Investigating the role of cryo-EM as a tool for structure-based drug design

Rachel May Johnson

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 2019

Intellectual property and publication

The candidate confirms that the work submitted is her 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.

Chapter 1 contains work from two jointly authored review papers and one technology of the month article:

Johnson, R. M., Gordon, J. A., Rawson, S., McPhillie, M. J., Fishwick, C. W. G., Muench, S. P. *The growing role of electron microscopy in anti-parasitic drug discovery*. Current medicinal chemistry. 25 (39), 5279-5290 (2018). The candidate wrote the initial draft of the manuscript, produced the figures and was involved in the subsequent editing. Muench helped to write and edit the manuscript and all authors were involved in proof-reading of the manuscript.

Rawson, S., McPhillie, M. J., **Johnson, R. M**., Fishwick, C. W. G. & Muench, S. P. *The potential use of single-particle electron microscopy as a tool for structure-based inhibitor design*. Acta Crystallogr. Sect. D, D73, 1–7 (2017). The candidate was involved in writing a small section of the paper as Rawson had already drafted the majority of the manuscript. The candidate was involved in proof reading and editing of the manuscript. Muench, Fishwick and McPhillie were involved in editing of the manuscript.

Johnson, R. M., Higgins, A. J., and Muench, S. P. *Emerging role of electron microscopy in drug discovery*. Trends in Biochemical Sciences, August 2019. *In press*. The candidate was involved in drafting, editing and proof-reading the manuscript with Higgins and Muench. The candidate made Figure 2 where part of the Figure has been used in Figure 1.11 within this thesis.

Chapter 2 contains work from one jointly authored review paper:

Drulyte, I., **Johnson, R. M.**, Hesketh, E. L., Hurdiss, D. L., Scarff, C. A., Porav, S. A., Ranson, N. A., Muench, S. P. & Thompson, R. F. *Approaches to altering particle distributions in cryo-electron microscopy sample preparation*. Acta Crystallogr. Sect. D Struct. Biol. 74, 1–12 (2018). The candidate was involved in writing her section of the manuscript as the majority of the manuscript had already been drafted by Drulyte, Muench and Thompson. All authors were responsible for their sections of the manuscript and the proof-reading/editing.

Chapter 3 contains data from one jointly authored paper published in 2018:

Amporndanai, K*., **Johnson, R. M*.**, O'Neill, P. M., Fishwick, C. W. G., Jamson, A. H., Rawson, S., Muench, S. P., Samar Hasnain, S. & Antonyuk, S. V. *X-ray and cryo-EM structures of inhibitor-bound cytochrome bc_1 complexes for structure-based drug discovery*. IUCrJ. 5, 200–210 (2018). The candidate completed all of the EM for the manuscript including optimising the EM grids, collecting the data, image processing and subsequent structure interpretation with guidance from Rawson and Muench. Amporndanai provided the protein, carried out the crystallography experiments and was also involved in data analysis. The candidate and Amporndanai were involved in the initial drafting of the manuscript and the candidate made all of the figures except for Figure 1A. Muench, Fishwick, Hasnain and Antonyuk supervised the project and were involved in writing/editing the manuscript.

The appendix contains data which has been published in one paper: Gopalasingam, C. C., **Johnson, R. M.**, Chiduza, G. N., Tosha T., Yamamoto M., Shiro, Y., Antonyuk, S. V., Muench, S. P., and Hasnain, S.S. *Dimeric structures of quinol-dependent Nitric Oxide Reductases (qNOR) revealed by cryo-Electron Microscopy.* Science Advances. 5, (8), eaax1803 (2019). The candidate was involved in cryo-grid optimisation, data acquisition and image processing. The protein purification, structure interpretation and drafting of the manuscript (including figure making) was performed by C.C.G.. The candidate was only involved in proof reading the manuscript. S.P.M, S.V.A. and S.S.H designed the experiments.

The right of Rachel May Johnson to be identified as Author of this work has been asserted by her in accordance with the Copyright, Designs and Patents Act 1988.

Acknowledgements

Firstly, I would like to thank my supervisors, Stephen Muench and Colin Fishwick, for not only providing me with the opportunity to work on this project but also for all of their excellent help and guidance throughout my PhD. I am also very grateful to Mike Harrison for helping me to optimise the V-ATPase purifications and to David Klebl for his fantastic assistance when completing the preps. Thanks to Martin McPhillie, Katie Simmons and Chris Fitzpatrick for all of their help and advice relating to the docking and synthesis.

Thank you also to everybody who I have collaborated with over the past few years. This has included Samar Hasnain, Svetlana Antonyuk and the members of their groups at the University of Liverpool who have visited Leeds with great protein samples. In particular, thanks to Kangsa Amporndanai for providing cytochrome bc_1 (for the work described in Chapter 3), George Chiduza for LAT1 and Chai Gopalasingham for qNOR.

Thank you to all members of the Muench, EM and Fishwick groups, past and present, for their help over the years. A huge thank you to Shaun Rawson for his patience when teaching me how to process data and for fixing all of my many IT-related issues. Dan Hurdiss, despite you never understanding why I don't like tea, I am very grateful to you for sharing your figure-making expertise in particular for helping to create the IUCrJ cover. To Becky Thompson, Emma Hesketh, Dan Maskell and Martin Fuller, thank you for your fantastic help and support within the facility. To all the members of all of the groups, (in particular Matt Iadanza, Dave Nicholson, Josh White, David Carrier and Sophie Hesketh), thank you for all of your help, the insightful discussions and for creating a happy environment to work in. A big thank you also to Brendan Farrell and Emma Cawood for always being there for me throughout the past four years. I have particularly valued our F.R.I.E.N.D.S.-hip, especially when we were on a break from science and enjoyed a day-of-fun outside of the lab.

Last but not least, I would like to thank all of my friends and family for their support in allowing me to get to where I am today. I hope that if you read this thesis you will finally have the answer to the question "What is it that you actually do again?".

Abstract

Structure based drug design (SBDD) programs have traditionally relied upon structural information derived from X-ray crystallography. However, despite the high-resolution structures, the information obtained is a static snapshot of flexible biological systems. As the proteins are not constrained by a crystal lattice, EM can provide unique insights into the dynamics and conformational states of the system. Moreover, the structures of biological systems which were previously intractable to X-ray crystallography projects, such as membrane proteins, have now been determined providing a wealth of structural information to be utilised in SBDD programs.

This thesis presents two examples of cryo-EM being used to aid SBDD for different membrane proteins. One example is Cytochrome bc_1, which is a validated anti-parasitic drug target. Cryo-EM has been used in a proof of principle approach to determine four structures of bovine bc_1 in the absence and presence of four different inhibitors to ~3.3-4.6 Å resolution. The resolutions attained have enabled the inhibitors to be unambiguously positioned in the density, thereby Chapter 3 provides one of the first examples of inhibitor density being visualised using cryo-EM.

The second system studied is the Vacuolar ATPase (V-ATPase) which is at an early-stage of the drug discovery pipeline. By using a virtual screening approach, several hit compounds have been identified to act upon V-ATPase. Derivatives of the hit compound have been synthesised, resulting in three compounds which have <100 µM IC_{50} values. Moreover, the cryo-EM structure of V-ATPase in the apo and inhibitor bound form were determined. Although resolutions were not sufficient to visualise inhibitor binding, a novel state of V-ATPase has been identified which shows the complex without the H-subunit present. Overall the work in this thesis shows that cryo-EM can play an important role in SBDD programs.

Contents Page

Abbreviations

ACIII	Alternative complex III
ADMET	Adsorption, metabolism and toxicity
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
bc_1	Cytochrome bc_1 complex
BCA assay	Bicinchoninic acid assay
CCD	Charge coupled device
COSY	Correlated spectroscopy
Cryo-EM	Cryogenic electron microscopy
crYOLO	cr-You Only Look Once
CryoSPARC	cryo-EM single-particle *ab initio* reconstruction and classification
CTF	Contrast transfer function
DCCD	Dicycloxylcarbodiimide
DHODH	Dihydroorotate dehydrogenase
DDM	n-dodecyl-beta-D-maltoside
DMEM	Dulbecco's modified Eagle's medium
DTT	Dithiothreitol
DQE	Detective quantum efficiency
EM	Electron microscopy
EMDB	Electron microscopy data bank
ETC	Electron transport chain
EPU	*E pluribus unum*
FBDD	Fragment-based drug design
FDA	Food and Drug Administration
FEG	Field emission gun
FITH	First time in human
FSC	Fourier shell correlation
FT	Fourier transform
GPCRs	G-protein coupled receptors
GSK	GlaxoSmithKline
HCMV	Human cytomegalovirus
HEK	Human embryonic kidney cells
HPLC	High performance liquid chromatography
HRMS	High resolution mass spectrometry
HSP	Heat shock protein
HSQC	Heteronuclear single quantum correlation

HTS	High throughput screening
IC_{50}	Half maximal inhibitory concentration
IR	Infrared
ITC	Isothermal calibration
K_d	Equilibrium dissociation constant
K_i	Equilibrium inhibition constant
LCMS	Liquid chromatography mass spectrometry
LDC	Lead Discovery Centre
LMNG	Lauryl Maltose Neopentyl Glycol
MAPs	Monolithic active pixel sensors
MDFF	Molecular dynamics flexible fitting
MTT	3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromidefor
NMR	Nuclear magnetic resonance
PA1b	Pea Albumin Subunit b
PDB	Protein data bank
PPI	Protein-protein interactions
ppm	Parts per million
RELION	Regularised likelihood optimisation
RMSD	Root mean square deviation
ROCs	Rapid overlay chemical screening
ROS	Reactive oxygen screening
rpm	Revolutions per minute
RT	Room temperature
SAR	Structure activity relationship
SBDD	Structure based drug design
SMALPs	Styrene maleic acid lipid particles
SNR	Signal-to-noise ratio
SPA	Single particle averaging
SPR	Surface plasmon resonance
TLC	Thin layer chromatography
TRP	Transient receptor potential channels
UV	Ultraviolet
V-ATPase	Vacuolar ATPase
vHTS	Virtual high throughput screens
VPP	Volta Phase Plate
YEPD	Yeast extract peptone dextrose

List of Figures

List of Tables

1   Introduction

1.1 Drug discovery pipeline

New therapeutic agents need to be continually developed to treat a wide range of disorders such as infectious diseases, multiple types of cancers and neurological conditions such as Alzheimer's and Parkinson's disease. However, the process of drug discovery can be challenging, containing numerous hurdles and many unforeseen difficulties[1]. From the discovery of the initial 'hit' compound, it often takes large pharmaceutical companies ten to fifteen years for the compound to reach the clinic[2]. This emphasises the large amount of compound development and research hours needed to allow the initial hit compound to be transformed into a drug candidate, thus explaining why typical drug discovery programs cost billions of pounds. A general overview of the drug discovery pipeline is shown in Figure 1.1.
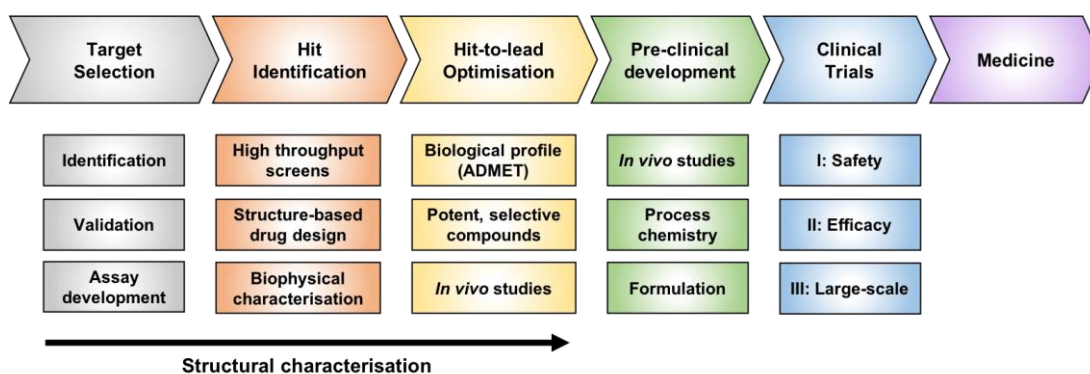
Figure 1.1: An overview of the drug discovery pipeline. *The keys steps in the process from target selection all the way through to medicine are highlighted within the arrow. The processes involved within each step are listed in the boxes below.*

The first step in drug discovery programs is to identify a suitable target for therapeutic agents to act upon. This step is called target selection and encompasses the identification and validation of the target which includes analysing the protein to identify a suitable ligand-binding site, whether it is druggable and if it can facilitate ligand binding[3].

