EM as a tool to study structure and function to guide inhibitor design

Shaun Deane Rawson

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

Faculty of Biological Sciences, School of Molecular and Cellular Biology

September 2016

Intellectual property and publication statement

The candidate confirms that the work submitted is his own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Components of this thesis contain work from jointly authored publications. Work in Chapter 1 contains figures from two jointly authored review papers published in 2016.

Rawson S., Davies S., Lippiat J.L. & Muench S.P. The changing landscape of membrane protein structural biology through developments in electron microscopy (2016) Mol. Mem. Biol. In press – The candidate devised the layout and wrote the initial draft and was heavily involved in editing and producing all of the figures bar figure 1. Simon Davies was involved in helping with the membrane protein case studies and producing figure 1 with Lippiat and Muench helping to edit and write the manuscript.

Rawson S., Harrison M.A. & Muench S.P. Rotating with the brakes on and other unresolved features of the vacuolar ATPase. (2016) Bioch. Soc. Trans. 44(3) 851-855. The candidate wrote the first draft of the review paper and was heavily involved in all subsequent editing. Both Harrison and Muench provided guidance and support with writing and proof read the manuscript along with minor editorial contributions.

Work in Chapter 3 contains data from three jointly authored publications:

Rawson S., Iadanza M.G., Ranson N.A. & Muench S.P. Methods to account for movement and flexibility in cryo-EM data processing. (2016) Methods. 100 35-41. The candidate carried out all of the flexibility analysis designing the experimental approach and carrying out data processing. The first draft of the manuscript and subsequent editing was carried out by the candidate. Muench provided supervision and guidance where needed. Iadanza and Ranson were responsible for writing the computer scripts, with help from the candidate and contributed to paper writing.

Rawson S., Phillips C., Huss M., Tiburcy F., Wieczorek H., Trinick J., Harrison M.A., Muench S.P. Structure of the vacuolar H+-ATPase rotary motor at subnanometer resolution. (2015) Structure 23(3) 461-471. The candidate made the EM grids, collected and processed the data. The candidate was heavily involved in all data analysis, manuscript writing and figure production. The sample was prepared by Huss M., Tiburcy F., and Wieczorek H. Clair Phillips carried out biochemical analysis including deglycosylation analysis along with Huss M., Harrison, Trinick and Muench provided project supervision and helped write the paper.

Song C.F., Papachristos K., Rawson S., Paci E., Trinick J., Harrison M.A. & Muench S.P. Flexibility within the rotor and stator structures of the vacuolar H+-ATPase: implications for energy conversion efficiencies (2013) PlosOne 8(12) e82207. The candidate was involved in a subsection of the paper and did all of the negative stain work relating to the addition of ATP and its role in flexibility and also the influence of mask size on the resulting data. All of the yeast data within the paper was collected and processed by a postdoc Chun-Feng Song. Papachristos and Paci were involved in running MD simulations. Trinick, Harrison and Muench designed the experimental approach and wrote a significant part of the manuscript. The candidate was involved in writing his section for the paper but not the majority as this had been written before he started.

Work in Chapter 4 contains data from a jointly authored publication:

Muench S.P., Rawson S., Eyraud V., Delmas A., Da Silva P., Phillips C., Trinick J., Harrison M.A., Gressent F., & Huss M. PA1b inhibitor binding to subunits c and e of the Vacuolar ATPase reveals its insecticidal mechanism (2014). J Biol Chem 289(23),16399-16408. The candidate collected the negative stain data, carried out data processing and carried out data analysis on the structural work along with writing that section of the paper. Eyraud, Delmas, Gressent and Da Silva produced PA1b, conducted biochemical analysis and wrote these sections of the paper. Markus Huss conducted radio-ligand binding experiments. Muench and Harrsion conducted data analysis and drafted the manuscript.

Acknowledgements

First and foremost, I would like to thank my supervisors Stephen Muench, Robin Bon and John Trinick for their excellent guidance and support through the years.

I am extremely grateful to all members of the Muench and Ranson labs for their help and assistance. In particular Matt Iadanza for scripting support, Dan Hurdiss for photoshop wizardry, Emma Hesketh for general nagging and substitute mothering, and Rebecca Thompson for the spoon.

I am grateful to the MRC-LMB for providing access to their Titan Krios, in particular Shaoxia Chen and Christos Savva for their assistance with data collection. In addition I am grateful to CAS, Beijing for access to their Titan Krios. I also need to thank; Frederic Gressent, Markus Huss, Claudine Bisson, Clair Phillips, Mike Harrison and all the others who have provided me with samples and biochemical evidence to back up my pretty picture taking.

Last but not least, my parents and Lindsey, without whom this work would not have been possible.

Abstract

Rational drug design is underpinned by structural biology, in particular X-ray crystallography. This work aimed to explore a simple question; can electron microscopy (EM) play a useful role in the drug discovery process? To this end, two biological systems, the V-ATPase and Imidazoleglycerol-phosphate dehydratase (IGPD), were studied via EM.

A modest (~1 nm) structure of the V-ATPase was solved via cryo-EM, permitting existing high resolution crystal structures to be accurately fitted in the context of the whole 1 MDa complex. This allowed new mechanistic insights to be uncovered. The dissociated V_1 domain was studied to see if the mechanism of ATP silencing, used as a regulatory control, could provide new inhibitor targets to be identified. Although the specific mechanism could not be identified the resulting structures have shown a much more complicated regulatory mechanism than previously thought. The ubiquitous nature of the V-ATPase makes selectivity a significant challenge. To address this the binding mode of the selective V-ATPase inhibitor, Pea Albumin 1 subunit b (PA1b), was studied at low resolution using negative stain EM. Combined with biochemical analysis this has successfully identified the subunits responsible for binding PA1b. Furthermore, IGPD was used as a model system for high resolution cryo-EM studies to directly visualise inhibitor binding. A 3.1 Å reconstruction was obtained which allowed the *de novo* building of the atomic model and the identification of a small molecule inhibitor within the EM map.

This work shows the potential of EM to provide valuable information for drug design at a wide range of resolutions, from rapid low resolution binding studies to direct visualisation of inhibitors at ~3 Å.

Table of Contents

List of Figures

List of Tables

Abbreviations

Adenosine triphosphate	ATP
Basic Local Alignment Search Tool	BLAST
Charge coupled device	CCD
Contrast transfer function	CTF
Detector quantum efficiency	DQE
Direct electron detector	DED
Electron microscopy	EM
Electron microscopy data bank	EMDB
Enterococcus hirae	*E. hirae*
Fast Fourier transform	FFT
Fourier shell correlation	FSC
Graphics processing unit	GPU
High throughput screening	HTS
Histidinol-phosphatase	HP
Horseradish peroxidase	HRP
Imidazoleglycerol-phosphate dehydratase	IGPD
Liquid nitrogen	LN2
Manduca sexta	*M. sexta*

Maximum likelihood	ML
Monolithic active pixel sensor	MAPS
Nuclear magnetic resonance	NMR
Pea Albumin 1 subunit b	PA1b
Protein data bank	PDB
Signal to noise ratio	SNR
Single particle analysis	SPA
Styrene maleic acid lipid particle	SMALP
Transmission electron microscopy	TEM
Vacuolar ATPase	V-ATPase

1. Introduction

"Almost all aspects of life are engineered at the molecular level, and without understanding molecules we can only have a very sketchy understanding of life itself."

Francis Crick - **What Mad Pursuit: A Personal View of Scientific Discovery** (1988)

"Does structure determine function, or does function determine structure, or does each determine the other, and if the latter, how can it be?"

E. Stanley Abbot - **The Causal Relations between Structure and Function in Biology** (1916)

1.1 Introduction to Structural Biology and Electron Microscopy

It has long been acknowledged that the understanding of structure is crucial to our understanding of biology. In the 1920's the field of structural biology as we know and understand it was born, with the work of Astbury and Bernal on X-ray fibre diffraction. A further milestone occurred in 1958, when building upon the pioneering X-ray studies of the Braggs in 1913 on crystalline salts, Kendrew showed the three dimensional structure of a protein for the first time (Kendrew et al., 1958), using myoglobin as his example. This was rapidly followed in 1960 by Perutz's crystal structure of haemoglobin (Perutz et al., 1960). Following these first protein crystal structures, structural biology has gone from strength to strength. Now detailed structural understanding underpins many advances in biological understanding and informs the design of new therapeutic agents. In the years that followed Perutz's and Kendrew's landmark work, X-ray crystallography has remained the most widely used technique for the study of protein structure; indeed the protein databank (PDB) contains more than 100000 protein X-ray crystal structures along with an additional ~10000 structures from other techniques including NMR.

An alternate technique for studying protein structure is electron microscopy (EM). Rather than studying the diffraction patterns formed from a highly ordered crystal lattice, EM is typically used to image individual particles of the specimen, removing the often difficult requirement for a well behaved crystal. While traditional light microscopy is limited to resolutions of ~200 nm by the wavelength of visible light, the wavelength of an electron accelerated to 100 kV is in the picometer range, allowing theoretical resolutions of smaller than an atom. The biological applications of EM were first explored in the 1960's and 70's with several papers showing stained biological specimens and crystalline proteins (Brenner and Horne, 1959; De Rosier and Klug, 1968; Unwin and Henderson, 1975). However, it was not until the work of Dubochet in the 1980's that EM began to show its true potential for structural studies of individual proteins. Dubochet's innovation was to image the protein in a thin layer of vitreous ice, rather than dried and embedded in a heavy metal stain (Adrian et al., 1984; Dobro et al., 2010; Dubochet et al., 1988; McDowall et al., 1983). This allowed finer details to be observed and avoided the harsh conditions previously required, which could introduce significant artefacts into the structure of the protein. From this point onwards improvements in technology have taken EM from producing low resolution envelopes or 'blobology' to obtaining near atomic resolution for a variety of biological macromolecules from molecular machines to viruses (Smith and Rubinstein, 2014).

1.2 Potential for structure based inhibitor design by EM

While, due to resolution restraints, the idea of EM as a tool for direct visualisation of inhibitor binding would have been far-fetched at best at the start of the project, a wealth of useful information is still available at the more modest resolutions (~5-10 Å) previously achievable. Given the developments in EM, leading to side chain assignment becoming achievable, what role could it play in the process of therapeutic discovery, in particular informing structure based inhibitor design?

1.2.1 Current approaches to drug discovery

The world is in constant need of new therapeutics, from the ongoing war against bacterial resistance to new therapies for diseases ranging from cancer to psychological conditions. However, the process of drug discovery is extremely slow and fraught with many hurdles and difficulties. One common approach to drug discovery consists of target-based drug design where a molecular target, for example an enzyme, is identified and modulated to hopefully produce the desired physiological effect. This can be achieved through screening techniques, including natural product libraries, fragment libraries and high throughput screening (HTS) (Erlanson et al., 2016). In HTS an extremely large library of compounds is screened against a biological target of interest and any hits are then taken forward through development. Alternately rational drug design can be used, where knowledge of the structure of the molecular target is used to design new molecules which are hoped to exhibit binding/inhibitory properties (Lounnas et al., 2013). An alternative to target-based drug discovery is phenotypic screening, where instead of aiming at a molecular target a phenotypic response is looked for in a model system of the disease (Moffat et al., 2014).

While HTS has proved to be valuable in several different inhibitor development programs it does have drawbacks (Macarron et al., 2011). Firstly in order to perform a high throughput screen a large and drug like chemical library (~100,000-1,000,000's) is needed, and while many large pharmaceutical companies possess in-house libraries ranging into the millions of compounds, this can be a hurdle for smaller companies or academic researchers. Indeed the costs associated with obtaining, curating and maintaining a compound library makes this process extremely difficult for many organisations. Secondly, within these large libraries only a small fraction of chemical space is covered, which may hinder discovery of inhibitors for targets with unusual binding sites. Furthermore, HTS is dependent on the ability to assay the target protein(s) in a high throughput manner while carefully tuning the assay conditions to ensure the best compound classes are identified as rapidly as possible. Additionally, most HTS campaigns will focus on identifying a molecule which either binds or inhibits the molecular target, but inhibition of a single protein does not necessarily lead to a phenotypic effect.

While phenotypic screening has the advantage that no molecular target needs to be identified in order to find a therapeutic possessing the desired effect, it is less commonly used relative to other methods. This is because it relies on possessing an extremely good model for the disease in question which can then be interrogated in a similar manner to HTS. However, only a small number of diseases have well researched, valid models and often the use of these is prohibitively expensive compared to an isolated enzyme for example. A further limitation is the difficulty in identifying the 'targets' which are being modulated by any molecules discovered in this way. While this knowledge is not required for a inhibitor to become a successful therapeutic, the lack of known target can hinder further optimisation efforts (Dominguez et al., 2014).

1.2.2 Rational Drug Design

In theory rational drug design overcomes some of the issues encountered by other approaches as a large compound library is not needed, nor is a detailed, expensive model system. By using prior knowledge of the target protein, small molecules can be designed which theoretically maximise binding potential (Clark, 2006). Although it must be noted that this assumes a much greater understanding of what happens upon ligand binding, and our ability to translate this to chemistry, than is perhaps true (Davis and Teague, 1999; Davis et al., 2008). Traditionally rational drug design has been underpinned by structural information primarily obtained from X-ray crystallography and NMR studies. While X-ray crystallography has provided a wealth of structural information which has been used in rational design pipelines it does suffer from several limitations, including difficulty in obtaining crystals of the protein of interest (Niedzialkowska et al., 2016), crystal packing artefacts in the structure (Davis et al., 2008; Steuber et al., 2006; Søndergaard et al., 2009), issues with ligand geometry (Kleywegt et al., 2003; Liebeschuetz et al., 2012; Reynolds, 2014), difficulties with low occupancy inhibitors, and showing only a static snapshot of the protein/inhibitor complex with restrictions in conformational changes that can be observed.

The requirement for high resolution structures is a major bottleneck in rational drug design. For example, despite the PDB containing over 100,000 X-ray crystal structures, only 2821 are membrane proteins, clearly illustrating the difficulty in expression, purification and obtaining crystals of this important class of proteins which make up to 60% of drug targets (Overington et al., 2006). Although great efforts have been made to improve this, with some success in the use of crystallization techniques such as lipidic cubic phase it still proves a major challenge to reliably crystallize most membrane proteins (Caffrey, 2015). This lack of structural information in turn then prevents rational drug design being performed on the vast majority of membrane proteins and thus hinders the development of potential therapeutics. This is not just the case for membrane proteins; other potential drug targets such as viruses and rotary motors have proven difficult to crystallize for a variety of reasons including flexibility and size. When a structure is obtained from crystallography there is also the possibility that some of the observed interactions are artificial and are a result of crystal packing artefacts. Crucially, a major drawback in using crystal structures for rational drug design is the rigid and static nature of the structure obtained (Teague, 2003). While inhibitor-bound crystal structures give the impression of a solid static protein binding to an equally static inhibitor the truth is that both are constantly moving and the protein behaves more like a firm gel than a rigid solid. As crystallography relies on each copy of the protein being identical in order to diffract in the lattice and only the most rigid/thermodynamically stable states of a protein typically crystallize, so any dynamic movements are lost.

1.2.3 Advantages of EM

Electron microscopy overcomes many of the hurdles and limitations experienced by crystallography. Instead of requiring large amounts of protein (mg scale) and a large amount of luck to crystallize one's protein, EM only requires small amounts (µg scale) and no need for crystals (other than for 2D crystallography). This widens the number of targets that can be studied, from large macromolecules such as viruses to membrane proteins where the presence of detergent makes crystallization a challenge. EM also offers potential advantages for drug discovery by being able to trap different conformational states of a protein, moving away from static crystal structures to a dynamic ensemble of different states (Dashti et al., 2014; Frank and Ourmazd, 2016). This could prove to be extremely powerful as inhibitor binding pockets can breathe and change as the protein samples different conformations. Additionally, while it may be possible to artificially trap a protein in other states in a crystal through the use of additives, it would be impossible to just add natural substrate and capture the protein in all stages of what may be a continuous process. EM has also been used to study proteins in more native conditions; for example TRPV1 with several bound ligands has been studied in the presence of native lipid through the use of nanodisks (Gao et al., 2016). Indeed the native lipid is thought to be crucial to the mechanism and to inhibition.

However, in the past EM has not been a viable technique for structure-based drug design, primarily due to the low resolutions that had previously been obtainable, leading to EM being known as 'blobology'. While in the past EM has been limited to around nanometer resolution at best for most samples, recent technological advances have now enabled near atomic resolutions (~3 Å) to be obtainable (Figure 1.1). Improvements in microscope stability and particularly the development of direct electron detectors (DED) have formed the basis for the renaissance of EM as a mainstream structural technique (Glaeser, 2016; Nogales, 2016; Smith and Rubinstein, 2014).

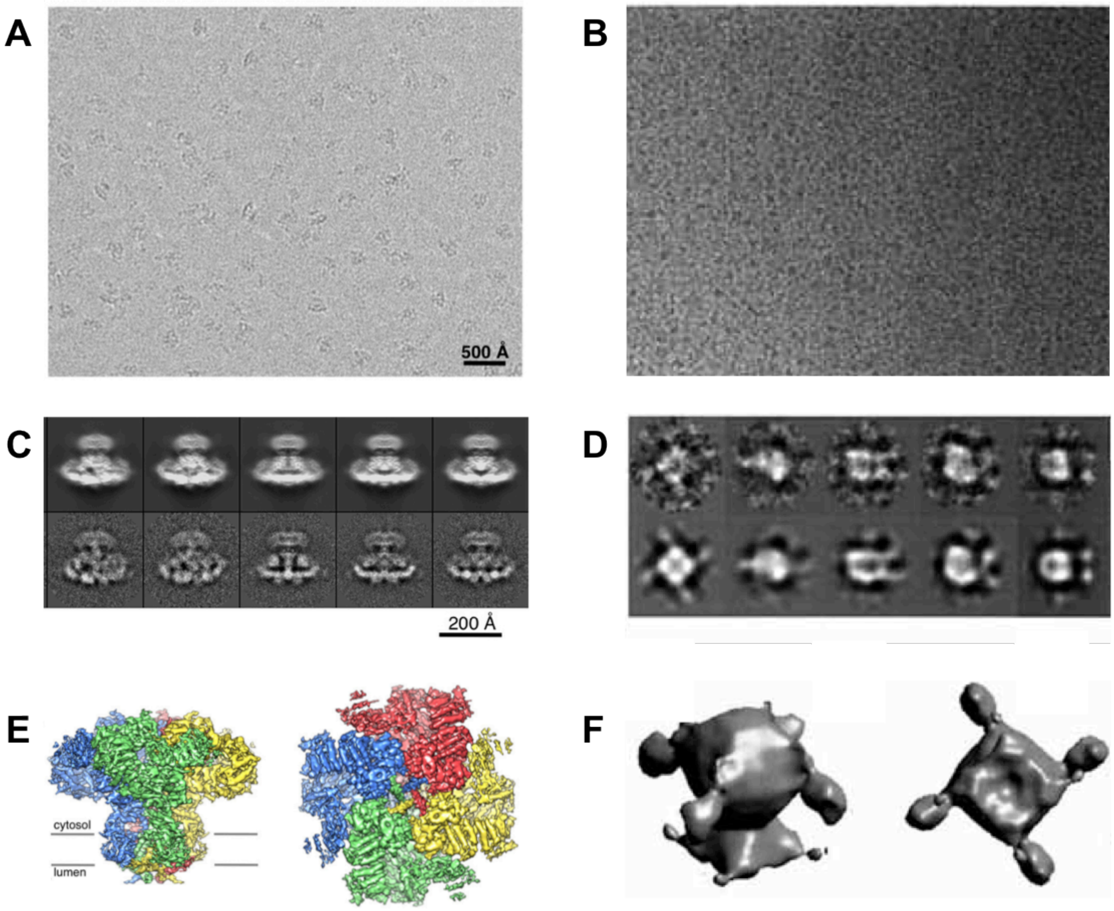


Figure 1.1 – Comparison of IP3R EM structures pre/post DED's. (A/B) Cryo micrograph of IP3R on DED (A) and CCD (B). (C/D) Projections and class averages on DED (C) and CCD (D). (E/F) 3D reconstructions on DED (E) and CCD (F). Images adapted from (Fan et al., 2015) and (Jiang et al., 2002).

Direct electron detectors (DEDs) have not only increased the sensitivity and contrast of the images but also allow capture of high frame rate movies. These movies, in conjunction with new processing algorithms allow the motion of the specimen during the exposure due to both mechanical and beam induced movements to be partially accounted for, reducing the blurring of the image and allowing higher resolution information to be recovered (Campbell et al., 2012; Li et al., 2013a). At the same time it also mitigates, to some extent, one of the largest remaining challenges in EM, radiation damage. With these improvements allowing higher resolution structures to be obtained there have already been a number of structures published in complex with ligands and inhibitors, from large soluble complexes like the proteasome and β-galactosidase to smaller membrane proteins including the TRP channels and γ-secretase (Figure 1.2) (Bai et al., 2015b; Bartesaghi et al., 2015; da Fonseca and Morris, 2015; Gao et al., 2016; Li et al., 2016).

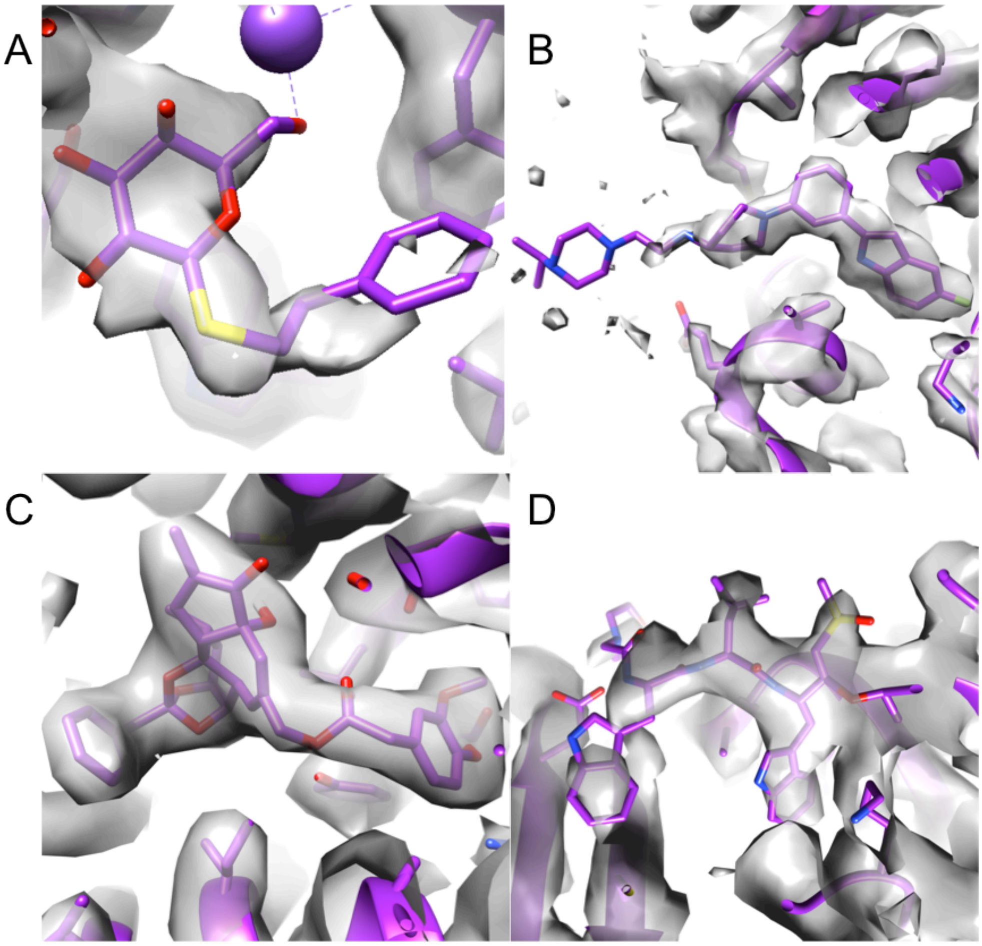


Figure 1.2 – Example EM reconstructions showing bound ligands. (A) β-galactosidase at 2.2 Å resolution showing bound PETG (Bartesaghi et al., 2015)(EMDB 2984, PDB 5A1A). (B) Human p97 at 2.3 Å with bound UPCDC30245 inhibitor (Merk et al., 2016)(EMDB 3295, PDB 5FTJ). (C) TRPV1 at ~3 Å showing bound resiniferatoxin (Gao et al., 2016)(EMDB 8117, PDB 5IRX). (D) Human proteasome at 3.6 Å with bound WLW-vs inhibitor (Li et al., 2016)(EMDB 3231, PDB 5FMG).

1.2.4 *Limitations and developments in EM*

Although EM offers many exciting new possibilities to play a role in, and enhance, structure-based drug design, there are still several significant challenges that may hinder its widespread use. Foremost amongst these obstacles is the limited resolution which is currently obtainable. Despite recent advances (Merk et al., 2016), EM still lags behind X-ray crystallography in that obtaining ~1 Å resolution for biological specimens is not currently possible. This is a severe limitation as the exact details of binding mode and precise interactions between side chains and the ligand are only observable at these high resolutions. It is well established that EM can obtain atomic resolution structures in the materials field and is routinely used for this (Krivanek et al., 2010; Spence, 1999; Urban, 2009). As such the main factors hindering obtaining high resolution in biological EM are twofold: the radiation damage sustained by the sample, and the movement of the specimen itself in the electron beam (Glaeser, 2016).

While X-ray crystallography studies are also affected by radiation damage the problem is significantly worse for EM. Indeed it has been compared to a nuclear explosion at the specimen scale (Glaeser and Taylor, 1978; Orlova and Saibil, 2011). As direct detectors allow the capture of movie frames this allows the use of all the data including later frames from the exposure to be used to align the images giving higher contrast and easier alignment and then to use a subset of these frames with lower dose to perform the reconstructions, reducing some of the effects of radiation damage. However, it has been estimated that significant loss of high resolution information occurs after doses of ~3 $e^{-}/Å^{2}$ (Baker et al., 2010; Grant and Grigorieff, 2015), and much of this dose occurs within the first few frames of the exposure. Moreover, the frames that would contain this high resolution information suffer from the worst effects of beam induced motion, whereby as the grid is first exposed to the electron beam, large movements occur through poorly understood combinations of charging and expansion/contraction of the vitreous ice and the support film, typically made of amorphous carbon (Brilot et al., 2012; Glaeser et al., 2011). This means that with current methodology they cannot be used, making extremely high resolution (~ 1 Å) structures unlikely in the near future, although research into new support substrates is ongoing (Russo and Passmore, 2014; 2016; Yoshioka et al., 2010). Radiation damage can also have a more subtle effect aside from the degradation of resolution, where damage causes protein side chains to shift. Previous studies have shown that negative side chains appear to suffer more from radiation damage (Allegretti et al., 2014; Grant and Grigorieff, 2015), which is potentially due to decarboxylation of these side chains (Weik et al., 2000), so if there are key negative residues in the binding pocket then this may induce a significant movement of the inhibitor or show a binding mode that is not physiologically relevant.

EM can tackle proteins that are intractable for crystal studies such as large complexes and membrane proteins. However, current limitations in technology impose a size limit on the proteins that can be observed. Currently proteins smaller than ~250 kDa prove a major challenge for EM, though there are exceptions to this, notably work carried by Subramanian and coworkers on isocitrate dehydrogenase and other small targets (Merk et al., 2016). While there are developments underway to improve this, and already significant progress through the use of direct detectors, this size limitation still hinders the use of EM for several important classes of drug targets including GPCRs. As such this is the subject of much ongoing research, particularly the development of phase plates. By inducing a quarter wave phase shift into the electron beam, the contrast close to – or at – focus, is greatly enhanced. While several designs of phase plate have been investigated in the past (Danev and Nagayama, 2001), the current technology is known as the Volta phase plate. Volta phase plates consist of a carbon film, heated to around 200 °C, placed in the path of the beam, though how this induces phase shift is not yet fully understood (Danev and Baumeister, 2016; Danev et al., 2014). This technology has already allowed the structure of the human Prx3 (~250 kDa) to be obtained at 4.4 Å resolution (Khoshouei et al., 2016). Both phase plates and direct detectors are still in their infancy; with further development and optimisation this technology could enable EM to look at a much wider range of proteins than are currently viable.

Another method in development to increase the range of viable proteins for study via EM is through the use of affinity grids (Yu et al., 2016). Structural study of several potentially therapeutically interesting proteins is precluded simply by difficulties in obtaining sufficient pure protein to carry out even EM studies, much less the orders of magnitude more required for crystal studies. This can be due to challenges in the overexpression including posttranslational modifications, toxicity or mislocalisation. To combat this, rather than performing traditional purification and studying the resulting protein via EM, work is ongoing to purify the sample directly onto the EM grid. Various methods are being developed to achieve this, including immobilising antibodies to the grid surface to tether the protein of interest to the grid (Benjamin et al., 2016) without any need for genetic modification. Alternatively Ni doped lipid monolayers can be applied to the grid surface to extract a His tagged protein directly. While currently the only examples of this work at high resolution have been highly symmetrical virus structures (Yu et al., 2016), with further development it could become a powerful technique for the structural study of hard-to-obtain proteins.

A further drawback of EM for drug design purposes is the speed at which structures can be obtained. While with crystallography it is routine to set up several co-crystallizations or soaks simultaneously and the data collection for each crystal obtained is very rapid (minutes) (Blundell et al., 2002; O'Reilly et al., 2007), EM data collection is still relatively slow (Thompson et al., 2016), typically taking several days of microscope time for a single sample. While automated data collection has improved the situation (Carragher et al., 2000; Suloway et al., 2005; Tan et al., 2016), by increasing throughput from hundreds of micrographs collected manually a day to thousands, this is still a slow procedure compared to data collection on a crystal at a beamline.

For structure based drug design this is a particular issue, as a large number of analogue structures will typically be needed to build the structure activity relationship required for structure based drug design (Blundell et al., 2002). Nonetheless, new technologies in the field have the potential to partially mitigate this. For example, work is ongoing to improve the reproducibility of grid preparation, reducing the time spent optimising freezing conditions, thereby reducing microscope time for screening. One example of this is Spotiton, an automated grid preparation system, which rather than relying on using a relatively large (µl) droplet of sample and blotting most away, instead sprays a small (nl) volume onto a small section of the grid (Jain et al., 2012; Razinkov et al., 2016). Then this small volume is rapidly wicked away to a thin film, with tuneable thickness. This was made possible by the development of new grid designs, containing copper nanowires, which wick the liquid from the holes. As the sample only occupies a small portion of the grid 'real estate' it is easy to envisage applying multiple samples, for example a protein mixed with various inhibitors, onto different areas of the same grid. Although the data collection for each sample will remain slow, this would allow multiple data collections to occur without need for repeatedly changing grids and setting up automated data collection separately for each.

Data processing is also typically significantly slower than crystallography, taking between weeks to months in order to extract the most from the raw data, possibly precluding EM as a high throughput structural technique. Tied to this is the volume of data produced in a single EM experiment, with a single micrograph being between 1-10 GB in size, leading to datasets of 10+ TB for a single specimen. This presents a major infrastructure challenge and without significant resources dedicated to processing and data storage, makes the initial barrier to entry for EM substantial (Cianfrocco and Leschziner, 2015; Patwardhan et al., 2012; Thompson et al., 2016). Nevertheless, improvements are being made in speed of data acquisition and enhancing the speed of image processing. Computational power is constantly increasing, meaning the bottlenecks that currently hinder rapid processing of EM data should reduce over time. At the same time graphical acceleration of the processing pipeline using GPU technology is an active area of research with several key steps already optimised in this manner (Zhang, 2016), cutting processing times for these steps from days to hours. Further work is ongoing in utilising this dramatic acceleration to speed up the currently slow steps of classification and refinement. With the rapid rate of change in computational power and developments such as GPU acceleration, it is easy to envisage a processing pipeline for EM where several processing steps are carried out 'on the fly' as data is collected in an analogous manner to X-ray data collection (Lander et al., 2009).

Advances in technology and computation may also allow the use of EM to study biological processes in a time-dependent manner, giving insight into the dynamic changes that occur within the system. Work is ongoing to develop time-resolved EM through a variety of methods, including rapid mixing/spraying in order to physically trap the protein of interest in a variety of functional states (Chen and Frank, 2016; Chen et al., 2015; Lu et al., 2009; M Walker, 1995), as well as computational manifold mapping approaches which have already been utilised to identify several functional conformational states of the ribosome (Dashti et al., 2014; Frank and Ourmazd, 2016).

1.3 Membrane protein structure via EM

It is estimated that 60% of drug targets are membrane proteins, including GPCRs, ABC transporters, ion channels and rotary ATPases, across a wide range of therapeutic areas from antibiotics to cancer (Overington et al., 2006). However, only ~3% of protein structures in the PDB are membrane proteins. This paucity of structural information hinders rational drug design efforts and slows the understanding of these vital systems.

Until recently, high resolution structural information on membrane proteins has been obtained primarily from X-ray crystallographic data. However, the requirement for highly ordered crystals has made many membrane proteins and protein complexes intractable to crystallographic approaches. Until the recent developments in the field of cryo-EM, high resolution membrane protein structures via EM were only obtainable through electron diffraction of 2D crystals, with the highest resolution single particle EM structure pre-2009 being the RyR1 calcium-release channel at a modest 9.6 Å resolution (Ludtke et al., 2005). This was not only a feature of the EM hardware but also the relatively small size of many membrane proteins of interest and their propensity to aggregate, which precluded many high resolution structures from being obtained. Thus, historically only very large membrane proteins such as the rotary ATPase family were routinely studied via EM as their size and flexibility meant crystallographic studies were very demanding (Ludtke et al., 2005; Muench et al., 2011).

With the modern developments in detectors and microscopes, the use of EM for the study of membrane proteins has been rapidly increasing. Since the widespread adoption of direct detectors, several membrane proteins have had high resolution structures solved by EM including TRPV1, γ-secretase, RyR1, Cav1.1, NPC1 and Piezo (Figure 1.3) (Bai et al., 2015c; Gao et al., 2016; Ge et al., 2015; Gong et al., 2016; Wu et al., 2015; Yan et al., 2015). Indeed the quality of the EM maps is now sufficient in some cases to permit the identification of a bound ligand or inhibitor as with TRPA1, TRPV1 and γ-secretase (Bai et al., 2015b; Paulsen et al., 2015), perhaps opening the way for rational drug design informed via EM structures, an area previously reserved for crystallographic or NMR studies.

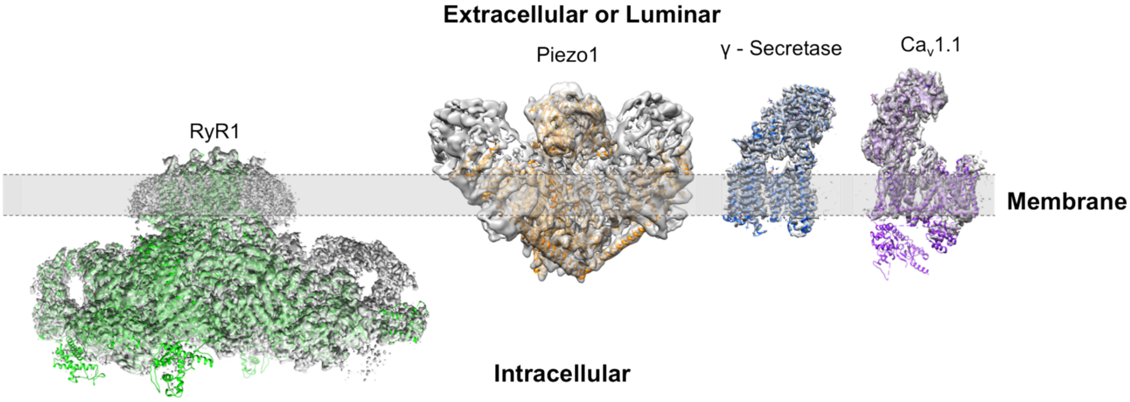


Figure 1.3 - Examples of sub nm membrane protein structures determined by cryo-EM. Examples showing EM reconstructions (grey density) of RyR1 (EMDB-2807), Piezo1 (EMDB-6343), γ–secretase (EMDB-3061) and Cav1.1 (EMDB-6475) with fitted atomic models. Figure adapted from (Rawson et al., 2016a).

In 2006 there were 8 membrane protein structures deposited in the EMDB with an average resolution of ~22 Å (Figure 1.4). In contrast, in 2015 the number of deposited structures increased to 52 and the average resolution of these has improved to 12 Å (Figure 1.4). While not all published membrane proteins have been deposited within the EMDB, this does provide an approximate representation of a general trend. At the same time the average molecular weight of the membrane proteins released has remained relatively unchanged, perhaps surprising given the technological advances, particularly the use of direct electron detectors. If the current trend in resolution was to continue, it is tempting to speculate that EM could become the structural technique of choice for membrane proteins in the future. However, it is worth noting that between

2013-2016 the highest resolution structure deposited has remained fairly stable, suggesting a plateau in resolution for membrane proteins at ~3.5 Å.

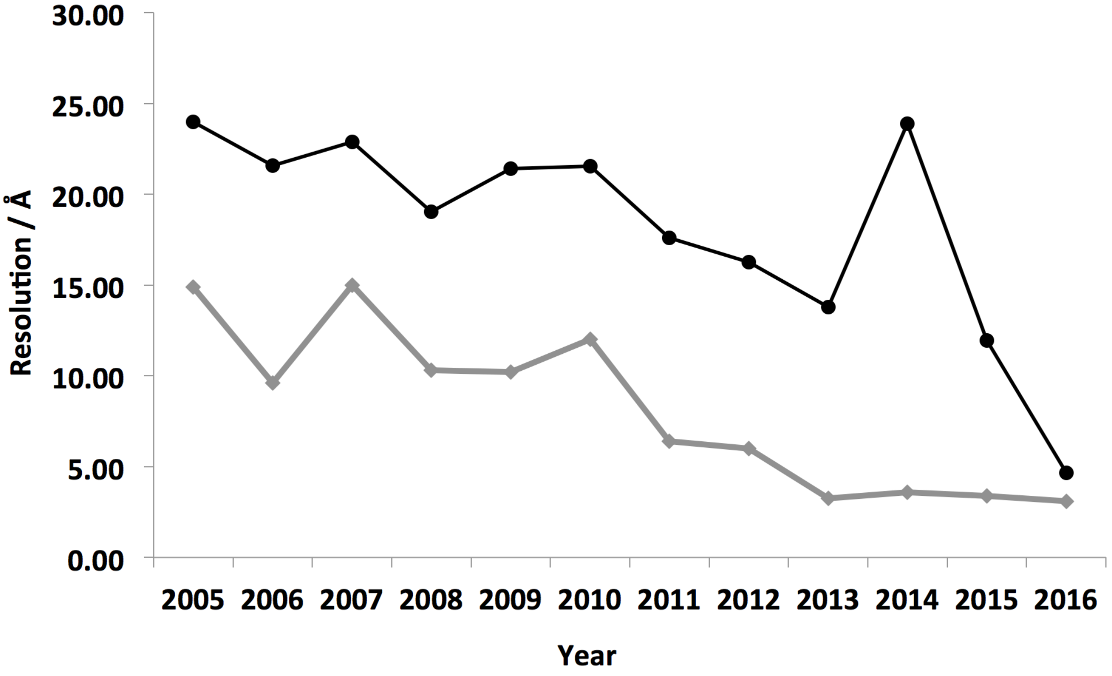


Figure 1.4 - Mean resolution vs year of deposition within the EMDB. Analysis of the average resolution of deposited membrane protein structures within the EMDB in each year is shown in black. Highest resolution reconstruction deposited each year is show in grey. The general trend shows the resolution steadily increasing over time. Figure taken from (Rawson et al., 2016a).

While advances in hardware and software within EM have enabled it to become a viable and valuable technique for structural studies of membrane proteins several challenges still remain. There will always be some of the same struggles faced by crystallography in obtaining and purifying sufficient amounts of protein and ensuring the protein's stability, although progress is being made in the biochemistry to aid this (Abdul-Hussein et al., 2013; Bill et al., 2011; Frauenfeld et al., 2016; Tate, 2010). A further complication is the need to solubilise the protein in detergent; several EM studies have already shown that the choice of detergent plays a large role in the final resolution obtainable and indeed on whether the project is viable at all (Gao et al., 2016; Hauer et al., 2015; Lu et al., 2014). Indeed the function of several proteins is known to undergo large changes upon purification within a detergent (Postis et al., 2015), and other studies have shown the vital role played by phospholipids in stability, mechanism and inhibitor binding (Gao et al., 2016). This casts some doubt on the physiological relevance of many of the membrane protein structures currently available from both EM and crystal studies.

While great progress has been made in producing and finding milder and more relevant detergents, they are still significantly different from the native membrane environment. Several groups have been developing methodology to combat this issue; of particular interest are the efforts to introduce a more native environment through the use of styrene maleic acid lipid particles (SMALPs) (Jamshad et al., 2011; Lee et al., 2016; Postis et al., 2015) and the progress being made in imaging proteins *in situ*, within a vesicle (Wang and Sigworth, 2009; 2010). These *in situ* studies would not only allow protein activity to be measured in the imaged sample, but also potentially allow different conformations of the protein to be observed by inducing a membrane potential or a chemical gradient through the addition of additives to the sample. While this is an active area of research, it is likely that there will be no ‘one size fits all’ general solution, and that the method of solubilisation will have to be optimised for each membrane protein in turn (Dörr et al., 2016; Kleinschmidt and Popot, 2014).

1.4 V-ATPase Introduction

A group of particularly interesting membrane proteins for therapeutic design is the rotary ATPase family, which are membrane bound molecular motors that act either by hydrolysing ATP to drive a proton gradient or utilise an existing proton gradient to carry out ATP synthesis. They consist of the ATP synthase (F-ATPase), vacuolar ATPase (V-ATPase) and archaeal ATPase (A-ATPase) (Figure 1.5). An inhibitor targeting the F-ATPase (Bedaquiline) has already been approved as a treatment for tuberculosis, the first new approved treatment for this disease in 40 years (Mahajan, 2013). The V-ATPase plays an essential role in several physiological processes (Marshansky et al., 2014), and therefore is found in all eukaryotic cells (Forgac, 2007). It is thus unsurprising that the complex is implicated in a variety of disease states (Hinton et al., 2009). For example its role in bone resorption means that it is involved in osteopetrosis (Blair et al., 1989). The V-ATPase has also been linked to cancer invasiveness and has been studied as a target for cancer therapies (Fais et al., 2007; Perez-Sayans et al., 2009; Sennoune et al., 2004). However, there are currently no therapeutically viable inhibitors targeting the V-ATPase and little previous effort has focussed on a rational design-led approach as complete structural information has proved extremely challenging to obtain.

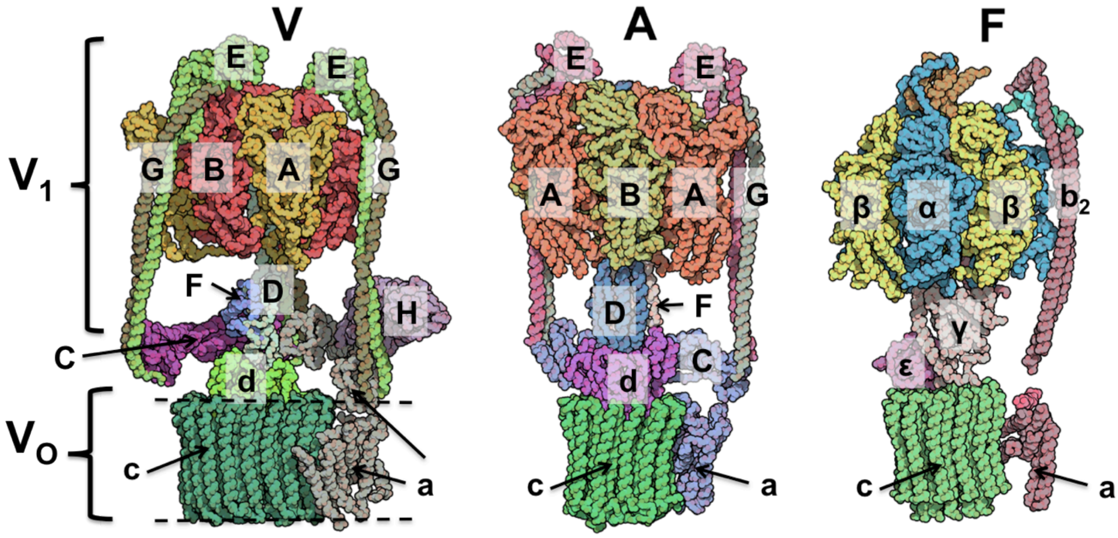


Figure 1.5 – Schematic representations of the rotary ATPase family. V,A and F-ATPase schematics shown from deposited atomic models with subunits labelled (Rawson et al., 2016b).

The common feature of the rotary ATPase family is the coupling of a membrane bound proton translocating domain and a soluble, ATP hydrolysing/synthesising motor. While each member of the family has a high degree of similarity with the others, they posses varying degrees of complexity (Muench et al., 2011). The best studied system; the F-ATPase (often referred to as the ATP synthase) (Walker, 2013), is the simplest of the family, possessing only one stator. It primarily uses the proton gradient and flow of H^+ through the F_o domain to generate torque and drive ATP synthesis in the F_1 region. ATP synthesis is carried out through the Boyer mechanism (Boyer, 1997) (Figure 1.6). This consists of a cycle where the three AB pairs move through open, loose and tightly bound states as ADP and phosphate is bound, ATP is formed and finally released. Functional and structural studies have identified several additional sub-states within the catalytic cycle (Masaike et al., 2008; Yasuda et al., 2001), with EM studies on the yeast complex finding 7 separate sub-states (Zhou et al., 2015).

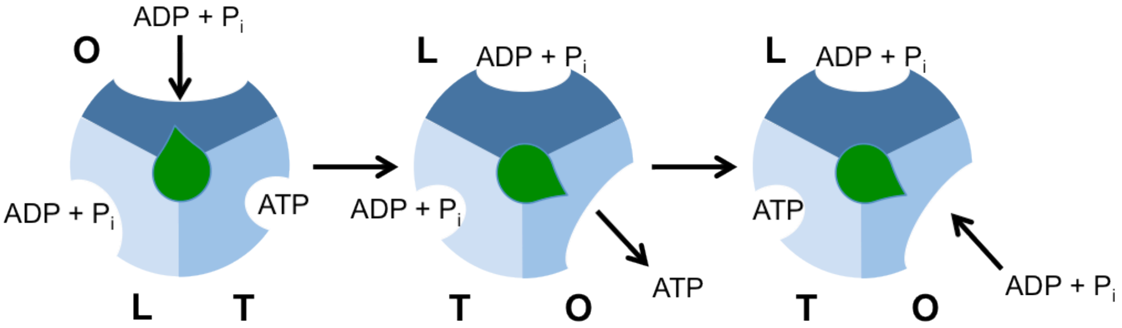


Figure 1.6 – Diagram of the Boyer Mechanism. Schematic of the Boyer mechanism of ATP synthesis showing the cycling between loose (L), tight (T), and open (O) sites as ADP and P_i binds and ATP is released.

Recent EM and crystallography studies have shown the interface where proton translocation occurs between the rotating *c*-ring and the static subunit *a* (Morales-Rios et al., 2015; Zhou et al., 2015). At this interface the proton enters one side of a half channel within *a*, is picked up by a conserved glutamate residue on the *c*-ring and following a full rotation of the rotor is then removed into the half channel on the opposite side (Figure 1.7).

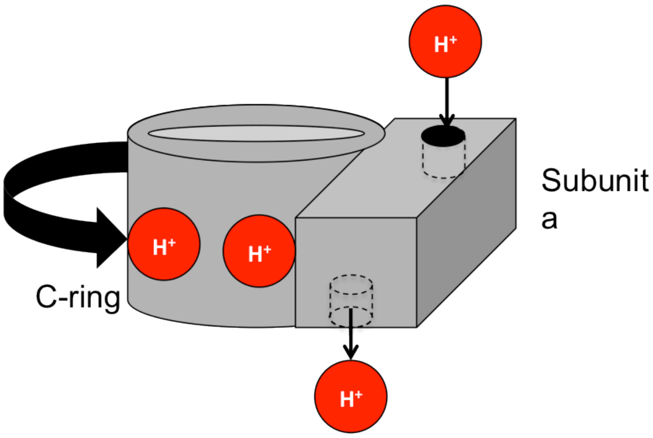


Figure 1.7 - Illustration of proton translocation. Diagram showing proton translocation across a membrane with the protons entering via a half channel, are driven around the *c*-ring and exit via a second half channel.

While generally shown as a monomer it has been shown that in many systems the F-ATPase exists as a dimer with each complex bridged by subunits e, g and k (Hahn et al., 2016). The F-ATPase has also been implicated in membrane remodelling and appears to induce ridges in membranes with the F-ATPase dimers lying in rows along the ridge (Figure 1.8) (Davies et al., 2012; Mühleip et al., 2016). Studies have shown that this V-shaped arrangement is crucial for the function of the complex and has been compared to the arrangement of pistons within an engine (Stewart et al., 2013).

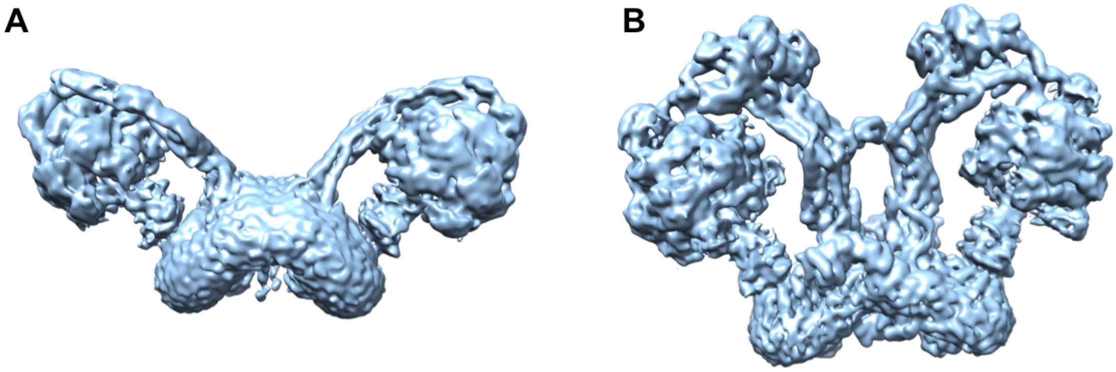


Figure 1.8 – F-ATPase dimers. Example dimers of the F-ATPase from (A) Yarrowia lipolytica (EMDB-8151) and (B) *Polytomella* sp. (EMDB-2852) (Hahn et al., 2016).

The second member of the ATPase family is the A-ATPase, which differs in organisation to the F-ATPase in having two stators rather than one (Lau and Rubinstein, 2012; Schep et al., 2016). Each of these stators is made of just two coiled coils rather than the multi-subunit stator found in the F-ATPase, similar to the V-ATPase. It is found in archaeal bacteria and unlike the other two members of the family can both hydrolyse and synthesise ATP depending upon the requirements of the organism. Despite the differences in function and in structure, the A-ATPase is still often referred to as the V/A-ATPase within the literature.

