.10.2 2D and 3D processing

The two sets of particles which were used in the maps above were combined to create a new dataset pf 39 000 particles.

An initial 2D classification of the particles was performed in order to select 10 classes (3000 particles) for an initial map. Once again, 2D classification shows a high proportion of side views and no intermediate angles (Figure 2-28). The side views show a good level of detail in the highly ordered equatorial region and there are no 18mer classes.

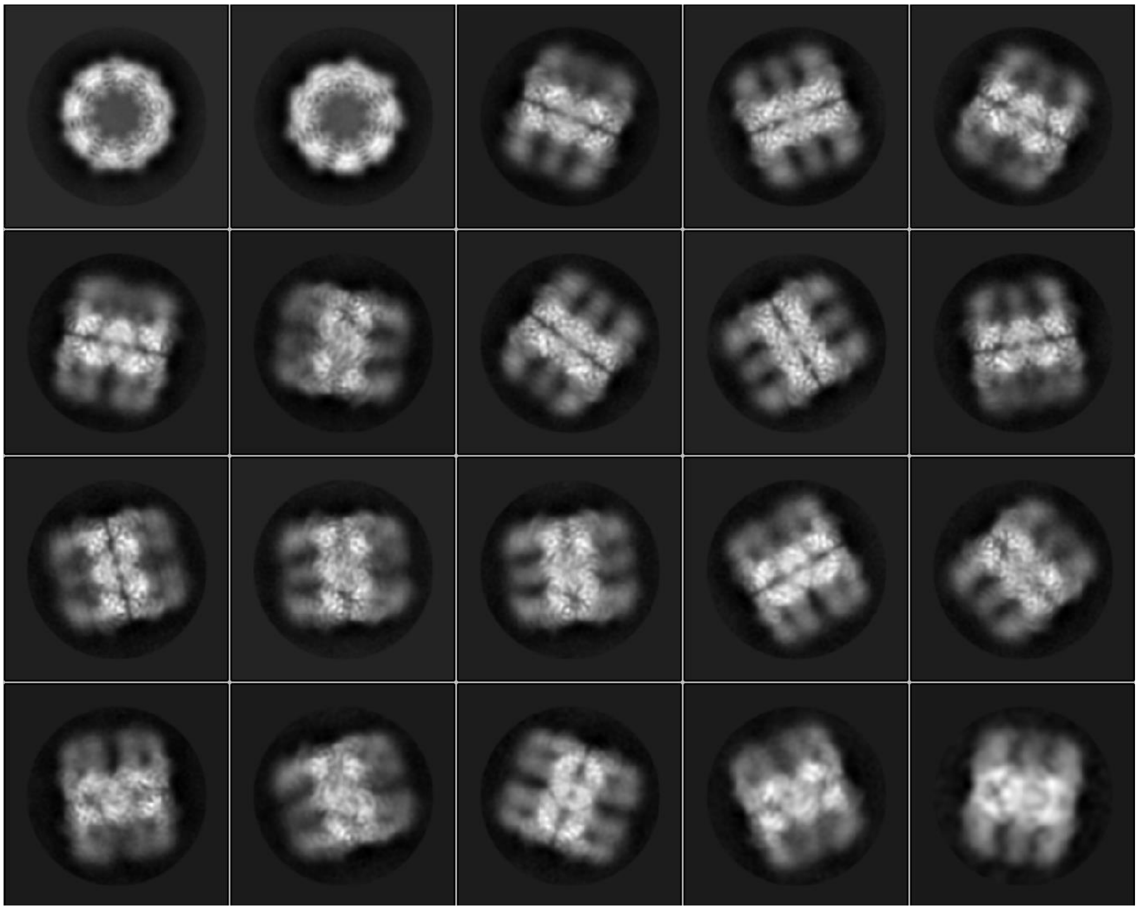


Figure 2-28: 2D classification of combined mutant datasets. 20 most populous classes showing a preference for side views.

Figure 2-29 shows the initial map, which has an appropriate barrel shape and apical domain-like features. Also shown is an overview of the 3D processing which followed the same protocols as described above. Both D4A and D4B maps were created but only one is shown for brevity. Resolutions in the equatorial regions reach around 3.5 Å which is comparable with the wild type map.

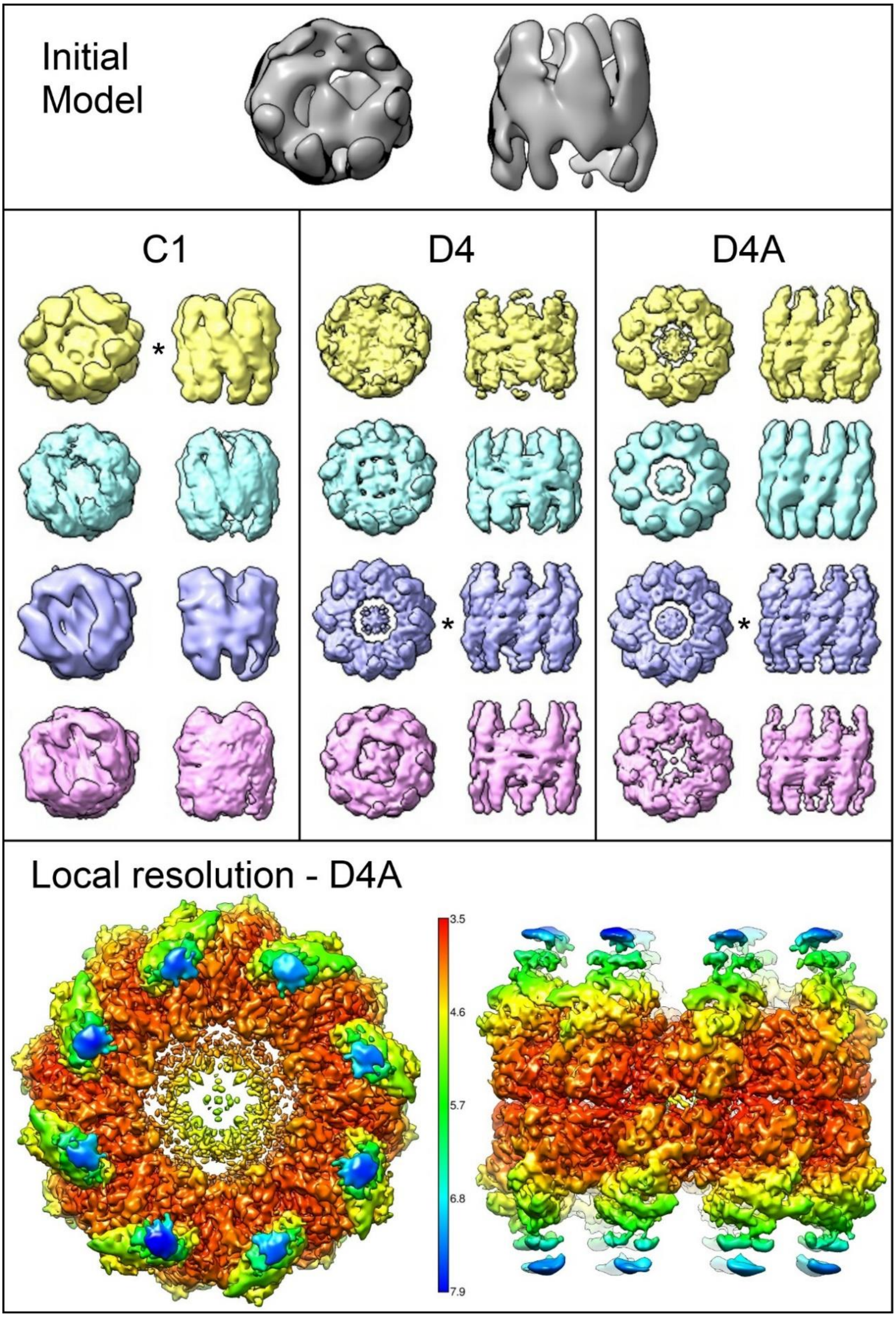


Figure 2-29: Combined mutant dataset 3D processing. Top: initial map, Middle: Sequential 3D classifications with C1, D4 and D4A symmetries. Classes marked with an asterisk were used as a reference for the following step. Bottom: refined, post processed map with D4A symmetry showing resolution in equatorial region reaching 3.5 Å.

2.5.10.3 Symmetry revisited

Given the improvement in resolution, we wanted to confirm that appropriate symmetry was being used for the data. To reiterate, dihedral symmetry would suggest that the two hemispheres were in the same conformation, while rotational symmetry order gives information on the organisation of the subunits within the rings. A similar study was therefore performed to that described in Section 2.5.7. This time however, given the greater number of particles available, subsets were made for each symmetry operation to control for the increase in effective particle numbers. Thus C2, C4 and C8 contained $\frac{1}{2}$, $\frac{1}{4}$ and $\frac{1}{8}$ of all particles respectively, while D1, D2, D4, and D8 contained $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$ and $\frac{1}{16}$. In each case, the subsets were formed of a random selection of particles, using RELION's random sort option.

Since the D8 subset contained only 2400 particles, it was not appropriate to perform 3D classification, but rather each subset was refined using the D4A map as a reference and mask. In order to most clearly illustrate the differences, the refined maps were then post processed and local resolution maps produced. These are shown in Figure 2-30 in which all maps are shown with the same colour scale and using the same thresholding cut-off.

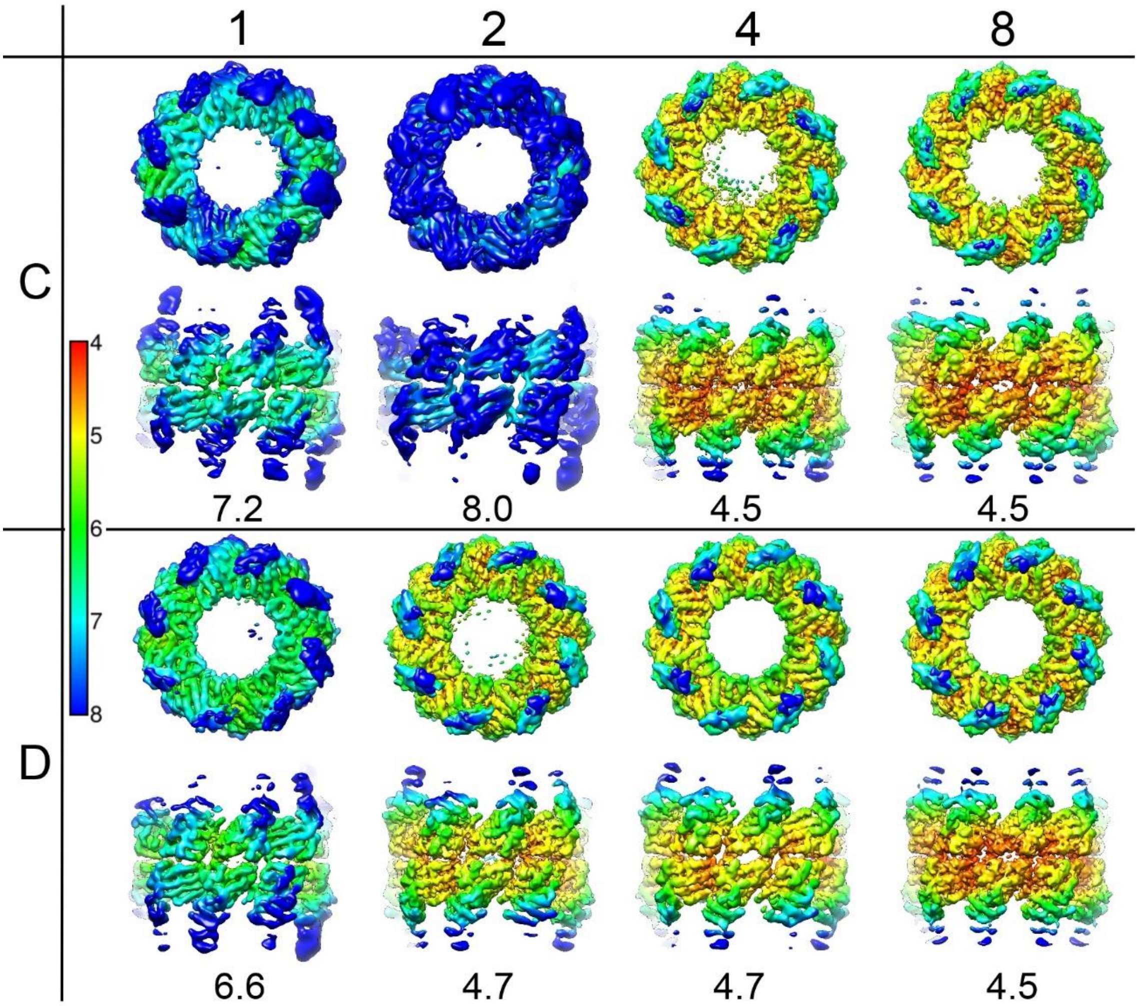


Figure 2-30: Subsets of combined mutant dataset refined using different symmetry operations. Overall resolution calculated from FSC shown below in Å.

Interestingly, in contrast to the previous result, the best resolution map at both the refinement and post processing stages was the C4 symmetrised map which reached an overall resolution of 4.48 Å, compared to D4 which was 4.7 Å.

It was therefore decided, based on these results to create final reconstructions based on both C4 and D4 A and B symmetries, in keeping with the D4 character shown in crystallographic studies.

In order to generate new masks for each symmetry operation, the full dataset was subjected to a final 3D classification (not shown). In each case, the best class was low pass filtered for use as a reference in the final refinement, and to create a mask. In addition to refinement of the full dataset, the best class from each was also refined on its own to ensure that there was not a mixture of structures present. Resolutions calculated for each are listed in Table 2-8.

	Refine	Post Process
D4A best class	4.5 Å	4.1 Å
D4A all	4.2 Å	4.0 Å
D4B best class	4.4 Å	4.1 Å
D4B all	4.2 Å	3.8 Å
C4 best class	5.9 Å	4.5 Å
C4 all	4.4 Å	4.1 Å

Table 2-8: Subset vs Full dataset resolutions

Overall it was decided that the benefit of the increase in resolution using the full dataset outweighed any negative impact of combining the classes and ultimately the full dataset was used.

2.5.10.4 Final reconstructions

The final reconstructions of the full dataset based on each of the three symmetry operations are shown in Figure 2-31. They are represented at both high and low thresholds to give an idea of the level of detail available in the different regions of the maps. At the lower threshold, the lower resolution apical domains are visible. The local resolution maps are extremely similar between the D4A and D4B while C4 drops to lower resolutions closer to the equatorial region. The entire equatorial region for the D4A and D4B are resolved to approximately 4.5 Å or better while the C4 falls to 5 in some areas.

At the higher threshold, the two D4 maps appear extremely similar, however the D4A has a slightly higher local resolution in the equatorial regions as shown in the magnified regions. This is not reflected in the post processing overall resolution estimates of 4.0 Å for the D4A and 3.8 Å for D4B.

The result that the C4 map does not resolve to as high resolution as the D4 maps (reaching 4.1 Å), is likely down to the fact that the D4 still has effectively twice as many particles as C4. However, given the result from the symmetry study above we may expect that a C4 map would refine to similar resolutions given a dataset of twice the size. We would expect that once a certain number of particles was reached, the D4 symmetry would cease to improve the resolution over C4. At this point informative differences between hemispheres may be revealed, but at the current resolutions these are not visible.

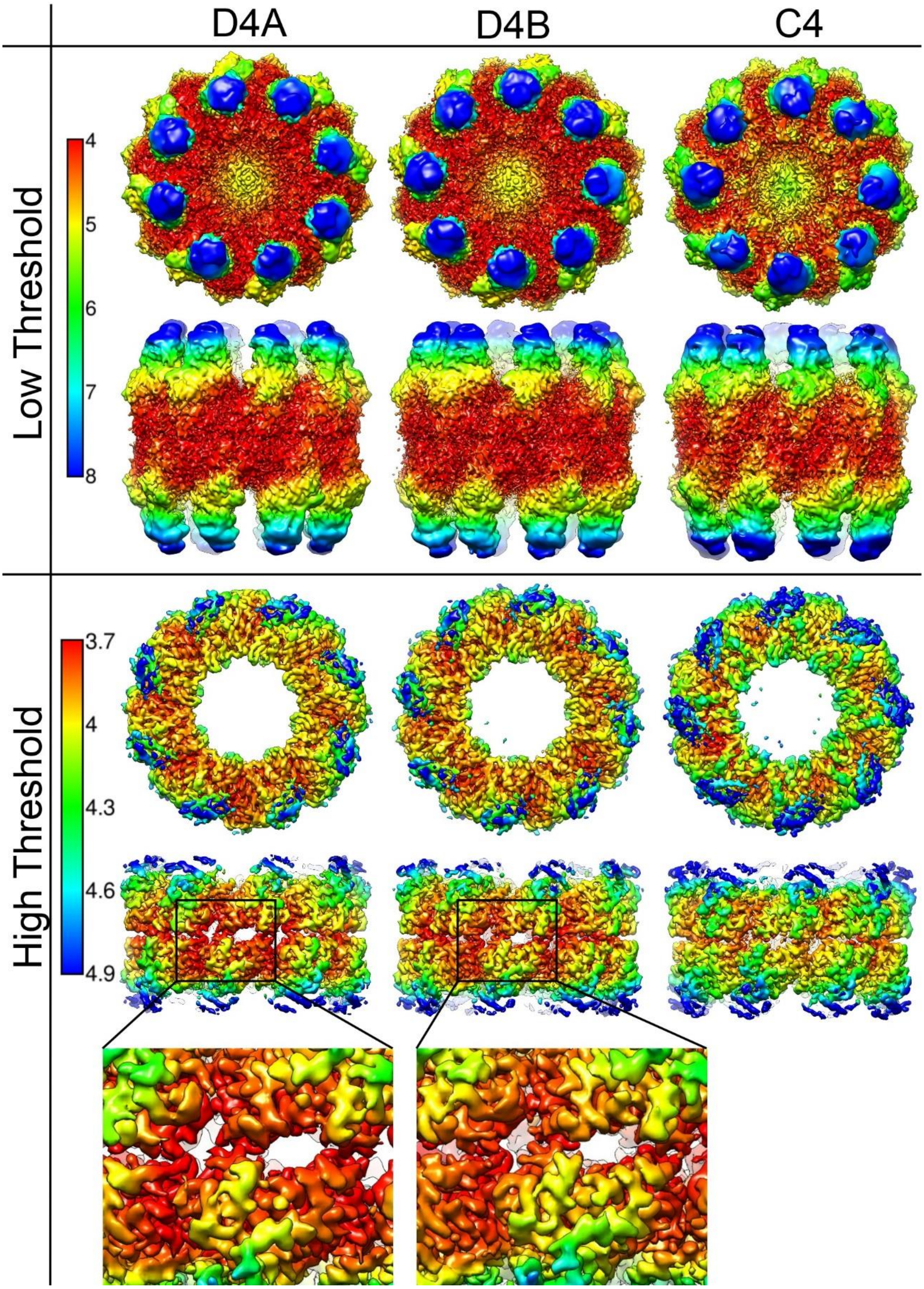


Figure 2-31: Combined mutant dataset final models at high and low threshold cut-offs.

The key results of combining datasets are as follows:

1. The overall resolution based on gold standard FSC has improved from 4.5 Å for 10 300 particles to 3.8 Å for 38 500 particles.
2. Initial and combined mutant datasets both had slightly better FSC resolutions for D4B than D4A. (A:4.7Å vs B:4.5Å and A:4.0 Å vs B:3.8Å)
3. Examination of local resolution maps for the combined datasets suggests resolution in the equatorial region may be higher for D4A.
4. There is no improvement in resolution for the apical regions.
5. D8 resolved better than D4 in subset symmetry analysis so there is no evidence that the beta subunit is any more or less open than the alpha.
6. C4 resolved the best in the subset symmetry analysis but this did not translate to an improvement for the full dataset.
7. The overall resolution achieved by combining the datasets is better than suggested by the B-factor plot in Figure 2-26.
8. Updated B-factor plot (Figure 2-32 – blue line) including all data points suggests that 244 000 particles would be required to reach 3.5 Å
9. Including only the two best datapoints, *i.e*. excluding the subset analysis, reduces this further to 160 000 particles.

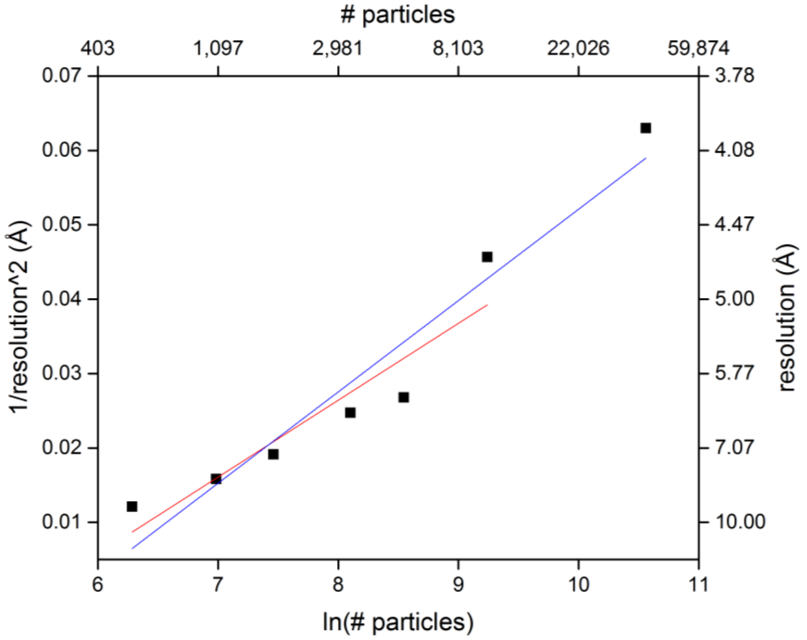


Figure 2-32: Updated B-factor plot with extra data point from combined dataset. Red line is original calculated fit. Blue line is new calculated fit for all datapoints.

2.5.11 Comparison of wild type and mutant maps

Given that the wild type maps calculated with D4A and D4B were indistinguishable, and that the mutant maps were similarly ambiguous, it was decided to proceed with D4A symmetry in line with the published crystal structure [9]. The final maps are therefore shown in Figure 2-33 along with their FSC curves and the angular distribution of contributing particles.

As seen in this, and indeed all stages of this study, the wild type is in a substantially closed conformation while the mutant is open. Measurements of differences in dimensions are discussed in the following chapter.

The resolutions achieved for the maps are both below 4 Å, with almost the entire equatorial domain of both maps being resolved to 4.5 Å or better. This is crucial for distinguishing between the subunits when model fitting as the crystal structure shows the secondary structure arrangement of the subunits are almost identical [9].

The angular distribution of the particles contributing to the two final reconstructions varies between maps. This is shown in Figure 2-33(C) where a longer, redder line indicates a greater number of particles, and peaks are arranged on the surface of a sphere corresponding to the orientation of the particles. Note that the applied D4 symmetry means only the 90º wedge is required to describe the full map. In both cases the majority of particles are oriented in either a top or side view. While the wild type sees particles in the full range of angles, this is much reduced for the mutant.

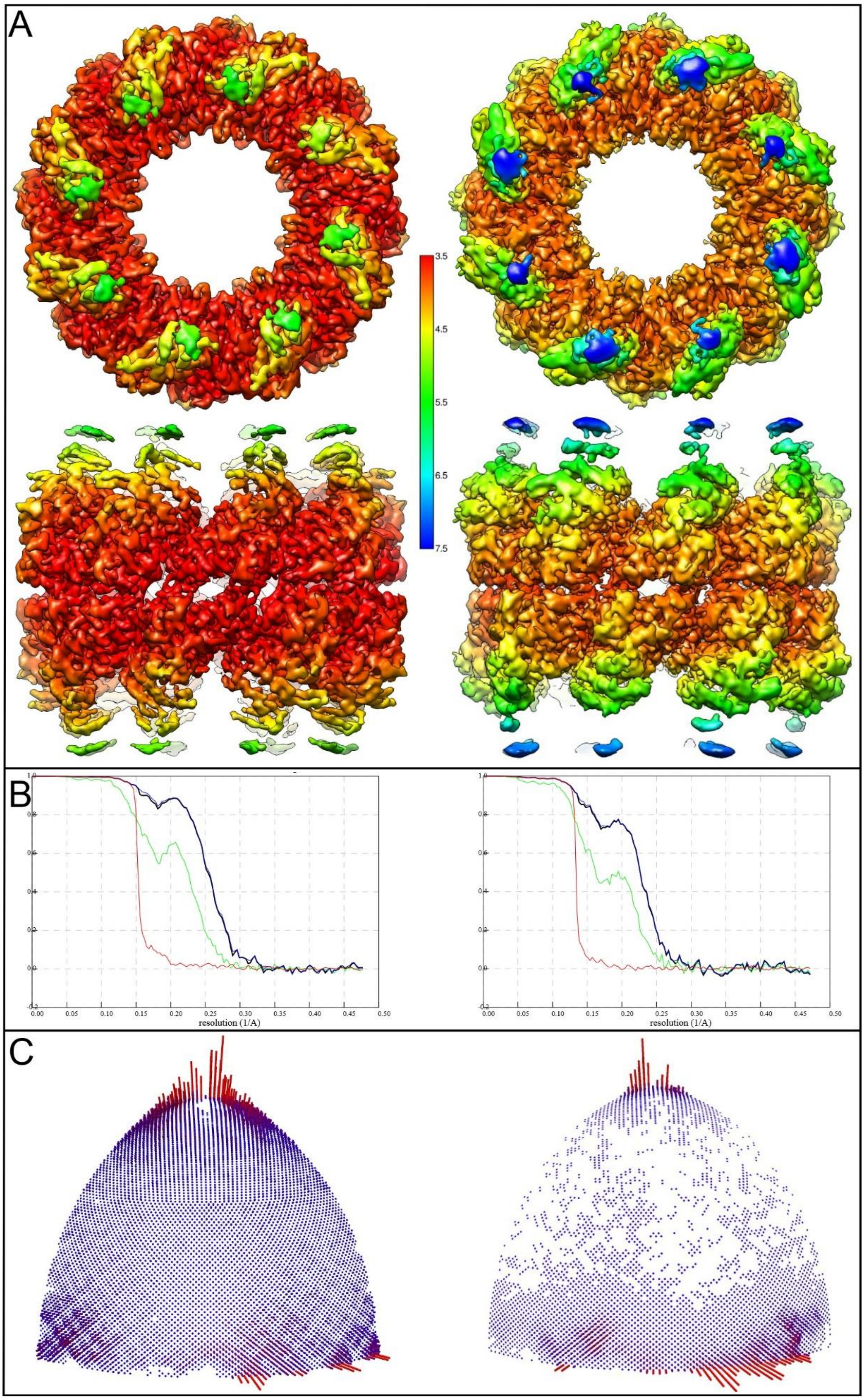


Figure 2-33: Final wild type (left) and mutant (right) maps. (A) Top and side views coloured on the same resolution scale. (B) Half-map FSC curves (see Appendix III for full scale). (C) Angular distribution of particles contributing to final map.

2.6 Conclusions

In this chapter I have shown the development of two maps to sub-4 Å resolution. Data collection and processing differed for the two samples so main differences and reasonings for these have been discussed.

Although every care has been taken to ensure that a true representation of the structure of the thermosome has been found, the software used in this analysis may not have been ideally suited to the system. Particles with strong pseudo-symmetry, such as the thermosome, may be classified based on their strong low resolution features when using multi-reference alignment methods such as those implemented in RELION. This leads to blurring at the resolution desired for the analysis discussed in this chapter, particularly when reference models are filtered. Creating an independent model in an alternative software package which does not make use of multi-reference classification could further confirm the following observations and provide strength to the analyses in this and the following chapters.

The first key outcome that we see is the effect that a thin layer of continuous carbon film has on the orientation of particles on the grid. In the wild type, where no carbon film was used, there is a slight preference for top views, but all angular distributions are available. The possibility of tilting the sample was explored in order to remedy this for the wild type, but since the particles are roughly spherical, a slight tilt did not have much effect on the particle projection image and it was decided that we would continue imaging without tilting. In the mutant dataset however, a continuous carbon layer was added to the Quantifoil grid which resulted in a wider range of views. Where previous works have employed detergents or removal of the helical protrusion in order to encourage an increase in side views, this provides a non-chemical, non-invasive method of producing the same result.

Secondly, it is shown that by combining multiple data sets, the final resolution of the output structure could be improved. This was done for the mutant dataset for two reasons: firstly, that the grids were prepared at a slightly lower concentration so there were fewer particles per image compared to the wild type data set. Secondly, the images were collected on a Talos Arctica, rather than a Krios, so the number of micrographs that could be collected during an imaging session was lower than that on the Krios. Despite these differences the overall resolution of the two final models was relatively similar.

Finally, the effect of the single point mutation in the ATP binding domain of the beta subunit has a drastic effect on the overall structure of the thermosome particles. While the wild type sample imaged was in a relatively closed conformation, very similar to that of the crystal structure, the mutant was in an open conformation.

In order to interrogate the structure at the molecular level we have taken the published crystal structure [9] as a starting point to generate atomistic models. The gross changes to the structure and the specific effects of the D93K mutation on the ATP binding pocket will be the focus of Chapter 3:. We will investigate what these models can teach us about the mechanism of action of the thermosome.

References

1. Yebenes, H., et al., *Chaperonins: two rings for folding.* Trends in Biochemical Sciences, 2011. **36**(8): p. 424-432.
2. Breiman, A. and A. Azem, *Editorial: Type I Chaperonins: Mechanism and Beyond.* Frontiers in Molecular Biosciences, 2018. **5**.
3. Lopez, T., K. Dalton, and J. Frydman, *The Mechanism and Function of Group II Chaperonins.* Journal of Molecular Biology, 2015. **427**(18): p. 2919-2930.
4. Kabir, M.A., et al., *Functional Subunits of Eukaryotic Chaperonin CCT/TRiC in Protein Folding.* J Amino Acids, 2011. **2011**: p. 843206.
5. Braig, K., et al., *The Crystal-Structure of the Bacterial Chaperonin Groel at 2.8-Angstrom.* Nature, 1994. **371**(6498): p. 578-586.
6. Chen, S., et al., *Location of a Folding Protein and Shape Changes in Groel-Groes Complexes Imaged by Cryoelectron Microscopy.* Nature, 1994. **371**(6494): p. 261-264.
7. Roh, S.H., et al., *Subunit conformational variation within individual GroEL oligomers resolved by Cryo-EM.* Proceedings of the National Academy of Sciences of the United States of America, 2017. **114**(31): p. 8259-8264.
8. Gestaut, D., et al., *The Chaperonin TRiC/CCT Associates with Prefoldin through a Conserved Electrostatic Interface Essential for Cellular Proteostasis.* Cell, 2019. **177**(3): p. 751-+.
9. Ditzel, L., et al., *Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT.* Cell, 1998. **93**(1): p. 125-138.
10. Skjaerven, L., et al., *Dynamics, flexibility, and allostery in molecular chaperonins.* FEBS Lett, 2015. **589**(19 Pt A): p. 2522-32.
11. de Groot, B.L., G. Vriend, and H.J.C. Berendsen, *Conformational changes in the chaperonin GroEL: New insights into the allosteric mechanism.* Journal of Molecular Biology, 1999. **286**(4): p. 1241-1249.
12. Pereira, J.H., et al., *Mechanism of nucleotide sensing in group II chaperonins.* EMBO J, 2012. **31**(3): p. 731-40.
13. Chaston, J.J., et al., *Structural and Functional Insights into the Evolution and Stress Adaptation of Type II Chaperonins.* Structure, 2016. **24**(3): p. 364-74.