Once a new target has been identified, the next step is to identify a compound which produces a response and is therefore termed a 'hit'[3]. Hit compounds can be identified through assays which either probe the protein directly or look for a cellular response. Alternatively, they are found through biophysical techniques such as surface plasmon resonance (SPR)[4] and isothermal calibration assays (ITC)[5]. Enzyme assays are often high throughput, use small amounts of protein and can be performed in 96-well plates allowing the activity of many compounds to be tested in a short time frame[6]. The assays rely on using light (either absorbance, fluorescence or luminescence) to measure the activity of the protein in the absence and presence of the compounds which determines whether the compounds are having the desired biological response[6]. These enzyme assays often generate IC_{50} curves. In comparison, the biophysical techniques such as SPR and ITC measure the enthalpic cost of the ligands upon binding to the target and therefore can be used to generate the inhibitory (K_i) and dissociation constants (K_d)[7]. To perform these techniques, generally more protein is required than the enzyme assays and they are lower throughput. Other biophysical and structural techniques such X-ray crystallography and Nuclear Magnetic Resonance (NMR) can also be used to identify hit compounds and are described in more detail in Section 1.2.2.

Once the initial hit compound has been identified, the next step is to turn this hit into a lead compound which is often termed 'hit-to-lead optimisation'[8]. At this stage, medicinal chemists can synthesise hundreds to thousands of compounds to fine-tune the chemical properties of the molecules in order to improve their potency, selectivity and pharmacokinetic properties. Moreover, the biological properties and safety of the lead compounds are investigated to determine the adsorption, metabolic and toxicity (ADMET) profiles of the compound[9]. Before the compound can be administered to humans, it is also subjected to *in vivo* testing in animal models (typically in rodents) which checks the compound is having the desired therapeutic effect and also investigates the toxicity of the compound. During the pre-clinical stage, process chemists analyse the synthesis of the desired compound so kilograms of material can be generated. After all of the rigorous biological and safety tests have been completed, the lead compound can then move forward through the pre-clinical stage and into clinical trials[10].

The clinical trials are split up into three main stages; Phase I, Phase II and Phase III[11]. Phase I enables the toxicity of the compound to be tested in healthy human volunteers by looking for any adverse side effects. Phase II then determines

whether the drug has the desired effect by administering the treatment to patients with the disease. This phase also establishes the correct dose to be administered to patients. Phase III involves using a much larger group of patients to determine the efficacy of the compound and compares the drug to existing treatments in the clinic. If the drug candidate passes all of these stages then it can be taken forward to the market. Due to the numerous safety and efficacy tests, there are many hurdles which the drug has to overcome before it can reach the market thus leading to high attrition rates. For instance, the likelihood of a drug being improved after entering Phase I clinical trials is less than 10%[12].

Identifying a hit compound is an important step at the beginning of the process. Traditionally, drug discovery programs relied upon high throughput screening (HTS) to identify novel inhibitor molecules but now structure based drug design (SBDD) has become a powerful complementary tool in the design of new therapeutic agents. For instance, between 2010-2016, 210 new drugs were approved by the FDA and for 88% of the protein targets, structures were published in the protein data bank (PDB)[13]. This exemplifies how useful the 3D structural information is when designing new therapeutic agents.

1.1.1 High-throughput Screens

High-throughput screening (HTS) is an automated process allowing for large compound libraries, consisting of millions of compounds, to be screened against a biological target. It has been the main technique for hit identification in the pharmaceutical industry for the past two decades[14,15]. Typically, 10,000 compounds are screened per day to identify hit compounds with those identified being taken forward into hit-to-lead development. A schematic of the process is shown in Figure 1.2.

High-throughput Screening:

Protein   Screening library of compounds   Hit identification
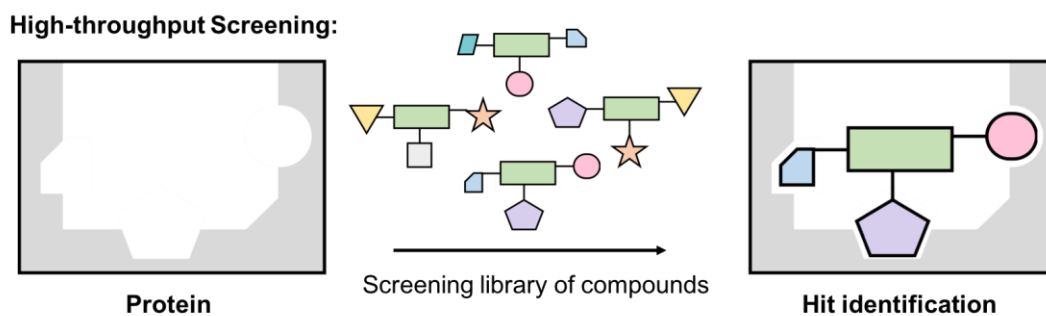
Figure 1.2: An overview of the HTS process*. A library of compounds is screened against the target of interest in order to identify a hit compound.*

As discussed above in Section 1.1, medicinal chemists can then improve the potency and selectivity of the compounds by conducting structure activity relationships (SAR) enabling a series of compounds to be designed which have a balanced profile of physicochemical properties. Despite the low probability of identifying a hit (~1%), this approach has been successful in numerous drug discovery programs[15] including the antiretroviral inhibitor Maraviroc[16], the protease inhibitor Tipranavir[17] and the antihyperglycemic inhibitor Sitagliptin[18] which are used to treat HIV infection and type II diabetes, respectively (Figure 1.3).

Maraviroc   **Tipranavir**   **Sitagliptin**
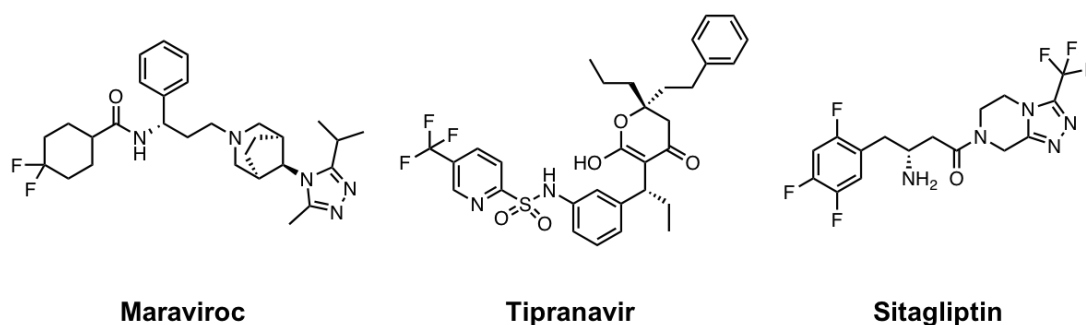
Figure 1.3: Example compounds identified through HTS. *The compounds include Maraviroc, Tipranavir and Sitagliptin.*

There are limitations to HTS which include the efficacy and effectiveness of the technique[19,20]. One of the main disadvantages is the large number of compounds (typically >1,000,000) which are screened in order to identify a hit compound. Accessing and maintaining these compound libraries are particularly challenging, especially to small pharmaceutical/biotech companies or academic groups. The

compounds within the libraries can also have poor physicochemical properties such as low solubility, or they could contain motifs which are associated with 'frequent hitter' behaviour, which results in a high percentage of false positive results[21,22]. Moreover, the large libraries only cover a small fraction of chemical space, which may hinder the discovery of inhibitors for targets with unusual binding sites such as allosteric sites or protein-protein interactions (PPIs). To reduce the cost of HTS, new strategies which involve using smaller numbers of more-diverse compounds are being developed[23–25].

1.1.2 Fragment-based drug design

Another technique which can be used to identify hit compounds is fragment-based drug design (FBDD) which has become an important tool in both academia and industry within the past ~20 years[26]. Fragment compounds, typically with a molecular weight of <250 kDa, are screened against the target protein with hit compounds being identified through biophysical techniques such as NMR or X-ray crystallography[27]. The fragments normally form a strong, high-quality interaction to the protein because there is a significant entropic barrier which has to be overcome to facilitate binding, yet the binding affinities are often only in the millimolar range[28]. X-ray crystallography is the mainstay for identifying hit fragments as hundreds of compounds can be screened per day. This is exemplified at the Diamond light source where there is a high-throughput XChem fragment screening program which allows users to screen ~500 crystals with different fragments bound per day[29]. The users can screen available fragment libraries or can alternatively use their own. The resulting data is automatically processed and algorithms such as panDAA have been developed to automatically detect density corresponding to the ligand[30].

The hit fragments identified in the screens are subsequently modified to improve the potency of the molecules[31]. This can be done by expanding the fragments or linking multiple hit fragments together which form interactions in different pockets within the binding site, subsequently improving the potency and binding affinity of the compound which in turn creates a lead molecule. An overview of this process is shown in Figure 1.4. The structural details of how the fragments bind to the protein, as identified by X-ray crystallography, can be utilised by the medicinal chemists in the hit-to-lead optimisation to create a high affinity compound which makes multiple interactions to the protein[31].

Fragment based drug design:

Protein — Screening library of fragments → Fragment hit identification — Synthesis to join fragments → Hit compound
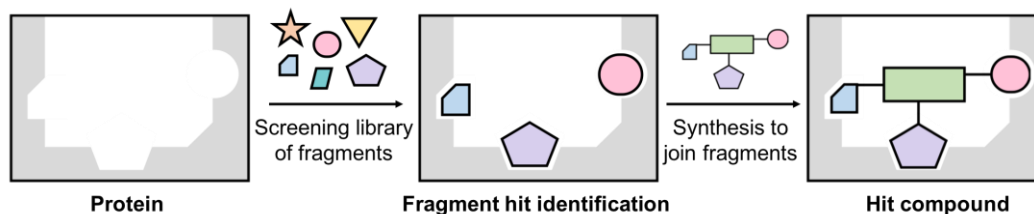
Figure 1.4: A schematic of the FBDD process. *A library consisting of low molecular weight compounds are screened for activity against the protein of interest. Compounds can bind in intricate pockets within the binding site and can then subsequently be expanded or joined together to form high affinity compounds.*

FBDD programs typically have a higher hit rate than HTS programs even though the library of compounds being screened is orders of magnitudes smaller than typical HTS compounds (thousands rather than millions)[26]. The higher hit rate is a result of the fragments sampling a wider range of chemical space than HTS compounds which means that fragments are more likely to bind to intricate pockets within the binding sites and the compounds are also tailored to interact optimally with multiple sites. There are four FDA approved drugs including the BRAF kinase inhibitor Vemurafenib[32] and the BCL-2 inhibitor Venetoclax[33], which are two examples of oncology drugs discovered using FBDD (Figure 1.5).