Of particular interest is the V-ATPase, which acts via the hydrolysis of ATP in order to drive proton translocation across a membrane. The V-ATPase has the same overall architecture as the other ATPase family members consisting of the soluble ATP hydrolysing domain (V_1) and the membrane bound proton-translocating domain (V_o). V_1 consists of three AB subunit repeats where ATP hydrolysis takes place, subunits D and F that make up the top of the rotating axle, and three stator stalks consisting of subunits E and G in a coiled coil. The three stators are linked by a series of 'collar' subunits, C, H and the soluble domain of subunit *a*. The membrane embedded V_o domain consists of the *c*-ring, made up of a variable number of *c* subunits and the remainder of subunit *a*. There can also be variations of the subunit composition of the *c*-ring with modified *c*′ and *c*′′ observed in some species. However, given the observed step size in the F-ATPase and the varying number of *c* subunits present in differing V-ATPases, this is calculated to result in a non integer number of protons crossing the membrane per ATP (Nakanishi-Matsui et al., 2010). In addition there is a symmetry mismatch between the two domains of the complex, with the V_1 domain having three sites for hydrolysis in comparison to the variable stoichiometries of the *c*-ring in the V_o domain. These factors combine to create a mismatch in the energies required. It has been proposed that some form of energy storage mechanism within the complex is required (Nakanishi-Matsui et al., 2010), and it has been speculated that the stators may play this role, storing excess energy through torsion, and releasing when required.

In some environments the V-ATPase is thought to consume up to 10% of cellular ATP, making regulation extremely important (Dow, 1984; Huss and Wieczorek, 2007). The complex therefore requires a mechanism to sense and respond to several factors including ATP and amino acid concentration (Cotter et al., 2015; Stransky and Forgac, 2015). Previous work on the complex from both yeast and *M. sexta* has shown that under low levels of ATP induced by starvation the complex is regulated via some degree of dissociation between the V_1 an V_o domains (Kane, 1995; Sumner et al., 1995). While the studies in yeast have suggested full dissociation of the domains, imaging studies of the *M. sexta* complex suggest that the localisation of the V_1 domain does not move from the membrane region, suggesting a more complex mechanism than simple dissociation (Tabke et al., 2014).

Following this dissociation the V_o domain becomes impermeable to protons, preventing leakage, and the V_1 domain can no longer hydrolyse ATP, preventing futile ATP consumption. However, the structural changes associated with the silencing of ATP hydrolysis in the V_1 domain are not yet fully understood. Several studies have suggested that this silencing is due to interaction between subunit H and the central rotor axle (Diab et al., 2009; Jefferies and Forgac, 2008). Indeed, crystal structures are now available which support this theory (Oot et al., 2016), although this conclusion may be due in part to artefacts introduced by crystal contacts, discussed in more detail in Chapter 4. Also, the resolution of the crystal structure is very low (6.2 Å, within the bottom ~0.1% of entries in the PDB), potentially leading to uncertainty in the phasing of the structure. Additionally, low resolution EM studies have not been able to identify any density which can be assigned to the H subunit in isolated V_1 samples (Muench et al., 2014b).

During normal activity, a mechanism must also be in place to prevent futile rotation of the ATPase or backwards rotation of the *c*-ring which would drastically lower the efficiency of proton translocation. In order to achieve this some form of brake or ratchet-like mechanism would be required. Indeed some additional mechanism, not found in A-ATPases, must be present to prevent the V-ATPase from acting as an ATP synthase under low ATP conditions.

In order to exploit the V-ATPase as therapeutic target, an additional level of complexity must be considered: the isoform specificity of the inhibitor. While many inhibitors for the V-ATPase already exist, including bafilomycin (Zhang et al., 1994), archazolid (Huss et al., 2005), and the enamide inhibitors (Huss and Wieczorek, 2009), they are extremely toxic and thus unsuitable as therapeutics due to their isoform promiscuity, with some exceptions (Niikura et al., 2004). For example, while it may be beneficial to inhibit the V-ATPase in osteoblasts in order to treat osteopetrosis, if other V-ATPase isoforms throughout the body are also inhibited the effect would likely be lethal. Due in part to a lack of detailed understanding of the mode of action of these inhibitors, and no high resolution structure of the intact complex, it has so far been challenging to design inhibitors which do not suffer from this drawback.

It is interesting to note that subtle differences between the rotary ATPases of different families may allow selective targeting of some species. For example, while the F-ATPase appears to be vital for the survival of the malaria parasite during the mosquito phase (Sturm et al., 2015), no homologue for the crucial subunit *a* required for proton translocation can be identified in BLAST searches (Balabaskaran Nina et al., 2011). This suggests that either subunit *a* is scavenged in some manner from the host organism or that it exhibits a drastically different sequence and fold than typical, so it cannot be found through BLAST searches, though some candidates for the *a* subunit have been proposed (Mogi and Kita, 2009). This may allow for it to be targeted with a degree of specificity. Subunit *a* also cannot be located within the related *Toxoplasma gondii* and *Neospora* parasites (Seeber et al., 2008), potentially hinting at a common sequence within the larger Apicomplexa family. Substantial differences also exist across species going from prokaryotic systems to simple eukaryotes such as yeast and onto higher eukaryotes. These differences come in both structure and function and are relevant to the crucial questions of regulation within the complex. In addition to reported differences in dissociation between the yeast and the *M. sexta* complex, additional factors have been identified within the yeast system that are required for re-assembly of the complex, including RAVE (Smardon et al., 2002; Smardon and Kane, 2007), although re-assembly has not been directly observed and is still poorly understood.

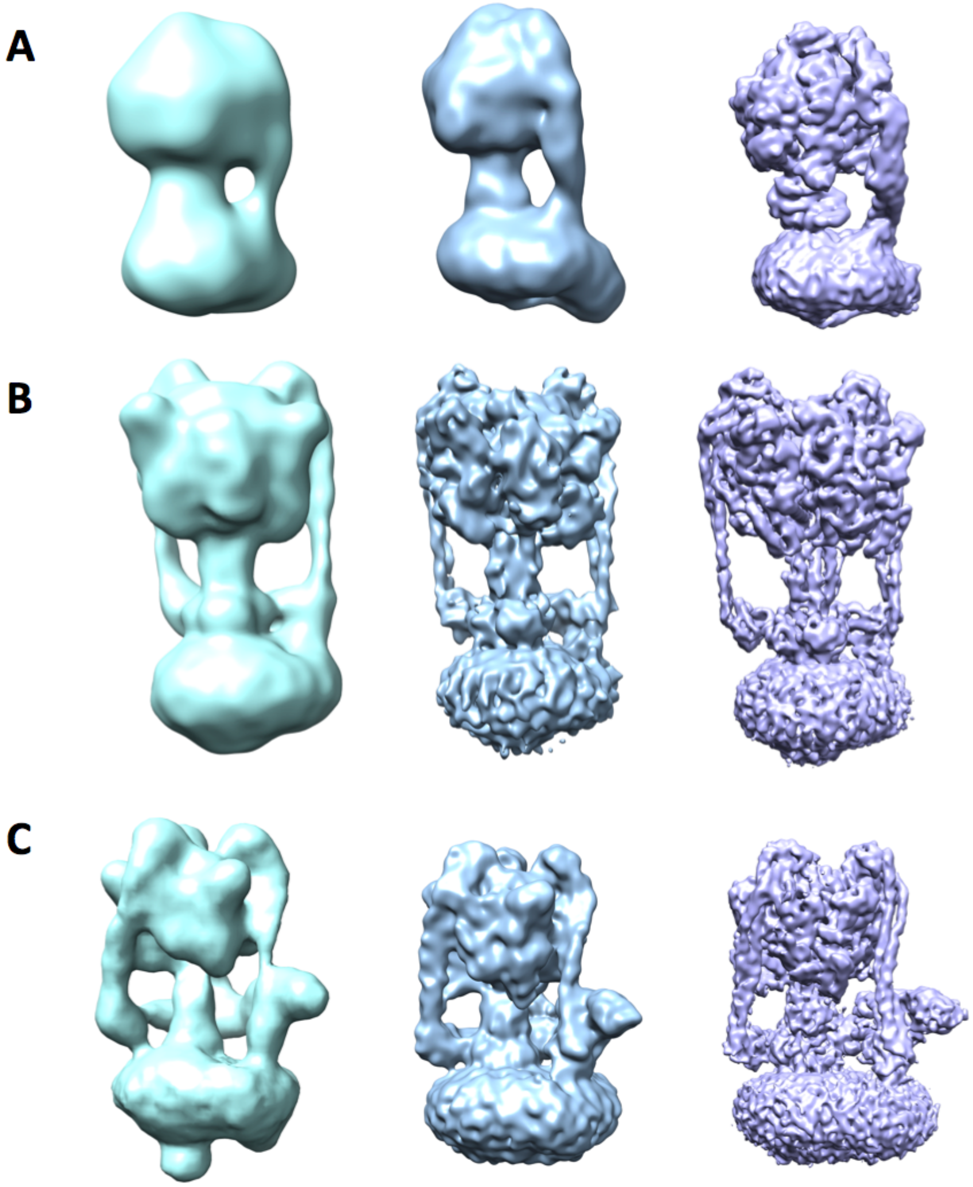


Figure 1.9 – Improvement in EM reconstructions of the rotary ATPase. (A) F-ATPase structures determined by EM at ~32 Å, 18 Å and 6.4 Å resolution (EMDB accession codes: 1357,2091,3169). (B) A-ATPase structures determined by EM at ~16 Å, 9.7 Å and 6.4 Å resolution (EMDB accession codes: 1888,5335,8016). (C) V-ATPase structures determined by EM at ~17 Å, 11 Å and 6.9 Å resolution (EMDB accession codes: 1590,5476,6284). Figure taken from (Rawson et al., 2016a).

In many respects structural studies of the rotary ATPase family have mirrored the progress within the wider EM field (Figure 1.9). This is due to the continuing evolution in EM, brought on by improvements in all areas of the technique; from the microscope stability, detector performance, to improved algorithms and crucially better sample preparation. These advances have allowed ever higher resolutions to be attained and in doing so have formed the basis for greater functional understanding of these vitally important molecular machines. When the work in this thesis was undertaken there were several unanswered questions and challenges relating to the V-ATPase including: a sub nm resolution structure of the intact complex, the structural identification of catalytic states within the complex, the architecture of the proton translocating *a* subunit, the location and purpose of subunit *e*, and the mechanism of regulation of the complex. This thesis aimed to address some of these issues.

Furthermore, this thesis will explore the potential of EM to inform drug design, focussing initially on the V-ATPase. EM will be used to improve structural knowledge of the V-ATPase, leading to insights into mechanism, regulation and ultimately inhibition. Studies on a model system will focus on the power of high resolution EM structures, made possible by advances in technology, to inform inhibitor design.

2. Materials and methods

2.1 Basic background/Instrumentation

2.1.1 Image Formation

All electron microscopes have the same broadly similar layout and components, consisting of an electron source, a series of magnetic lenses, a specimen holder and a detection system, all of which are held under vacuum (Figure 2.1).

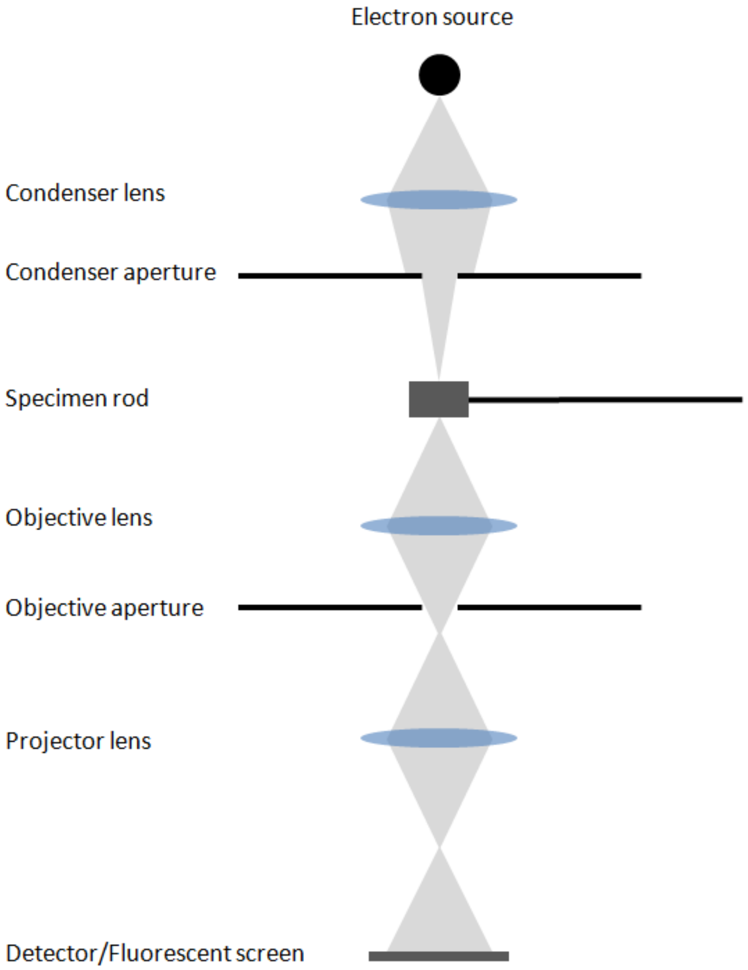


Figure 2.1 – Electron microscope schematic. Schematic drawing of an electron microscope showing beam path (grey) and lenses (blue).

In transmission electron microscopy (TEM) the image observed is a projection of the specimen made from electrons which have been transmitted through the sample and scattered (Orlova and Saibil, 2011). This scattering can occur in two ways; elastically scattered electrons, where the path of the electron is altered but no energy is transferred, and inelastic scattering, where energy is deposited from the incident electron to the sample. In standard TEM, while both types of scattered electrons are collected, only elastically scattered electrons are useful for image formation. However, it is estimated that for each elastic scattering event 3-4 inelastic events occur (Henderson, 1995). The energy deposited into the sample by these electrons can cause many detrimental effects including direct damage to the sample, formation of free radicals and bond rearrangement. These damaging effects are a key limitation in EM of biological specimens, due to the exposure times needed to form a usable image and the inability to recover the information from the earliest part of the exposure (Glaeser, 2016), which will be discussed in more detail elsewhere (Chapter 1). It is also possible for a single electron to undergo multiple scattering events as it passes through the sample layer, adding to the noise in the image. To minimise this it is thus important that the specimen layer is kept as thin as possible to lower the chances of multiple scattering events.

A further complication in the imaging of biological samples via EM is contrast. Again contrast can be split into two categories; phase contrast and amplitude contrast. In materials with a high atomic number, for example commonly used stains containing uranium, amplitude contrast dominates. This results from the loss of electrons due to high angle scattering by the heavy metal. However, for light atoms for example O,N,C,H that make up the majority of protein, there is very little absorption and thus amplitude contrast is minimal for an unstained protein. In this case phase contrast dominates, which results from interference between the scattered and unscattered beam. Unfortunately biological samples are also classified as weak phase objects, meaning that they induce not only minimal amplitude contrast, but also very little phase shift leading to little phase contrast (Orlova and Saibil, 2011). To counter this, the ideal optics of the microscope must be sacrificed through the use of defocus to generate a larger phase shift and thus a higher contrast image. The need to defocus further complicates the processing of the images and high defocus can limit the practically obtainable resolution. Thus data collection becomes a balance between achieving sufficient contrast through dose and defocus to see, pick and align particles and imaging close enough to focus and with as small a dose as possible to preserve as much high resolution information as possible.

2.1.2 Detectors

The method of recording the image is of vital importance to the quality of the data obtained and thus the potential resolution possible. There have been three widely used methods of recording images; film, charge-coupled device (CCD) cameras and direct electron detectors (Figure 2.2), each with distinct advantages and disadvantages. The oldest and simplest detection system is film where the image is formed through interaction of the scattered electron with a thin layer of silver halide causing a chemical reaction/rearrangement of the surface material at the point of electron interaction. Until the advent of direct detectors, film was the recording medium of choice for high resolution studies (Bammes et al., 2011). Its fine grain size gave a high sampling while also giving a large field of view and a greater detector quantum efficiency (DQE) at high spatial frequencies than a typical CCD camera (McMullan et al., 2009a). However, film is expensive to buy, difficult to work with, limits the number of images that can be taken in a single session, and is slow to process, requiring first processing in a dark room and then digitising before use in EM processing packages.

In comparison, CCD cameras are relatively cheap, easy, and quick to use. CCDs operate through the conversion of the incoming electrons to photons via a scintillator layer. The photons are then carried via fibre optic to a CCD chip, digitized into an electronic signal and read out. Due to this series of conversions a large amount of noise is introduced into the image and typically the DQE of a CCD camera is significantly lower than film, particularly at high resolutions.

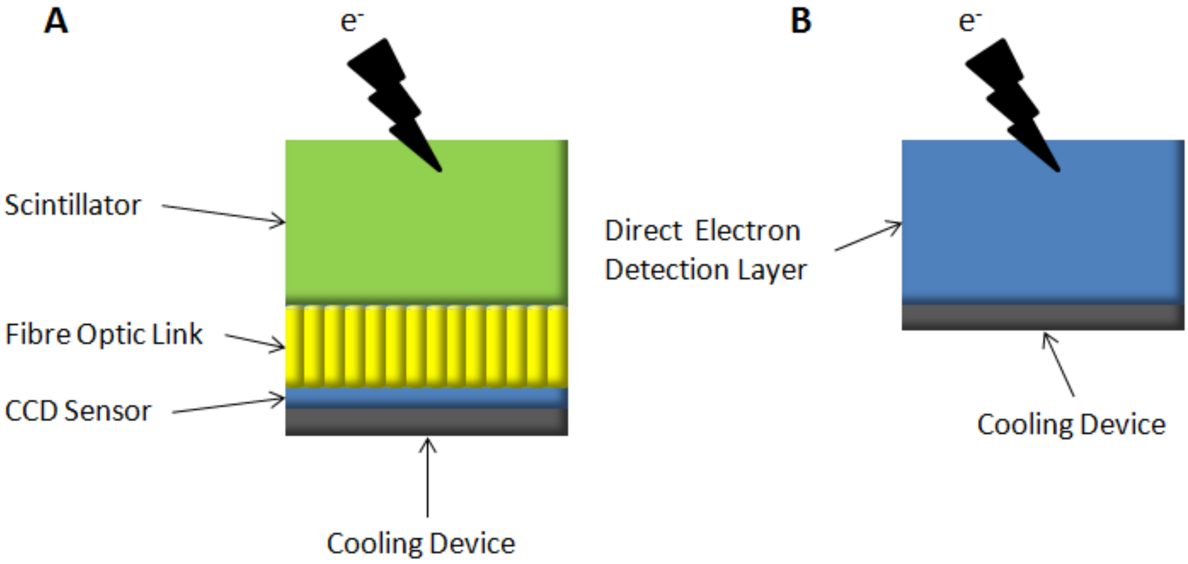


Figure 2.2 – Detector Schematics. Schematic representation of a CCD camera (A) and a DED (B)

Recent advances in detector technology have allowed the direct detection of electrons (Faruqi and McMullan, 2011; McMullan et al., 2014). Increases in the robustness of detectors, brought on through monolithic active pixel sensor (MAPS) technology (Faruqi et al., 2005), increased radiation hardness over traditional CCDs. Substantial thinning of the sensing layer minimises spreading of the signal, and backscattering events, which results in higher resolutions being obtainable (McMullan et al., 2009c). Another advantage of this method of detection is the high frame rate of the detectors, which allows images to be recorded as multiple frame movies rather than a single integrated image. This makes dose fractionation and drift correction possible by utilising the 'movie frames' which will be discussed in more detail in Section 2.4.1. Direct detectors currently act in one of two modes, integrating or counting, each with advantages and drawbacks. Integrating detectors such as the FEI Falcon II combine the signal from all electrons hitting the detector, which due to variability in electron energy can add noise to the final output, but this mode allows short exposure times and higher dose rates. In comparison, counting mode detectors like the Gatan K2 count individual electrons as they hit the detector (McMullan et al., 2014; 2009b), which with robust counting algorithms can more accurately localise the position and path of each electron. Indeed some detectors have 'super resolution' capability where sub-pixel localisation allows information beyond the Nyquist limit to be obtained. However, counting and super resolution modes rely on extremely low dose rates to minimise the potential for multiple electrons hitting the same pixel at any given time (Li et al., 2013b), and this in turn leads to much longer exposure times, which slows data acquisition. Thus these modes may suffer more from any mechanical specimen drift.

In order to image a biological specimen, it must be held on some form of support in the microscope. The grids that are generally used for EM studies consist of a metal mesh, generally copper, at a set spacing in a circular grid ~3 mm in diameter. Depending on the nature of the individual experiment, a variety of different support films are then added ranging from a continuous layer of amorphous carbon or regular arrays of holes in a carbon film to more exotic materials such as gold or graphene films. The hole size chosen, as well as the material used, can have profound effects upon how the sample is distributed on the grid. The quality of the image can also be directly affected by the support chosen. For example, different amounts of beam induced motion and drift have been observed between the standard carbon support film and so-called UltraFoil grids made from gold (Russo and Passmore, 2014).

2.2 Negative stain EM

Negative staining is used for low resolution EM studies of protein samples where the protein of interest is first coated in a thin layer of a heavy metal staining solution (Ohi et al., 2004). This heavy metal stain provides contrast in the microscope through amplitude contrast as described above. The sample is typically prepared on a grid with a continuous carbon support film which has been made hydrophilic by treatment either by irradiation under a UV lamp or exposure to a plasma in the form of plasma cleaning or glow discharge. A key advantage of negative stain analysis is the small amount of protein required, with typical protein concentrations in the nM range. The protein is first applied to the grid surface and allowed to adsorb onto the support film before blotting to remove excess liquid, washing if desired, and application of the staining solution (Booth et al., 2011). Depending on the precise method used, the stain is either blotted away and reapplied a number of times or flooded over the grid repeatedly. A variety of different heavy metal stains are available, including different uranyl and molybdenum salts, with the grain size of the staining solution limiting the maximum resolution obtainable. The most widely used stain, uranyl acetate, gives a maximum resolution of ~15 Å. In order to obtain good quality negative stain grids the particle distribution needs to be optimised by altering concentration. Crucially the stain depth must also be varied, which, depending on technique, can either be by varying staining time or number of applications of the stain. A key advantage of negative stain is the relative speed and ease with which it can be performed, allowing the user to rapidly assess sample quality before performing more arduous cryo-EM studies.

2.3 Cryo EM

Cryo-EM studies involve imaging the specimen of interest in a thin film of vitreous solution maintained at cryo temperatures (< -180 °C) to prevent formation of crystalline ice which would damage the fragile protein and interfere with the imaging conditions. The preparation of cryo samples is more challenging than negative staining due to the constant temperature requirement and sensitivity to contamination during grid preparation and loading (Grassucci et al., 2007). Holey carbon films are typically used for cryo-EM rather than the continuous carbon coated grids used for negative stain, although the size and organisation of these holes can vary as can the support material. Before application of sample, the grid must first be treated with glow discharge or UV in the same way as a negative stain grid in order to make the surface hydrophilic, although the exact method of treatment and time will vary depending on the sample in question and the support film used. In order to obtain the vitreous film, the sample is suspended between holes on a support film by applying the sample and then blotting away excess so only a thin film remains. The grid is then plunged into a cryogen to rapidly vitrify the sample (Cabra and Samso, 2015; Grassucci et al., 2007; Orlova and Saibil, 2011)(Figure 2.3).

The process of blotting and plunge freezing is usually carried out with specialised equipment, for example the FEI Vitribot. Depending on the equipment, the user can alter a number of parameters with the aim of optimising ice thickness on the grid. This includes the number of blots, how long the sample is blotted for, and any wait time between blotting and the plunge into the cryogen. Typically, liquid ethane is used as a cryogen rather than nitrogen because this avoids the formation of an insulating vapour layer which would slow the cooling process and lead to crystalline ice formation, although ethane/propane mixtures can also be used. Following vitrification, the grid is transferred to a storage puc under liquid nitrogen until required for imaging. In order to be imaged the grid must be transferred into a specially designed sample holder and inserted into the microscope without warming above approximately -160°C, significantly complicating the procedure in comparison to negative stain (Grassucci et al., 2008).

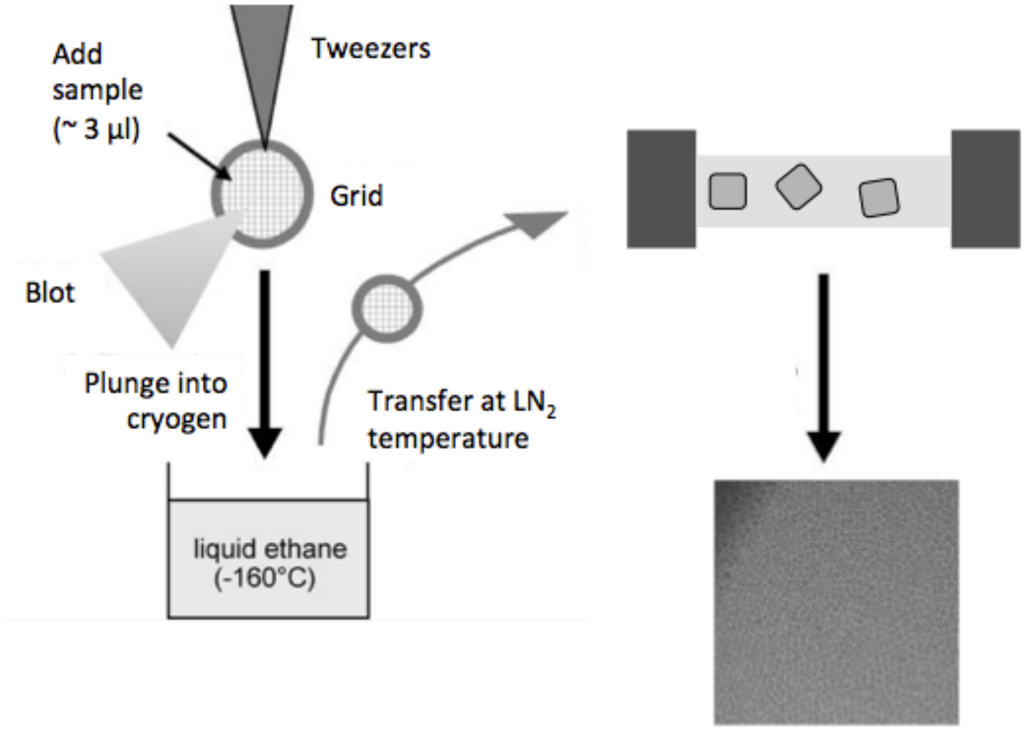


Figure 2.3 – Cryo Grid preparation schematic. Workflow of cryo-grid preparation adapted from (Orlova and Saibil, 2011). Initially sample is applied to the grid, then blotted to a thin layer prior to plunge freezing into the cryogen. This captures particles in random orientations in the vitreous ice and the grid is transferred into a microscope for imaging.

As biological specimens are extremely sensitive to radiation damage it is vital to control and limit the electron dose received by the sample (Baker et al., 2010; Grant and Grigorieff, 2015; Stark et al., 1996). In typical cryo data collection conditions this is achieved through a technique known as low dose imaging. For this reason microscopes are generally equipped with a low dose mode. In this technique the grid is first viewed in 'search mode' that uses low magnification and minimal dose to find the desired imaging area. Once an area has been selected, the beam is focused away from the area of interest, at a set distance and angle so that the specimen is not pre-exposed, before imaging the desired area.

Software developments and improved hardware integration have allowed data to be collected in a semi-automatic fashion. Rather than the user collecting each micrograph by hand then finding another imaging area and moving manually before focussing and imaging again, this can now be performed by software. Programs including Leginon (Carragher et al., 2000), SerialEM (Mastronarde, 2005), and EPU are all able to provide some degree of automation. While different software packages exist, each with different levels of user interaction depending on budget and hardware manufacturer, the general principles are consistent for all. In most cases a series of low magnification images are taken of the grid to build up a mosaic overview. From this, the user selects areas of interest based on desired criteria, for example thin ice, no cracking, and low contamination. If a grid type with a regular array is used, then a template can be specified so that viable holes for imaging are detected automatically. After selection of the desired areas, the program proceeds to image each of these in turn, checking defocus and correcting at set intervals. This allows data collections to continue for long periods without user intervention, maximising the efficiency of the time available. As well as allowing longer data collection sessions and making data collection significantly easier, another advantage is that the software is sometimes able to image in ways challenging for humans, for example taking several images in each hole using beam deflectors to increase total data obtained.

2.4 Image processing

Once the images have been collected, several processing steps are then required to go from noisy, low contrast micrographs to a high resolution structure. Each of these steps aims to either correct for imperfections in the imaging conditions or to deal with extracting the relevant data, aligning and averaging the images. While the exact order of these steps may vary slightly depending on the user's preference or the nature of the data obtained, it is possible to simplify this into a rough schematic (Figure 2.4). Each of these steps will be described in more detail in the subsequent sections below.

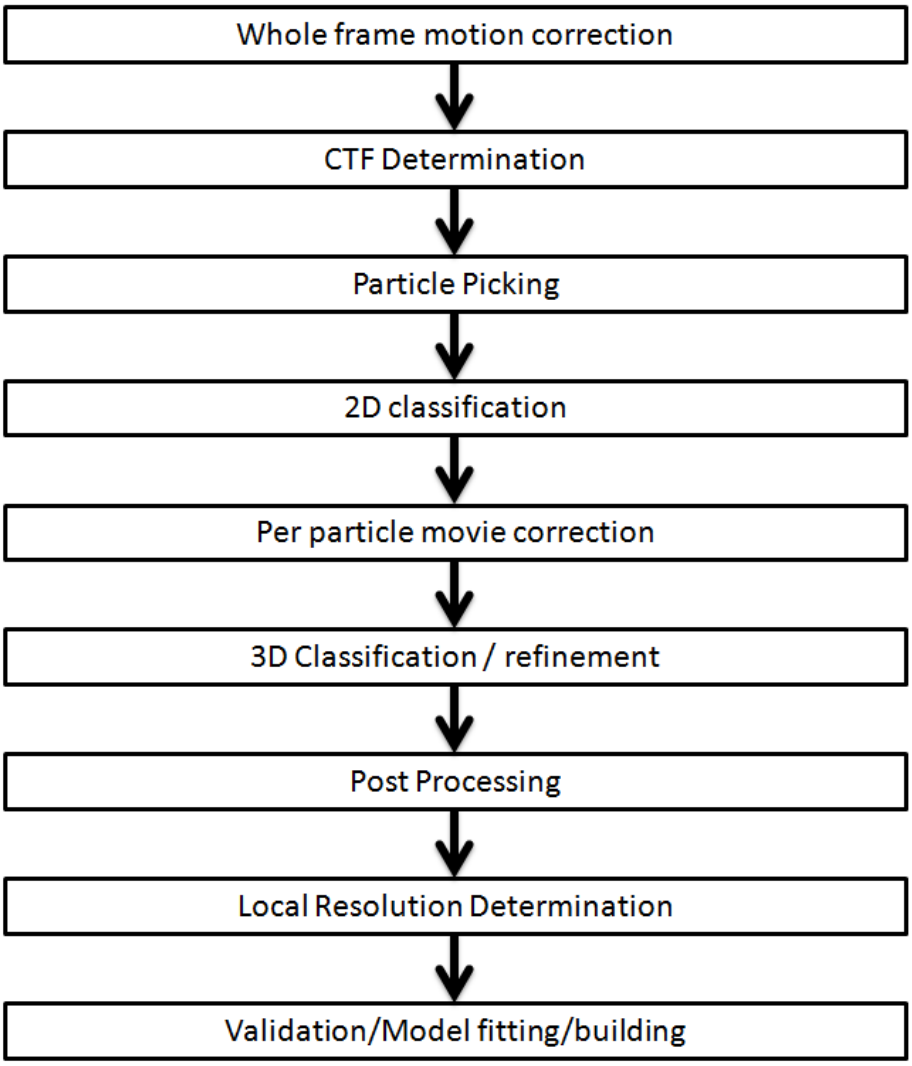


Figure 2.4 - EM processing workflow. Workflow showing typical processing steps taken in a cryo-EM project.

2.4.1 *Movie correction*

Due to the high frame rates of the new generation of DEDs it is possible to obtain a series of frames or a 'movie' from a single exposure rather than the single sum of the whole exposure, which was obtained using older recording media including film and CCDs. By utilising the movie frames obtained from DEDs, both the movement of the sample and the damaging effects of the electron beam can be mitigated to some extent. Large movements due to mechanical stage drift can be corrected for at the whole frame level, while more subtle and complex movements of the particles, which result from exposure to the electron beam, can be corrected for on a particle by particle basis. In addition to this motion correction, some of the effects of radiation damage can be lessened by the removal or weighting of later frames from the exposure, leaving only the frames containing the high resolution, undamaged information for reconstruction (Figure 2.5). This has allowed a new data collection strategy to be used, where very high (70-100 $e^-/Å^2$) doses can be used in order to generate higher contrast and improve picking and low resolution orientational assignments, before removal of the later, damaged frames, for use in high resolution reconstructions.

To carry out dose weighting, programs can use a standard curve of radiation damage at different resolutions, either from studies carried out on 2D crystal diffraction patterns (Baker et al., 2010), or more recently, using observed damage in high resolution single particle reconstructions (Grant and Grigorieff, 2015). This curve can then be directly used to weight each frame dependent on the dose received at that point. Alternatively, the data can be assessed on a case-by-case basis, for example particle polishing within RELION (Scheres, 2014). This module calculates a 'B-factor' for each individual frame based upon the quality of a 3D reconstruction from each frame individually. As the precise extent of radiation damage is likely to be subtly different on a case-by-case basis, this method may be potentially more accurate. However, if the quality of each frame reconstruction is not sufficient to accurately calculate a B-factor then the results may be unreliable. To combat this, frames can be averaged to improve the signal for estimation, but this comes at the expense of precision of the weighting obtained.

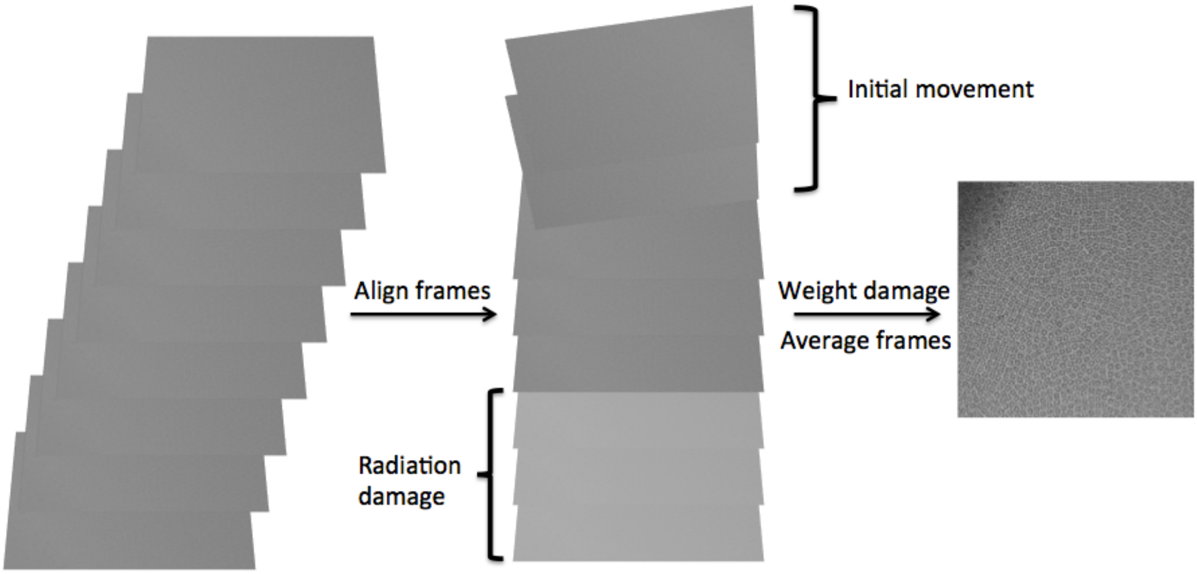


Figure 2.5 - Motion correction overview. Overview of motion correction showing first alignment of frames followed by weighting/removal of damaged frames prior to averaging to final corrected micrograph.

At the whole-frame level, there are several software packages available to carry out this motion correction including MOTIONCORR (Li et al., 2013a), alignframes_LMBFGS (Rubinstein and Brubaker, 2015), Unblur (Grant and Grigorieff, 2015), and modules within EMAN2 (Tang et al., 2007). While the details of each program differ slightly, the core principle is that each frame is compared to the other frames in the stack and then shifted to maximise the correlation between the stack. Variations include using a single frame as a fixed point of reference or aligning each frame to the sum of all frames and iterating this process after each frame has been moved. Whole frame motion correction is typically the first operation performed on movie frames as by accounting for this motion the accuracy of all subsequent steps in processing is improved, especially for long exposure movies often obtained from K2 detectors. While this is a valuable method for removing large, generally unidirectional shifts, it fails to capture the complex motions of individual particles.

In order to better capture the complex ‘swirling’ motions that particles within a micrograph undergo each particle must be considered both individually and correlated to its neighbouring particles (Figure 2.6). This can be carried out by a variety of software packages including the particle polishing module within RELION, alignparts_LMBFGS and Unblur.

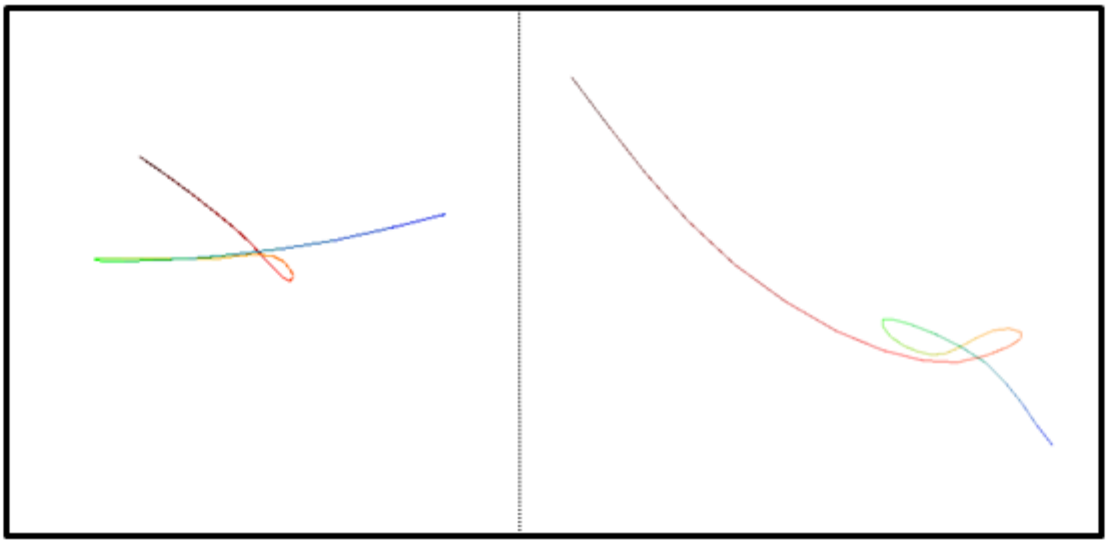


Figure 2.6 - Particle motion example. Example of complex trajectories taken by single V-ATPase particles in ice following exposure to electron beam (Chapter 3). Generated in alignparts_LMBFGS (Rubinstein and Brubaker, 2015).

2.4.2 CTF correction

The need for contrast transfer function (CTF) correction is a direct result of the inability to create the perfect microscope. CTF can be thought of as the mathematical description of the microscope's point spread function. It results from the combined effect of all the imperfections within the imaging conditions, including chromatic and spherical aberrations within the lens systems as well as any other flaws within the microscope. Thus images that are recorded can be thought of as a convolution between the information in the object imaged and the CTF of the microscope, with the addition of noise from imperfect detectors. As the effects of CTF occur in a periodic manner related to resolution, the resulting power spectrum is generally plotted in Fourier space as an oscillating curve with decaying amplitudes at higher resolutions. As previously discussed, defocus is a necessary evil in cryo-EM; in order to generate contrast within weak phase biological specimens defocus must be introduced in the imaging process. However, increasing defocus leads to higher frequency oscillations and the shifting of the curve towards lower resolutions, acting to reduce the amount of high-resolution information recovered.

The first step of CTF correction is the determination of the defocus and CTF parameters of the data. Several programs, including CTFFIND (Rohou and Grigorieff, 2015) and gCTF (Zhang, 2016), estimate the defocus of each micrograph and attempt to accurately model the power spectrum of the micrograph. This is then accounted for during subsequent CTF correction steps (Mindell and Grigorieff, 2003; Rohou and Grigorieff, 2015; Zhang, 2016). It is also possible to calculate a defocus value for each particle within a micrograph, which can account for a defocus gradient across a micrograph and/or differences in each particle's z position within the ice layer.

The oscillating nature of the CTF leads to several spatial frequencies with zero contrast at any given defocus value and negative values at several points. Thus the simplest form of CTF correction is phase flipping to remove any negative values. The simplest method to remove the zero points is to vary the defocus value across a range of a few microns during data collection so that when the data are averaged together no zero points will remain. There are also more complex treatments including amplitude correction, which seek to account for the decaying amplitudes at higher spatial frequencies. However this approach requires an extra layer of complexity due to the low SNR of EM images of biological specimens, which could lead to an amplification of noise rather than signal unless the data are optimally filtered. This complex treatment is automatically carried out within several software packages including XMIPP and RELION (Scheres, 2012; Scheres et al., 2008).

2.4.3 *Particle Picking*

The first step of single particle processing is selecting the particles themselves. Particle picking can be manual, semi-automatic or fully automated. While manual picking, where the user selects each particle individually, allows greater control and confidence in the particles picked, it is extremely slow and laborious, particularly for low contrast cryo datasets often consisting of 1000s of micrographs. The throughput can be increased by using automated methods including swarm picking in EMAN2 (Tang et al., 2007), and automatic particle picking in RELION (Scheres, 2015), and Gautomatch (In development by Kai Zhang (Urnavicius et al., 2015)). These programs require the user to pick a number of particles which are classified and averaged to make templates used for automated particle selection. The programs then attempt to find any density in a micrograph that matches the templates provided, with additional user parameters to minimise the selection of contamination, aggregation or particles lying on the support film. However, the introduction of a template can potentially introduce bias into the data, particularly if a template is obtained from an existing structure or is not filtered to low resolution (Henderson, 2013;

van Heel, 2013). Additionally, if the templates only represent a subset of the particle orientations, this may cause only that subset of views to be chosen and thus introduce the corresponding artefacts into the dataset.

2.4.4 Alignment and classification

In EM, single particle analysis (SPA) is fundamentally an averaging technique, conceptually similar to crystallography. While the crystal lattice orients each copy of the protein for crystallographers, in EM this alignment and averaging must be performed *in silico*. It is this averaging process that allows the extraction of near-atomic resolution information from extremely low contrast, low SNR images. The key step is determining the orientation for each individual particle and calculating its angular relationship to the other particles in the dataset. Particles are then classified based on orientation prior to averaging. Several approaches have been used to implement this in alignment and classification in 2D including 'K means classification', 'equal distributed K means' and 'maximum likelihood (ML)' approaches

(Sigworth, 1998; Sigworth et al., 2010). This alignment and classification can be performed either using user-selected references or in a reference-free manner. In reference-based classification the user first selects a series of particles to which the dataset is then aligned, often through iterative cross-correlation steps. Reference free programs instead act in an unsupervised manner without the user defining the initial references. This removes some potential user bias as the particles cannot be forced into an incorrect alignment or class by the user. In both approaches the user selects the number of classes they require and these are computed based on similarity of the particles and each particle will contribute to only the class to which they are assigned. In the ML approach each particle will contribute to all classes, with a probability weighting assigned to each class so that in practice a particle will contribute strongly to one class or a small subset of similar classes and then negligibly to the other classes. This weighting can be used to identify particles which do not align well to any class, or rather equally well to many classes as this can indicate the particle does not represent any particular view and may be contaminant or degraded in some way. The advantage of the maximum likelihood approach is the way in which it deals with poor quality data. Because SNR is so low in single particle images, there is a degree of uncertainty in the orientational assignments. By ‘smearing’ the data across many classes, ML approaches can account for some of this error. Indeed, for a perfect image with high SNR where orientational assignment and classification is perfect, ML would be equivalent or slightly worse than traditional one particle, one class approaches.

This classification approach can be expanded to three dimensions and has been implemented in several software packages. This can be a particularly powerful mechanism for sorting heterogeneous datasets and identifying homogeneous subsets suitable for high-resolution refinements. In the ML 3D classification within RELION, the user selects the number of desired classes, typically 3-5 depending on overall particle number, bearing in mind that for asymmetric objects a minimum of ~5-10k particles are required to generate a good quality 3D reconstruction. Initially particles are randomly assigned to a 3D class based upon an initial model leading to a series of subtly different classes. Classification then proceeds in an iterative fashion, typically ~25 rounds until the classification has stabilised or reached convergence. This classification can be repeated on particle subsets which have been thus selected in a hierarchical manner, allowing the identification of lowly occupied populations (Fernández et al., 2013)(Figure 2.7). In addition, focussed masks can be applied to an area of particular interest to discover any subtle variations in this region, which may not be observed when considering the entire molecule.

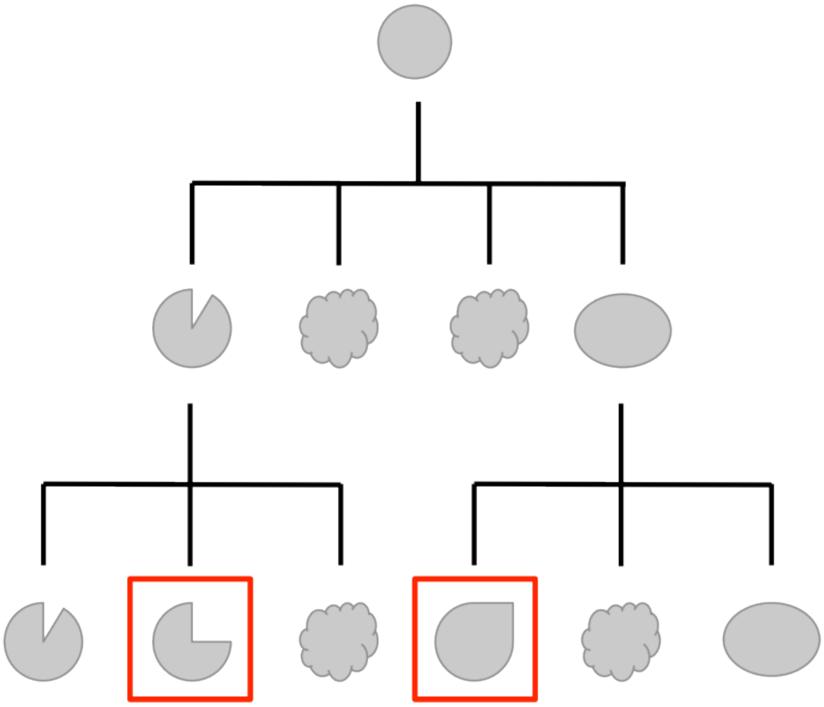


Figure 2.7 - Example 3D hierarchical classification. Example of 3D classification being used in an iterative fashion from a consensus model to identify lowly populated states (boxed in red) within a larger dataset (Clouds represent 'poor' quality classes).

2.4.5 3D reconstruction

In order to move from a 2D projection or class to a 3D volume, SPA uses Fourier slice theorem, which states that for a given object, the Fourier transform of a projection represents a central slice through the Fourier transform of the 3D object. In order for this approach to be successful, the orientation of each projection needs to be estimated prior to reconstruction. How this estimation is achieved varies depending upon the software package used. A typical approach is projection matching, where a rough starting model is backprojected and then compared to the experimental data. This is then used to bootstrap a new model, and this process iterated to move away from the starting model. The angular assignment can also be carried out in a maximum likelihood manner where, as in classification and alignment, the angular assignment is 'smeared' across several possibilities to account for uncertainty in estimation of the Euler angles. As with classification steps, the reconstruction process is iterated several times to move away from the starting model towards purely experimental data and hopefully convergence. It is also important to note that in most modern reconstruction programs the data are split into two independent subsets and each refined separately within the program. The agreement between the two halves is monitored as a surrogate for resolution which will be discussed in more detail below (Section 2.5).

One potential pitfall for users generating 3D reconstructions from EM data is model bias. As the SNR of the individual images is very low, it is vital that initial models used in the reconstruction process are not only accurate, but are also filtered sufficiently that the high-resolution features of the final map are a product of the experimental data rather than simply carried over from the starting model. Therefore, the generation of a good starting model is important in reconstructions efforts. A common method of generating an accurate starting model for a refinement is to use an existing atomic model of the sample derived from crystallography. However, this is potentially extremely dangerous, as if unfiltered, the high-resolution crystal structure will allow alignment of high resolution noise in the dataset, leading to a nominally 'high-resolution' result that is based on incorrect alignments and is entirely due to model bias. Ideally the starting model used would be derived directly from the experimental data with little or no input from the user or prior knowledge of the sample. Several programs exist that can accomplish this including EMAN2 and SIMPLE-PRIME (Elmlund et al., 2013; Tang et al., 2007). EMAN2 for example, works by assigning selected classes to a random starting model and checking for agreement between the reprojected map and the reference-free 2D classes.

Once the refinement process has finished, the user is left with a 3D volume that hopefully represents the experimental data. However, there are still a number of steps that can be performed to improve the quality of this map. The simplest of these steps is masking of the final volume to remove background noise so that when the Fourier shell correlation (FSC) is calculated (discussed further in Section 2.5), only regions corresponding to the sample are compared. This leads to removal of noise around the model and potentially a higher reported resolution by FSC, as uncorrelated noise will be removed. It is also common for the map to be filtered to the average resolution of the reconstruction as determined by FSC, removing high-resolution noise from the final reconstruction. Nevertheless, in maps with significant disparities in resolution across the volume this can still result in areas containing high-resolution noise and small 'popcorn' density. To combat this, programs have been developed to utilise local resolution measures to filter different areas of the map according to the resolution in that region (personal communication Dr CHS Aylett), which can lead to a cleaner looking map.

It is often apparent that in the raw reconstruction, a map of a given resolution will appear less featured than expected based purely on the resolution value given. This is because high-resolution features are often lost due to imperfections in the imaging, errors in orientational assignment, and the Gaussian blurring of the image giving rise to a B-factor decay of the structure factors. In order to regain these features, the high resolution signal must be boosted, often called map or B-factor sharpening. While the user can estimate an ad hoc B-factor, which is then applied to the reconstruction, this can lead to oversharpening where high resolution noise rather than signal is amplified, often visually identified through 'sharp' and fragmented density. To avoid this approaches have been developed that seek to fit the decay of structure factors in a given map to a theoretical scattering curve and use a weighting function to calculate the appropriate levelling of sharpening required independently of the user (Rosenthal and Henderson, 2003). This is unreliable at more modest resolutions, requiring <9 Å maps as a minimum.

2.4.6 *Model building/Flexible fitting*

Even at modest resolutions, EM allows important biological insights to be obtained through combining high-resolution structural information obtained via other techniques such as crystallography or NMR. Existing structures can be rigidly fit into a lower resolution EM map using several tools, the simplest being the Fit in Map function within Chimera (Goddard et al., 2007), although there are a number of more sophisticated tools including Situs and DockEM (Roseman, 2000; Wriggers, 2012). These programs generally work by maximising the correlation between the atomic model and the EM map, and following this rigid body fit can provide a pseudo-atomic model of a complex from multiple atomic models into the EM density.

In maps of higher resolution (< 9 Å), where secondary structure becomes visible, it becomes possible to begin to fit atomic models into EM density with much greater accuracy at the secondary structure level. This resolution allows the model to be flexibly fit into the density rather than simple rigid body fitting. While the individual crystal structures will be of higher resolution, often the structures of individual subunits are not representative of their organisation in the complex as a whole. This can be due to crystal packing artefacts or the result of lacking neighbouring subunits within the complex (Rawson et al., 2015). By combining lower resolution EM information with high-resolution crystal structures it is possible to obtain new insights into complex organisation. There are several programs that can be used for flexible fitting including MDFF

(Trabuco et al., 2011), FlexEM (Topf et al., 2008) and the relax protocols within Rosetta (DiMaio et al., 2015). While each program functions slightly differently, one approach is to perform a MD simulation on the input atomic model, allowing the protein to relax with a weighting applied to the EM density to encourage the model to move into the experimental density. However, at these intermediate resolution ranges it is important to avoid overfitting, or over-interpreting any deviations observed from the higher resolution structural information.