14. Zhang, J., et al., *Mechanism of folding chamber closure in a group II chaperonin.* Nature, 2010. **463**(7279): p. 379-83.
15. Douglas, N.R., et al., *Dual Action of ATP Hydrolysis Couples Lid Closure to Substrate Release into the Group II Chaperonin Chamber.* Cell, 2011. **144**(2): p. 240-252.
16. Zhang, J.J., et al., *Cryo-EM Structure of a Group II Chaperonin in the Prehydrolysis ATP-Bound State Leading to Lid Closure.* Structure, 2011. **19**(5): p. 633-639.
17. Schoehn, G., et al., *Domain rotations between open, closed and bullet-shaped forms of the thermosome, an archaeal chaperonin.* Journal of Molecular Biology, 2000. **301**(2): p. 323-332.
18. Bigotti, M.G., S.R. Bellamy, and A.R. Clarke, *The asymmetric ATPase cycle of the thermosome: elucidation of the binding, hydrolysis and product-release steps.* J Mol Biol, 2006. **362**(4): p. 835-43.
19. Bigotti, M.G. and A.R. Clarke, *Chaperonins: The hunt for the Group II mechanism.* Arch Biochem Biophys, 2008. **474**(2): p. 331-9.
20. Bigotti, M.G. and A.R. Clarke, *Cooperativity in the thermosome.* J Mol Biol, 2005. **348**(1): p. 13-26.
21. Paul, D.M., et al., *Internal (His)(6)-tagging delivers a fully functional hetero-oligomeric class II chaperonin in high yield.* Sci Rep, 2016. **6**: p. 20696.
22. Shoemark, D.K., et al., *Intraring allostery controls the function and assembly of a hetero-oligomeric class II chaperonin.* FASEB J, 2018. **32**(4): p. 2223-2234.
23. Fenton, W.A., et al., *Residues in Chaperonin Groel Required for Polypeptide Binding and Release.* Nature, 1994. **371**(6498): p. 614-619.
24. Grant, T. and N. Grigorieff, *Measuring the optimal exposure for single particle cryo-EM using a 2.6 angstrom reconstruction of rotavirus VP6.* Elife, 2015. **4**.
25. Rohou, A. and N. Grigorieff, *CTFFIND4: Fast and accurate defocus estimation from electron micrographs.* Journal of Structural Biology, 2015. **192**(2): p. 216-221.
26. Ripstein, Z.A. and J.L. Rubinstein, *Processing of Cryo-EM Movie Data.* Methods Enzymol, 2016. **579**: p. 103-24.

27. Brilot, A.F., et al., *Beam-induced motion of vitrified specimen on holey carbon film.* Journal of Structural Biology, 2012. **177**(3): p. 630-637.

28. Scheres, S.H.W., *Classification of Structural Heterogeneity by Maximum-Likelihood Methods.* Methods in Enzymology, Vol 482: Cryo-Em, Part B: 3-D Reconstruction, 2010. **482**: p. 295-320.

29. Scheres, S.H.W., *Processing of Structurally Heterogeneous Cryo-EM Data in RELION.* Methods in Enzymology, Vol 482: Cryo-Em, Part B: 3-D Reconstruction, 2016. **579**: p. 125-157.

30. Scheres, S.H.W. and S.X. Chen, *Prevention of overfitting in cryo-EM structure determination.* Nature Methods, 2012. **9**(9): p. 853-854.

31. Fernandez, J.J., et al., *Sharpening high resolution information in single particle electron cryomicroscopy.* Journal of Structural Biology, 2008. **164**(1): p. 170-175.

32. Scheres, S.H. *Single-particle processing in RELION-3.0*. 2018.

33. Glaeser, R.M., *Specimen Behavior in the Electron Beam.* Methods Enzymol, 2016. **579**: p. 19-50.

34. Cong, Y. and S.J. Ludtke, *Single Particle Analysis at High Resolution.* Methods in Enzymology, Vol 482: Cryo-Em, Part B: 3-D Reconstruction, 2010. **482**: p. 211-235.

35. *CTFFIND 4*. 2017 [cited 2019 19/07/2019]; Available from: http://grigoriefflab.janelia.org/ctffind4.

36. Klumpp, M. and W. Baumeister, *The thermosome: archetype of group II chaperonins.* FEBS Letters, 1998. **430**(1): p. 73-77.

37. Chen, D.H., et al., *Visualizing GroEL/ES in the Act of Encapsulating a Folding Protein.* Cell, 2013. **153**(6): p. 1354-1365.

38. Levy, E.D., et al., *3D complex: A structural classification of protein complexes.* Plos Computational Biology, 2006. **2**(11): p. 1395-1406.

39. Rosenthal, P.B. and R. Henderson, *Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy.* J Mol Biol, 2003. **333**(4): p. 721-45.

40. Nitsch, M., et al., *The thermosome: alternating alpha and beta-subunits within the chaperonin of the archaeon Thermoplasma acidophilum.* J Mol Biol, 1997. **267**(1): p. 142-9.

Chapter 3: Modelling the structure of the *T. acidophilum* Thermosome

3.1 Introduction

In the previous chapter, two sub 4 Å electron density maps were calculated for the *T. acidophilum* thermosome-ATP-AlF_x complex. We have captured the system in two distinct conformations, one a more closed and the other a more open conformation.

To gain a deeper understanding of the biological states that these two maps represent, the transition between states, and the mechanism of ATP hydrolysis, we use atomic models to interpret the electron densities. This chapter focuses on identifying models which best describe the EM maps using both molecular dynamics flexible fitting (MDFF) and simulated annealing molecular dynamics (FlexEM).

The key aims are to:

1. Describe global changes in system: dimensions, cavity size *etc.*
2. Identify differences between subunits in the EM maps in order to assign alpha and beta.
3. Identify the differences in motion of alpha and beta subunits.
4. Identify the nucleotide binding pocket and its occupancy state.

3.2 Methods

3.2.1 Molecular dynamics flexible fitting

Molecular dynamics flexible fitting, or MDFF [1], takes an initial molecular model, usually a crystal structure, and fits it into an electron density map. In order to do this the density map is converted into a potential energy field. Regions of greatest electron density correspond to energy minima. The potential is then converted into a force acting on atoms in the model. This force can be scaled by a user defined ‘force scaling’ constant ζ, or G factor as it is referred to in this thesis.

This is combined with a molecular dynamics simulation such that the motion of the molecule is effectively biased into a conformation which matches the electron density map. Over fitting is prevented by the introduction of harmonic restraints to the internal angles of the protein. Progress of the simulation is tracked by monitoring the root mean squared deviation (RMSD) of the atomic positions relative to their initial positions. Once this converges the simulation can be considered complete.

In this work, MDFF simulations were performed using the VMD (Visual Molecular Dynamics) software package [2].

All simulations were run for 500 frames corresponding to 500 ps. The temperature was set to 300 K and G-factors tested were 0.1, 0.3, 0.5 and 1.0.

3.2.2 Simulated annealing molecular dynamics – FlexEM

Similar to MDFF, the aim of FlexEM [3] is to refine an existing model into an EM density map. A flow diagram of a simulation is shown in Figure 3-1.

Initially a density map and an initial model are input into the workflow. An inbuilt implementation of ribfind [4] identifies rigid bodies in the input model. It does this by first identifying secondary structure elements and grouping them based on their proximities. These groups are then fitted as rigid bodies by a random sampling (Monte-Carlo) method (stage 1 in the figure). An iterative fitting step then follows in which the rigid bodies are fitted and then reduced in size, until rigid bodies consist of single residues (stage 2). Finally, the model undergoes a cycle of simulated annealing (heating up and cooling down) molecular dynamics steps to reach a stable conformation (stage 3).

Each FlexEM run in this work was performed with 20 iterations of simulated annealing molecular dynamics although they tended to converge within the first 5.

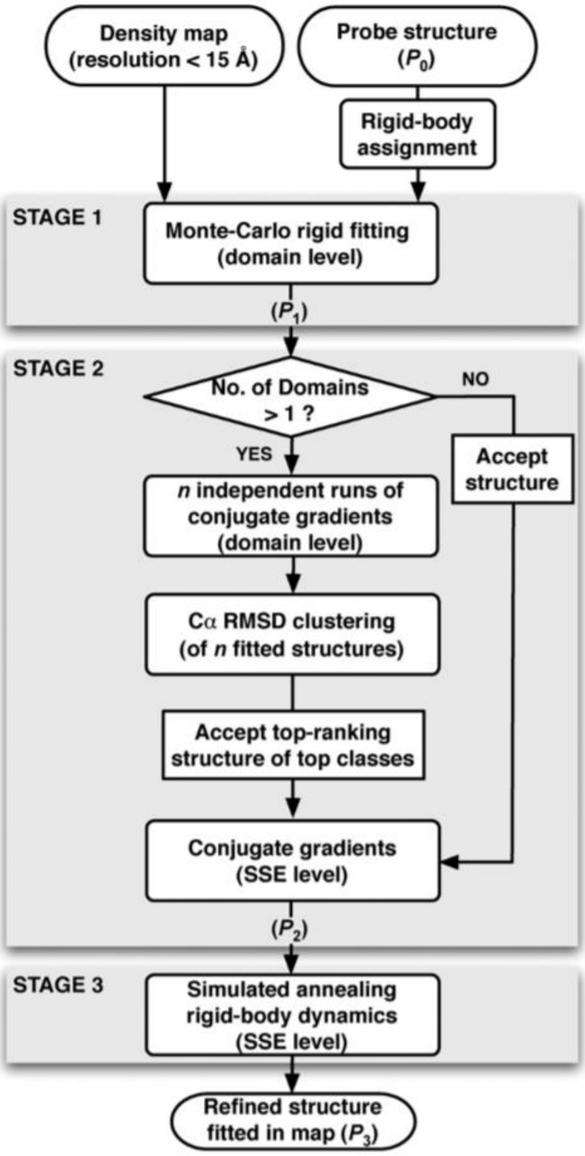


Figure 3-1: FlexEM overview. Rigid bodies from a model are fitted into an EM map by random sampling (stage 1) Rigid bodies are iteratively fitted and reduced in size (stage 2). Molecular dynamics simulated annealing leads to final structure refinement. (stage 3). Reproduced from [3]

3.3 Results

3.3.1 Global fitting

Molecular dynamics flexible fitting (MDFF) was used to show global fitting of the high-resolution crystal structure by Ditzel *et. al*. [5] into both of the electron density maps calculated in Chapter 2:.

The simulation relies on having starting models which are reasonably well fitted into the density. The difference in conformation between the closed crystal structure and the 'open' mutant map meant that the former could not be used as a starting point for this simulation.

For this reason, the crystal structure was modified prior to running molecular dynamics simulations for the mutant map. This was perfomed by Dr Deborah Shoemark and Dr Richard Sessions. Briefly, each subunit was divided into several domains which were then manually fitted as rigid bodies. Finally, the connections between domains were regenerated and the His-tag inserted into the alpha subunit.

The starting model is shown in Figure 3-2(A) fitted into the map. 95% of the known residues are present in the model. The missing 25 residues are those at the tip of the apical domain which was not visible in the EM density. This region has been suggested to be unstructured in the open state and therefore we would not expect to resolve it in our mutant map [6].

Outputs from an initial MDFF run for the mutant map are shown in Figure 3-2(B). The run was repeated 4 times with a range of force scaling values: G = 0.1 (red), 0.3 (yellow), 0.5 (green) and 1 (blue). The run with the lowest restraint on the model, *i.e*. the most flexible model shown in red, did not result in a well fitted structure, with arrows in the figure illustrating where loops in this model are considerably worse than the others. The other three G values all gave similar results. All four scaling values were used for each run, however in the following figures, only the output of the G = 0.3 is shown for clarity.

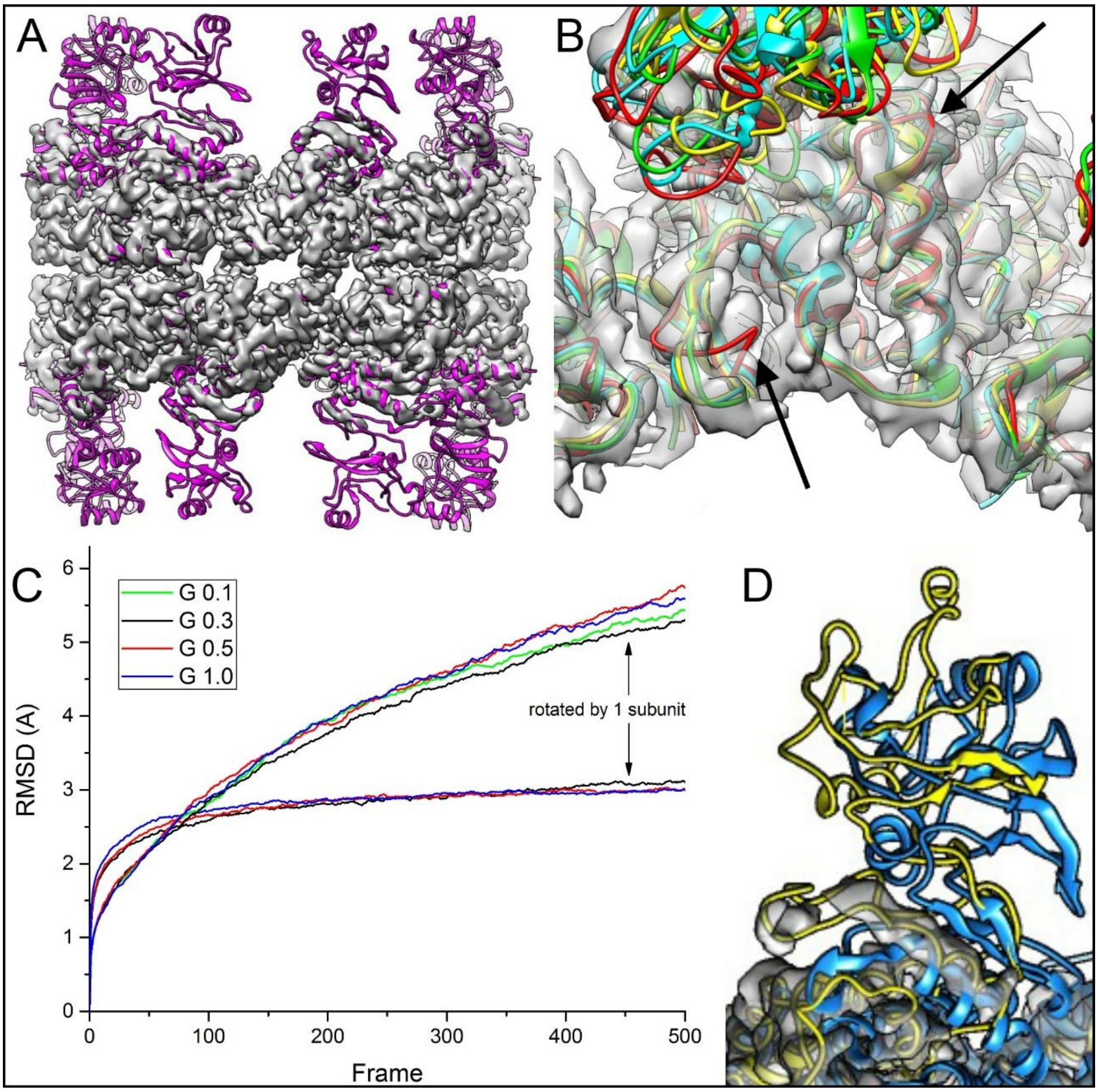


Figure 3-2: Initial MDFF for mutant map. (A) starting model fitted into EM density map. (B) Outputs for G values: red = 0.1, yellow = 0.3, green = 0.5, blue = 1. Arrows indicate poor fitting of red model. (C) RMSD of atoms from their initial positions for two sets of runs rotated by one subunit relative to each other. One set converged while the other did not over the simulation timeframe. (D) Example outputs from the runs in (C) showing collapse of the apical domain.

The initial aim of the MDFF analysis was to determine if there were interpretable differences between the alpha and beta subunits. The fitting was performed twice for each map, rotating the model by one subunit in between runs. It was hoped that differences between subunits could be quantified. It was found that only one of two fits ran to convergence over the time frame modelled. This is shown in Figure 3-2(C) where one set of runs converged in fewer than 100 frames while the other did not in the full 500. Inspection of the output models indicated that the apical domains in the rotated run were collapsing towards the equatorial domains and not improving their fit to the map. This can be seen in Figure 3-2(D) where the apical domain of the blue model has collapsed down towards the equatorial domain. Thus, it was not possible to distinguish between subunits using the output of these simulations as one had become unstable.

For this reason, it was decided to further remove the apical domains from the starting model, as focussing on the better-resolved regions of the map may highlight the differences between subunits. The models were truncated (alpha Gly215-Asn365; beta Ile198 – Ala369) for both the wild type and mutant models and are shown in Figure 3-3(A) alongside the original mutant model for reference.

In order to ensure that the newly created termini were well behaved through the modelling, the separations between C-alpha atoms were measured. These are shown in Figure 3-3(B) with the measured separations shown for the initial model, the output of the control run and the output of the truncated run. The results show that while the mean separation did change for both subunits, neither was outside of the margin of error for the control. This, in conjunction with visual inspection of the output models, confirmed that truncation of the models did not cause unacceptable changes to the structure.

Results of the simulations are shown in Figure 3-3(C-F). In these cases, all models ran to convergence over the timeframe of the simulation (D), but neither conformation appeared to be better fitted than the other. (C) shows an example of outputs from wild type fits in which the red and blue were rotated relative to each other. The coloured arrows indicate regions where the model of the corresponding colour shows better fitting than the other. (E) and (F) show the same region from mutant runs, indicating a region in which the secondary structure has been lost, resulting in poor fitting in both runs.

In general, the secondary structure fitting of the models into the maps were good with most of the secondary structure being well fitted into the density. However, these small deviations sticking out of the density, combined with the fact that the alpha and beta subunits are extremely similar, meant that it was not possible to determine which subunit was which based on this global fitting method.

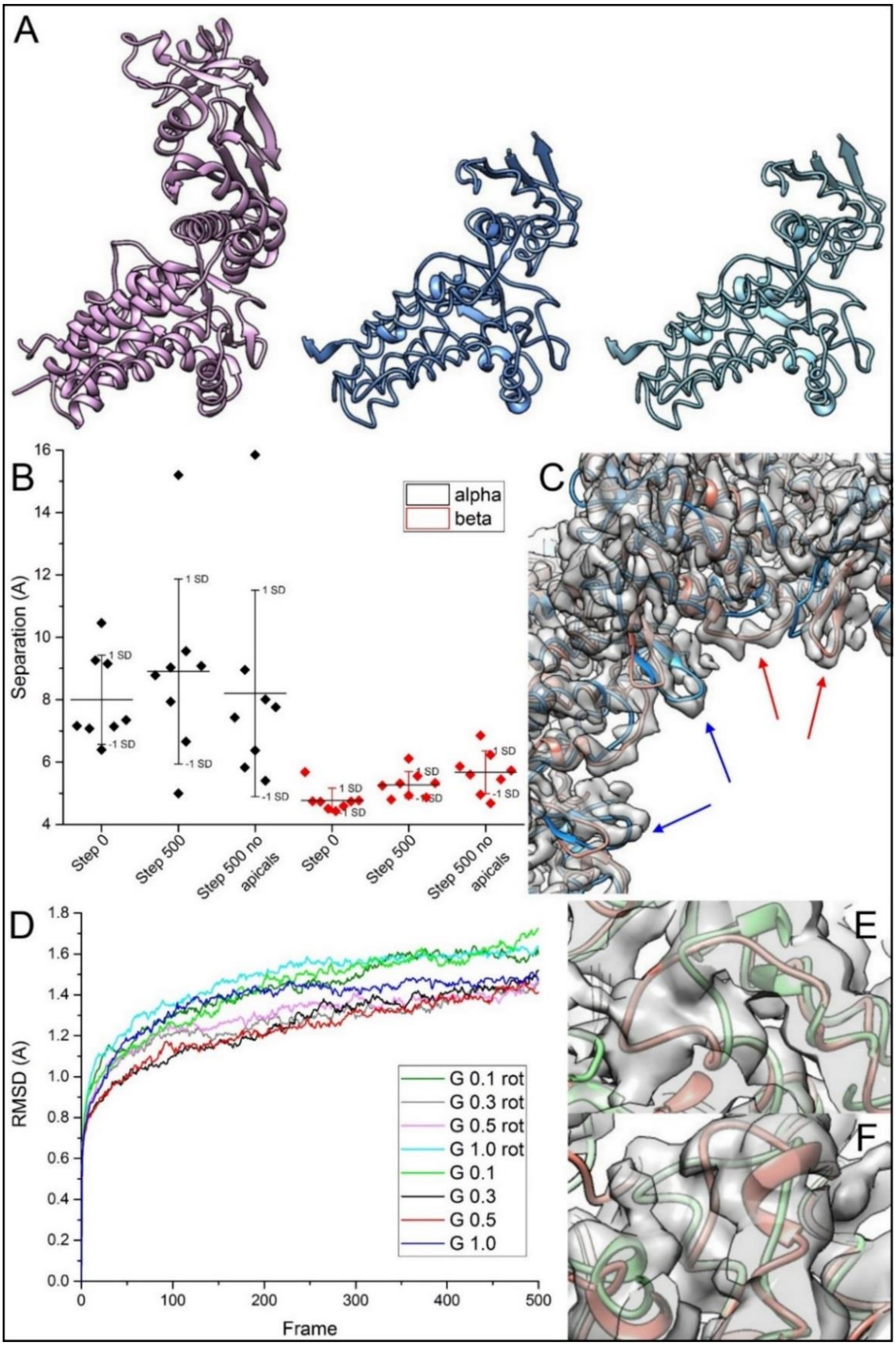


Figure 3-3: MDFF of truncated models. (A) Truncated wild type (middle) and mutant (right) models shown with the starting mutant model for reference (left). The two truncated models are the same as the equatorial domain was not modified from the crystal structure for mutant fitting. (B) Separation of new termini residues for a G=0.1 mutant simulation. 'Step 0' is starting model, 'step 500' is un-truncated model and 'step 500 no apicals' is truncated model. (C) Wild type models rotated one subunit relative to each other. Red and blue arrows indicate which map is better fitted at each point. (D) Convergence of truncated models shown by RMSD. (E-F) mutant models rotated one subunit to each other showing region of poor fitting.

Although not conclusive in subunit assignment, MDFF was used to gain insight into the motion of the molecule. The output of a truncated wild type simulation was used as a starting model for fitting into the mutant EM maps.

The wild type model was first rigid body fitted into the mutant map and MDFF performed. Since the apical regions had already been removed, the starting fit was sufficiently good for this to run well. Taking a series of frames from throughout the run prior to convergence, we could see the secondary structures being pulled into the density of the map. There appeared to be an overall rocking motion which was not limited to the intermediate domain, but was also seen in the equatorial domain. This is seen in Figure 3-4 in which the starting model is shown in red, with sequential intermediate stages in orange, yellow and finally yellow/green.

The side view shown (Figure 3-4(A)) illustrates that while the intermediate domain appears to rock backwards, the equatorial domain shifts into the plane of the page. (B) shows a view of the three equatorial alpha helices seen in top views of the map demonstrating the shift in the equatorial domain.

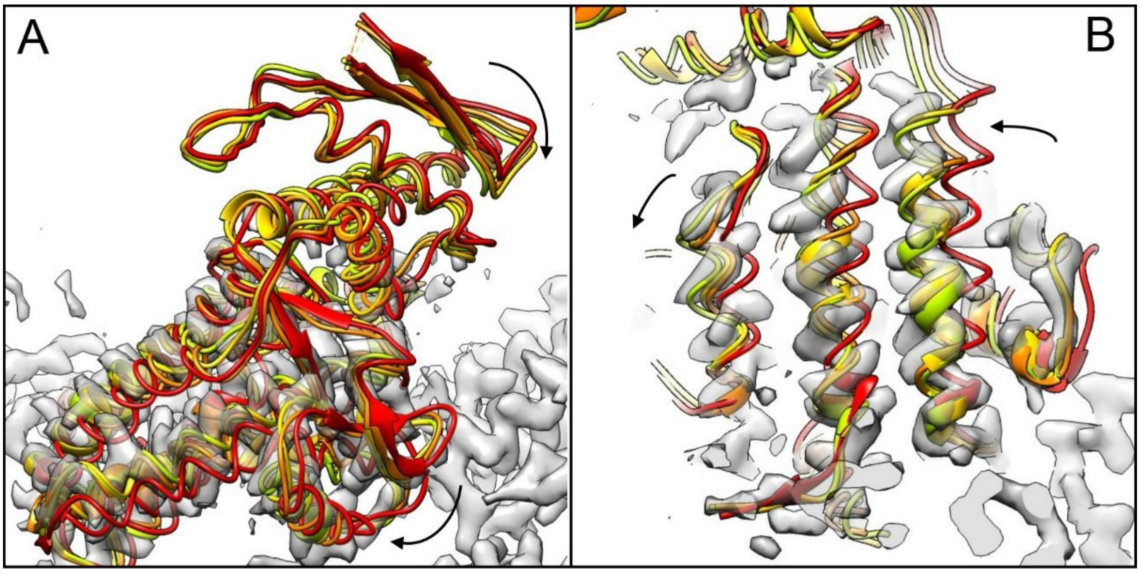


Figure 3-4: MDFF of wild type model into mutant map. (A) side view and (B) top view of H4, H18 and H6. Arrows indicate direction of motion of model. Starting model is red (mutant model), subsequent time points shown in orange, yellow and yellow-green to show improvement in fitting.

This appears to correspond well with the results from Mm-cpn which shows tilting in the equatorial domain between prehydrolysis and hydrolysis states [7]. While motion in the intermediate region was expected based on the change in overall shape of the maps, the extent of the motion in the equatorial domain was surprising as the closing mechanism of GroEL only results in small shifts within this region [8, 9].

This simulation gave some initial insight into the overall conformational changes between the different states of the thermosome, however, even in the best models there were regions which were not well fitted into the density. Furthermore, the dynamic nature of the modelling leads to different results for each subunit which is difficult to interpret. While it is possible to symmetrise these simulations, it was decided to proceed with an alternative method.

3.3.2 Individual Subunit Fitting

A new full His-tagged alpha subunit was created using the SWISS-MODEL homology modelling service [10]. This online programme takes an input sequence and identifies a target structure based on a search of the Protein Data Bank (PDB). It then aligns the sequences and models inclusions before finally assessing the quality of the output model.

Figure 3-5 shows the output structure of the new alpha subunit with colours indicating how similar it is to the crystal structure. The His-tag insertion is indicated with a black arrow. In addition to the His-tag itself, there were only two regions which showed significant differences to the crystal structure and a third which shows a small deviation (labels). Each of these are regions which we would expect to be flexible as they are found in unstructured loops. This model was therefore accepted as a good starting point for the fitting.

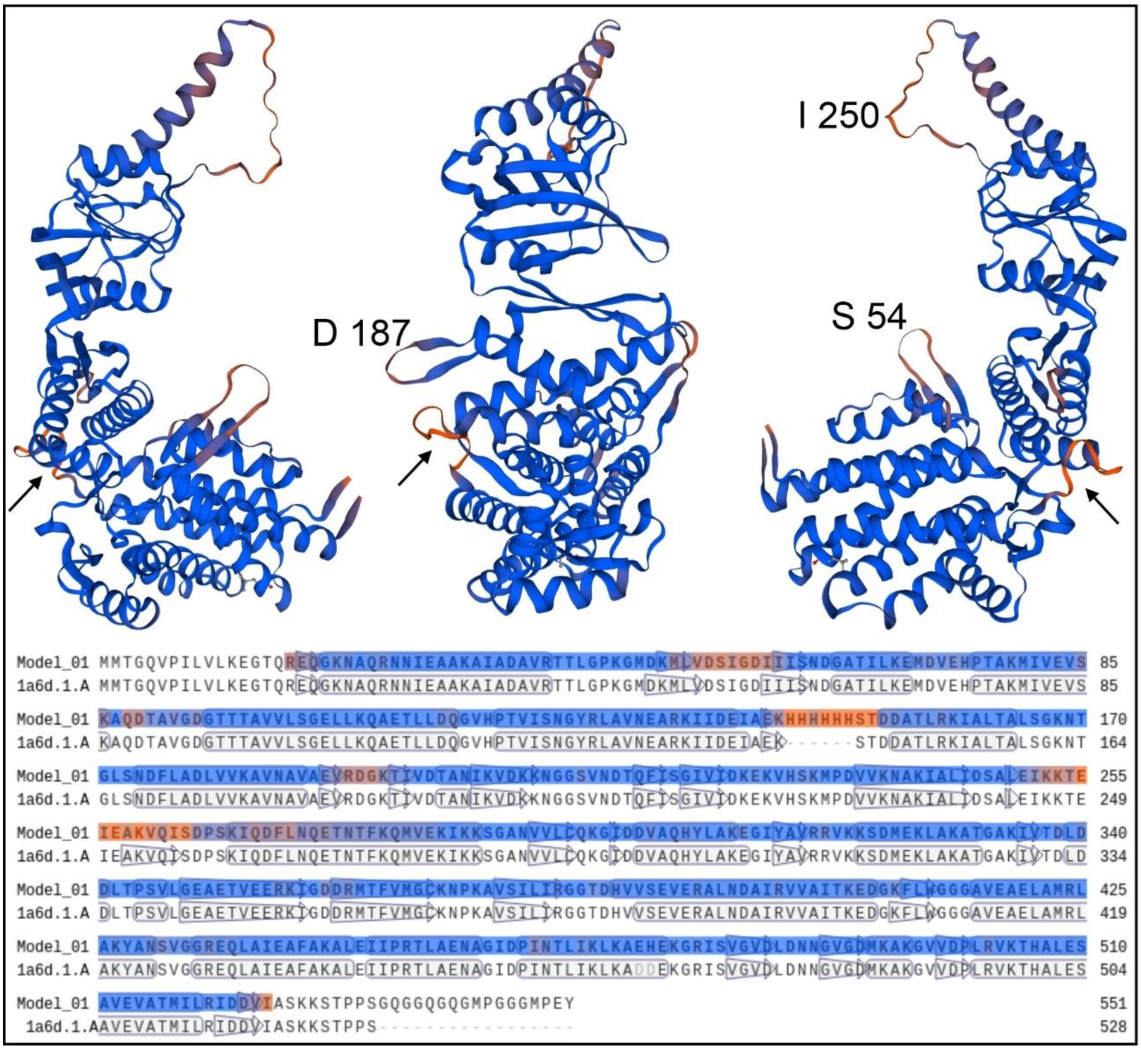


Figure 3-5: Insertion of His-tag into alpha subunit. Top – three views of the new alpha subunit. The His-tag is indicated by a black arrow in each. Labels correspond to regions with poor similarity to the crystal structure, as shown in the sequence alignment (bottom).

3.3.2.1 **Cutting out individual subunits from the EM map**

Before running FlexEM, densities corresponding to two individual subunits were extracted from each map. This was done using the Segger tool in Chimera [11] as shown in Figure 3-6.

The starting model was rigid body fitted into the EM map using the fit-in-map tool (A) and the threshold set so that as much of the model as possible was enclosed within the density (B). The map was then segmented without any smoothing or grouping steps (C). The segments were grouped according to whether they overlapped with the estimated volume of the fitted model (D). Since the model used had not yet been flexibly fitted, the resolution of this was set to 8 Å to minimise bias. The density enclosed within these grouped segments was then extracted (E) and the single subunit can be seen at low threshold (F) with the crystal structure fitted.

In both cases, an initial fitting run was performed to approximately fit the model, and then the extraction process repeated with both this and the crystal structure. By using both model structures, it was ensured that all densities resulting from the protein would be extracted using the above method. This was particularly important for the mutant as the large changes in conformation could have resulted in density from the protein being missed if extraction was only performed relative to the crystal structure.

Two adjacent subunits were extracted in this manner from each map. Since D4 symmetry had been applied to the map (see Chapter 2:) this should ensure that one alpha and one beta subunit had been extracted.