Vemurafenib
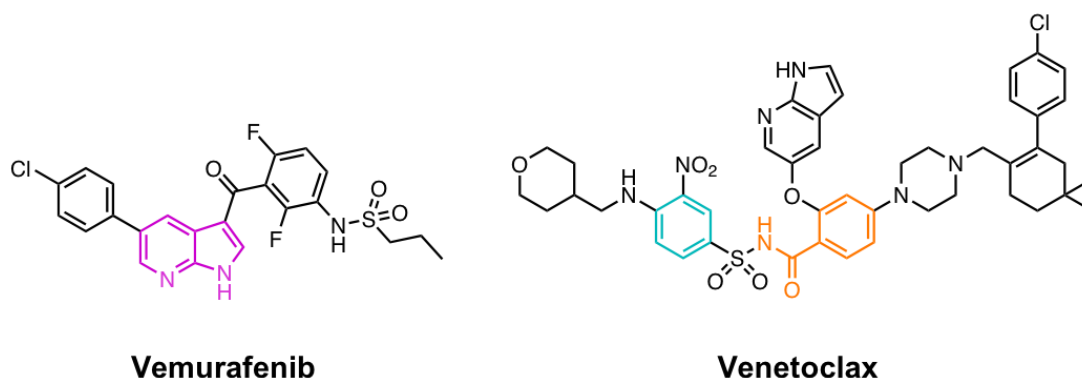
Venetoclax

Figure 1.5: Example compounds from FBDD programs. *Vemurafenib and Venetoclax were designed from fragment screens. The initial fragments are highlighted in colour within the final drug molecules.*

1.1.3 Structure Based drug design

Drug discovery campaigns can be significantly enhanced, through time and cost, when the 3D structural information of the protein target is known, especially if

the protein is bound to an inhibitor allowing the mode of inhibitor binding to be established[34]. Structure-based drug design (SBDD) utilises prior structural knowledge of the target system to design new inhibitors and can be used to complement HTS methods[35]. This can be carried out via the structural development of a hit identified through HTS or FBDD, or via an independent approach where new hits are identified through docking approaches or by designing molecules from scratch (*de novo* design). Molecular docking, such as virtual high throughput screening (vHTS), can be used to virtually screen large libraries of compounds from databases, such as the commercial libraries ZINC15[36], against the desired target and identifies compounds which are predicted to bind with high affinity[37]. Common programs which run vHTS screens are GLIDE[38], GOLD[39] and Autodock[40] which use differing algorithms to position the molecules into the binding site and subsequently score those compounds based upon their predicted interactions. The hit compounds identified through vHTS can represent the starting point for drug discovery programs as they are developed into lead-like molecules. SBDD by vHTS has played a pivotal role in close to 20 drugs in clinical use[41]. Examples include the HIV protease inhibitors nelfinavir[42], amprenavir[43] and lopinavir[44] (Figure 1.6).
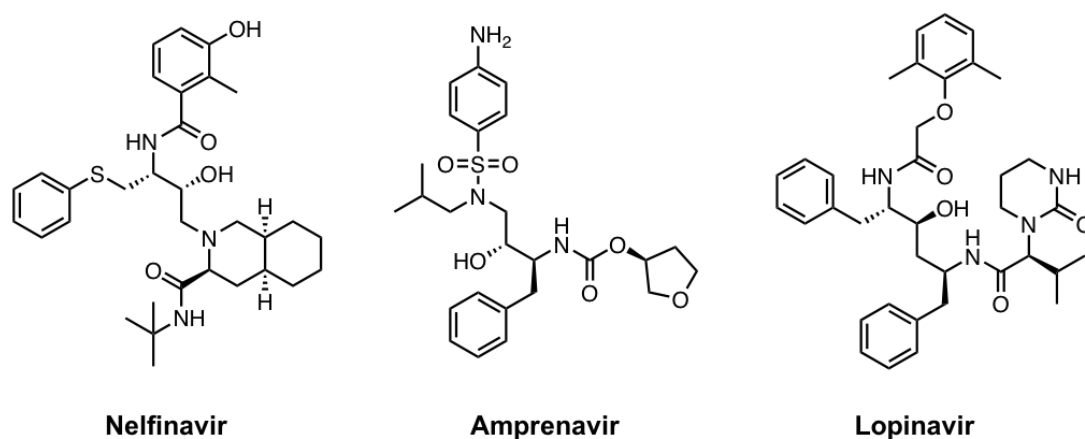
Figure 1.6: Example compounds identified through virtual screening approaches. *Nelfinavir, Amprenavir and Lopinavir were developed to treat HIV after hit compounds were initially identified in vHTS screens.*

De novo design is an alternative approach which allows inhibitors to be designed from scratch, and is similar to the FBDD approach[45]. There are a number of programs which can be used to carry out *de novo* design such as SPROUT[46] and LUDI[47]. The general principle behind the technique is that the inhibitor binding site on

the protein is analysed and regions which could form bonding interactions to the inhibitor molecule, such as H-bond acceptor/donor sites or hydrophobic side chains, are identified[45]. Fragments or molecular building blocks are then positioned at these sites to create favourable interactions between the ligand and receptor. These fragments can subsequently be linked together to form a larger, drug-like molecule which is predicted to bind to the protein. The fragments can be linked together one at a time or by positioning multiple fragments at the favourable sites and linking them together all at once[45]. A schematic of the *de novo* design process is shown in Figure 1.7.
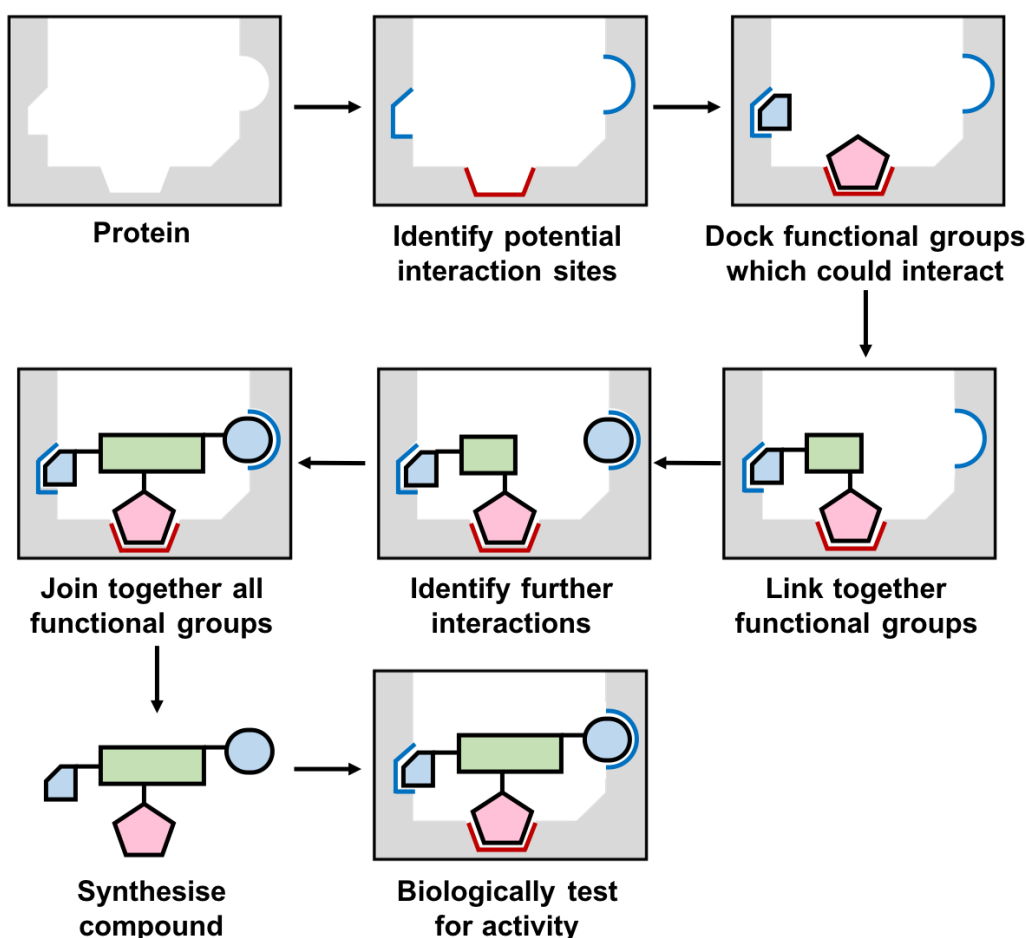
Figure 1.7: An overview of the de novo design process. *Initially, potential sites of interaction such as H-bond acceptor or donor sites are identified and functional groups which could form interactions are positioned at these sites. Computationally the functional groups are linked together in a variety of possibilities. Additional binding sites can then be identified and the process can be repeated until all of the functional groups are linked together. The programs subsequently score the compounds based upon the predicted binding affinity. Compounds are then synthesised and biologically screened to determine their activity towards the protein of interest.*

De novo design can sometimes result in large numbers of compounds being identified. One issue is that some of these compounds could also contain synthetically intractable functional groups, therefore programs have implemented scoring functions based upon the predicted binding affinities and ligand complexity. This allows the user to decide which compounds to take forward for chemical synthesis and subsequent biological evaluation. SPROUT, developed at the University of Leeds, has successfully identified compounds to bind to antibacterial drug targets such as the bacterial RNA polymerase, which has resulted in highly potent lead molecules to be produced[48].

1.2   Obtaining protein for structural studies

SBDD techniques such as vHTS screens or *de novo* design are dependent upon reliable structural models of the biological target being obtained. A pre-requisite for structure determination is the ability to obtain sufficient quantities of pure protein. Sufficient protein quantities can be achieved in two ways; overexpressing the protein in suitable hosts such as *E. coli* and mammalian cells or by extracting the protein from the native source organism. Traditionally, obtaining large amounts of purified protein has relied upon the ability to overexpress proteins in more amenable hosts, such as bacteria or yeast which can be cultured into high volumes[49]. However, a disadvantage of this approach is in the study of large protein complexes which can be challenging due to the need for specific chaperones and/or complex formation pathways which are not always replicated in the host organisms. Therefore, for some systems the preferred route is to purify the proteins from the native source organism.

Obtaining the protein from the native source organism is exemplified for members of the mitochondrial electron transport chain as the protein can be directly isolated from bovine heart tissue. For instance, the crystal structures of bovine complex I[50], cytochrome *bc₁* (complex III)[51] and the F-ATPase (complex V)[52] were solved after the protein was extracted from heart muscle directly. This approach is reliant on a high natural abundance of the protein and an appropriate method of purification that does not rely on engineered tags, such as His or FLAG tags. However, for some species, this method of purification is very challenging. For instance, in parasitic organisms obtaining sufficient quantities of protein is challenging as the parasites are relatively small and are difficult to culture[53,54]. Therefore, obtaining large (litres) volumes of culture can become unfeasible or too costly for most projects. One way to overcome this is to use the CRISPR-CAS9 gene editing technology which allows tags to be

introduced into the genome of the protein of interest thus facilitating the purification of the desired protein[55]. This approach has recently been successful in determining the cryo-EM structure of the *Plasmodium* translocon of exported proteins (PTEX) where FLAG tags were engineered onto HSP101 allowing the purification from parasite culture[56].

SBDD programs rely upon the structure of the target protein being known. If purification of the desired protein cannot be obtained, then homology modelling can be used to generate a model of the protein of interest and the subsequent inhibitor-binding site. This method generally relies upon having an accurate starting model to act as a template for the modelling and also having appropriate levels of sequence similarity which can improve the confidence in the resulting model[57].