If the EM map obtained is of sufficient quality and resolution (~3.5 Å or better), it becomes possible to build an atomic model into the density *de novo*. There are several methods to do this, many adapted from existing crystallography programs including Coot for model building and visualisation (Emsley et al., 2010), Buccaneer (Cowtan, 2006) for initial model generation and REFMAC5 for refinement into the EM density map (Brown et al., 2015). In addition, Rosetta can attempt to automatically build the entire model into the EM map *ab initio* from only sequence information by breaking the sequence into a series of overlapping short segments. The most likely conformations for each short segment are predicted and then a search is performed over the entire EM volume for the best fit consistent with the location of previously placed fragments. This process is then iterated several times with outputs scored based on agreement with the EM map and a series of metrics related to likely bond geometries and prior knowledge of typical protein structure (DiMaio et al., 2015; Wang et al., 2015).

2.5 Structure analysis/validation

Unlike light microscopy, the term resolution frequently used in biological EM does not refer to a physical property, or the ability to discriminate between two nearby points. When used in respect to 3D reconstructions, generally an FSC value is referred to as the resolution. FSC is a measure of correlation at a given resolution and is typically calculated by splitting the data into two and measuring consistency between the halves at each resolution shell. This has been an area of intense debate within the field as a map could be entirely wrong yet, if self-consistent at high spatial frequencies, would be reported as a 'high-resolution' structure. In addition if the data were processed as a whole and only split before the FSC determination, it is possible to introduce spurious correlations at resolutions beyond the actual information contained in the data. Thus the independent processing of each half is vital (Scheres and Chen, 2012). A range of validation steps are required for reconstructions produced at all resolution ranges. The value at which the FSC should be measured has also been the subject of heated debate within the field, with many studies using a correlation cutoff of 0.5 between two half-maps, but a consensus value of 0.143 correlation has been established.

The resolution determined by FSC refers to the map as a whole, but it is well established that biological specimens have some regions better-resolved than others. Tools have been developed to calculate this variability in resolution across a map, including ResMap (Kucukelbir et al., 2013). A feature of X Å is detected in a voxel by testing for the presence of a X Å sinusoid above the noise calculated from the surrounding voxels. This is followed by applying controls for false detection and following testing across a range of values of X. The local resolution is given as the smallest value of X where the sinusoid is detectable above the noise level at a given confidence value. Another way of modelling the variability within the sample is through the use of B-factors. In any built or fitted atomic model B-factors reflect the uncertainty in the position of any atom or residue, and can be obtained during refinement of an atomic model into the map. An alternate method of visualising this is through the use of consensus models, where the fitted model is used to generate an ensemble of models that are then scored against the map prior to a subset undergoing refinement. RMSD values between all refined models can then be plotted across the sequence; areas with high values are then associated with poor map density and greater uncertainty in the position of the atomic co-ordinates.

As previously discussed, EM is particularly vulnerable to reference bias at several different steps within the processing pipeline. This can cause the well-established problem of ‘Einstein from noise’ where a reasonable-looking structure is obtained from pure noise due to the incorporation of reference bias (Henderson, 2013; Shatsky et al., 2009). The first stage where model bias can be introduced is during particle picking, particularly if an automated picking program that relies on template matching is used. If a reference containing high resolution information, such as one generated from a high-resolution crystal, it is possible that noise images that match these ‘high-resolution’ features will be selected. If taken forward, these ‘particles’ will provide a ‘high-resolution’ structure of the search template, despite having no basis in reality (van Heel, 2013). This problem persists into classification steps. For example if a series of projections from a crystal structure were used as references then it is likely that classes will be obtained that resemble these projections, even if that does not represent the actual data. In addition, if a high resolution starting model is used for a 3D reconstruction it can heavily bias the resulting structure. Thus starting models for initial refinements are heavily filtered to a resolution significantly lower than the expected resolution of the map obtained, so that only size and gross shape information is left, although if the starting model was produced from the data itself then model bias is less of an issue.

The simplest form of validation is the comparison of the classes or raw data to the final reconstruction. If several projections of the map have little resemblance to reference-free classes or the raw data from which it was obtained, then it is likely there is some error in the final reconstruction. Likewise, if many of the classes obtained are not found in the projections, this could indicate either a mixed population, impurity in the sample, or the presence of start model bias.

A tilt pair test is an additional extension of simply comparing classes/particles with projections and is a key tool for validation, particularly for maps at lower resolutions. This is achieved by collecting a micrograph at 0° and again at tilt angle, often ~30° (Henderson et al., 2011). The 0° particle from each tilt-pair is compared to projections of the reconstruction to assign initial angles and then the tilted particle against the corresponding projection with that tilt applied. Over a series of micrographs and particles this agreement can then be plotted and the distribution and clustering of this can provide a validation of the reconstruction. This process has been added as a module in EMAN2 and there is also a web server available (Wasilewski and Rosenthal, 2014). However, this can be difficult for samples with very similar views where assignment can be challenging, although is still worth doing as a valuable form of validation.

The true gold standard of validation is direct comparison to existing high resolution structures of the protein of interest. Indeed it has been argued that rather than comparison of two half-maps a better measure of a 3D reconstruction would be its correlation to the corresponding atomic model, although this is obviously not possible in many cases where either the resolution obtained is simply too low or there is no atomic model for comparison. If the proper precautions against model bias have been undertaken, by strongly filtering any models used, then the appearance of chemically-sensible high resolution features itself can serve as part of the validation. For example, if the initial model contains no information beyond global shape at ~30 Å and the final map shows clear secondary structure elements or side chain detail consistent with typical protein structure, it is likely these features are truly present in the structure. However, the best form of validation still remains common sense, if the structure seems too good to be true, it likely is.

2.6 Equipment and Materials Used

All continuous carbon and lacy carbon grids were purchased from Agar Scientific and all Quantifoil grids purchased from Quantifoil. All negative stain grids were treated under a UV lamp for 20-40 minutes and all cryo grids were glow discharged in air for 10-20 s using either a Cressington 208 or a Pelco EasiGlow immediately prior to use. All cryo grids were prepared using a Vitribot IV (FEI) at nominal 100% humidity and 4°C with a range of blotting conditions dependent upon the sample, typically ranging between a blot time of 3-8 s and a blot force of 3-5, before transfer, under liquid nitrogen (LN_2), to storage dewars prior to imaging.

The following microscopes were used for this project:

Microscope	**Location**	**Voltage (keV)**	**Electron Source**	**Detector**
FEI Tecnai T12	University of Leeds	120	LaB6	Gatan US1000XP CCD
FEI Tecnai F20	University of Leeds	200	FEG	Gatan US4000/SP CCD
FEI Titan Krios	MRC LMB, Cambridge	300	FEG	FEI Falcon II DED
FEI Titan Krios	CAS, Beijing	300	FEG	FEI Falcon II DED

Table 2.1 – Electron microscopes and detectors used during the PhD project.

Software Package	**Function**	**Version(s) used**
Buccaneer	Automated protein chain tracing	CCP4 7.0020
Coot	Graphics program for protein model building	0.8.6
CTFFIND	Estimation of defocus	3, 4

	and CTF parameters	
EMAN	EM image processing suite	1, 2.1
gCTF	GPU accelerated CTF parameter estimation	0.50
MDFF	Flexible fitting of protein structures into EM density	VMD 1.9.1
MOTIONCORR	Whole frame alignment and motion correction	2.0
RELION	Maximum likelihood based EM image processing suite	1.3, 1.4
ResMap	Estimation of local resolution in EM reconstructions	1.1.4
Rosetta	Modelling software suite	2016.13.58602

Table 2.2 – Processing software used during the PhD project.

2.6.1 *General Image Processing Workflow*

A typical processing workflow for a cryoEM project in this thesis consisted of first aligning the raw frames using MOTIONCORR (Li et al., 2013a). CTF determination was then carried out using either CTFFIND3, CTFFIND4 within the RELION GUI or gCTF (Mindell and Grigorieff, 2003; Rohou and Grigorieff, 2015; Zhang, 2016). Particles were then picked either manually or automatically using the boxer function within EMAN2 and coordinates exported as box files. Particles were then inverted, normalised, dust corrected (using typical values of 3 or 5 sigma for removal of both black and white pixels) and extracted within the extract particle module of RELION. 2D classification was then carried out within RELION, initially the number of classes used was ~(total number of particles/200) up to a maximum of ~300 classes. Particles belonging to 'good' classes were selected by eye and used in a consensus 3D auto-refinement within RELION, using a starting model low-pass filtered to ~60 Å. 3D classification in RELION was generally trialled at this point using the consensus structure from auto-refinement as a starting model.

Depending on results of 3D classification, 2D and 3D classification was then iterated until a homogeneous particle stack is found. Following a further 3D auto-refinement of this subset, movie frames were then extracted and the refinement continued utilising these movie frames. The output _data.star file from this process was then used in the particle polishing module of RELION using both linear fit of particle movement and damage weighting. The calculated B-factors from the damage weighting were then plotted against frame number to ensure the B-factors follow the expected damage trend. If the plot is highly inconsistent with expected shape (for example a flat line) then the particle polishing was re-run omitting the damage weighting with frames to be removed specified manually. The 'shiny' particles given as an output of particle polishing were then re-refined within auto-refinement. Finally the output was post processed within RELION. The unfiltered half-maps are masked, using a soft mask expanded by around >9 pixels from the model with the initial binarisation threshold determined by the noise level of the half-maps. The maps were also filtered to the global resolution and have map sharpening applied through either an automatically estimated or ad hoc B-factor in this post processing procedure.

3. V-ATPase Structure and Mechanism

Structural understanding of the V-ATPase complex has progressed rapidly following the widespread adoption of new developments in EM, primarily the use of direct electron detectors (DEDs). At the time the following studies were carried out, there were no sub nm structures of the V-ATPase, only limited structural knowledge of the catalytic states of the full complex and an incomplete understanding of the flexibility inherent within the rotary mechanism. The work in this chapter aimed to address this lack of information through the use of cryo-EM to provide a higher resolution structural model of the V-ATPase as well as understand the role flexibility can play in the mechanism of the V-ATPase.

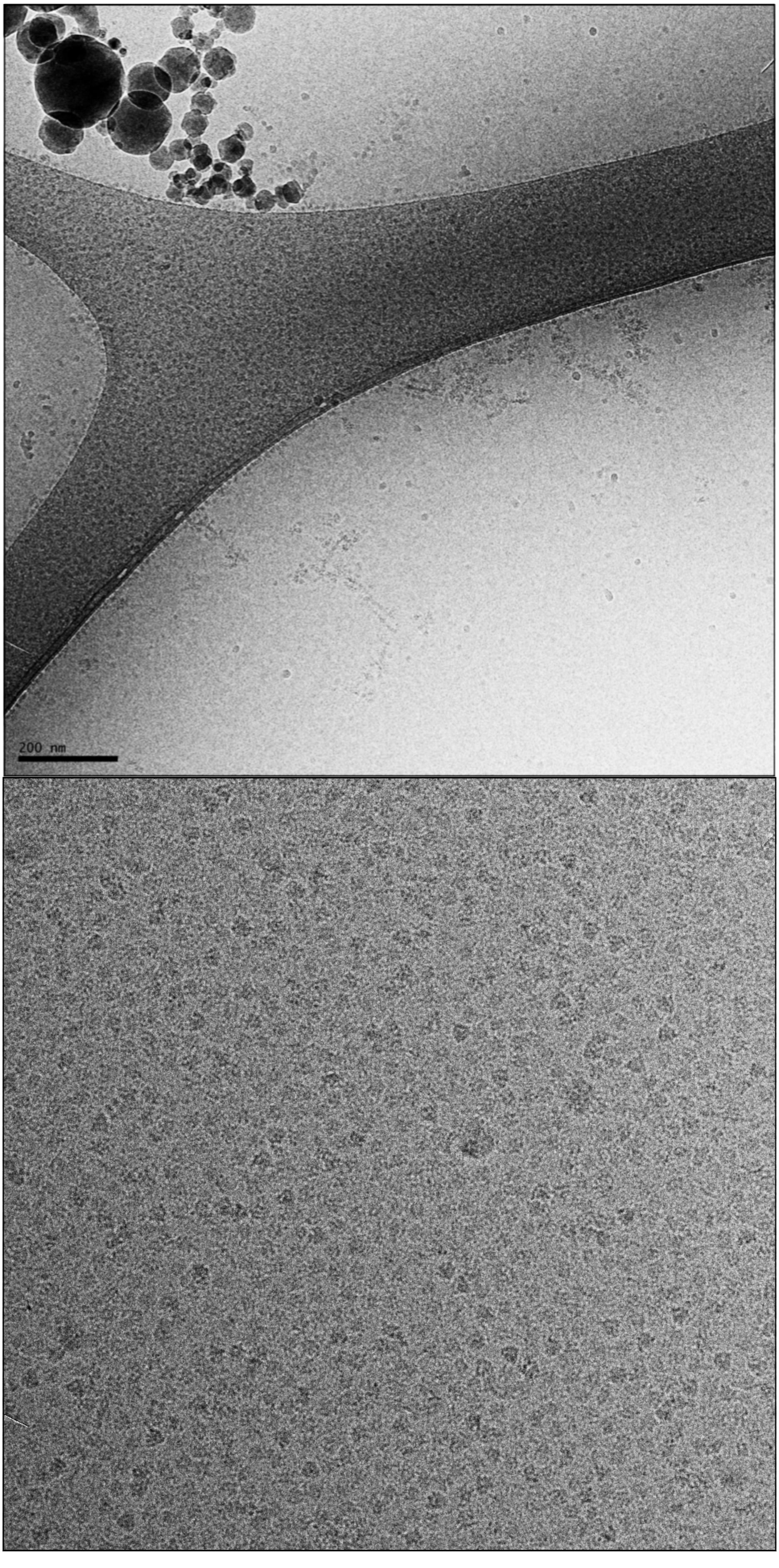
3.1 Results and Discussion

3.1.1 V-ATPase EM Sample Preparation

V-ATPase complex was obtained from *Manduca sexta* (*M. sexta*) midgut and kindly provided by the group of Prof. Helmut Wiezorek, University of Osnabruck, Germany. *M. sexta* was used as the protein source as it has several advantages over the commonly used yeast system; not only has extensive biochemical characterisation been carried out, but as it is a higher eukaryote any structural information obtained may be more applicable to therapeutically relevant homologues. For example, the V-ATPase complex from yeast apparently lacks the *e* subunit which is present in the *M. sexta* sample (Bueler and Rubinstein, 2015), and in the Bovine complex. *M. sexta* V-ATPase is sensitive to some selective V-ATPase inhibitors which do not affect the yeast complex (Muench et al., 2014a). In addition the midgut of *M. sexta* is highly enriched in the V-ATPase complex, allowing large amounts of protein to be purified in a single preparation and by isolating from only a single organ this preparation may be enriched with a single isoform of the complex, which should reduce heterogeneity and allow for higher resolutions to be obtained. However, the use of *M. sexta* complex does present several disadvantages in comparison to the yeast system, chiefly that it is genetically intractable, so potentially valuable mutational studies are not possible in the same way.

3.1.2 Cryo Grid Preparation and Optimisation

Negative stain EM was first used to check sample quality and purity to assess for aggregation or degradation before cryo grid optimisation was carried out. The sample was checked for the distinctive dumbbell shape side view which is commonly observed in negative stain due to a strongly preferred orientation on the carbon film. Thus single spherical particles were attributed to belonging to degraded or disassociated V_1/V_o domains. Screening for grid optimisation was performed on an FEI F20 microscope with a Gatan 4K x 4K CCD camera. Initial attempts at preparing grids suitable for high resolution cryo-EM studies proved unsuccessful with standard lacey carbon grids due to the complex preferentially resting on the carbon film and clumping around the edges of the holes rather than distributing evenly. As previous studies, for example on ribosomes, have shown that a thin carbon layer can be used to improve the particle distribution, a thin (~10 nm) layer of carbon was placed over the quantifoil support film. First carbon was deposited onto a mica surface and this layer then floated onto the grids by inserting the mica sheet at a ~45° angle into a water bath containing the grids to be coated. Using the previous protein concentration of ~1 mg/ml this caused overlapping particles on the grid but acceptable ice thickness. To reduce the crowding a dilution with buffer (20 mM Tris-HCl, 9.6 mM 2-mercaptoethanol, 0.01% $C_{12}E_{10}$, pH 8.1) of ~20x was used to give a reasonable particle distribution (Figure 3.1).


200 nm

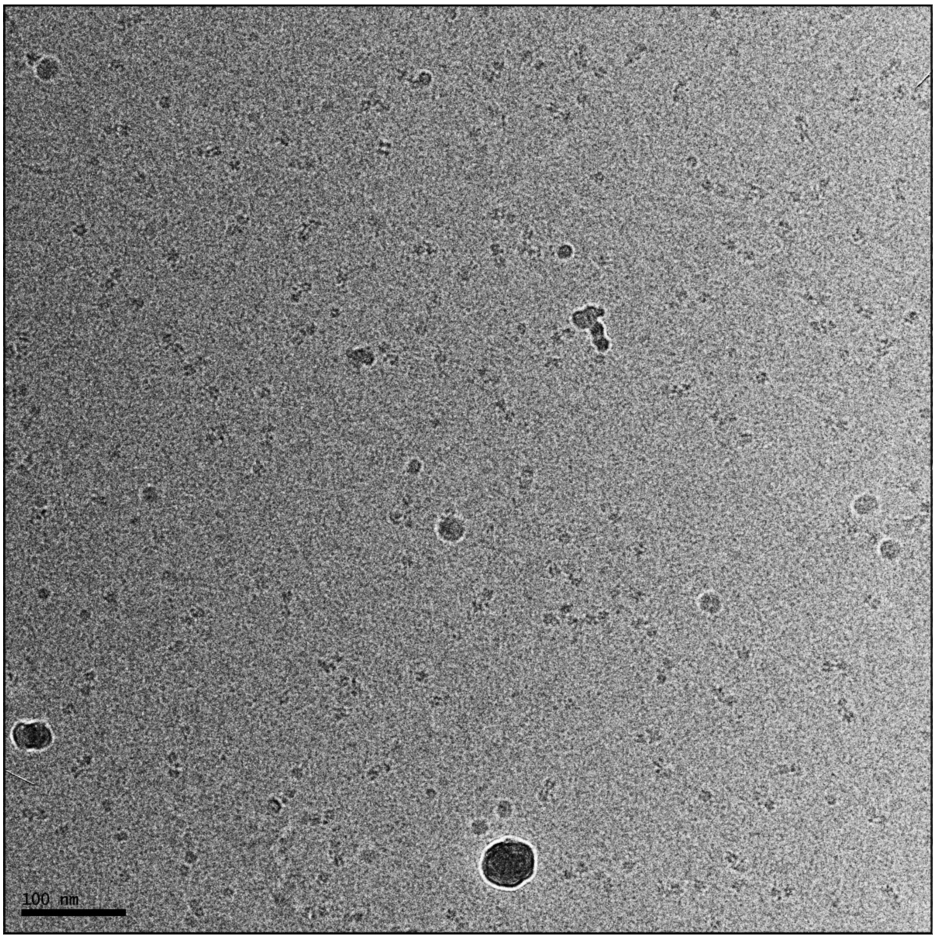
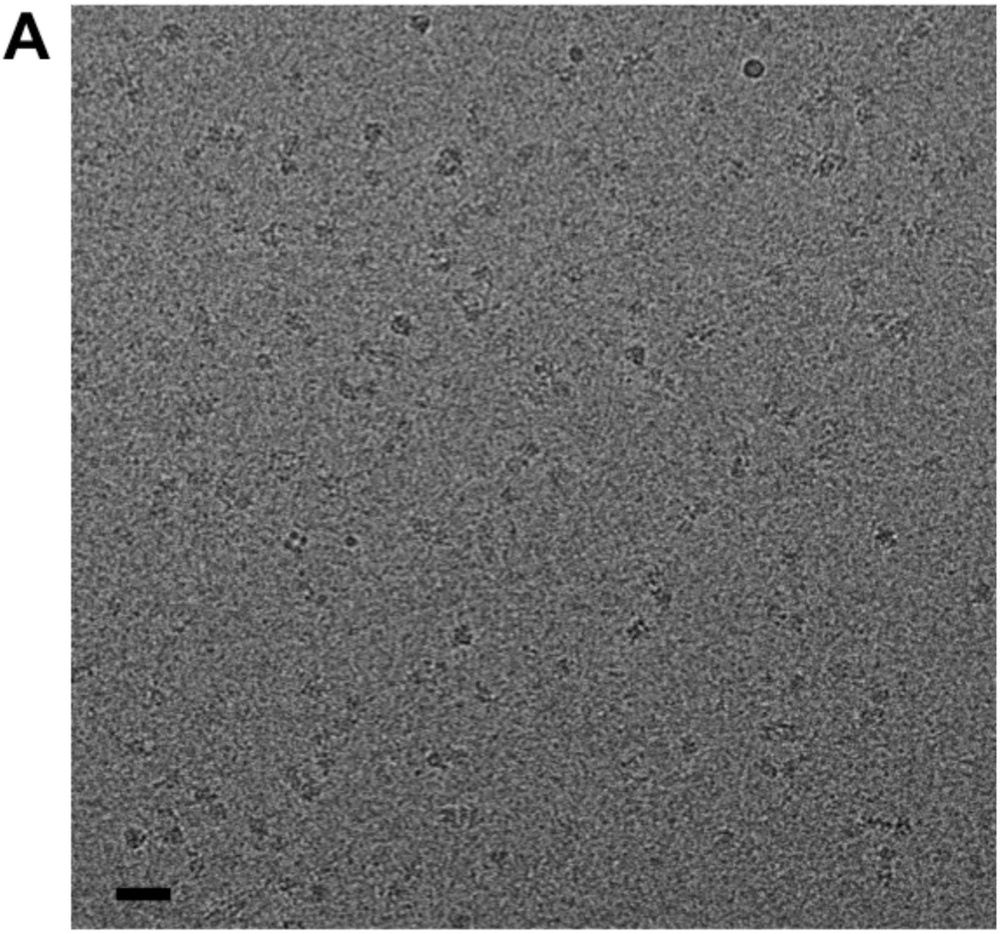


Figure 3.1 – Example V-ATPase micrographs. Example cryo micrographs taken on a CCD equipped F20 on lacey grids (top), carbon coated quantifoil (centre) and carbon coated quantifoil following ~20x dilution (bottom).

High resolution data collection was subsequently carried out on the optimised grids at the MRC-LMB using a FEI Titan Krios equipped with a Falcon II direct electron detector at a sampling of 1.35 Å/pix and a total dose of ~50 e^-/A^2 over a 2 s exposure with 34 frames in each micrograph 'movie'. In total 1366 micrographs were collected using the FEI EPU automated data collection software. Micrographs were collected with acceptable particle distribution and minimal drift or astigmatism observed in the Fast Fourier Transforms (FFT) (Figure 3.2).


A

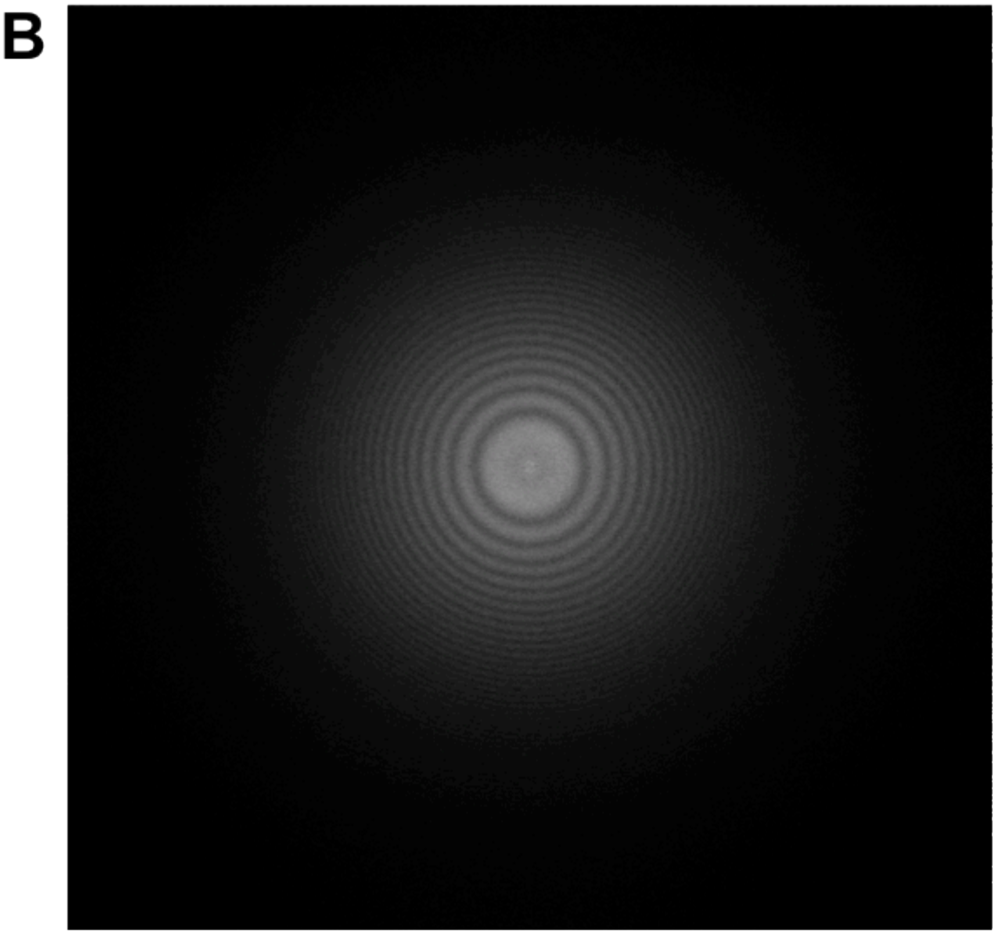

B

Figure 3.2 – Example V-ATPase cryo data. (A) Example V-ATPase micrograph from Krios collected on a Falcon II (Scale bar represents 300 Å). (B) Example FFT showing no drift or astigmatism

3.1.3 Image Processing

Whole frame motion correction was carried out on all micrographs using MOTIONCORR to minimise the effect of any mechanical drift over the course of the exposure and then sum the aligned frames (Li et al., 2013a). Defocus parameters were calculated for the aligned image sums using CTFFIND3; micrographs had a defocus range of 1.7-5.5 μm (Mindell and Grigorieff, 2003). 30730 particles were manually picked in EMAN2 e2boxer before iterative rounds of reference free 2D classification in RELION (Figure 3.3) (Scheres, 2012; 2015; Tang et al., 2007). The classification identified poorly aligned particles which were subsequently removed and also identified a large number of dissociated V_o and V_1 domains, likely due to limited degradation of the sample prior to grid preparation (Figure 3.3). Particles belonging to these classes were removed leaving 7160 particles. Additionally a well defined class was observed with clear 4-fold symmetry which did not correspond to any known V-ATPase view and is a contaminant, potentially alpha-ketoacid dehydrogenase complex which exhibits a very similar view (Wagenknecht et al., 1990).

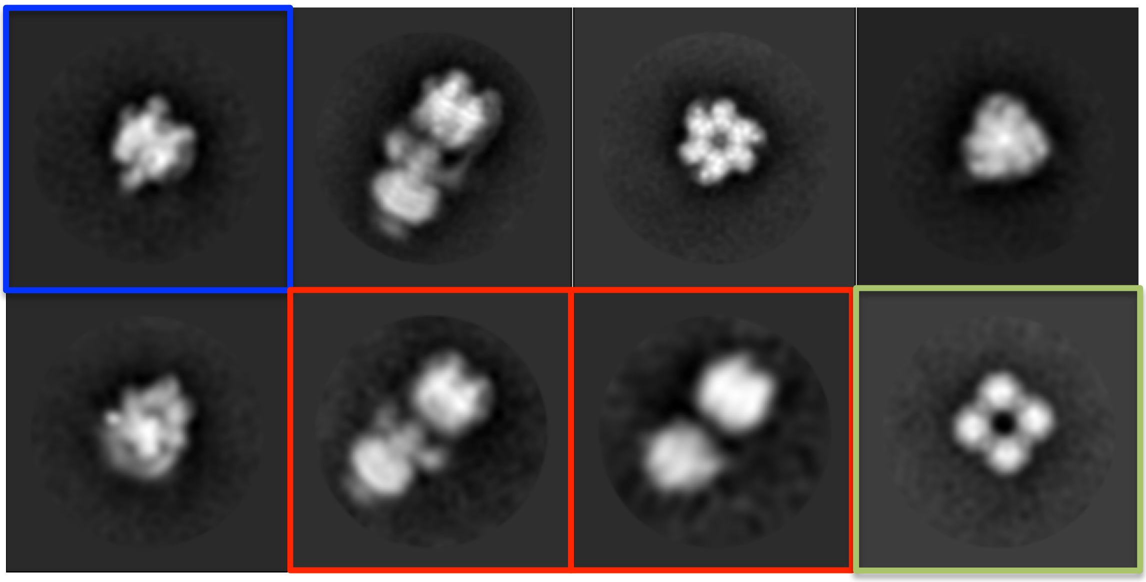


Figure 3.3 – Representative 2D classes from V-ATPase cryo dataset. 2D V-ATPase classes from RELION 2D classification showing top, side and high angle views. Disassociated V_1 side view shown in blue and unknown symmetric contaminant shown in green with poorly aligning 'bad' particles shown in red.

Following 2D classification a 3D auto-refinement was run using a previously published V-ATPase reconstruction filtered to 60 Å as a starting model. This produced a 12 Å resolution reconstruction, which displayed 'typical features' of the V-ATPase and clear deviation from the low pass filtered starting model. By removing particles based on poor LogLikelihood scores to leave 6714 particles the resolution was increased to 11 Å. Finally per particle movie processing was then carried out to account for the effects of beam-induced movement during the exposure, further increasing the resolution to 9.4 Å by gold standard 0.143 FSC (Figure 3.4). 3D classification was attempted to identify a more homogeneous sub-population, but all reconstructions obtained were of lower resolution following refinement, likely due to the low particle numbers. Nevertheless, focussed classifications on the V_1 domain did reveal small sub populations with the open site apparently located in different positions, similar to those observed by Zhao et al in the yeast structures (Zhao et al., 2015). However, due to the small particle number these reconstructions were of lower overall quality and resolution than the consensus structure.

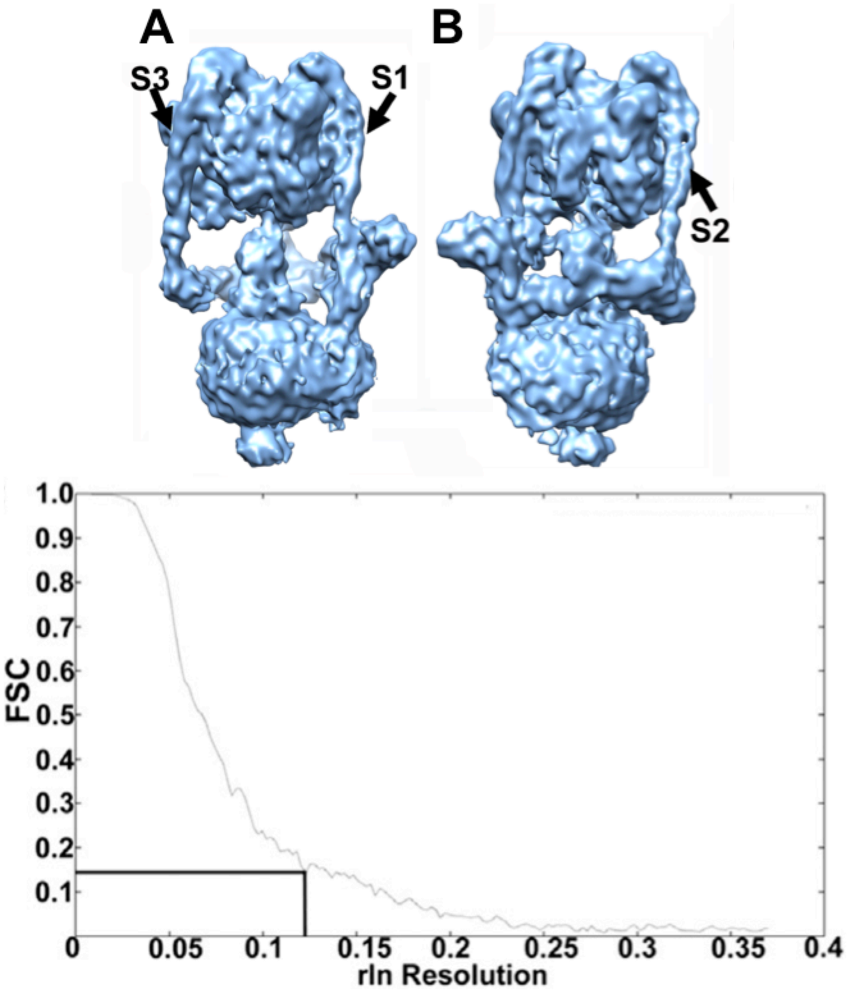


Figure 3.4 – 3D reconstruction of V-ATPase. (A) V-ATPase at 9.4 Å resolution viewed from front. (B) Rear view of V-ATPase reconstruction (C) Fourier Shell Correlation (FSC) curve showing gold standard (0.143) resolution. Figure adapted from (Rawson et al., 2015).

3.1.4 *Flexible Fitting of Crystal Structures*

In order to better understand the detailed structure of the V-ATPase complex, homology models were generated of each individual subunit in Phyre2 (Kelley and Sternberg, 2009; Kelley et al., 2015), with the exception of subunit *a* where insufficient structural information existed at the time of study. These homology models were then used to perform flexible fitting into the whole complex using MDFF (Trabuco et al., 2011) to give an optimal fit and, although the resolution does not permit accurate positioning of side chains, the Cα positions in the secondary structure elements could be approximated with confidence (Figure 3.5). However, due to the large number of subunits within the V-ATPase causing difficulties with the computation, the flexible fitting was carried out in sections consisting broadly of the V_1 domain and the V_o domain and the two halves were then combined.

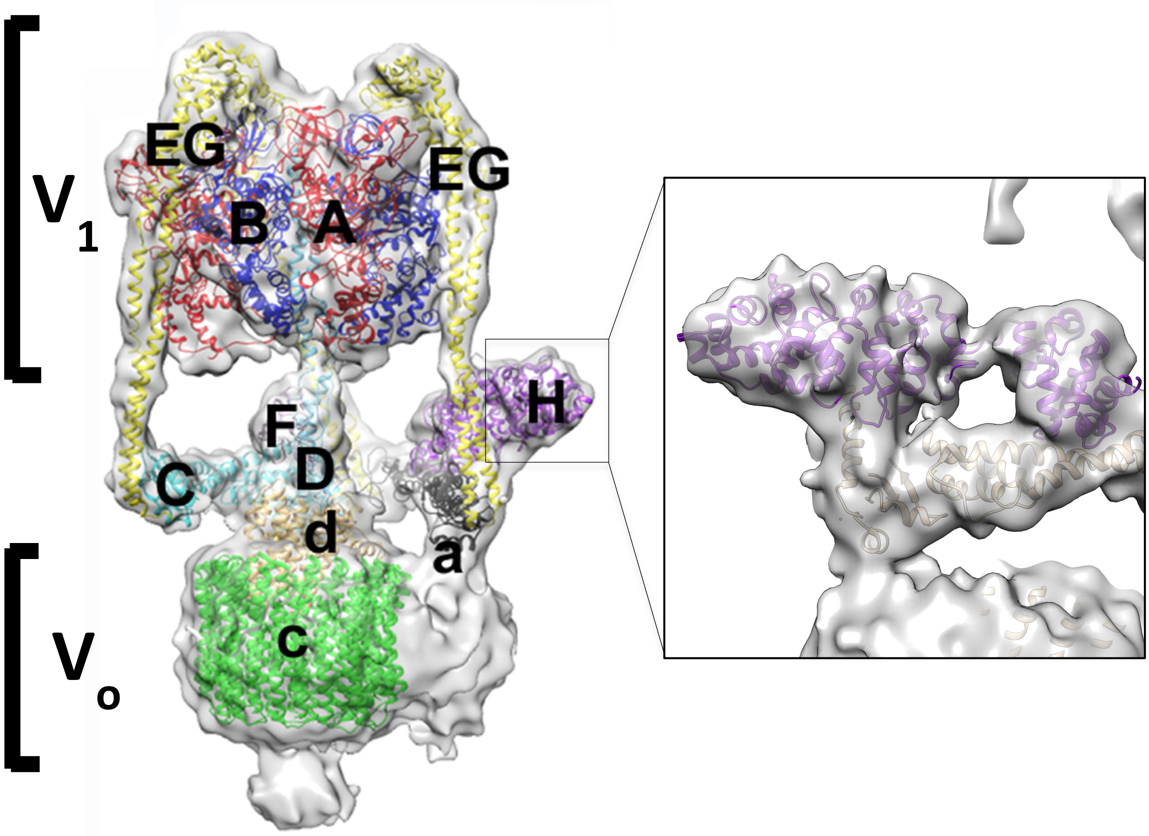


Figure 3.5 – V-ATPase reconstruction with fitted homology models. Labelled *M. sexta* homology models fitted into 9.4 Å V-ATPase reconstruction with MDFF. Inset shows density around H subunit to illustrate quality of fit (Trabuco et al., 2011). Figure adapted from (Rawson et al., 2015).

3.2 Mechanistic insights

3.2.1 Comparison between yeast and M. sexta V-ATPase structures

Analysis of the V_1 domain clearly shows the open AB dimer to be positioned above subunit H within the *M. sexta* reconstruction, with the remaining two sites showing a closed conformation. Importantly, analysis of the ~11 Å yeast V-ATPase reconstruction shows the open site to be situated above subunit C (Benlekbir et al., 2012), 120° rotated from that of *M. sexta*. This has significant implications for understanding the catalytic mechanism of the V-ATPase and rotary ATP family as a whole, as comparisons can be made of the rotary motor as a whole in two defined catalytic states. One significant feature is that despite the two V-ATPase systems being at different stages of their rotary cycle the 3 stator elements are superimposable (Figure 3.6). From the different conformations seen in the stator crystal structures, it has been proposed that the cycling between the open, loosely bound and tightly bound states results in a change in stator conformation, meaning that as the ATPase cycles the stators will accommodate the conformational change within the A/B dimers. Also the stators have been proposed to play a role in energy storage, potentially compensating for any symmetry mismatch between the V_1 and V_o domains.

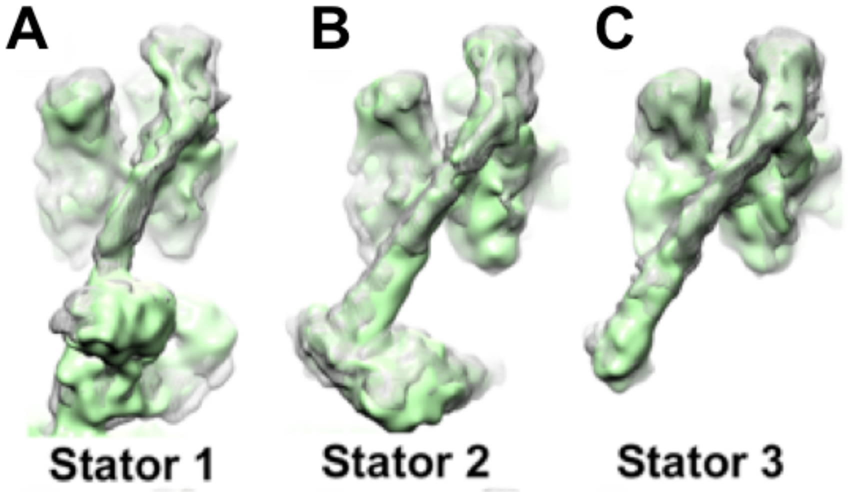


Figure 3.6 – Superposition of stators from yeast and *M. sexta* V-ATPase.

Comparison of stators from yeast V-ATPase (EMDB 2781) in grey mesh and *M. sexta* V-ATPase (EMDB 2781) in green. Figure adapted from (Rawson et al., 2015).

This change in conformation of the stators would account for the ~7° flexing seen directly by EM (see Section 3.2.5), within crystal structures and calculated through MD simulations (Oot et al., 2012; Stewart et al., 2012). As the stator position appears to be unchanged when comparing the *M. sexta* complex to the yeast system despite apparently being in a different catalytic state, it is likely that any deformation of the stators, either for energy storage or to accommodate changes in the AB dimers, are transient. Indeed, this comparison suggests that the occupancy of the A/B dimer does not dictate the stator position or conformation. Thus the conformation of each stator appears to be determined, not by the occupancy of the associated A/B dimer but by the contact it makes with subunits *a*, C and H which make up the collar region. Interestingly upon inspection, subunit H does not interact with both EG subunits, only subunit E. The interface between the H subunit and stator shows a large charged patch with a hydrophobic core flanked by a positive and negative patch of residues; it is tempting to speculate that this interface plays a role in the apparent mobility of subunit H in the isolated V_1 domain (Oot et al., 2016). It is also interesting that when the structures of all three catalytic states from yeast were compared, the authors report small changes in the conformation of the stator network depending upon nucleotide occupancy. This discrepancy is likely to be a result of the lower resolution of the *M. sexta* study, or could potentially be a thought-provoking species difference, highlighting the dangers of extrapolating across species.

3.2.2 Apparent linkage between subunits d and C

Key for the rotary mechanism is the need for the central rotor axle to couple to the *c*-ring with no interactions with the static stator complex. However, within both the yeast and *M. sexta* systems this is not the case with significant contact made between subunits *d* and C, observed both directly in the EM reconstructions and in the atomic surfaces of fitted atomic models (Benlekbir et al., 2012; Rawson et al., 2015; Zhao et al., 2015) (Figure 3.7). The removal of this linkage has been shown to increase the flexibility of the system and it was hypothesised that this may act in a ratchet like mechanism (Richardson et al., 2014). In the absence of any ATP hydrolysis this 'catch' prevents the axle rotating either forwards or in reverse, thus silencing proton leakage and preventing futile backwards rotation of the rotor.

Importantly, both yeast and *M. sexta* V-ATPase complexes show an apparent 'resting state', which based on particle number is the predominant population (~47.5% within the yeast system (Zhao et al., 2015)). Interestingly, this coincides with the level of interaction between subunit *d* and subunits C and H, with the largest to smallest interfaces having approximately 47.5%, 36% and 16.5% of particles associated with them, respectively. This is suggestive that the interface may potentially provide stability within the complex, although the exact purpose is unknown. One hypothesis could be that this asymmetry in population is related to regulation, whereby ATP induces dissociation of the complex (Huss and Wieczorek, 2007), so it is possible that the detachment of V_1 from V_o can only occur in certain rotational states and thus is mediated by this interface. Indeed, X-ray studies of the isolated V_1 domain which have show this domain being consistent with 'state 2' from the intact complex.

Figure 3.7 – Subunit *d* interactions. Atomic surfaces from yeast V-ATPase subunits *d* (orange), C (green) and H (pink) showing differences in interfaces depending on catalytic state. Figure taken from (Rawson et al., 2016b).

3.2.3 *Electrostatic bearing region*

The mechanism of catalytic cycling within the F/A-ATPase has been well characterised through crystal structures of differing states and biophysical analysis. However, at the time this work was carried out no crystal structure had been reported for a Eukaryotic V_1 domain, only that of the related bacterial Enterococcus hirae (*E. hirae*) A/V-ATPase (Arai et al., 2013). A large degree of structural conservation is found between the A and V-ATPase ATP hydrolysing AB domains especially within those regions directly involved in catalysis (Muench et al., 2011). Following flexible fitting of the homology models using MDFF, the AB domains were examined. Superposition of the fitted AB structures showed close similarity towards the top of each subunit but a large difference at the base, consistent with crystal studies on the open and closed states. This change is focussed within the area equivalent to the DELSEED loops of the F-ATPase (Watanabe et al., 2015). Interestingly, a slice through of the V_1 domain shows that the central rotor axle passes through the central cavity in an asymmetric manner making two points of contact with the AB domains. The first point of contact is at the 'DELSEED' loops, which is expected due to their role in applying torque to the rotor axle as part of the mechanical function of the complex. However, a second close contact is formed roughly two thirds through the structure of the V_1 domain, interacting with both the A and B subunits (Figure 3.8). The region of interaction is in a highly conserved region with a P-A/G-D/E-X-G (subunit A) and P-G/S-R-R/K-G-Y/F (subunit B) motif. Analysis of 300 A/V-ATPase sequences within Consurf (Goldenberg et al., 2009) shows full conservation of Arg and 95% conservation of Arg/Lys within subunit B with full sequence conservation for a negatively charged D/E in subunit A. Moreover these loops form a tight hairpin structure which is only permitted by the conserved Pro and Ala/Gly that flank the motif. This motif is fully conserved between the A- and V-ATPase family and results in a highly polar collar of density which alternates between a positive and negative charge. The charge faces the highly polar central rotor axle with the highly conserved nature potentially suggesting that it performs an important role in ATPase function. Interestingly this region is also highly conserved in other rotary motors, including the bacterial flagella motor, and the F-ATPase which may suggest a more general role for this bearing motif.

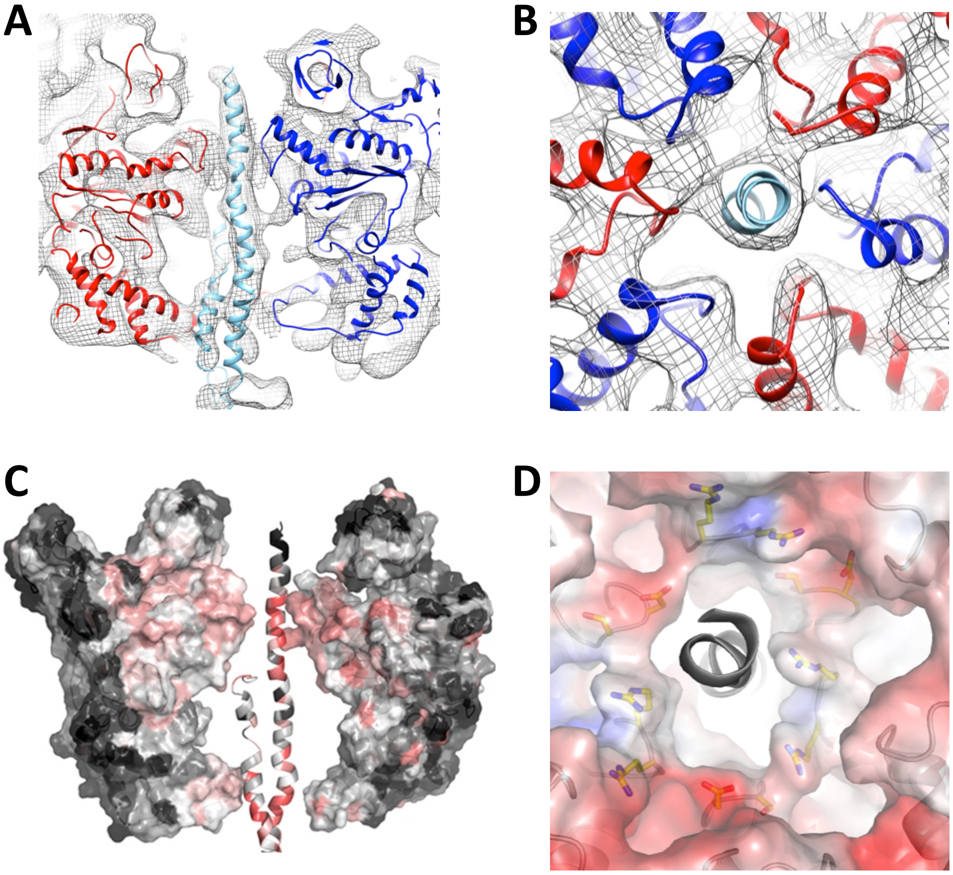


Figure 3.8 – Electrostatic bearing region in *M. sexta* V-ATPase V_1 domain. (A) Section through V_1 domain showing fitted subunits A (blue) B (red) and D (cyan). (B) Horizontal section of V_1 at bearing region. (C) Sequence conservation calculated in Consurf ranging from high (pink) to low (black) conservation. (D) Electrostatic surfaces calculated in APBS from -5.0 (red) to 5.0 (blue). Figure adapted from (Rawson et al., 2015)

The loops regions show little movement in the open and closed states in the *E. hirae* crystal structure, suggesting the loop regions do not operate in manner similar to the DELSEED loops. While this feature was identified in the F-ATPase complex by Abrahams et al., it was described as a hydrophobic collar region and the charged characteristics of these loops were not studied

(Abrahams et al., 1994). A potential explanation of the alternating charge is to form opposing interactions with the central axle allowing it to 'hover' within the collar structure, reducing drag on the central axle. Thus it is tempting to speculate that this bearing may play some role in the extraordinary efficiency of the rotary ATPases (Junge et al., 2009). It is also interesting to note that this bearing region coincides with the highest resolution elements identified in the reconstruction by local resolution assessment in ResMap (Kucukelbir et al., 2013), indicating its stability within this dynamic motor domain.

3.2.4 d/c-ring interface and subunit e/Ac45

Another poorly understood feature of the A/V-ATPase family is the ability to accommodate different stoichiometries of *c*-ring, using the same D/*d* subunits and maintaining the same overall architecture. This accommodation can be achieved through a striking feature in the reconstruction where the contact made between subunit *d* and the *c*-ring, 'socket' is minimal despite this being an interface for torque transmission. By only interfacing with a small surface area within the *c*-ring, rather than plugging the hole in the centre of the *c*-ring, the *d* subunit will be able to interface with a range of differing *c*-ring stoichiometries, rather like a guide wheel (Figure 3.9). This would explain how the A/V-ATPase can adapt to differing symmetries within the *c*-ring without the need to change the *d* subunit architecture (Meier et al., 2005; Pogoryelov et al., 2005; Vollmar et al., 2009). This feature is consistent with the position of the *d* subunit in yeast where its position in all three catalytic states can be observed. However, with this highly tilted contact a large hole is visible into the centre of the *c*-ring, which would need to be plugged in some manner to prevent free diffusion of protons and other ions across the membrane. While some studies have suggested that lipid plays this role this is a potentially interesting area of further research, with new technologies such as styrene maleic acid lipid particles (SMALPs) potentially allowing the complex to be imaged in the native lipid environment and lipidomics experiments to be performed on the same sample (Prabudiansyah et al., 2015).

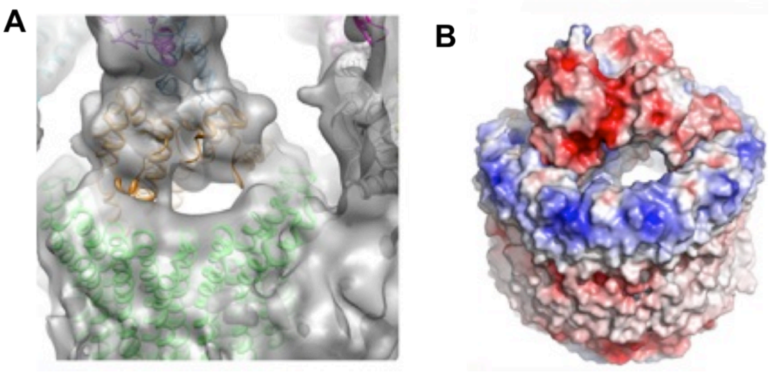


Figure 3.9 – Subunit *d/c*-ring interface. (A) EM density showing *d/c*-ring interface. (B) Electrostatic surfaces of interface showing areas of complementary charge. Figure adapted from (Rawson et al., 2015).