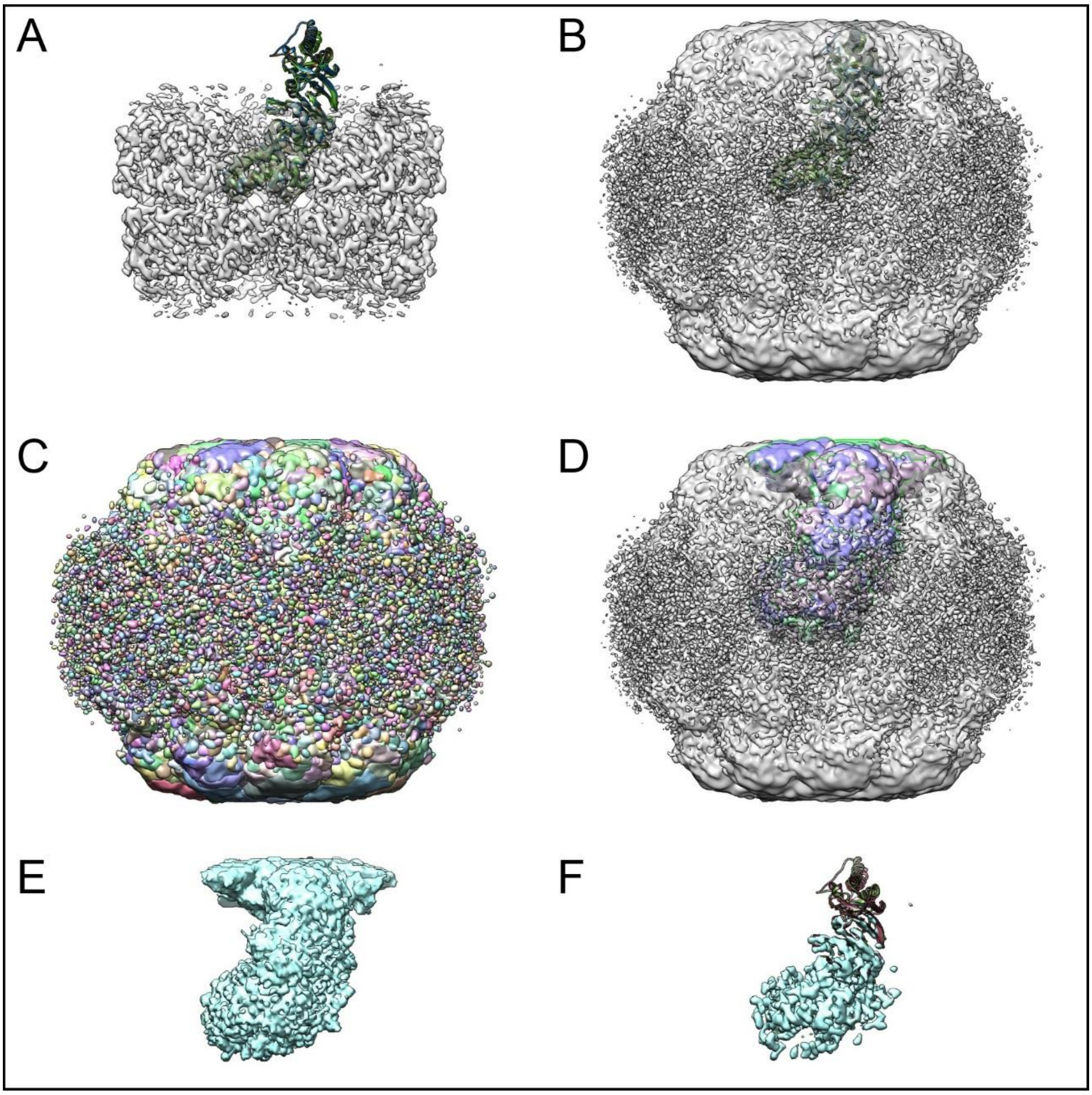


Figure 3-6: Extraction of a single subunit from wild type map. (A-B) crystal structure is rigid body fitted into the map. (C) Segmented Map. (D) Segments which overlap with predicted volume of model are selected. (E) Density from within the selected segments is extracted leaving a map corresponding to a single subunit. (F) The same extracted density at a lower threshold showing the fitted model from (A).

3.3.2.2 Flexible Fitting

Fitting of the crystal structure alpha and beta subunits into individual model subunits was performed using the FlexEM software implementation in ccpem [3].

In order to discern which subunit was which, four fits were performed for each of the structures, *i.e*. alpha in subunits 1 and 2, and beta in subunits 1 and 2. The rigid body assignment during the initial steps of the simulation meant that manual manipulation of the structure (as performed for MDFF) was not required.

Each run consisted of 20 iterations, although in general convergence was reached within the first 5. An example output of several iterations of a wild type FlexEM run is shown in Figure 3-7 with input (red) and output (dark blue) models, and three intermediates. (A) shows two views of the overall subunit along with indications of locations of the magnified views of alpha helices fitting into the EM density (B-D).

The initial model is clearly not well fitted into the density, however in most regions each of the other models has been pulled into the density. This is shown for several secondary structure elements, illustrating both the successful fitting of backbone, and the rapid convergence.

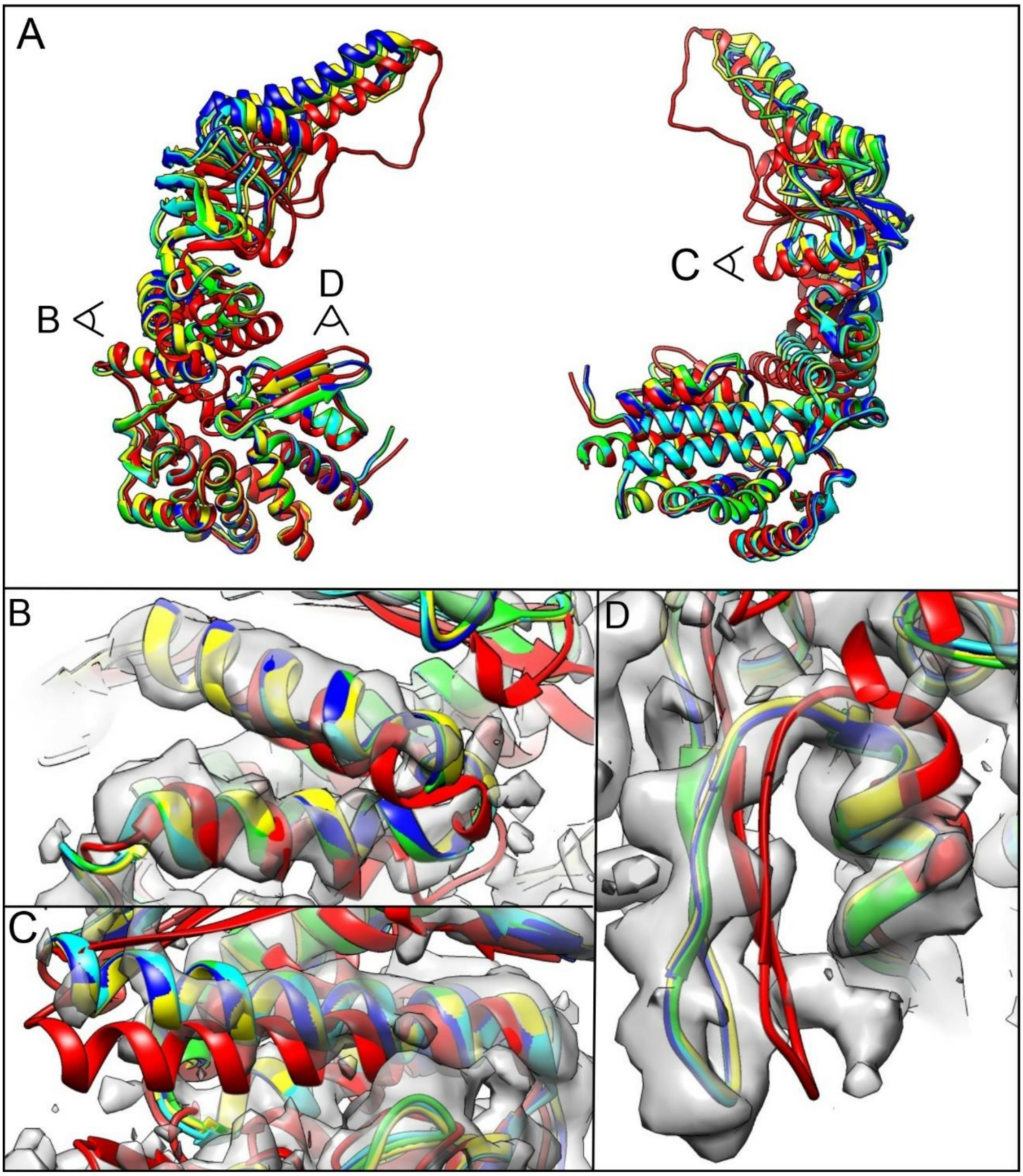


Figure 3-7: FlexEM fitting with intermediate iterations. (A) side views of full subunits, (B-D) magnified views of the alpha helical regions labelled in panel A with experimental map. Initial model shown in red, final model in dark blue and intermediate iterations 5, 10 and 15 in yellow, green and light blue respectively.

3.3.2.3 Identification of the His-tag

It was hypothesised that the presence of the His-tag may facilitate the identification of the alpha subunit as it should provide extra, low resolution density on the outer surface of the thermosome. The fitted models were used to identify the approximate location of the His-tag in the EM density maps. Local resolution maps of the subunits were then inspected to see if the alpha subunit could be identified in this way. Figure 3-8 shows these regions for each of the extracted subunit densities. In both cases, no clear feature could be identified. Local resolutions were similar between subunits and features varied between every subunit in the full map *i.e*. no consistent alternating pattern going around a ring was visible.

The lack of conclusive result here is not entirely surprising as we would expect the His-tag to be flexible. Furthermore, its position within a loop region will increase the flexibility such that the density it provides could be averaged out and not be seen in the map.

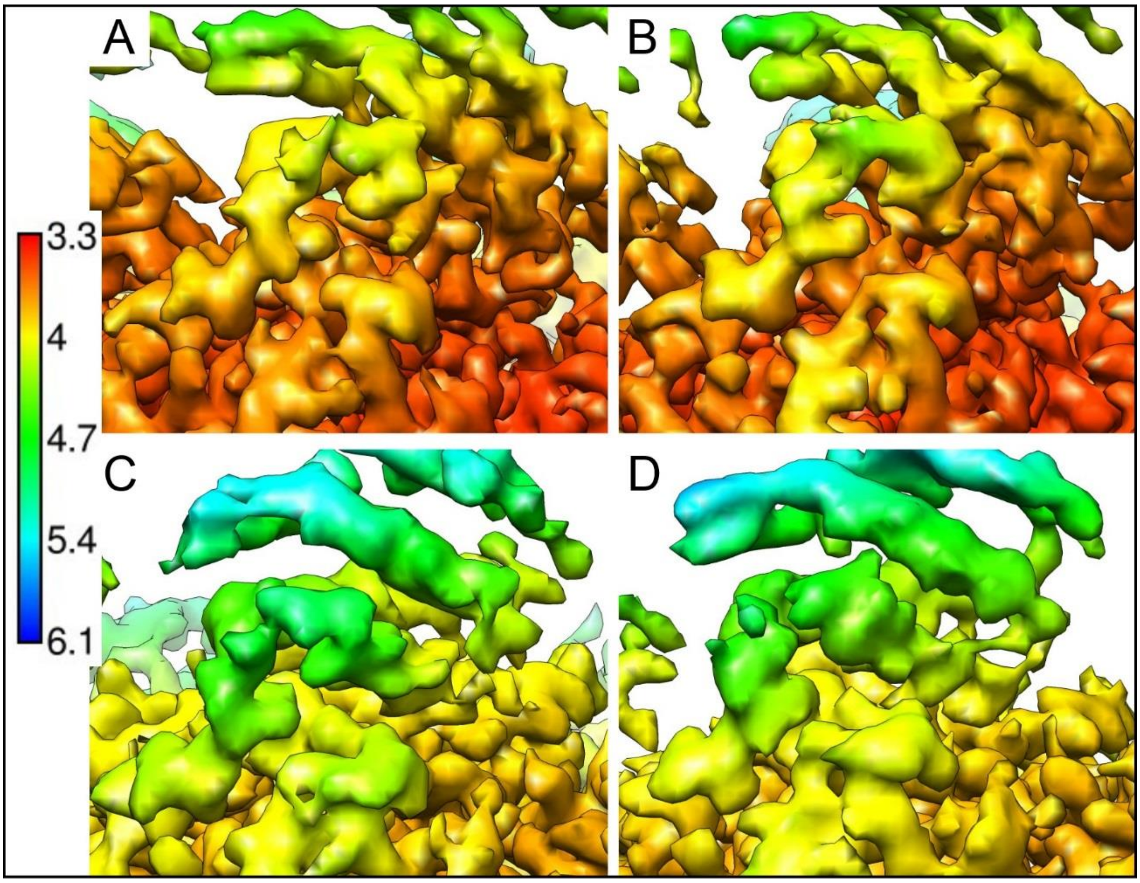
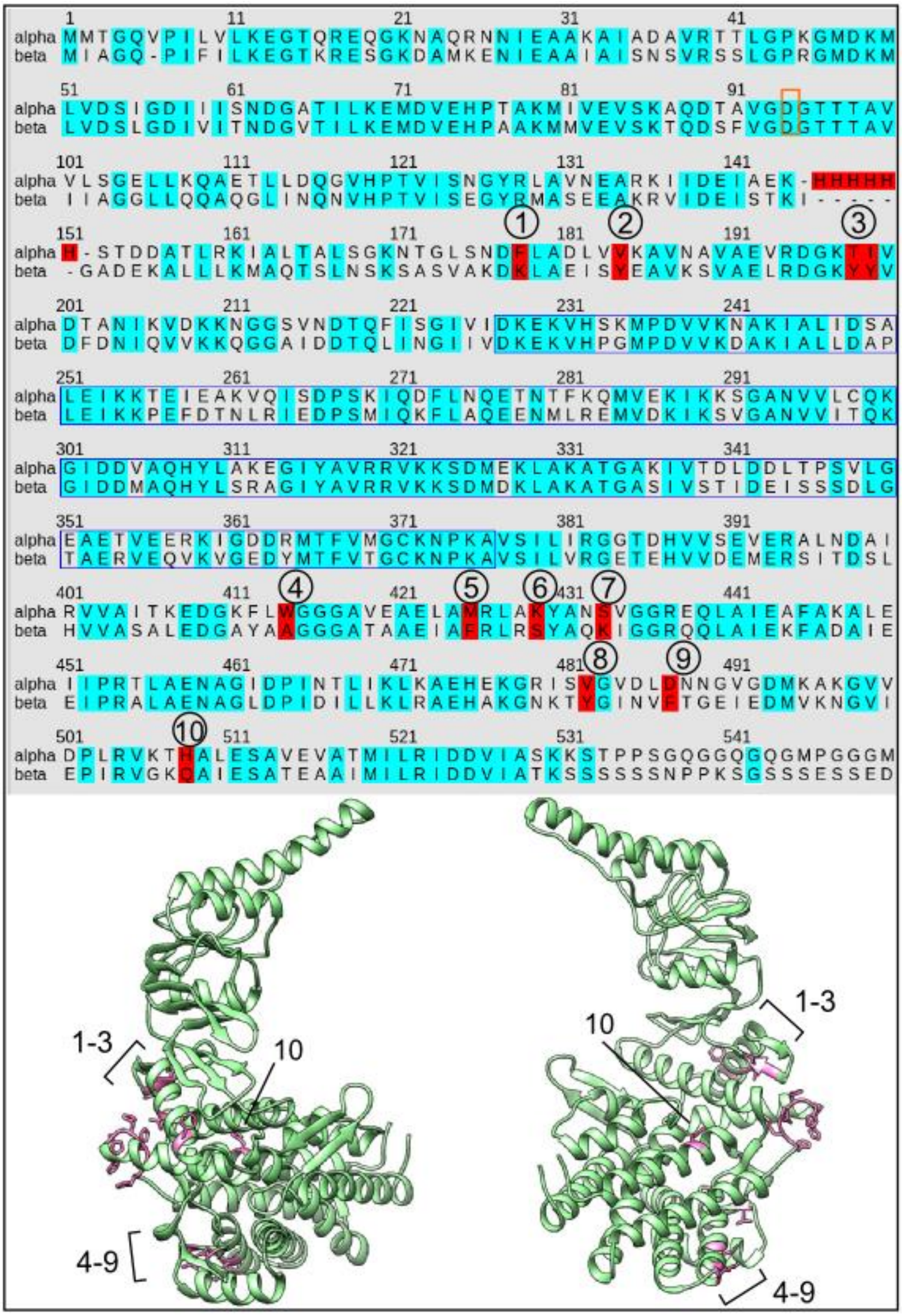


Figure 3-8: Local resolution maps at His-tag locations in the wild type (A-B) and mutant (C-D) EM density maps. Scale on the left indicates local resolution in Å.

3.3.2.4 Comparison of individual residues



```
      1         11        21        31        41
alpha MMTGQVPILVLKEGTQREQGKNAQRNNIEAAKAIADAVRTTLGPKGMDKM
beta  MIAGQ-PIFILKEGTKRESGKDAMKENIEAAIAISNSVRSSLGPRGMDKM

      51        61        71        81        91
alpha LVDSIGDIIISNDGATILKEMDVEHPTAKMIVEVSKAQDTAVGDGTTTAV
beta  LVDSLGDIVITNDGVTILKEMDVEHPAAKMMVEVSKTQDSFVGDGTTTAV

      101       111       121       131       141
alpha VLSGELLKQAETLLDQGVHPTVISNGYRLAVNEARKIIDEIAEK-HHHHH
beta  IIAGGLLQQAQGLINQNVHPTVISEGYRMASEEAKRVIDEISTKI-----

      151       161       171       181       191
alpha H-STDDATLRKIALTALSGKNTGLSNDFLADLVVKAVNAVAEVRDGKTIV
beta  -GADEKALLLKMAQTSLNSKSASVAKDKLAEISYEAVKSVAELRDGKYYV

      201       211       221       231       241
alpha DTANIKVDKKNGGSVNDTQFISGIVIDKEKVHSKMPDVVKNAKIALIDSA
beta  DFDNIQVVKKQGGAIDDTQLINGIIVDKEKVHPGMPDVVKDAKIALLDAP

      251       261       271       281       291
alpha LEIKKTEIEAKVQISDPSKIQDFLNQETNTFKQMVEKIKKSGANVVLCQK
beta  LEIKKPEFDTNLRIEDPSMIQKFLAQEENMLREMVDKIKSVGANVVITQK

      301       311       321       331       341
alpha GIDDVAQHYLAKEGIYAVRRVKKSDMEKLAKATGAKIVTDLDDLTPSVLG
beta  GIDDMAQHYLSRAGIYAVRRVKKSDMDKLAKATGASIVSTIDEISSSDLG

      351       361       371       381       391
alpha EAETVEERKIGDDRMTFVMGCKNPKAVSILIRGGTDHVVSEVERALNDAI
beta  TAERVEQVKVGEDYMTFVTGCKNPKAVSILVRGETEHVVDEMERSITDSL

      401       411       421       431       441
alpha RVVAITKEDGKFLWGGGAVEAELAMRLAKYANSVGGREQLAIEAFAKALE
beta  HVVASALEDGAYAAGGGATAAEIAFRLRSYAQKIGGRQQLAIEKFADAIE

      451       461       471       481       491
alpha IIPRTLAENAGIDPINTLIKLKAEHEKGRISVGVDLDNNGVGDMKAKGVV
beta  EIPRALAENAGLDPIDILLKLRAEHAKGNKTYGINVFTGEIEDMVKNGVI

      501       511       521       531       541
alpha DPLRVKTHALESAVEVATMILRIDDVIASKKSTPPSGQGGQGQGMPGGGM
beta  EPIRVGKQAIESATEAAIMILRIDDVIATKSSSSSSNPPKSGSSSESSED
```

Figure 3-9: Sequence alignment of alpha and beta subunits. Residues in low resolution regions of the map are outlined in blue. Residues which are identical between the subunits are highlighted in cyan and those with very different structural characteristics highlighted in red.

The sequences of the alpha and beta subunits have a 60% identity and the secondary structure arrangement is almost identical in the crystal structure. The sequence alignment is shown in Figure 3-9 with identical residues highlighted in cyan. The region outlined in blue represents the upper-intermediate and apical domains, which were resolved to a resolution at which we would not expect side chains to be defined, and therefore were not used to distinguish between subunits. Finally, highlighted in red are pairs of residues which we may expect to be sufficiently different to be distinguishable in the electron density maps. In particular, large residues such as tryptophans or phenylalanines in one subunit, aligned with small residues such as glycine or alanine in the other. These regions were highlighted in the output maps for each FlexEM run and compared relative to the subunit densities to determine which was which. In Figure 3-9 and Figure 3-10 the models are coloured according to the scheme shown in Table 3-1. The two subunit densities are coloured grey and yellow to distinguish from each other, but it should be noted that wild type grey may not relate to mutant grey. In each case all four model fits are shown in the two densities separately for ease of comparison.

Maps were initially viewed in 3D using sequential stereo mode in Chimera and 3D viewing glasses over a range of thresholding values. Once regions had been viewed and compared in this way, 2D images were recorded.

	Backbone	Residue
Alpha	Green	Pink
Beta	Blue	Orange

Table 3-1: Colour scheme for individual residue comparison.

Mutant comparison

The residue comparison for the mutant data is shown in Figure 3-10 approximately ordered according to how convincing the conclusion drawn from each was. In the following section, numbers relate to the highlighted residues from the sequence alignment.

These views show density in the grey for the beta and yellow for the alpha. In (A) there is clear density in the grey for the tyrosine residue (2) of the beta subunit which is not visible in the yellow density demonstrating one of the most convincing differences. (B) shows the lysine residues (6 and 7) which are on different subunits. (C) and (D) show views of the same residue (3) at different map thresholds. At the higher threshold (C) it appears that the grey density more clearly describes the Tyr group. However, the lower threshold in (D) shows a less focused view which also reveals a neighbouring lysine group which is also on the beta subunit. Combined these two residues could perhaps be interpreted as fitting better in the yellow density. (E) shows two regions of density which are very similar between the two maps. However, the grey appears to have an additional narrow region for the lysine in the beta (1).

(F) shows density for an arginine residue which is slightly clearer in the yellow density. This weak evidence suggests an opposite arrangement to the previous examples. (G) (9) shows a region of extra density in the grey to the right, which would be available for the phenylalanine group, however the fitting suggests that this may go into the other region to the left of this. It is possible that this is a small region of noise that was included in the grey density extraction.

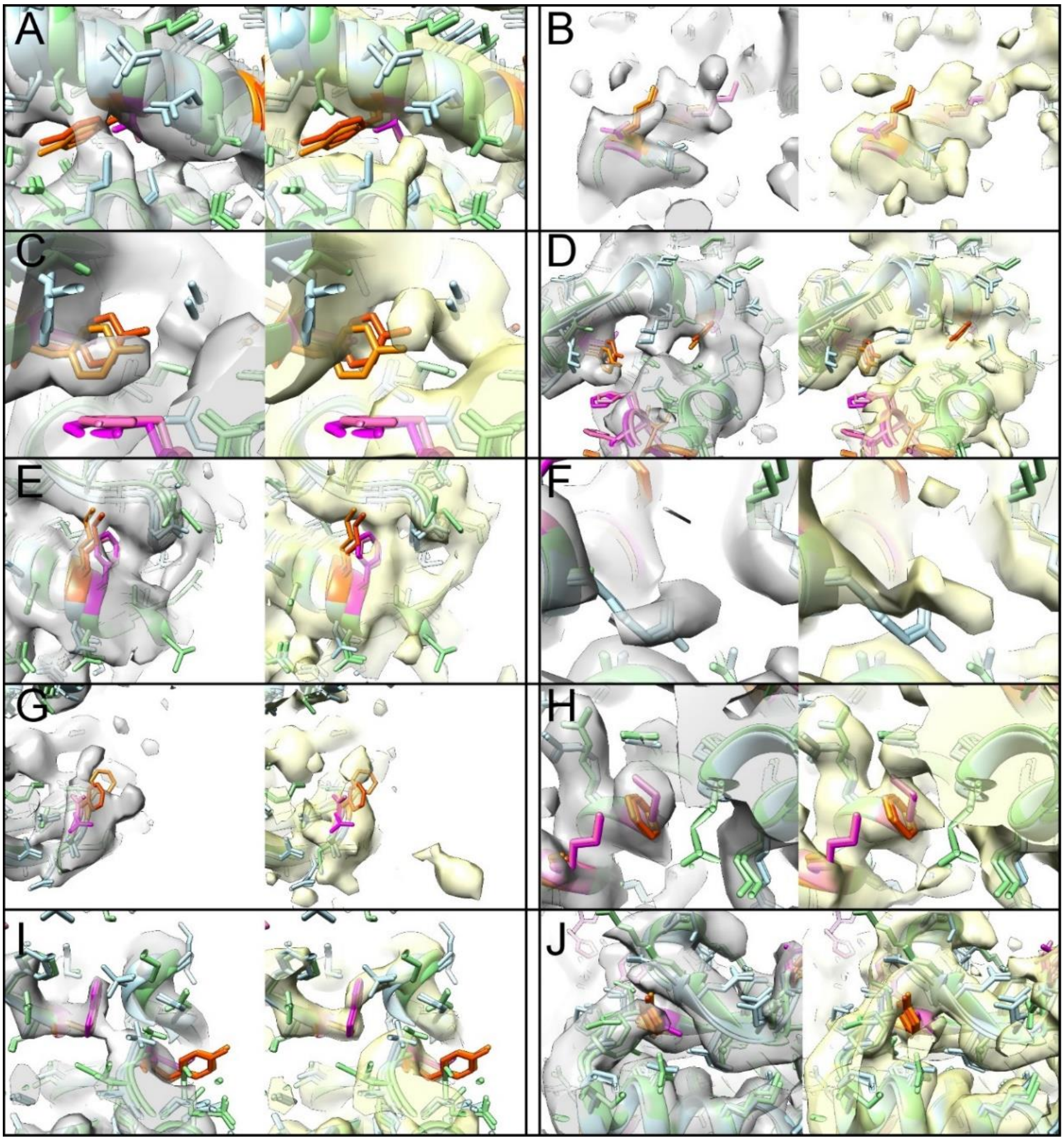


Figure 3-10: Single residue comparison for mutant maps. Residues being compared are highlighted in pink (alpha subunit – main chain in green) and orange (beta subunit – main chain in blue) (A) α-Val vs β-Tyr, (B) α-Lys vs β-Ser and vice versa, (C-D) α-Thr vs β-Tyr at different thresholding levels (E) α-Phe vs β-Lys, (F) α-Ala vs β-Arg (G) α-Asp vs β-Phe, (H) α-Met vs β-Phe, (I) α-Trp vs β-Ala, (J) α-Val vs β-Tyr.

Some regions of the subunits were so similar it was not possible to make a distinction between which subunit was better fitted as seen in (H-J) (5, 4 and 8), perhaps a particularly surprising example of this is the tryptophan of the alpha subunit shown in (I).

Taking all the evidence from the mutant map comparison into account it was decided that the beta subunit was best described by the grey density. The strongest evidence for this is seen in (A, B and E). While there were no strong examples of the alpha subunit being described by the yellow density, there was also no strong evidence against it.

Wild type comparison

The same analysis was performed for the wild type maps and models (Figure 3-11). These proved to be slightly less conclusive than the mutant with a higher number of sites being extremely similar (E-H) (5, 9, 10 and 2). (A-D) on the other hand gave some indication of differences. Site 8 this time shows a strong preference for alpha in yellow and beta in grey (A). The alpha-tryptophan (4) this time showed a slight preference for the yellow density (B). As in the mutant, (C) shows a slight additional narrow density corresponding to the lysine in beta from 1 and (D) shows the alpha-lysine in 6 with yellow density available.

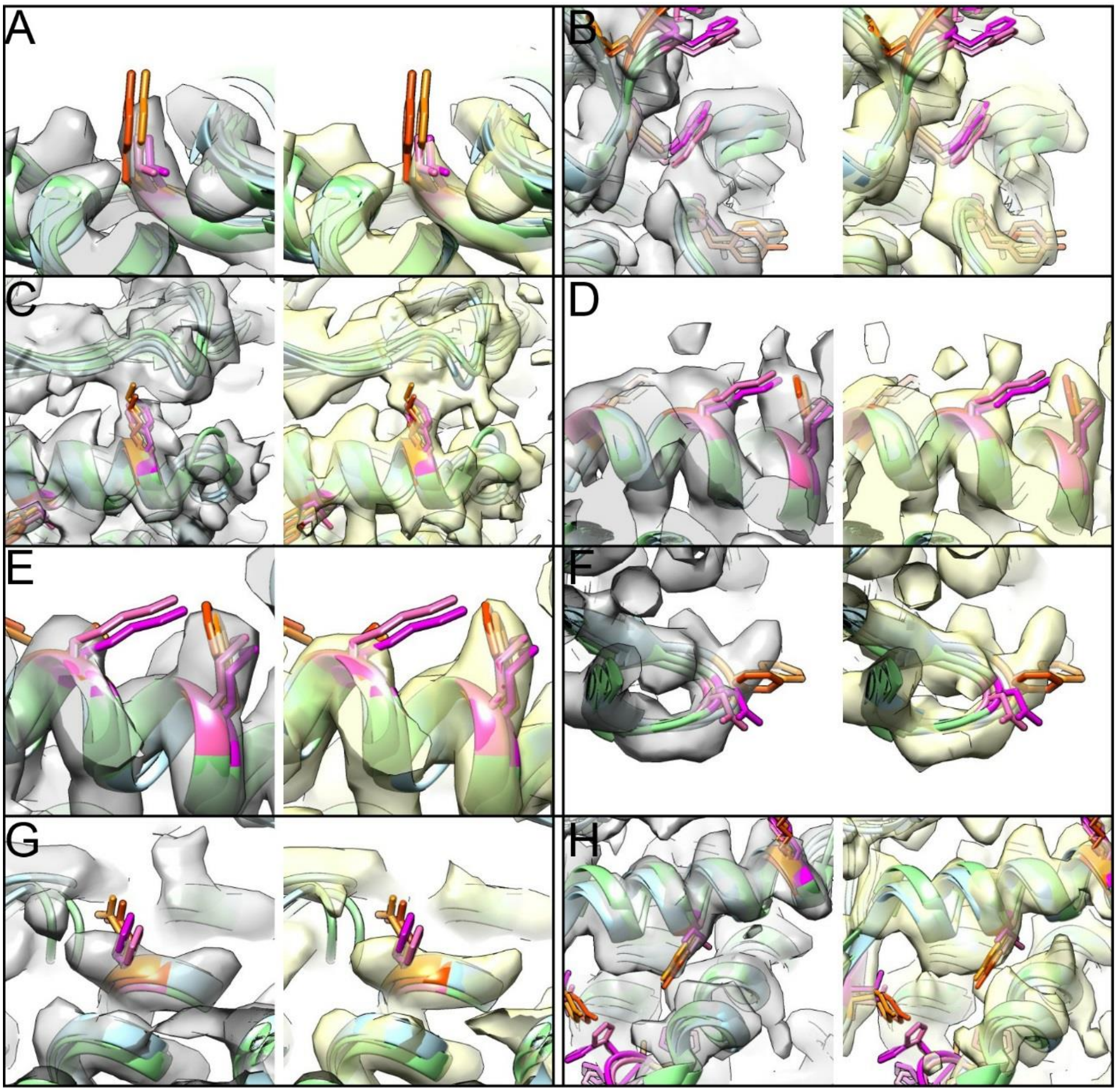


Figure 3-11: Single residue comparison for wild type maps. Residues being compared are highlighted in pink (alpha subunit – main chain in green) and orange (beta subunit – main chain in blue) (A) α-Val vs β-Tyr, (B) α-Trp vs β-Ala, (C) α-Phe vs β-Lys, (D) α-Lys vs β-Ser, (E) α-Met vs β-Phe, (F) α-Asp vs β-Phe, (G) α-His vs β-Gln, (H) α-Val vs β-Tyr.

For ease of comparison, each of the regions of potential difference is listed in Table 3-2 along with the corresponding figure reference and confidence level.

	Mutant – Figure 3-10	Wild type - Figure 3-11
1	E - moderate	C - moderate
2	A - good	H - inconclusive
3	C,D differing	Not shown - inconclusive
4	I - inconclusive	B - moderate
5	H - inconclusive	E - inconclusive
6	B - good	D - moderate
7	B - good	Not shown - inconclusive
8	J - inconclusive	A - good
9	G - inconclusive	F - inconclusive
10	Not shown - inconclusive	G - inconclusive

Table 3-2: Key to equivalent residues for single residue comparison in wild type and mutant figures.

Interestingly, the most conclusive residues were not the same between the two models, as demonstrated by the confidence levels in the table. Some potential differences such as 5 and 10 were in fact too similar to show variation. It is unclear why the mutant was slightly more easily interpreted, since the overall map resolutions were very similar. The majority of residues which were checked were clustered around H15 and S22 which are at the bottom of equatorial domain but not in regions which form contacts.