1.2.2 Methods to determine structures for SBDD programs

Previously X-ray crystallography and NMR were the principle techniques used to elucidate high resolution structural information and underpin SBDD. However, Electron Microscopy (EM) is now playing a pivotal role as more EM maps are being deposited into the Protein Data Bank (PDB) each year (Figure 1.8). Since 2012, the number of EM derived structures being deposited each year has been increasing dramatically; from 65 in 2012 to 846 in 2018. In comparison, the number of Nuclear Magnetic Resonance (NMR) structures being released has reduced from 537 to 395 in the same time period which highlights how important cryo-EM has become as a tool for structural determination.
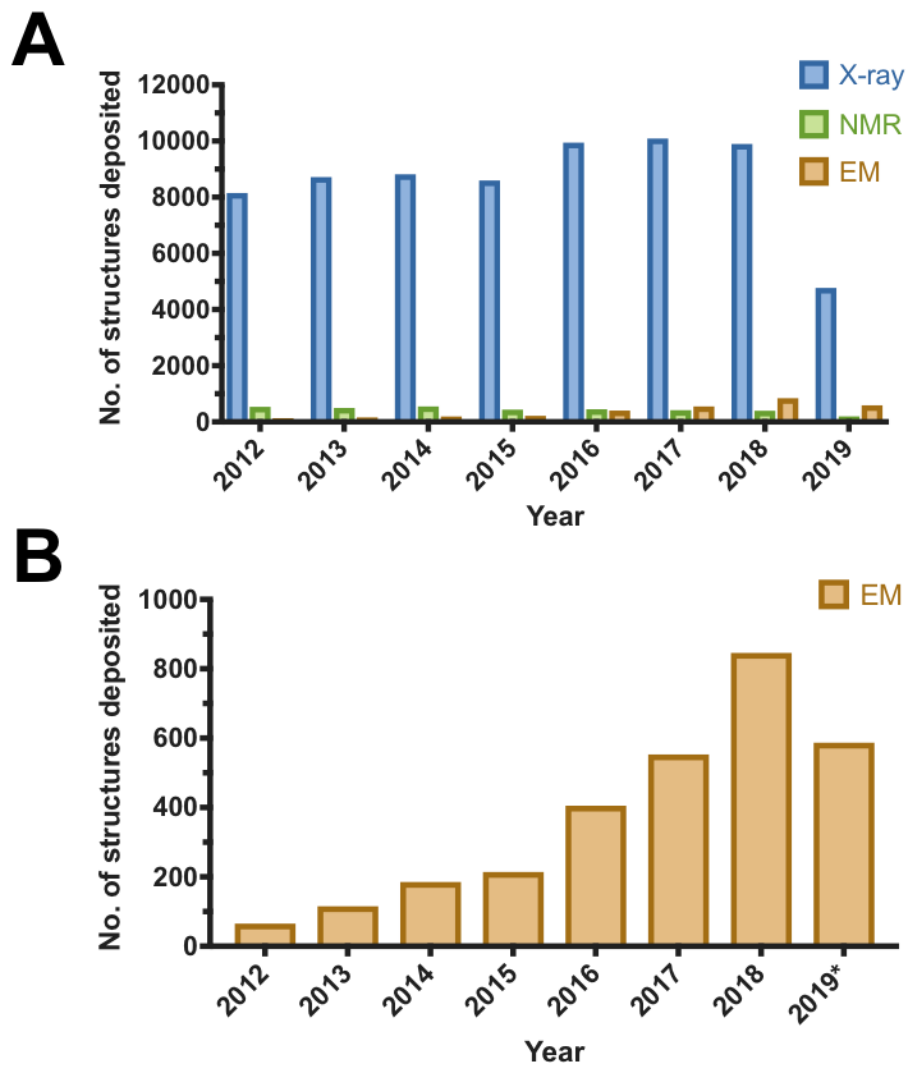
Figure 1.8: The number of structures deposited into the PDB each year by experimental method.
A) The no. of X-ray (blue), NMR (green) and EM (gold) deposited each year. Since 2012, the number of EM depositions have been increasing each year. In 2017 the number of EM maps overtook the number of NMR-derived structures whilst X-ray crystallography has remained relatively constant. B) The no. of EM structures deposited which shows that there has been a dramatic rise in the past few years. Values accurate as of June 2019.

X-ray crystallography is a powerful and predominant tool in structural biology with over ~128,000 protein structures listed in PDB that have been solved using this technique[58]. The proteins are often solved to near-atomic resolution permitting the amino acid side chains, metal ions and water molecules within the sample to be resolved. Furthermore, in many cases an inhibitor is present within the protein structure and the orientation of the inhibitor and how it interacts within the protein can be determined[59].

11

One of the main limitations of X-ray crystallography is the requirement to form ordered crystals which for some proteins, such as large protein complexes or membrane proteins, can be particularly challenging[60]. For instance, membrane proteins exist in a lipid bilayer and are stabilised by these interactions. Once the membrane protein has been solubilised from the native environment, which is commonly achieved using detergents, these favourable interactions are typically lost resulting in the protein losing its structural integrity thus making it challenging to solve the structure. This is shown by there being only ~2,000 membrane protein structures solved using X-ray crystallography deposited in the PDB (<2%)[60]. Moreover, upon the formation of a crystal lattice, the protein adopts a fixed conformation which may be unnatural for the protein in its native solution[61]. Therefore, the resulting structure obtained is a static snapshot of the protein of interest which may not represent the natural conformation of the protein. Additionally, the formation of the crystal lattice could either prevent the ligand from binding or cause it to bind in an artificial position which would subsequently hinder SBDD programs[59].

NMR can not only be used to determine the structure of proteins but it can also play an important role in drug discovery[62,63]. One of the advantages of using NMR is that the protein remains in solution so the structure is determined in its native state and is therefore physiologically relevant. Additionally, NMR is a powerful tool at studying the dynamics of the protein and for looking at intermolecular interactions[64]. The protein does need to be pure, highly stable and is used at high concentrations so large amounts of the protein (often mg's) are needed for structure determination. Moreover, the proteins need to be labelled which can be time consuming and expensive. Another limitation is that NMR cannot be used to study proteins which have a molecular weight greater than ~50 kDa due to spectral crowding and peak broadening making the spectra complicated to interpret[64].

Nevertheless, NMR is a powerful tool in drug discovery as it can detect fragments which bind weakly to the protein (eg K_d values in the millimolar range) or it can be used to complete SAR by characterising the binding affinities for a range of compounds[65]. The chances of obtaining false positives are also very low as NMR spectroscopy observes the ligand binding to the protein directly thus reducing the likelihood of artefacts from enzymatic assays used in HTS screening. Moreover, the ligand binding site is characterised which allows 'hot-spots' to be identified thus revealing novel binding pockets or allosteric sites for drug discovery programs to target[65]. Furthermore, there are different approaches which can be utilised in

screening ligands such as protein-observed or ligand-observed NMR[66]. Ligand observed NMR can be used to screen fragments on larger proteins, with molecular weights between 15-100 kDa typically used, and does not require the protein to be isotopically labelled. In comparison, for protein-observed NMR the binding affinity and binding pose of the ligand can be identified but it is only suitable for smaller proteins (<50 kDa). Therefore, ligand observed NMR is a useful screening tool and hits can subsequently be characterised using protein-based NMR[66].

Cryo-Electron Microscopy (cryo-EM) is starting to play a much bigger role in determining the structure of protein and macromolecular complexes[67]. Protein molecules are suspended in vitreous ice and imaged inside of the electron microscope, therefore removing the need to form protein crystals. Typically, less protein is needed for EM than for X-ray crystallography and NMR, and structures can be obtained for challenging proteins which have previously evaded structural characterisation[68]. However, cryo-EM currently cannot be used routinely as a screening method for the identification of hit fragments. This is because data collections typically last 48-72 hours and processing the resulting data adds additional time, therefore the throughput is much lower than for X-ray crystallography and NMR[68]. The resolutions attained using cryo-EM are now good enough to visualise inhibitor binding so it can play an important role in drug discovery programs. The following sections will discuss the developments in cryo-EM which has enabled the technique to become an important tool in drug discovery.

1.3 Electron Microscopy: A tool in Structural Biology

In 1995, Richard Henderson stated that the laws of physics would allow ~100 kDa biological structures to be resolved to atomic level detail using cryo-EM[69]. Rapid advances in the cryo-EM field have now allowed this goal to become more feasible as there are examples of sub 100 kDa proteins achieving <4 Å resolution[70,71]. These rapid advances have been facilitated by developments in camera technology, microscope stability and to the data processing software available which has given rise to the 'resolution revolution'[72]. The increase in the resolution of the structures being solved by EM has enabled this technique to be thought of as a powerful tool in drug discovery[73]. An overview of developments within the field and the number of maps deposited within the Electron Microscopy Data Bank (EMDB) each year is shown in Figure 1.9.
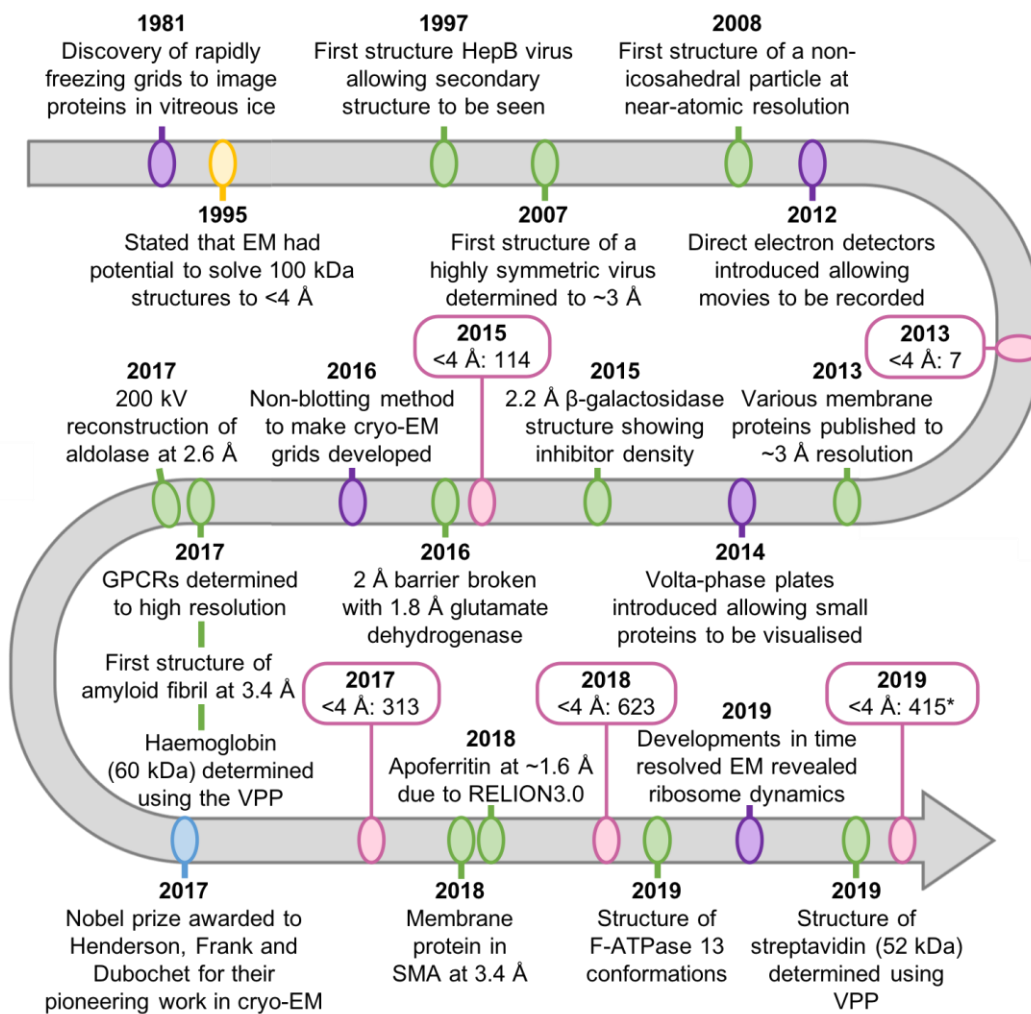
Figure 1.9: A timeline of the key developments within the EM field. *Purple circles indicates developments in technology, green shows notable structures and pink shows the number of maps deposited into the EMDB each year with resolutions less than 4 Å. Yellow indicates when the potential for cryo-EM to be used as a powerful tool for structure determination was first stated and blue is when the Nobel prize was awarded. References are included in the text, adapted from*[74].

For many years EM was unable to resolve protein structures to high resolution therefore the technique was often referred to as 'blobology'. In the 1980's, Dubochet and colleagues discovered that freezing the sample in a thin layer of vitreous ice allowed the protein to be imaged in the high vacuum of the electron microscope[75,76]. However, it wasn't until the 1990's that maps displaying secondary structure could be resolved using EM. At this time, due to the high level of symmetry, the highest resolution structures in the EM field were predominantly large virus particles[77]. In 1997, α-helices were resolved for the first time in the Hepatitis B core virus particle and in 2008 the amino acid backbone could be traced in a number of viruses. By 2010

the density maps showed enough detail for *de novo* atomic model building which led to the publication of a 3.3 Å structure for a non-enveloped virus[78]. For non-symmetric systems, such as membrane proteins, the progress was much slower[79]. The main limitation for these smaller systems was the requirement to collect a much larger data set to overcome the lack of symmetry. In 2012, direct electron detectors were introduced which resulted in an avalanche of structures, of a much broader range of biological molecules, to be determined. Some examples include, an *E. coli* ribosome-EF-TU complex solved to 2.9 Å[80], the human γ-secretase membrane protein resolved to 3.4 Å[81] and the TRPV1 ion channel resolved to 2.9 Å[82].