Another prominent feature observed in the V_o domain is an additional density at the base of the complex, present within the *M. sexta* sample but not within the yeast reconstructions (Figure 3.10). This protrusion is shown to be glycosylated and following treatment of the complex with deglycosylation enzymes this density reduces (Rawson et al., 2015). Previous studies have speculated that this density is due to the accessory protein ac45, although recent studies suggest that this is not the case as extensive mass spectrometry analysis has been unable to identify ac45 within the purified *M. sexta* sample (Smith et al., 2016). Alternatively subunit *e* has been suggested to make up this density. Subunit *e* has been suggested to be vital for the formation of the *c*-ring and has been shown to be heavily glycosylated and has been identified as part of the purified *M. sexta* complex (Compton et al., 2006; Ludwig et al., 1998; Sambade and Kane, 2004). However, biochemical studies have suggested that it is not present in the yeast complex and is not required for function (Bueler and Rubinstein, 2015), explaining the absence of this additional density from the yeast structure. Low resolution EM inhibitor binding studies discussed in detail in Chapter 4.2 have subsequently supported this theory.

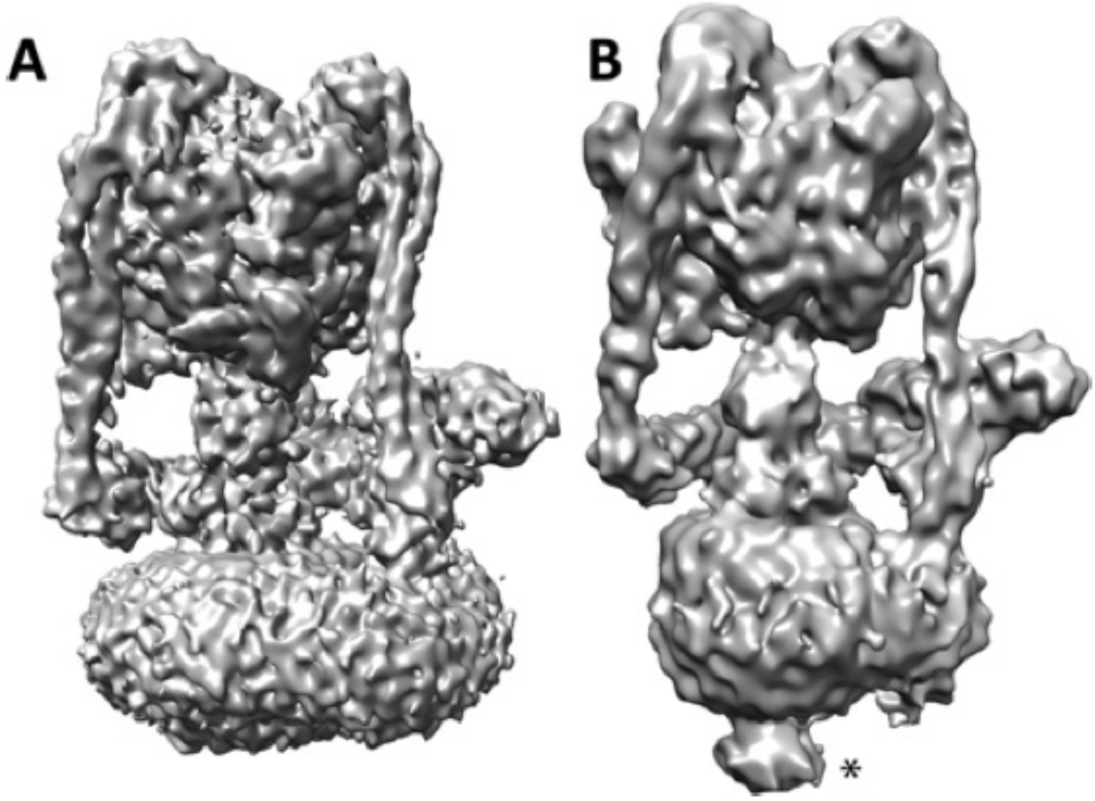


Figure 3.10 – Cryo-EM reconstructions of yeast and *M. sexta* V-ATPase. (A) Yeast V-ATPase reconstruction at ~7 Å. (B) *M. sexta* reconstruction with heavily glycosylated protrusion at base denoted by *. Figure taken from (Rawson et al., 2016b).

3.2.5 Effect of Flexibility

Despite the use of a Titan Krios equipped with a DED, the resolution of the V-ATPase reconstruction was only a modest 9.4 Å. While this is likely in part due to the relatively small dataset collected and the sub-optimal thickness of the additional carbon film, resolution estimates of individual micrographs in gCTF (Zhang, 2016) suggested that the quality of the data itself was high, with the majority of micrographs estimated to have Thon rings extending beyond 3 Å

(Figure 3.11). Another potential factor limiting resolution could be sample heterogeneity as we were unable to separate the conformational states corresponding to the 3 stages of the catalytic cycle. Furthermore several previous studies had suggested that the V-ATPase exhibits a degree of continuous flexibility in addition to the discrete catalytic states.

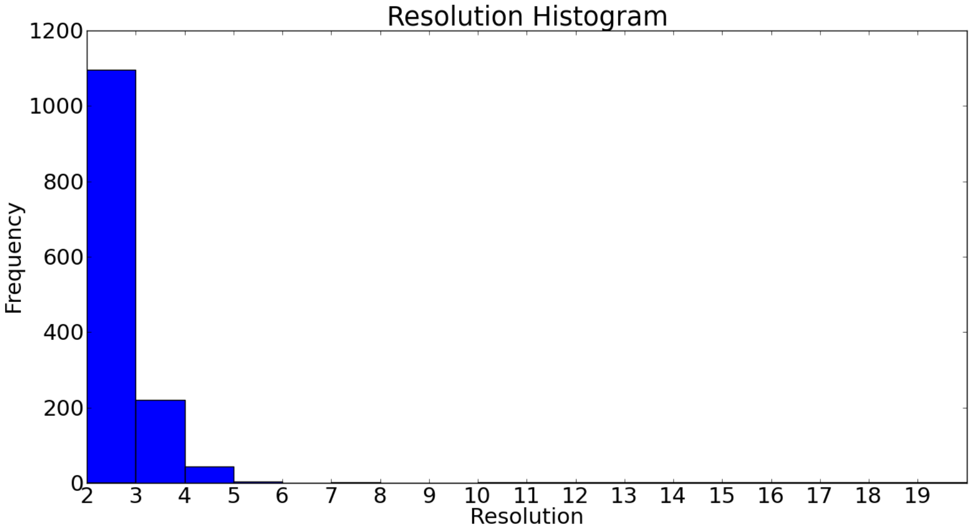


Figure 3.11 – Histogram of estimated micrograph resolution. Histogram of resolution of each V-ATPase micrograph estimated by gCTF (Zhang, 2016).

In order to attempt to quantify this flexibility, prior to my involvement in the project Dr Chun Feng Song used negative stain EM to assess the level of flexibility in the yeast V-ATPase complex. Negative stain was used as it was hoped it would simplify the flexibility analysis, as the complex is forced to lie flat on the carbon film, and may induce more strained states due to the staining procedure, which while potentially introducing artefacts may also allow lowly populated flexed states to be observed. In addition it was hoped the higher contrast of negative stain would aid in the processing. Following 2D classification in IMAGIC-5 of negatively stained V-ATPase analysis (van Heel et al., 1996), through first alignment and classification of the full complex to remove poorly aligning or degraded particles followed by alignment and masking of each domain individually, showed a large degree of flexibility within the sample with flexing up to ~30° observed. While it is possible the staining itself induced some of this observed flexing through flattening the complex onto the carbon film, a similar feature was observed in cryo-EM classes from the same sample (**Figure 3.12**). I then repeated this work using a larger, circular mask alongside the single domain mass to ensure that the observed flexing was not an artefact of masks used during the data processing. The same feature was observed with the same magnitude of flexing seen in classes processed using all masks trialled.

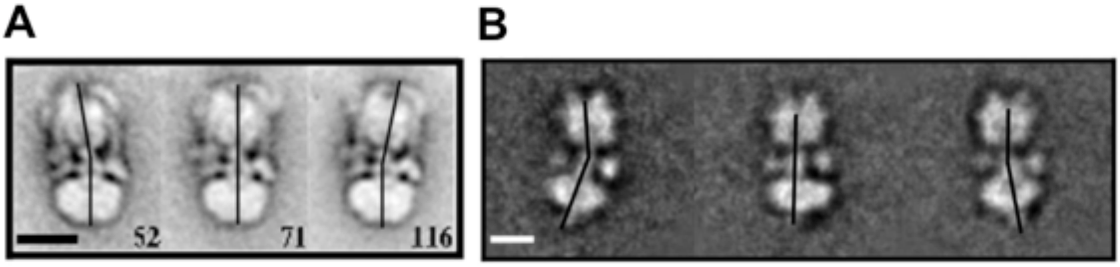


Figure 3.12 – V-ATPase 2D flexibility analysis. (A) Negative stain Yeast V-ATPase classes (B) Single particle cryo-EM analysis of the *M. sexta* V-ATPase. Figure adapted from (Song et al., 2013).

I then investigated the effect of ATP on the flexibility of the V-ATPase complex by collecting two additional data sets in the presence or absence of 5mM Mg.ATP. All data were collected at the same magnification and processed in the same manner, using the same alignment references and mask to ensure that any differences seen are related to the addition of ATP and not data processing artefacts. The resulting classes for the non-ATP sample showed the same classes as generated previously for the *M. sexta* sample with a number of classes showing flexion to a maximum angle of 30°. Interestingly, the addition of ATP resulted in the loss of classes that display extreme flexing of V_1 relative to V_o with only classes displaying a maximum flexion of 10° being seen (Figure 3.13.B). Approximately 10-20% of ATP-treated particles showed this degree of flexion, comparable to that seen in particles without substrate, although this is only from a single experiement.

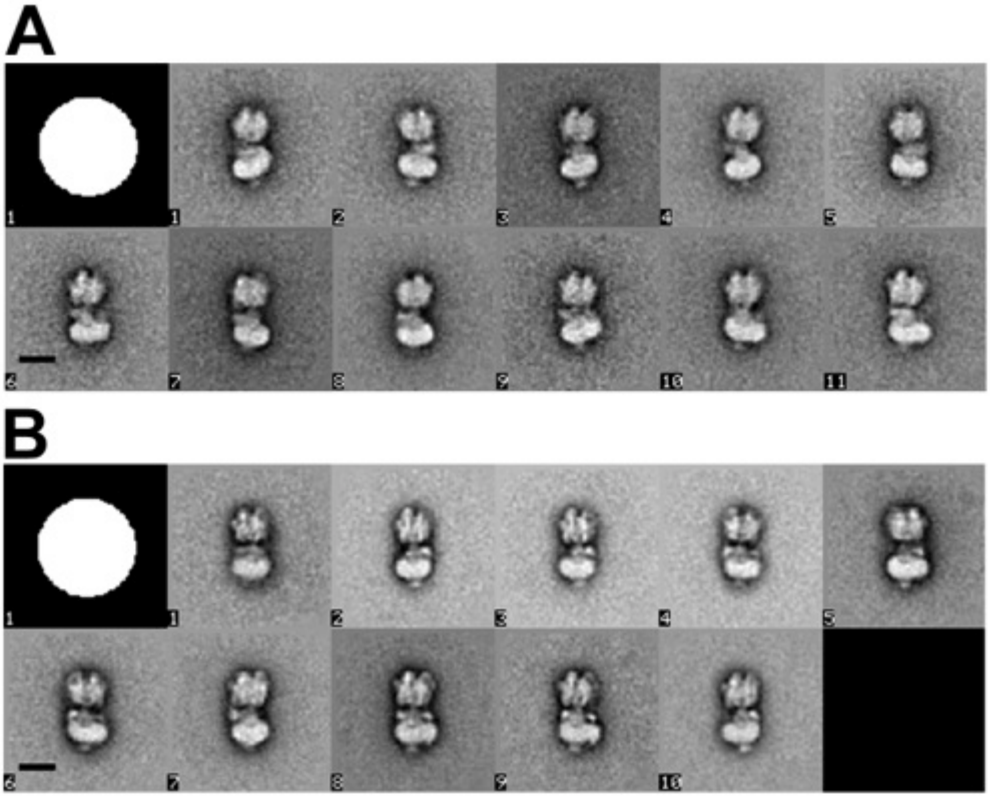


Figure 3.13 - The influence of ATP on V-ATPase flexibility. Classes showing flexing between V_1 and V_o in the absence (A) and presence (B) of 5 mM ATP. All data were aligned and classified in IMAGIC-5, using the circular mask shown in the top right corner. The scale bar represents 150 Å. Figure taken from (Song et al., 2013).

It is also important to consider the role that this flexing motion may play in the regulatory mechanism of the V-ATPase. Both the yeast and *M. sexta* V-ATPase have been shown to be regulated through controlled dissociation, whereby the V_1 domain separates from the V_o domain through a series of currently unresolved structural changes. Importantly, the addition of Mg.ATP substrate does not increase either the range of the flexion or the proportion of particles that are flexed as may have been expected. Instead, it limits the flexibility to a maximum of ~10°, the angle that is most consistent with the proposed bending and observed variations in crystal structures of the subunit E/G heterodimer (Oot et al., 2012). This more subtle movement is also consistent with the predicted flexibility within the V-ATPase from MD simulations (Stewart et al., 2012). Ion mobility mass spectrometry experiments have also subsequently suggested that addition of ATP can reduce conformational flexibility within the A-ATPase (Zhou et al., 2014). The larger bending motion is more consistent with the large angle flexing seen in the early stage proposed to immediately precede disassembly of the *Thermus thermophilus* A-ATPase (Tani et al., 2013). However, it is important to note that the disassembly in the A-ATPase study was triggered through large changes in temperature or pH and may not be representative of the controlled process of disassociation observed following starvation conditions.

As ATP is required for disassociation (Huss and Wieczorek, 2007), the larger bending motion could be due to particles that are primed for disassembly but are unable to disassociate due to the lack of ATP. This has important implications for crystallographic or further cryo-EM experiments on the rotary ATPase family as priming the sample with ATP may allow for a more homogeneous sample, allowing higher resolution structures to be obtained.

3.2.5.1 3D flexibility and masking

The highly dynamic V-ATPase mechanism is likely to involve some structural flexibility and there is evidence for this in variability between 2D image classes from EM as shown in Section 3.2.5, and it is also supported by molecular dynamics (Song et al., 2013). Compared to a series of discrete states, the continuum nature of flexing makes sorting of the heterogeneity difficult by conventional 3D classification. As an alternative, masking based on V_1 and V_o was investigated to attempt to improve the resolution of the resulting models and to asses to what extent this continuous flexibility was impacting the obtainable resolution.

2D classification of the previously obtained cryo-EM dataset (see 3.1.2) was carried out again in RELION, resulting in 13,083 particles in well defined classes. The 3D reconstruction had a tighter mask than previously used for these data, which resulted in a global resolution of 9.2 Å, but with a clear difference in detail between V_1 and V_o, as previously described with helices well resolved in the V_1 domain, but no secondary structure observed in the V_o region. 3D refinement was then carried out separately with the same data stack, focussing on V_1 and V_o through performing local angular searches based around previously determined Euler angles and imposing domain-specific masks in RELION (Figure 3.14). Accounting for this flexibility resulted in a significant increase in V_1 resolution (8.2 Å), with no apparent improvement to V_o (~1 nm).

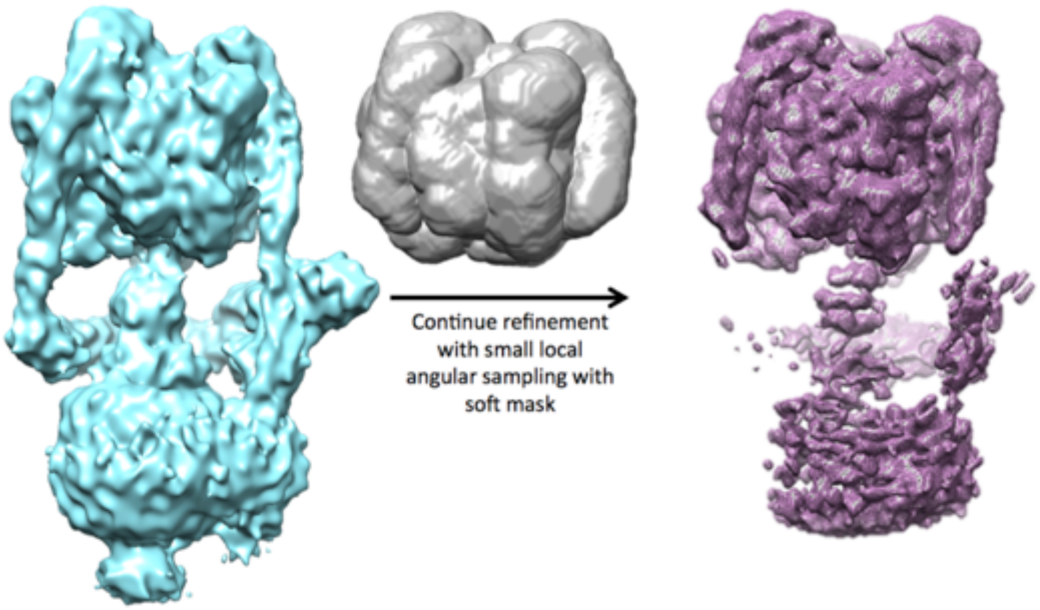


Figure 3.14 – Schematic of localised masking procedure. Refinement of original 9.4 Å reconstruction (cyan) is carried on with local angular sampling and a tight mask imposed (grey) to produce a reconstruction taking account of flexibility (purple). Figure taken from (Rawson et al., 2016c).

Analysis of the masked V_1 shows an improved resolution of 8.2 Å, with local resolution determination using ResMap (Kucukelbir et al., 2013) showing a range from 5 Å to 12 Å, with a significant proportion <7 Å (Figure 3.15). The base of V_1 shows clear asymmetry with each A/B pair making a different interaction with the central rotor axle.

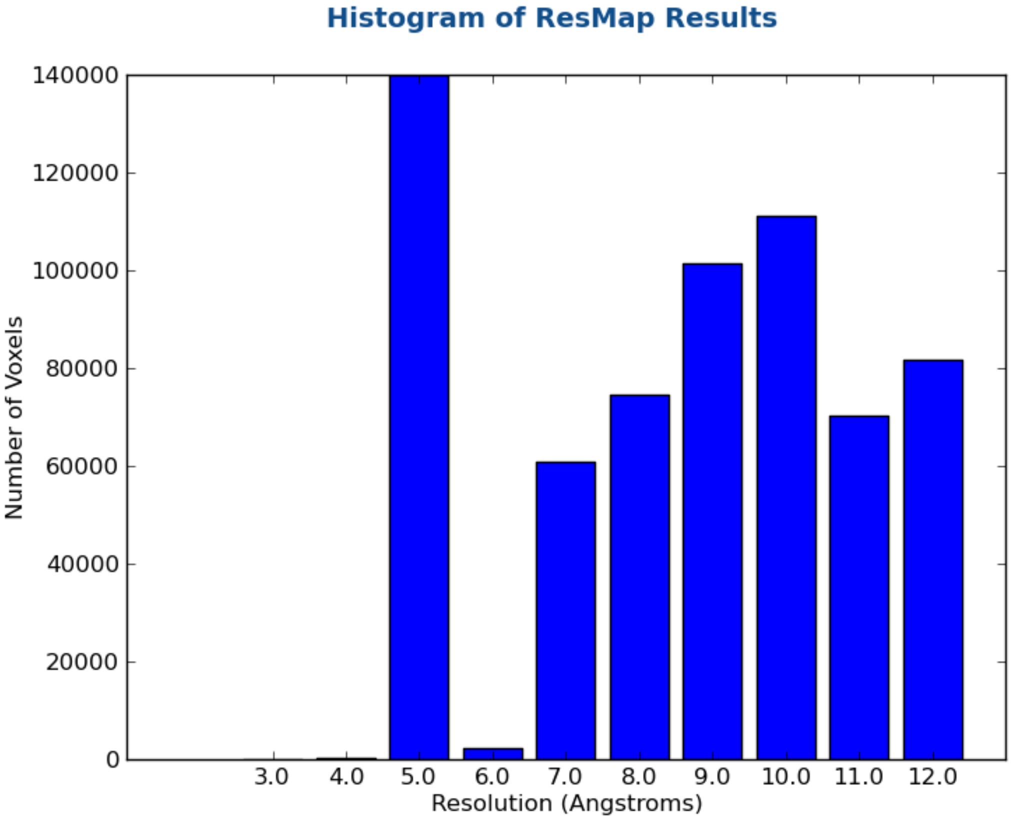


Figure 3.15 – ResMap Histogram of masked V$_1$. Histogram generated by ResMap showing estimated resolutions of masked V$_1$ domain showing large regions of < 7 Å voxels.

The stators show a clear separation of the two α-helices in the masked V$_1$, in contrast to the reconstruction obtained masking the whole complex, while the highest resolution feature observed was the electrostatic bearing region previously identified, with this high stability perhaps suggesting a functional role (Figure 3.16). However, the base of the stators becomes poorly resolved, as does the central rotor axle and V$_o$ domain when the alignment and reconstruction is based on the V$_1$ domain, which indicates flexing between the two domains. Masking of the central region alone resulted in a low quality reconstruction, likely due to its small volume.

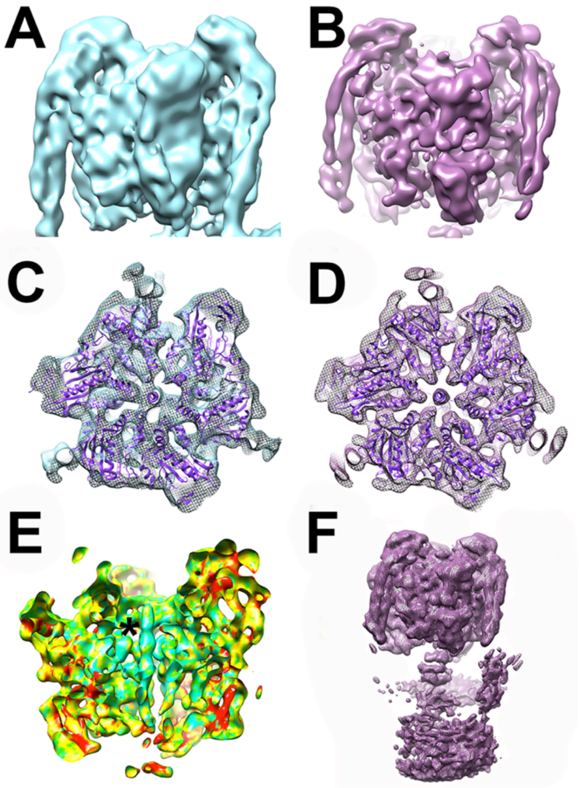


Figure 3.16 – Effect of localised masking on V_1 reconstruction. (A & B) Density of V_1 domain before (cyan) and after (purple) masked refinement. (C & D) Slice through V_1 density showing improved resolution in core region and improved fit of model. (E) Local resolution of V_1 domain following masked refinement showing highest resolution on core with * denoting electrostatic bearing region. (F) Full V-ATPase reconstruction following masking showing V_o smearing due to movement. Figure taken from (Rawson et al., 2016c).

Similar masking of V_o had a negligible effect on the resolution obtained (Figure 3.17). This could be limited by two factors: firstly heterogeneity within V_o caused by rocking of the *c*-ring, relative to subunit *a*; alternatively, the smaller size of V_o and the presence of bound detergent makes alignment more challenging.

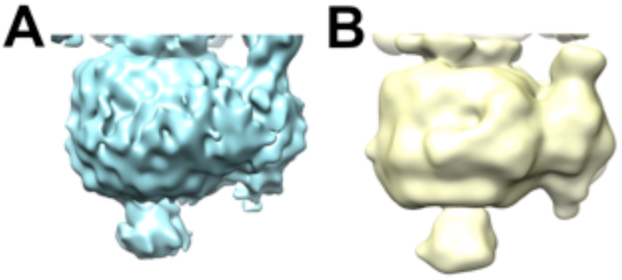


Figure 3.17 – Effect of localised refinement on V_o domain. (A) V_o domain before localised refinement. (B) V_o domain following masking procedure showing no improvement in resolution.

Subtle heterogeneity seen in the *M. sexta* V-ATPase images can be treated by masking, resulting in significant resolution improvement. Architectural and mechanical similarities across the rotary ATPase family suggest that flexibility may also account for resolution differences between the motor and pump domains of other systems. The use of masking to identify different conformational states in RELION has previously been reported for other systems such as the ribosome (Bai et al., 2015b). This work used a similar procedure to account for continuous rather than discrete motions in the rotary ATPase, which may prove a powerful technique in conjunction with the new technologies and conformational sorting to account for discrete heterogeneity within a sample.

Flexion of ~7° about the long axis of the complex has been measured, but the precise reasons for this remain uncertain. One possibility is that it may be linked to the symmetry mismatch between the 3-fold symmetric ATP motor and the variable sized *c*-ring, necessitating some energy buffering between ATPase and proton pump by a flexible rod or spring element. The stator filament network could fulfil this function, although recent cryo-EM 3D reconstructions show relatively little of the flexing in the different catalytic states of the structure. The stator network appears to be relatively rigid, with wobble in the motor resulting from conformational changes in the ATPase during catalysis being accommodated by flexing of subunits C and *a* working as a 'torsion spring' and within the rotor, or by bending at one of the subunit couplings.

This flexibility is further shown when other V-ATPase datasets are examined. For example, in the yeast V-ATPase reconstructions published by Zhao and co-workers over 250k particles were originally identified as V-ATPase but only 50503, 38347, and 17595 particles were assigned to the final reconstructions (Zhao et al., 2015). While many of the particles that were removed are likely to have been contaminant or degraded, it is possible that the majority of these data represented the complex in the continuum of flexed states. In order to attain the side chain resolution now almost routine within the cryo-EM community it may be necessary to mitigate for this flexibility either through addition of additives including ATP which may give a more rigid sample of through finding more amenable (rigid) systems. Though this raises the question of the value of static structural snapshots of what is obviously a highly dynamic complex.

3.3 Conclusions

The V-ATPase is a dynamic molecular motor, with significant flexibility and movement inherent in its function. This work has shown several important features of the V-ATPase mechanism through the production of the first sub nm V-ATPase reconstruction via cryo-EM. This reconstruction allowed the fitting of homology models for the complete *M. sexta* V-ATPase apart from the key subunit *a*, involved in proton translocation.

One feature observed following this fitting is an electrostatic collar region in the V_1 domain, consisting of alternating positive and negative charges in a ring around the central rotor axle, which is also the highest resolution region of the complex, suggesting a highly stable, functionally important region. While a contact between the AB subunits and subunit D in this region had been observed previously in the F-ATPase, it was termed a hydrophobic region and largely ignored (Abrahams et al., 1994). However, this charged bearing is shown to be highly conserved across a large number of species, suggesting it may play a functional role within the complex.

While other studies have focussed on classifying the various catalytic states of the V-ATPase, little attention has been paid to the role of continuous flexibility in the mechanism of the complex. Through negative stain analysis this work shows that a large amount of flexing is present in a sample of V-ATPase. While intuitively the level of flexing may be expected to increase when ATP is added and the complex is functioning, when ATP is added only a small amount of flexing remains. This is consistent with the theory that ATP is required for dissociation and may prove useful if the addition of ATP to ATPase samples prior to structural studies this may reduce the flexibility, producing a more homogeneous sample, potentially leading to a higher resolution reconstruction.

Flexibility was also examined in 3D cryo reconstructions and partially mitigated through a localised refinement procedure, enabling the continuum of flexibility to be accounted for resulting in improved resolution in the V_1 domain. In combination with larger datasets to enable conformational sorting this provides another tool to increase the resolution of EM reconstructions and may be particularly powerful in cases where continuous flexibility is observed rather than a series of defined states. An alternate method for examining this continuum of states would be the manifold embedding technique pioneered in the Frank group which would allow this continuous motion to be represented as a trajectory through the energy landscape and give further insight into the mechanical properties of the system (Dashti et al., 2014; Frank and Ourmazd, 2016).

Despite the findings discussed here and subsequent structural studies by several other groups, including a number of ~6 Å cryo-EM structures of both the V-ATPase and related F- and A-ATPases as well as determination of the architecture of subunit *a* through crystallography and EM studies, there are still many unresolved questions in the V-ATPase field. Key among these questions is the precise arrangement of subunit *a* from a high resolution structure, which would help to elucidate the mechanism of proton translocation. It is also interesting to note the highly unusual horizontal architecture of the *a* subunit, and how this differs from the vertical 4-helical bundle model previously proposed and observed in the modest resolution yeast structures (Lau and Rubinstein, 2012), though several subsequent studies suggest that this horizontal feature is ubiquitous across the rotary ATPase family (Allegretti et al., 2015; Morales-Rios et al., 2015; Zhao et al., 2015; Zhou et al., 2015). In addition there are several questions still remaining regarding the complex mechanisms of regulation within the complex, particularly how ATP silencing in the isolated V_1 domain is achieved following disassociation, although a crystal structure of the intact V_1 domain from yeast now suggests that this silencing is caused through a structural rearrangement of the H subunit (Oot et al., 2016).

This structural work acts to underpin further studies into the mechanisms of regulation of the V-ATPase and also forms the starting point for studies into V-ATPase inhibition discussed in more detail in Chapter 4.

4. V-ATPase Regulation and Inhibition

The V-ATPase can consume a significant proportion of cellular ATP. For example, it is estimated to use ~10% of cellular ATP in the *M. sexta* midgut (Huss and Wieczorek, 2007). Thus it is vitally important that its activity can be regulated under low ATP conditions such as starvation, to minimise futile ATP turnover. The primary method of regulation is reported to be by disassociation of the V-ATPase complex, with the soluble ATP hydrolysing V_1 domain detaching from the membrane bound, proton translocating V_o domain. Following this dissociation the V_o domain becomes impermeable to protons and ATP hydrolysis no longer occurs in the V_1 domain. However, the precise mechanism of ATP silencing in the dissociated V_1 is poorly characterised and understood. By understanding the nature of ATP silencing it may be possible to design small molecules that mimic this interaction, resulting in V-ATPase inhibition. The current working model is that ATP silencing is a result of a rearrangement of the H subunit which contacts the rotor axle, preventing rotation and hydrolysis (Figure 4.1) (Jefferies and Forgac, 2008; Oot et al., 2016). However, alternative mechanisms have been proposed, such as the movement of the terminal helices in subunit F to interact in the AB dimer (Hildenbrand et al., 2010; Makyio et al., 2005). If this were the case, then it may be possible to target ATP silencing using a helical mimetic.

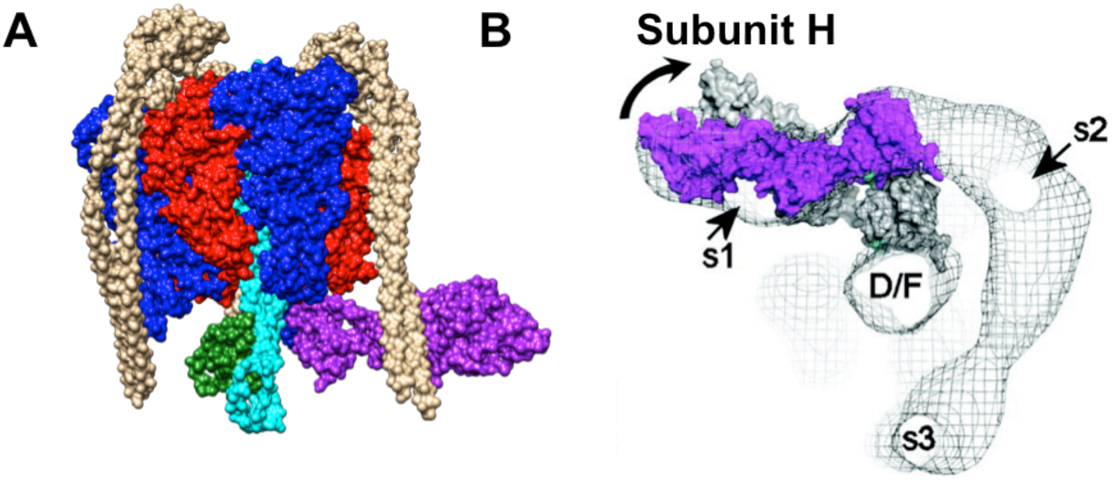


Figure 4.1 - V_1 schematic illustrating proposed H subunit interaction. (A) Surface representation of yeast V_1 X-ray structure (PDB 5D80) showing AB subunits (red and blue), stator EG subunits (beige), rotor DF (cyan and green) and H subunit (purple). (B) Proposed swing of H subunit to contact central rotor axle in context of whole complex (grey mesh). Adapted from (Muench et al., 2011).

As discussed in Chapter 3, V-ATPase would be an attractive therapeutic target for a range of diseases. While several inhibitors of the V-ATPase have already been identified, most suffer from a lack of selectivity (Michel et al., 2013), particularly the ability to discriminate between different V-ATPase isoforms within the body, preventing their use as effective therapeutics. However, Pea Albumin 1 subunit b (PA1b) has been shown to display high selectivity against the V-ATPase from different weevil populations

(Chouabe et al., 2011; Eyraud et al., 2013). It was hoped that understanding the mechanism of action that leads to this specificity could inform future efforts to design new selective V-ATPase inhibitors.

This work aimed to inform on V-ATPase inhibitor design through two distinct approaches. The first was to see if by understanding the structural changes necessary for silencing ATP hydrolysis in the V_1 domain through cryo-EM these changes could be mimicked, thus inhibiting the V-ATPase. This would be a new mode of inhibition for the V-ATPase family. The second was to explore the mode of action of PA1b with the aim of understanding the basis for selectivity and the location of inhibitor binding. This would allow further development of existing potent inhibitors which could be altered using this information to increase their specificity.

4.1 V-ATPase Regulation

4.1.1 Cryo-EM structure of V_1 from attached and degraded V-ATPase

Data for attached and degraded *M. sexta* V_1 samples were obtained from the intact V-ATPase cryo grids, prepared and collected as previously described (Section 3.1.2). When attached to the full complex, the V_1 domain reconstructions had a resolution of ~9 Å, which was improved to ~8 Å following accounting for flexibility with tightly masked refinements as detailed in Section 3.1.2 (Figure 4.2.A).

The degraded V_1 particles were selected through 2D classification within RELION, identifying side views of isolated V_1 as well as clear top views. This led to the selection of 11701 particles identified as isolated V_1. This is an imperfect way of selecting the isolated V_1 domain as some of the apparent top view classes may actually be top views of the entire complex, which would be detrimental to any resulting reconstruction. Following 3D auto-refinement and particle polishing of this particle set within RELION a ~9 Å reconstruction was obtained (Figure 4.2.B). However, care must be taken when interpreting the reconstruction obtained from the degraded V_1 sample. Not only is it possible that particles belonging to the intact complex remain in the data, it is also inherently a damaged and degraded sample, which is likely to not be fully representative of any natural regulation or dissociation.

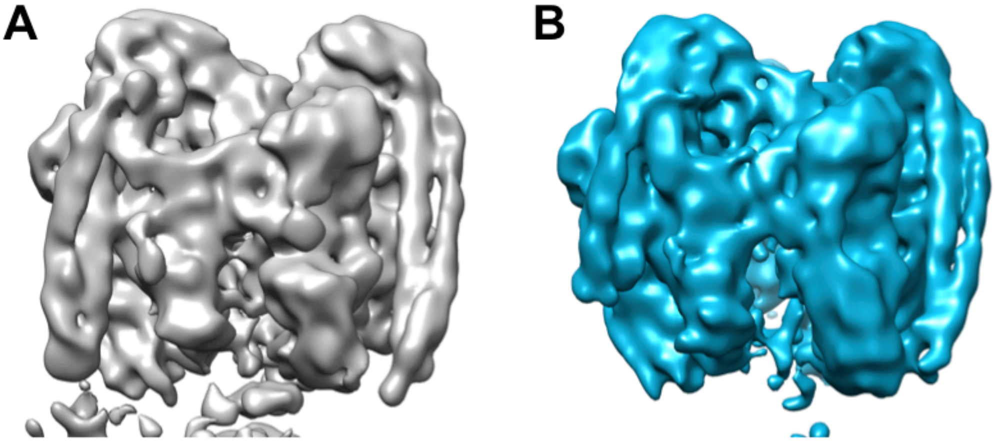


Figure 4.2 - 3D reconstructions of V_1 attached to the V-ATPase and degraded.

(A) V_1 domain of the V-ATPase at ~8 Å resolution. (B) V_1 domain from degraded V-ATPase particles at ~9 Å resolution.

Nevertheless, some initial insights can be gained from looking at this structure in comparison with the attached complex. Most noticeable is the lack of density corresponding to subunit H, despite this subunit being heavily linked with ATP silencing by biochemical and X-ray studies (Jefferies and Forgac, 2008; Oot et al., 2016; Parra et al., 2000). This was thought likely to be a result of degradation within the sample rather than showing a different mechanism for regulation. Interestingly, despite the sample not being the result of 'natural' dissociation the overall position and architecture of the AB subunits that make up the core of the V_1 domain is very similar to that of the attached, intact V_1 domain. This suggests that, even following this detachment, there are no large scale structural changes within the core of the particle.

4.1.2 Cryo-EM structure of V_1 from disassociated V-ATPase

4.1.2.1 V_1 Sample Preparation

Isolated V_1 (~1 mg/ml in 20 mM Tris–HCl and 100 mM NaCl, pH 7.9) from *M. sexta* midgut was kindly provided by the group of Prof. Helmut Wiezorek, University of Osnabruck, Germany. V_1 was detached from the V-ATPase complex and purified as described in Muench et al 2014 (Muench et al., 2014b; Vitavska et al., 2003). As the midgut is isolated during the molt stage of larval development, it is enriched in naturally dissociated V_1 (Gräf et al., 1996), because the glucose levels in the larva drop sharply prior the beginning of the molt stage (Siegert, 1995). Moreover, since they do not eat at this stage, the midgut V-ATPase is not required to regulate the pH within the midgut lumen for leaf digestion.

4.1.2.2 Cryo Grid Preparation and Optimisation

Prior to cryo grid preparation the V_1 sample was first assessed for sample quality and purity as well as aggregation or degradation by negative stain EM. Grids were UV treated for ~30 mins immediately prior to use and stained using uranyl acetate. The grids were screened on the in-house Technai T12 microscope, operating at 120 kV and at a nominal magnification of 30k. This showed homogeneous particles with little aggregation and no obvious degradation (Figure 4.3). Cryo grid preparation was carried out on an FEI vitribot IV according to standard procedures using lacey grids which had been glow-discharged for 30 s prior to use with blot times varying between 4 and 6 s. These were screened using a FEI F20 microscope equipped with a Gatan 4K x 4K CCD camera. Grids showed reasonable particle distribution, with acceptable ice thickness, although particles were observed to cluster around the edges of the holes.

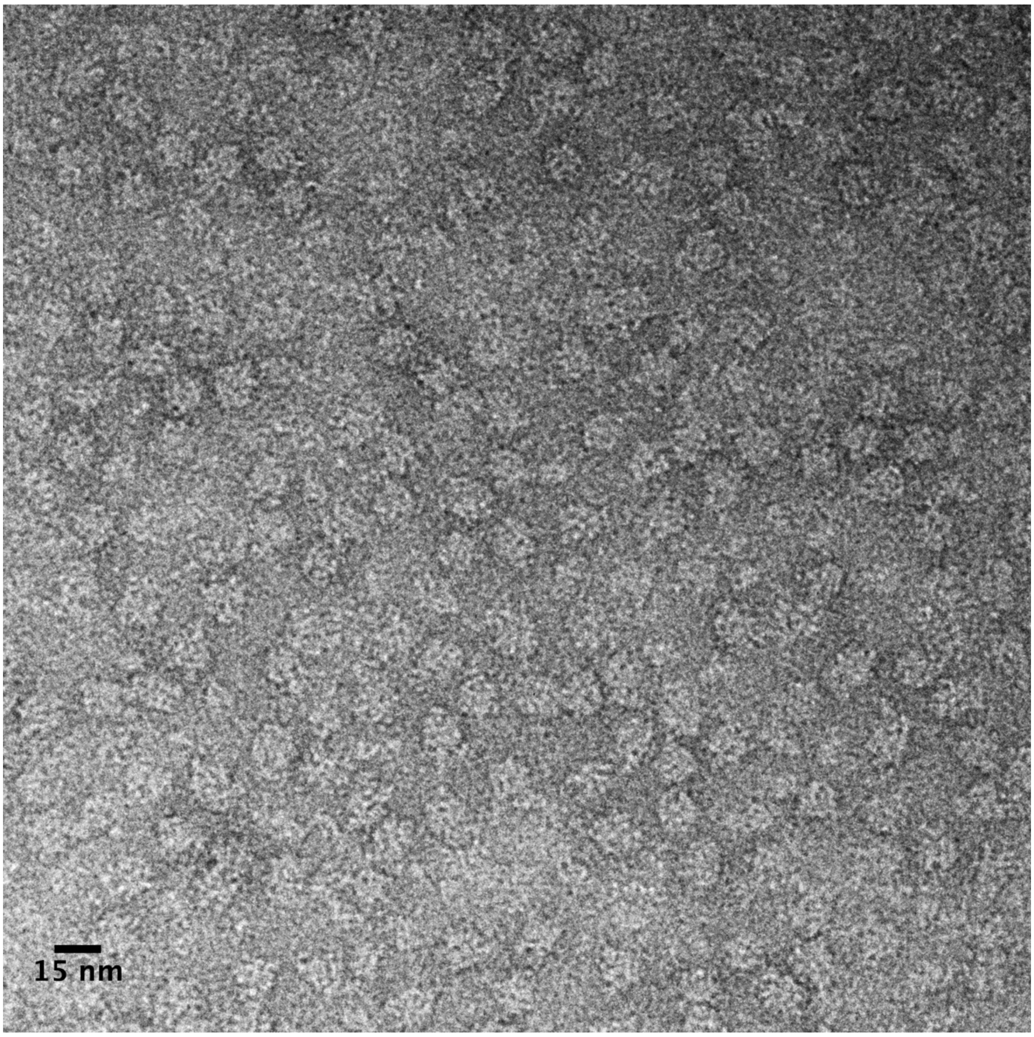


Figure 4.3 - V_1 negative stain micrograph. Negative stain micrograph of dissociated V_1 at 50k magnification.

Data were collected in a semi-automated fashion using SerialEM at the Chinese Academy of Sciences, Beijing on a FEI Titan Krios equipped with a Falcon II camera. 986 micrographs consisting of 32 frames each were collected at a total dose of ~60 $e^-/Å^2$ over 2 s at 1.76 Å/pixel sampling (Figure 4.4).

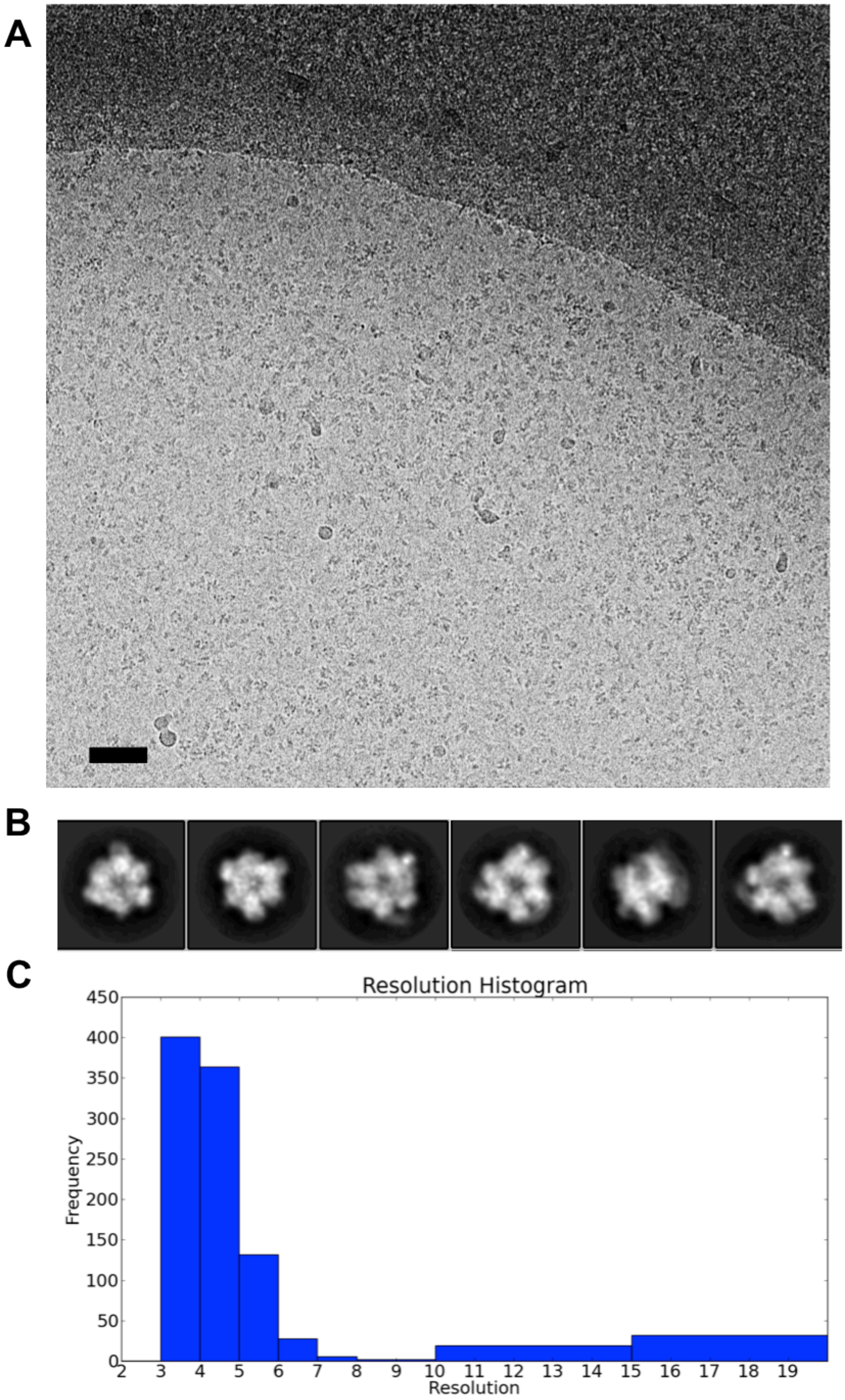

A
B
C
Resolution Histogram
Frequency
0
50
100
150
200
250
300
350
400
450
Resolution
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

Figure 4.4 - Dissociated V_1 cryo-EM raw data. (A) Cryo micrograph of dissociated V_1 (scale bar 500 Å). (B) Reference free class averages of V_1 showing clear top and side views. (C) Estimated resolution of micrographs from gCTF.

4.1.2.3 Image Processing

Whole-frame motion correction was carried out on all micrographs using MOTIONCORR (Li et al., 2013a) and defocus parameters calculated with CTFFIND4 in the first instance (Rohou and Grigorieff, 2015), with micrographs having a defocus range of ~2-4 µm. 43326 particles were manually picked in EMAN2 (Bell et al., 2016; Tang et al., 2007), prior to 2D classification within RELION (Scheres, 2012). Well-defined classes showing the characteristic top view of the V_1 domain were observed as well as side views with clear AB domains and some stator density (Figure 4.4.B). Particles belonging to poorly-defined classes were removed to leave 29905 particles. 3D auto-refinement of this subset led to a 8.1 Å reconstruction, which was improved to 7.5 Å following per particle motion correction and polishing within RELION. A further resolution improvement to 7.0 Å was obtained through per particle defocus determination through gCTF (Zhang, 2016).

4.1.3 Results and Discussion

The final 7.0 Å reconstruction of V_1 is consistent with the previous cryo-EM studies of the domain, clearly showing the position of the open site (Muench et al., 2014b). The improved resolution allows secondary structure to be identified, with the helices in the core of the region well-defined (Figure 4.5). However, in contrast to the previous cryo-EM study, stator density was poor in the final map, with only one stator being clearly visible.

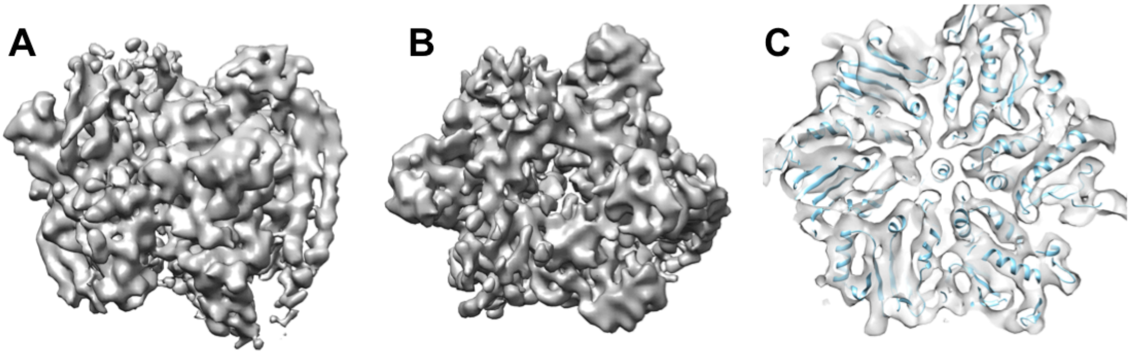


Figure 4.5 - Dissociated V_1 structure. (A) Side view of dissociated V_1 reconstruction at 7 Å, facing open site, showing a stator well resolved on the right of the complex. (B) Top view of the V_1 domain. (C) Slice through the core of V_1 domain around electrostatic-bearing region showing helical secondary structure well-resolved.

The most notable feature of both the 2D classes and the 3D reconstruction is the lack of any density for the H subunit thought to be responsible for ATP silencing within V_1. While this feature was not previously observed within *M. sexta* V_1, it has been detected in a crystal structure from the yeast complex (Oot et al., 2016). Whereas in the previous *M. sexta* model the lack of subunit H density could be attributed potentially to the low resolution, the current 7 Å reconstruction is very comparable to the 6.5 Å crystal structure. Indeed, once phase bias is accounted for, it could be argued that a 7 Å EM reconstruction would be superior in quality to a 6-7 Å X-ray map. Indeed a 6.5 Å structure is in the bottom ~0.1% of all structures in the PDB by resolution. The 7 Å EM map clearly shows secondary structure resolved within the AB subunits, and local resolution estimates that portions of the reconstruction extend to ~6 Å (Figure 4.6).