Taking all of the evidence into account it was deemed that the wild type alpha was best described by the yellow density while grey described the beta subunit.

3.3.2.5 Full model fitting

Having used the evidence from fitting models into individual subunit densities, we have now assigned which are alphas and betas for both the wild type and the mutant EM data. Full thermosome models were then created by rigid body fitting the appropriate subunits into the full maps according to D4A symmetry (See section 2.5.8)

The locations of differences between alpha and beta subunits discussed above were checked for the full map. Examples for the mutant are shown in Appendix IV.

The full models are shown in Figure 3-12 fitted into the local resolution coloured maps at various thresholds. On the left, the low threshold images demonstrate that the apical domains, while having been fitted to some extent, are still around 8 Å in resolution. For both models all but the very tips of the apical regions are enclosed within the density. Nevertheless, the apical regions were not included in the analysis into the motion of the structures due to the high error associated. Upper-intermediate domains are resolved to around 6 Å while the lower-intermediate and equatorial domains are at worst 5.5 Å and mostly significantly better than this. At these resolutions we see very good fitting of secondary structure features which are then used to explore the overall motion.

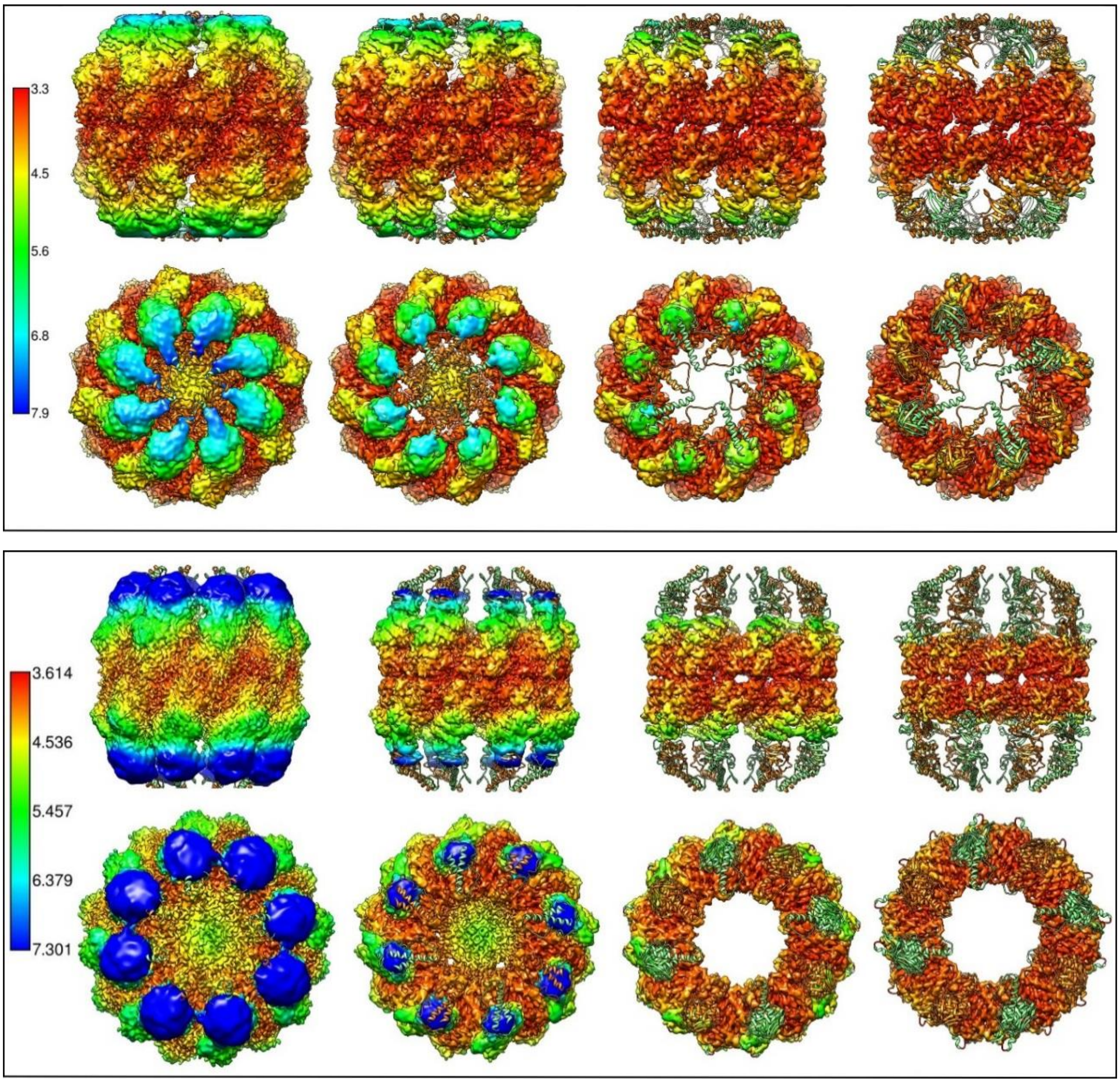


Figure 3-12: Wild type (top) and Mutant (bottom) full models fitted into local resolution coloured electron density maps at a variety of thresholds, increasing from lower (left) to higher (right). Upper line of each panel is side views; lower line is top views. Bars on the left indicate local resolution in Angstroms.

3.3.3 Thermosome dimensions

Cross sectional views of the wild type and mutant thermosome structures are shown in Figure 3-13. The diameter at various heights was measured by finding the mean distance between equivalent internal atoms on opposite sides of the ring, as illustrated by the mutant top view in (C). The equator was defined as the midpoint between the lowest measurement point, and the equivalent point on the opposite hemisphere. Scale cartoon depictions of single hemispheres with measurements are shown in (B). Values are shown to the nearest Å although, as discussed above, the error is greater further from the equator. In the equatorial domain, diameters are very similar, with the narrowest point reaching 63 and 68 Å for the wild type and mutant respectively. Above this pinch point the cavity opens out greatly for the mutant which is 30% wider across the cavity than the wild type. The increase in cavity volume from the closed to open conformation is greater than two-fold, increasing from approximately 184 000 Å^3 to 433 000 Å^3. In reality, it is likely that the additional bulk contributed by the lid domains in the wild type would reduce its internal volume compared to this measurement.

The wild type measurements were very similar to equivalent measures performed for the crystal structure (see Appendix V). The overall height of the crystal structure was slightly larger than the wild type, although the contacts between hemispheres are more compact.

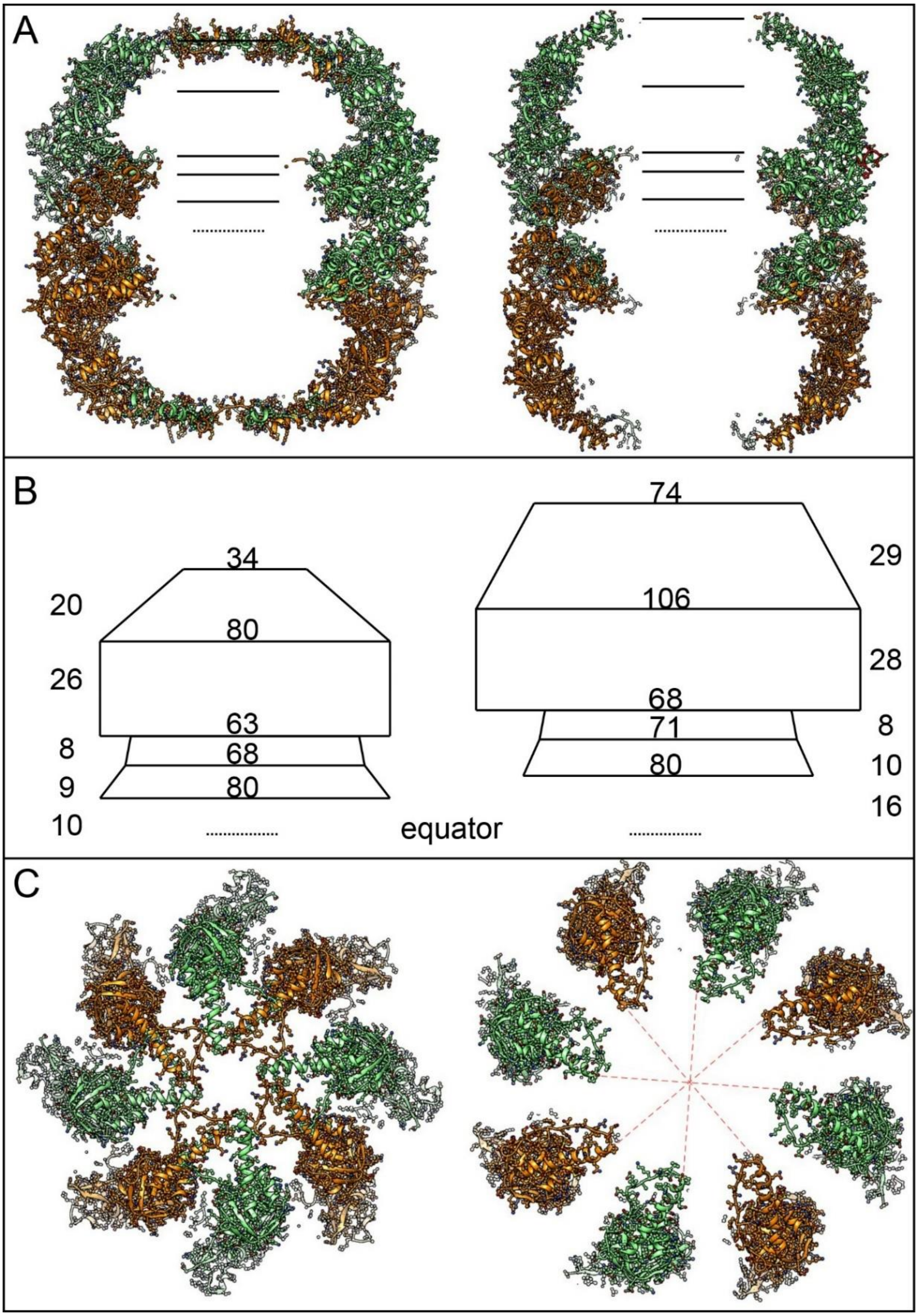


Figure 3-13: Measurements of the full fitted wild type (left) and mutant (right) models. (A) solid lines indicate positions of diameter measurements, dotted line indicates equator. (B) Cartoon representations showing measurements of the models. Numbers indicate distances in Å (C) Top views of the models. Dotted lines illustrate how the distances are measured for the mutant.

3.3.4 Subunit motion

Examining how the individual subunits compare in each of the two conformations reveals the extent of the motion in the equatorial and intermediate domains.

In order to align the models, the maps were first overlaid with alpha subunits overlapping. The grid centres were aligned and fitted together using Chimera's 'fit in map' tool with transitions limited to rotations and no translations. This ensured that one hemisphere was not preferentially aligned over the other. The models were then fitted into the maps resulting in the overlaid models shown in Figure 3-14 (A) where wild type is shown in yellow and mutant in blue. To measure the motion of the domains, each alpha helix was fitted using the 'Define Axes' tool. This identifies a vector for each alpha helix in each model as shown in (B). (C) shows these helices for two adjacent subunits, in both the wild type and mutant conformations labelled according to domain. (D) shows a top view of the models with equatorial helices fitted while (E) shows only the equatorial fitted helices. This view clearly demonstrates the tilting motion of the equatorial domain.

The angle between each pair of helices (i.e. open *vs* closed) was measured and plotted in Figure 3-15, grouped according to domain and subunit. The helix in the lid region was excluded from these measurements as it was not well resolved in the data. There is no significant difference between alpha and beta subunits in either the equatorial or intermediate domains. The apical domains show a variation of around 5º however these consist of only two data points and represent the lowest resolution region of the maps.

There is a significant difference between the motion of the equatorial and intermediate domains, with overall tilts of approximately 10º and 22º respectively.

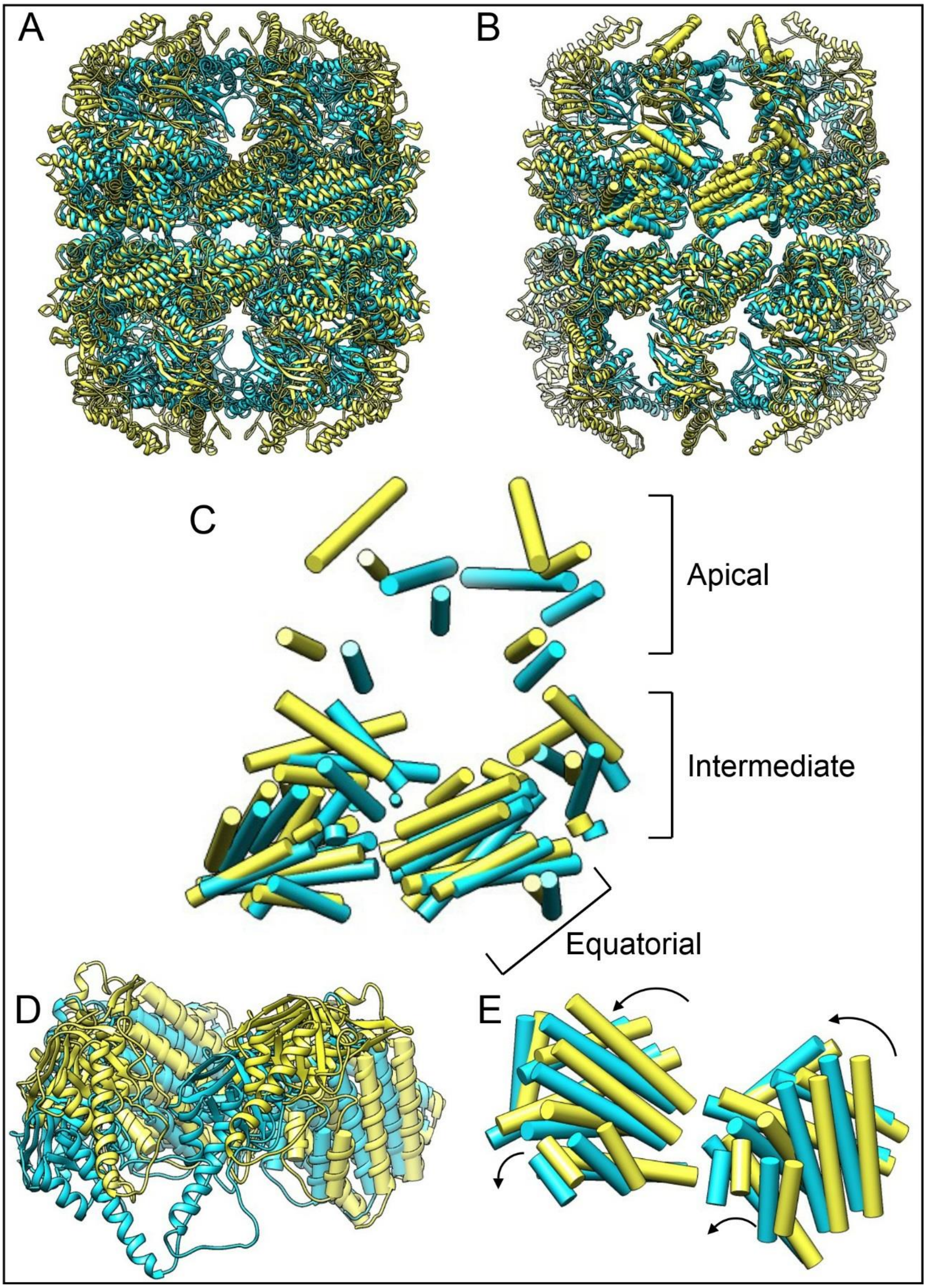


Figure 3-14: Comparison of domain motion. Wild type shown in yellow, mutant in blue. (A) Full models are fitted together ensuring equivalent subunits are overlaid. (B) Axes are fitted to each alpha helix in two adjacent subunits. (C) The same view as (A-B) showing only the fitted alpha helices and labelled by domain. (D) Top view of models with fitted equatorial helices. (E) Equatorial helices showing motion of the whole domain. Arrows indicate closing motion.

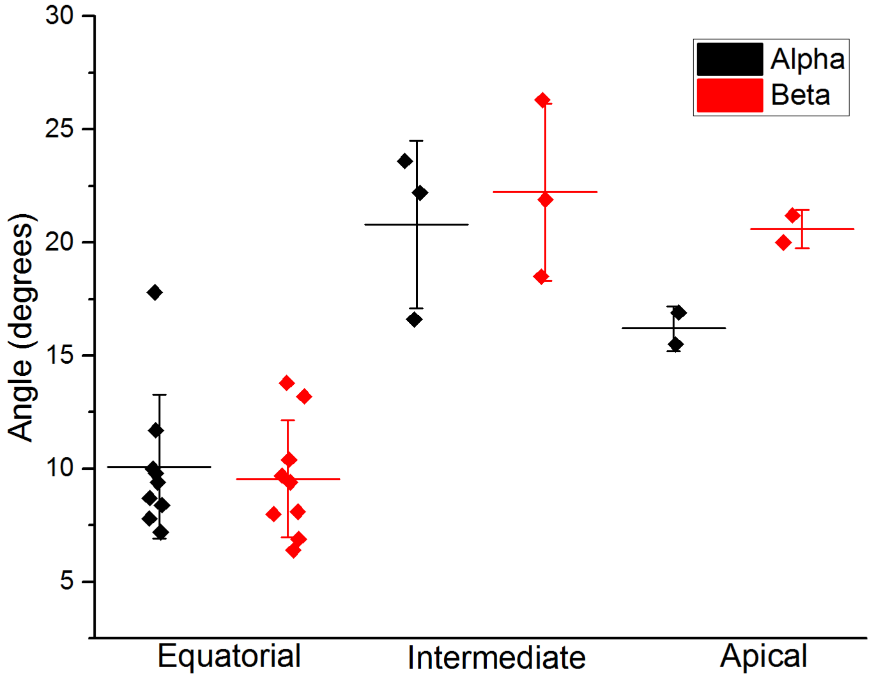


Figure 3-15: Motion of domains for alpha (black) and beta (red) subunits. Horizontal lines indicate mean angle. Whiskers represent 1 standard deviation from the mean.

3.3.5 Difference Mapping

Difference mapping is a technique whereby any regions of density found in the experimental map which cannot be attributed to the fitted model may be identified. In order to do this, a simulated electron density map was generated from the full model at a resolution of 4 Å, corresponding to the approximate resolutions of the EM maps. This was then subtracted from the experimental density map using the chimera_diffmap.py script from the Topf lab available at http://topf-group.ismb.lon.ac.uk/Software.html.

For each map several such non-attributed regions were identified and are shown in Figure 3-16. (A) and (B) show portions of the EM maps for the wild type and mutant respectively. Within these maps there are solid purple regions which correspond to signal in the difference map and are therefore regions which are not explained by the fitted model. Many of these could be attributed to slight mis-fitting of the model, however there were two key regions of interest which were visible in every subunit in both wild type and mutant maps.

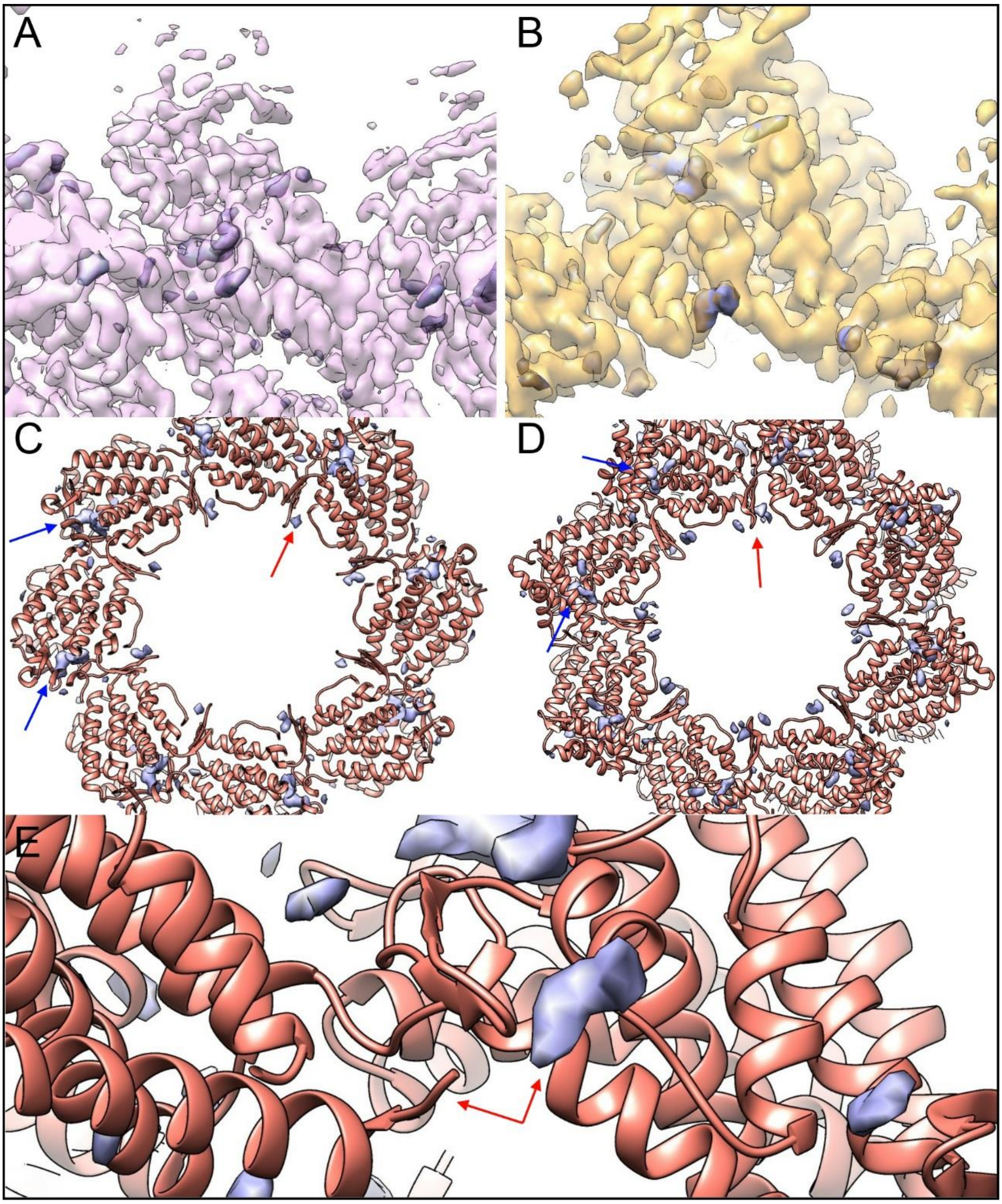


Figure 3-16: Difference mapping. Full (transparent) and difference (solid purple) maps of wild type (A) and mutant (B). (C-D) Regions of extra density in equatorial region. Blue arrows indicate density in nucleotide binding pocket. Red arrows indicate extra density adjacent to central beta hairpin. (E) Magnified view of c-terminus and extra density adjacent to central beta hairpin.

The first of these is found adjacent to the central beta hairpin, indicated by red arrows in (C) and (D). In the wild type maps this is consistently found on the clockwise side in the figure while for the mutant it is found on both sides. We speculate that this region of density may be caused by the c-terminus of the adjacent subunit. These terminal residues were not seen in the crystal structure and since this was used as a starting point for our modelling, they are also not included here. Thus, as far as we are aware, this is the first time these residues have been visualised. This is illustrated in (E) for the mutant structure where the terminus of the subunit on the left is pointing towards the region of extra density indicated in purple.

The poor fitting of some of the beta hairpins in the mutant may be attributed to the extra density pulling the hairpin towards it. Since each of these subunits was fitted in the absence of surrounding subunits, there would be nothing there to fill this density, therefore allowing the beta hairpin to be pulled out of place.

The second region of interest is found in the ATP binding pocket. Two of these are indicated by blue arrows in (C) and (D) although they can be seen in every subunit.

In order to study these in more detail the difference maps were segmented and the resulting segments sorted by volume. The regions in the binding pockets were systematically the largest and are shown on the left-hand side of Figure 3-17. ADP- AlF_3 was fitted into the structure using chimera's 'fit in map' tool. This was performed with multiple starting positions and the most favourable one (by visual inspection) is shown. The right-hand side of the figure shows that ADP-AlF_3 fits well into this density for all four of the subunits. Although ADP-AlF_3 has been fitted here, we would expect the beta subunit of the mutant to contain ATP as the mutation prevents hydrolysis. For further discussion see Chapter 4:.

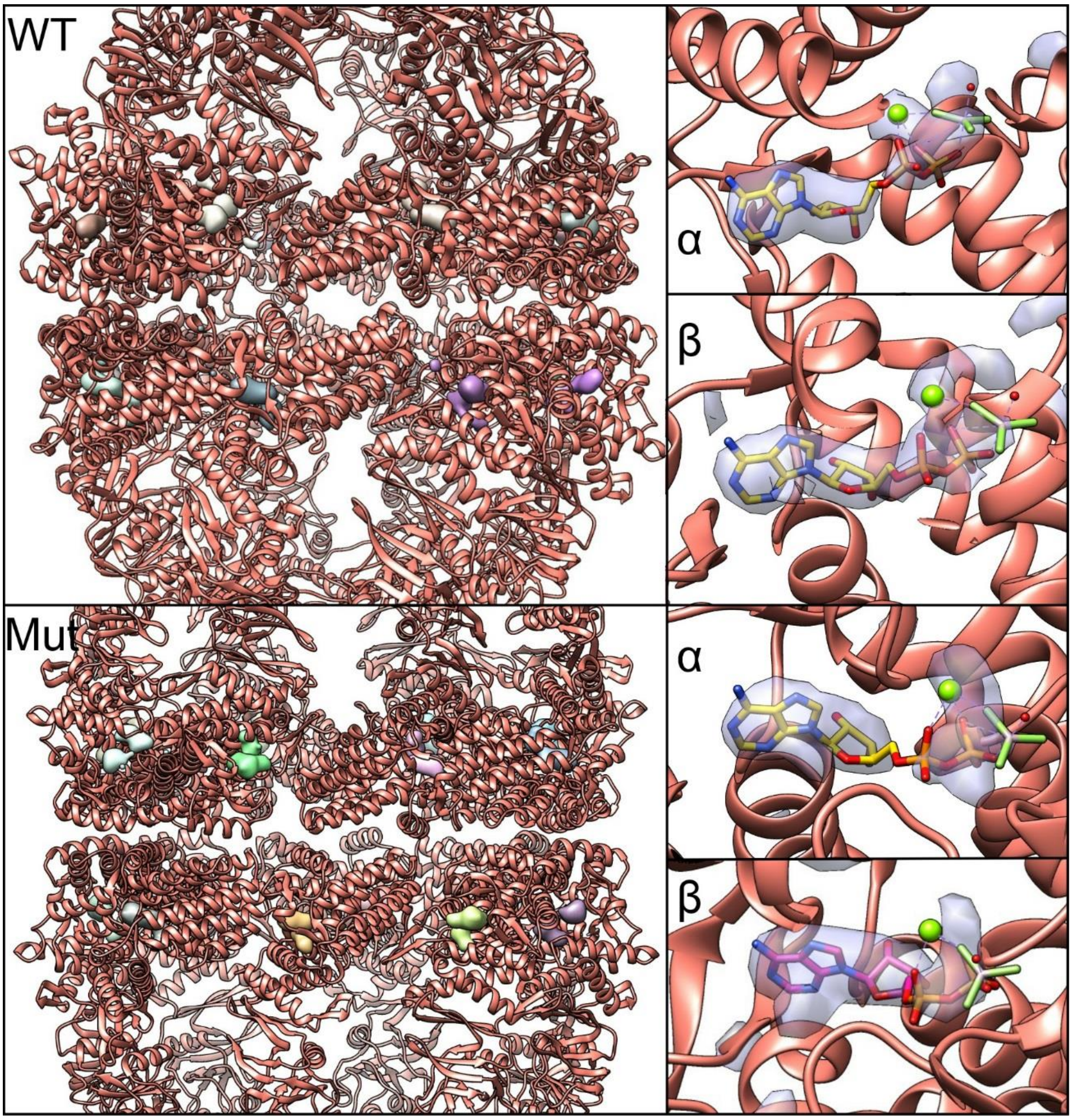


Figure 3-17: Nucleotide binding pockets. (Left) overview of wild type and mutant models with positions of extra density in nucleotide binding pocket highlighted. (Right) examples of densities with ADP- AlF_3.Mg fitted. Different colours in mutant beta subunit to highlight that we may expect a different nucleotide to be present.

3.3.6 Model Verification

Finally, the quality of the models was checked using the MolProbity web server [12]. This records statistics on geometries such as backbone angles and rotamers and outputs quality measures including Ramachandran plots and cis proline counts. For each measure there is a statistically acceptable number of outliers based on their prevalence in nature. The programme outputs both an overall statistical report and a full report for each residue in the protein.

Of particular interest to this study were the CaBLAM results [13] which give a measure of low resolution fit quality by highlighting unusual conformations in the Cα geometry. Since the models developed in this work are outputs from molecular dynamics simulations and have not been fully refined, statistics such as Ramachandran outliers would likely change following refinement. For example, we have seen from the individual residue comparison that many side chains are not optimally fitted into the data. Without further model refinement we are therefore most interested in validation of the low resolution, *i.e*. backbone and secondary structure.

CaBLAM recommends lower that 1% outliers are acceptable within a structure. While only the mutant beta subunit satisfied this, the other three all had 2 % or less.

For full model validation statistics from MolProbity, see Appendix VI.

3.4 Summary

In this chapter, molecular dynamics simulations have allowed the fitting of molecular models into the EM maps developed in the previous chapter.

A close comparison of adjacent subunits in each map allowed us to discern which density corresponds to which subunit.

From this, a full model was built following the symmetry of the EM map. Allowing us to quantify the changes in morphology from the fully closed crystal structure, the partially closed, wild type EM structure, and the open mutant EM structure.

The key results of this are summarised as follows:

1. Between the wild type and mutant structures, there is a greater than 2-fold increase in cavity volume and 25% increase in hemisphere height.
2. The equatorial domain acts as a rigid body which tilts by approximately 10°, while the intermediate and apical domains tilt by approximately 22°.
3. All subunits in both maps contain regions of extra density located within the nucleotide binding pocket. This extra density is well modelled by ADP-AlF_3.
4. Furthermore, there is additional density which may be explained by the c-terminus of the adjacent subunit which had not been visualised in the published crystal structure.

The implications and interpretation of these results are discussed in the following chapter.

References

1. Trabuco, L.G., et al., *Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics.* Structure, 2008. **16**(5): p. 673-683.
2. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics.* Journal of Molecular Graphics & Modelling, 1996. **14**(1): p. 33-38.
3. Topf, M., et al., *Protein structure fitting and refinement guided by cryo-EM density.* Structure, 2008. **16**(2): p. 295-307.
4. Pandurangan, A.P. and M. Topf, *RIBFIND: a web server for identifying rigid bodies in protein structures and to aid flexible fitting into cryo EM maps.* Bioinformatics, 2012. **28**(18): p. 2391-2393.
5. !!! INVALID CITATION !!! {}.
6. Zhang, J.J., et al., *Cryo-EM Structure of a Group II Chaperonin in the Prehydrolysis ATP-Bound State Leading to Lid Closure.* Structure, 2011. **19**(5): p. 633-639.
7. Zhang, J., et al., *Mechanism of folding chamber closure in a group II chaperonin.* Nature, 2010. **463**(7279): p. 379-83.
8. Chen, D.H., et al., *Visualizing GroEL/ES in the Act of Encapsulating a Folding Protein.* Cell, 2013. **153**(6): p. 1354-1365.
9. de Groot, B.L., G. Vriend, and H.J.C. Berendsen, *Conformational changes in the chaperonin GroEL: New insights into the allosteric mechanism.* Journal of Molecular Biology, 1999. **286**(4): p. 1241-1249.
10. Waterhouse, A., et al., *SWISS-MODEL: homology modelling of protein structures and complexes.* Nucleic Acids Research, 2018. **46**(W1): p. W296-W303.
11. Pintilie, G., et al., *Identifying Components in 3D Density Maps of Protein Nanomachines by Multi-scale Segmentation.* 2009 Ieee/Nih Life Science Systems and Applications Workshop, 2009: p. 44-+.
12. Davis, I.W., et al., *MolProbity: all-atom contacts and structure validation for proteins and nucleic acids.* Nucleic Acids Research, 2007. **35**: p. W375-W383.
13. Williams, C.J., et al., *MolProbity: More and better reference data for improved all-atom structure validation.* Protein Sci, 2018. **27**(1): p. 293-315.