Further developments within the EM field included the introduction of new instruments to make cryo-EM grids which moved away from the traditional blotting approaches[83]. Furthermore, phase plates have allowed smaller proteins to be imaged to high resolution. Examples include haemoglobin (~60 kDa)[70], streptavidin (~52 kDa)[71] and therapeutically relevant GPCRs[84,85]. Recently, it has also been shown that high resolution structures of <100 kDa samples can be obtained by imaging at 200 kV, without the use of the phase plate, which further expands the scope of samples which cryo-EM can study[86].

Improvements in image-processing methodologies and particle polishing algorithms have not only improved the resolution of the maps but have also allowed the structures of proteins in multiple different conformations to be obtained. This is exemplified for the bacterial F-ATPase which was shown to exist in 13 different unique conformational states[87]. The improvements within the EM field mean that a wide range of different systems can now be studied using cryo-EM. This includes the 'traditional' EM specimens such as viruses[88–91] and ribosomes[92,93], large protein complexes such as the phycobilisome (~16 MDa)[94], and helical proteins such as tau[95,96] and α-synuclein[97,98] which are implicated in Alzheimer's disease. Moreover, the resolution is constantly improving and the highest resolution structure deposited into the EMDB is a 1.6 Å reconstruction of apoferritin (June 2019) where the quality of the map is high enough to visualise water molecules and clearly define side chain rotamers. The resolution of these structures and the wide range of structures which can be studied proves that EM has the capabilities to rival X-ray crystallography as a tool in drug discovery. In 2017, the developments and progress in the EM field led to the award of the Nobel prize in chemistry to Richard Henderson, Joachim Frank and Jacques Dubochet. Examples of some of the structures obtained using cryo-EM are shown in Figure 1.10.
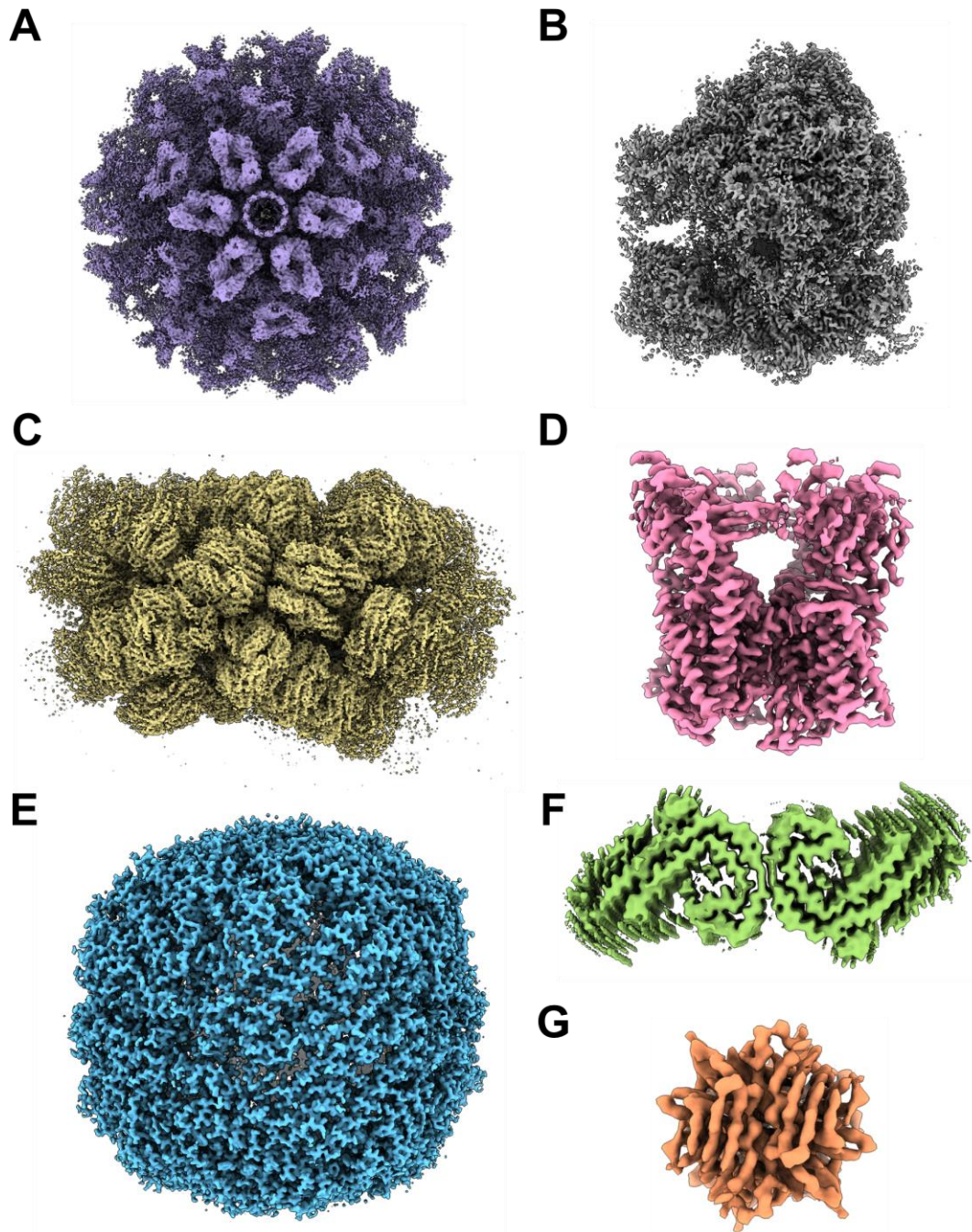
Figure 1.10: Example cryo-EM structures of a wide range of protein systems. *Examples of the 'traditional' systems studied by cryo-EM include A) viruses (EMD0046)[88] and B) ribosomes (EMD2847)[80]. Cryo-EM can also be used to study systems which had previously evaded structural characterisation which is shown for C) large protein complexes such as the phycobilisome (EMD6769)[94] and D) membrane proteins such as TRPV1 (EMD8117)[82]. E) The highest resolution map deposited into the EMDB is a 1.6 Å structure of apoferritin (EMD9599). F) Example of a helical reconstruction of the tau filament (EMD3742)[95]. G) The smallest protein determined by cryo-EM to high resolution is Streptavidin at 52 kDa (EMD0690)[71].*

The renaissance in cryo-EM has particularly benefited the membrane protein field which make up ~60% of all drug targets[99]. Membrane proteins play key roles in cells as they are involved in processes such as synthesising ATP and the transport of ions and molecules across a cell membrane. Determining the structure of membrane proteins can be the bottleneck to many SBDD programs as they are notoriously difficult to crystallise[100]. Due to removing the need for crystallisation, cryo-EM represents a viable alternative to structure determination which is demonstrated by ~175 membrane protein structures being deposited into the EMDB (single particle processed, protein, <12 Å resolution) thus giving structural insights into these important therapeutic targets[79]. These include ion channels[101–104] and transport proteins[105–107] which in some cases had not been structurally characterised before. A further advantage of using an EM approach is that membrane proteins can be imaged in not only a wide range of detergents but also more native systems such as nanodiscs[108] or SMALPs[109]. A cryo-EM structure of TRPV1 in a nanodisc was determined to 2.9 Å resolution and showed a key interaction with native lipids[82]. The first cryo-EM structure of an SMA-solubilised protein was AcrB which was determined to ~9 Å[110]. Higher resolution structures have now been obtained such as the alternative bc_1 complex[111] and AcrB[112] to resolutions of 3.4 Å and 3.2 Å, respectively, thus allowing the interactions to the native lipids to be probed.

Cryo-EM can now play an important role in SBDD pipelines due to the improvements in the resolution of the maps which are routinely being determined[73,113,114]. The first example of inhibitor binding being visualised was in 2015 when a 2.2 Å structure of β-galactosidase was published[115], which included inhibitor density for a PETG inhibitor molecule. Since then, there have been many examples of proteins in complex with small-molecule inhibitors, natural substrates or neutralizing antibodies. Two examples of inhibitor bound structures include a 3.2 Å structure of a plant histidine biosynthesis protein, IGPD[116], and an A_{2A} adenosine GPCR determined to 3.8 Å which was stabilised by the binding of a nanobody[117]. These two structures are shown in Figure 1.11 and emphasise how cryo-EM can visualise inhibitor density in therapeutically relevant proteins and thus can play a role in SBDD programs.
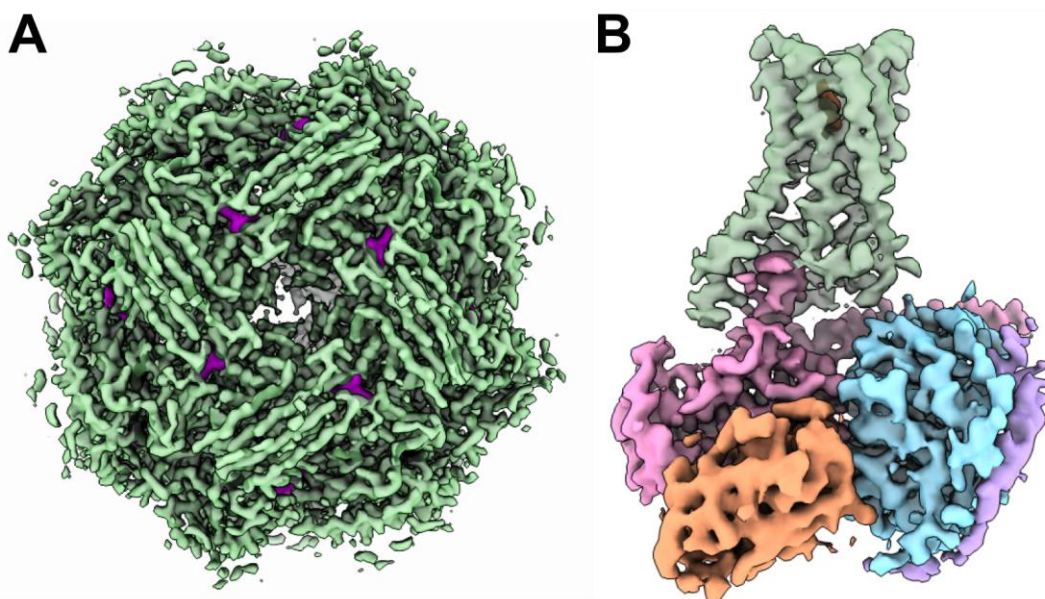
Figure 1.11: Examples of inhibitor-bound cryo-EM maps. *A) The cryo-EM map of IGPD coloured green with inhibitor density coloured in purple (EMD4160)[116]. B) The A$_{2A}$ adenosine GPCR with a small molecule inhibitor (red) bound within the transmembrane region (green). The nanobody which acted to stabilise the structure and allow structure determination is shown in orange (EMD4390)[117].*

Chapter 2 will subsequently discuss the underlying principles and theory behind cryo-EM, from sample preparation, imaging the specimen under the electron microscope and the resulting image processing steps needed to obtain a structure.

1.4   Biological Targets

The SBDD timeline often takes over ten years to complete from start to finish, therefore to experience different aspects of the drug discovery pipeline two systems were studied during this project; cytochrome *bc$_1$* and the Vacuolar ATPase (V-ATPase). Cytochrome *bc$_1$, a* validated anti-malarial target, is at an advanced stage of the drug discovery pipeline and there has been lots of research into the system, resulting in highly potent compounds being developed within the Fishwick group. There are existing crystal structures of *bc$_1$* from mammalian homologues, which show how inhibitors bind to the target. Moreover, an anti-malarial agent, atovaquone, acts upon this enzyme. In comparison, there has been very little inhibitor-design work into V-ATPase which is therefore at an earlier stage of the pipeline than *bc$_1$*. There are existing cryo-EM structures of the protein which show the dynamic nature of the complex and there are also highly potent and selective inhibitors. However, there is

no structural information which details how the inhibitors bind to the protein so this information cannot be utilised in SBDD programs. The biological backgrounds to both cytochrome bc_1 and V-ATPase will be described in Sections 1.5 and 1.6 and will form the main introductions of Chapter 3 and Chapter 4, respectively.