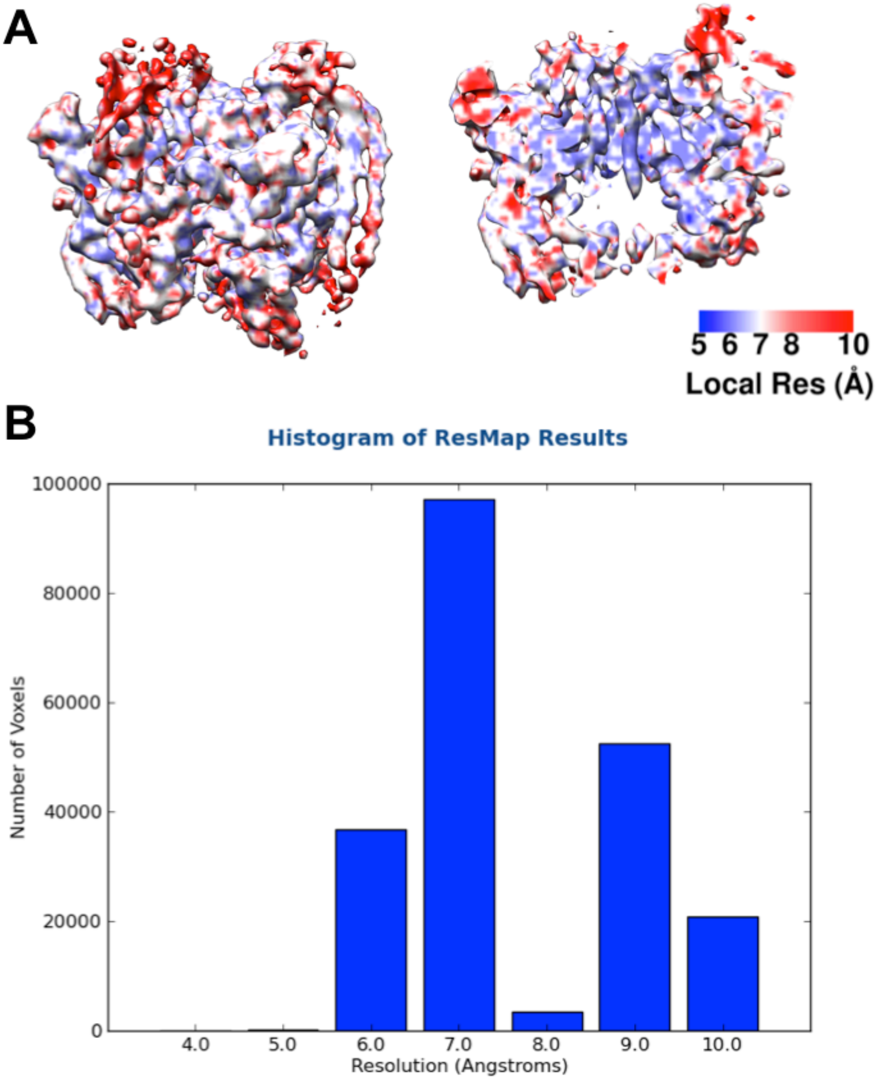


Figure 4.6 - Dissociated V_1 local resolution. (A) 3D reconstruction of dissociated V_1 coloured by local resolution calculated by ResMap (Kucukelbir et al., 2013). (B) Histogram showing voxels assigned to each resolution bin.

As the H subunit is thought to be crucial for silencing of ATP hydrolysis the V_1 crystal structure, which had a defined position for subunit H, was filtered to 30 Å and used as a starting model for 3D classification (Figure 4.7). This would check if there was a subset of particles where subunit H could be observed by giving it the extra density to align with. It should be noted that this is not normally a sensible approach, as it could add bias into the model, but it does indicate whether the subunit may be present. However, no well-defined 3D classes were obtained with any density corresponding to the H subunit. In addition the X-ray model was used as a starting model for the auto-refinement procedure in an attempt to bias the reconstruction towards the H subunit, but again the resulting reconstruction at 7.5 Å showed no signs of the H subunit (reconstruction not shown).

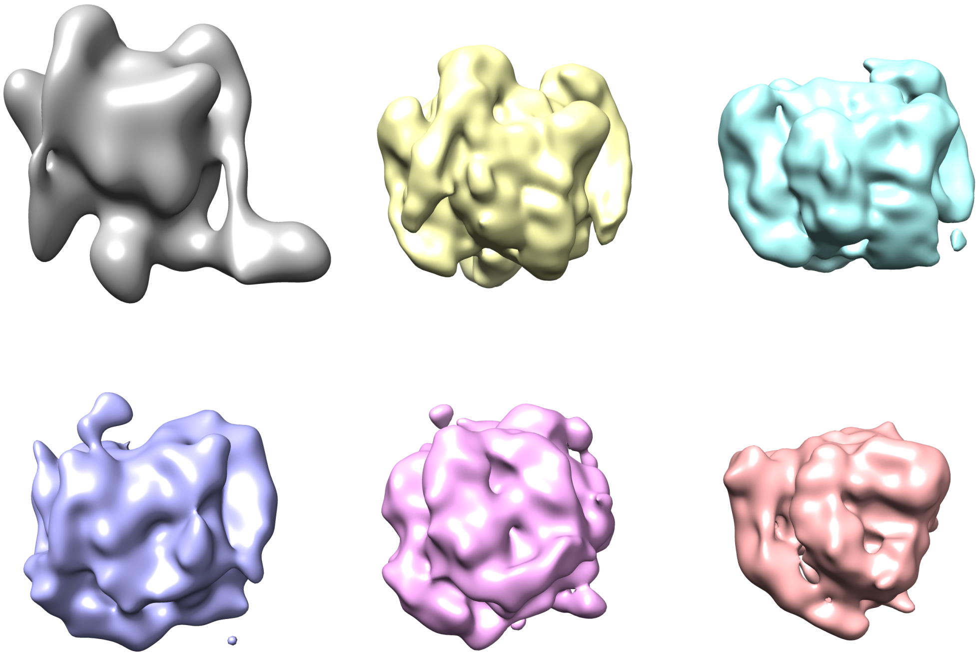


Figure 4.7 - Biased 3D classification of V_1. Results of 3D classification within RELION using a filtered crystal structure (PDB 5D80) (grey) start model (Oot et al., 2016). Additional density at base corresponding to subunit H was not observed within any resulting class.

Subunit H was present in the V_1 sample used to prepare the cryo grids and has been observed on gels (data not shown). Previous work on the V_1 domain also showed that the H subunit is present in gels and negative stain classes of V_1 (Muench et al., 2014b). To test that this was not sample-specific, the negative stain grids that were made on the same day as the cryo samples were re-analysed and data collected under my supervision by Rachel Johnson, a PhD student in the group. In total 181 micrographs were collected at a nominal magnification of 50k resulting in 15218 particles. Reference-free classification in RELION resulted in classes that showed clear additional density at the base corresponding with the location of V_1 (Figure 4.8). This observation may be due to the surface tension forces of the stain, forcing H into a single defined position or limiting its movement. Without this stabilising interaction, the H subunit may be mobile and thus when captured in vitreous ice will be in a range of positions and therefore lost in the averaging process.

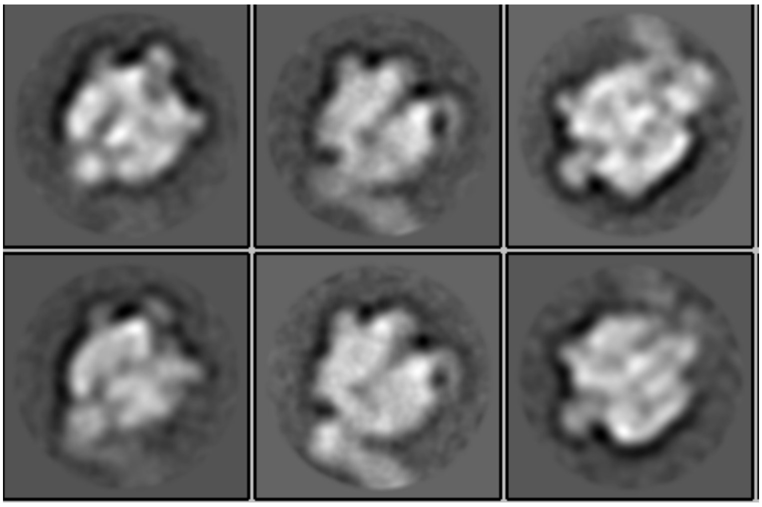


Figure 4.8 - Negative stain V_1 class averages. Class averages of V_1 showing extra density at base of several classes, corresponding to location of H subunit. Data processing carried out by Rachel Johnson.

It is important to note that, in the X-ray structure subunit H makes extensive crystal contacts with neighbouring subunits, which may mirror the stabilising effect of the stain pressing onto the carbon film (Figure 4.9). Thus, while the interaction between H and the central rotor axle, shown by X-ray studies and cross linking experiments may indeed occur, it is likely transient interaction rather than a stable interaction based on the EM data.

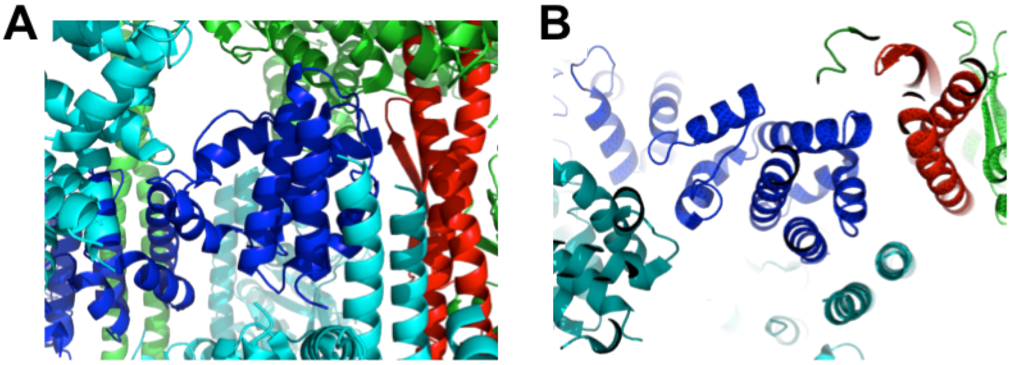


Figure 4.9 - H subunit crystal packing. Crystal packing around the H subunit in V_1 X-ray structure (PDB 5D80) (Oot et al., 2016). H subunit (blue) packs against the central axle subunit D (red) and EG stators and H subunits from neighbouring molecules (cyan).

As only limited stator density was observed in the EM reconstruction and no subunits were visible in the collar region, it is possible that the alignment of the V_1 domain is dominated by the position of the open site. Thus subunit H could become rotationally averaged across the three different positions. However, if this were the case, some smeared density from subunit H in three positions should be visible around the base of the V_1 domain, but nothing was observed even at high contour levels. Also when using the X-ray starting model, subunit H was more prominent than the position of the open site and thus would be expected to 'lock in' the alignment on this feature. The X-ray study also suggested that isolated V_1 was only present in one conformational state, corresponding to the 'state 2' in yeast (Oot et al., 2016; Zhao et al., 2015). As the EM reconstruction lacks any rotor or collar density it is not possible to say which state the *M. sexta* V_1 structure is residing in. The H subunit is ~55 kDa which given the overall mass of V_1 (550kDa), means it makes up ~10% of the structure. Given other examples in the literature (Bai et al., 2015b; Hesketh et al., 2015), it would be expected that a mass of this size would be observed through 3D classification if present in a stable position.

4.1.3.1 Comparison of V_1 models

If the H subunit was not playing a role in ATP silencing, the next step was to assess if there were any clear differences between the different V_1 domains. Although this could not be done at the atomic level, the resolution was sufficient to be able to see large conformational changes in the secondary structure. To accomplish this, MDFF (McGreevy et al., 2016; Trabuco et al., 2011) was used to flexibly fit homology models of the AB subunits (generated in Phyre2 (Kelley et al., 2015)) into each map in turn using the approach described previously (Section 3.1.4). The fitted models were then split into single AB pairs corresponding to the open, loose and tight sites. AB pairs corresponding to each site were then overlaid using the MatchMaker and MatchAlign tools within Chimera (Meng et al., 2006). Backbone RMSD values were then calculated within Chimera and plotted onto the secondary structure to attempt to show an objective measure of difference between the fitted models.

It is worth noting that only minimal changes are observed in the AB subunits of V_1, with the fitted models from the three states (attached, dissociated and degraded) all showing very similar backbone positions within the maps. The majority of changes between the attached and dissociated AB subunits are located at the base of the open site (Figure 4.10). However, it is not obvious how these small shifts would cause ATP silencing to occur. While there are some differences in the EM maps in this region, due to the modest resolution of the EM reconstructions, these minor differences could be a result of errors within the fitting, particularly in loop and terminal regions.

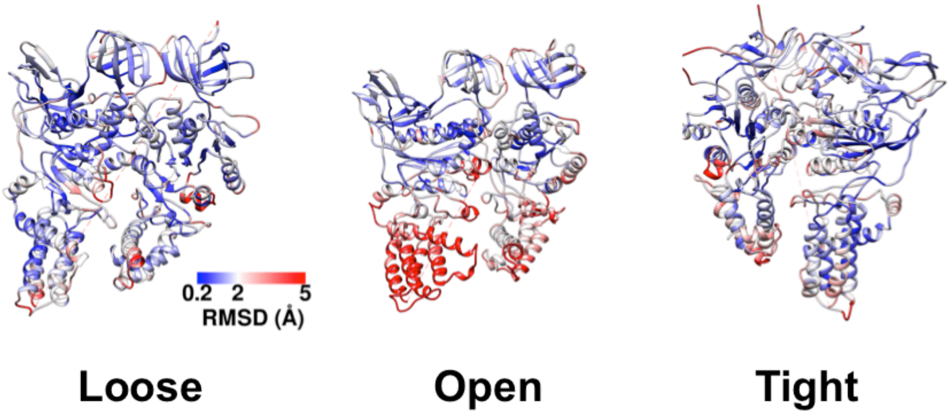


Figure 4.10 - Differences in AB pairs between dissociated and attached V_1. Fitted atomic models of AB pairs from each state coloured by backbone RMSD values between the dissociated and attached structures. Large differences are concentrated at the base of the open site.

While not directly relating to regulation, several other features are apparent within the reconstruction, notably the lack of density for two out of the three stators. This could be the result of degradation of the sample (although it is unclear why one stator would remain relatively intact), an image processing artefact, or an indication that these regions are highly mobile within the sample. As it is known that collar subunit C is lost from V_1 upon dissociation, it is possible that the two stators that contact this subunit become more mobile following its removal and that they are therefore not visible in the reconstruction. Additionally it is interesting to note that the density for the central rotor axle subunit D becomes weak in the central cavity in the lower portion of V_1, near the 'DELSEED' loop region (Figure 4.11). This suggests that removal of the connection between V_1 and V_o causes this region to become mobile, with subunit D only held rigidly around the electrostatic bearing region previously described (Section 3.2.3). This also re-enforces the theory that no stable interaction with the H subunit is present, as this would be predicted to stabilise the position of the central rotor axle.

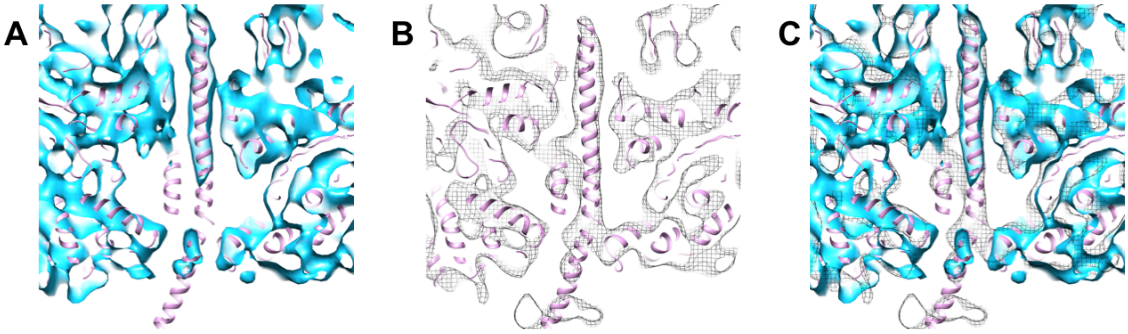


Figure 4.11 - Comparison of rotor axle density between dissociated and attached V_1. (A) Central slice through dissociated V_1 reconstruction showing lack of density for rotor axle in the central cavity towards the base. (B) Rotor axle density in attached V_1 domain. (C) Overlay of dissociated (blue) and attached (grey mesh).

One interesting and potentially worrying possibility is that the wrong question is being asked. Is the dissociated sample obtained from ATP starved organisms, either yeast or *M. sexta*, actually relevant? If the reconstruction from this supposedly natural dissociated form is compared with that from a specimen obtained from degradation of intact V-ATPase *in vitro,* very few differences are observed in the fitted models. Indeed the AB subunits are almost identical aside from minor differences primarily centred around loop and terminal regions, suggesting that there is little or no difference between V_1 that has undergone 'natural' dissociation and V_1 which has dissociated *in vitro* without any prior regulatory signal and so is assumed to be 'degraded' (Figure 4.12). A further speculative theory is that the V-ATPase is in equilibrium between apparent 'dissociated' and full complex, with cellular signals simply shifting this equilibrium towards a more dissociated state. This would explain why the 'natural' V_1 domain and the 'degraded' V_1 domain are identical in structure at the resolution studied. Its interesting to note that it was the idea that both domains dissociate was originally based upon studies showing that tagged subunit C diffuses away from the membrane upon starvation (Sumner et al., 1995; Vitavska et al., 2003). However a recent study by the same group showed that labelling of the A subunit does not result in a loss of signal at the membrane (Tabke et al., 2014). This suggests the V_1 domain is never truly dissociated from the membrane in *M. sexta*, implying that the actual process of regulation and ATP silencing is potentially more complex and subtle than simply breaking the V-ATPase complex in two. This is perhaps unsurprising given that previous studies have shown that while ATP turnover is silenced in starvation conditions, in order to disassemble the complex ATP is in fact required (Huss and Wieczorek, 2007).

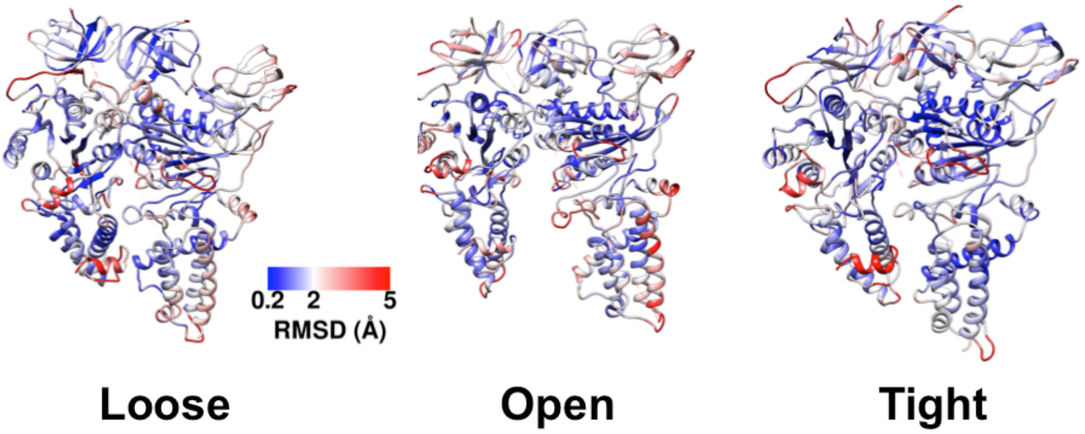


Figure 4.12 - Comparison of degraded and dissociated models. Comparison of atomic models flexibly fitted into both dissociated and degraded domains coloured by RMSD. Minimal differences are seen in each AB pair.

An additional concern is knowing where the V_1 sample was obtained in the 'lifetime' of the complex. While it is assumed that the V_1 is obtained results from dissociation and degradation, it is possible that a proportion of this sample was actually from newly assembled V_1, which has not yet been coupled with V_o to form an intact complex, and may be regulated in a different manner entirely. However, this is less likely as it is known that glucose starvation induces the dissociation of the V-ATPase and due to the larval stage at which V_1 is isolated being depleted in glucose it is assumed that the majority of free V_1 is a result of this process (Gräf et al., 1996; Siegert, 1995). It is also important to note that the full complex is purified by size exclusion chromatography and so no free V_1 should be purified at the same time (Huss et al., 2002; Rawson et al., 2015). Therefore any V_1 found in the sample as 'degraded' cannot be newly assembled V_1.

So the question remains: what is the cause of ATP silencing following dissociation of the V_1 domain? As the EM reconstructions show no sign of the expected rearrangement of the H subunit, perhaps ATP silencing is a result of a more subtle change in the structure of V_1. However this would require obtaining higher resolution structures than those currently available of both the intact complex and the dissociated V_1 domain. Subunit F was also proposed as an alternate possibility for the regulation of V_1, but this is not observed in either the X-Ray or EM structures, so this seems unlikely to be the case (Hildenbrand et al., 2010; Makyio et al., 2005). It is also possible that the regulation of the V-ATPase differs across species, which would explain the differences in the *M. sexta* and yeast structures. This would be potentially unsurprising as several aspects of dissociation apparently differ between the two systems, with *M. sexta* V_1 not thought to move from the membrane while the yeast is thought to completely dissociate. Given that no clear structural change was observed to trigger ATP silencing, unfortunately this work did not provide any potential inhibitor binding sites. Nevertheless, it has shown that the simple picture of dissociation as a mechanism for silencing ATP turnover may be more complex than was previously thought.

Negative stain data collection of V_1 at 50k magnification and classification was carried under my supervision by a PhD student in the group, Rachel Johnson, who also aided in preparation of the V_1 homology models in Phyre2 and subsequent fitting in MDFF with my supervision.

4.2 V-ATPase Inhibition

The V-ATPase is a potentially attractive therapeutic target for a number of disease states including cancer invasiveness and osteopetrosis. However, the lack of selectivity in existing V-ATPase inhibitors precludes its development into a viable drug target. This work aims to understand the basis of selectivity for a plant toxin which inhibits the V-ATPase in a selective manner with the aim of using this to inform future efforts in therapeutic developments against this target.

4.2.1 V-ATPase inhibition by PA1b

Pea albumin 1 subunit b (PA1b) is a 37-residue peptide isolated from pea plants (Jouvensal et al., 2003). Incorporating three disulfide bridges, it displays a high degree of stability (Figure 4.13) (Da Silva et al., 2009; Eyraud et al., 2013; Jouvensal et al., 2003).

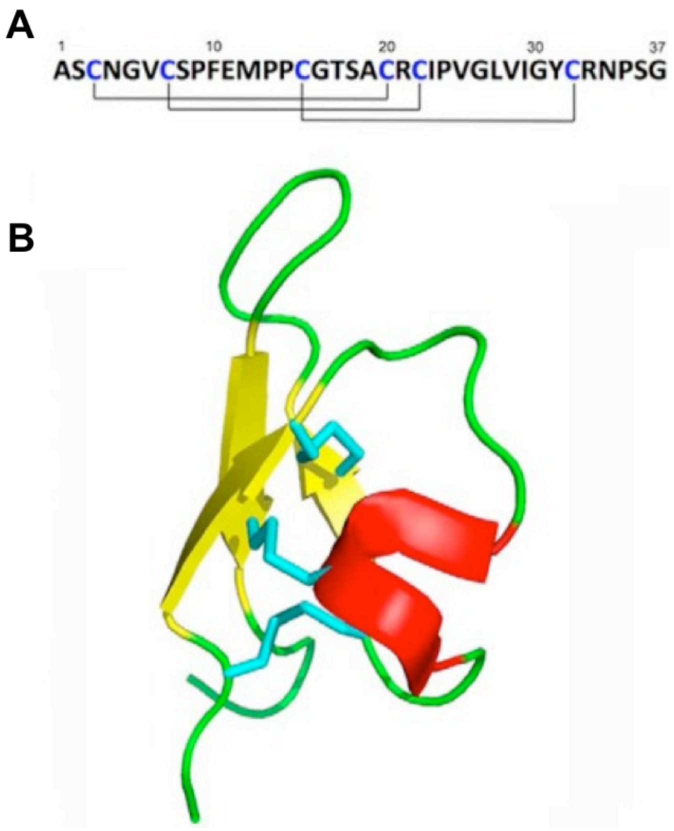


Figure 4.13 – PA1b sequence and structure. (A) PA1b sequence with disulfide bridges shown. (B) PA1b structure (PDB: 1P8B) with disulphide bridges shown in cyan. Figure adapted from (Muench et al., 2014a).

In plants it is produced as a natural insecticide, acting to kill weevils. Its toxic effect was found to be through the inhibition of the V-ATPase complexes of these species (Chouabe et al., 2011; Gressent et al., 2007). In contrast to the majority of other V-ATPase inhibitors, PA1b displays high levels of specificity; for example, it shows no inhibition of the V-ATPase from yeast or humans and it is even highly selective among different weevils, which show little sequence variation (Gressent et al., 2007; 2011). However, its precise binding location on the complex was unknown as was its mode of action. Although it was known to bind to the membrane bound V_o domain and thus could access this domain directly from the insect gut. At the time this work was carried out DEDs were not yet widely in use, thus the realistic resolution obtainable for the V-ATPase was in the region of 11-16 Å. Therefore, negative stain EM was used to rapidly identify the binding site of PA1b on the V-ATPase complex due to its ease and speed as the expected resolution gain from cryo-EM was not thought likely to provide much additional information.

4.2.1.1 Sample preparation and data collection

The PA1b sample was kindly provided by Dr Frédéric Gressent at Université de Lyon, France, while V-ATPase was provided by the group of Prof. Helmut Wiezorek, University of Osnabruck, Germany as previously described (Chapter 3). The resolution typically obtainable through negative stain EM is 15-25 Å, thus the small PA1b toxin would likely be too small to directly observe using this technique (Figure 4.14).

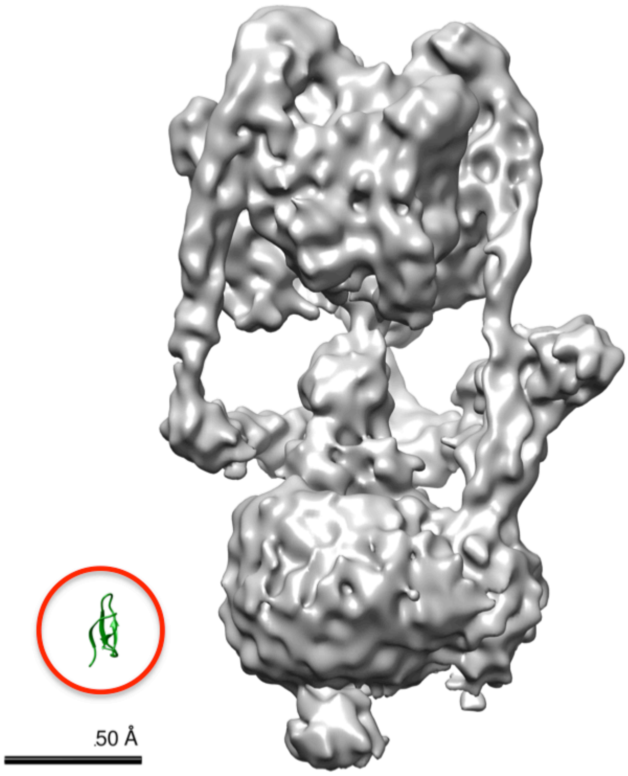


Figure 4.14 - V-ATPase and PA1b size comparison. Size comparison of V-ATPase (grey) and PA1b toxin (circled in red). PDB: 1P8B (Jouvensal et al., 2003) EMDB: 2781

To combat this, the PA1b was first conjugated via a biotin tag to a streptavidin-functionalised horseradish peroxidase (Strep-HRP) in order to increase the steric bulk and bring the total size of the inhibitor from ~4 kDa to ~100 kDa. To check if conjugation with biotin had a detrimental effect on inhibition, binding studies were carried out by Dr Markus Huss, University of Osnabruck, Germany, which showed no overall changes to the binding affinity following the addition of the biotin tag (Figure 4.15), with K_i of 6.5 ± 1.13 nM for PA1b and 5.4 ± 6.6 nM for PA1b-biotin.

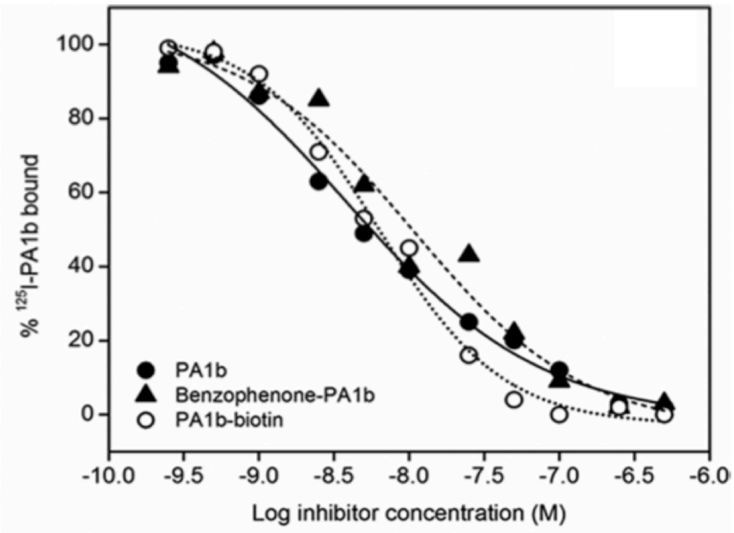


Figure 4.15 - Binding curves of PA1b ± biotin tag. Binding curved carried out by Dr Markus Huss showing no differences between PA1b binding after addition of biotin tag. Figure adapted from (Muench et al., 2014a).

All negative stain grids were prepared according to standard procedures using continuous carbon-coated grids that had been UV treated for ~30 minutes prior to use. Data were then collected using a FEI Technai T12 microscope equipped with a 2k CCD camera (Figure 4.16). 527 and 841 micrographs were collected for the PA1b and PA1b + ATP samples, respectively.

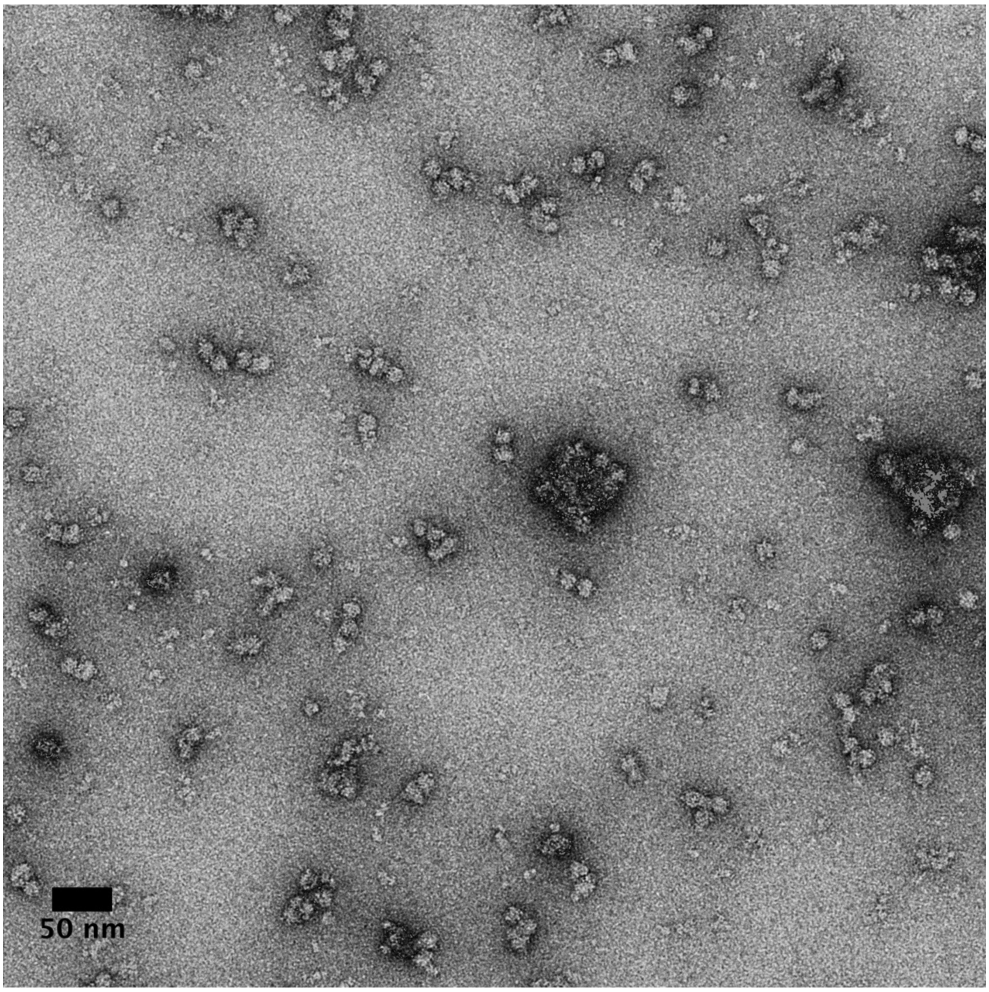


Figure 4.16 - Negative stain micrograph of V-ATPase + tagged PA1b. Example micrograph of V-ATPase + PA1b dataset showing reasonable particle distribution.

4.2.1.2 *Results and Discussion*

Prior to my involvement in this project preliminary studies had been carried out by Dr Stephen Muench who collected an initial five datasets and controls used to confirm PA1b binding. To avoid any bias within the data processing five samples were initially collected, V-ATPase alone, V-ATPase + Strep-HRP-PA1b (conjugated for 1 hour), V-ATPase + PA1b, V-ATPase + Strep-HRP and

V-ATPase + Strep-HRP + PA1b (not pre conjugated). Each of these was made by Dr Clair Phillips and labelled anonymously so that all data were collected and processed 'blind'. All data were picked in EMAN and the reference-free classification was done in IMAGIC-5. Following classification, only the samples containing both PA1b and Strep-HRP showed any additional density. Following this I prepared all subsequent grids and performed all imaging and image processing.

A further 527 micrographs of the V-ATPase with the pre-incubated Strep-HRP tagged PA1b were then collected, leading to 10056 particles (Figure 4.16). Following reference-free classification within IMAGIC-5 the classes clearly showed an additional density at the base of the V_o domain of the V-ATPase (van Heel et al., 1996) (Figure 4.17).

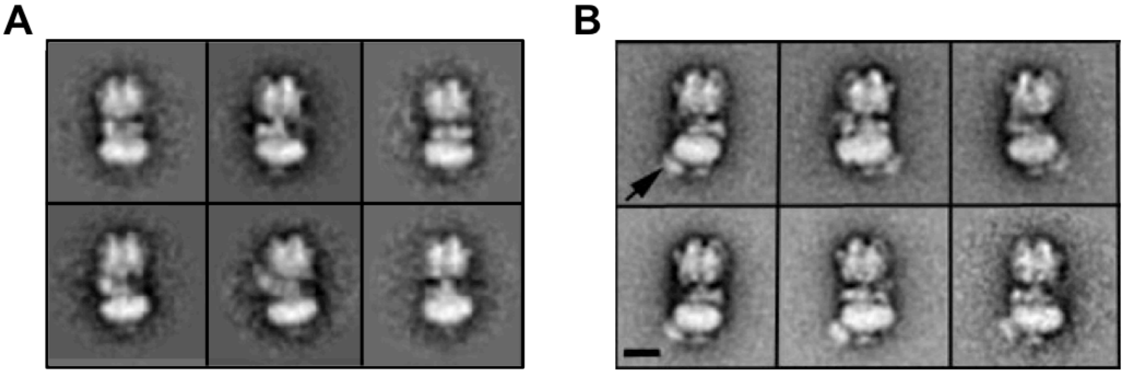


Figure 4.17 - Negative stain classes ± PA1b. (A) Apo V-ATPase negative stain classes. (B) V-ATPase + PA1b classes, arrow denotes additional PA1b density (scale bar 150 Å). Figure adapted from (Muench et al., 2014a).

Following an asymmetric 3D refinement within the refine module of EMAN, a 3D model was obtained showing an arc of density extending ~100° around the base of V_o (Ludtke et al., 1999) (Figure 4.18). At the time the study was carried out, 3D classification was not routine, thus no effort was made to separate out separate bound states within the data.

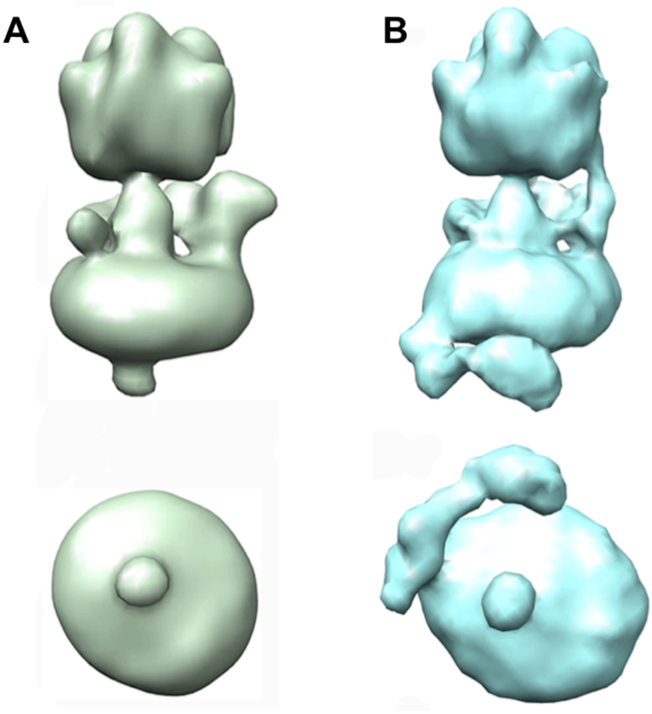


Figure 4.18 - 3D reconstructions ± PA1b. (A) 3D reconstruction of apo V-ATPase. (B) 3D reconstruction of V-ATPase + PA1b showing additional density in arc at base of V_o domain. Figure adapted from (Muench et al., 2014a).

The binding location suggested that the inhibitor was binding to the *c*-ring of V_o, which would be consistent with the binding location of many previously identified inhibitors, including bafilomycin (Osteresch et al., 2012; Wang et al., 2005). These *c*-ring binders are thought to act as a 'spanner in the works' by binding to the *c*-ring, being rotated during the complex's normal activity, and then preventing further rotation when they reach the *c/a* subunit interface (Wang et al., 2005) (Figure 4.19). If this were the case, then upon addition of ATP to the sample the additional density would be expected to localise to a single point at the base of the V_o domain, consistent with the location of the *c/a* subunit interface.

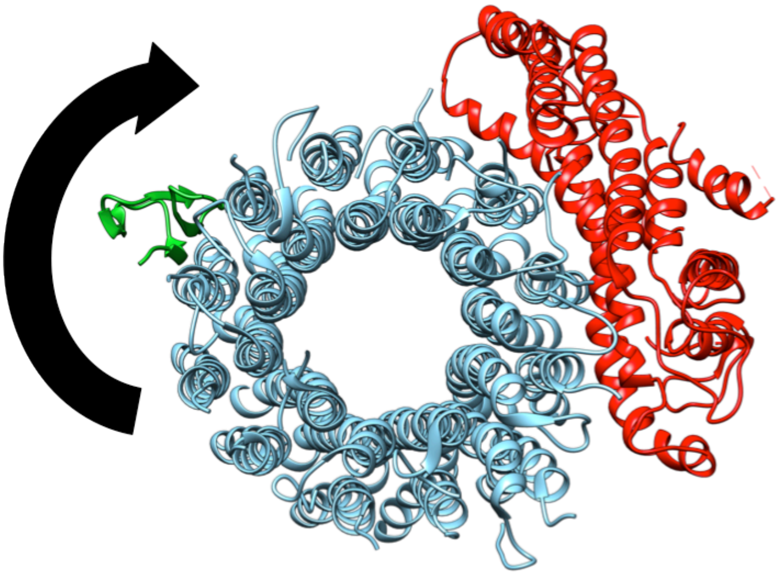


Figure 4.19 - PA1b expected mechanism schematic. Schematic of expected PA1b mechanism with PA1b (green) bound to *c*-ring (cyan) driven into subunit *a* (red) by rotation of the *c*-ring.

To test if PA1b was operating via this mode of action ATP was added to the sample, negative stain grids were prepared and the data collection repeated, leading to 841 micrographs containing 11341 particles. Upon classification in IMAGIC-5, additional density was again observed at the base of V_o. 3D refinement in EMAN showed clear density at the base, approximately the correct size for a single bound PA1b-Strep-HRP, localised to a single point as expected. However, instead of being positioned at the *c/a* subunit interface, the density appeared to be localised ~90° away (Figure 4.20). This was an unexpected result, showing that PA1b acts on the V-ATPase through a novel mechanism. Efforts were made to bias the resulting reconstructions through the use of starting models containing a smear of density for the +ATP set and a localised density for the –ATP dataset but in both cases no difference was found in the final result.

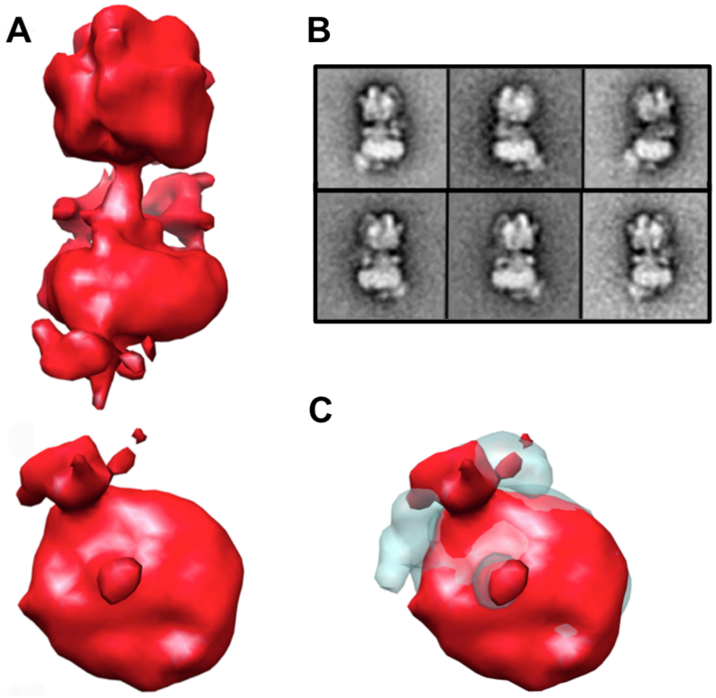


Figure 4.20 - V-ATPase and PA1b + ATP. (A) 3D reconstruction of V-ATPase + PA1b + ATP showing localised density at base of V_o. (B) 2D classes showing additional density at the base of the V-ATPase. (C) Comparison of PA1b binding +ATP (red) and -ATP (cyan). Figure adapted from (Muench et al., 2014a).

In order to further examine the reason for the unexplained binding location, photoaffinity labelling studies were carried out by Dr Markus Huss, Osnabruck, using radiolabelled PA1b-benzophenone. This showed that binding appeared to be to the subunits making up the *c*-ring, as expected, but also to subunit *e* (Figure 4.21). This finding was particularly interesting as the location of subunit *e* within the V-ATPase complex was not known at the time, as it is expected to be disordered and heavily glycosylated, making structural studies extremely challenging. The binding location of PA1b suggests that subunit *e* is positioned at the base of V$_o$, distant from subunit *a*, although the modest resolution obtained does not allow any further interpretation of the precise location of the subunit.

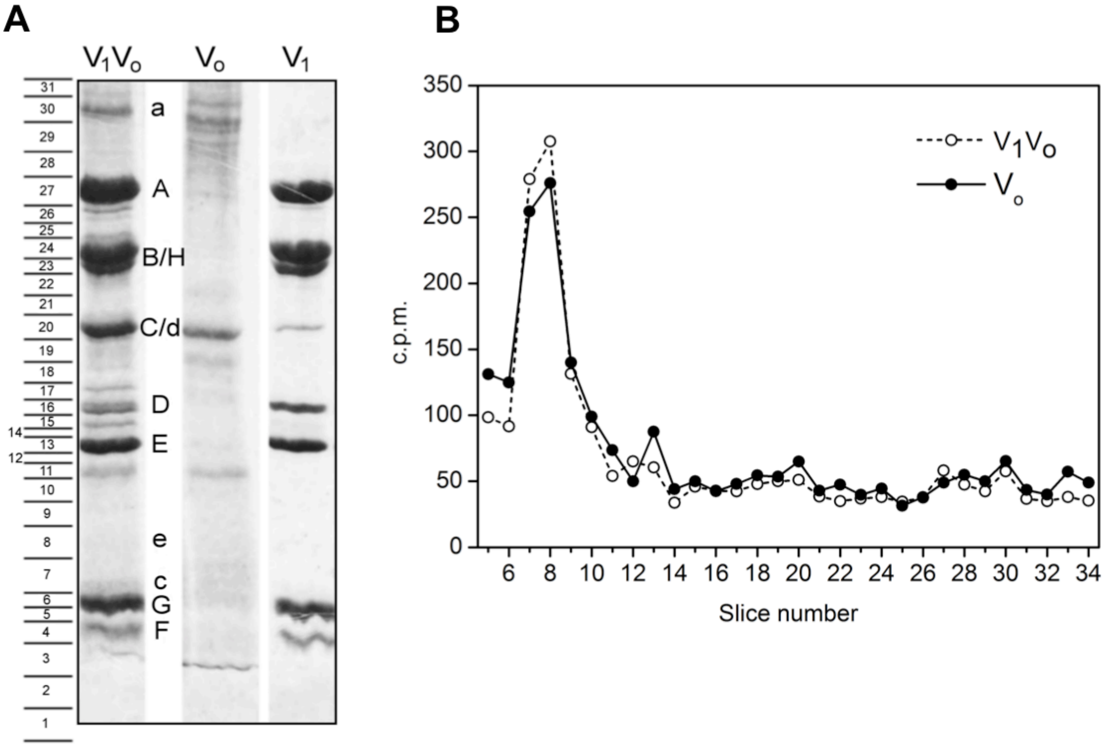


Figure 4.21 - PA1b radiolabelling data. Radiolabelling analysis performed by Dr Markus Huss. (A) SDS-PAGE showing V-ATPase subunits following labelling with ^{125}I-PA1b with slice numbers used for scintillation counting. (B) Scintillation counting against slice number showing majority of counts in bands corresponding to subunits *c* and *e*. Figure adapted from (Muench et al., 2014a).

The binding to subunit *e* may be the basis of PA1b's selectivity. For example, it has been suggested that subunit *e* is not present in the purified yeast complex, which may explain why PA1b has no effect (Bueler and Rubinstein, 2015). If this is the case it may enable the development of specific V-ATPase inhibitors by targeting these non-conserved subunits. However, currently subunit *e* is unlikely to be an attractive target for future inhibitors, as very little is currently known about it. While this work has identified its rough location within the complex, its precise location is still not well-established, nor are its structure or function within the V-ATPase. It would perhaps be beneficial to perform cryo-EM studies on PA1b binding as recent advances would allow the binding site to be determined with much greater accuracy. This would also potentially shed further light on the position and function of subunit *e* within the V-ATPase complex.

4.3 Conclusions

The V-ATPase is involved in a large range of disease states including cancer invasiveness and osteopetrosis and as such has been a long-sought-after therapeutic target. To investigate this further, two approaches were used. The first was to investigate the structural changes that accommodate ATP silencing of the V-ATPase during glucose starvation to see if they could be mimicked by small molecules. The second was to develop our understanding of the species specific inhibitor PA1b, to characterise the basis of its selectivity, which is currently the biggest hurdle in V-ATPase inhibitor development.

To explore the structural changes that trigger this silencing in isolated V_1, cryo-EM reconstructions were obtained of V_1 in naturally dissociated state and compared to similar resolution reconstructions of the V_1 domain in the intact complex. The current theory in the field suggests that a rearrangement of subunit H is responsible for ATP silencing (Jefferies and Forgac, 2008; Oot et al., 2016), and is supported by biochemical and X-ray data. However, the cryo-EM reconstructions show no density corresponding to this subunit. Biochemical and negative stain analysis both show that subunit H is present within the sample used for the EM studies, suggesting that subunit H is highly flexible and any interaction with the rotor axle must be transient. Unfortunately, while small changes in the AB subunits were observed between the attached and dissociated state, they were not sufficient to explain the silencing.

In addition, a cryo-EM reconstruction of V_1 from *in vitro* degraded V-ATPase was compared to the dissociated sample. This showed only minor differences between the two isolated V_1 structures, potentially suggesting that the dissociated sample does not fully represent the physiologically relevant state of the protein.

Together, this work suggests that the H subunit rearrangement is possibly either only present in lower Eukaryotes or does not fully explain this process. Moreover, the role of silencing during association is also poorly understood and often overlooked when studying ‘isolated’ domains. Although this work does not provide any firm conclusions on the exact cause, it does provide evidence that a simple structural rearrangement is not the cause. Unfortunately no new insights into a possible site for inhibition were generated, but understanding function and mechanism is essential if new therapeutics are to be developed.

In addition to identifying new inhibitor binding sites, a second important aspect of drug design is understanding the mode of action of existing inhibitors, so that this insight can be incorporated into future drug design efforts. To this end the binding site of a selective V-ATPase inhibitor, PA1b, was investigated through negative stain EM. This used a novel tagging approach for EM studies, which permitted the identification of PA1b at the base of the V_o domain, which in conjunction with radiolabelling showed the inhibitor bound to the *c*-ring and subunit *e*. This work also showed the approximate location of *e* within the complex for the first time. This showed the utility of even low resolution EM studies to rapidly identify inhibitor binding sites before carrying out more detailed, focussed studies on the area of interest.

Although this work did not provide any new V-ATPase inhibitors, it has provided insight into the mechanism of V-ATPase regulation, which could prove valuable in future efforts to identify novel inhibitors. It also demonstrated the utility of low resolution tagging techniques to rapidly identify inhibitor binding sites, which can be used to guide future structural characterisation through co-crystallization for example.

5. Visualising Inhibitor Binding using EM

The technical advances occurring in the field of cryo-electron microscopy (cryo-EM) have resulted in the resolution expectations for single particle EM reconstructions being significantly improved from the identification of secondary structural elements, to the resolving of side chain positions (Bai et al., 2015a; Vinothkumar and Henderson, 2016). For example, the number of sub 4 Å structures in EMDB has grown from 32 to 112 in recent years (Kühlbrandt, 2014). This improvement in resolution has opened the door to a host of new research directions driven by single particle cryo-EM experiments. One such example is structure-based inhibitor design, a field previously underpinned by X-ray crystallography. However, the recent step change in the capabilities of EM, has seen cryo-EM used to determine the structure of the TRPA1, TRPV1, β-galactosidase, human p97 and the proteasome in complex with their respective inhibitors (Bartesaghi et al., 2015; da Fonseca and Morris, 2015; Gao et al., 2016; Li et al., 2016; Merk et al., 2016; Paulsen et al., 2015). In the majority of these cases high-resolution X-ray crystallographic data were used as a foundation to build atomic models within the EM map.

The versatility of EM for large protein complexes, which are often intractable to crystallization, makes it advantageous over traditional routes for structure-guided drug design in a number of areas. Structural biology can underpin ligand design and it is becoming clear that with advances in EM, obtaining the structure of previously intractable systems is now possible.

5.1 IGPD Structural Studies

Imidazoleglycerol-phosphate dehydratase (IGPD) is an essential enzyme complex, found in both plants and microorganisms, that catalyzes the sixth step in histidine biosynthesis, where dehydration of imidazoleglycerol-phosphate (IGP) occurs to form imidazoleacetol-phosphate (IAP) (Bisson et al., 2015; Gohda et al., 1998) (Figure 5.1). The ~565 kDa catalytically active 24-mer, which exhibits 432 symmetry, is formed from inactive trimers of the apo-protein upon addition of Mn^{2+} (Hawkes et al., 1995; Petersen et al., 1997; Tada et al., 1995; Wilkinson et al., 1995). Given that histidine is synthesized *de novo* by plants, IGPD is an attractive target for herbicides that act after the emergence of the crop, for which there are presently a limited number of biological targets. For this reason IGPD is currently being studied as a new herbicide target.