Chapter 4: **Discussion**

Chaperonins are a class of molecular chaperones that are ubiquitous and indispensable for all forms of life. Of the two classes, type II, found in archaea and the eukaryotic cytosol, are the most mysterious and challenging to study. Despite being the focus of many recent studies, there is still much that is not understood about the mechanisms of these protein folding machines. This study focusses on the thermosome from the archaeon *T. acidophilum,* which is a hetero-oligomeric class II system composed of two types of subunit. This feature allows us to investigate the role of subunit heterogeneity in a system with reduced complexity relative to eukaryotic CCT with its 8 distinct subunits.

Prior to the work reported in this thesis, a large panel of ATP-binding and/or hydrolysis site directed mutants in each of the subunits had been developed in our lab [1].

A thorough biochemical analysis allowed us to describe different behaviours of the two subunits towards ATP. This explains the pattern of bi-phasic ATP hydrolysis shown by this thermosome, analogous to the behaviour of CCT.

To understand how changes in structure influence the differential behaviour between the two subunits, we chose a mutant in which ATP hydrolysis was blocked in only the beta subunit. For comparison we performed the same analyses on the wild type. We chose to analyse the complex in its post-hydrolysis (ATP-AlF_x bound) state, which in all class II chaperonins has a fully closed conformation, in order to understand the structural effects of abolishing such hydrolysis in one type of subunit. The discovery that the wild type and mutant show a fundamentally different conformation in the ATP-AlF_x state also gave us the unexpected and invaluable opportunity to examine the dynamics of closing of this macromolecular complex by comparing our two sub-4 Å cryo-TEM maps.

We have identified 6 key results of this study which will be discussed in this chapter.

1. A single D93K point mutation in the beta subunit results in a significant structural change of the whole complex, giving a thermosome in an open conformation.
2. The cryo-TEM map of the closed wild type Thermosome has strong similarities to the crystal structure by Ditzel et al. [2].

3. The individual domain movements leading to the transition from closed to open state are in line with those which have been determined, based on lower resolution models, for the *Methanococcus maripaludis* thermosome (Mm-cpn) [3].
4. There is no structural evidence of negative cooperativity between rings.
5. There is no evidence of large conformational differences between subunits in the open post-hydrolysis state of the mutant, indicating strong cooperativity within a ring.
6. The nucleotide binding pocket of both subunits is occupied.

4.1 Significant structural rearrangements from a single point mutation

Observing the mutant thermosome in an open conformation in the presence of ATP-AlF$_x$ is a significant result. We did not know what structural effect a point mutation that blocks ATP hydrolysis in the beta subunit would have, either locally at the site of the mutation (*i.e.* in the ATP binding pocket), or on the whole complex.

Since the ATP-hydrolysis activity of the mutant has been attributed to only the alpha subunits, we may expect to observe alternating open and closed subunits within a ring. We would also then expect ADP-AlF$_x$ to be present in the nucleotide binding pocket of the alpha subunit and ATP bound but unable to be hydrolysed in the beta subunit.

However, detailed symmetry analysis showed no evidence of this, with no gross conformational differences between subunits, which were both in a partially open state.

Although this specific mutation blocks the hydrolysis in a single subunit, it should be noted that the molecular machinery is still intact, ATP is hydrolysed (possibly by the alpha subunit only) and it refolds substrate to a certain extent. Taking this information into account, along with the structural evidence, it is clear that the two subunits are working in a cooperative way, and that ATP-hydrolysis on half of the sites is not sufficient to close the folding chamber.

The transition from open to closed state in our data corresponds to a 2-fold decrease in cavity volume, and a 20% decrease in the height of a single hemisphere. The ATP free state was also seen to be in an open state with a sizable (45º) rotation of the apical domains in the homo-oligomeric thermosome from Mm-cpn [3]. Preliminary data from our group indicates that the *T. acidophilum* thermosome is similarly open although at this time the volume data and molecular modelling have not yet been carried out.

4.2 Confirmation of the high-resolution crystal structure and recovery of flexible domains

The 1998 crystal structure of the *T. acidophilum* thermosome from Ditzel *et al*. has provided an important reference point in the type II chaperonin structural and functional analysis over the past 20 years. Thus, it was critical to compare our cryo-TEM map of the hydrated protein with the crystal structure for differences and similarities. The crystal structure and our WT map were both prepared with ATP-AlF_x giving us the ability to draw direct comparisons.

The comparison between the internal measurements of the crystal structure and the wild type cryo-TEM derived map revealed that all diameter measurements were the same to within a 10% margin. There was slightly greater variation between the height measurements, both at the equator and in the lid regions. The overall result is a slight compaction at the equator of the crystal structure, and extension in the apical domains, leading to a slightly taller and narrower structure than the cryo-TEM map.

The internal cavity volume for the crystal structure is approximately 170 000 $Å^3$ while the cryo-TEM model suggests 184 000 $Å^3$, once again within a 10% margin.

A second key difference between the crystal and cryo-TEM structures is the presence in the latter of additional density adjacent to the stem loop. We propose that this density is due to the additional residues of the termini of the adjacent subunits which wasn’t visualised in the crystal structure. It has previously been reported that these termini and their interaction with the stem loop are significant in substrate binding in GroEL and subunit cooperation in Mm-cpn [3, 4].

In their current state, the full models in this study are formed from subunits which have been individually refined. In order to further explore this stem loop-termini interaction it would be beneficial to refine the full model within the EM density, allowing this interaction to be more accurately modelled.

4.3 Molecular machine in motion

Due to the unexpected open state of the mutant thermosome revealed by the cryo-TEM map, we have had the opportunity to look at the dynamics of the ATP-hydrolysis driven closure of the complex.

Although we do not have the 3D reconstructions representing each step in the chaperonin's mechanistic cycle, we can begin to visualise the motion of this molecular machine through the transition between the closed and open states.

Motion of the equatorial domains is similar to that described for Mm-cpn in that motion is observed in all three domains upon ATP hydrolysis. This was not quantified for the equatorial domain of Mm-cpn, however we found that it moved approximately as a rigid body with a tilt towards the centre of the thermosome of around 10°. The intermediate domains showed a greater movement with a tilt on the order of 20°. Motion of the apical domains was not specifically studied due to the lower resolution of the map, particularly that of the open mutant, in these regions, however a significant rocking towards the centre forms a closed lid.

We have been able to see the 'closing of the iris' or closing of the inbuilt lid that occurs upon ATP hydrolysis. This movement stems from the initial tilt of the equatorial domain, which propagates and amplifies in the intermediate domain and finally closes the 'lid' with a rocking of the apical domain. To see how the binding of ATP and hydrolysis itself causes this movement, other structural states (such as the ATP bound state in the presence of a non-hydrolysable ATP analogue) would need to be solved at a similar sub-nanometer resolution.

4.4 No structural evidence of negative cooperativity between rings

Negative cooperativity between rings is a known feature of type I chaperonins. In GroEL, this manifests as a ‘bullet’-like conformation in which one ring is open and the opposite is closed as in complex with GroES.

In our case, we did not see evidence of ATP-driven conformational rearrangements in one ring inhibiting the other at any stage of the image processing.

An in-depth analysis of symmetries was performed for all data. This included a comparison of cyclic and dihedral symmetries to see how the two hemispheres related to each other. In general, when accounting for effective particle numbers, dihedral symmetry described the data as well as the cyclic symmetry.

For the mutant dataset, subset refinement revealed a slight preference for C4 symmetry, however when this was extended to the full dataset, the effect was lost. Furthermore, no conformational differences were identified between the C4 and D4 final refinements.

These results were crucial in establishing that the two rings act in the same way in response to ATP. The single class of ATP-hydrolysis sites identified in ATP-ase measurements can be confidently assigned to the alpha subunit. In other words, the ATP hydrolysis being measured for the mutant must be occurring in only the alpha subunits, rather than all subunits in a single ring. We therefore examined the differences between the two subunits to see if any conformational differences could be found.

4.5 Data indicate strong cooperativity within a ring

During TEM map reconstruction, we demonstrated that the datasets could be described as having pseudo-8-fold symmetry. This indicates that there were no significant conformational differences between subunits at those resolutions. The maps were refined using D4 symmetry, and although they appear extremely similar, it was possible to determine which subunit was which by identifying and comparing individual residues of differing sizes.

As discussed above, one possible hypothesis to account for the D93K mutant steady-state data was that we might see alternating open and closed subunits upon ATP-hydrolysis. This was not the case, and the similarity of the subunit conformations suggests strong cooperativity within the ring. Blocking hydrolysis in the beta subunit prevents closure of either subunit, despite the continued ability of the alphas to hydrolyse ATP, as measured at steady-state.

Previous studies of Mm-cpn suggested that the majority of apical domain rotation occurs upon ATP binding, while hydrolysis results in an inward tilt. We can describe the beta subunit of our mutant as being analogous to the ATP bound state in those studies. Our beta subunit appears to be confining the un-mutated alpha subunit into a matching conformation. Although we do not see a differential tilt in the two subunits which might occur if they were acting sequentially or ‘independently’, we might be underestimating possible differences in the apical domain that may not be sufficiently resolved in that low-resolution region.

Again, as there is no significant difference between the subunits in terms of domain rotations or shifts, the two appear to be working in concert.

4.6 Presence of nucleotide

A crucial requirement for the discussion of the results above is the unequivocal confirmation of the presence of density attributable to nucleotide in all the subunits of both complexes.

Indeed, difference mapping revealed density in the nucleotide binding pocket of all subunits which is not modelled by the fitted structures. This density is well modelled by nucleotide in all cases. While it is not possible to distinguish between ATP and ADP-AlF_3 densities at these resolutions, we propose that the beta subunit of the mutant would likely be occupied by un-hydrolysed ATP while ADP-AlF_3 would be present in the alpha subunit and both wild type subunits.

We cannot rule out partial occupancy of this binding pocket, in particular due to the use of symmetry however it is a significant density in a high-resolution area of the map and as such we are confident that we are visualising nucleotide.

4.7 Future Work

The work reported within these chapters has allowed us to make several new observations on the structure and ATP-dependent conformational rearrangements of the *T. acidophilum* thermosome, however experimental avenues are by no means exhausted for these data.

New methods in data processing are constantly being developed. Indeed, a new version of the RELION software has been released since this work began, which could allow for improvements in maps resolutions. Additional features include improved motion correction determination for both whole micrographs and individual particles (known as particle polishing), and CTF refinement which makes use of both phase and amplitude information in CTF estimation, rather than just the amplitude information present in the power spectra [5, 6].

Other image processing techniques such as focussed classification and refinement could enhance resolution in desired regions of the map. This could result in the two subunits being more easily distinguishable. One such study has recently been performed on GroEL revealing three distinct subunit conformations present within individual GroEL oligomers [7].

As mentioned in Chapter 2 it would also be beneficial to create a refined model using an alternative software package which does not make use of multi-reference alignment methods. This would reinforce the evidence for the distinction of subunits as discussed in this thesis.

Of particular importance would then be to refine the fit of the full model rather than single subunits to allow for more accurate representation of these contacts and interactions.

The models reported in this thesis are known to have regions where manual manipulation of the model could improve the fit. This is evident in the individual residue comparison in which there were densities in the maps available for side chains which were unoccupied. The next stage of this would be to further refine the structure in order to explore different rotamers, torsion angles and real space fit of the model to the map. This would allow further investigation of the communication between subunits and hemispheres.

In order to gain a better understanding of the thermosome's allosteric mechanism we propose some further structural investigations.

Firstly, to study the D93K mutant discussed in this thesis in the absence of nucleotide. This would be analogous to the nucleotide free state studied for Mm-cpn. Since we know, from our analysis of the ATP-AlF$_x$ state, that the mutation does not affect nucleotide binding, we may expect the structure to potentially be in a slightly more open state. However, if this is true, we would also expect that such a conformation would have increased flexibility in the apical domains, thus increasing the difficulty of resolving them.

Similarly, studying both the wild-type and the D93K mutant in the presence of a non-hydrolysable analogue such as ATPγs, should yield thermosomes with all subunits in ATP bound state. This would inform us on the conformational rearrangements associated with ATP binding to all subunits, and would allow a comparison with the open state observed for the mutant, when ATP can be hydrolysed in one subunit.

Finally, within the original panel of mutants generated [1] the αD94K mutant has an equivalent hydrolysis blocking mutation in the alpha subunit but not the beta. Previous work indicates that this mutant is not able to hydrolyse ATP at all, indicating a clear difference in the behaviour of the two subunits. Structural analysis of such a mutant in a similar way to that reported here would be very informative in our quest to understand the differing role of the two subunits in the ATP-cycle of the whole complex.

References

1. Shoemark, D.K., et al., *Intraring allostery controls the function and assembly of a hetero-oligomeric class II chaperonin.* FASEB J, 2018. **32**(4): p. 2223-2234.
2. Ditzel, L., et al., *Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT.* Cell, 1998. **93**(1): p. 125-138.
3. Zhang, J.J., et al., *Cryo-EM Structure of a Group II Chaperonin in the Prehydrolysis ATP-Bound State Leading to Lid Closure.* Structure, 2011. **19**(5): p. 633-639.
4. Weaver, J. and H.S. Rye, *The C-terminal Tails of the Bacterial Chaperonin GroEL Stimulate Protein Folding by Directly Altering the Conformation of a Substrate Protein.* Journal of Biological Chemistry, 2014. **289**(33): p. 23219-23232.
5. Scheres, S.H. *Single-particle processing in RELION-3.0*. 2018.
6. Zivanov, J., et al., *New tools for automated high-resolution cryo-EM structure determination in RELION-3.* Elife, 2018. **7**.
7. Roh, S.H., et al., *Subunit conformational variation within individual GroEL oligomers resolved by Cryo-EM.* Proceedings of the National Academy of Sciences of the United States of America, 2017. **114**(31): p. 8259-8264.

Chapter 5: **Self-Assembling Peptide Cages**

In this final chapter, I will explore how cryo-TEM can be used in conjunction with several complementary techniques.

As discussed in previous chapters, single particle and sub-tomogram averaging experiments are becoming increasingly common. This is predominantly due to the recent advances in camera technology and accessible software tools. However, these approaches are not the only way in which cryo-TEM can be used. In this chapter I will discuss how a project, initially expected to make use of a single particle workflow, evolved to incorporate a wide variety of experimental methods to characterise a biological nanoparticle as fully as possible. I will discuss how cryo-TEM was initially used to aid in optimisation of sample conditions and later in conjunction with a variety of other methods such as dynamic light scattering, atomic force microscopy, and small angle x-ray scattering to improve our understanding of a complex system of *de novo* designed peptide nanoparticles.

5.1 Background

The natural world provides many examples of encapsulation in a wide variety of contexts, such as transportation, storage and protection. We have already explored the role of chaperonins in Chapters 2-4. Virus capsids often form regular shapes which store and transport DNA into a host cell, while ferritin is a ubiquitous family of proteins which stores soluble iron in cells [1, 2]. These natural protein systems evolved to be highly effective but can also be highly complex, specialised and difficult to replicate or modify. However, given their stability, homogeneity and potential for modification they are often used as a starting point for biotechnological applications [3].

Several review papers are available on the topic of artificial and biomimetic nano-encapsulation, with resulting materials being used in applications such as catalysis, drug delivery and nanoparticle synthesis [3-5].

The goal of the Woolfson group at the University of Bristol, was to develop a modular peptide-based system, in which features such as size, shape, targeting ability, type and volume of contents and content release could all be modified and adapted to suit a wide range of applications.

5.1.1 Initial development and characterisation

First reported by Fletcher *et al.* in 2013 [6], Self-Assembling peptide caGEs (or SAGEs) are 100 nm spherical particles comprised exclusively of *de novo* designed peptides. A homotrimeric and a heterodimeric coiled-coil were developed to form the basis of this system. By conjugating the trimeric and one or other of the dimeric peptides via a disulphide bridge and allowing the trimers to oligomerise, two separate building blocks, or hubs, are produced. When the two hubs are mixed in solution, the dimeric peptides come together to form coiled coils, resulting in the growth of a lattice (Figure 5-1(A)).

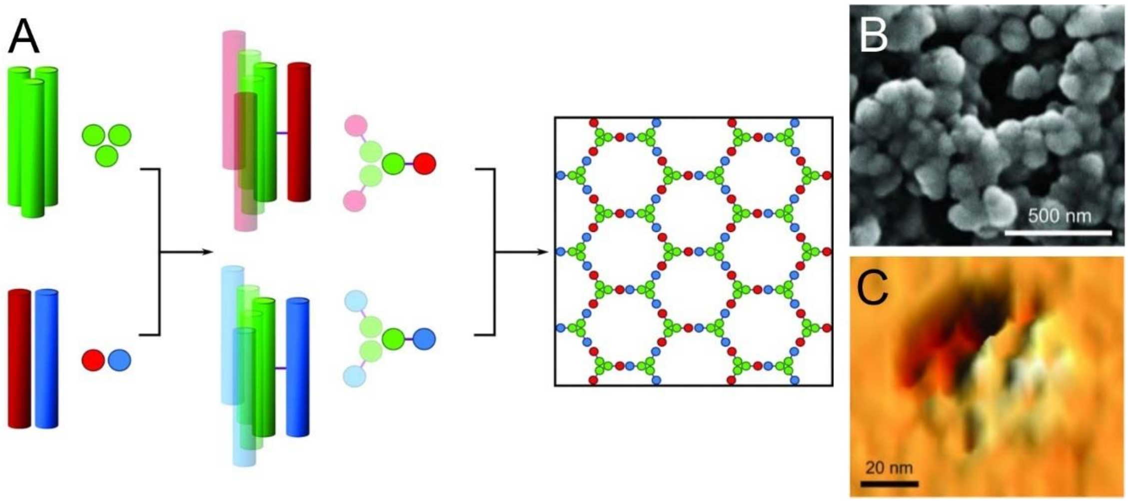


Figure 5-1: SAGE design and characterisation. (A) schematic of modular SAGE design with homotrimer in green, heterodimer peptides in red and blue. Left: initial coiled coils, centre: hubs conjugated with disulphide bonds, right: mixed hubs form extended hexagonal network. (B) scanning electron microscopy and (C) lateral molecular force microscopy images. Modified from [6].

When imaged by scanning electron microscopy (SEM) (Figure 5-1(B)), the resulting particles were approximately spherical, with a diameter of 97 ± 19 nm (including heavy metal coating), n = 135 particles. Tapping mode-atomic force microscopy (AFM, not shown) revealed that particles deposited and dried onto a mica substrate had a height of 9.2 ± 1.0 nm (5 particles) for particles of diameter 95 ± 14 nm.

Lateral molecular force microscopy (LMFM), a variant of AFM in which the surface of a particle is imaged in solution, revealed surface structure which was interpreted as hexagons with edges of 7 ± 2 nm (Figure 5-1(C))

These three observations in combination led to the conclusion that the predicted hexagonal lattice was curving to form spherical particles. This would necessitate the inclusion of some non-hexagonal defects in order to give appropriate curvature. In the simplest case this would mean the inclusion of 12 pentamers. Furthermore, atomistic modelling of a small patch of hexagons indicated that the hubs have intrinsic splay due to the non-central location of the disulphide bond, and the presence of positively charged lysine residues.

Thus, a modular system consisting of two simple building blocks was developed.

5.1.2 Modification and functionalisation

In order to make use of this system, it was necessary to add functionality. Subsequent papers from the group have described this process in detail, demonstrating that a variety of functionalities may be introduced to the system, in useful quantities, without disrupting the formation of particles.

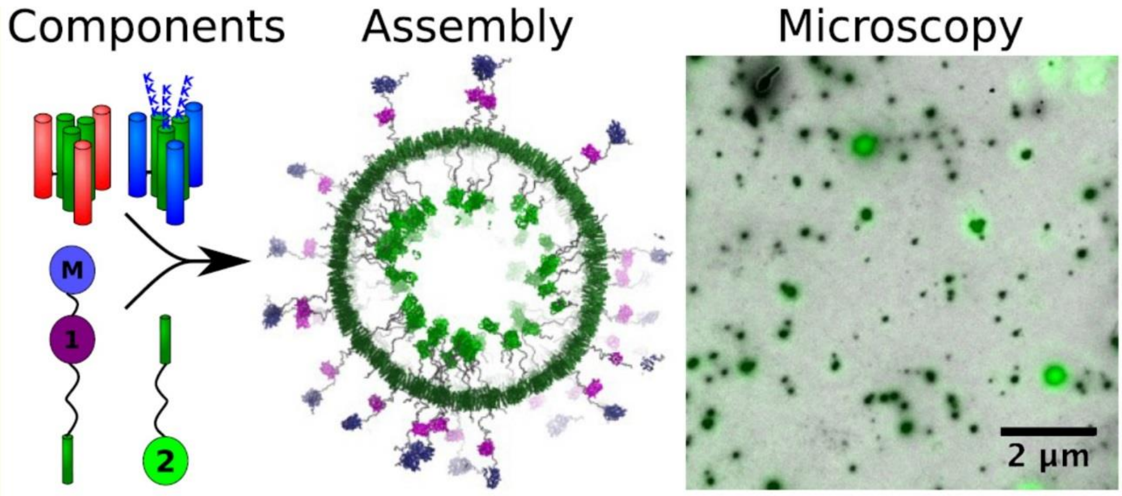


Figure 5-2: Modification of SAGEs. Design, modelling and TEM of fluorescent protein labelled SAGE particles. Reproduced from [7]

In a 2017 paper, the incorporation of multiple proteins, such as green fluorescent protein (GFP) and maltose binding protein (MBP), into the SAGE system was demonstrated [7]. Proteins were conjugated via linkers to either the N or C terminus of the trimer peptide and decorated hubs were mixed in various quantities. Thus, it was shown that the modularity of the system allows inclusion of multiple functionalities, and that up to 15% of the hubs could be modified with large proteins without breaking the system.

In addition to large proteins, functionalisation with small fluorescent molecules, short charged peptides and antigenic peptides have been discussed for applications in imaging, drug delivery and vaccines [8-10].

5.1.3 Applications

As mentioned above, one of the aims of developing a modular system was to allow the incorporation of multiple functionalities, thus making it suitable for a range of applications. As such, several studies have been published showing the versatility of SAGEs for a variety of functions.

The incorporation of multiple proteins discussed previously, provides a proof of concept for their development as nanoreactors. This could be done by immobilising multiple enzymes to form a catalytic pathway, with the active encapsulation resulting in increased presentation of substrate to the enzymes [7]. SAGEs have also been functionalised with antigenic peptides and shown to drive immune responses in mice [10].

A 2018 paper [8] shows that SAGE particles are internalised by cells, in a process which can be increased or inhibited by tuning the surface charge. SAGEs were internalised intact and were non-toxic to the cells. Work on delivery of cargo into cells is ongoing along with studies of how the internalised particles are trafficked.

5.1.4 Modelling

Alongside the experimental investigations, SAGEs have also been studied *in silico*. Due to the computational requirements, it would not be possible to perform atomistic modelling of the self-assembly process. As such, a coarse grained model was developed in which peptides were treated as rigid bodies with attractive patches [11]. The hub flexibility was reduced compared to the values predicted by atomistic simulations in order to increase the likelihood of formation of a hexagonal lattice. Despite this, the resulting self-assembled particle featured a high proportion ‘of non-hexagonal defects’ as shown in Figure 5-3(A-B). The range of angles found shows peaks for hexagons and squares but also features a broad tail to higher angles indicating the potential for formation of larger pores.

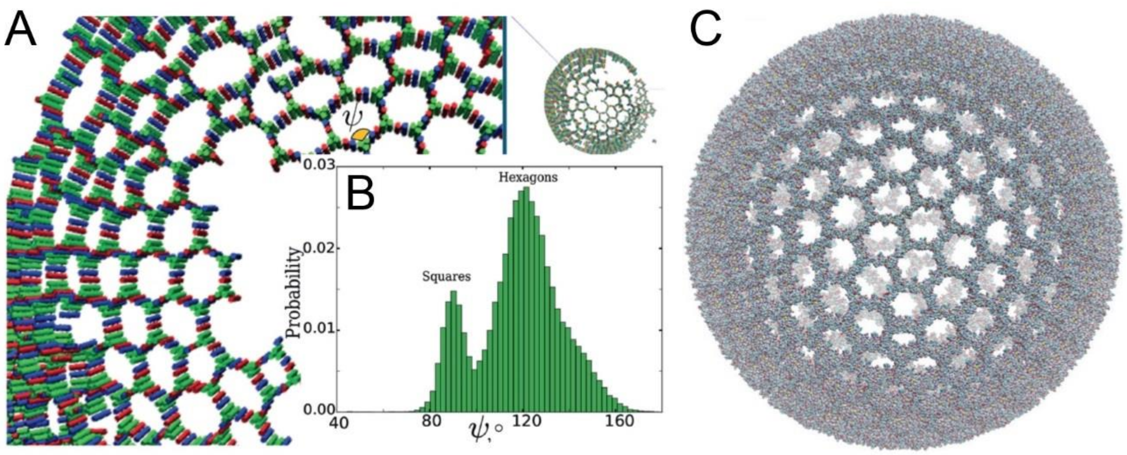


Figure 5-3: Modelling of SAGEs. (A) Coarse grain modelling of self-assembly. (B) Angular distribution for internal angles of surface polygons as shown in (A). (C) Atomistic model of ideal SAGE. Reproduced from [11, 12]

Secondly, full atomistic simulation of pre-assembled SAGE in the presence of a range of solutes from ions to macromolecules, was performed covering a timeframe of up to 1 μs [12]. Figure 5-3(C) shows the model particle prior to the simulation, which contains 12 equally spaced pentagons to allow curvature. Results from this study showed that the particles were stable over this time frame and allowed the free transfer of ions throughout the system while small molecules and proteins tended to stick to the SAGE surfaces. A further finding of this study was that while the trimer orientation remained constant with the N-termini on the external surface, the dimers were able to rotate, sampling 87% of all possible rotations.

5.1.5 Aims

The current literature on SAGE particles demonstrates that they are a modular system, capable of being modified with multiple functionalities to varying degrees. They show potential for applications in delivery of cargo into cells and as vaccines. In order for such applications to truly be realised however, a full, comprehensive, structural characterisation of the system is essential. Thus far, evidence for the proposed structure is limited to that seen in the initial paper. AFM scans over dried down SAGEs support the hypothesis that they are hollow, unilamellar particles.

Tomography of silicified SAGEs however suggests that there may be more internal material [9], while modelling studies demonstrate the ability for the formation of many non-hexagonal defects [11]. Coupled with the result that the dimeric peptides are able to rotate relative to the trimers, it is possible that the assembly process may be considerably more complex than previously suggested.

Cryo-transmission electron microscopy provides a unique route to visualising SAGEs as it allows them to be seen in their native hydrated state. This allows us to ask questions relating to structure, including the presence of defects, the porosity, and the effect of adding functional groups to the particles.

In order to address these questions, several complementary techniques were employed alongside cryo-TEM, demonstrating how the technique can contribute to a larger scientific problem as well as a specific structural analysis as described in previous chapters.

5.2 Methods

5.2.1 Synthesis Methods

5.2.1.1 Peptide synthesis

Peptides were prepared by solid state synthesis according to the protocol described by Fletcher *et al*. [6]. In short, a Liberty^{TM} microwave synthesizer was used to prepare peptides on a Rink amide resin by repeated coupling and deprotecting of Fmoc-L-amino acids with dimethyformamide (DMF) washes. The peptide was then acetylated before being cleaved from the resin and remaining protecting groups removed using trifluoroacetic acid (TFA), triisopropylsilane and ethanedithiol. The peptide solution was separated by filtration, and the peptide itself isolated by precipitating and resuspending in 50:50 H_2O:acetonitrile (MeCN) before freeze drying. Crude peptide was purified by high performance liquid chromatography. Conjugation of the trimeric and dimeric peptides was done by the formation of a disulphide bond and the final product purified again to remove any residual unreacted peptide.

Fluorescein and GFP conjugated hubs were prepared by Dr Joseph Beesley and Dr James Ross respectively as described in [7].

5.2.1.2 Salt substitution

In order to produce SAGEs in their purest possible state, a desalting protocol was tested. It had been hypothesised based on previous, unpublished studies in the group that following the standard hub synthesis and purification protocol, a number of TFA molecules remain associated with the peptides. In order to quantify this, a ^{19}F NMR experiment was performed by adding a known quantity of Trifluoroethanol (TFE) and comparing the measured quantities of TFE and TFA.

A peptide solution of known concentration was prepared as described. Based on previous work done in the group, an initial estimate of 30 TFA molecules per peptide was used. Trifluoroethanol (TFE) was then added in equimolar quantities to this assumed number of TFA molecule. ^{19}F NMR was then performed and the TFA and TFE peaks integrated and compared to give the true number of TFA molecules associated with each peptide molecule.

After freeze drying to remove TFE, the peptide was redissolved and shaken with a quantity of the ion exchange resin (Dowex® 1X8 chloride form) for 1-2 hours. The resin was filtered off and the solution prepared for NMR analysis with TFE as described above. The success of this process was demonstrated by the reduction in number of TFA molecules associated with the peptide after desalting compared to the as-prepared peptide. ^{19}F NMR was performed by Dr Guto Rhys. The quantity of TFA present in each sample was found by normalising the TFE peak to 1 and multiplying the number of TFE molecules added per hub (30) by the area under the TFA peak.

5.2.1.3 SAGE formation

Aliquots of hubs were diluted in the desired buffer to approximately 2x the desired concentration and then spun at 12 000 rpm for 30 minutes to remove any dust particles. The concentration of the supernatant was measured using a Nanodrop2000 (ThermoFisher Scientific) and further buffer added to reach desired concentration. Hub solutions were then mixed to desired ratios and the SAGE solution left at room temperature to form particles.

When functionalities such as fluorescence were incorporated into the SAGE, it was ensured that the like-hubs were mixed together initially before adding the oppositely charged hubs. For example, in the case of addition of 5% Alexa488 conjugated hubB, the hubB and Alexa488 hubB were mixed in the desired ratio before addition of hubA. This was to ensure that all functionalities were equally distributed among the formed SAGEs.

Unless otherwise stated, all SAGEs in this report are 50 % hubA + 50 % hubB SAGE – also known as parent SAGE or undecorated SAGE, at a concentration of 100 μM in 25 mM HEPES buffer.

5.2.2 Characterisation Methods

5.2.2.1 Room temperature Transmission Electron Microscopy

Hubs were mixed for 1 hour before pipetting 5 μl onto a Formvar/Carbon finder grid from EM Resolutions. The drop was left to dry on the grid overnight in the dark. Images were collected on 4k x 4k Eagle CCD camera mounted on the bottom of a Tecnai T12 (FEI company, Eindhoven, NL) with a tungsten filament operating at 120 keV.

5.2.2.2 Cryo-Transmission Electron Microscopy

Grids were plunge frozen using either a Leica GP or a Vitrobot Mark IV. 3 μl of sample was pipetted onto a glow discharged, 300 mesh lacey carbon grid in the humidified sample chamber. The excess was removed by blotting for 1-2 s depending on sample and plunged immediately into pure liquid ethane.

Cryo-TEM images were collected on a Tecnai T20 (FEI company, Eindhoven, NL) with a LaB_6 filament operating at 200 keV under low dose conditions using FEI software. The search and exposure magnifications were set at 2500x and 25 000x respectively throughout and images collected at ~2-6 μm underfocus. Images were collected on a bottom-mounted FEI Eagle 4k x 4k CCD camera.

5.2.2.3 Cryo-Tomography

High resolution images and tomograms were collected on a FEI Titan Krios with 300 keV FEG source at the electron Bioimaging Centre (eBic) at Diamond Light Source in Oxfordshire. Images were collected using either a FEI Falcon II detector or a Gatan Quantum Summit filter and detector. See Table 5-1 for tomography collection details.