1.5 Cytochrome bc_1

1.5.1 Electron transport chain

The cytochrome bc_1 complex (bc_1) resides in the inner mitochondrial membrane and acts as Complex III in the mitochondrial electron transport chain (ETC)[118]. The ETC consists of four key components; NADH:ubiquinone oxidoreductase (Complex I), succinate dehydrogenase (Complex II), cytochrome bc_1 (Complex III) and cytochrome c oxidase (Complex IV)[119]. These proteins are responsible for generating an electrochemical gradient across the mitochondrial membrane, via electron transfer and proton translocation. This facilitates the synthesis of ATP, via oxidative phosphorylation, by the F-ATPase synthase rotary enzyme (Complex V)[120]. The proton gradient established through the ETC is shown in Figure 1.12.
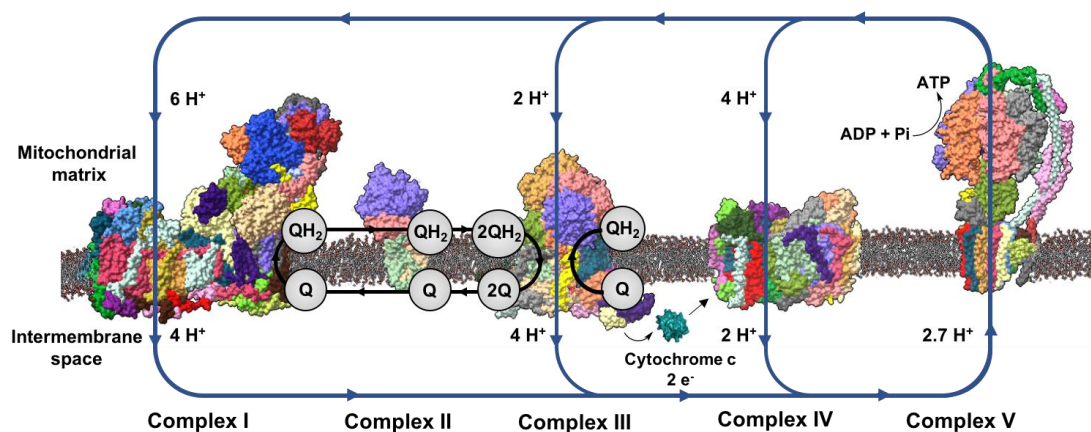
Figure 1.12: The mitochondrial electron transport chain. The five complexes are shown in the membrane and are adapted from[121]. The flow of electrons is controlled by a series of redox reactions between Complexes I-III in the Q cycle. This involves multiple oxidation/reduction reactions of ubiquinone (Q) and ubiquinol (QH$_2$) to generate a surplus of protons in the intermembrane space. This allows rotation of the c-ring in the F-ATPase which drives the synthesis of ATP.

All of the individual components of the ETC have been structurally characterised using both X-ray crystallography and cryo-EM. There are existing crystal structures for each of the components of the ETC with the exception of Complex I. Instead, X-

ray crystallography has been used to determine the structure of the individual domains within Complex I before cryo-EM was used to determine the structure of the full in-tact complex[122,123,124]. As the structures of the individual domains were known, they could be fitted into the cryo-EM map which exemplifies how the structural techniques are complementary and the insights gained can be integrated in order to determine the full structure of the complex. The structural details gleaned has enabled the mechanism of electron transfer from NADH to ubiquinone, via a chain of seven iron-sulphur clusters, to be established thus demonstrating how cryo-EM can be used to study large membrane protein complexes[125]. It has also been shown that the individual components of the ETC can arrange to form higher-ordered super complexes with defined stoichiometry[126]. The supercomplexes are important as they prevent the formation of reactive oxygen species (ROSs) and increase the stability of the ETC complexes.

The recent renaissance in cryo-EM has led to the structure of different supercomplexes being obtained from a variety of different species including porcine[127], ovine[128] and human[129]. This has provided structural details of how cytochrome bc_1 carries out its function. For instance, Guo et al identified two complexes from human mitochondria; the super-complex and the mega-complex[129]. Within the super-complex, the protein ratio was $CI_1CIII_2CIV_1$ and it was resolved to 3.9 Å[129]. By isolating different components of the complex through masking Complex I and Complex III, the resolution of the individual components were improved to 3.7 Å and 3.4 Å respectively, thus permitting side chains to be accurately modelled into the map. This level of structural detail is invaluable as it can reveal differences in amino acid side chain residues between the host and target species allowing SBDD programs to utilise this information to design compounds which are highly selective to the target species. The low resolution (17 Å) map of an additional mega-complex was obtained with protein architecture $CI_2CIII_2CIV_2$[129]. Despite the secondary structure of the complex not being known, the overall globular shape of the density allowed the individual components of the chain to be fitted into the map. The dimeric form of cytochrome bc_1 was positioned in the centre of the mega-complex and interacted with two copies of Complex I and Complex IV with all of the enzymes sharing the same ubiquinone pool. Additional supercomplexes containing dimeric bc_1 and Complex IV have also been identified[130].

During the course of the PhD project, cryo-EM structures of alternative Complex III (ACIII), which reside in the photosynthetic electron transport chains of bacteria, were determined. The protein is homologous to mitochondrial bc_1 yet shows little similarity in structures despite catalysing the oxidation of ubiquinol. They also pass electrons directly to Complex IV without the need for an electron carrier such as cytochrome c in mammals. Examples include ACIII isolated from *Flavobacterium johnsoniae*, which was purified in the SMA co-polymer to 3.4 Å resolution[111] and a 3.9 Å structure of ACIII from *Rhodothermus marinus*[131]. In both structures, ACIII exists in a supercomplex with an alternative complex IV, highlighting the importance of the supercomplexes in facilitating electron transfer as they can utilise the same ubiquinol/ubiquinone pool. Example structures of the supercomplexes are shown in Figure 1.13.
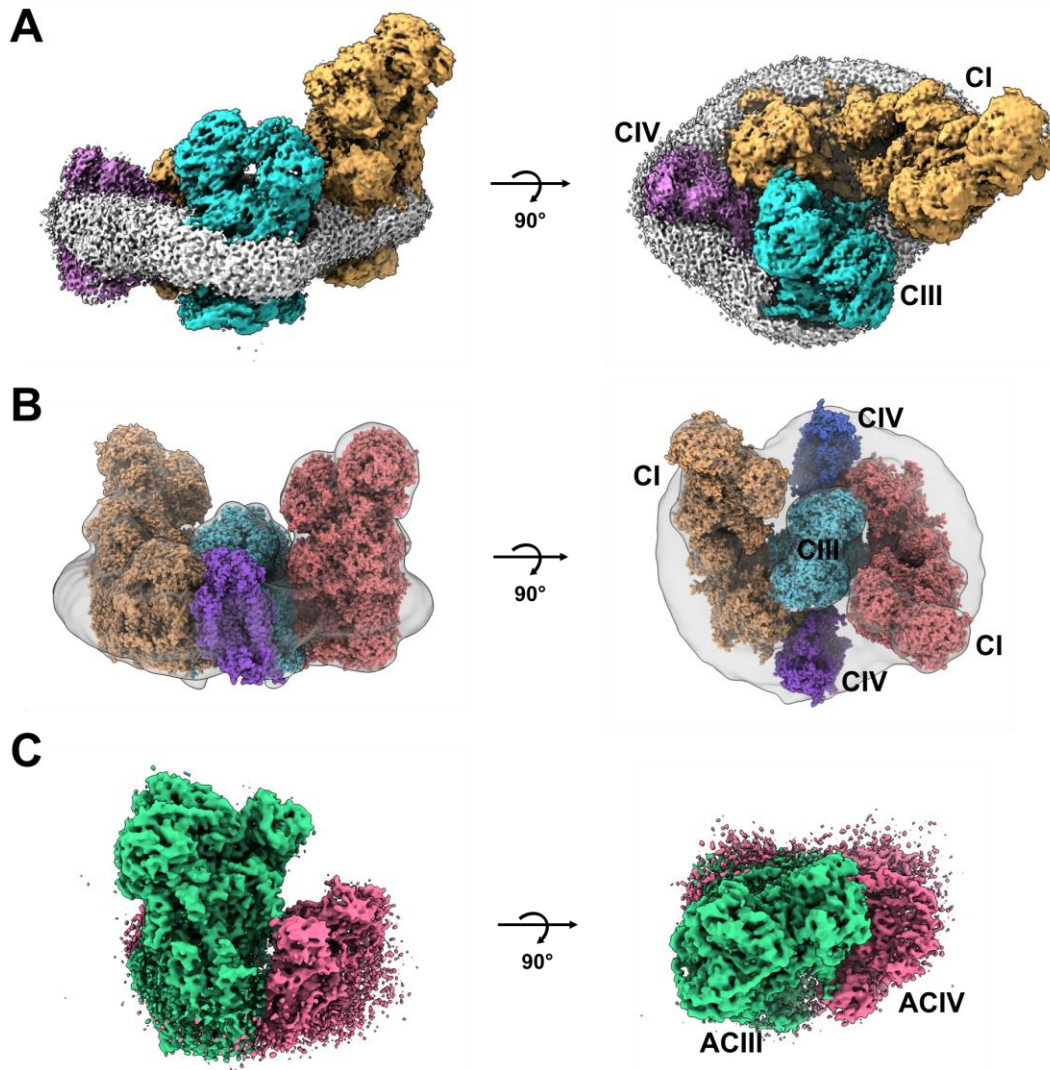
Figure 1.13: Cryo-EM structures of ETC supercomplexes. A) The human super respirasome complex ($CI_1CIII_2CIV_1$) with a global resolution of 3.9 Å (EMD-6775)[129]. Cytochrome bc_1 is coloured in cyan and interacts with both Complex I (gold) and Complex IV (pink). B) A low-resolution map of the human mega-complex at 17 Å resolution. Bc_1 (cyan) is at the centre of the circular complex and interacts with two copies of Complex I (orange and red) and Complex IV (dark blue and purple) (EMD-6776)[129]. C) The structure of the alternative Complex III (teal) interacting with an alternative Complex IV (pink) from Flavobacterium johnsoniae. The supercomplexes highlight how central bc_1 is for facilitating electron transfer (EMD-7286)[111].

1.5.2 Structure of cytochrome bc_1

Cytochrome bc_1 has been extensively studied by X-ray crystallography. The first crystal structure of cytochrome bc_1 from bovine mitochondria was solved in 1997 to 3.0 Å resolution[132]. Since then there have been 265 X-ray crystallography entries into the PDB for the full complex and individual subunits from a wide variety of different