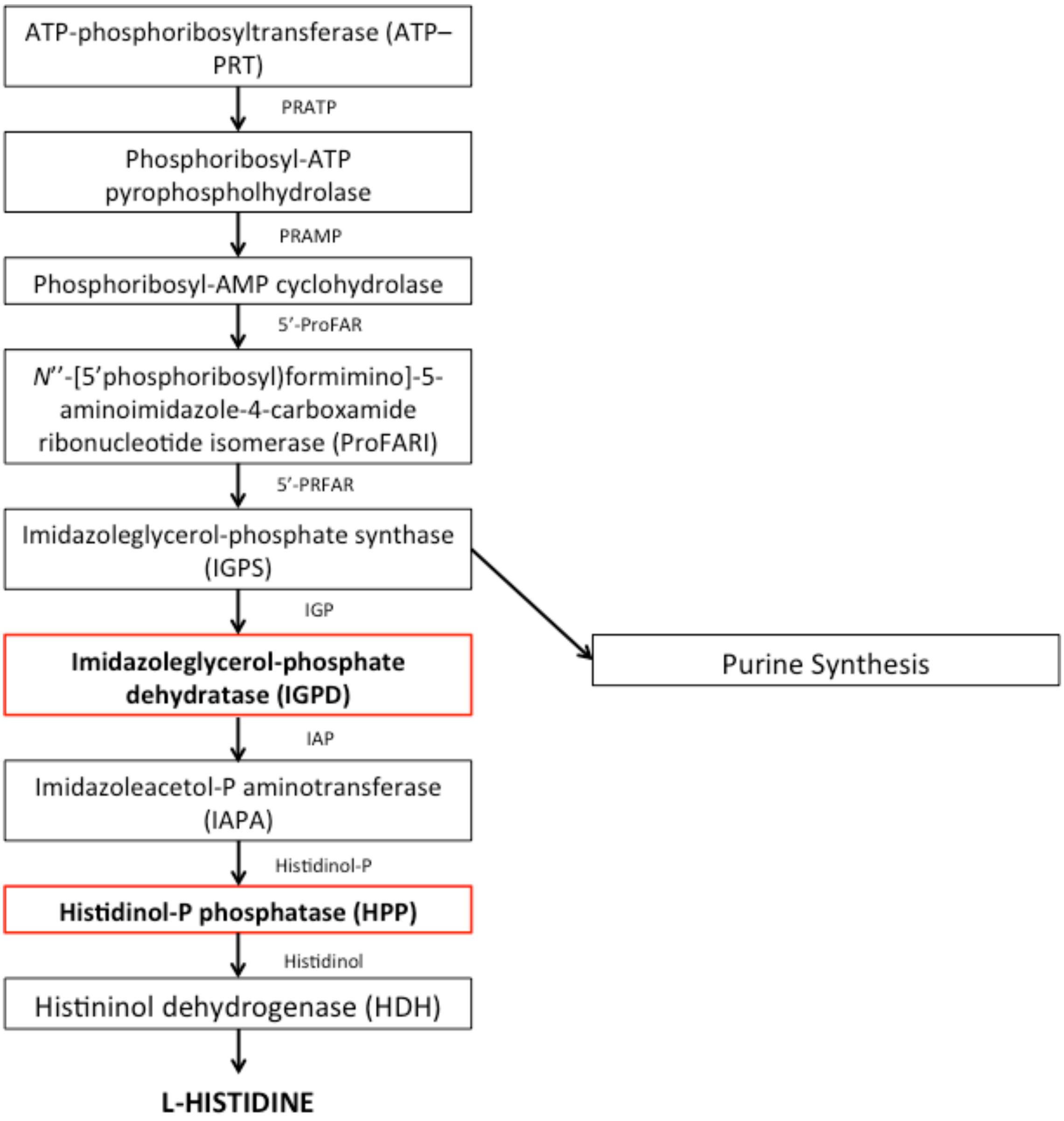


Figure 5.1 - Histidine biosynthesis pathway. Schematic of histidine biosynthesis pathway with IGPD and HPP highlighted. Based on (Ingle et al., 2005).

IGPD was used as a model system to study inhibitor binding via EM due to its high symmetry and stability, both of which were hoped to aid in obtaining the resolutions required to directly visualise inhibitor binding and allow *de novo* model building. In addition, the availability of highly potent (nM) inhibitors was hoped to minimise issues relating to binding site occupancy, thus aiding in locating the inhibitor within the complex. Finally, access to high-resolution crystal structures allowed validation of any binding observed to give confidence in any results obtained (Bisson et al., 2015) (personal communication Dr. Claudine Bisson, Sheffield).

5.1.1 EM Structure of IGPD

IGPD protein from *Arabidopsis thaliana* (*At*) and (2hydroxy-3-[1,2,4]triazole-l-yl-propyl)-phosphonic acid inhibitor was kindly provided by Dr. Claudine Bisson from the University of Sheffield. Cryo grids were prepared in a standard manner using Quantifoil 2.2 grids, which were glow discharged for 30 seconds before use. Following this, 3 µl of protein/inhibitor complex in buffer (50 mM Bis-Tris propane buffer pH 8.0, 50 mM NaCl, 10 mM β-mercaptoethanol, and 0.4 mM $MnCl_2$) was added at ~0.5 mg/ml concentration and the grid blotted and frozen using a Vitribot mark IV (FEI) according to general methods (Chapter 2). Grids were initially screened on a F20 equipped with a CCD camera operating at 60k magnification. The resulting micrographs showed a closely packed distribution with several characteristic square views predominating, as well as other views in areas with thicker ice towards the edges of the holes. Indeed the particles formed semi-ordered layers readily in the ice layer towards the centre of the holes where the ice was thinnest (Figure 5.2). Initially a small number (57) of micrographs were collected and picked semi-automatically using the EMAN2 swarm tool. This dataset of 19351 particles led to a very low resolution cube like structure when processed within

RELION using octahedral symmetry, potentially due to difficulties in alignment of the smooth, relatively unfeatured structure as previously observed in apoferritin EM studies (Figure 5.2) (Henderson and McMullan, 2013; Russo and Passmore, 2014). Additionally, the particles were likely too closely packed due to the magnification used, although recent studies have shown even extremely closely packed particles can achieve high resolution (Liu et al., 2016).

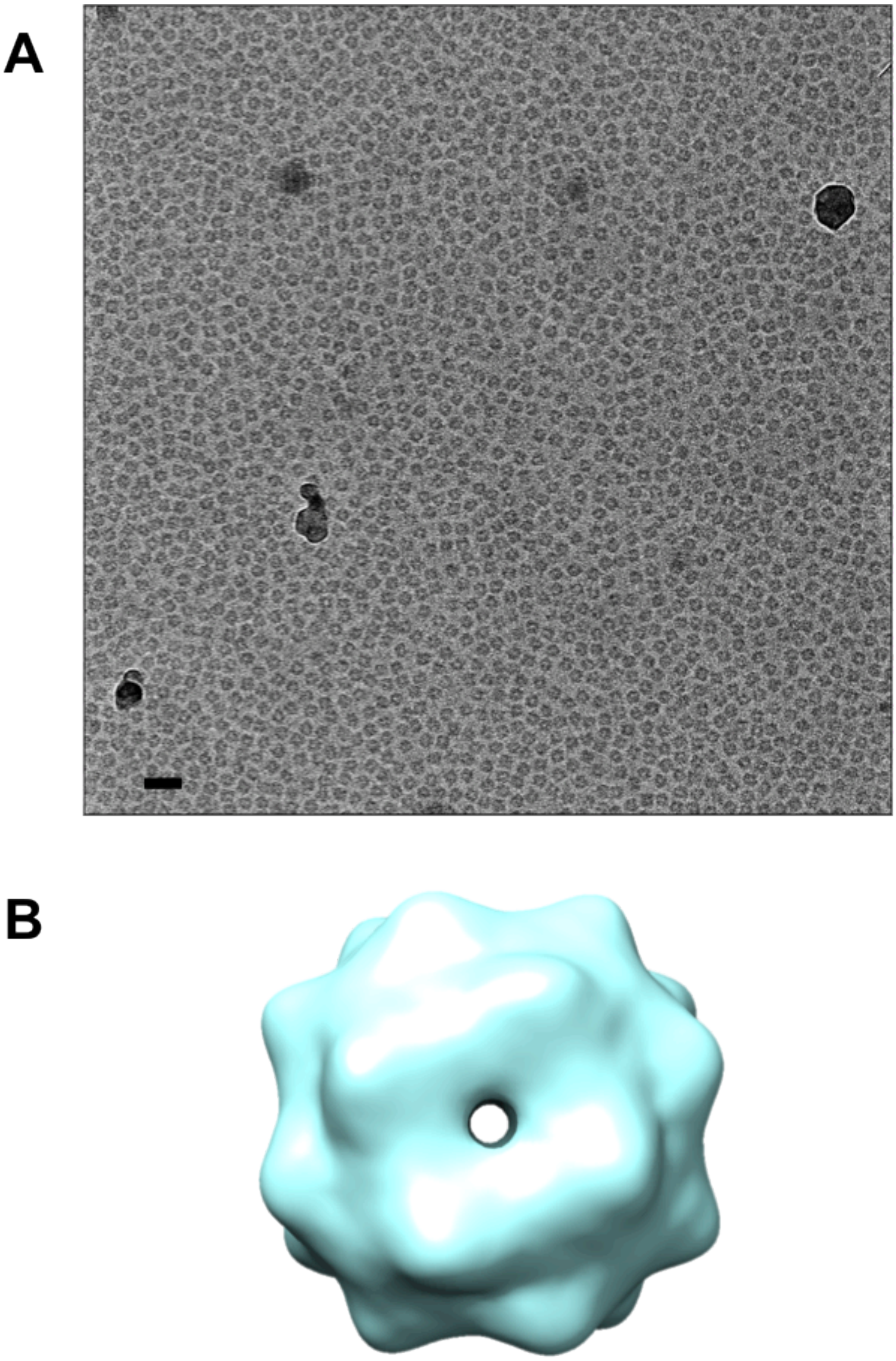


Figure 5.2 – IGPD data from CCD-equipped F20. (A) Micrograph of IGPD taken on a F20 equipped with a CCD (scale bar 200 Å). (B) 3D reconstruction of IGPD showing low-resolution features and no detail.

Following this, a new data set was collected (~80 micrographs) of the same batch of IGPD grids on the same F20 fitted with a K2 DED at 1.8 Å/pixel sampling. The micrograph quality was noticeably higher using the K2 camera, with more detail observed within individual particles (Figure 5.3). Following whole-frame motion correction using MOTIONCORR, 3049 particles were picked. After refinement of the particle stack using the known crystal structure filtered to 40 Å as the starting model, the resulting structure from RELION was post-processed using a mask extending 9 pixels from the model and a 6.6 Å resolution reconstruction was obtained. This showed clear density corresponding to α-helices following B-factor sharpening using an ad hoc B-factor of -350. This was in good agreement with the existing crystal structure of the complex, although this resolution was too low to identify any bound inhibitor. Particle polishing of this small dataset led to no improvement in resolution. 3D classification was also trialled, leading to a 2538 particle subset, but following refinement of these data no improvement in resolution was observed and the resulting map appeared of slightly lower quality. Automated particle picking using the swarm tool within EMAN2 was also trialled to increase the particle number, leading to 33614 particles picked initially. However, 2D classification showed many neighbouring particles in the classes and following further classification no improvement was made in the resulting reconstructions.

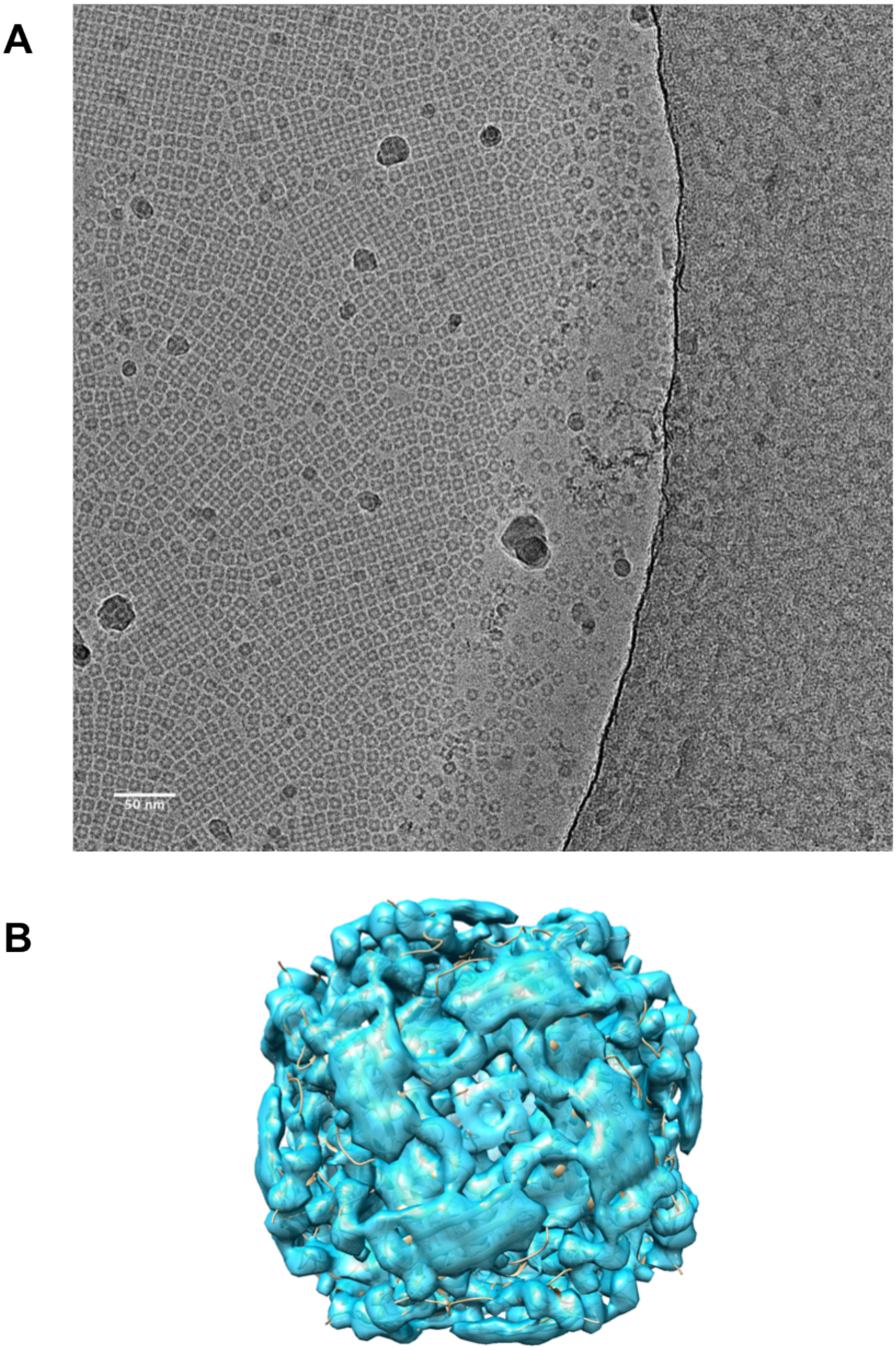


Figure 5.3 - IGPD from F20 equipped with a K2 DED. (A) Micrograph of IGPD taken on a F20 equipped with a K2 DED, scale bar represents 50 nm. (B) 3D reconstruction of IGPD showing secondary structure with flat β-sheets observed on the surface.

It is interesting to note the dramatic effect of motion correction on the quality of the data obtained using the K2-equipped F20. When the same particle co-ordinates are extracted from non-motion-corrected micrographs, the resulting classes and reconstruction lose all detail, with no secondary structure. Indeed, the resulting reconstruction appears similar to the CCD reconstruction obtained on the same microscope (Figure 5.4).

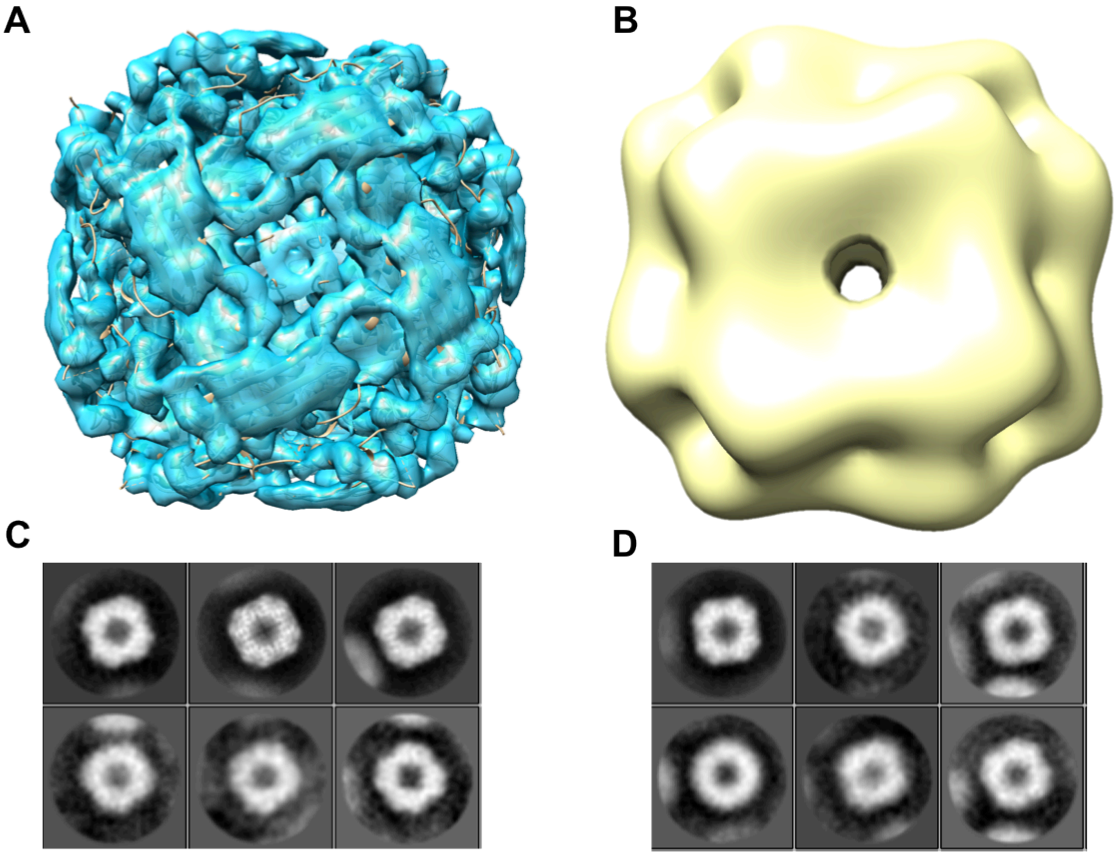


Figure 5.4 - Comparison of aligned and unaligned micrographs. (A) 3D reconstruction using aligned K2 frames and associated classes (C). (B) 3D reconstruction using unaligned K2 frames and associated classes (D).

To push the resolution further, a final dataset was collected using the same grids on a Titan Krios equipped with a Falcon II DED at the Chinese Academy of Sciences, Beijing using SerialEM to collect the data in an automated fashion (Mastronarde, 2005). 682 micrographs at 1.10 Å/pixel sampling were obtained with intentional variation in the area of the hole imaged in order to capture a wider range of views in the thicker ice towards the hole edges (Figure 5.5.A).

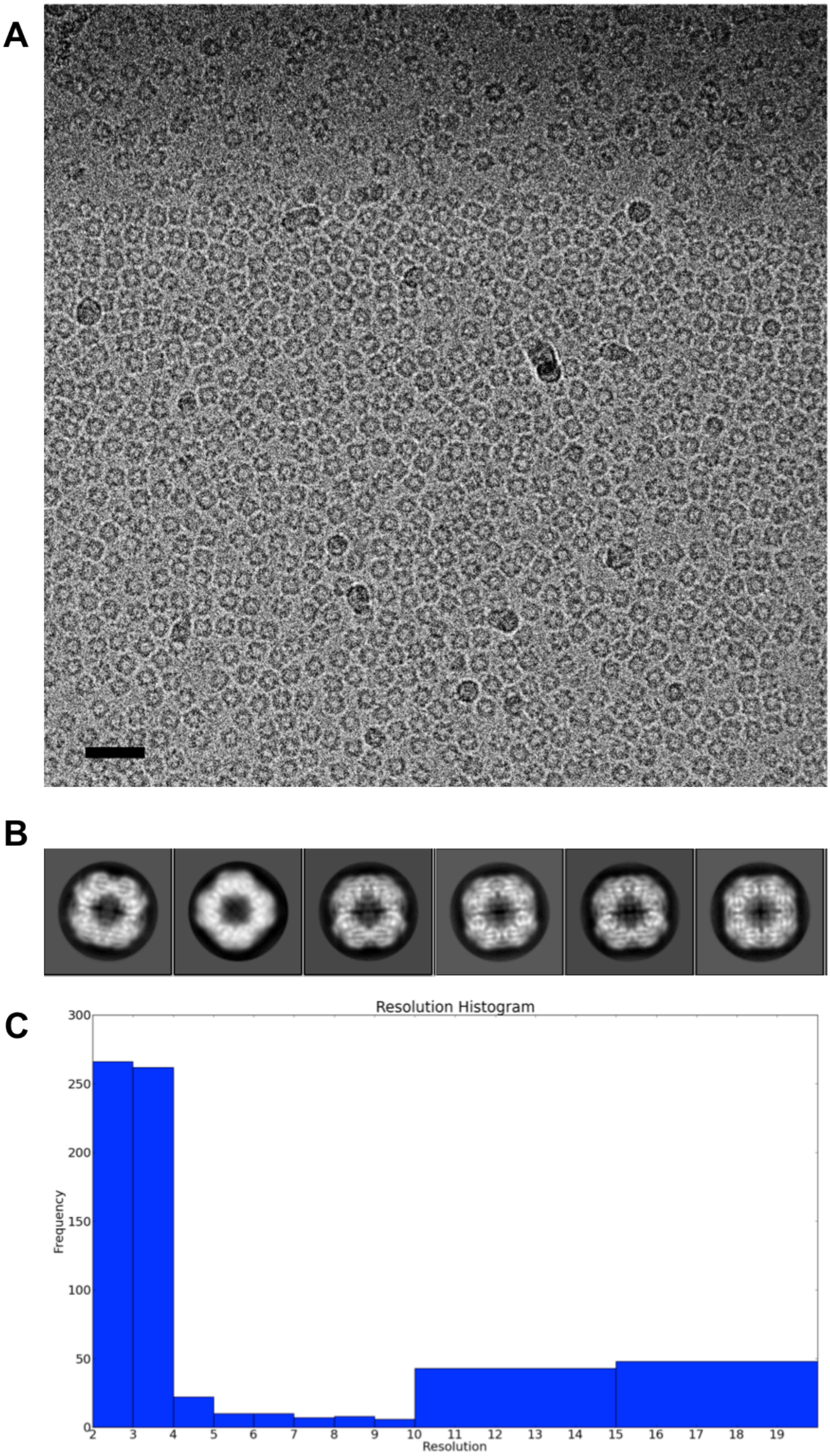

A
B
C
Resolution Histogram
Frequency
0
50
100
150
200
250
300
Resolution
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

Figure 5.5 - IGPD data from Titan Krios. (A) Motion-corrected micrograph from Titan Krios IGPD dataset showing thicker ice and wider distribution of views towards the edge of the hole (scale bar 300 Å). (B) 2D class averages from these data showing clear detail but limited views. (C) Estimated resolution of the micrographs following motion correction with majority <4 Å.

Following semi-automated picking using the swarm tool and manual culling to remove particles too closely packed, 110977 particles were obtained. 2D classification within RELION resulted in well-defined classes (Figure 5.5.B), although views were still limited and poorly aligned classes or classes showing density from neighbouring particles were then removed to leave 83466 particles. At this point a 3D refinement gave a reconstruction at 3.8 Å resolution, showing clear separation of β-sheets and larger bulky sidechains. This dataset was then subjected to further rounds of 2D and 3D classification leaving 55481 particles. Following per-particle motion correction in RELION's particle polishing utility and removing all but frames 5-18 (~27 $e^-/Å^2$ total dose received), the resolution was improved to ~3.4 Å. Frames were removed rather than weighted, as upon inspection of the estimated per-frame B-factors, the expected initial sharp rise and then a steady fall in B-factor was not observed, indicating the weighting was likely incorrect (Figure 5.6.A).

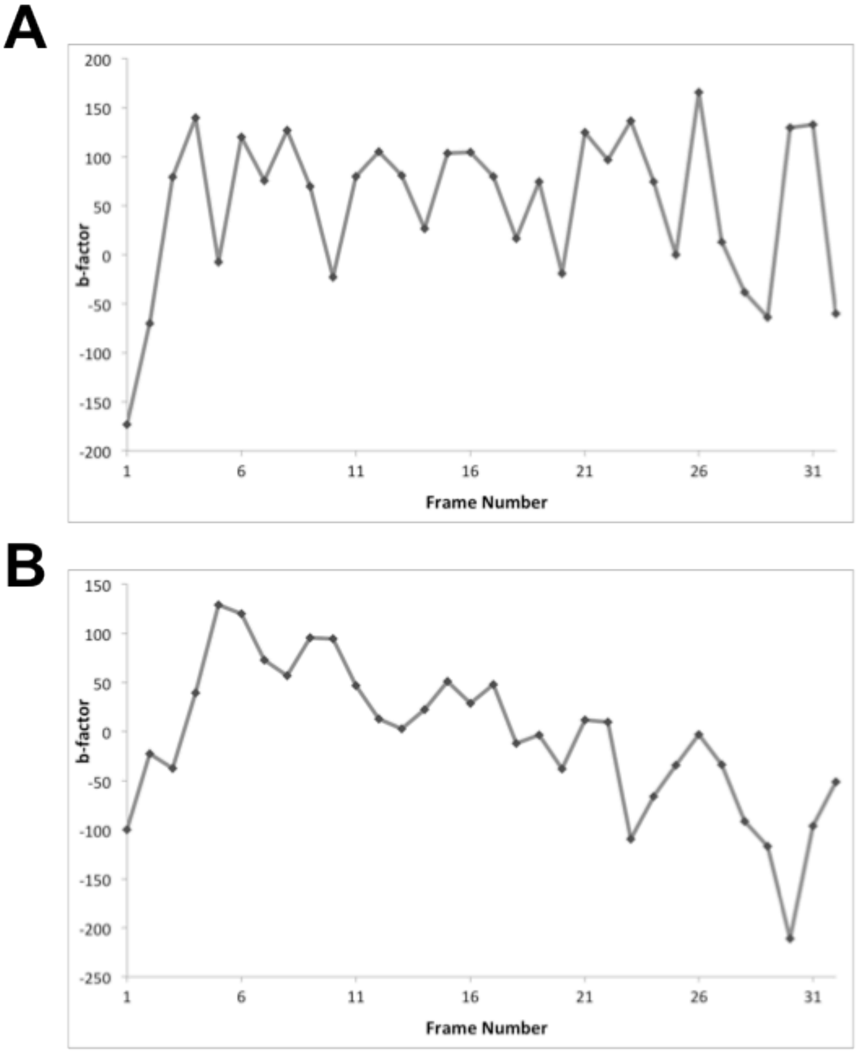


Figure 5.6 – B-factor plots of IGPD data. B-factor plots of IGPD data using no frame averaging (A) showing no clear trend, and 5 frame rolling average (B) giving expected decay curve.

By increasing the number of frames averaged to calculate the B-factor, the expected curve was obtained though this reduces the accuracy of the weighting (Figure 5.6.B). By removing further frames, leaving only frames 3-9 (~14 e^{-}/Å^{2} total dose) the resolution was again improved to ~3.2 Å, causing side chains to be clearly visible. With the rapid developments within the EM field new programs are constantly becoming available that permit small gains in resolution. An example of this is gCTF which can calculate defocus on a per-particle basis (Zhang, 2016). This may potentially mitigate any changes in defocus across the micrograph due to the grid not being flat or subtle differences in the particles z-position within the ice layer. The subsequent use of gCTF produced a small increase in resolution to ~3.1 Å although no clear differences between the maps were visible.

By using three different camera/microscope combinations several IGPD reconstructions were solved across a wide resolution range, illustrating the power of new developments in microscope stability and crucially the role of DEDs (Figure 5.7).

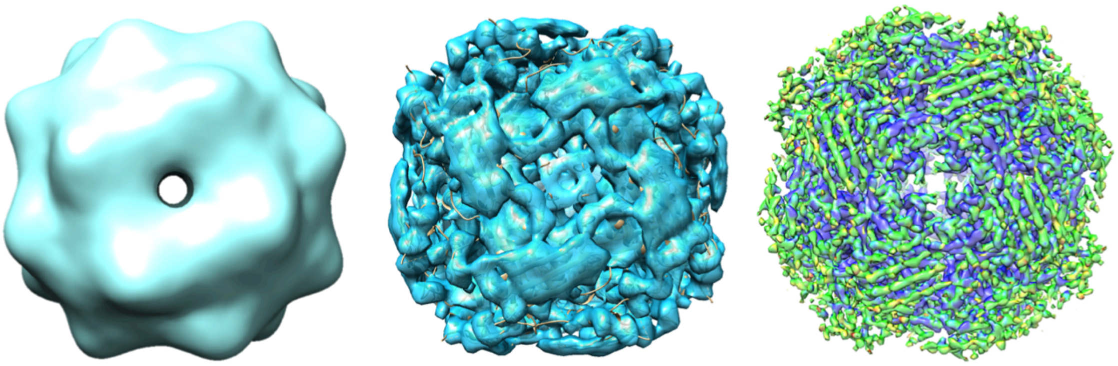


Figure 5.7 - IGPD comparison. Comparison of IGPD reconstructions from CCD camera on F20 (left), K2 on F20 (centre) and Titan Krios with a Falcon II (right).

5.1.2 IGPD Model Building and Fitting

Initially the existing crystal structure of IGPD was flexibly fitted into the ~6.6 Å IGPD map using MDFF (Trabuco et al., 2011). Secondary structure showed good agreement between the map and model, with α-helical regions clearly defined and the model was clearly traceable through the map.

To demonstrate how EM can be a powerful tool for the determination of ligand-protein complexes of unknown structure, *de novo* model building was conducted in Buccaneer and Coot using the IGPD 3.4 Å EM map (Cowtan, 2006; Emsley et al., 2010). To reduce computational time, the map was segmented in Chimera to isolate one of the 24 subunits, guided by the clear density of the external β-sheet. Buccaneer was run on the segmented map, with the resulting model clearly identifying four helical components and 6 β-strands (Figure 5.8).

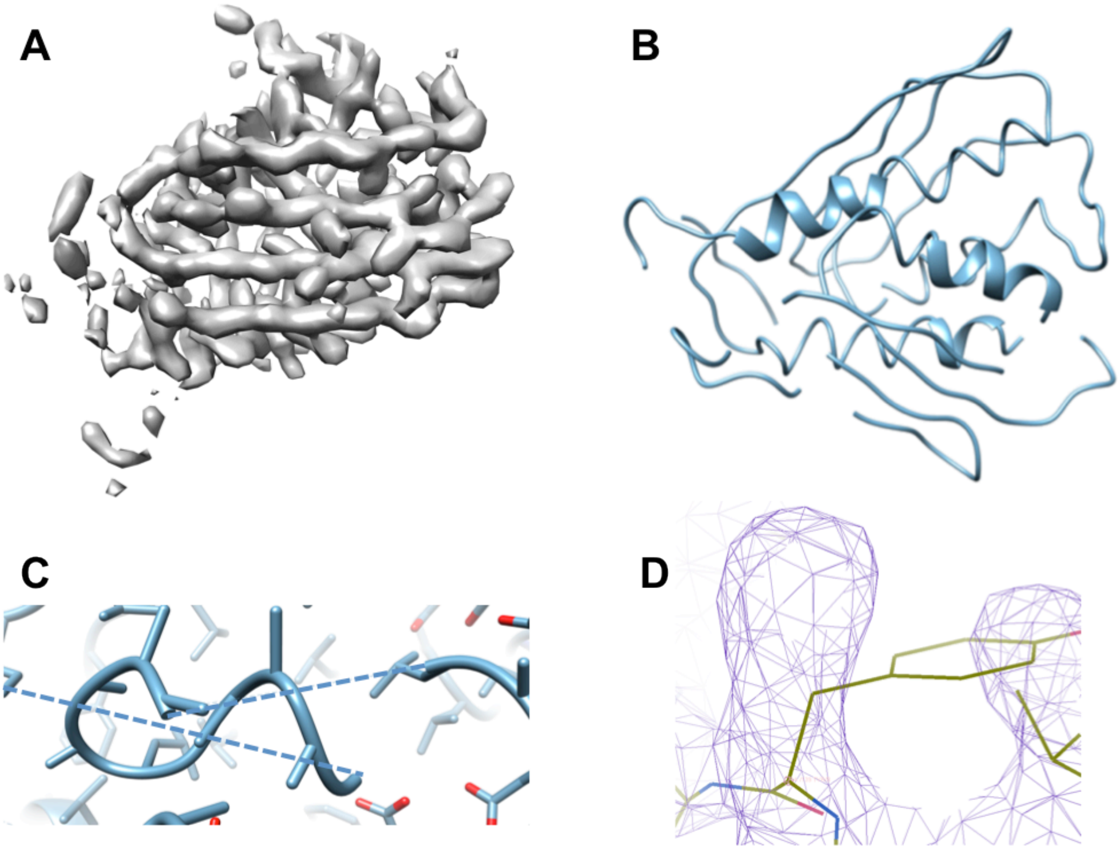


Figure 5.8 - Map segmentation and initial Buccaneer build. (A) Segmented monomer density used for the model building. (B) Initial model obtained from Buccaneer showing some helices assigned with several breaks in connectivity. (C) Example of automatically built model positioned backwards. (D) Example of poorly fitting automatically assigned sidechain.

Independently density-guided *de novo* model building was carried out within Rosetta3 (DiMaio et al., 2015; Wang et al., 2015). This process consists of breaking the full sequence into a series of overlapping short fragments, based on the common conformations these short sequences are known to adopt. These fragments are then placed into the density and scored. This process is then iterated using the positions of highly-scoring fragments to guide the placement of subsequent parts based upon the sequence. Finally a consensus map is created from these assembled and scored fragments, which can then be used for further iterations or manual building/refinement. In the first instance, this produced a consensus model containing the four core α-helices with side chains appearing well-matched to the density. Following several iterations of the process one of the four-strand β-sheets was also well-placed within the density by reducing the search space through subtracting the previously built model from the EM density (Figure 5.9).

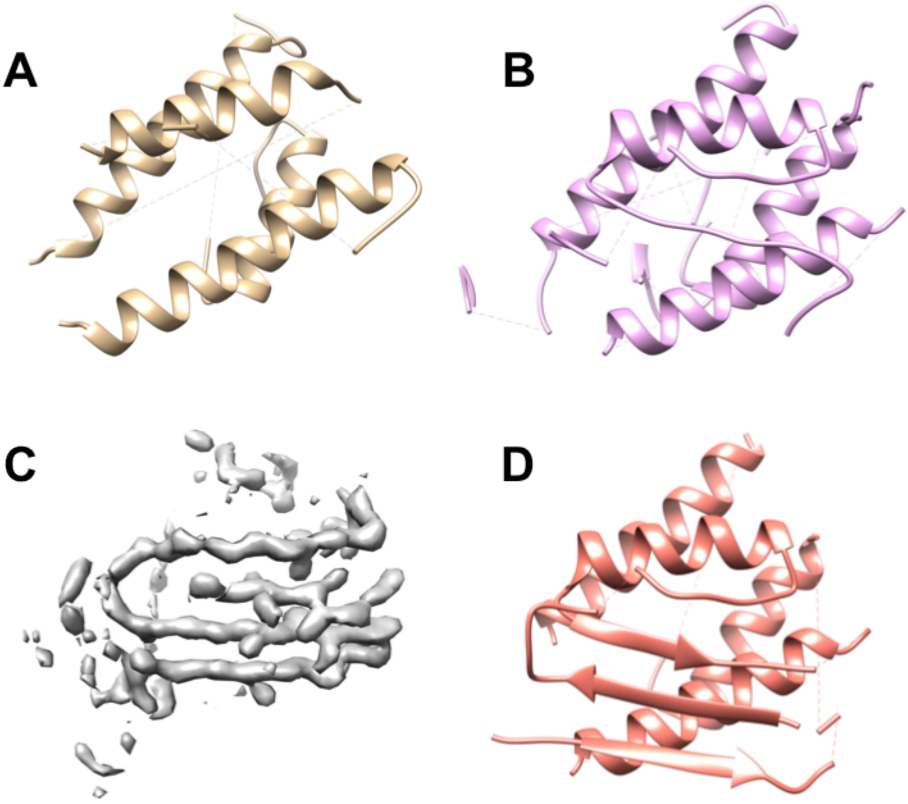


Figure 5.9 – *De novo* model building within Rosetta. (A) Initial round of Rosetta building showing core helices mostly assigned. (B) Third round of iterative Rosetta assignment with β-sheets starting to be built. (C) β-sheet density remaining following subtraction of built model from experimental density to reduce fragment search space. (D) Rosetta model following 5 rounds of building with core helices and one β-sheet well assigned.

Both models from Buccaneer and Rosetta3 corresponded well with the predicted secondary structure as calculated in Phyre2 (Kelley et al., 2015), independently of the known crystal structure. Visual inspection in Coot was subsequently used to edit clear errors in the initial model produced by Buccaneer, such as poor chain tracing and connectivity errors (Figure 5.10.A). Several areas of the sequence also appeared to be placed backwards into the density meaning extensive rebuilding was required. The sequence was assigned in Coot using both the EM side chain density and predicted secondary structure as a guide, although loop regions and the outer β-sheet proved challenging to model with confidence. The resulting model contained two β-sheets, flanking a four-helical bundle (Figure 5.10.B). Analysis of the bond angles and Ramachandran plot showed good overall geometry (~93% favourable, ~5% allowed, ~2% outlier).

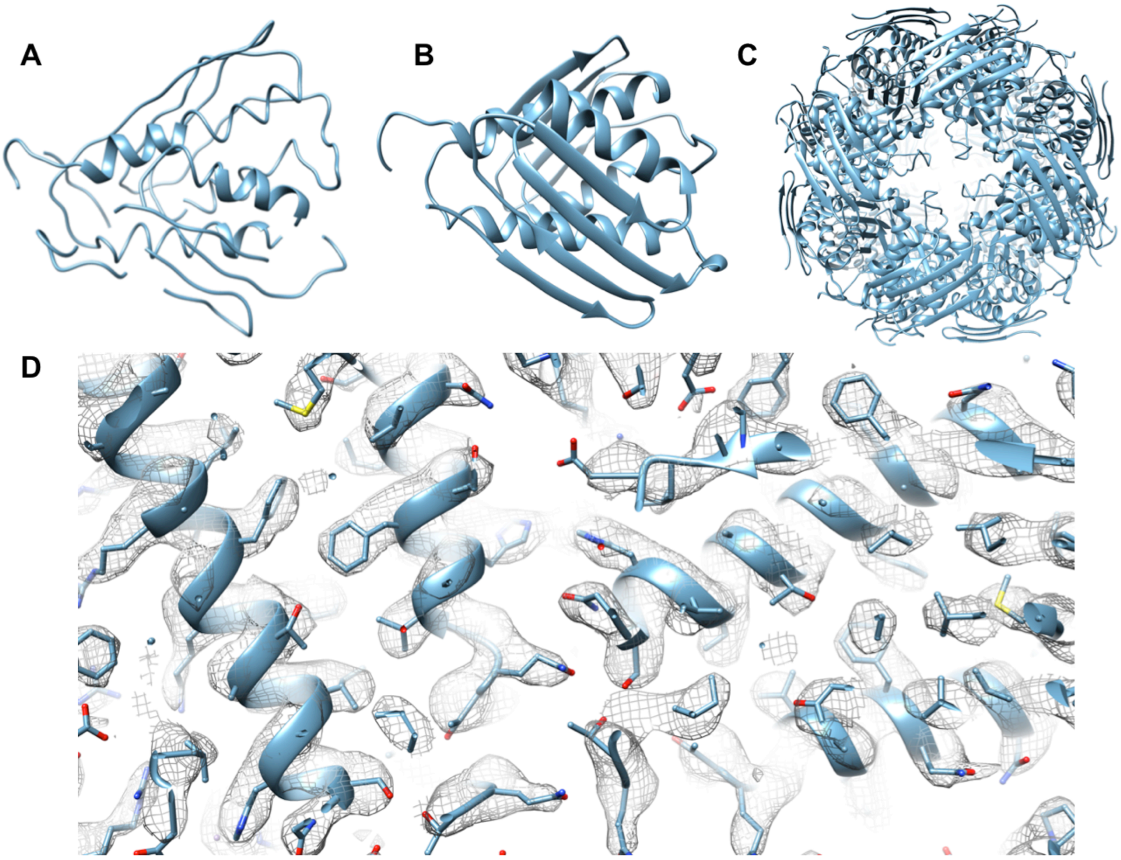


Figure 5.10 - Refined IGPD model and representative density. (A) Automatically built model from Buccaneer. (B) Refined monomer from REFMAC5 showing correct connectivity and assigned secondary structure. (C) Refined complex model into complete map. (D) Representative density following REFMAC5 refinement showing clear side chain density(Brown et al., 2015; Murshudov et al., 2011).

5.1.3 IGPD Inhibitor Docking and Visualisation

To identify any regions of unexplained density within the IGPD map, which might account for the bound inhibitor, the built model was filtered to 3.5 Å and a difference map between the 3.5 Å EM map and the *At*IGPD2 *de novo* built model was generated within Chimera using the 'vop subtract' command.

Strong positive density in the difference map was located within the dimer interface with weaker density flanking this region (Figure 5.11). This weaker feature in the difference map represents the C-terminal region of the polypeptide, identified after comparison to the X-ray crystal structure, discussed in Section 5.1.4.

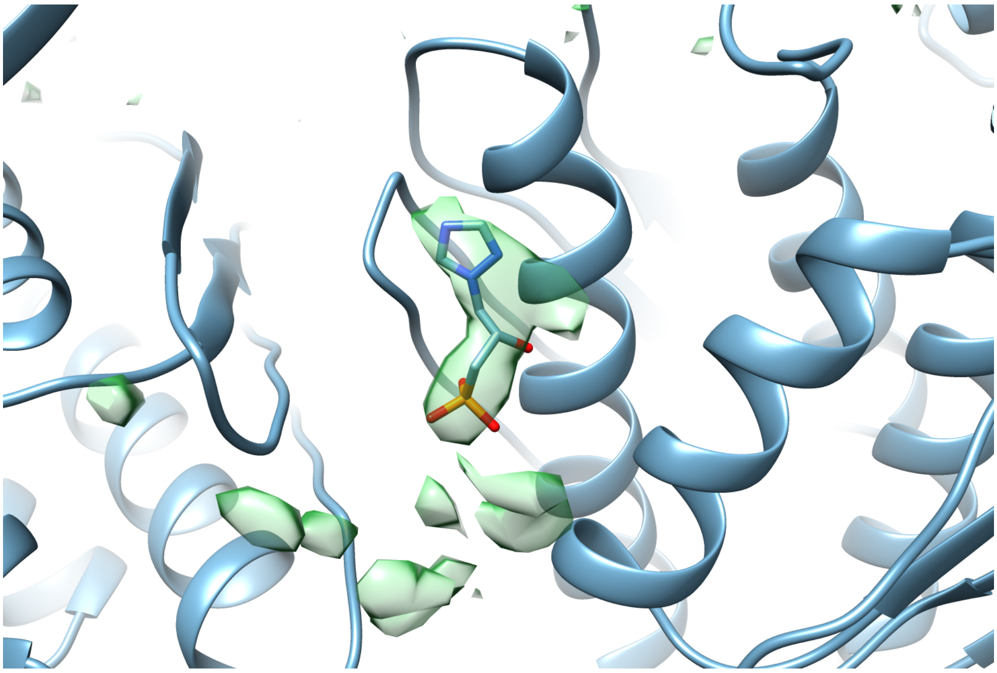


Figure 5.11 - Use of difference density to identify inhibitor binding site. Difference density (green) between symmetrised built model and experimental density showing good agreement with the size and shape of the bound inhibitor.

Strong density is seen in both the cryo-EM map and difference map within the heart of the dimer interface, which is flanked by a number of His residues at the top of the pocket and positively charged residues at the base. Calculation of the local resolution in Resmap (Kucukelbir et al., 2013), showed this region of positive density to be one of the higher resolution areas within the reconstruction, at approximately 2.6 Å local resolution. Therefore a new map was generated which was filtered to 2.6 Å and used to analyse the region of unaccounted-for density. Contouring of this map showed two strong density peaks within the pocket ~2.5 Å away from the histidine residues within the binding pocket (Figure 5.12), features consistent with the manganese ions that co-purify with the protein. However, it is not possible to unambiguously assign these density peaks to manganese in a standard EM experiment, although complimentary techniques including scanning transmission electron microscopy (STEM) could be used to identify the locations of metal ions and electron energy loss spectroscopy (EELS) or energy-dispersive X-ray spectroscopy (EDX) could be utilised to show the elements present. Thus biochemical information relating to the role of the Mn^{2+}, known co-ordination geometry and distances are required to assign this density.

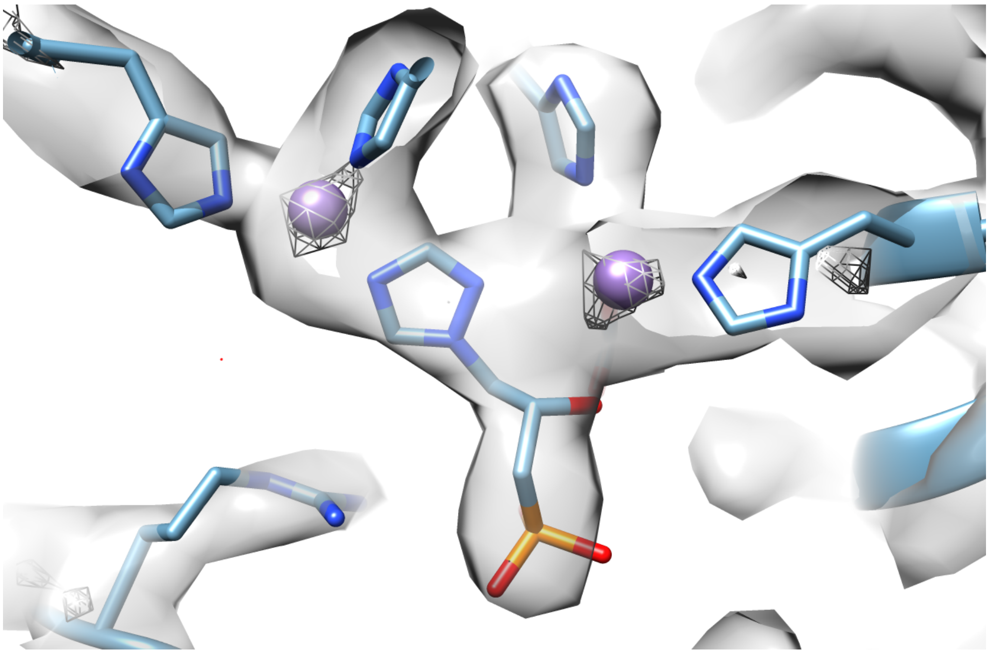


Figure 5.12 - Inhibitor binding site with Mn^{2+} density peaks. Inhibitor binding site showing fit of refined atomic model (blue) into experimental density (grey). Additional density from EM density map filtered to 2.6 Å shown in mesh is in good agreement with Mn^{2+} positions.

After fitting the metal ions, the remaining density was sufficient to accommodate the triazole-phosphonate inhibitor, with the three ring nitrogens occupying a position within the coordination sphere of each of the bound metal ions (Figure 5.12). Moreover, the map showed strong density consistent with the negatively charged phosphate group of the triazole-phosphonate inhibitor, which is located within a cavity surrounded by the positively charged residues (Figure 5.13.B). However, the resolution was not sufficient to assign the pose or stereochemistry of the inhibitor, but was sufficient to assign the approximate orientation and location. To further investigate the finding *in silico* molecular docking using SymBiosis' eHiTs software was performed by Dr Martin

McPhillie as a form of map-independent validation, using only the *de novo* built model and no information from the EM map (Zsoldos et al., 2007). The docking for the S-isomer was consistent with that of the manually-modelled inhibitor within IGPD, with the top-ranked pose suitably positioned for the trizole group for interaction with two Mn^{2+} ions, and the phosphate with the positively charged pocket. The docked inhibitor was also in good agreement with the EM density, despite this not being used as a constraint during the docking protocol.

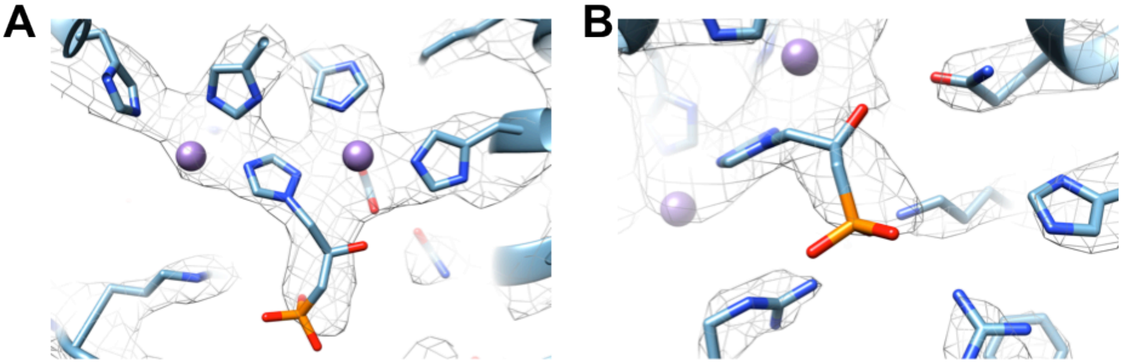


Figure 5.13 - Position of inhibitor within EM density. Refined atomic model and inhibitor (blue) shown in EM density (grey mesh). (A) Triazole ring of inhibitor positioned so coordination with Mn^{2+} is maintained. (B) Phosphonate moiety of inhibitor positioned in close proximity to positively charged sidechains giving confidence to assigned inhibitor pose.

The fit of the *in silico* docked inhibitor within the EM density, which was not used as a constraint within the docking program, gave a clear validation of the results. As such, even in structures where the resolution does not allow the unambiguous determination of the inhibitor pose, valuable information can be extracted through an integrated approach utilising computational docking.

Following initial model building in Coot, and addition of the fitted ligand with co-ordinating Mn^{2+} into the model, refinement of the model into the map was then carried out in REFMAC5 (Brown et al., 2015; Murshudov et al., 2011). Initially, refinement was carried out on a monomer, into the segmented monomer density used for the initial building. However, this did not properly reflect the co-ordination of the bound inhibitor at the interface between monomers. Thus refinement was performed using a symmetrised initial model into the full experimental map by Daniel Hurdiss, a rotation student under my supervision (Figure 5.10.C). Following refinement on the whole complex a R-factor of 0.39 was obtained and the inhibitor co-ordination well placed with respect to the surrounding histidine and Mn^{2+}.

5.1.4 Validation and comparison to X-Ray Structures

The resulting structure was validated by comparing it to a 1.1 Å X-ray crystal structure of the same complex (PDB 5EKW). Overall the two structures agree very well, with an overall RMSD value of 0.73 Å. Minor differences of ~1 Å RMSD are concentrated within the loop and terminal regions (Figure 5.14). This could be reflective of the minor changes seen between a protein in the crystalline state and the frozen hydrated state in cryo-EM, as well as errors in the model building due to the modest resolution.

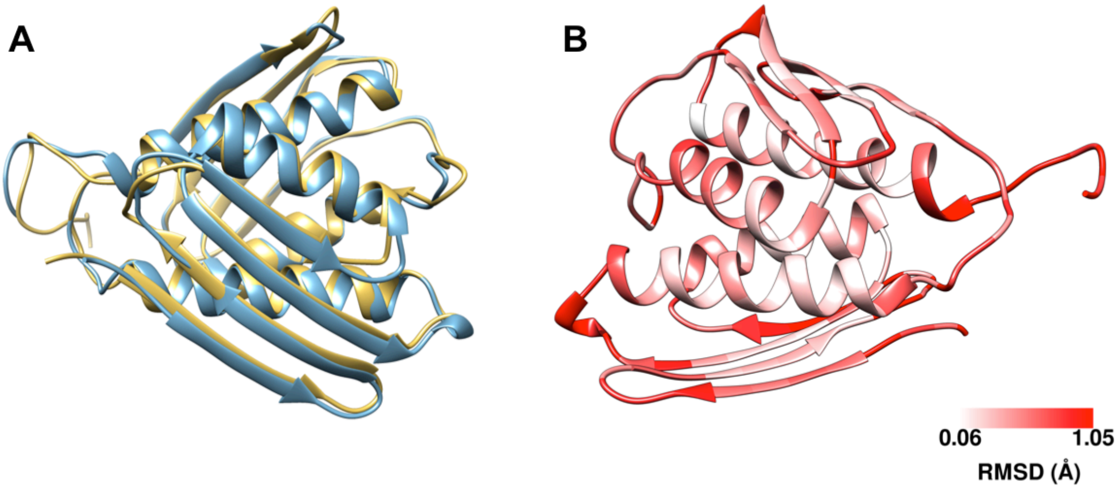


Figure 5.14 - Comparison between crystal and EM models. (A) Superposition of crystal structure (PDB 5EKW) (gold) and EM derived model (blue) showing good general agreement between the models. (B) IGPD model coloured by C-α RMSD showing the majority of differences are concentrated in terminal or loop regions.