Number of Tomograms	7
Max positive tilt	60º
Max negative tilt	-60º
High tilt switch	45º
High tilt step	3º
Low tilt step	4º
Tilt scheme	Start high angle, continuous
Total dose	~ 60 $e^-/Å^2$
Defocus	4 µm

Table 5-1: Cryo-Tomography collection details

5.2.2.4 Cryo-TEM time course image analysis

In order to track changes in SAGE structure and stability over time, a time course experiment was developed in collaboration with Dr Harriet Bray.

Dynamic Light Scattering

SAGEs were initially prepared as described in Section 5.2.1.3 in a quartz DLS cuvette. Following hub mixing, the cuvette was immediately placed in the DLS (Malvern, Zetasizer). Measurements were recorded (6 x 10 s each minute) for the first hour of SAGE formation. After the first hour, 150 µl of sample was removed from the cuvette and kept separately to make cryo-TEM grids and AFM stubs. 10-minute DLS measurements were recorded every 24 hours for the next 7 days.

Cryo-TEM grids and AFM stubs were prepared using the same sample solution and at the same time points to ensure correlation between experiments.

Atomic Force Microscopy

AFM stubs were prepared by spotting 5 µl of solution onto freshly cleaved mica surface and leaving for two minutes before washing with water (3 x 1 ml) and drying under nitrogen gas. AFM imaging was performed by Dr Harriet Bray on a Bruker Multimode with Nanoscope V controller and Picoforce Extender. All AFM image analysis was performed by Dr Harriet Bray.

Cryo-TEM Grid Preparation

Grids were prepared and images collected as described in Section 5.2.2.2.

Cryo-TEM Image Analysis

Approximately 100 images were manually collected for each time point at a magnification of 25 000x. Since particle projections were not perfectly circular, particles were manually selected and measured using the elliptical area selection tool in the FIJI software package [13]. The measured area for each particle was then used to calculate the mean diameter for each particle and timepoint.

5.2.2.5 Cryo-Correlative Light Electron Microscopy

Correlative Light-Electron Microscopy (CLEM) utilises two complementary microscopy techniques to image the same features over a range of magnifications. Particles which are both fluorescent and electron dense may be visualised by both light and electron microscopy in order to gain high resolution insight often in a cellular context. SAGE particles could therefore be identified by combining these techniques and correlating between the two.

Cryo-fluorescence microscopy was performed using a Leica cryo-CLEM stage (Leica microsystems) mounted on an optical microscope with a low working distance Leica EM Cryo-CLEM objective lens. Grids were held at liquid nitrogen temperature during transfer to the cryo-stage using the transfer shuttle. This system is shown in Figure 5-4.

Figure 5-4: Leica Cryo-Stage. Components from left to right: Cooling unit control system, liquid nitrogen tank which pumps nitrogen vapour into the cryo-stage, lamp and power units, microscope control unit and widefield microscope with cryo-stage attachment and transfer system[14].

In general, CLEM experiments make use of finder grids, in which a TEM grid has an in-built location system, in order to facilitate finding the same region of interest in multiple imaging modalities. However, a lack of access to lacy carbon film on finder grids and the difficulty of preparing high quality vitreous ice on such grids meant that this was done manually using standard 200 mesh lacy carbon grids. In short, this required locating the centre of the grid and manually noting the relative positions of the regions of interest. Following image collection, the grid was transferred into the TEM, the region of interest found, and imaged under low dose conditions as described previously. Overlays of the FM and EM images were created using the TurboReg plugin in FIJI [15].

More recent developments have considerably improved this process including the addition of a motorised stage to the cryo-fluorescence system and software which allows automated correlation between the fluorescence and EM images.

5.2.2.6 Electron Energy Loss Spectroscopy

SAGE solution (100 μM) which had been forming for 4 days was spotted onto lacey carbon on 300 mesh copper grids, left for several minutes and then blotted and allowed to dry in air. EELS measurements were performed using a JEOL 2100F with integrated GATAN PEELS spectrometer with the help of Judith Mantell (University of Bristol) and Neil Wilkinson (Gatan, Inc.).

5.2.2.7 Small Angle X-ray Scattering

SAXS experiments were performed on two separate instruments: B21 Solution state SAXS beamline at Diamond Light Source, and a SAXSLAB Ganesha 300XL instrument at the University of Bristol.

Three samples were prepared for the former: 100 μM, 200 μM and 100 μM +1% sucrose SAGE in HEPES buffer. In each case, the hubs were combined 1 hour prior to measurement. 18 x 10 s exposures were taken for each sample and corresponding buffer (for background subtraction) giving a total exposure time of 3 minutes per sample. This experiment was performed by Dr Robert Rambo through a submission service.

Only the 100 μM + 1% sucrose sample was run in Ganesha. It was prepared immediately before being loaded into the machine; however, the blank and buffer measurements were performed prior to sample measurement so the sample was effectively mixed for 14 hours before measurement. Each measurement is recorded as an average over one 7-hour exposure. This experiment was done with the help of Dr Annela Seddon.

Data analysis was performed using the ScÅtter software package, developed by Dr Robert Rambo at Diamond Light Source (http://www.bioisis.net/tutorial/9).

5.3 Results and Discussion

As discussed in the introduction to this chapter, SAGEs had previously been imaged by both SEM and AFM as reported by Fletcher *et al.* [6] and shown to be spherical particles with a radius of approximately 100 nm. It was also known that variations in composition could alter the mean diameter of particles. In order to begin to identify SAGEs in the TEM, we performed some preliminary imaging as described below.

5.3.1 Initial Sample Characterisation

In the initial stages of the project, SAGEs with a variety of functionalisations were checked by TEM and fluorescence microscopy (FM) both in preparation for cryo-TEM and in collaboration with other projects. The example shown below shows positively charged SAGEs generated for silicification experiments [9].

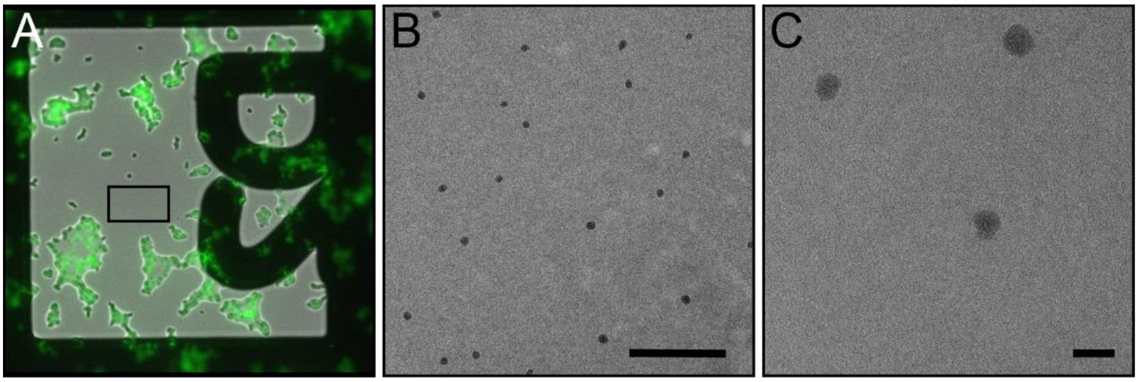


Figure 5-5: FM and TEM of SAGEs. (A)Widefield fluorescence and bright field overlay of SAGE on a finder grid. (B-C) TEM of unstained SAGE taken within the black box from (A). Scale bars = 500 nm (B) and 100 nm (C).

Figure 5-5 shows SAGE containing 5% fluorescent (carboxyfluorescein-hubB), and 45% positively charged (R4-hubB), prepared in 0.01 M phosphate buffer. (A) shows a fluorescence microscopy image collected by Dr Lorna Hodgson, in which large regions of fluorescent aggregates can be seen. Figures (B-C) show TEM images at two magnifications of regions within the black rectangle where well dispersed, individual particles can be seen.

These particles are of approximately the expected size, ~100 nm in diameter, although it was not possible to distinguish them by fluorescence. Since heavy metal staining causes fluorescence quenching, this imaging was performed without any stain. Despite this, particles showed an unexpected high level of contrast, particularly for the predicted unilamellar structure. There was a degree of heterogeneity in particle size and shape, however this was expected as studies of negatively stained liposomes show collapse and irregular edges due to the drying process [16]. A similar effect is expected for SAGE particles based on the original AFM experiments which found the height after drying to be approximately 10% of the diameter [6].

Based on these observations, it was predicted that cryo-TEM would provide much more useful structural information than standard room temperature TEM.

5.3.2 Cryo-Correlative Light Electron Microscopy

To locate SAGE particles in a cryo-TEM image, it was decided to use cryo-correlative light and electron microscopy (cryo-CLEM). Since SAGEs are easily functionalised with small fluorescent tags, we initially began work with 5% fluorescein labelled SAGEs.

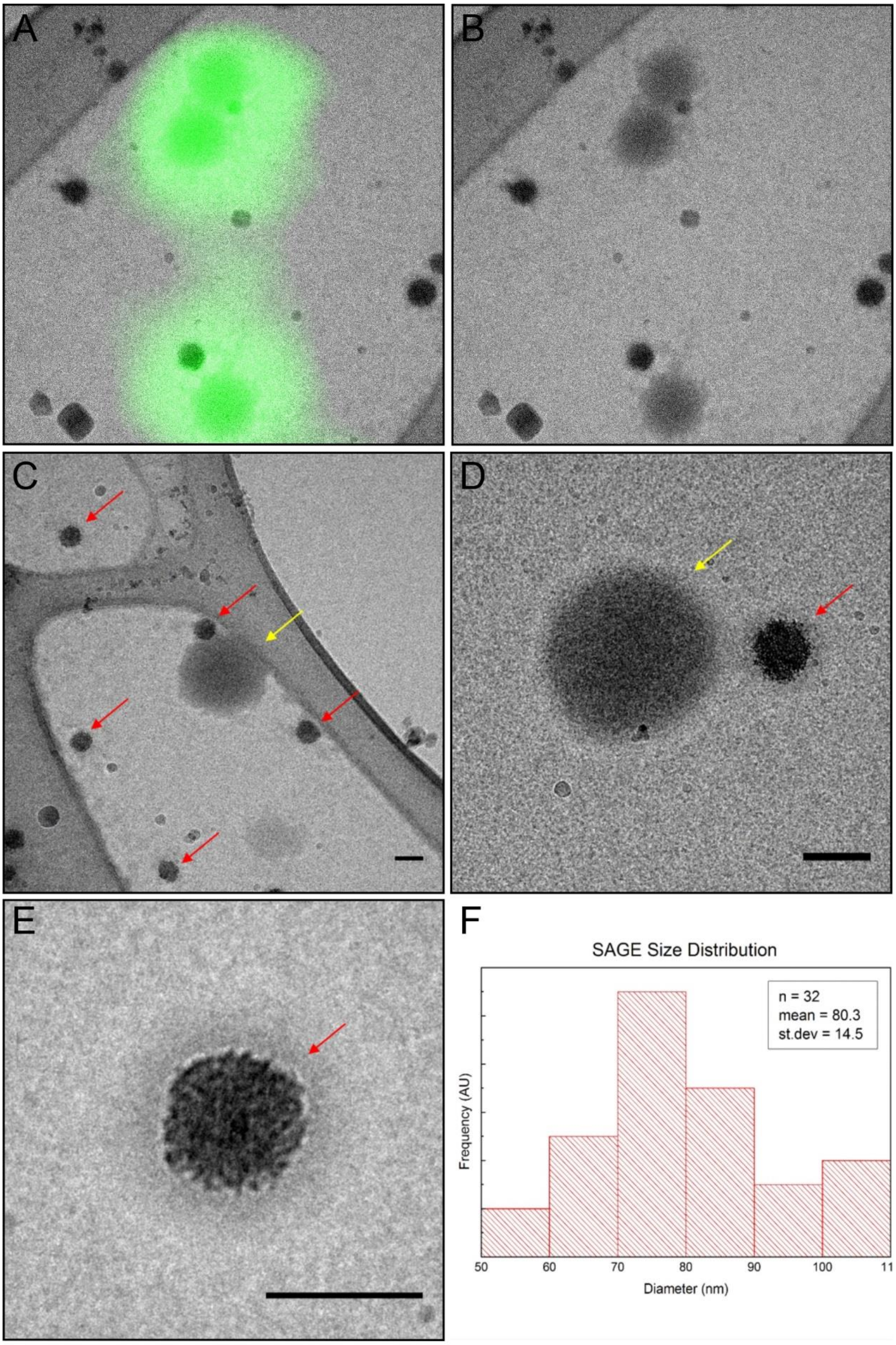


Figure 5-6: Cryo-TEM with fluorescence overlay (A) and TEM only showing particles suspended in vitreous ice (C-E). Red arrows indicate suspected SAGE particles while yellow arrows indicate areas that correlate with bright spots in cryo-fluorescence microscopy. Scale = 100nm. (D) Size distribution of particles similar to those indicated by red arrows.

Initial experiments were performed on SAGEs prepared in cell medium (preliminary work for [8]) and results are shown in Figure 5-6. This grid contained few areas of well frozen ice, however particles were clearly visible held suspended within the ice. These particles, indicated by red arrows, appear very dark, again suggesting unexpectedly high electron density for organic material. A number of images were taken, and the mean diameter found to be approximately 80 nm for a total of 32 particles. The size distribution is displayed in Figure 5-6(D) showing a similar particle size to that reported [6].

Prior to imaging by TEM, this grid had been imaged by cryo-fluorescence microscopy (cryo-FM) in order to correlate the two. Surprisingly it was found that the fluorescent regions correlated with the more diffuse areas (yellow arrows). Very few of these particles were imaged so a size distribution has not been measured. It was thought at the time that the lack of correlation between the dark particles and the fluorescence could be due to the fluorescent hubs not being fully integrated into the particles. It is also possible that the dark particles were fluorescent but not bright enough to be visible above the background signal from the ice [7]. Large regions of aggregated fluorescence were also seen (not shown), in agreement with the room temperature experiment in Figure 5-5.

Figure 5-7 shows TEM images taken from the paper ‘Large-deformation and high-strength amorphous porous carbon nanospheres’ by Yang *et. al.* [17]. This demonstrates the type of contrast which would be expected from a hollow particle (A) compared to a filled, porous one (C). The edges of a hollow sphere would be expected to have increased contrast compared to the centre resulting in a ring-shaped projection in the TEM. This is the type of contrast that would be expected from SAGEs according to the model, which suggests a thin layer of peptide forming a hollow sphere. This is not the case in the recorded images however, for either type of particle shown, which have approximately constant contrast throughout suggesting a solid rather than hollow particle.

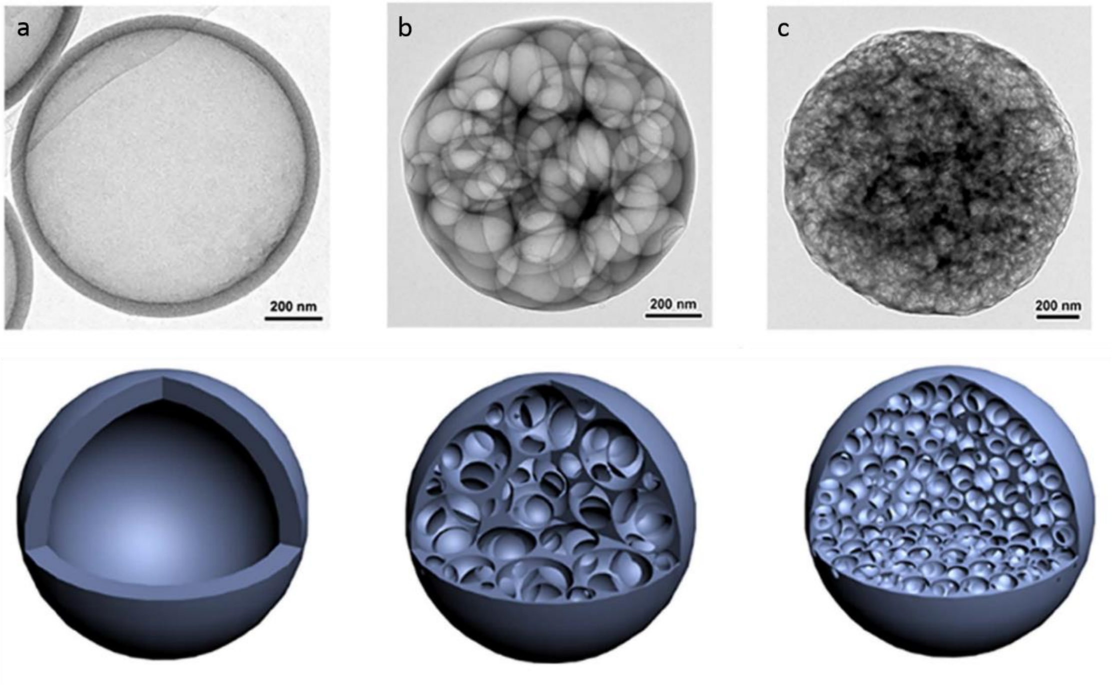


Figure 5-7: TEM images of Carbon Nanospheres with various pore structures and corresponding schematic diagrams below. Modified from [17].

5.3.3 Cryo-Tomography

Many variations of SAGE composition and freezing conditions were tested, however none yielded a sufficiently homogeneous population of particles to be appropriate for single particle analysis (Figure 5-8(A-D)). Cryo-tomography was therefore performed in order to gain an understanding of the internal structure and composition. In this case a composition of 50 % hubA + 50 % hubB SAGE in a 50:50 mixture of phosphate buffer and serum free cell medium, at a SAGE concentration of 10 µM. K4 hubs have an additional four lysine residues on the N-terminus of each trimer peptide giving the particles an overall positive charge. Images of the particles formed are shown in Figure 5-8(A-B). (A) shows a lacey carbon grid with a good coverage of very dark particles. Interestingly, this sample also appeared to contain a mixed population of particles, with both large particles seen in the centre and smaller, more aggregated ones towards the left-hand side of the image. (B) shows a higher magnification image of the same particles, this time on a Quantifoil grid.

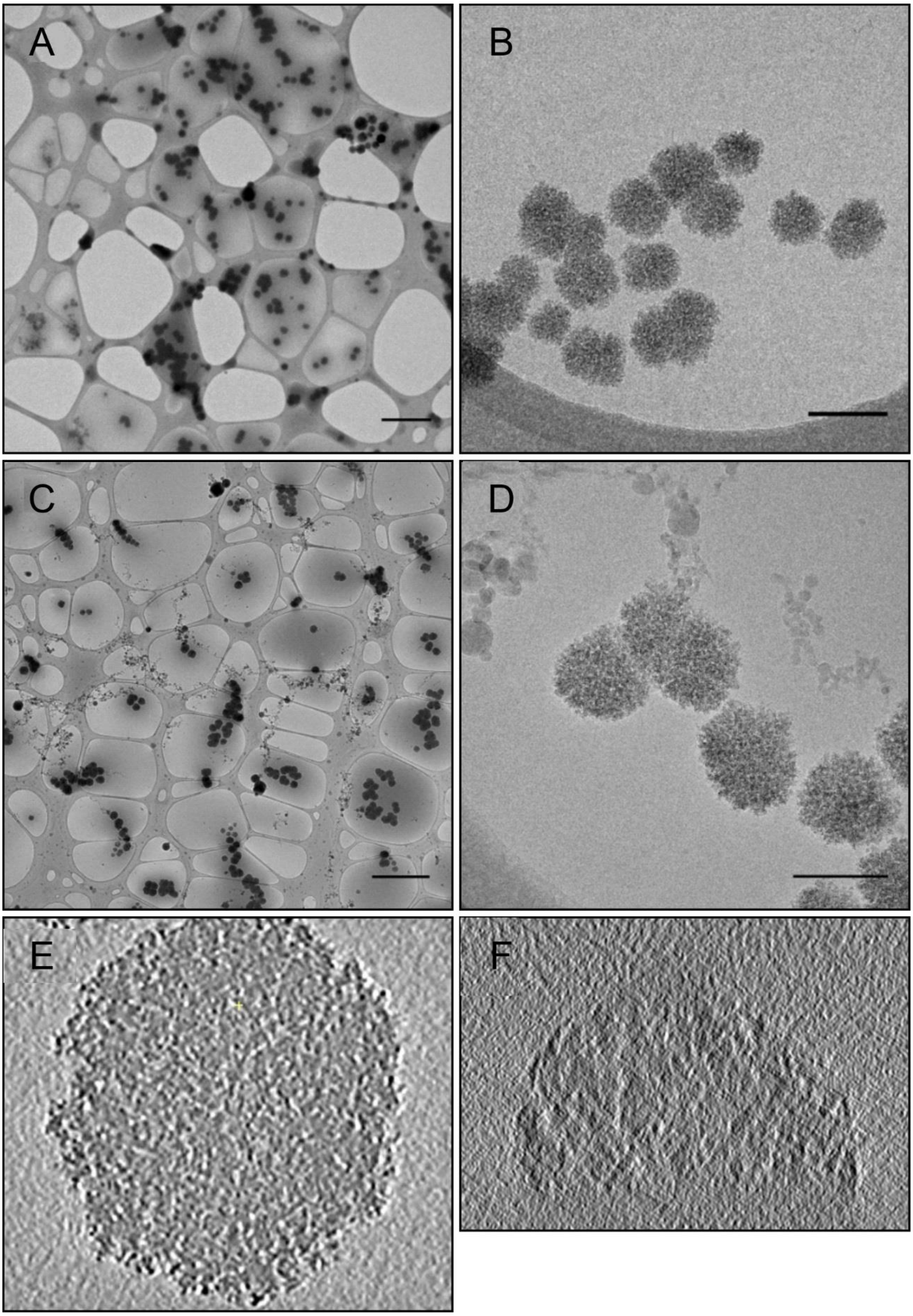


Figure 5-8: Cryo-TEM images and tomograms. (A-B) Cryo-TEM images collected on T20 microscope at the University of Bristol. (C-D) Cryo-TEM images collected on the Titan Krios at Diamond Light Source. (E-F) Slices of a reconstructed tomogram (E) in the plane of the ice, (F) perpendicular to the plane of the ice. Scale (A,C) 2 µm, (B,D) 200 nm.

Images collected on the Titan Krios microscope at Diamond Light Source are shown in Figure 5-8(C-D) alongside similar magnification images taken on the in-house T20 for comparison. Image resolution is greatly enhanced, thanks to a more coherent electron source (FEG vs LaB_6) and better cameras (Direct electron vs CCD), and the heterogeneous nature of the particles is clear. Comparison with example projection images in Figure 5-7 clearly demonstrates that these particles are not hollow. Several tomograms were collected, each containing multiple particles. Following reconstruction, it was confirmed that the particles were indeed filled throughout. A slice taken from a reconstruction of one particle is shown in (E). It shows inhomogeneous density with channels or pores through the full volume. There is no indication of a shell structure surrounding this density which would suggest that it is not an impurity being encapsulated within a shell. Fourier transforms of the images did not reveal any order within the structure indicating a lack of any extended crystalline structure.

A further observation, shown in (F), was that the particles were not fully spherical but rather had a flat surface along the ice-water interface. Adsorption of protein to this interface is a well-known phenomenon and has been discussed in detail in a recent paper by Robert Glaeser [18]. The relatively large size of the SAGE particles compared to many which are imaged by cryo-TEM may mean that this effect is much more visible.

5.3.4 'Clean' SAGE preparation

Two possible origins for the unexpectedly high electron density and non-hollow nature of the particles were proposed. Firstly, that salts in the buffer were being sequestered by the peptides, leading to a dense, amorphous particle. The second was that there are dense salts which remain associated with the peptide following the synthesis.

5.3.4.1 Buffers

In order to investigate the former, a screen of different buffers was performed by Dr Harriet Bray. It was found by AFM that SAGE particles were better distributed on a mica sheet when deposited from HEPES buffer than from phosphate buffer, with much less aggregation observed. As a result of this, all SAGEs were subsequently prepared in HEPES buffer at pH 7.2 and without salt. Cryo-TEM images of SAGEs prepared in this way shown in Figure 5-9(A-B).

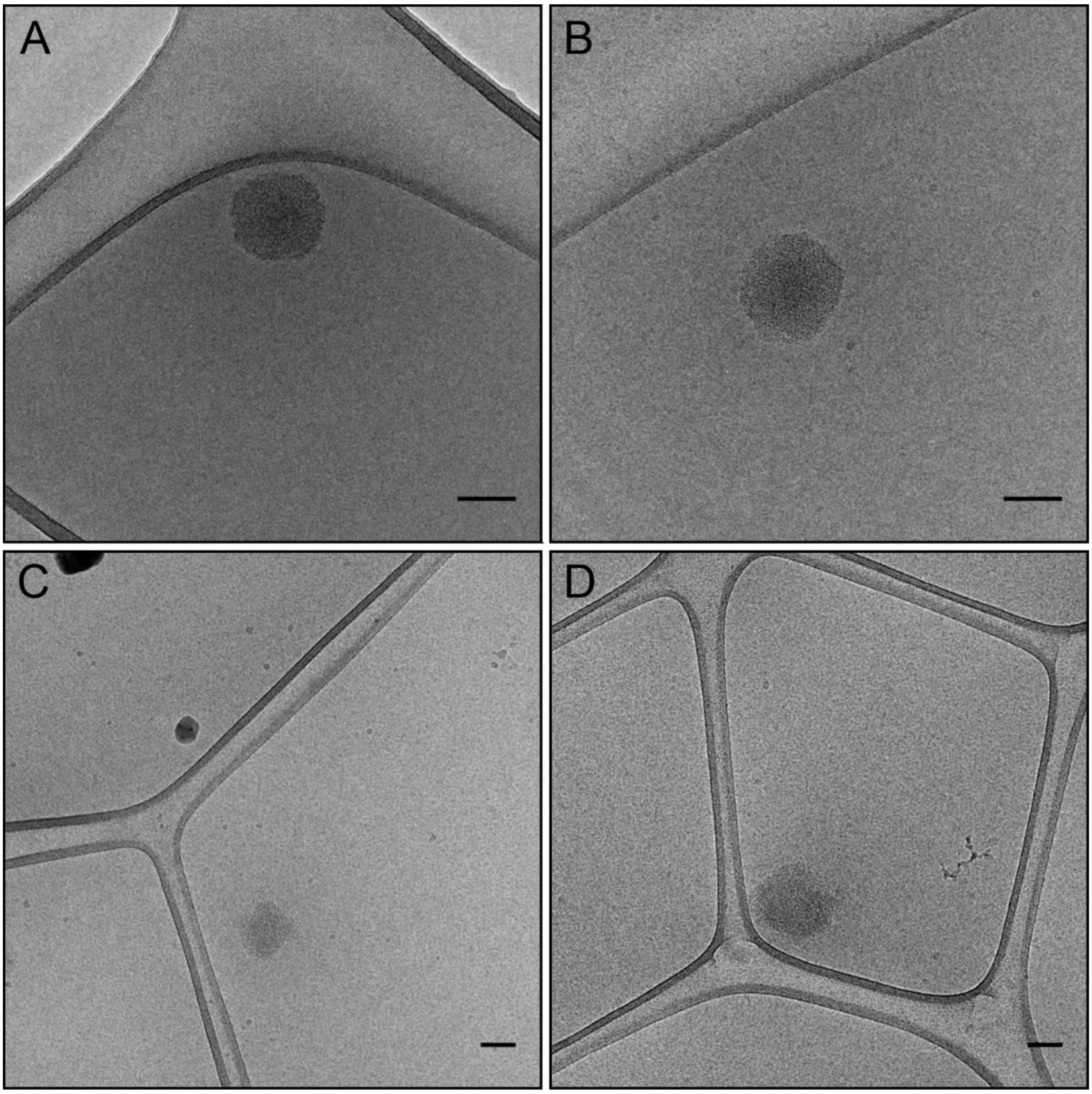


Figure 5-9: Cryo-TEM of 'clean' SAGE. (A-B) 10 µM SAGE in HEPES, (C-D) 10 µM desalted SAGE in HEPES showing less regular size and shape. Scale bars = 100 nm.

5.3.4.2 Salt Substitution

Secondly, a salt substitution experiment was performed as follows in order to minimise any contrast in the TEM arising from residual salt from the peptide synthesis process. Salt substitution was performed on purified, unfunctionalized hubs. The amount of TFA present after peptide synthesis and purification by HPLC was assumed to be 30 TFA molecules per peptide. After dilution, the concentration of peptide in solution was determined and a 30x molar equivalence of TFE added to solution. ^{19}F NMR, a technique used to identify and quantify relative abundance of fluorine nuclei in different chemical states, revealed the number of TFA molecules associated with each hubA and hubB molecule to be 2.4 and 8.9 respectively. The lower value for hubA is as expected due to the differing charges of the heterodimer peptides.

After desalting, ^{19}F NMR revealed TFA levels of 0.84 and 2.2 molecules per hubA and hubB molecule respectively. Thus, for both hubs the amount of TFA was decreased by 3-4 times (Figure 5-10). Cryo-TEM of the resulting ‘cleaned’ particles was performed and is shown in Figure 5-9(C-D).

It should be further noted that while this is labelled a ‘desalting’ experiment, the TFA (CF_3CO_2H, MW: 114) molecules originally present are simply being replaced by chloride ions. (Cl-, MW: 35).

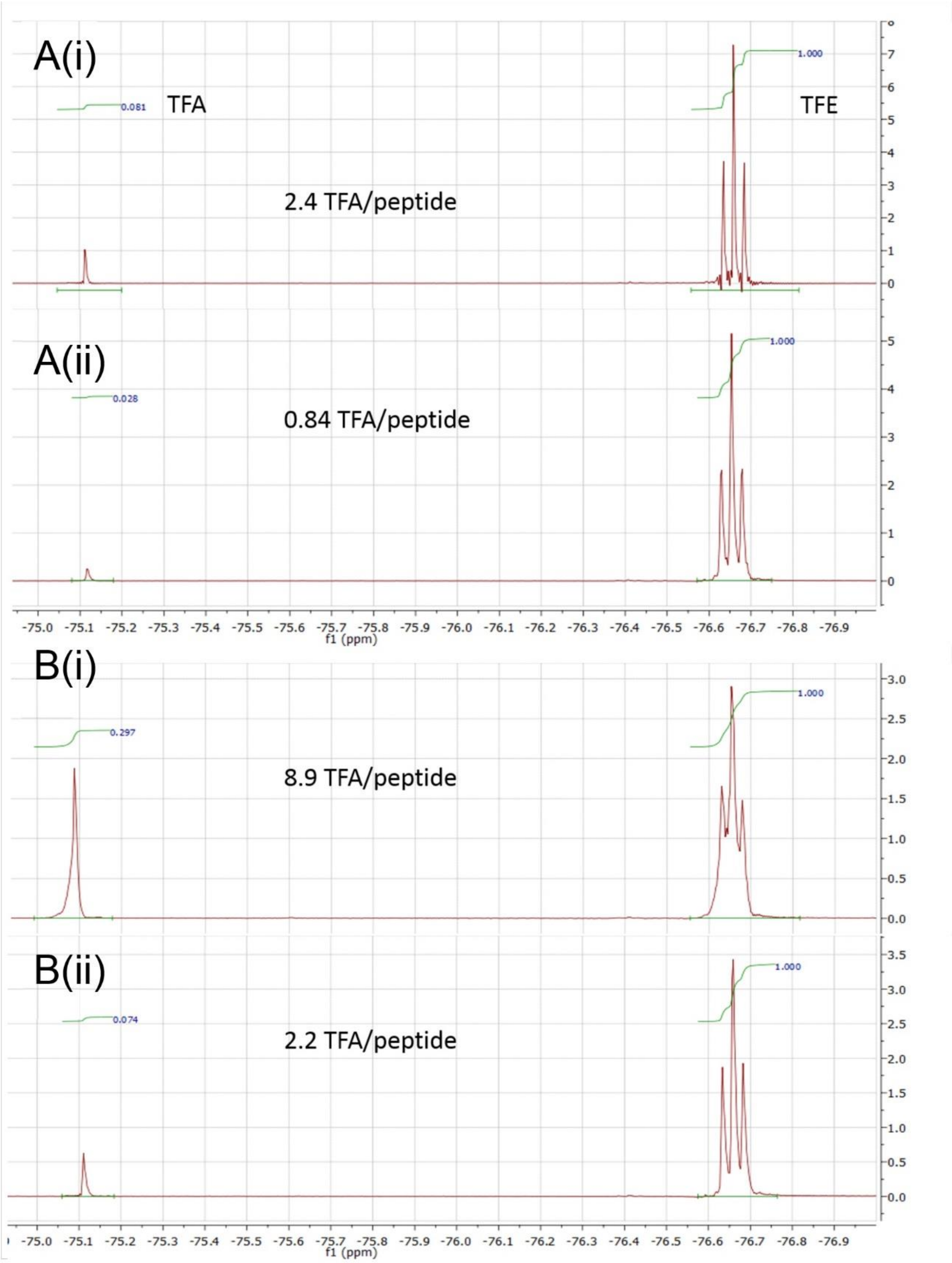


Figure 5-10: ^{19}F NMR spectra showing amount of TFA present in HubA (a) and hubB (b) before (i) and after (ii) desalting.