species. Examples include a 2.7 Å structure of chicken bc_1[133], yeast bc_1 solved to 1.9 Å[134], bovine bc_1 to 2.1 Å[135] and bc_1 from rhodobacteria at 2.4 Å[136]. The structures have provided a detailed understanding of the function and mechanism of the complex. The overall architecture of the protein including the different subunits is shown in Figure 1.14.
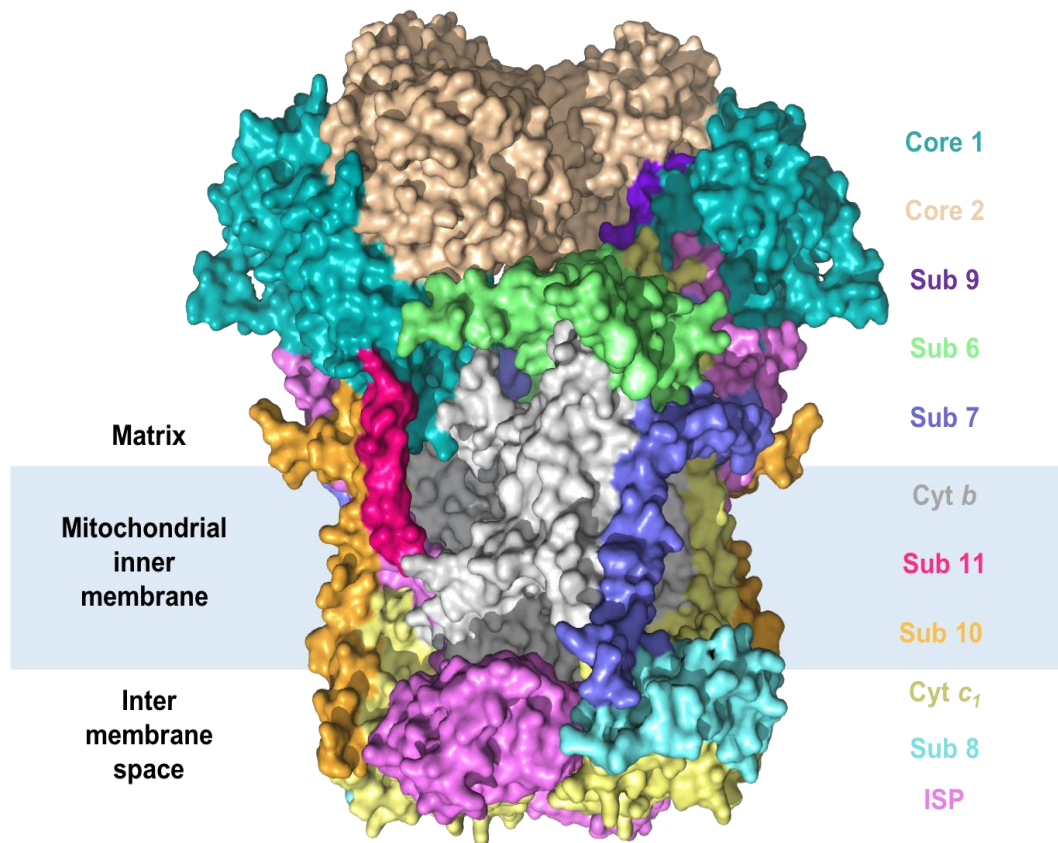
Figure 1.14: Crystal structure of dimeric cytochrome bc₁. Example of an existing crystal structure coloured according to subunit. The complex resides in the inner mitochondrial membrane and the large cytochrome b subunit is embedded in the mitochondrial membrane. The catalytic Rieske domain (ISP) and cytochrome c₁ are in the intermembrane space.

Cytochrome bc_1 subunit composition can vary between different species, with mammalian species having 11 subunits[132], yeast 10 subunits[137] and bacteria 3 subunits[138]. Between all species there are three conserved catalytic subunits; cytochrome b, cytochrome c_1 and the Reiske iron-sulphur protein[139]. In mammals, bc_1 exists as an 11-subunit heterodimer with a molecular weight of ~480 kDa[140]. The complex couples electron transfer from ubiquinol to Cytochrome C with the generation of a proton gradient across the mitochondrial membrane[11]. This occurs

through the Q cycle which involves a series of redox reactions to facilitate electron transfer. The Q cycle occurs at two distinct catalytic sites in cytochrome b, the Q_i and Q_o sites, with both sites containing a catalytic heme group. Ubiquinol (QH_2) is oxidised to ubiquinone (Q) in the oxidative (Q_o) site and releases two protons into the intermembrane space. Meanwhile, ubiquinone binds to the reductive (Q_i) site where it is reduced to ubiquinol and takes two protons from the matrix[142,143]. The overall net reaction for the Q cycle is:

QH_2 + 2 ferricytochrome C $\leftrightarrows$ Q + 2 ferricytochrome C + 2H$^+$
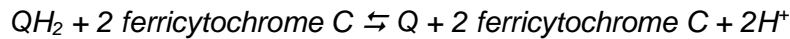
An overview of the Q cycle mechanism is shown in Figure 1.15.
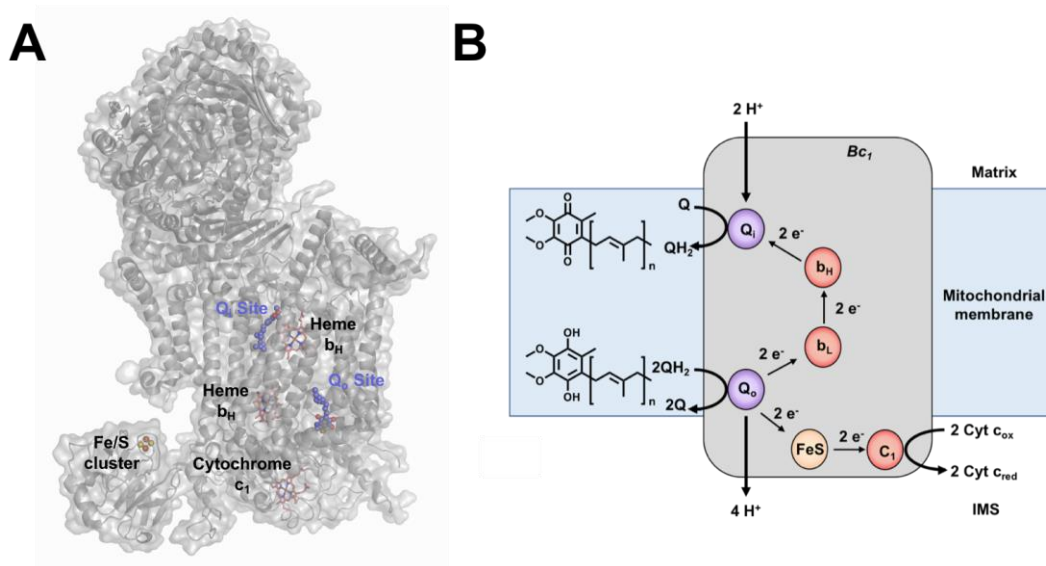
Figure 1.15: The structure and mechanism of cytochrome bc₁. *Adapted from[70.] A) A crystal structure of a monomer subunit of cytochrome bc₁. The positions of the catalytic heme groups are shown as pink sticks and the Q_i and Q_o inhibitor binding sites are shown with purple spheres. B) The flow of electrons through bc₁ with the structure of the natural substrates shown at their binding sites.*

1.5.3 Cytochrome *bc₁* as a therapeutic target

As the ETC is involved in respiration and providing energy to cells, disrupting the chain in pathogens can be fatal to the organism, therefore the individual components represent attractive therapeutic targets. For instance, inhibitors against cytochrome *bc₁* have been developed to act as antimicrobial agents[144]. Examples of highly specific inhibitors of cytochrome *bc₁* are antimycin and stigmatellin which bind to the

cytochrome b Q_i and Q_o sites, respectively[135]. As these compounds are highly toxic they cannot be used as medicines. There are existing crystal structures showing how the drugs bind to the protein highlighting that the compounds occupy the ubiquinone or ubiquinol natural substrate binding site[135]. An example of how antimycin and stigmatellin bind to cytochrome bc_1 is shown in Figure 1.16.
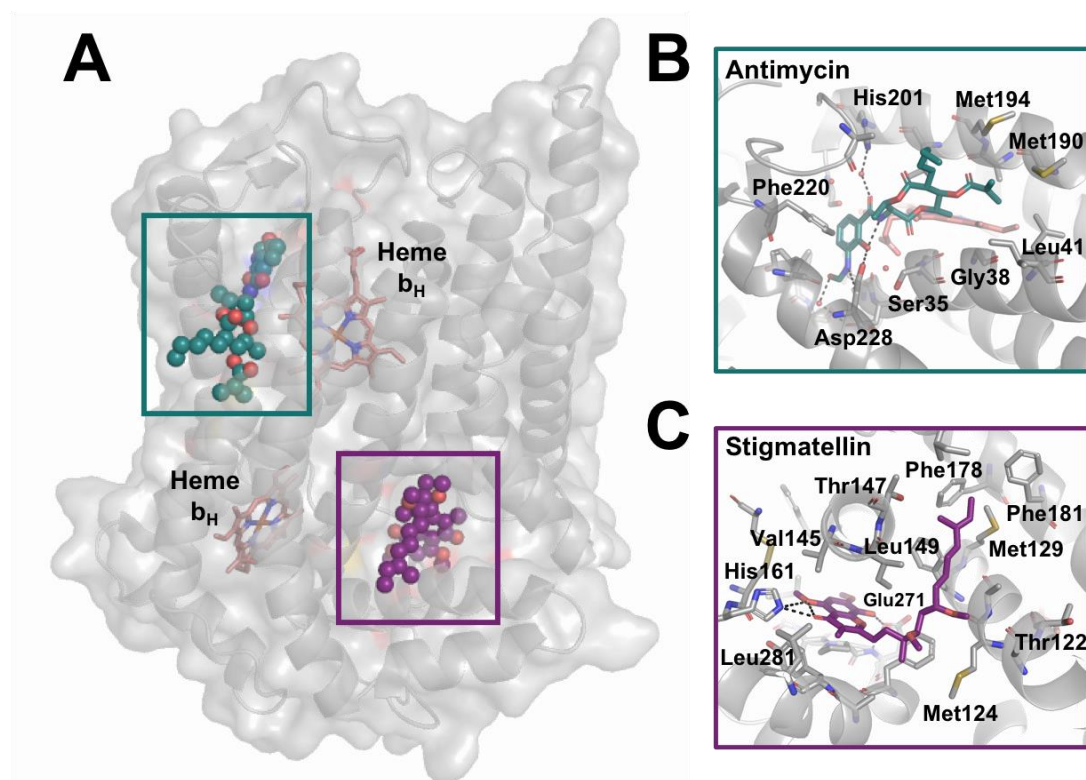
Figure 1.16: Known inhibitors of cytochrome bc₁. *A) The cytochrome b portion of a crystal structure (pdb 1ppj)[135] showing how Antimycin (teal) and Stigmatellin (purple) bind at the Q_i and Q_o sites, respectively. B) Antimycin (teal) bound to the protein where the compound forms H-bonds to His201 and Asp228. C) Stigmatellin (purple) bound at the Q_o site where it can make two H-bonds to His161 and Glu271.*

The structure of the catalytic binding site is highly conserved between mammalian species. However, there are differences in pathogens which can be exploited[144]. This is exemplified by a number of different fungicides which target the Q_o site in pathogenic bc_1 such as azoxystrobin, famoxadone and fenamidone[145]. One problem is that resistance to treatments is now emerging which could be due to cytochrome b being encoded by the mitochondrial genome which means that mutations in the protein can arise quickly[144].

Cytochrome bc_1 is an established drug target against *Apicomplexan* parasites, such as *P. falciparum* (malaria)[146,147] and *T. gondii* (toxoplasmosis)[148,149]. In *Apicomplexan* parasites, oxidised ubiquinone from cytochrome bc_1 is used by dihydroorotate dehydrogenase (DHODH) to generate orotate, which is an essential intermediate for pyrimidine biosynthesis[150]. Therefore, the inhibition of cytochrome bc_1 collapses pyrimidine production causing parasite death. The existing crystal structures from mammalian systems have been solved in the presence of both Q_i and Q_o site inhibitors[132,151]. One of these structures shows the anti-malarial drug, atovaquone, bound at the Q_o site[152]. However, resistance is now emerging to this treatment and new drug molecules are urgently required for the treatment of malaria[153]. Using this crystal structure, a series of potent compounds were designed by GlaxoSmithKline (GSK) to target the Q_o site of bc_1. One compound was highly active against atovaquone-resistant *P. falciparum* but failed in a first time in human (FTIH) trial because of acute cardiotoxicity[154]. A crystal structure of this compound (GSK932121) bound to bovine bc_1 revealed that it actually occupied the Q_i site rather than the Q_o site[151] and this has led to different series of compounds being developed against this site for both malaria and toxoplasmosis.