Comparisons of the inhibitor binding region show high similarities with regards to the positioning and co-ordination of the bound inhibitor (Figure 5.15). Minor differences are seen between the two structures within the binding site, with minor shifts of the inhibitor and bound Mn^{2+}. These differences are most likely due to the disparity in resolutions. However, it is possible that this is through subtle differences within the EM map compared to the crystal structure, and not the modelling of IGPD or fitting of the inhibitor.

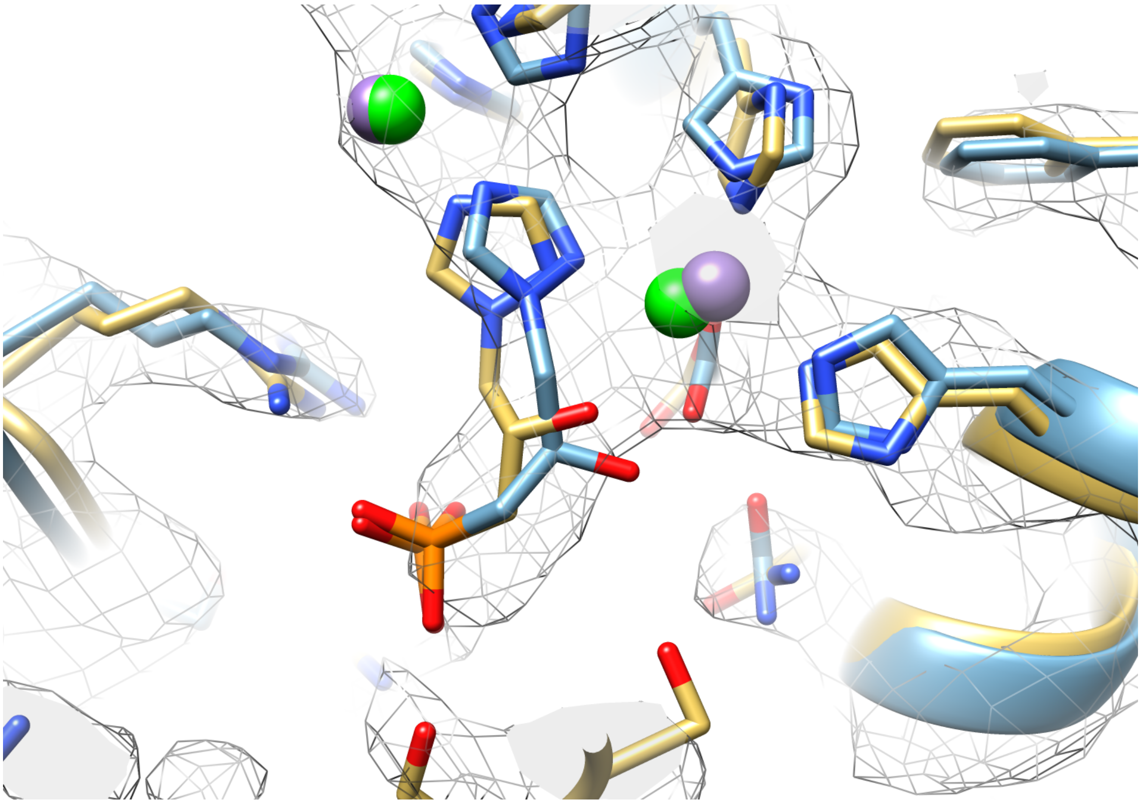


Figure 5.15 - Comparison of binding sites in EM and X-ray models.

Comparison of crystal structure (PDB 5EKW) (gold) and EM derived model (blue) showing minor differences in inhibitor pose despite near identical sidechain positions. However, rotomers cannot be accurately assigned due to the resolution of the EM density.

Another possible cause of the difference in inhibitor position is the level of radiation damage in both structures. As in EM negative side chains are known to degrade even after extremely low electron doses (Allegretti et al., 2014; Grant and Grigorieff, 2015), the model that is actually obtained is actually a high-resolution structure of a damaged, potentially altered protein. Indeed, when viewed at the same contour level, negatively-charged side chains within the map are much weaker than the aliphatic or aromatics ones, though when a reconstruction is produced from a lower dose subset of the data, some of this density does return (Figure 5.16). This differential damage may explain the subtle shifts in the inhibitor position as two Glu (E77 and E173) residues are present in the binding site, potentially stabilising the Mn^{2+}. In the EM structure there is little density for these side chains, so movement/damage could cause a shift in the Mn position, in turn shifting the inhibitor, which is co-ordinated by these ions.

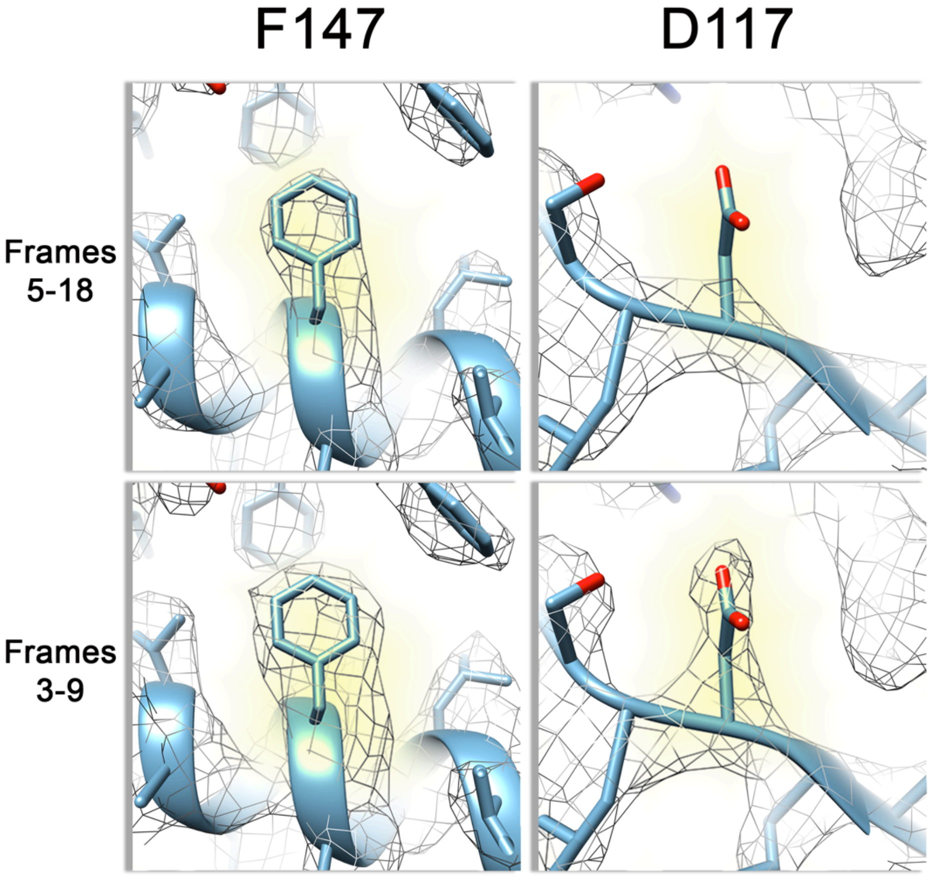


Figure 5.16 - Comparison of side chains at varying dose levels. Comparison of negatively charged side chain D117 density (grey mesh) from reconstructions using frames 5-18 (~ 27 e$^-$/Å^2 total dose) and frames 3-9 (~14 e$^-$/Å^2 total dose). F147 shown as reference for contour level used and general side chain density.

However, it is worth nothing that there are no major changes in the density around the inhibitor binding site (Figure 5.17). The lower dose map shows the same overall pocket of density for the inhibitor, although the density for the phosphonate group is perhaps slightly stronger.

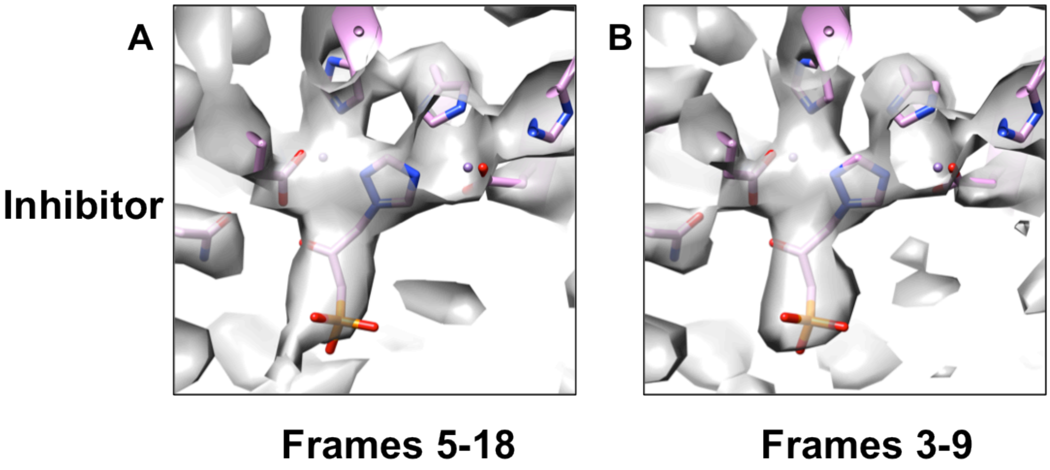


Figure 5.17 - Comparison of inhibitor binding site at varying dose levels. (A) Inhibitor binding site with density in grey at higher dose showing minimal difference from the lower dose map (B).

The crystal structures of IGPD have shown that the C-terminal 15 residues form a loop that closes down on the substrate (or inhibitor) as they bind, which is disordered in the unbound form (Bisson et al., 2015). There is clear density around this region in the EM map (Figure 5.18) and the corresponding difference map, but it is not of sufficient quality to permit the unambiguous modelling of this region, and likely represents a more mobile area of the structure (Bisson et al., 2015). While 3D classification was carried out on the IGPD data set, this binding loop was not subjected to focussed classification, so it is possible that a subset of particles exists with a more ordered loop region. Both the crystal structure and EM structure were solved in the presence of a mix of R and S forms of the inhibitor, which surprisingly display similar binding affinities (Bisson et al., 2016). Within the crystal structure this results in little difference in the position of the triazole or phosphate groups but does result in changes to the C3 position and hydroxyl group, though the resolution of the EM map does not allow this subtle change to be observed.

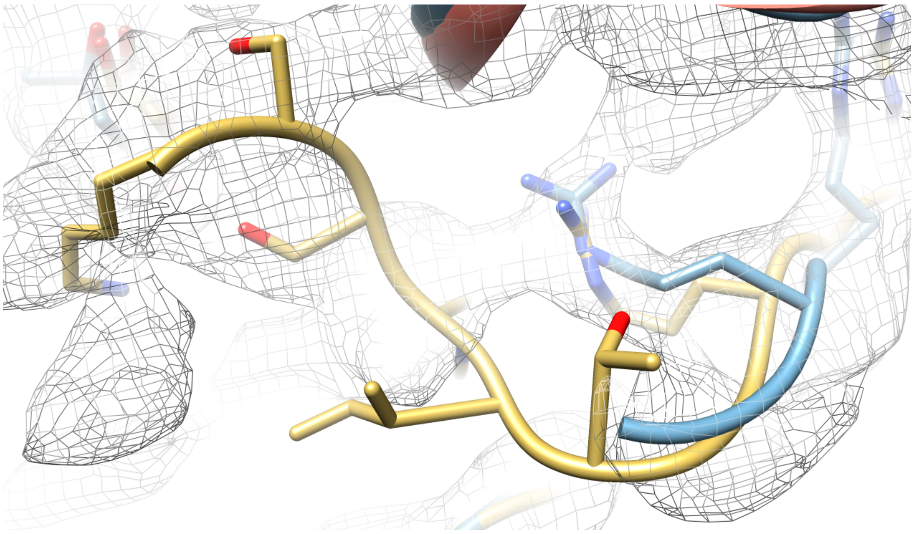


Figure 5.18 - Terminal differences between X-ray and EM IGPD models. Comparison between X-ray (PDB 5EKW) (gold) and EM (blue) C-terminal loop region of IGPD model showing additional residues built into X-ray map.

Initial image processing of the Titan Krios dataset and refinement of the resulting atomic model (with assistance from Garib Murshadov) was carried out under my supervision by a rotation student, Daniel Hurdiss. Inhibitor docking was carried out by Dr Martin McPhillie, Leeds.

5.2 HisB Structural Study

While in plants IGPD is expressed as a single protein that catalyses the 6th step in histidine biosynthesis, in *E. coli* a fusion protein is used. This larger complex (~900 kDa), HisB, combines the C-terminal (residues 167-355) IGPD core (6th stage of the pathway) with a N-terminal (residues 1-166) histidinol phosphatase (HP) domain, which carries out the 8th step in histidine biosynthesis (Figure 5.1) and is suggested to still form a 24-mer (Rangarajan et al., 2006). Although high-resolution crystal structures of the IGPD core and the truncated HP domains are available (PDB 2FPU) (Rangarajan et al., 2006), the larger HisB complex has proven challenging for X-ray crystallography studies since well diffracting crystals could not be obtained. As such, this protein was studied via EM in order to decipher the arrangement of the HP domains and to compare the IGPD core structure with the known plant homologue which has ~50% sequence identity (Rangarajan et al., 2006). The ability to differentiate between the plant and bacterial IGPD complex may provide insights into the design of more selective inhibitors.

5.2.1 HisB Structure

HisB protein samples were provided by Dr. Claudine Bisson at the University of Sheffield and the quality of sample assessed via negative stain EM. Briefly, grids were made by applying ~20 μg/ml HisB protein (2mM $MnCl_2$, 0.1M KCl, 50mM Hepes pH 7.5) onto a glow-discharged carbon-coated grid which was subsequently negatively stained as described in Section 4.1.2.2. Grids were assessed on a Technai T12 microscope, fitted with a 2k Gatan CCD and operating at a nominal magnification of 30k. The protein appeared relatively monodisperse and homogeneous, although some aggregation was seen, and as such was taken into cryo studies (Figure 5.19).

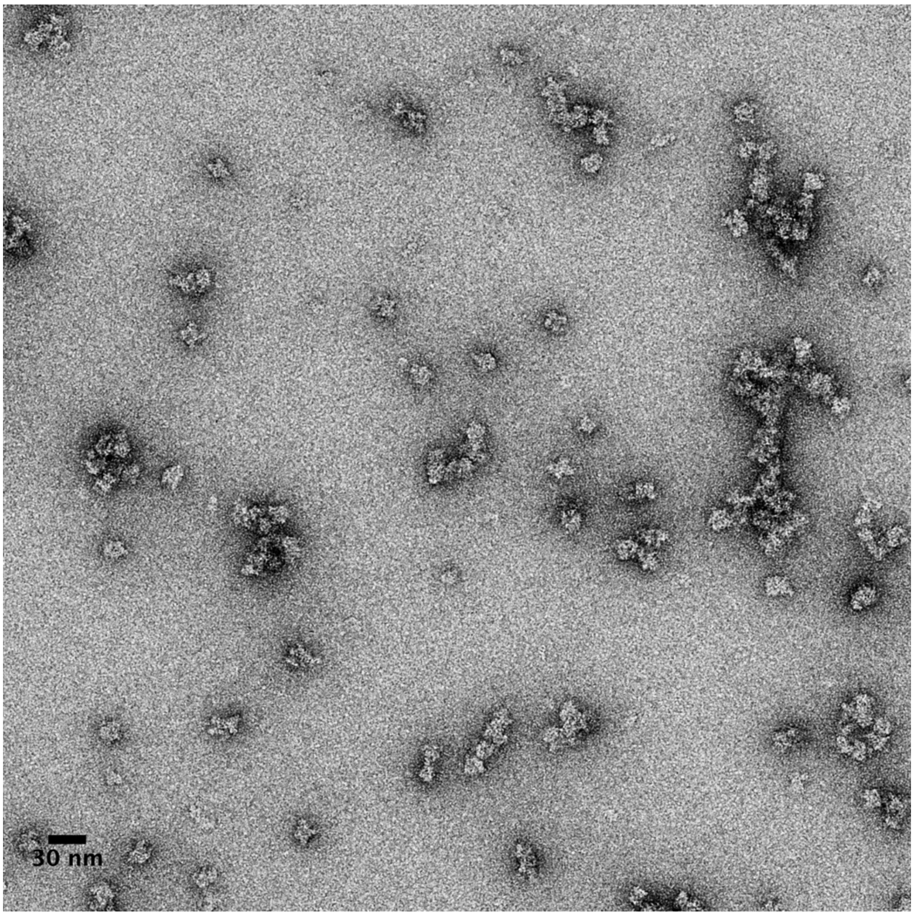


Figure 5.19 - Representative negative stain micrograph of HisB. Negative stain micrograph of HisB showing individual particles with minimal aggregation.

Grids were prepared using standard methods (Chapter 1.6) on 2:2 quantifoil grids following 30 s glow discharge using a Vitribot IV and screened using a F20 equipped with a CCD. Again the protein appeared relatively well-distributed and monodisperse. Data collection was then carried out using a Titan Krios equipped with a Falcon II at LMB Cambridge. 843 micrographs were collected at 1.35 Å/pixel sampling over a 2 s exposure spread across 34 frames (Figure 5.20).

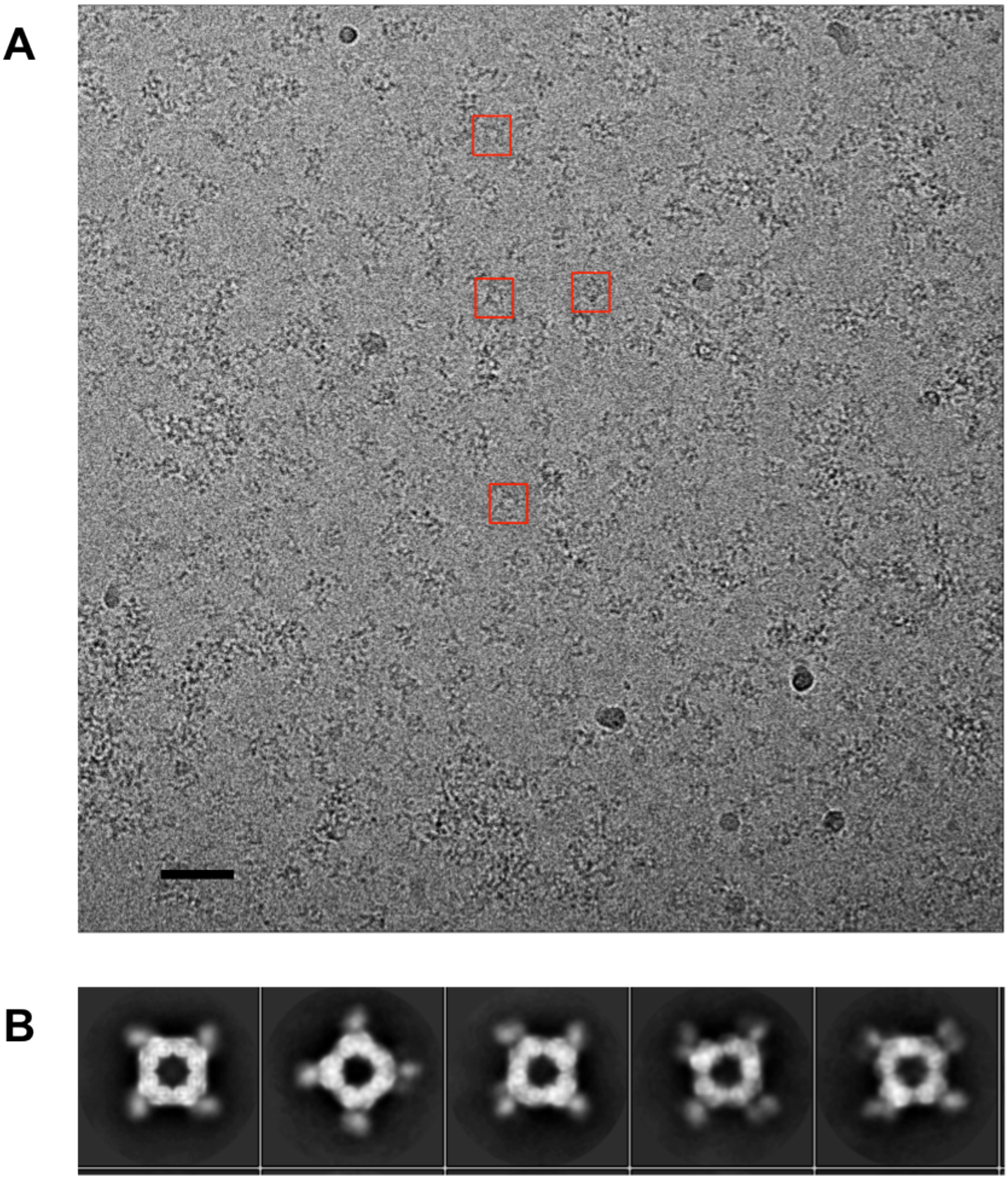


Figure 5.20 - HisB raw data from Titan Krios. (A) Motion-corrected micrograph of HisB collected on Falcon II equipped Titan Krios with example particles boxed in red (scale bar 400 Å). (B) Reference-free 2D classes of HisB showing characteristic IGPD square core with additional less well-defined density on corners.

Whole frame motion correction carried out in MOTIONCORR and manual particle picking was then carried out in EMAN2 and CTF parameters estimated in CTFFIND3. Initial particle picking gave 35654 particles. An initial model with octahedral symmetry was obtained in EMAN2 following 2D classification and used for initial 3D refinement within RELION. Following iterative rounds of 2D and 3D classification to remove particles belonging to poorly defined classes, the particle number was reduced to 15765, and a reconstruction at 6.9 Å resolution was obtained. The resolution was then increased to 6.7 Å following the RELION particle polishing procedure. Again, upon inspection the B-factor plots showed a random distribution, suggesting there was not enough signal in a single frame to robustly estimate B-factors. Frames 5-18 were then used in the reconstruction process with no B-factor weighting applied resulting in a 6.4 Å reconstruction. ResMap and localfsc showed a strong variability in local resolution across the complex (Cardone et al., 2013), the IGPD core having large regions of ~4-5 Å density and the peripheral HP domains at a much lower resolution of ~7-10 Å (Figure 5.21). This is likely a result of heterogeneity and flexibility in the HP domain, indeed similar smearing of mobile peripheral features has been observed in other systems including pyruvate dehydrogenase (Milne et al., 2002).

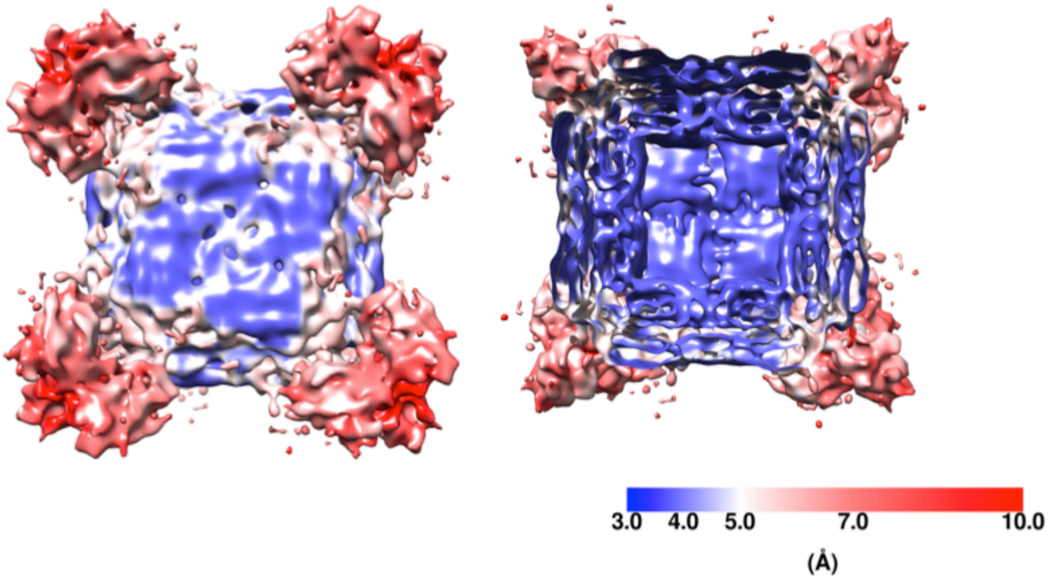


Figure 5.21 - Local Resolution of HisB reconstruction. Local resolution of HisB complex calculated in ResMap showing well-resolved core and poorly defined peripheral domains on the corners, seen as a surface (left) and a slice through section (right)

The IGPD core was then masked and re-refined using only small local searches in the procedure previously used for the V-ATPase (Section 3.2.5). Following this focussed refinement, a IGPD core structure was obtained at 4.5 Å resolution following post processing. In this map β-sheet separation was beginning to appear as expected at this resolution range. The resolution in this region suggests that the IGPD core is not significantly destabilised in the fusion protein relative to the isolated core enzyme, and that the overall resolution of the complex is being severely affected by the highly mobile HP domains (Figure 5.22).

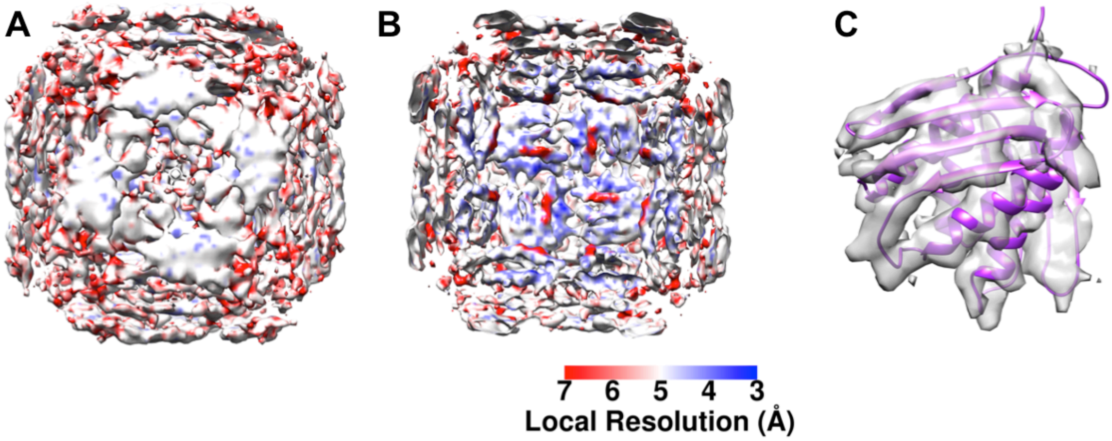


Figure 5.22 - Local resolution of HisB core domain following masking and focused refinement. (A & B) IGPD core of HisB coloured by local resolution values from ResMap following focussed refinement of the core domain. (C) Segmented monomer with fitted atomic model showing β-strands beginning to resolve.

However, focussed classification and refinement of the HP domain did not lead to any noticeable improvement in this region, with no additional detail observed and resolution values ~10-15 Å in these areas. As previously discussed in Section 3.2.5, this could be due to the degree of heterogeneity in these domains in addition to any flexibility between the core and this region.

Signal subtraction as developed in the γ-secretase study (Bai et al., 2015b) was also trialled on both the IGPD core and the HP peripheral domains. Following subtraction of the HP domains, no alternate conformations of the IGPD core were observed and a reconstruction of the core region led to a 5.1 Å reconstruction similar to the reconstruction obtained from masked refinement.

However, no improvement was observed to the HP domain density following this procedure and no alternate conformations were observed. It is also worth noting that when a starting model lacking density for the HP domains was provided, density for these peripheral domains returned following the refinement procedure.

5.2.1.1 Organisation of HP domain

Previous crystallography studies had shown the HP domain as a dimer (Rangarajan et al., 2006), with the protein also existing as a dimer in solution. Thus the HP domain was assumed to lie as a dimer along the twofold edge of the cube-like core. However, reference-free 2D classification clearly showed that additional density was observed on the corners of the central cube. As there are three copies of the IGPD core making up this vertex, with the termini nearby it is clear that this dimer model is incorrect, with a trimer forming instead. Following reconstruction with octahedral symmetry imposed there was a clear addition of a three-fold density at each corner of the IGPD core when compared to the plant IGPD structure (Figure 5.23). As the HisB protein is a fusion protein of the IGPD monomer and the HP enzyme, the position of three IGPD termini on the corner of the complex means there has to be three copies of the HP domain present. The size and shape of this additional density is consistent with the addition of three copies of the HP domain being attached at this point, although the resolution of these domains is too low to be able to unambiguously show the organisation of the HP domain or the nature of its exact attachment to the IGPD core.

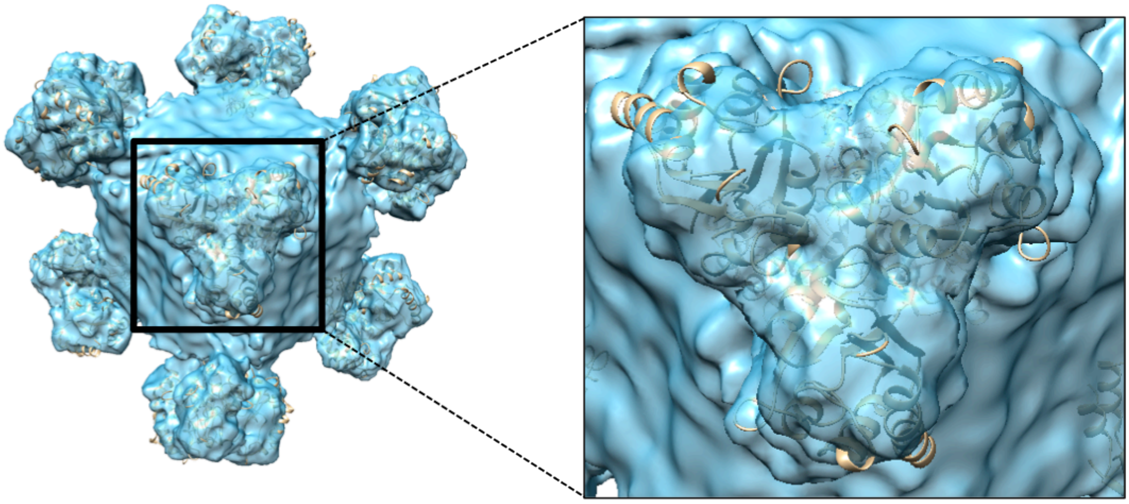


Figure 5.23. - HisB HP trimer organisation. HisB reconstruction showing peripheral HP domain as a trimer on each corner of IGPD core rather than expected dimer. Insert shows zoom of trimer region. Atomic model adapted from PDB 2FPU (Rangarajan et al., 2006).

As the final reconstruction had octahedral symmetry applied, the observed HP trimer may have been a symmetry artefact. In an attempt to improve the reconstruction of the HP domains, and to verify the proposed HP trimer, localised refinement, using protocols developed by Juha Huiskennen (Ilca et al., 2015), was trialled. Sub-particles consisting of only the HP domains were extracted from the final particle stack giving ~126,000 sub-particles. 3D classification into 6 classes with no alignments was then carried out. This classification led to several classes showing features consistent in size and shape with a dimer of HP domain as well as one class showing a trimer and one other appearing to show a dimer with extra density alongside, potentially representing a lone monomer (Figure 5.24). This is perhaps not surprising as crystal structures have shown the molecule to form a dimer. However, as three copies of this protein must be present due to its location on the three-fold vertex of the IGPD core, there will be a symmetry mismatch between the HP domain and the overall complex.

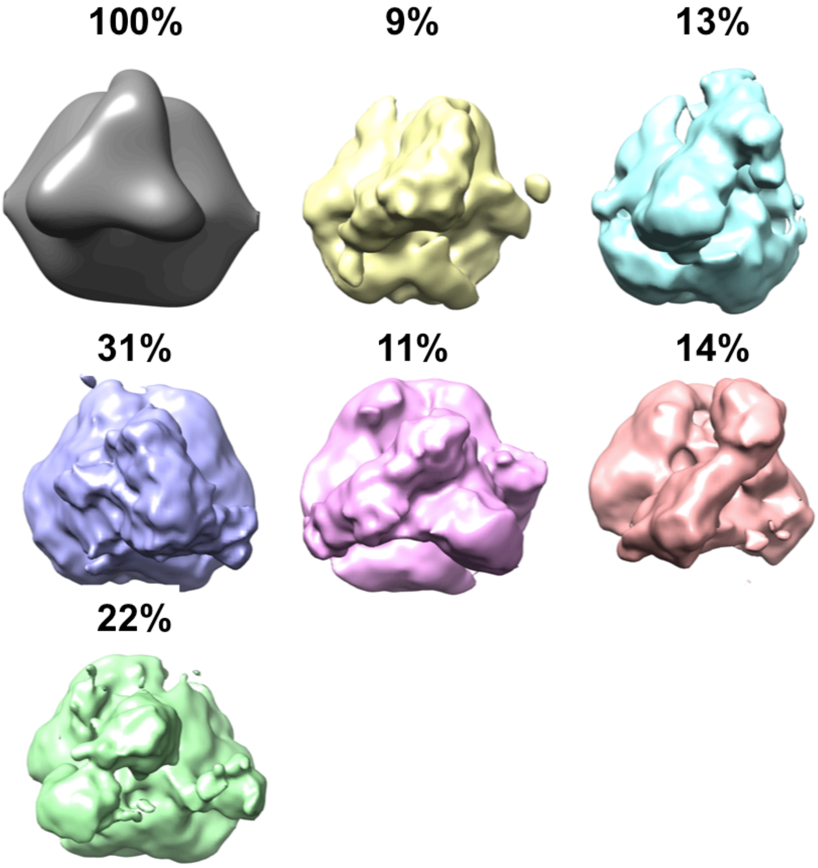


Figure 5.24 - Localised classification of HP trimer subparticles. Results of subparticle classification into six classes with C1 symmetry showing reduction of the trimeric shape from the starting model (grey) with percentage of particles in each class shown above.

It is possible that at any given time, two copies of the HP domain exist as a more rigid dimer, with the third copy as a flexible monomer, which could be why it is not observed in the EM density. It is unclear why an enzyme that forms a native dimer would be fused such that it lies on a three-fold axis of the larger complex. The sub-particle extraction was also trialled using signal subtraction to remove density corresponding to the IGPD core particle prior to extraction of the HP sub-particle. However, in this case the HP domain showed poor density following 3D classification, potentially due to the smaller density left following the core subtraction, and no useful information was obtained.

5.3 IAPA Binding Study

It is interesting to note that this HisB fusion complex is formed of steps 6 and 8 in the histidine biosynthesis pathway, rather than 6/7 or 7/8. If this complex existed in isolation, this would require the product of the 6th step (catalysed by IGPD) to diffuse away from the complex, be processed by the 7th element and then return to the HisB complex again, a seemingly inefficient process (Figure 5.1). As such the binding of the 7th protein, IAPA, to this complex was then investigated as formation of a supercomplex would allow more efficient transfer of substrate between the elements of the pathway.

IAPA was cloned, expressed and purified by Dr. Clair Phillips. Crosslinking studies were performed, but no binding was observed upon addition of IAPA to the HisB sample. In addition, negative stain classes of HisB showed no extra density upon incubation with IAPA. This shows that simple addition of the components together is not sufficient to initiate formation of any larger species and suggests that either a more complex formation pathway is required or that no larger complex is formed.

5.4 Conclusions

While traditionally structure-guided inhibitor design has been underpinned through X-ray studies, the increasing power of EM has allowed it to become a viable method for ligand identification. Resolutions sufficient to visualise side chain density have become viable for many systems and the work in this section has shown inhibitor density at ~2.6 Å resolution. The global 3.1 Å resolution allowed an atomic model to be built in *de novo* fashion from only EM density, illustrating the potential of EM for identifying inhibitor binding sites and informing inhibitor design. At the start of this study there were few examples of EM being used to directly visualise binding of small molecule inhibitors to a protein. This work has shown one of the first examples of EM being used to build a *de novo* atomic model using only EM information and directly observe a bound inhibitor within the map. This work represents the smallest inhibitor (205 Da) to be visualised directly by EM to date and it is also the first use of difference maps to identify a bound inhibitor within an EM reconstruction.

Despite the resolution of the EM reconstruction not allowing unambiguous assignment of the inhibitor pose and position, this could be established using *in* *silico* docking procedures with the final result in good agreement with a known crystal structure used for validation purposes. It is, however, interesting to note that the inhibitor position observed in the crystal structure does not precisely match the EM reconstruction. This may be due to simply resolution, subtle differences between the crystalline and frozen hydrated state, or an artefact of radiation damage in the EM structure.

Furthermore, this work has also explored the structure of the HisB protein, which has proven intractable via crystallographic techniques and has shown the trimeric arrangement of the HP domain for the first time.

However, it is worth noting that the IGPD system used for this study is perhaps an unrepresentative test case for the utility of EM in drug discovery efforts. The complex forms an extremely stable, highly symmetric structure, reducing the amount of data required and increasing the likelihood of obtaining high resolution. In addition the inhibitor used was already highly potent, avoiding potential issues with binding site occupancy which may prove challenging in many cases, particularly in early stages of compound development or fragment-based methods. Thus, while this study has shown promise, it is likely that for many other targets of interest, for example the V-ATPase, EM studies will prove significantly more difficult. It is worth noting that for membrane proteins, a key target for drug discovery efforts, there have been no <3 Å EM structures to date, illustrating the challenges still ahead in the routine use of EM for drug discovery.

6. Discussion

The initial aim of this PhD was to investigate the use of EM as a complementary technique to aid the drug discovery process. At the onset of the project the resolutions typically obtainable via EM for non symmetric objects (~1 nm resolution at best) meant a tagging approach was envisaged. By adding a large bulky tag to the inhibitor of interest it was hoped it could be easily localised via EM to provide approximate information on the binding location, which could be expanded further through other approaches. For example, identifying the specific subunit within a larger complex by tagging in EM, which could then be analysed through co-crystallization experiments. The goal was to then functionalise the tag through tag transfer chemistry (Fujishima et al., 2012; Tamura et al., 2012), so that the modified area of the protein could then be easily identified via mass spectrometry analysis, giving both global and local sequence location of the binding site. This tagging approach was trialled using PA1b, a potent toxin inhibitor of the V-ATPase. By labelling the toxin with a bulky horseradish peroxidase tag the binding site was identified at the base of the V_o domain, binding to subunits *c* and *e*. This also allowed the identification of the approximate location of the elusive *e* subunit for the first time.

However, around this time what is now sometimes referred to as the 'Resolution Revolution' in EM was taking place (Kühlbrandt, 2014). It was apparent that while this tagging approach could potentially be a powerful tool, with improving resolutions came the opportunity to study inhibitor binding via

EM in a more direct fashion. At that point EM had only given modest reconstructions of the V-ATPase at 11 and 16 Å from yeast and *M. sexta*, respectively. As such cryo-EM studies were carried out on the V-ATPase in order to improve this resolution and determine if EM could be used to inform the design of new V-ATPase inhibitors. As a result, the first sub nm structure of the V-ATPase was obtained from *M. sexta*, showing a different catalytic state to that previously observed in yeast. While this structure was superseded by yeast structures in three catalytic states at higher resolution, it did provide insight into the mechanism of the V-ATPase. This included analysis of a previously ignored electrostatic bearing region. It is likely that the resolution of the *M. sexta* reconstruction was limited largely due to particle number, with only ~12k V-ATPase particles collected as opposed to the ~250k initially picked particles in the yeast dataset. However, even with more data available the yeast system still is limited to ~7 Å resolution. Thus other causes for the limited resolution were investigated.

As part of this work flexibility within the V-ATPase was studied. This showed flexing of the V_1 and V_o domains of ~7° in class averages, consistent with MD simulations and crystallographic studies. A larger flex of ~30° was also observed which was affected by the addition of ATP and could represent a degraded or dissociating complex. This flexibility exists in addition to the multiple conformational states which exist in the sample as a result of catalytic cycling. The resolution degradation due to the flexibility along the molecule was partially mitigated in the cryo-EM data set through the use of tight local masks and fine angular sampling, improving the V_1 domain by ~2 Å, with local resolution extending to ~6 Å. However this had no effect on the smaller V_o domain.

In addition to studying the full V-ATPase complex, cryo-EM studies were carried out on the isolated V_1 domain in order to probe the mechanism by which regulation and ATP silencing occurs following dissociation of the complex. As such structures of V_1 from intact complex, naturally dissociated complex and degraded complex were solved by cryo-EM and compared. However, no obvious structural changes which could cause ATP silencing were identified, although some variability in the AB domains in the open site was observed. Subunit H had been proposed to play a role in ATP silencing through a structural rearrangement allowing it to contact the central rotor axle (Oot et al., 2016), but this was not observed in our EM reconstructions of either V_1 domain. This maybe due to differences in higher and lower eukaryotes (*M. sexta* vs yeast), crystallization artefacts or limited resolution but suggest that regulation is not through a simple rearrangement of subunit H. Surprisingly there appeared to be little difference between the supposedly natural and relevant dissociated sample and the V_1 which was obtained as a result of degradation of the full complex. This suggests that potentially studies carried out on isolated V_1 do not fully reflect the natural process of regulation.

While the moderate resolution obtained for the V-ATPase does allow for fitting of known atomic coordinates into the map, potentially allowing binding sites at subunit interfaces to be identified, it was clear that it would be insufficient for traditional structure based drug design approaches. As such a new model system was identified which was hoped would be more amenable to achieving high resolution reconstructions, illustrating the potential for EM as a tool for drug discovery. *At*IGPD2 was chosen as a model system as it possessed octahedral symmetry, thus reducing the amount of data needed drastically compared to an asymmetric system. Several potent (~20 nM) inhibitors were also available, mitigating any potential issues arising from partial occupancy within the complex. In addition it was known to be a very stable complex with several high resolution crystal structures available to aid in validation of the EM structure. This allowed a ~3.1 Å reconstruction of the complex to be obtained, allowing *de novo* building of an atomic model into the EM reconstruction. A difference map was then used to identify the inhibitor binding site, with clear density corresponding to the inhibitor visible. Validation with existing crystal structures showed extremely good agreement, with minor differences likely a product of differing resolutions and the radiation damage present in the EM reconstruction. In addition the structure of the related HisB complex from *E. coli* was solved by cryo-EM to ~7 Å showing the trimeric arrangement of the catalytic histininol phosphatase domains for the first time.

6.1 Concluding remarks and future perspectives

This thesis aimed to explore a simple question; could cryo-EM play a useful role in the drug discovery process? In short, the answer is yes. This work has shown that cryo-EM can provide valuable information at all resolution ranges; from low resolution binding site identification, to direct visualisation of a bound small molecule inhibitor.

However, there are important caveats which must be considered. Firstly, while always expanding, there is a limited range of proteins where high resolution EM studies are viable. Indeed at the present time it seems unlikely that structures of large, dynamic systems like the V-ATPase will be obtained at resolutions where direct visualisation of inhibitor binding will be viable. The model system used in this work (IGPD) represents the best case scenario, combining high symmetry with stability and potent inhibitors. This is unlikely to be the case for many interesting molecular targets. Secondly, and perhaps more importantly, how valid is the information that we obtain from these reconstructions? As the information contained within the earliest moments of an exposure is currently unusable due to apparent beam induced movement, EM is limited to producing reconstructions made from protein which has already undergone a significant radiation dose. Nevertheless, when comparing atomic models obtained from EM reconstructions with their X-ray counterparts little difference is observed, suggesting that while damage may be occurring, the answer obtained is little changed at the resolution ranges observed. Finally a practical concern, throughput. While crystallographers can routinely collect diffraction data and solve preliminary structures in a matter of minutes, in EM, data collection takes hours or days and the subsequent data processing weeks or months after this. Thus the concept of screening several different analogues of a compound to explore subtle changes in binding becomes far less viable.

The good news is that all of these challenges are being addressed by the field. Work is on-going from improvements in imaging conditions and specimen supports, to retain as much of the early undamaged information as possible, to novel grid preparation techniques and automated data collection/processing pipelines offering the possibility of rapid collection of multiple samples. EM can also provide valuable structural information on proteins which are intractable by other techniques like crystallography or NMR across a wide resolution range. The tagging approaches originally planned may prove to be another powerful tool for studying inhibitor binding via EM. It is also important to remember the wealth of complimentary information EM can provide away from simple single static structures, from identifying conformational states in dynamic proteins to visualising proteins in more natural environments.

In conclusion, while challenges remain, the field of EM is constantly evolving and improving. Soon the most important question left to answer may be “What do we do now we’ve solved all of the structures?”

7. References

Abdul-Hussein, S., Andréll, J., and Tate, C.G. (2013). Thermostabilisation of the Serotonin Transporter in a Cocaine-Bound Conformation. J. Mol. Biol. *425*, 2198–2207.

Abrahams, J.P., Leslie, A.G., Lutter, R., and Walker, J.E. (1994). Structure at 2.8 A resolution of F1-ATPase from bovine heart mitochondria. Nature *370*, 621–628.

Adrian, M., Dubochet, J., Lepault, J., and McDowall, A.W. (1984). Cryo-electron microscopy of viruses. Nature *308*, 32–36.

Allegretti, M., Klusch, N., Mills, D.J., Vonck, J., Kühlbrandt, W., and Davies, K.M. (2015). Horizontal membrane-intrinsic α-helices in the stator a-subunit of an F-type ATP synthase. Nature *521*, 237–240.

Allegretti, M., Mills, D.J., McMullan, G., Kühlbrandt, W., and Vonck, J. (2014). Atomic model of the F(420)-reducing [NiFe] hydrogenase by electron cryo-microscopy using a direct electron detector. eLife *3*, e01963.

Arai, S., Saijo, S., Suzuki, K., Mizutani, K., Kakinuma, Y., Ishizuka-Katsura, Y., Ohsawa, N., Terada, T., Shirouzu, M., Yokoyama, S., et al. (2013). Rotation mechanism of Enterococcus hirae V1-ATPase based on asymmetric crystal structures. Nature *493*, 703–707.

Bai, X.-C., McMullan, G., and Scheres, S.H.W. (2015a). How cryo-EM is revolutionizing structural biology. Trends Biochem. Sci. *40*, 49–57.

Bai, X.-C., Rajendra, E., Yang, G., Shi, Y., and Scheres, S.H.W. (2015b). Sampling the conformational space of the catalytic subunit of human γ-secretase. eLife *4*, e11182.

Bai, X.-C., Yan, C., Yang, G., Lu, P., Ma, D., Sun, L., Zhou, R., Scheres, S.H.W., and Shi, Y. (2015c). An atomic structure of human γ-secretase. Nature *525*, 212–217.

Baker, L.A., Smith, E.A., Bueler, S.A., and Rubinstein, J.L. (2010). The resolution dependence of optimal exposures in liquid nitrogen temperature electron cryomicroscopy of catalase crystals. J Struct Biol *169*, 431–437.

Balabaskaran Nina, P., Morrisey, J.M., Ganesan, S.M., Ke, H., Pershing, A.M., Mather, M.W., and Vaidya, A.B. (2011). ATP synthase complex of Plasmodium falciparum: dimeric assembly in mitochondrial membranes and resistance to genetic disruption. J. Biol. Chem. *286*, 41312–41322.

Bammes, B.E., Rochat, R.H., Jakana, J., and Chiu, W. (2011). Practical performance evaluation of a 10k × 10k CCD for electron cryo-microscopy. J Struct Biol *175*, 384–393.

Bartesaghi, A., Merk, A., Banerjee, S., Matthies, D., Wu, X., Milne, J.L.S., and Subramaniam, S. (2015). 2.2 Å resolution cryo-EM structure of β-galactosidase in complex with a cell-permeant inhibitor. Science *348*, 1147–1151.

Bell, J.M., Chen, M., Baldwin, P.R., and Ludtke, S.J. (2016). High resolution single particle refinement in EMAN2.1. Methods 1–10.

Benjamin, C.J., Wright, K.J., Hyun, S.-H., Krynski, K., Yu, G., Bajaj, R., Guo, F., Stauffacher, C.V., Jiang, W., and Thompson, D.H. (2016). Nonfouling NTA-PEG-Based TEM Grid Coatings for Selective Capture of Histidine-Tagged Protein Targets from Cell Lysates. Langmuir *32*, 551–559.

Benlekbir, S., Bueler, S.A., and Rubinstein, J.L. (2012). Structure of the vacuolar-type ATPase from Saccharomyces cerevisiae at 11-Å resolution. Nature Structural & Molecular Biology *19*, 1356–1362.

Bill, R.M., Henderson, P.J.F., Iwata, S., Kunji, E.R.S., Michel, H., Neutze, R., Newstead, S., Poolman, B., Tate, C.G., and Vogel, H. (2011). Overcoming barriers to membrane protein structure determination. Nat. Biotechnol. *29*, 335–340.

Bisson, C., Britton, K.L., Sedelnikova, S.E., Rodgers, H.F., Eadsforth, T.C., Viner, R.C., Hawkes, T.R., Baker, P.J., and Rice, D.W. (2015). Crystal Structures Reveal that the Reaction Mechanism of Imidazoleglycerol-Phosphate Dehydratase Is Controlled by Switching Mn(II) Coordination. Structure *23*, 1236–1245.

Bisson, C., Britton, K.L., Sedelnikova, S.E., Rodgers, H.F., Eadsforth, T.C., Viner, R.C., Hawkes, T.R., Baker, P.J., and Rice, D.W. (2016). Mirror-Image Packing Provides a Molecular Basis for the Nanomolar Equipotency of Enantiomers of an Experimental Herbicide. Angewandte Chemie International Edition *55*, 13485–13489.

Blair, H.C., Teitelbaum, S.L., Ghiselli, R., and Gluck, S. (1989). Osteoclastic bone resorption by a polarized vacuolar proton pump. Science *245*, 855–857.

Blundell, T.L., Jhoti, H., and Abell, C. (2002). High-throughput crystallography for lead discovery in drug design. Nat Rev Drug Discov *1*, 45–54.

Booth, D.S., Avila-Sakar, A., and Cheng, Y. (2011). Visualizing Proteins and Macromolecular Complexes by Negative Stain EM: from Grid Preparation to Image Acquisition. e3227.

Boyer, P.D. (1997). The ATP synthase--a splendid molecular machine. Annu. Rev. Biochem. *66*, 717–749.

Brenner, S., and Horne, R.W. (1959). A negative staining method for high resolution electron microscopy of viruses. Biochim. Biophys. Acta *34*, 103–110.

Brilot, A.F., Chen, J.Z., Cheng, A., Pan, J., Harrison, S.C., Potter, C.S., Carragher, B., Henderson, R., and Grigorieff, N. (2012). Beam-Induced Motion of Vitrified Specimen on Holey Carbon Film. J Struct Biol *177*, 630–637.

Brown, A., Long, F., Nicholls, R.A., Toots, J., Emsley, P., and Murshudov, G. (2015). Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. Acta Crystallographica. Section D, Biological Crystallography *71*, 136–153.

Bueler, S.A., and Rubinstein, J.L. (2015). Vma9p need not be associated with the yeast V-ATPase for fully-coupled proton pumping activity in vitro. Biochemistry-Us *54*, 853–858.