Resulting particles from these two methods were approximately the same size as those imaged previously, however the electron density appears to be greatly reduced. It was noted however, that the desalted particles (Figure 5-9(C-D)) exhibited a less regular shape and size. It was therefore decided to continue with the standard preparation method with HEPES buffer.

5.3.5 Cryo-CLEM with GFP SAGE

Thus far it has been demonstrated that we have particles of approximately the size as SAGEs shown by SEM [6], which can be prepared in a salt-free buffer, reducing their contrast in TEM to what would be expected for organic matter. It was suggested that there may also be particles in the solution which were not being visualised in the TEM due to extremely low electron density. Since the proposed model suggests a hollow shell only 3 nm thick, these may not produce enough contrast to be seen.

Dr James Ross in the Woolfson group produced protein SAGEs which, rather than containing small fluorescently tagged hubs, contained hubs conjugated with fluorescent proteins such as green fluorescent protein (GFP) [7]. This protein could help increase both the density of the shell and the brightness of the particle in cryo-FM due to the higher quantum yield of GFP compared to fluorescein.

As such, cryo-CLEM was performed on SAGE particles consisting of 10% GFP linked to the c-terminus (internal surface of cage according to modelling) of the hubB trimer peptide.

Results of this correlative microscopy are shown in Figure 5-11. Firstly, a cryo-FM image was recorded. The same area was then located in the TEM and a map of the grid square of interest was overlaid with the fluorescence image (A). Higher magnification images of the fluorescent regions were also collected (B-H).

The SAGEs functionalised with GFP have a very similar morphology to other SAGEs previously imaged and, again, appear to have a uniform density rather than a shell type structure. We can clearly see very bright fluorescent spots which correlate to aggregates (B-C) as well as smaller ones corresponding to individual particles (D-E). There were no fluorescent spots where a SAGE-like particle or aggregate was not found, however there were a few particles which did not appear to be fluorescent (G-H). These were the smallest particles found, with diameters of 130 nm or less. This may simply indicate that at this proportion of fluorescent hub, the smallest of particles do not have enough GFP molecules to give a signal higher than the background fluorescence of the ice.

These data serve to confirm that the particles imaged are indeed peptide constructs, and that the fluorescent protein is incorporated into the particle. There is no indication that there are additional, low-contrast, hollow particles in solution.

This experiment proved much more successful than the previous cryo-CLEM experiment shown in Section 5.3.2. One cause of this was simply user familiarity with both plunge freezing and cryo-fluorescence systems. However, it was also clear that the use of GFP rather than carboxyfluorescein resulted in improved fluorescence signal. Furthermore the ‘cleaner’ SAGE preparation with minimal salt means that we no longer see the dual particle population, with all particles other than the very smallest being visible in the fluorescence.

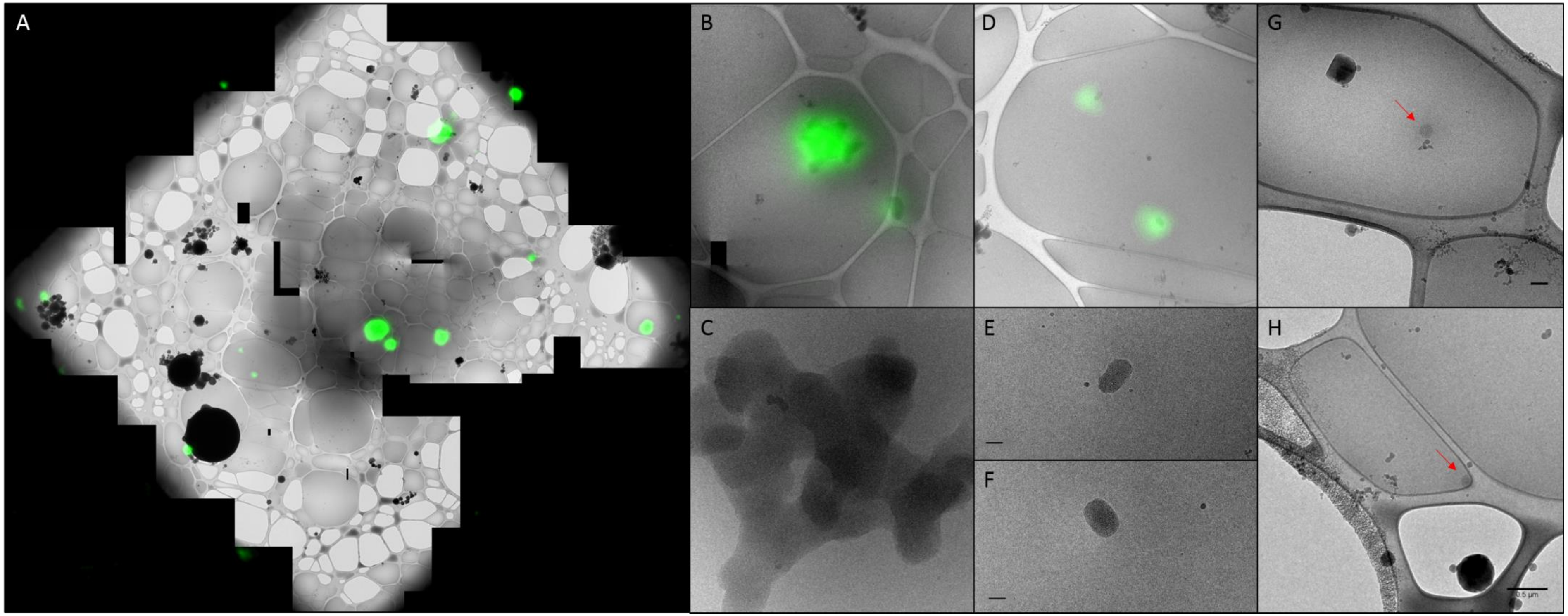


Figure 5-11: Cryo-CLEM of GFP functionalised SAGE. (A) Overlay of cryo-fluorescence microscopy image with mosaic of Cryo-TEM images. (B-C) overlay and EM only images of fluorescent aggregate. (D-F) overlay and EM only images of individual fluorescent particles. (G-H) SAGE-like particles (indicated by red arrows) which did not have detectable fluorescent signal. Scale bars (B-G)=100nm, (H)=500nm.

5.3.6 Time Course

To understand the longer-term dynamics of the system, data were collected using three complementary techniques daily over the course of a week. Parent SAGEs (50 % hubA + 50% hubB, 100 µM in 25 mM HEPES) were formed in a quartz DLS cuvette and half of the solution decanted into an Eppendorf for TEM after 1 hour of DLS measurement. In the following data, T=0 indicates 1 hour after mixing. DLS, AFM and TEM data were collected daily over 7 days and are shown below.

5.3.6.1 Dynamic Light Scattering

Dynamic light scattering, or DLS, is a common technique which gives information on the size of particles in solution. Particles scatter a laser beam while undergoing Brownian motion. This scattering is detected over time which gives information on the speed of motion which is then used to calculate the particle size [19].

DLS data showed that the particles maintained a roughly stable size over the course of the first three days, with a slight decrease in days 1 and 2, followed by a period of growth (Figure 5-13). We found that this trend was repeatable however the absolute size and rate of growth was not consistent between different runs. This was likely due to changes in ambient conditions such as the temperature in the laboratory which was not controlled. The decrease in the first 2-3 days may be explained by the fact that particles appear initially to be very ill-defined, ‘clumpy’ aggregates, while later they form discrete spheres (see discussion of TEM data below). There is also a spike on day 7 which may be due to the growth of bacteria as seen by TEM.

5.3.6.2 Atomic Force Microscopy

Atomic force microscope (AFM) is an imaging technique in which a probe is scanned over a surface and detects changes in height as well as surface characteristics such as stiffness [20]. AFM data were collected by Dr Harriet Bray and are in agreement with the DLS results until day 2. Unfortunately following this timepoint the particles became very ‘sticky’, meaning the probe was not able to scan cleanly over the surface, making AFM size measurements unreliable after this point. Thus, only days 0 to 2 are reported here. Further analysis of these data performed by Dr Harriet Bray, shows that particle height has a linear dependence on diameter. This contradicts the claims of the 2013 paper [6] which states that particles have a height approximately equal to that of two ideal monolayers. This original data however, was only collected from a small number of particles (n=5) while new data is taken from a much greater number (n ≈ 5000). Days 1 and 2 are in excellent agreement with height being roughly 50% of diameter, while particles measured after only 1 hour appear to collapse more.

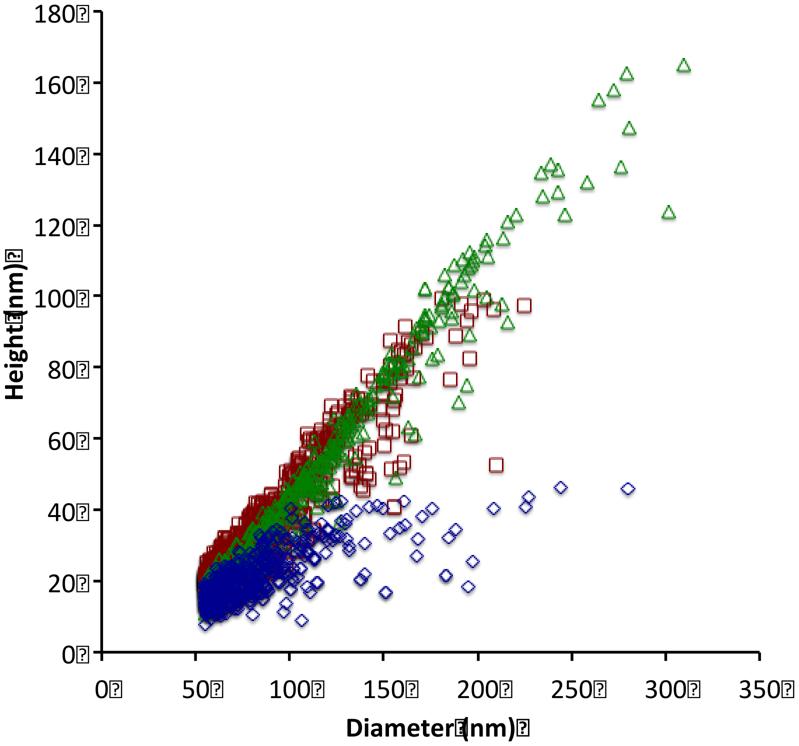


Figure 5-12: AFM of particle height vs diameter measured after 0 days (blue), 1 day (red) and 2 days (green).

5.3.6.3 Cryo-TEM

Representative images of SAGEs frozen at each time point are shown in Figure 5-13. At both 0 (not shown) and 1 days, the SAGE material appears to form non-discrete network type structures. These include large aggregates and smaller, irregularly shaped particles. It was not possible to make an estimation of the size of these particles from the TEM data due to the irregularity and overlapping nature. Particles frozen on subsequent days had roughly circular projections, which combined with intensity profiles would suggest solid, spherical particles.

The diameter of the particles tends to increase over time and at day 5 we begin to capture connecting particles. This indicates that the particles are coalescing and rearranging to form larger spherical particles in order to minimise surface area.

By days 6 and 7, particles have a mean diameter of 358 nm and 407 nm respectively however some are as large as 1 µm. At these later timepoints some particles appear to have denser cores compared to the outside, the density gradient is no longer consistent with that of a sphere of uniform density. We are unsure of the cause of this, however one explanation could be that a dense core has formed in the centre while the peptides on the periphery continue to rearrange following joining of particles. Alternatively, it may be that those peptides on the outside are beginning to degrade, which had been observed previously by performing HPLC on hubs which had been in solution for upwards of a week.

Bacteria were also seen in the sample at the day 7 time point. This could have been introduced while taking out a sample for an earlier time point as none had been seen at other points. This may also be the cause of the anomalous spike in diameter measured by DLS on day 7.

The time course was concluded after one week of measurements.

In order to quantitatively describe the cryo-TEM data, particles were selected and measured by hand using the Fiji software package [13]. Figure 5-13 shows the histograms of measured particle diameters for each time point showing the clear growth over time. The polydispersity (standard deviation^2/mean^2) is on the order of 15 - 25% at all time points with the spread in diameters increasing with the mean.

For comparison, the mean particle diameter and standard deviation have been reported for each measurable time point for the three techniques used and are displayed in Figure 5-13. Day three is the only timepoint at which it was possible to measure by all three methods. We can see that at this point, all methods are in good agreement. The TEM and DLS then follow the same trend of increasing diameter however they deviate significantly with DLS showing a sharper growth. The DLS data may be skewed due to aggregates in solution which would be approximated to a single spherical particle. Such aggregates would be discounted in the TEM as particles of interest were selected manually, while very large particles may not be well suspended in the ice and therefore may not be imaged at all.

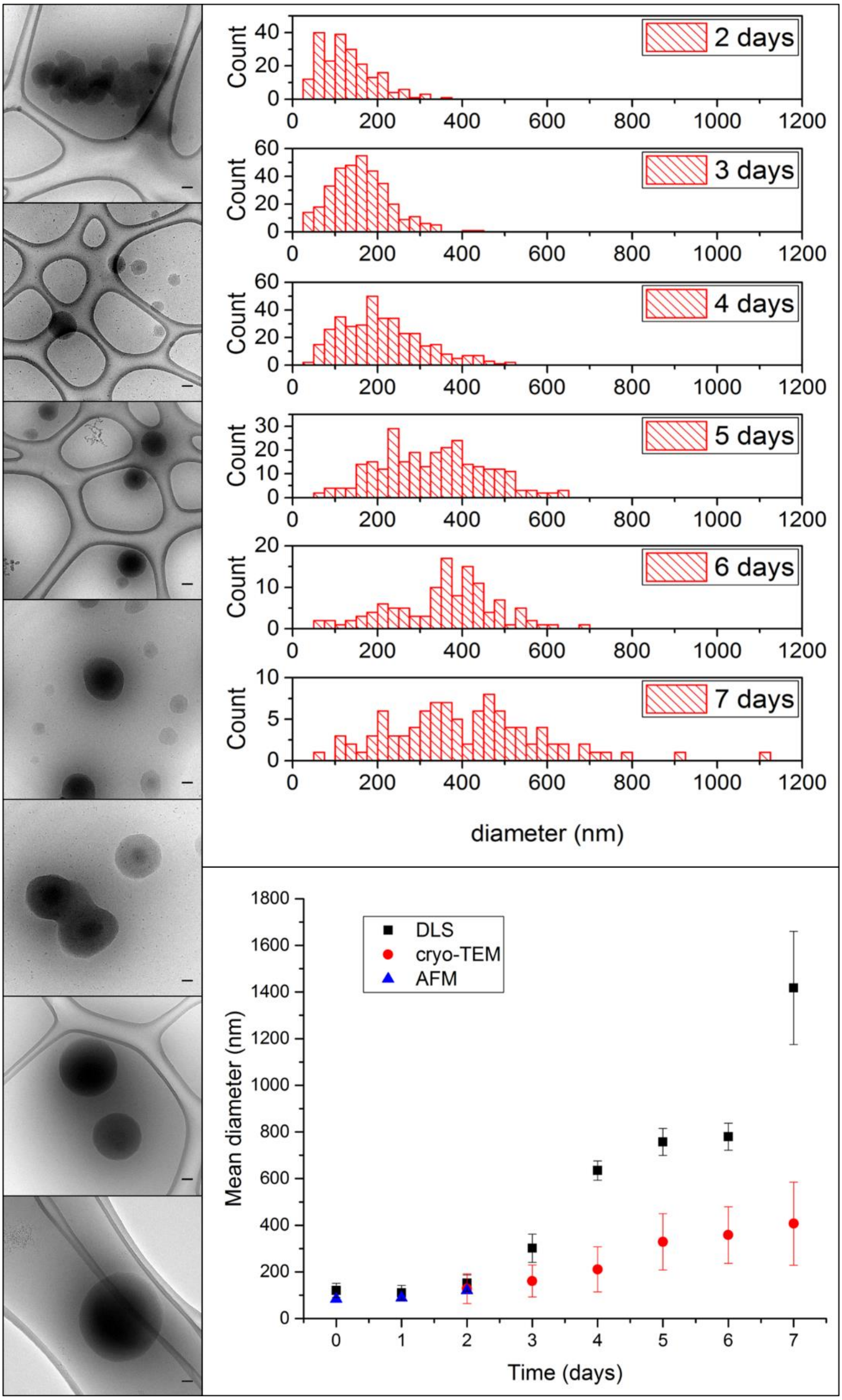


Figure 5-13: Time course. (Left) Example cryo-TEM images from day 1 to 7. Scale = 100nm. (Right, top) histograms showing particle diameters as measured from cryo-TEM images. (Right, bottom) Comparison of particle diameter as measured by DLS, cryo-TEM and AFM.

5.3.7 Electron Energy Loss Spectroscopy

Electron Energy Loss Spectroscopy (EELS) is a technique in which the energy of inelastically scattered electrons is recorded after passing through a sample in a TEM. The resulting energy spectrum then provides information about the types of atoms present in the sample [21]. This experiment was performed in order to gain more insight into the composition of the particles being imaged thus far.

The SAGEs used were prepared identically to those used in the time course and dried onto a lacey carbon film. Areas where SAGE particles appeared to be suspended over holes were found as shown in Figure 5-14(A). It was possible to identify peaks relating to carbon, nitrogen and oxygen but there was no measurable evidence of elements such as phosphorus, sulphur or sodium. This suggests that the SAGEs are not filled with salts of any of these elements to any measurable degree, but rather peptide only. Elemental mapping, Figure 5-14(B-D), also shows that while oxygen can be found all over the grid, the nitrogen is found exclusively at the location of the SAGE particle. This confirms the presence of peptide here and that these are indeed the SAGE particles which had been imaged using other techniques.

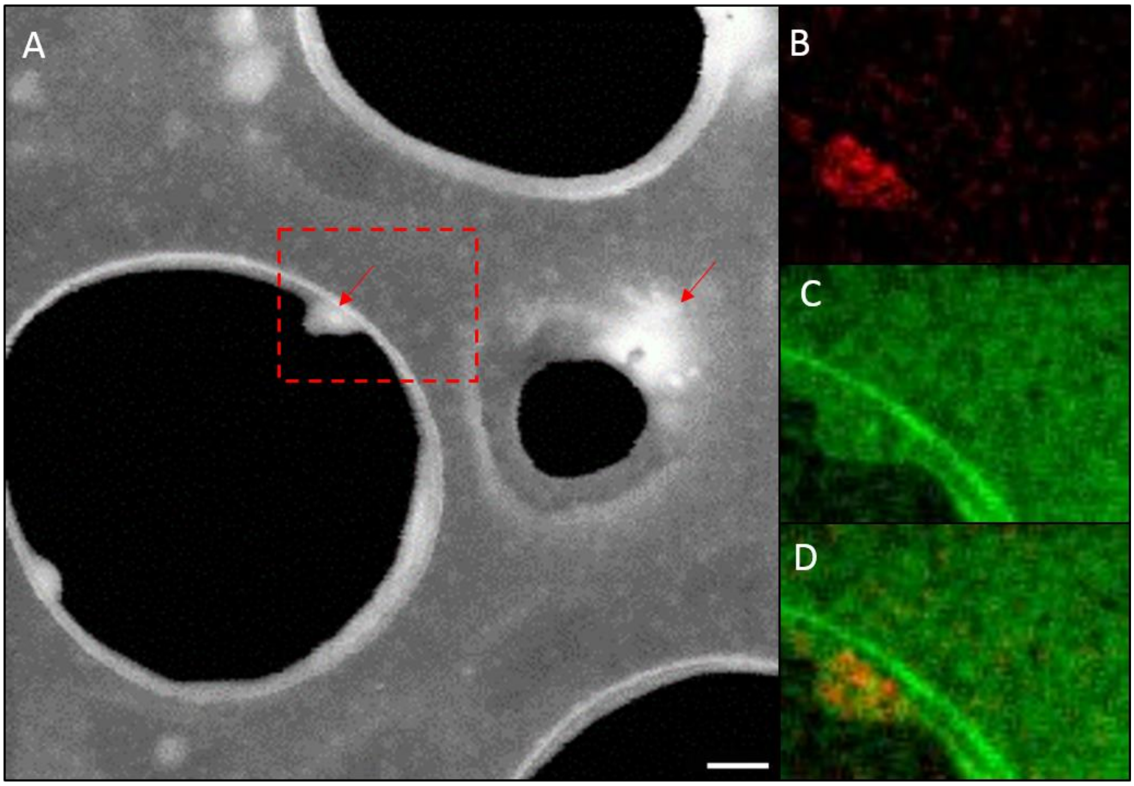


Figure 5-14: Elemental Mapping. (A) dark field overview of an area of lacey carbon grid. Black areas are holes in the grid, white patches indicated by red arrows are SAGE particles. Scale = 100nm. (B) Nitrogen, (C) Oxygen, and (D) overlaid elemental maps of area indicated by red box showing presence of Nitrogen only at location of particle.

5.3.8 Small Angle X-ray Scattering

Small Angle X-ray Scattering (SAXS) is another technique commonly used to characterise the size, and overall shape of biological macromolecules. An x-ray beam is passed through the sample in solution and the scattered beam is detected. The scattering pattern at small angles yields information on molecular weight and radius of gyration which can inform on how compact a particle is [22].

Two Small Angle X-ray Scattering experiments were performed. The first at diamond light source on the solution state SAXS beamline (B21) at Diamond Light Source. Samples were submitted at 100 µM and 200 µM, along with another at 100 µM containing 1% sucrose to minimise radiation damage [23].

Raw data from the three samples is shown in Figure 5-15(A). Each colour represents a separate 10 s measurement which would usually be averaged together and the background subtracted. However we can see that at low q values, the signal shifts over time for the samples containing no sucrose. This is a typical indication of radiation damage over time by the high energy of the beam. This is not the case for the 100 µM + sucrose sample however, indicating that the sucrose is indeed reducing this beam damage. Consequently, only the data from this sample was examined further.

A follow up experiment was performed on the University of Bristol’s instrument, Ganesha, which can perform Ultra small angle to wide angle x-ray scattering measurements, allowing it to probe a much wider range of length scales. In the ESAXS (Extremely Small Angle X-ray Scattering) mode, it has a q range equivalent to 3 nm to 200 nm.

The background subtracted curves for the two experiments are shown in Figure 5-15(B). The y-offset of the two data sets is due to the fact that absolute intensities were not measured in either experiment. We can see that the data recorded at the synchrotron extends to much higher q values, which equates to shorter length scales, while data from Ganesha extends to shorter q values. Unfortunately, neither dataset displays any clear features which may be used to extract useful information about the sample.

Figure 5-15(C) shows model scattering patterns for hollow, 100 nm spheres with a 3 nm thick shell, kindly provided by Dr Majid Mosayebi. Adding an appropriate level of particle polydispersity, we would expect the features to be masked somewhat, however we are still unable to see any similar features in our measured data.

We would interpret this in the following way: that the high resolution structure of the particles as measured at the synchrotron is not regular, as we would expect from the original model, but heterogeneous, with all length scales being explored and measured rather than those of a regular hexagonal lattice. It is known from the time course measurements reported in Section 5.3.6 that SAGEs change in size and morphology over time so the measurements made on Ganesha, which were an average of 7 hours of data recording, would mask any features which may have been visible at any one time during this period. Thus, unfortunately, the time resolution and the distance resolution advantages provided by the synchrotron and Ganesha respectively could not be taken advantage of with this particular system.

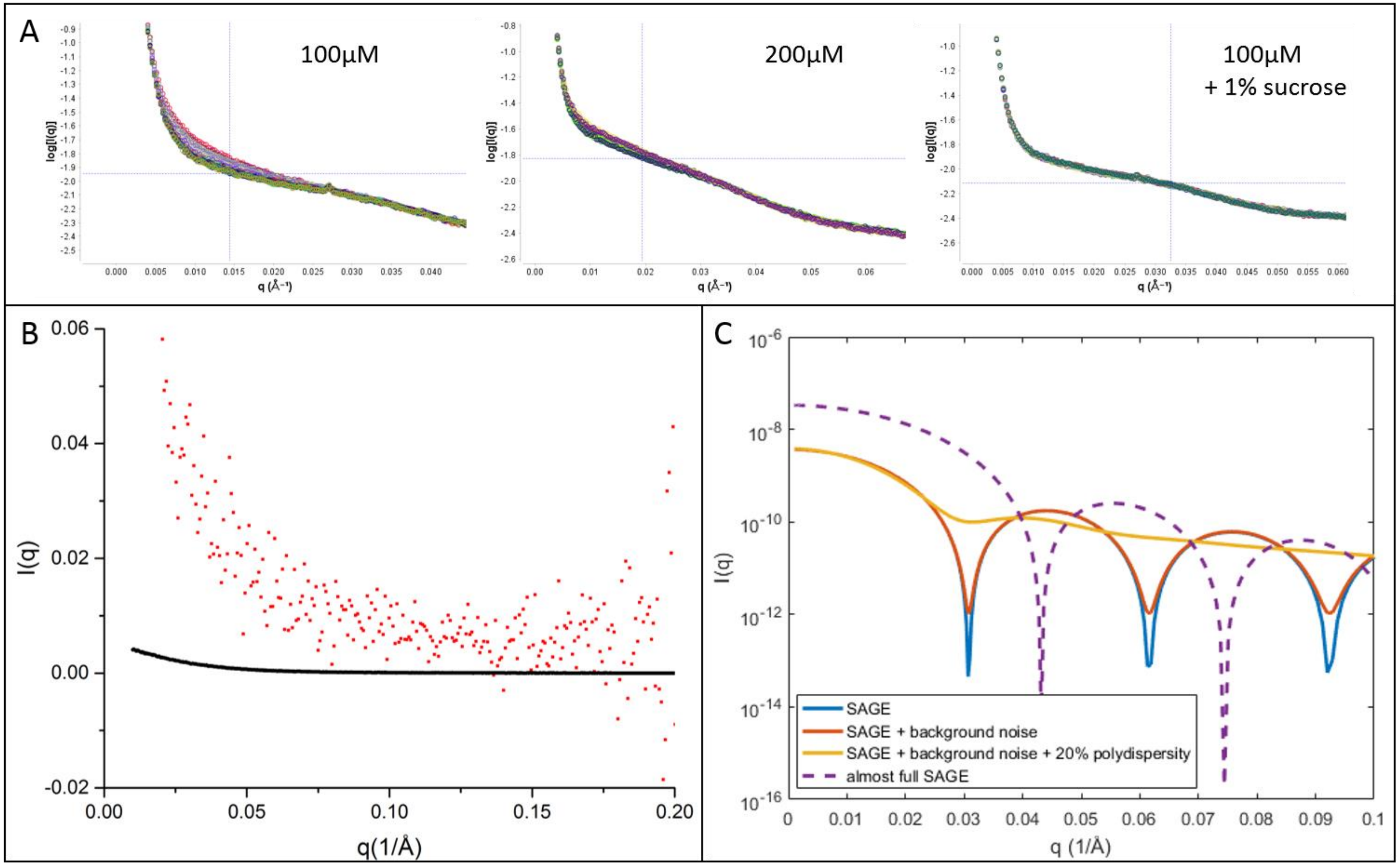


Figure 5-15: SAXS data. (A) Raw data collected at synchrotron showing radiation damage to the sample. (B) Background subtracted data collected at synchrotron (black) and on Ganesha (red). (C) Predicted scattering patterns from SAGE particles showing the effect of noise and polydispersity courtesy of Dr Majid Mosayebi.

5.4 Discussion

The aim of this study was to gain a deeper understanding of the structure of the *de novo* designed peptide nanoparticles first reported in [6]. Cryo-TEM was used alongside an array of other techniques included DLS, cryo-CLEM, SAXS and EELS, revealing that the structure may be considerably more complex than previously thought.

Early experiments showed particles with much higher than expected contrast in cryo-TEM. This prompted a change in buffers and particular focus on minimising any salts present in SAGE preparations. This served to reduce the contrast, however did not result in the projection pattern which would be expected from a unilamellar particle.

The published data showed surface structure and partly collapsed particle profiles which suggest a unilamellar cage-like structure, however the cryo-TEM projection and tomography data do not fully support this theory. Furthermore, the particles' ability to coalesce into larger structures, suggests an ability for internal bonds and structure to rearrange dynamically. In light of these surprising results, several alternative structural models were suggested, as illustrated in Figure 5-16.

The first possibility (A) is that the particles consist simply of aggregated peptide. Biochemical analysis of the peptides shows that the alpha helicity increases upon SAGE formation, indicating that the majority of peptides are forming coiled-coil interactions [6,24].

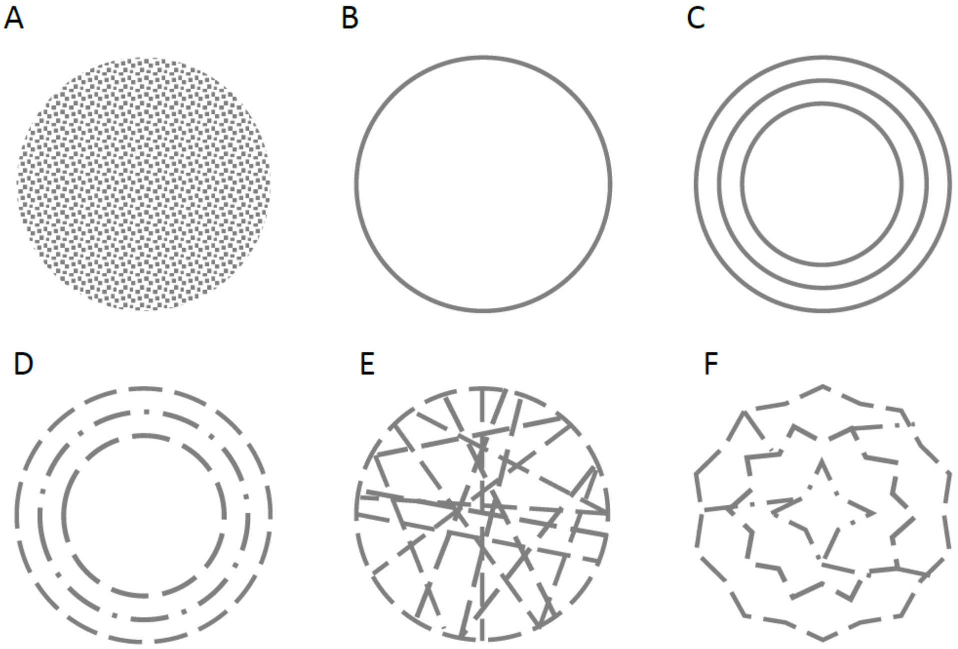


Figure 5-16: Schematic of possible SAGE structures. (A) aggregate, (B) unilamellar, (C) multilamellar, (D) layered patches, (E) sponge, (F) corrugated. Reproduced with permission from [24]

The second model shown is the unilamellar (B), originally proposed model which may be discounted as discussed previously. An alternative however could be a multilamellar structure (C). Such a particle would show the same surface structure as a unilamellar object but contain more material. However, in this case we may expect to see evidence on concentric rings in the cryo-TEM which were not visible. Drying down of such structures may also be expected to result in discreet values according to the number of layers present, while the data in fact show a continuum of values.

Aggregation of individual flakes of hexagonal lattices could result in a layered structure as seen in (D). This was not visible in AFM data studied by Dr Harriet Bray where there was no evidence of edges or steps on the particle surface [24].

The observation that the dimeric coiled-coils could rotate relative to the trimers [12], suggests that the assembly process may propagate in 3 dimensions rather than 2. This could result in a sponge-like 3D lattice (E) in which the majority of alpha helices would still be able to form coiled-coil interactions. Such a lattice may still be expected to display the surface structure visualised previously, while giving projection images similar to those seen by cryo-TEM. Such structures would likely have unfavourable, unbonded peptides on the surface which would explain why the particles are able to stick together and rearrange to form larger spheres. It is unclear how densely packed such a particle would be and may depend on the conditions during formation (*e.g*. temperature, concentration), but this could also account for the linear dependence of dried down particle height on diameter as seen by AFM.