Cabra, V., and Samso, M. (2015). Do"s and Don"ts of Cryo-electron Microscopy: A Primer on Sample Preparation and High Quality Data Collection for Macromolecular 3D Reconstruction. e52311.

Caffrey, M. (2015). A comprehensive review of the lipid cubic phase or in meso method for crystallizing membrane and soluble proteins and complexes. Acta Crystallogr F Struct Biol Commun *71*, 3–18.

Campbell, M.G., Cheng, A., Brilot, A.F., Moeller, A., Lyumkis, D., Veesler, D., Pan, J., Harrison, S.C., Potter, C.S., Carragher, B., et al. (2012). Movies of Ice-Embedded Particles Enhance Resolution in Electron Cryo-Microscopy. Structure *20*, 1823–1828.

Cardone, G., Heymann, J.B., and Steven, A.C. (2013). One number does not fit all: Mapping local variations in resolution in cryo-EM reconstructions. J Struct Biol *184*, 226–236.

Carragher, B., Kisseberth, N., Kriegman, D., Milligan, R.A., Potter, C.S., Pulokas, J., and Reilein, A. (2000). Leginon: An Automated System for Acquisition of Images from Vitreous Ice Specimens. J Struct Biol *132*, 33–45.

Chen, B., and Frank, J. (2016). Two promising future developments of cryo-EM: capturing short-lived states and mapping a continuum of states of a macromolecule. Microscopy *65*, 69–79.

Chen, B., Kaledhonkar, S., Sun, M., Shen, B., Lu, Z., Barnard, D., Lu, T.-M., Gonzalez, R.L., Jr., and Frank, J. (2015). Structural Dynamics of Ribosome Subunit Association Studied by Mixing-Spraying Time-Resolved Cryogenic Electron Microscopy. Structure *23*, 1097–1105.

Chouabe, C., Eyraud, V., Da Silva, P., Rahioui, I., Royer, C., Soulage, C., Bonvallet, R., Huss, M., and Gressent, F. (2011). New mode of action for a knottin protein bioinsecticide: pea albumin 1 subunit b (PA1b) is the first peptidic inhibitor of V-ATPase. J. Biol. Chem. *286*, 36291–36296.

Cianfrocco, M.A., and Leschziner, A.E. (2015). Low cost, high performance processing of single particle cryo-electron microscopy data in the cloud. eLife *4*, e06664.

Clark, D.E. (2006). What has computer-aided molecular design ever done for drug discovery? Expert Opinion on Drug Discovery *1*, 103–110.

Compton, M.A., Graham, L.A., and Stevens, T.H. (2006). Vma9p (subunit e) is an integral membrane V0 subunit of the yeast V-ATPase. Journal of Biological Chemistry *281*, 15312–15319.

Cotter, K., Stransky, L., McGuire, C., and Forgac, M. (2015). Recent Insights into the Structure, Regulation, and Function of the V-ATPases. Trends Biochem. Sci. *40*, 611–622.

Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallographica. Section D, Biological Crystallography *62*, 1002–1011.

da Fonseca, P.C.A., and Morris, E.P. (2015). Cryo-EM reveals the conformation of a substrate analogue in the human 20S proteasome core. Nat Commun *6*, 7573–7576.

Da Silva, P., Strzepa, A., Jouvensal, L., Rahioui, I., Gressent, F., and Delmas, A.F. (2009). A folded and functional synthetic PA1b: an interlocked entomotoxic miniprotein. Biopolymers *92*, 436–444.

Danev, R., and Nagayama, K. (2001). Transmission electron microscopy with Zernike phase plate. Ultramicroscopy *88*, 243–252.

Danev, R., and Baumeister, W. (2016). Cryo-EM single particle analysis with the Volta phase plate. eLife *5*, 439.

Danev, R., Buijsse, B., Khoshouei, M., Plitzko, J.M., and Baumeister, W. (2014). Volta potential phase plate for in-focus phase contrast transmission electron microscopy. Proceedings of the National Academy of Sciences *111*, 15635–15640.

Dashti, A., Schwander, P., Langlois, R., Fung, R., Li, W., Hosseinizadeh, A., Liao, H.Y., Pallesen, J., Sharma, G., Stupina, V.A., et al. (2014). Trajectories of the ribosome as a Brownian nanomachine. Proceedings of the National Academy of Sciences *111*, 17492–17497.

Davies, K.M., Anselmi, C., Wittig, I., Faraldo-Gómez, J.D., and Kühlbrandt, W. (2012). Structure of the yeast F1Fo-ATP synthase dimer and its role in shaping the mitochondrial cristae. Proceedings of the National Academy of Sciences *109*, 13602–13607.

Davis, A.M., and Teague, S.J. (1999). Hydrogen bonding, hydrophobic interactions, and failure of the rigid receptor hypothesis. Angewandte Chemie International Edition *38*, 736–749.

Davis, A., St-Gallay, S., and Kleywegt, G. (2008). Limitations and lessons in the use of X-ray structural information in drug design. Drug Discov. Today *13*, 831–841.

De Rosier, D.J., and Klug, A. (1968). Reconstruction of Three Dimensional Structures from Electron Micrographs. Nature *217*, 130–134.

Diab, H., Ohira, M., Liu, M., Cobb, E., and Kane, P.M. (2009). Subunit Interactions and Requirements for Inhibition of the Yeast V-1-ATPase. Journal of Biological Chemistry *284*, 13316–13325.

DiMaio, F., Song, Y., Li, X., Brunner, M.J., Xu, C., Conticello, V., Egelman, E., Marlovits, T.C., Cheng, Y., and Baker, D. (2015). Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement. Nat. Methods *12*, 361–365.

Dobro, M.J., Melanson, L.A., Jensen, G.J., and McDowall, A.W. (2010). Chapter Three - Plunge Freezing for Electron Cryomicroscopy. In The Resolution Revolution: Recent Advances in cryoEM, G.J. Jensen, ed. (Academic Press), pp. 63–82.

Dominguez, E., Galmozzi, A., Chang, J.W., Hsu, K.-L., Pawlak, J., Li, W., Godio, C., Thomas, J., Partida, D., Niessen, S., et al. (2014). Integrated phenotypic and activity-based profiling links Ces3 to obesity and diabetes. Nature Chemical Biology *10*, 113–121.

Dow, J. (1984). Extremely High Ph in Biological-Systems - a Model for Carbonate Transport. Am. J. Physiol. *246*, R633–R636.

Dörr, J.M., Scheidelaar, S., Koorengevel, M.C., Dominguez, J.J., Schäfer, M., van Walree, C.A., and Killian, J.A. (2016). The styrene-maleic acid copolymer: a versatile tool in membrane research. European Biophysics Journal : EBJ *45*, 3–21.

Dubochet, J., Adrian, M., Chang, J.J., Homo, J.C., Lepault, J., McDowall, A.W., and Schultz, P. (1988). Cryo-electron microscopy of vitrified specimens. Q. Rev. Biophys. *21*, 129–228.

Elmlund, H., Elmlund, D., and Bengio, S. (2013). PRIME: probabilistic initial 3D model generation for single-particle cryo-electron microscopy. Structure *21*, 1299–1306.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallographica. Section D, Biological Crystallography *66*, 486–501.

Erlanson, D.A., Fesik, S.W., Hubbard, R.E., Jahnke, W., and Jhoti, H. (2016). Twenty years on: the impact of fragments on drug discovery. Nat Rev Drug Discov *15*, 605–619.

Eyraud, V., Karaki, L., Rahioui, I., Sivignon, C., Da Silva, P., Rahbe, Y., Royer, C., and Gressent, F. (2013). Expression and biological activity of the cystine knot bioinsecticide PA1b (Pea Albumin 1 Subunit b). PLoS ONE *8*, e81619.

Fais, S., De Milito, A., You, H., and Qin, W. (2007). Targeting vacuolar H+-ATPases as a new strategy against cancer. Cancer Res. *67*, 10627–10630.

Fan, G., Baker, M.L., Wang, Z., Baker, M.R., Sinyagovskiy, P.A., Chiu, W., Ludtke, S.J., and Serysheva, I.I. (2015). Gating machinery of InsP3R channels revealed by electron cryomicroscopy. Nature *527*, 336–341.

Faruqi, A.R., and McMullan, G. (2011). Electronic detectors for electron microscopy. Q. Rev. Biophys. *44*, 357–390.

Faruqi, A.R., Henderson, R., Pryddetch, M., Allport, P., and Evans, A. (2005). Direct single electron detection with a CMOS detector for electron microscopy. Nuclear Instruments and Methods in Physics Research Section a: Accelerators, Spectrometers, Detectors and Associated Equipment *546*, 170–175.

Fernández, I.S., Bai, X.-C., Hussain, T., Kelley, A.C., Lorsch, J.R., Ramakrishnan, V., and Scheres, S.H.W. (2013). Molecular architecture of a eukaryotic translational initiation complex. Science *342*, 1240585–1240585.

Forgac, M. (2007). Vacuolar ATPases: rotary proton pumps in physiology and pathophysiology. Nat Rev Mol Cell Biol *8*, 917–929.

Frank, J., and Ourmazd, A. (2016). Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-EM. Methods *100*, 61–67.

Frauenfeld, J., Löving, R., Armache, J.-P., Sonnen, A.F.-P., Guettou, F., Moberg, P., Zhu, L., Jegerschöld, C., Flayhan, A., Briggs, J.A.G., et al. (2016). A saposin-lipoprotein nanoparticle system for membrane proteins. Nat. Methods *13*, 345–351.

Fujishima, S.-H., Yasui, R., Miki, T., Ojida, A., and Hamachi, I. (2012). Ligand-Directed Acyl Imidazole Chemistry for Labeling of Membrane-Bound Proteins on Live Cells. Journal of the American Chemical Society *134*, 3961–3964.

Gao, Y., Cao, E., Julius, D., and Cheng, Y. (2016). TRPV1 structures in nanodiscs reveal mechanisms of ligand and lipid action. Nature *534*, 347–351.

Ge, J., Li, W., Zhao, Q., Li, N., Chen, M., Zhi, P., Li, R., Gao, N., Xiao, B., and Yang, M. (2015). Architecture of the mammalian mechanosensitive Piezo1 channel. Nature *527*, 64–69.

Glaeser, R.M., and Taylor, K.A. (1978). Radiation damage relative to transmission electron microscopy of biological specimens at low temperature: a review. Journal of Microscopy *112*, 127–138.

Glaeser, R.M., McMullan, G., Faruqi, A.R., and Henderson, R. (2011). Images of paraffin monolayer crystals with perfect contrast: minimization of beam-induced specimen motion. Ultramicroscopy *111*, 90–100.

Glaeser, R.M. (2016). How good can cryo-EM become? Nat. Methods *13*, 28–32.

Goddard, T.D., Huang, C.C., and Ferrin, T.E. (2007). Visualizing density maps with UCSF Chimera. J Struct Biol *157*, 281–287.

Gohda, K., Kimura, Y., Mori, I., Ohta, D., and Kikuchi, T. (1998). Theoretical evidence of the existence of a diazafulvene intermediate in the reaction pathway of imidazoleglycerol phosphate dehydratase: design of a novel and potent heterocycle structure for the inhibitor on the basis of the electronic structure-activity relationship study. Biochim. Biophys. Acta *1385*, 107–114.

Goldenberg, O., Erez, E., Nimrod, G., and Ben-Tal, N. (2009). The ConSurf-DB: pre-calculated evolutionary conservation profiles of protein structures. Nucleic Acids Res. *37*, D323–D327.

Gong, X., Qian, H., Zhou, X., Wu, J., Wan, T., Cao, P., Huang, W., Zhao, X., Wang, X., Wang, P., et al. (2016). Structural Insights into the Niemann-Pick C1 (NPC1)-Mediated Cholesterol Transfer and Ebola Infection. Cell *165*, 1467–1478.

Grant, T., and Grigorieff, N. (2015). Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. eLife *4*, e06980.

Grassucci, R.A., Taylor, D.J., and Frank, J. (2007). Preparation of macromolecular complexes for cryo-electron microscopy. Nature Protocols *2*, 3239–3246.

Grassucci, R.A., Taylor, D., and Frank, J. (2008). Visualization of macromolecular complexes using cryo-electron microscopy with FEI Tecnai transmission electron microscopes. Nature Protocols *3*, 330–339.

Gräf, R., Harvey, W.R., and Wieczorek, H. (1996). Purification and Properties of a Cytosolic V1-ATPase. Journal of Biological Chemistry *271*, 20908–20913.

Gressent, F., Da Silva, P., Eyraud, V., Karaki, L., and Royer, C. (2011). Pea Albumin 1 subunit b (PA1b), a promising bioinsecticide of plant origin. Toxins *3*, 1502–1517.

Gressent, F., Duport, G., Rahioui, I., Pauchet, Y., Bolland, P., Specty, O., and Rahbe, Y. (2007). Biological activity and binding site characteristics of the PA1b Entomotoxin on insects from different orders. J. Insect Sci. *7*, 1–10.

Hahn, A., Parey, K., Bublitz, M., Mills, D.J., Zickermann, V., Vonck, J., Kühlbrandt, W., and Meier, T. (2016). Structure of a Complete ATP Synthase Dimer Reveals the Molecular Basis of Inner Mitochondrial Membrane Morphology. Molecular Cell *63*, 445–456.

Hauer, F., Gerle, C., Fischer, N., Oshima, A., Shinzawa-Itoh, K., Shimada, S., Yokoyama, K., Fujiyoshi, Y., and Stark, H. (2015). GraDeR: Membrane Protein Complex Preparation for Single-Particle Cryo-EM. Structure *23*, 1769–1775.

Hawkes, T.R., Thomas, P.G., Edwards, L.S., Rayner, S.J., Wilkinson, K.W., and Rice, D.W. (1995). Purification and characterization of the imidazoleglycerol-phosphate dehydratase of Saccharomyces cerevisiae from recombinant Escherichia coli. Biochemical Journal *306*, 385–397.

Henderson, R. (1995). The potential and limitations of neutrons, electrons and

X-rays for atomic resolution microscopy of unstained biological molecules. Q. Rev. Biophys. *28*, 171.

Henderson, R. (2013). Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. Proceedings of the National Academy of Sciences *110*, 18037–18041.

Henderson, R., and McMullan, G. (2013). Problems in obtaining perfect images by single-particle electron cryomicroscopy of biological structures in amorphous ice. Microscopy (Oxf) *62*, 43–50.

Henderson, R., Chen, S., Chen, J.Z., Grigorieff, N., Passmore, L.A., Ciccarelli, L., Rubinstein, J.L., Crowther, R.A., Stewart, P.L., and Rosenthal, P.B. (2011). Tilt-Pair Analysis of Images from a Range of Different Specimens in Single-Particle Electron Cryomicroscopy. J. Mol. Biol. *413*, 1028–1046.

Hesketh, E.L., Meshcheriakova, Y., Dent, K.C., Saxena, P., Thompson, R.F., Cockburn, J.J., Lomonossoff, G.P., and Ranson, N.A. (2015). Mechanisms of assembly and genome packaging in an RNA virus revealed by high-resolution cryo-EM. Nat Commun *6*, 10113.

Hildenbrand, Z.L., Molugu, S.K., Stock, D., and Bernal, R.A. (2010). The C-H Peripheral Stalk Base: A Novel Component in V1-ATPase Assembly. PLoS ONE *5*, e12588–e12589.

Hinton, A., Bond, S., and Forgac, M. (2009). V-ATPase functions in normal and disease processes. Pflugers Arch. *457*, 589–598.

Huss, M., Ingenhorst, G., Konig, S., Gassel, M., Drose, S., Zeeck, A., Altendorf, K., and Wieczorek, H. (2002). Concanamycin a, the specific inhibitor of V-ATPases, binds to the V-o subunit c. Journal of Biological Chemistry *277*, 40544–40548.

Huss, M., and Wieczorek, H. (2007). Influence of ATP and ADP on dissociation of the V-ATPase into its V1 and VO complexes. FEBS Lett. *581*, 5566–5572.

Huss, M., and Wieczorek, H. (2009). Inhibitors of V-ATPases: old and new players. J. Exp. Biol. *212*, 341–346.

Huss, M., Sasse, F., Kunze, B., Jansen, R., Steinmetz, H., Ingenhorst, G., Zeeck, A., and Wieczorek, H. (2005). Archazolid and apicularen: novel specific V-ATPase inhibitors. BMC Biochem. *6*, 13.

Ilca, S.L., Kotecha, A., Sun, X., Poranen, M.M., Stuart, D.I., and Huiskonen, J.T. (2015). Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes. Nat Commun *6*, 8843.

Ingle, R.A., Mugford, S.T., Rees, J.D., Campbell, M.M., and Smith, J.A.C. (2005). Constitutively high expression of the histidine biosynthetic pathway contributes to nickel tolerance in hyperaccumulator plants. Plant Cell *17*, 2089–2106.

Jain, T., Sheehan, P., Crum, J., Carragher, B., and Potter, C.S. (2012). Spotiton: A prototype for an integrated inkjet dispense and vitrification system for cryo-TEM. J Struct Biol *179*, 68–75.

Jamshad, M., Lin, Y.-P., Knowles, T.J., Parslow, R.A., Harris, C., Wheatley, M., Poyner, D.R., Bill, R.M., Thomas, O.R.T., Overduin, M., et al. (2011). Surfactant-free purification of membrane proteins with intact native membrane environment. Biochem. Soc. Trans. *39*, 813–818.

Jefferies, K.C., and Forgac, M. (2008). Subunit H of the vacuolar (H+) ATPase inhibits ATP hydrolysis by the free V1 domain by interaction with the rotary subunit F. Journal of Biological Chemistry *283*, 4512–4519.

Jiang, Q.-X., Thrower, E.C., Chester, D.W., Ehrlich, B.E., and Sigworth, F.J. (2002). Three-dimensional structure of the type 1 inositol 1,4,5-trisphosphate receptor at 24 A resolution. Embo J. *21*, 3575–3581.

Jouvensal, L., Quillien, L., Ferrasson, E., Rahbe, Y., Guéguen, J., and Vovelle, F. (2003). PA1b, an insecticidal protein extracted from pea seeds (Pisum sativum): 1H-2-D NMR study and molecular modeling. Biochemistry-Us *42*, 11915–11923.

Junge, W., Sielaff, H., and Engelbrecht, S. (2009). Torque generation and elastic power transmission in the rotary FOF1-ATPase. Nature *459*, 364–370.

Kane, P.M. (1995). Disassembly and Reassembly of the Yeast Vacuolar H+-Atpase in-Vivo. Journal of Biological Chemistry *270*, 17025–17032.

Kelley, L.A., and Sternberg, M.J.E. (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nature Protocols *4*, 363–371.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols *10*, 845–858.

Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H., and Phillips, D.C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature *181*, 662–666.

Khoshouei, M., Radjainia, M., Phillips, A.J., Gerrard, J.A., Mitra, A.K., Plitzko, J.M., Baumeister, W., and Danev, R. (2016). Volta phase plate cryo-EM of the small protein complex Prx3. Nat Commun *7*, 10534.

Kleinschmidt, J.H., and Popot, J.-L. (2014). Folding and stability of integral membrane proteins in amphipols. Arch. Biochem. Biophys. *564*, 327–343.

Kleywegt, G.J., Henrick, K., Dodson, E.J., and van Aalten, D.M.F. (2003). Pound-Wise but Penny-Foolish. Structure *11*, 1051–1059.

Krivanek, O.L., Chisholm, M.F., Nicolosi, V., Pennycook, T.J., Corbin, G.J., Dellby, N., Murfitt, M.F., Own, C.S., Szilagyi, Z.S., Oxley, M.P., et al. (2010). Atom-by-atom structural and chemical analysis by annular dark-field electron microscopy. Nature *464*, 571–574.

Kucukelbir, A., Sigworth, F.J., and Tagare, H.D. (2013). Quantifying the local resolution of cryo-EM density maps. Nat. Methods *11*, 63–65.

Kühlbrandt, W. (2014). Biochemistry. The resolution revolution. Science *343*, 1443–1444.

Lander, G.C., Stagg, S.M., Voss, N.R., Cheng, A., Fellmann, D., Pulokas, J., Yoshioka, C., Irving, C., Mulder, A., Lau, P.-W., et al. (2009). Appion: An integrated, database-driven pipeline to facilitate EM image processing. J Struct Biol *166*, 95–102.

Lau, W.C.Y., and Rubinstein, J.L. (2012). Subnanometre-resolution structure of the intact Thermus thermophilus H+-driven ATP synthase. Nature *481*, 214–218.

Lee, S.C., Knowles, T.J., Postis, V.L.G., Jamshad, M., Parslow, R.A., Lin, Y.-P., Goldman, A., Sridhar, P., Overduin, M., Muench, S.P., et al. (2016). A method for detergent-free isolation of membrane proteins in their local lipid environment. Nature Protocols *11*, 1149–1162.

Li, H., O'Donoghue, A.J., van der Linden, W.A., Xie, S.C., Yoo, E., Foe, I.T., Tilley, L., Craik, C.S., da Fonseca, P.C.A., and Bogyo, M. (2016). Structure- and function-based design of Plasmodium-selective proteasome inhibitors. Nature *530*, 233–236.

Li, X., Mooney, P., Zheng, S., Booth, C.R., Braunfeld, M.B., Gubbens, S., Agard, D.A., and Cheng, Y. (2013a). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nat. Methods *10*, 584–590.

Li, X., Zheng, S.Q., Egami, K., Agard, D.A., and Cheng, Y. (2013b). Influence of electron dose rate on electron counting images recorded with the K2 camera. J Struct Biol *184*, 251–260.

Liebeschuetz, J., Hennemann, J., Olsson, T., and Groom, C.R. (2012). The good, the bad and the twisted: a survey of ligand geometry in protein crystal structures. Journal of Computer-Aided Molecular Design *26*, 169–183.

Liu, Z., Guo, F., Wang, F., Li, T.-C., and Jiang, W. (2016). 2.9 Å Resolution Cryo-EM 3D Reconstruction of Close-Packed Virus Particles. Structure *24*, 319–328.

Lounnas, V., Ritschel, T., Kelder, J., McGuire, R., Bywater, R.P., and Foloppe, N. (2013). Current progress in Structure-Based Rational Drug Design marks a new mindset in drug discovery. Comput Struct Biotechnol J *5*, e201302011–e201302014.

Lu, P., Bai, X.-C., Ma, D., Xie, T., Yan, C., Sun, L., Yang, G., Zhao, Y., Zhou, R., Scheres, S.H.W., et al. (2014). Three-dimensional structure of human γ-secretase. Nature *512*, 166–170.

Lu, Z., Shaikh, T.R., Barnard, D., Meng, X., Mohamed, H., Yassin, A., Mannella, C.A., Agrawal, R.K., Lu, T.-M., and Wagenknecht, T. (2009). Monolithic microfluidic mixing-spraying devices for time-resolved cryo-electron microscopy. J Struct Biol *168*, 388–395.

Ludtke, S.J., Baldwin, P.R., and Chiu, W. (1999). EMAN: Semiautomated software for high-resolution single-particle reconstructions. J Struct Biol *128*, 82–97.

Ludtke, S.J., Serysheva, I.I., Hamilton, S.L., and Chiu, W. (2005). The Pore Structure of the Closed RyR1 Channel. Structure *13*, 1203–1211.

Ludwig, J., Kerscher, S., Brandt, U., Pfeiffer, K., Getlawi, F., Apps, D.K., and Schägger, H. (1998). Identification and characterization of a novel 9.2-kDa membrane sector-associated protein of vacuolar proton-ATPase from chromaffin granules. Journal of Biological Chemistry *273*, 10939–10947.

M Walker, J.T.H.W. (1995). Millisecond time resolution electron cryo-microscopy of the M-ATP transient kinetic state of the acto-myosin ATPase. Biophysical Journal *68*, 87S–212.

Macarron, R., Banks, M.N., Bojanic, D., Burns, D.J., Cirovic, D.A., Garyantes, T., Green, D.V.S., Hertzberg, R.P., Janzen, W.P., Paslay, J.W., et al. (2011). Impact of high-throughput screening in biomedical research. Nat Rev Drug Discov *10*, 188–195.

Mahajan, R. (2013). Bedaquiline: First FDA-approved tuberculosis drug in 40 years. Int J Appl Basic Med Res *3*, 1–2.

Makyio, H., Iino, R., Ikeda, C., Imamura, H., Tamakoshi, M., Iwata, M., Stock, D., Bernal, R.A., Carpenter, E.P., Yoshida, M., et al. (2005). Structure of a central stalk subunit F of prokaryotic V-type ATPase/synthase from Thermus thermophilus. Embo J. *24*, 3974–3983.

Marshansky, V., Rubinstein, J.L., and Grüber, G. (2014). Eukaryotic V-ATPase: Novel structural findings and functional insights. Biochim. Biophys. Acta *1837*, 857–879.

Masaike, T., Koyama-Horibe, F., Oiwa, K., Yoshida, M., and Nishizaka, T. (2008). Cooperative three-step motions in catalytic subunits of F1-ATPase correlate with 80° and 40° substep rotations. Nat Struct Mol Biol *15*, 1326–1333.

Mastronarde, D.N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. J Struct Biol *152*, 36–51.

McDowall, A.W., Chang, J.J., Freeman, R., Lepault, J., Walter, C.A., and Dubochet, J. (1983). Electron microscopy of frozen hydrated sections of vitreous ice and vitrified biological samples. Journal of Microscopy *131*, 1–9.

McGreevy, R., Teo, I., Singharoy, A., and Schulten, K. (2016). Advances in the molecular dynamics flexible fitting method for cryo-EM modeling. Methods

100, 50–60.

McMullan, G., Chen, S., Henderson, R., and Faruqi, A.R. (2009a). Detective quantum efficiency of electron area detectors in electron microscopy. Ultramicroscopy *109*, 1126–1143.

McMullan, G., Clark, A.T., Turchetta, R., and Faruqi, A.R. (2009b). Enhanced imaging in low dose electron microscopy using electron counting. Ultramicroscopy *109*, 1411–1416.

McMullan, G., Faruqi, A.R., Clare, D., and Henderson, R. (2014). Comparison of optimal performance at 300keV of three direct electron detectors for use in low dose electron microscopy. Ultramicroscopy *147*, 156–163.

McMullan, G., Faruqi, A.R., Henderson, R., Guerrini, N., Turchetta, R., Jacobs, A., and van Hoften, G. (2009c). Experimental observation of the improvement in MTF from backthinning a CMOS direct electron detector. Ultramicroscopy *109*, 1144–1147.

Meier, T., Polzer, P., Diederichs, K., Welte, W., and Dimroth, P. (2005). Structure of the Rotor Ring of F-Type Na+-ATPase from Ilyobacter tartaricus. Science *308*, 659–662.

Meng, E.C., Pettersen, E.F., Couch, G.S., Huang, C.C., and Ferrin, T.E. (2006). Tools for integrated sequence-structure analysis with UCSF Chimera. BMC Bioinformatics *7*, 339.

Merk, A., Bartesaghi, A., Banerjee, S., Falconieri, V., Rao, P., Davis, M.I., Pragani, R., Boxer, M.B., Earl, L.A., Milne, J.L.S., et al. (2016). Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. Cell *165*, 1698–1707.

Michel, V., Licon-Munoz, Y., Trujillo, K., Bisoffi, M., and Parra, K.J. (2013). Inhibitors of vacuolar ATPase proton pumps inhibit human prostate cancer cell invasion and prostate-specific antigen expression and secretion. Int. J. Cancer *132*, E1–E10.

Milne, J.L.S., Shi, D., Rosenthal, P.B., Sunshine, J.S., Domingo, G.J., Wu, X., Brooks, B.R., Perham, R.N., Henderson, R., and Subramaniam, S. (2002). Molecular architecture and mechanism of an icosahedral pyruvate dehydrogenase complex: a multifunctional catalytic machine. Embo J. *21*, 5587–5598.

Mindell, J.A., and Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. J Struct Biol *142*, 334–347.

Moffat, J.G., Rudolph, J., and Bailey, D. (2014). Phenotypic screening in cancer drug discovery - past, present and future. Nat Rev Drug Discov *13*, 588–602.

Mogi, T., and Kita, K. (2009). Identification of mitochondrial Complex II subunits SDH3 and SDH4 and ATP synthase subunits a and b in Plasmodium spp. Mitochondrion *9*, 443–453.

Morales-Rios, E., Montgomery, M.G., Leslie, A.G.W., and Walker, J.E. (2015). Structure of ATP synthase from Paracoccus denitrificansdetermined by X-ray crystallography at 4.0 Å resolution. Proceedings of the National Academy of Sciences of the United States of America 201517542–201517546.

Muench, S.P., Rawson, S., Eyraud, V., Delmas, A.F., Da Silva, P., Phillips, C., Trinick, J., Harrison, M.A., Gressent, F., and Huss, M. (2014a). PA1b Inhibitor Binding to Subunits c and e of the Vacuolar ATPase Reveals Its Insecticidal Mechanism. J. Biol. Chem. *289*, 16399–16408.

Muench, S.P., Scheres, S.H.W., Huss, M., Phillips, C., Vitavska, O., Wieczorek, H., Trinick, J., and Harrison, M.A. (2014b). Subunit positioning and stator filament stiffness in regulation and power transmission in the V1 motor of the Manduca sexta V-ATPase. J. Mol. Biol. *426*, 286–300.

Muench, S.P., Trinick, J., and Harrison, M.A. (2011). Structural divergence of the rotary ATPases. Q. Rev. Biophys. *44*, 311–356.

Murshudov, G.N., Skubak, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallographica. Section D, Biological Crystallography *67*, 355–367.

Mühleip, A.W., Joos, F., Wigge, C., Frangakis, A.S., Kühlbrandt, W., and Davies, K.M. (2016). Helical arrays of U-shaped ATP synthase dimers form tubular cristae in ciliate mitochondria. Proceedings of the National Academy of Sciences of the United States of America *113*, 8442–8447.

Nakanishi-Matsui, M., Sekiya, M., Nakamoto, R.K., and Futai, M. (2010). The mechanism of rotating proton pumping ATPases. Biochim. Biophys. Acta *1797*, 1343–1352.

Niedzialkowska, E., Gasiorowska, O., Handing, K.B., Majorek, K.A., Porebski, P.J., Shabalin, I.G., Zasadzinska, E., Cymborowski, M., and Minor, W. (2016). Protein purification and crystallization artifacts: The tale usually not told. Protein Science *25*, 720–733.

Niikura, K., Takano, M., and Sawada, M. (2004). A novel inhibitor of vacuolar ATPase, FR167356, which can discriminate between osteoclast vacuolu ATPase and lysosomal vacuolar ATPase. British Journal of Pharmacology *142*, 558–566.

Nogales, E. (2016). The development of cryo-EM into a mainstream structural biology technique. Nat. Methods *13*, 24–27.

Ohi, M., Li, Y., Cheng, Y., and Walz, T. (2004). Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy. Biol Proced Online *6*, 23–34.

Oot, R.A., Huang, L.-S., Berry, E.A., and Wilkens, S. (2012). Crystal Structure of the Yeast Vacuolar ATPase Heterotrimeric EGChead Peripheral Stalk Complex. Structure *20*, 1881–1892.

Oot, R.A., Kane, P.M., Berry, E.A., and Wilkens, S. (2016). Crystal structure of yeast V 1-ATPase in the autoinhibited state. Embo J. e201593447–13.

Orlova, E.V., and Saibil, H.R. (2011). Structural Analysis of Macromolecular Assemblies by Electron Microscopy. Chem. Rev. *111*, 7710–7748.

Osteresch, C., Bender, T., Grond, S., Zezschwitz, von, P., Kunze, B., Jansen, R., Huss, M., and Wieczorek, H. (2012). The Binding Site of the V-ATPase Inhibitor Apicularen Is in the Vicinity of Those for Bafilomycin and Archazolid. Journal of Biological Chemistry *287*, 31866–31876.

Overington, J.P., Al-Lazikani, B., and Hopkins, A.L. (2006). How many drug targets are there? Nat Rev Drug Discov *5*, 993–996.

O'Reilly, M., Vinković, M., Sharff, A., and Jhoti, H. (2007). High throughput protein crystallography: Developments in crystallisation, data collection and data processing. Drug Discov Today Technol *3*, 451–456.

Parra, K.J., Keenan, K.L., and Kane, P.M. (2000). The H subunit (Vma13p) of the yeast V-ATPase inhibits the ATPase activity of cytosolic V1 complexes. Journal of Biological Chemistry *275*, 21761–21767.

Patwardhan, A., Carazo, J.M., Carragher, B., Henderson, R., Heymann, J.B., Hill, E., Jensen, G.J., Lagerstedt, I., Lawson, C.L., Ludtke, S.J., et al. (2012). Data management challenges in three-dimensional EM. Nature Structural & Molecular Biology *19*, 1203–1207.

Paulsen, C.E., Armache, J.-P., Gao, Y., Cheng, Y., and Julius, D. (2015). Structure of the TRPA1 ion channel suggests regulatory mechanisms. Nature *520*, 511–517.

Perez-Sayans, M., Manuel Somoza-Martin, J., Barros-Angueira, F., Gandara Rey, J.M., and Garcia-Garcia, A. (2009). V-ATPase inhibitors and implication in cancer treatment. Cancer Treatment Reviews *35*, 707–713.

Perutz, M.F., Rossmann, M.G., Cullis, A.F., Muirhead, H., Will, G., and North, A.C. (1960). Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis. Nature *185*, 416–422.

Petersen, J., Hawkes, T.R., and Lowe, D.J. (1997). The metal-binding site of imidazole glycerol phosphate dehydratase; EPR and ENDOR studies of the oxo-vanadyl enzyme. Journal of Biological Inorganic Chemistry *2*, 308–319.

Pogoryelov, D., Yu, J., Meier, T., Vonck, J., Dimroth, P., and Muller, D.J. (2005). The c15 ring of the Spirulina platensis F-ATP synthase: F1/F0 symmetry mismatch is not obligatory. EMBO Reports *6*, 1040–1044.

Postis, V., Rawson, S., Mitchell, J.K., Lee, S.C., Parslow, R.A., Dafforn, T.R., Baldwin, S.A., and Muench, S.P. (2015). The use of SMALPs as a novel membrane protein scaffold for structure study by negative stain electron microscopy. Biochim. Biophys. Acta *1848*, 496–501.

Prabudiansyah, I., Kusters, I., Caforio, A., and Driessen, A.J.M. (2015). Characterization of the annular lipid shell of the Sec translocon. Biochim. Biophys. Acta *1848*, 2050–2056.

Rangarajan, E.S., Proteau, A., Wagner, J., Hung, M.-N., Matte, A., and Cygler, M. (2006). Structural Snapshots of Escherichia coli Histidinol Phosphate Phosphatase along the Reaction Pathway. Journal of Biological Chemistry *281*, 37930–37941.

Rawson, S., Davies, S., Lippiat, J.D., and Muench, S.P. (2016a). The changing landscape of membrane protein structural biology through developments in electron microscopy. Mol. Membr. Biol. 1–11.

Rawson, S., Harrison, M.A., and Muench, S.P. (2016b). Rotating with the brakes on and other unresolved features of the vacuolar ATPase. Biochem. Soc. Trans. *44*, 851–855.

Rawson, S., Iadanza, M.G., Ranson, N.A., and Muench, S.P. (2016c). Methods to account for movement and flexibility in cryo-EM data processing. Methods *100*, 35–41.

Rawson, S., Phillips, C., Huss, M., Tiburcy, F., Wieczorek, H., Trinick, J., Harrison, M.A., and Muench, S.P. (2015). Structure of the Vacuolar H(+)-ATPase Rotary Motor Reveals New Mechanistic Insights. Structure *23*, 461–471.

Razinkov, I., Dandey, V.P., Wei, H., Zhang, Z., Melnekoff, D., Rice, W.J., Wigge, C., Potter, C.S., and Carragher, B. (2016). A new method for vitrifying samples for cryoEM. J Struct Biol *195*, 190–198.

Reynolds, C.H. (2014). Protein–Ligand Cocrystal Structures: We Can Do Better. ACS Med. Chem. Lett. *5*, 727–729.

Richardson, R.A., Papachristos, K., Read, D.J., Harlen, O.G., Harrison, M., Paci, E., Muench, S.P., and Harris, S.A. (2014). Understanding the apparent stator-rotor connections in the rotary ATPase family using coarse-grained computer modeling. Proteins *82*, 3298–3311.

Rohou, A., and Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J Struct Biol *192*, 216–221.

Roseman, A.M. (2000). Docking structures of domains into maps from cryo-electron microscopy using local correlation. Acta Crystallographica. Section D, Biological Crystallography *56*, 1332–1340.

Rosenthal, P.B., and Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. J. Mol. Biol. *333*, 721–745.

Rubinstein, J.L., and Brubaker, M.A. (2015). Alignment of cryo-EM movies of individual particles by optimization of image translations. J Struct Biol *192*, 188–195.

Russo, C.J., and Passmore, L.A. (2014). Electron microscopy: Ultrastable gold substrates for electron cryomicroscopy. Science *346*, 1377–1380.

Russo, C.J., and Passmore, L.A. (2016). Ultrastable gold substrates: Properties of a support for high-resolution electron cryomicroscopy of biological specimens. J Struct Biol *193*, 33–44.

Sambade, M., and Kane, P.M. (2004). The yeast vacuolar proton-translocating ATPase contains a subunit homologous to the Manduca sexta and bovine e subunits that is essential for function. Journal of Biological Chemistry *279*, 17361–17365.

Schep, D.G., Zhao, J., and Rubinstein, J.L. (2016). Models for the a subunits of the Thermus thermophilusV/A-ATPase and Saccharomyces cerevisiaeV-ATPase enzymes by cryo-EM and evolutionary covariance. Proceedings of the National Academy of Sciences of the United States of America *113*, 3245–3250.

Scheres, S.H.W. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol *180*, 519–530.

Scheres, S.H.W. (2014). Beam-induced motion correction for sub-megadalton cryo-EM particles. eLife *3*, e03665.

Scheres, S.H.W. (2015). Semi-automated selection of cryo-EM particles in RELION-1.3. J Struct Biol *189*, 114–122.

Scheres, S.H.W., and Chen, S. (2012). Prevention of overfitting in cryo-EM structure determination. Nat. Methods *9*, 853–854.

Scheres, S.H.W., Nunez-Ramirez, R., Sorzano, C.O.S., Maria Carazo, J., and Marabini, R. (2008). Image processing for electron microscopy single-particle analysis using XMIPP. Nature Protocols *3*, 977–990.

Seeber, F., Limenitakis, J., and Soldati-Favre, D. (2008). Apicomplexan mitochondrial metabolism: a story of gains, losses and retentions. Trends Parasitol. *24*, 468–478.

Sennoune, S.R., Bakunts, K., Martinez, G.M., Chua-Tuan, J.L., Kebir, Y., Attaya, M.N., and Martinez-Zaguilan, R. (2004). Vacuolar H+-ATPase in human breast cancer cells with distinct metastatic potential: distribution and functional activity. American Journal of Physiology-Cell Physiology *286*, C1443–C1452.

Shatsky, M., Hall, R.J., Brenner, S.E., and Glaeser, R.M. (2009). A method for the alignment of heterogeneous macromolecules from electron microscopy. J Struct Biol *166*, 67–78.

Siegert, K.J. (1995). Carbohydrate metabolism during the pupal molt of the tobacco hornworm, Manduca sexta. Arch. Insect Biochem. Physiol. *28*, 63–78.

Sigworth, F.J. (1998). A Maximum-Likelihood Approach to Single-Particle Image Refinement. J Struct Biol *122*, 328–339.

Sigworth, F.J., Doerschuk, P. C., Carazo, J.M., and Scheres, S.H.W. (2010). An Introduction to Maximum-Likelihood Methods in Cryo-Em. Method Enzymol *482*, 263–294.

Smardon, A.M., Tarsio, M., and Kane, P.M. (2002). The RAVE complex is essential for stable assembly of the yeast V-ATPase. Journal of Biological Chemistry *277*, 13831–13839.

Smardon, A.M., and Kane, P.M. (2007). RAVE is essential for the efficient assembly of the C subunit with the vacuolar H+-ATPase. Journal of Biological Chemistry *282*, 26185–26194.

Smith, G.A., Howell, G.J., Phillips, C., Muench, S.P., Ponnambalam, S., and Harrison, M.A. (2016). Extracellular and Luminal pH Regulation by Vacuolar H+-ATPase Isoform Expression and Targeting to the Plasma Membrane and Endosomes. J. Biol. Chem. *291*, 8500–8515.

Smith, M.T.J., and Rubinstein, J.L. (2014). Structural Biology. Beyond blob-ology. Science *345*, 617–619.

Song, C.F., Papachristos, K., Rawson, S., Huss, M., Wieczorek, H., Paci, E., Trinick, J., Harrison, M.A., and Muench, S.P. (2013). Flexibility within the rotor and stators of the vacuolar H+-ATPase. PLoS ONE *8*, e82207.

Spence, J.C.H. (1999). The future of atomic resolution electron microscopy for materials science. Materials Science and Engineering: R: Reports *26*, 1–49.

Stark, H., Zemlin, F., and Boettcher, C. (1996). Electron radiation damage to protein crystals of bacteriorhodopsin at different temperatures. Ultramicroscopy *63*, 75–79.

Steuber, H., Zentgraf, M., Gerlach, C., Sotriffer, C.A., Heine, A., and Klebe, G. (2006). Expect the Unexpected or Caveat for Drug Designers: Multiple Structure Determinations Using Aldose Reductase Crystals Treated under Varying Soaking and Co-crystallisation Conditions. J. Mol. Biol. *363*, 174–187.

Stewart, A.G., Lee, L.K., Donohoe, M., Chaston, J.J., and Stock, D. (2012). The dynamic stator stalk of rotary ATPases. Nat Commun *3*, 687.

Stewart, A.G., Sobti, M., Harvey, R.P., and Stock, D. (2013). Rotary ATPases: models, machine elements and technical specifications. Bioarchitecture *3*, 2–12.

Stransky, L.A., and Forgac, M. (2015). Amino Acid Availability Modulates Vacuolar H+-ATPase Assembly. J. Biol. Chem. *290*, 27360–27369.

Sturm, A., Mollard, V., Cozijnsen, A., Goodman, C.D., and McFadden, G.I. (2015). Mitochondrial ATP synthase is dispensable in blood-stage Plasmodium bergheirodent malaria but essential in the mosquito phase. Proceedings of the National Academy of Sciences of the United States of America *112*, 10216–10223.

Suloway, C., Pulokas, J., Fellmann, D., Cheng, A., Guerra, F., Quispe, J., Stagg, S., Potter, C.S., and Carragher, B. (2005). Automated molecular microscopy: the new Leginon system. J Struct Biol *151*, 41–60.

Sumner, J.P., Dow, J.A.T., Earley, F.G.P., Klein, U., Jager, D., and Wieczorek, H. (1995). Regulation of Plasma Membrane V-ATPase Activity by Dissociation of Peripheral Subunits. Journal of Biological Chemistry *270*, 5649–5653.

Søndergaard, C.R., Garrett, A.E., Carstensen, T., Pollastri, G., and Nielsen, J.E. (2009). Structural Artifacts in Protein−Ligand X-ray Structures: Implications for the Development of Docking Scoring Functions. Journal of Medicinal Chemistry *52*, 5673–5684.

Tabke, K., Albertmelcher, A., Vitavska, O., Huss, M., Schmitz, H.-P., and Wieczorek, H. (2014). Reversible disassembly of the yeast V-ATPase revisited under in vivo conditions. Biochem. J. *462*, 185–197.

Tada, S., Hatano, M., Nakayama, Y., Volrath, S., Guyer, D., Ward, E., and Ohta, D. (1995). Insect Cell Expression of Recombinant Imidazoleglycerolphosphate Dehydratase of Arabidopsis and Wheat and Inhibition by Triazole Herbicides. Plant Physiology *109*, 153–159.

Tamura, T., Tsukiji, S., and Hamachi, I. (2012). Native FKBP12 Engineering by Ligand-Directed Tosyl Chemistry: Labeling Properties and Application to Photo-Cross-Linking of Protein Complexes in Vitro and in Living Cells. Journal of the American Chemical Society *134*, 2216–2226.

Tan, Y.Z., Cheng, A., Potter, C.S., and Carragher, B. (2016). Automated data collection in single particle electron microscopy. Microscopy (Oxf) *65*, 43–56.

Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., and Ludtke, S.J. (2007). EMAN2: An extensible image processing suite for electron microscopy. J Struct Biol *157*, 38–46.

Tani, K., Arthur, C.P., Tamakoshi, M., Yokoyama, K., Mitsuoka, K., Fujiyoshi, Y., and Gerle, C. (2013). Visualization of two distinct states of disassembly in the bacterial V-ATPase from Thermus thermophilus. Microscopy (Oxf) *62*, 467–474.

Tate, C.G. (2010). Practical considerations of membrane protein instability during purification and crystallisation. Methods Mol. Biol. *601*, 187–203.

Teague, S.J. (2003). Implications of protein flexibility for drug discovery. Nat Rev Drug Discov 2, 527–541.

Thompson, R.F., Walker, M., Siebert, C.A., Muench, S.P., and Ranson, N.A. (2016). An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology. Methods *100*, 3–15.

Topf, M., Lasker, K., Webb, B., Wolfson, H., Chiu, W., and Sali, A. (2008). Protein Structure Fitting and Refinement Guided by Cryo-EM Density.

Structure *16*, 295–307.

Trabuco, L.G., Schreiner, E., Gumbart, J., Hsin, J., Villa, E., and Schulten, K. (2011). Applications of the molecular dynamics flexible fitting method. J Struct Biol *173*, 420–427.

Unwin, P.N.T., and Henderson, R. (1975). Molecular structure determination by electron microscopy of unstained crystalline specimens. J. Mol. Biol. *94*, 425–440.

Urban, K.W. (2009). Is science prepared for atomic-resolution electron microscopy? Nat Mater *8*, 260–262.

Urnavicius, L., Zhang, K., Diamant, A.G., Motz, C., Schlager, M.A., Yu, M., Patel, N.A., Robinson, C.V., and Carter, A.P. (2015). The structure of the dynactin complex and its interaction with dynein. Science *347*, 1441–1446.

van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R., and Schatz, M. (1996). A new generation of the IMAGIC image processing system. J Struct Biol *116*, 17–24.

van Heel, M. (2013). Finding trimeric HIV-1 envelope glycoproteins in random noise. Proceedings of the National Academy of Sciences *110*, E4175–E4177.

Vinothkumar, K.R., and Henderson, R. (2016). Single particle electron cryomicroscopy: trends, issues and future perspective. Q. Rev. Biophys. *49*.

Vitavska, O., Wieczorek, H., and Merzendorfer, H. (2003). A Novel Role for Subunit C in Mediating Binding of the H+-V-ATPase to the Actin Cytoskeleton. Journal of Biological Chemistry *278*, 18499–18505.

Vollmar, M., Schlieper, D., Winn, M., Büchner, C., and Groth, G. (2009). Structure of the c14 rotor ring of the proton translocating chloroplast ATP synthase. J. Biol. Chem. *284*, 18228–18235.

Wagenknecht, T., Grassucci, R., and Schaak, D. (1990). Cryoelectron microscopy of frozen-hydrated alpha-ketoacid dehydrogenase complexes from Escherichia coli. Journal of Biological Chemistry *265*, 22402–22408.

Walker, J.E. (2013). The ATP synthase: the understood, the uncertain and the unknown. Biochem. Soc. Trans. *41*, 1–16.

Wang, L., and Sigworth, F.J. (2009). Structure of the BK potassium channel in a lipid membrane from electron cryomicroscopy. Nature *461*, 292–295.

Wang, L., and Sigworth, F.J. (2010). Liposomes on a Streptavidin Crystal: a System to Study Membrane Proteins by Cryo-Em. Method Enzymol *481*, 147–164.

Wang, R.Y.-R., Kudryashev, M., Li, X., Egelman, E.H., Basler, M., Cheng, Y., Baker, D., and DiMaio, F. (2015). De novo protein structure determination from near-atomic-resolution cryo-EM maps. Nat. Methods *12*, 335–338.

Wang, Y.R., Inoue, T., and Forgac, M. (2005). Subunit a of the yeast V-ATPase participates in binding of bafilomycin. Journal of Biological Chemistry *280*, 40481–40488.

Wasilewski, S., and Rosenthal, P.B. (2014). Web server for tilt-pair validation of single particle maps from electron cryomicroscopy. J Struct Biol *186*, 122–131.

Watanabe, R., Koyasu, K., You, H., Tanigawara, M., and Noji, H. (2015). Torque Transmission Mechanism via DELSEED Loop of F1-ATPase. Biophysj *108*, 1144–1152.

Weik, M., Ravelli, R.B., Kryger, G., McSweeney, S., Raves, M.L., Harel, M., Gros, P., Silman, I., Kroon, J., and Sussman, J.L. (2000). Specific chemical and structural damage to proteins produced by synchrotron radiation. Proceedings of the National Academy of Sciences of the United States of America *97*, 623–628.

Wilkinson, K.W., Baker, P.J., Rice, D.W., Rodgers, H.F., Stillman, T.J., Hawkes, T., Thomas, P., and Edwards, L. (1995). Crystallization and analysis of the subunit assembly and quaternary structure of imidazoleglycerol phosphate dehydratase from Saccharomyces cerevisiae. Acta Crystallographica. Section D, Biological Crystallography *51*, 845–847.

Wriggers, W. (2012). Conventions and workflows for using Situs. Acta Crystallographica. Section D, Biological Crystallography *68*, 344–351.

Wu, J., Yan, Z., Li, Z., Yan, C., Lu, S., Dong, M., and Yan, N. (2015). Structure of the voltage-gated calcium channel Cav1.1 complex. Science *350*, aad2395–aad2395.

Yan, Z., Bai, X.-C., Yan, C., Wu, J., Li, Z., Xie, T., Peng, W., Yin, C.-C., Li, X., Scheres, S.H.W., et al. (2015). Structure of the rabbit ryanodine receptor RyR1 at near-atomic resolution. Nature *517*, 50–55.

Yasuda, R., Noji, H., Yoshida, M., Kinosita, K., and Itoh, H. (2001). Resolution of distinct rotational substeps by submillisecond kinetic analysis of F1-ATPase. Nature *410*, 898–904.

Yoshioka, C., Carragher, B., and Potter, C.S. (2010). Cryomesh: a new substrate for cryo-electron microscopy. Microsc. Microanal. *16*, 43–53.

Yu, G., Li, K., and Jiang, W. (2016). Antibody-based affinity cryo-EM grid. Methods *100*, 16–24.

Zhang, J., Feng, Y., and Forgac, M. (1994). Proton conduction and bafilomycin binding by the V0 domain of the coated vesicle V-ATPase. Journal of Biological Chemistry *269*, 23518–23523.

Zhang, K. (2016). Gctf: Real-time CTF determination and correction. J Struct Biol

193, 1–12.

Zhao, J., Benlekbir, S., and Rubinstein, J.L. (2015). Electron cryomicroscopy observation of rotational states in a eukaryotic V-ATPase. Nature *521*, 241–245.

Zhou, A., Rohou, A., Schep, D.G., Bason, J.V., Montgomery, M.G., Walker, J.E., Grigorieff, N., and Rubinstein, J.L. (2015). Structure and conformational states of the bovine mitochondrial ATP synthase by cryo-EM. eLife *4*, e10180.

Zhou, M., Politis, A., Davies, R.B., Liko, I., Wu, K.-J., Stewart, A.G., Stock, D., and Robinson, C.V. (2014). Ion mobility-mass spectrometry of a rotary ATPase reveals ATP-induced reduction in conformational flexibility. Nat. Chem. *6*, 208–215.

Zsoldos, Z., Reid, D., Simon, A., Sadjad, S.B., and Johnson, A.P. (2007). eHiTS: A new fast, exhaustive flexible ligand docking system. Journal of Molecular Graphics & Modelling *26*, 198–212.