The final proposed structure is composed of a combination of the layered and sponge models (F). At present there is no clear way of distinguishing between such particles due to their small size and heterogeneity. Clearly there is flexibility within the system which would allow such a combination would be present.

Cryo-TEM has proved vital in gaining a deeper insight into the structure of SAGE particles. While they proved unsuitable for a single particle-type analysis, the technique revealed key structural information which may otherwise not have been known.

References

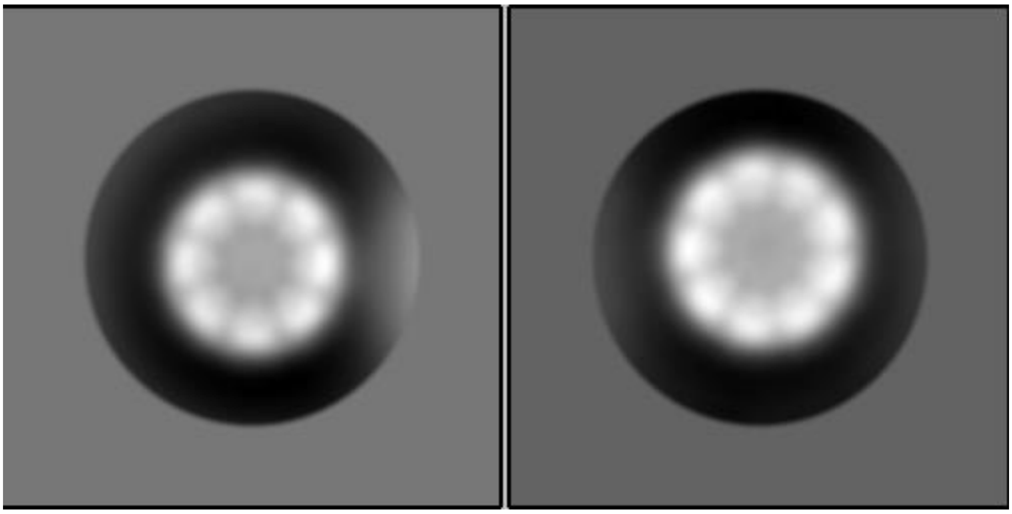
1. Roos, W.H., et al., *Viral capsids: Mechanical characteristics, genome packaging and delivery mechanisms.* Cellular and Molecular Life Sciences, 2007. **64**(12): p. 1484-1497.
2. Chasteen, N.D. and P.M. Harrison, *Mineralization in ferritin: an efficient means of iron storage.* J Struct Biol, 1999. **126**(3): p. 182-94.
3. He, D.D. and J. Marles-Wright, *Ferritin family proteins and their use in bionanotechnology.* New Biotechnology, 2015. **32**(6): p. 651-657.
4. Matsuura, K., *Synthetic approaches to construct viral capsid-like spherical nanomaterials.* Chemical Communications, 2018. **54**(65): p. 8944-8959.
5. Matsuura, K., *Construction of Functional Biomaterials by Biomolecular Self-Assembly.* Bulletin of the Chemical Society of Japan, 2017. **90**(8): p. 873-884.
6. Fletcher, J.M., et al., *Self-Assembling Cages from Coiled-Coil Peptide Modules.* Science, 2013. **340**(6132): p. 595-599.
7. Ross, J.F., et al., *Decorating Self-Assembled Peptide Cages with Proteins.* Acs Nano, 2017. **11**(8): p. 7901-7914.
8. Beesley, J.L., et al., *Modifying Self-Assembled Peptide Cages To Control Internalization into Mammalian Cells.* Nano Letters, 2018. **18**(9): p. 5933-5937.
9. Galloway, J.M., et al., *Bioinspired Silicification Reveals Structural Detail in Self-Assembled Peptide Cages.* Acs Nano, 2018. **12**(2): p. 1420-1432.
10. Morris, C., et al., *A Modular Vaccine Platform Combining Self-Assembled Peptide Cages and Immunogenic Peptides.* Advanced Functional Materials, 2019. **29**(8).
11. Mosayebi, M., et al., *Beyond icosahedral symmetry in packings of proteins in spherical shells.* Proceedings of the National Academy of Sciences of the United States of America, 2017. **114**(34): p. 9014-9019.
12. Shoemark, D.K., et al., *The dynamical interplay between a megadalton peptide nanocage and solutes probed by microsecond atomistic MD; implications for design.* Physical Chemistry Chemical Physics, 2019. **21**(1): p. 137-147.

13. Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis.* Nature Methods, 2012. **9**(7): p. 676-682.
14. Microsystems, L. *Correlative Light and Electron Microscope EM Cryo CLEM*. 17/02/2016]; Available from: http://www.leica-microsystems.com/products/em-sample-prep/biological-specimens/low-temperature-techniques/details/product/leica-em-cryo-clem/.
15. Thevenaz, P., U.E. Ruttimann, and M. Unser, *A pyramid approach to subpixel registration based on intensity.* IEEE Trans Image Process, 1998. **7**(1): p. 27-41.
16. Ruozi, B., et al., *AFM, ESEM, TEM, and CLSM in liposomal characterization: a comparative study.* Int J Nanomedicine, 2011. **6**: p. 557-63.
17. Yang, W.Z., et al., *Large-deformation and high-strength amorphous porous carbon nanospheres.* Scientific Reports, 2016. **6**.
18. Glaeser, R.M., *Proteins, Interfaces, and Cryo-Em Grids.* Curr Opin Colloid Interface Sci, 2018. **34**: p. 1-8.
19. Goldburg, W.I., *Dynamic light scattering.* American Journal of Physics, 1999. **67**(12): p. 1152-1160.
20. Rugar, D. and P. Hansma, *Atomic Force Microscopy.* Physics Today, 1990. **43**(10): p. 23-30.
21. Egerton, R.F., *Electron energy-loss spectroscopy in the TEM.* Reports on Progress in Physics, 2009. **72**(1).
22. Kikhney, A.G. and D.I. Svergun, *A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins.* FEBS Lett, 2015. **589**(19 Pt A): p. 2570-7.
23. Skou, S., R.E. Gillilan, and N. Ando, *Synchrotron-based small-angle X-ray scattering of proteins in solution.* Nat Protoc, 2014. **9**(7): p. 1727-39.
24. Bray, H.E.V., *Design and characterisation of self-assembling coiled-coil peptide cages.*, in *School of Chemistry*. 2018, University of Bristol.

Appendix I - All-alpha subunits

It was hypothesised that the 18mers found could consist of all-alpha particles as they are known to be able to form thermosomes in the absence of beta subunits [40]. In order to investigate this, a preparation of alpha-only thermosomes was prepared and studied by negative stain TEM.

2D classification of top views of resulting particles shows that both 16mers and 18mers were present with 18mers representing 32% of these particles. The 2D analysis was performed on RELION by Marston Bradshaw.



All-alpha 2D classification of top views with 16mer class on the left and 18mer on the right.

The fact that the all-alpha particles prefer to form 16mers particles than 18mers is not compatible will the hypothesis that the 18mers visible in the cryo-TEM data are all-alphas. In order to contribute 35% of the good particles, this would require either a very large mismatch in the relative concentrations of subunits during the preparation, or a very large number of surplus beta subunits in solution (beta subunits cannot assemble on their own).

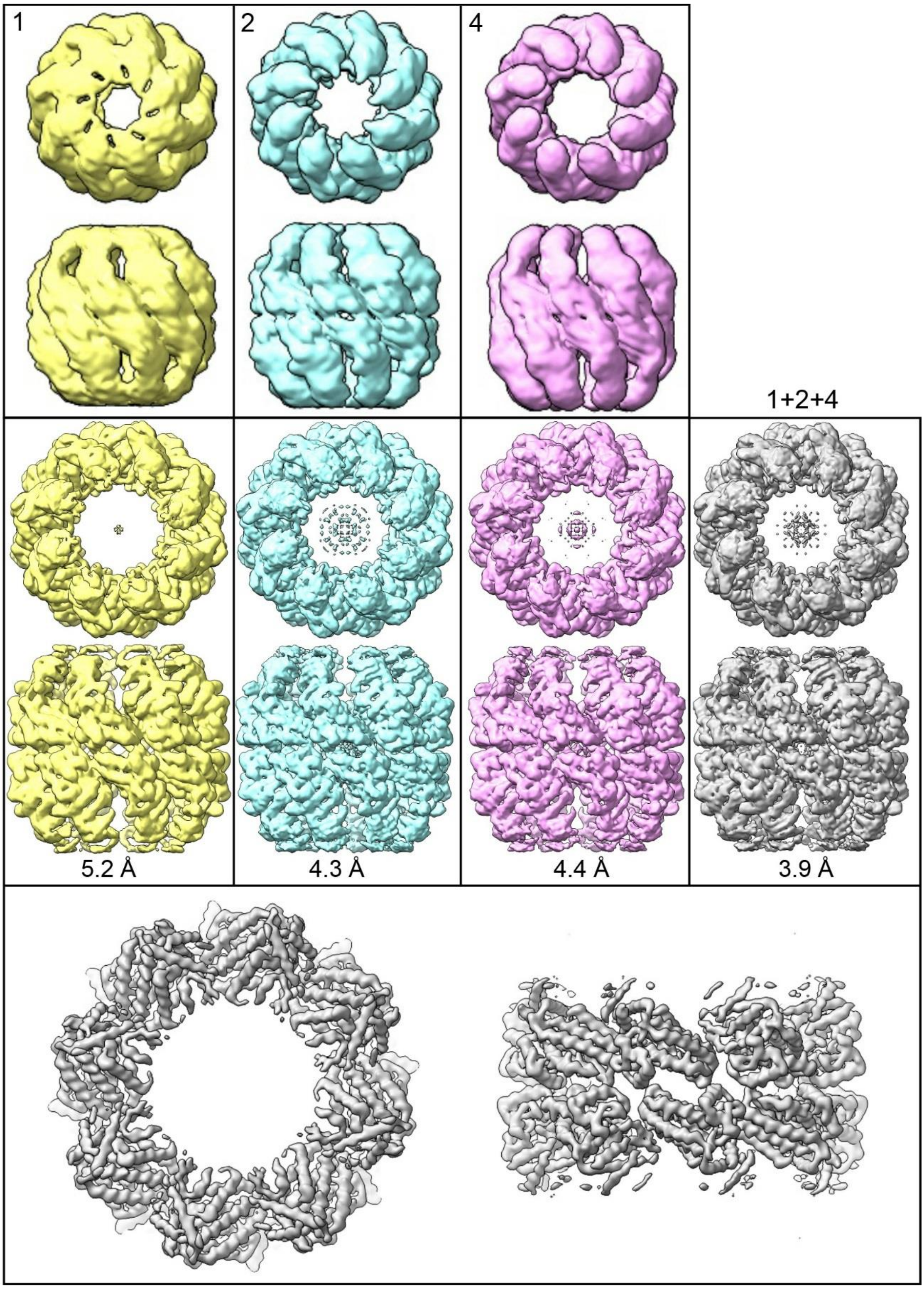
To assess this hypothesis further it is suggested that the all alpha samples are re-classified. Potentially using a 2D classification routines that are designed for negative stain data.

Further structural analysis of the 18mers was not taken forward after this point due to the relatively small numbers of particles, particularly for the mutant dataset.

Appendix II - Consistency checking 3D classifications of the Wild Type thermosome

Following on from section 2.5.8.1, the particles from each of the individual classes were refined using the low pass filtered map from class 2, giving resolutions of 5.2, 4.3 and 4.4 Å respectively. Once again, despite the higher number of particles in class 1, the resulting refined map had lower overall resolution than the two smaller ones: likely a result of a slightly mixed population included in this class.

Further refinements were therefore performed on a combination of particles from these three classes as well as individually. These results are shown in the figure below.



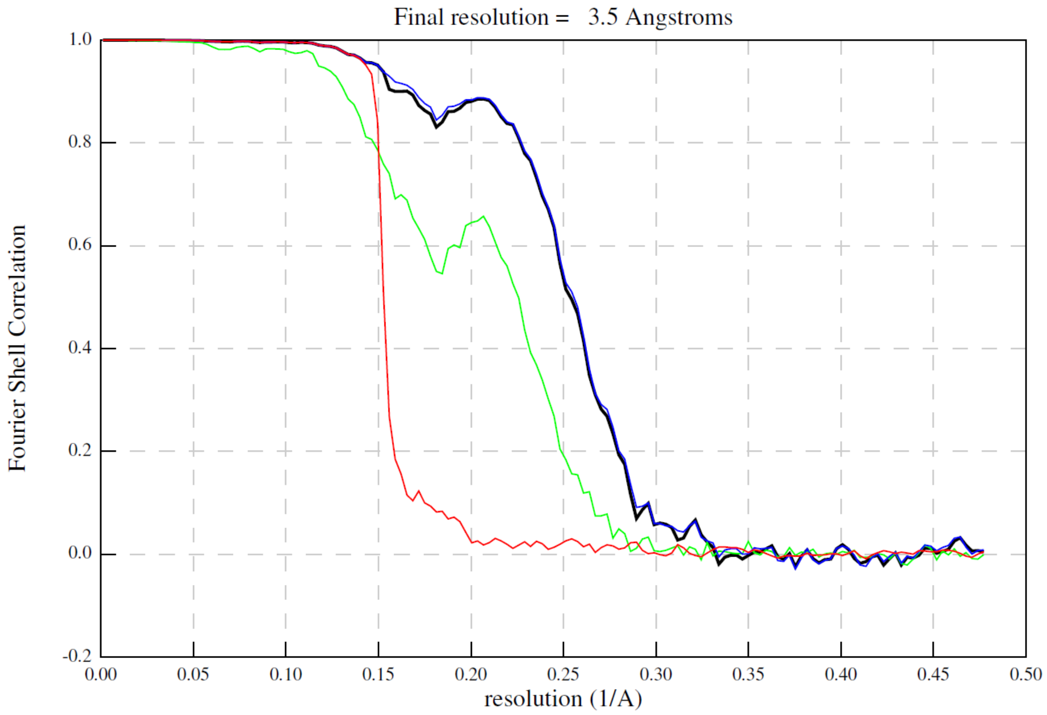
WT refinements: (Top) D8 classes 1,2 and 4. (Middle) Autorefinements of individual classes and combined classes with estimated resolutions. (Bottom) Higher threshold of the combined class map showing clear alpha helices.

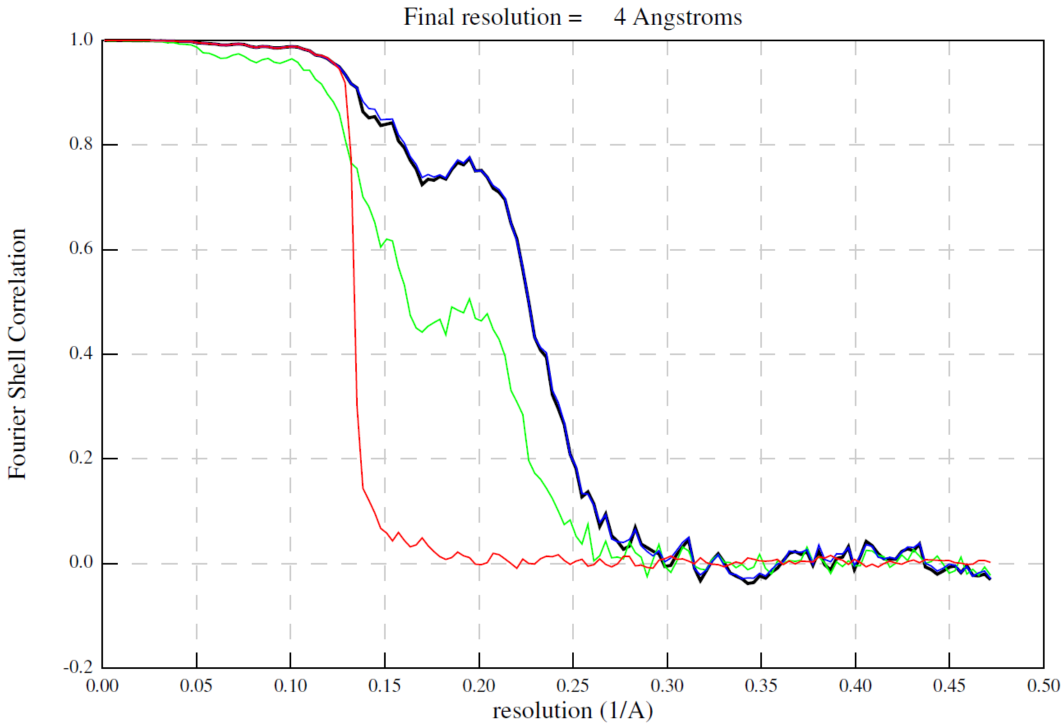
Class	Resolution after refinement
1	5.2 Å
2	4.3 Å
4	4.4 Å
2+4	4.0 Å
1+2+4	3.9 Å

Table 5-2: Final resolutions of maps from individual and combined classes.

The overall resolution of the three classes combined was 3.9 Å, a significant improvement upon any of the three on their own. Classes 2 and 4 were also combined (not shown), and although this led to an improvement in the overall resolution, it was not better than for all three. This suggests there are no significant differences between the structures refined from the classes and it remains beneficial to treat them all together rather than individually. (bottom) shows the overall refined map from the combined classes. This is shown at a lower threshold level, demonstrating that we can now clearly see secondary structure within the equatorial region.

Appendix III - FSCs for final maps





Half map Fourier Shell Correlation curves for wild type (top) and mutant (bottom) maps. Green: Unmasked, blue: masked, red: phase randomised and black: corrected masked maps

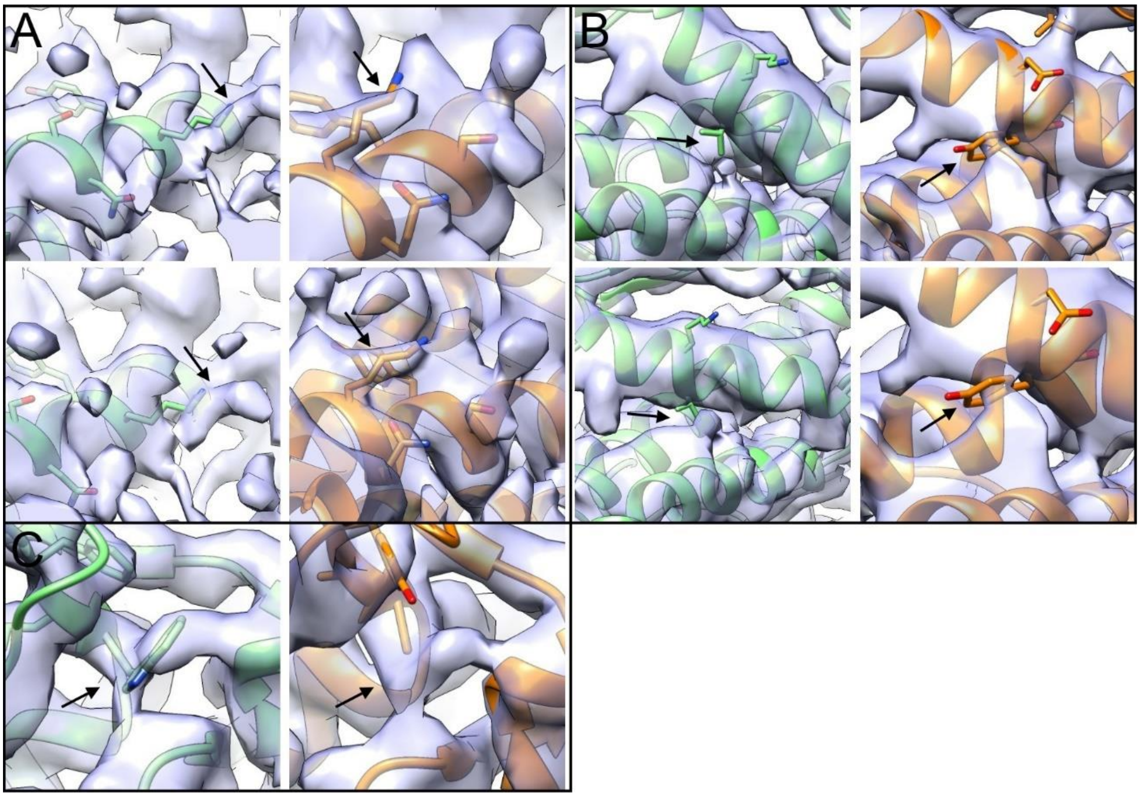
Appendix IV – Confirmation of individual residue density features in other subunits

Since the selection of subunits used for single subunit modelling was arbitrary, the presence (and absence) of features used to distinguish between subunits was checked in other subunits. Some examples are shown for the mutant below.

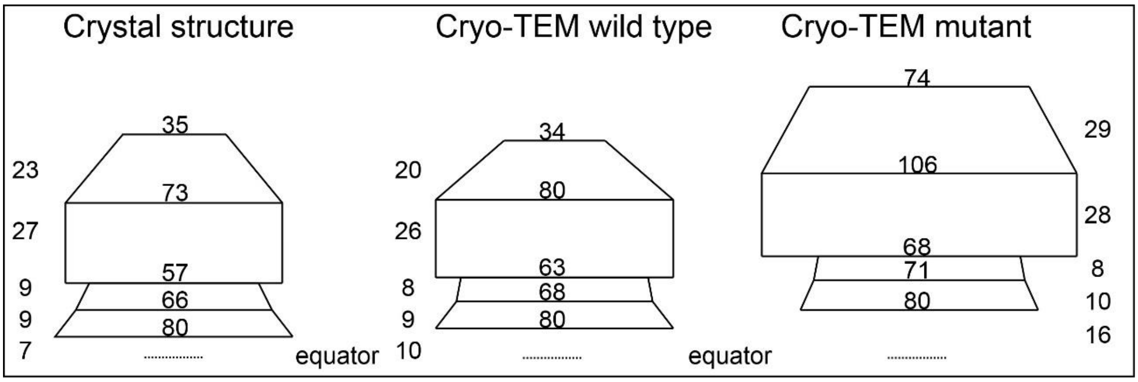
In (A) and (B), the upper images correspond to subunits in the same hemisphere as those originally tested. Lower images (and those in C) correspond to random subunits in the other hemisphere to those originally tested.

Alpha subunits are green while beta subunits are orange. (A) shows residues 6 and 7 as labelled in Figure 3-9. The arrow indicates the lysine residues fitted within the density in each case. (B) shows residue 2, with tyrosine density present in the beta subunit and absent in alpha.

(C) shows the tryptophan which was inconclusive in the original check. Once again it is inconclusive for these subunits in the opposite hemisphere.



Appendix V - Internal measurements

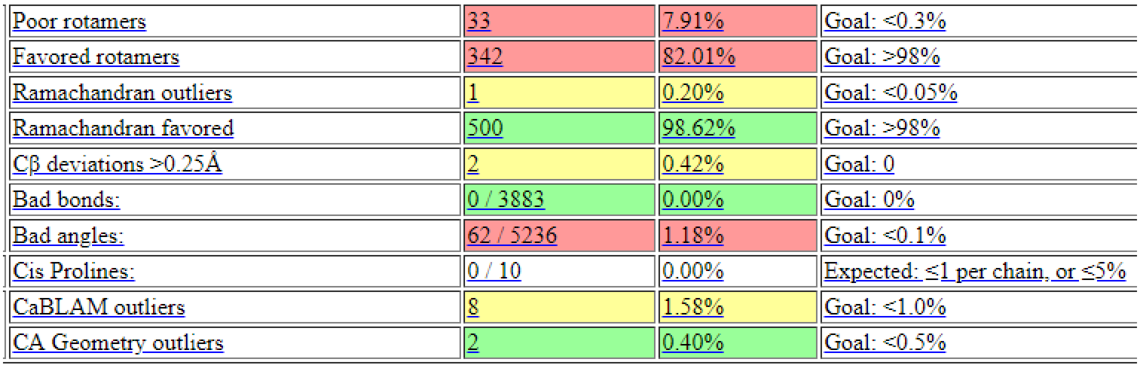


Internal measurements of crystal and cryo TEM models. This is an extension of Figure 3-13(B) to include the crystal structure. All measurements are to the closest Å.

Appendix VI - MolProbity results

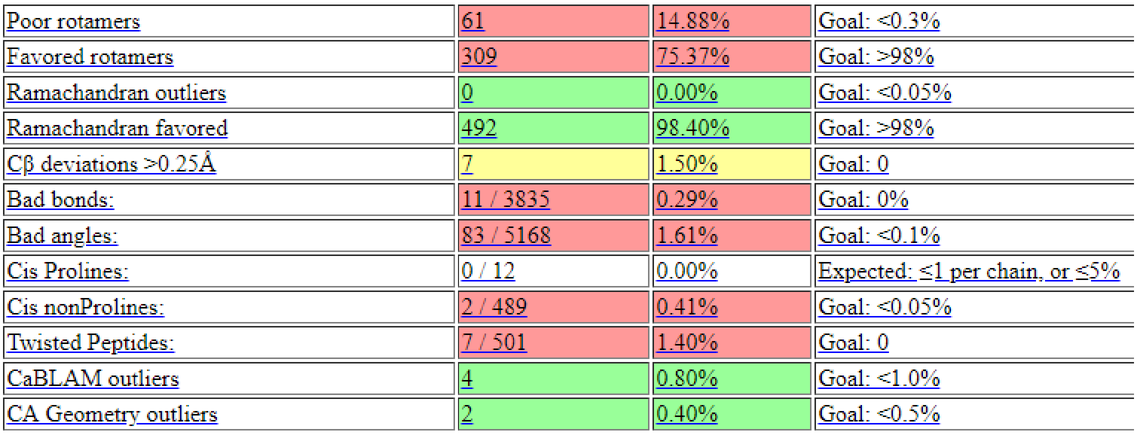
The overall MolProbity statistical results are reported below.

Mutant alpha subunit



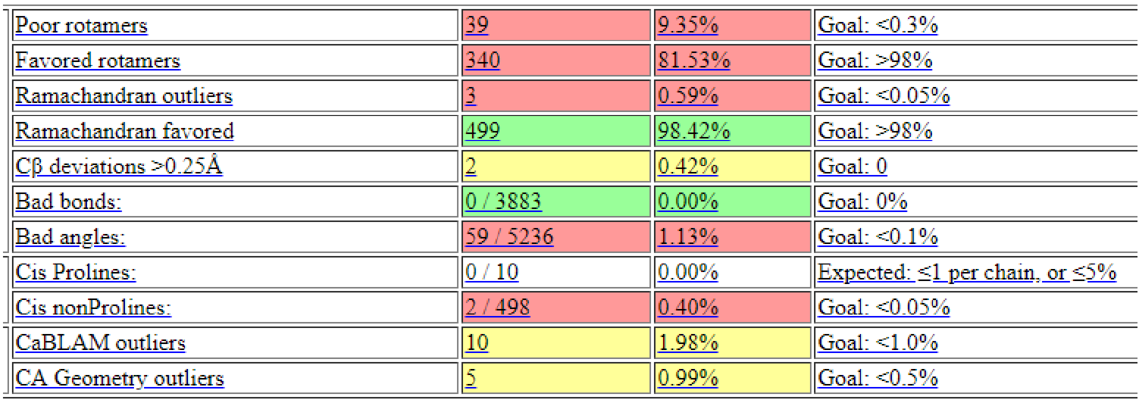
Poor rotamers	33	7.91%	Goal: <0.3%
Favored rotamers	342	82.01%	Goal: >98%
Ramachandran outliers	1	0.20%	Goal: <0.05%
Ramachandran favored	500	98.62%	Goal: >98%
Cβ deviations >0.25Å	2	0.42%	Goal: 0
Bad bonds:	0 / 3883	0.00%	Goal: 0%
Bad angles:	62 / 5236	1.18%	Goal: <0.1%
Cis Prolines:	0 / 10	0.00%	Expected: ≤1 per chain, or ≤5%
CaBLAM outliers	8	1.58%	Goal: <1.0%
CA Geometry outliers	2	0.40%	Goal: <0.5%

Mutant beta subunit



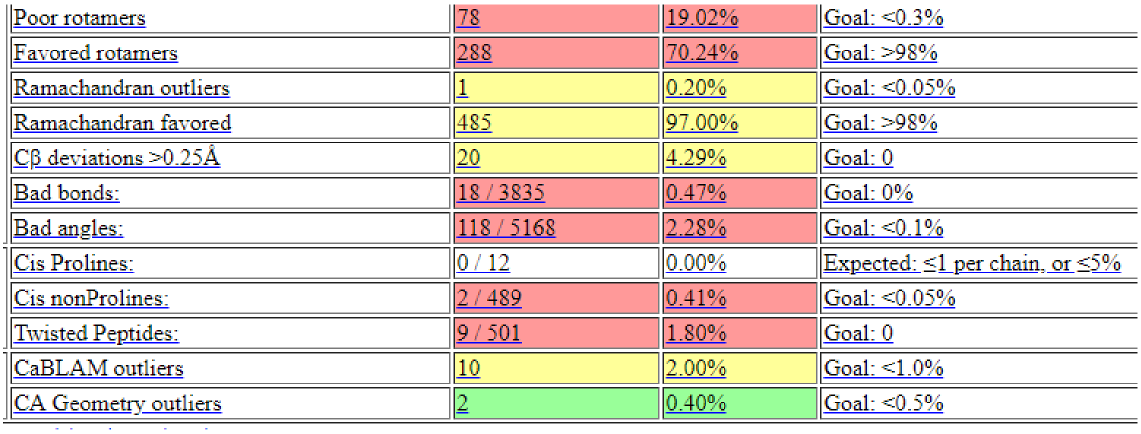
Poor rotamers	61	14.88%	Goal: <0.3%
Favored rotamers	309	75.37%	Goal: >98%
Ramachandran outliers	0	0.00%	Goal: <0.05%
Ramachandran favored	492	98.40%	Goal: >98%
Cβ deviations >0.25Å	7	1.50%	Goal: 0
Bad bonds:	11 / 3835	0.29%	Goal: 0%
Bad angles:	83 / 5168	1.61%	Goal: <0.1%
Cis Prolines:	0 / 12	0.00%	Expected: ≤1 per chain, or ≤5%
Cis nonProlines:	2 / 489	0.41%	Goal: <0.05%
Twisted Peptides:	7 / 501	1.40%	Goal: 0
CaBLAM outliers	4	0.80%	Goal: <1.0%
CA Geometry outliers	2	0.40%	Goal: <0.5%

Wild type alpha subunit



Poor rotamers	39	9.35%	Goal: <0.3%
Favored rotamers	340	81.53%	Goal: >98%
Ramachandran outliers	3	0.59%	Goal: <0.05%
Ramachandran favored	499	98.42%	Goal: >98%
Cβ deviations >0.25Å	2	0.42%	Goal: 0
Bad bonds:	0 / 3883	0.00%	Goal: 0%
Bad angles:	59 / 5236	1.13%	Goal: <0.1%
Cis Prolines:	0 / 10	0.00%	Expected: ≤1 per chain, or ≤5%
Cis nonProlines:	2 / 498	0.40%	Goal: <0.05%
CaBLAM outliers	10	1.98%	Goal: <1.0%
CA Geometry outliers	5	0.99%	Goal: <0.5%

Wild type beta subunits



Poor rotamers	78	19.02%	Goal: <0.3%
Favored rotamers	288	70.24%	Goal: >98%
Ramachandran outliers	1	0.20%	Goal: <0.05%
Ramachandran favored	485	97.00%	Goal: >98%
Cβ deviations >0.25Å	20	4.29%	Goal: 0
Bad bonds:	18 / 3835	0.47%	Goal: 0%
Bad angles:	118 / 5168	2.28%	Goal: <0.1%
Cis Prolines:	0 / 12	0.00%	Expected: ≤1 per chain, or ≤5%
Cis nonProlines:	2 / 489	0.41%	Goal: <0.05%
Twisted Peptides:	9 / 501	1.80%	Goal: 0
CaBLAM outliers	10	2.00%	Goal: <1.0%
CA Geometry outliers	2	0.40%	Goal: <0.5%