The sequence identity of cytochrome bc_1 is highly conserved across bovine, human and *P. falciparum.* However, the N-terminus of parasitic cytochrome *b* is shorter than in mammalian species leading to differences in the Q_i binding site (Figure 1.17)[151]. There is no structural information available for parasite-derived bc_1, which has hindered the design of new therapeutics targeting these species. Obtaining sufficient quantities of parasite protein for crystallography has proven to be a significant challenge because the protein cannot be over-expressed in traditional expression systems such as *E. coli*[53]. Therefore, it has to be obtained from the native parasite organism which will provide significantly less protein than the mammalian species. As cryo-EM uses less protein than X-ray crystallography, this technique could present an alternative method in elucidating how the inhibitor interacts with the target organism.
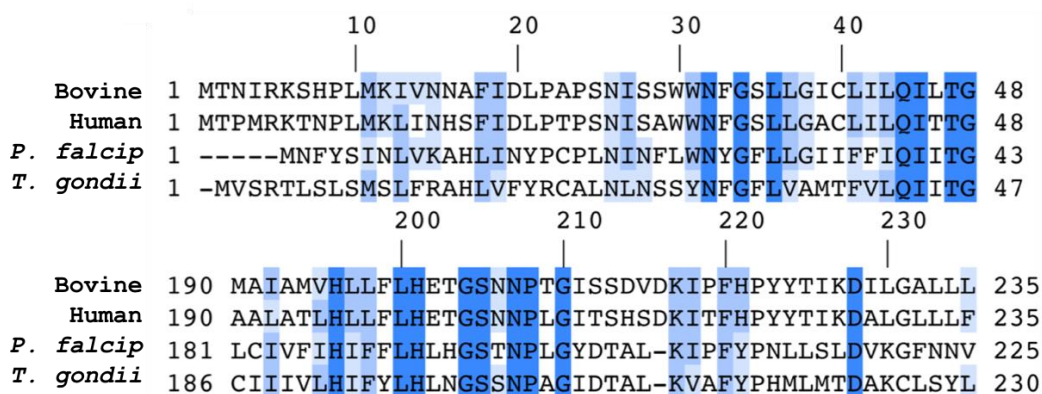
Figure 1.17: Sequence alignment at the Q$_i$ inhibitor binding site. *Residues fully conserved between human and parasite are illustrated in deep blue, partially conserved in light blue, and non-conserved in white.*

Chapter 3 will focus on determining whether cryo-EM can be used to study inhibitor binding to *bc$_1$* which is a therapeutically relevant target. To this end, bovine *bc$_1$* will be studied in a proof-of-principle approach to see whether the resolution attained by cryo-EM is high enough to visualise small molecules bound to the protein. After this has been achieved, work can then begin to obtain the parasite protein to allow inhibitor binding to the target species to be visualised. The novel structural insights gained can then be utilised in the design of new inhibitor molecules to make them more selective to the target parasite organism, thus reducing the toxic off-target effects.

1.6 Vacuolar ATPase

1.6.1 V-ATPase as a drug target

The Vacuolar ATPase (V-ATPase) is a 1 MDa protein complex formed from 14 subunits, which uses ATP hydrolysis to drive proton transport across the cell membrane, maintaining the pH of intracellular and extracellular compartments[155]. It is a member of the rotary family of enzymes which also include the F-ATPase and A-ATPase enzymes. V-ATPase is found in all eukaryotic cells and its major function is to maintain and regulate the pH within cells including within the cytoplasm and different organelles such as endosomes, lysosomes, vesicles and central vacuoles of fungi and plants[156]. It therefore plays a key role in cellular processes such as receptor mediated endocytosis, protein processing and degradation and the transport of small

molecules and ions. Moreover, V-ATPase which are found in plasma membranes, are important in facilitating bone resorption by osteoclasts and acid secretion in the kidneys therefore the complex has been implicated in a vast number of diseases such as osteopetrosis, renal tubular acidosis, male infertility, diabetes and cancer[157].

V-ATPase also exists in different isoforms within the body. For instance, in human V-ATPase, there are four subunit *a* isoforms which have 40-60% amino acid similarity[158]. The different isoforms are also localized to different organelles and can therefore be implicated in different diseases. This is exemplified by isoform *a*3 being found in osteoclasts and if mutated is implicated in osteoporosis. In comparison, isoform *a*4 is found exclusively in renal cells and is linked to renal diseases[158]. Therefore, V-ATPase is a complex drug target because it not only exists in different isoforms within the body but it also carries out a wide range of functions. Without having any structural information of the different isoforms, it is challenging to design a selective inhibitor to target one particular isoform and complete inhibition of the enzyme could be fatal.

1.6.2 Current V-ATPase structural information

The structure of the V-ATPase complex has been studied using both X-ray crystallography and cryo-EM. It is composed of two domains; the soluble V_1 domain and the membrane embedded V_0 domain, as shown in Figure 1.18. The V_1 domain, is responsible for the hydrolysis of ATP and consists of eight different subunits (A-H). The AB subunits exist as a circular hexameric ring, with three catalytic sites at the AB interface[156]. The V_1 and V_0 domains are joined together through central (subunits D and F) and peripheral (subunits E and G) stalks[159]. The central D subunit acts as a rotor and uses energy generated from ATP hydrolysis to drive rotation of the V_0 *c*-ring. Subunits E and G prevent rotation of the AB hexameric ring as ATP is hydrolyzed, therefore they act as stators. The subunits which make up the V_0 domain are: *a*, *c*, *d* and *e*. The *c*-subunits form a membrane embedded ring which rotates to transport protons across the membrane. Proton translocation occurs at the *c*-ring-subunit *a* interface and the number of *c*-subunits within the ring can vary between species[156].
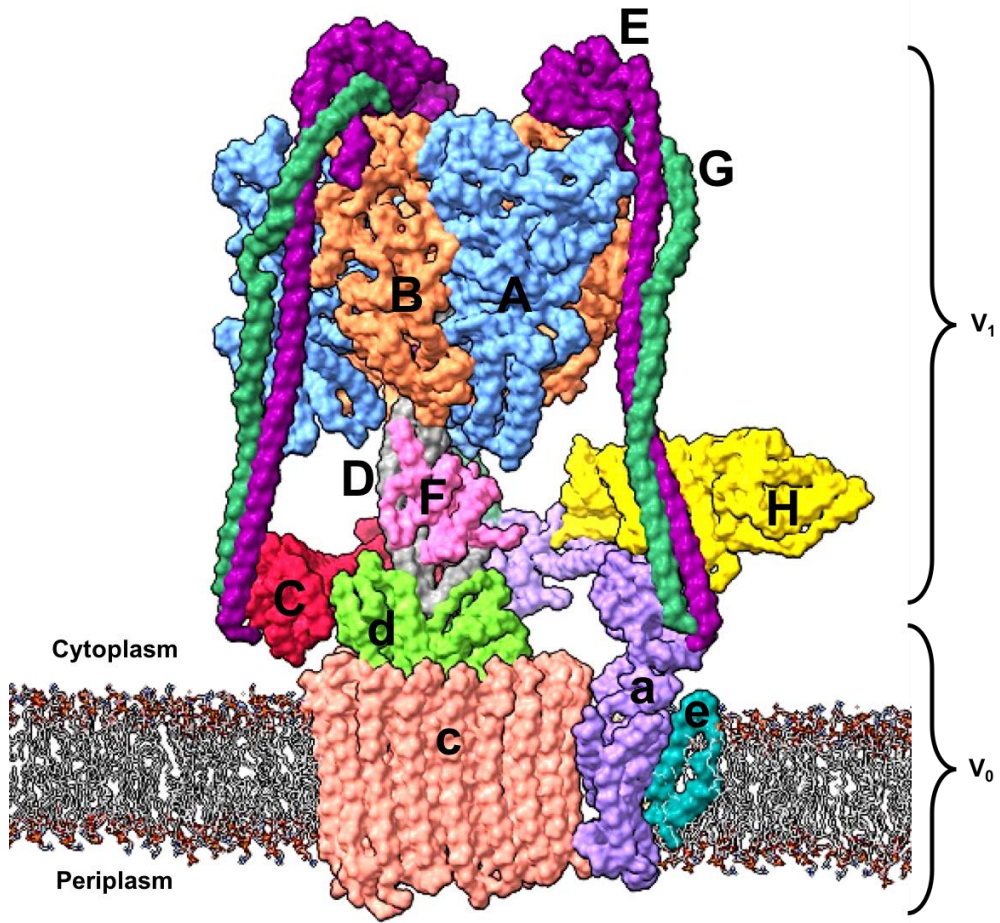
Figure 1.18: **The structure of the V-ATPase complex**. ATP hydrolysis occurs at the interface of the A and B subunits which resides as a hexameric ring (AB)₃. The membrane embedded c-ring is responsible for pumping protons across the membrane which occurs at the *c*-ring subunit *a* interface.

There are high-resolution crystal structures of the individual subunits of the complex and moderate resolution cryo-EM structures of the full complex showing how these individual subunits assemble to form the complex. One such example is from Zhao *et al* (2015) who determined the structure of yeast V-ATPase using cryo-EM[160]. Three different conformational states of the complex were identified which achieved resolutions of 6.9 Å, 7.6 Å and 8.3 Å (Figure 1.19).

State 1
47% of images

State 2
36% of images
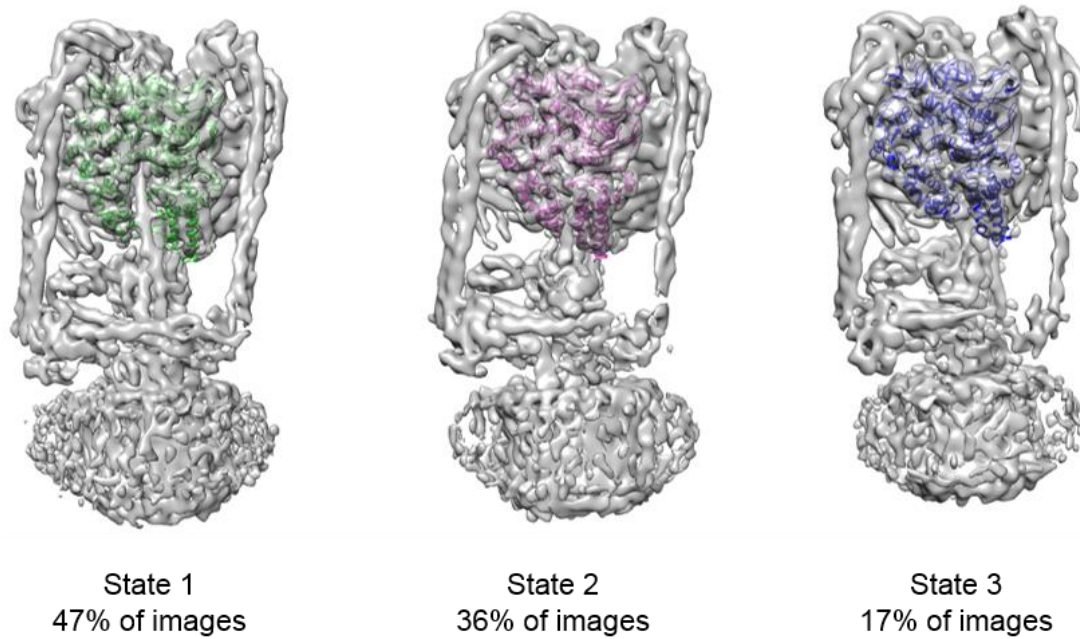
State 3
17% of images

Figure 1.19: Three catalytic states of the V-ATPase identified by cryo-EM. *The AB subunits in each state are shown in colour and represent how this domain changes conformation between all three states. In state one the AB domain above the C-subunit is in the open state, in state two it is in the tight state and finally in state three it is in the loose conformation.*

The number of particles which went into each state was calculated, showing state one had the highest percentage of particles and could therefore represent the most physiologically prevalent conformation which inhibitors should be designed against[160]. Designing inhibitors against particular states could be a strength of an EM approach. Because the protein molecules are trapped within a thin layer of ice and not locked into any particular conformation, the natural abundance of the different conformations can be determined. Furthermore, the presence of an inhibitor could reduce the dynamic nature of the complex, subsequently locking it into one conformational state which could enable the resolution of the full complex to be improved. If resolutions are good enough to visualise inhibitor binding then these will prove extremely beneficial to SBDD programs.

Within each state of the V-ATPase, the AB domain resides in three different catalytic conformations, open, loose and tight, which all contain one nucleotide binding site[160]. It is proposed that ATP binding results in conformational changes between the A_3B_3 hexamer creating 120° anti-clockwise rotations as the ATP binds,

30

is hydrolysed and subsequently released[161]. A schematic describing the process is shown in Figure 1.20. Briefly, ATP becomes tightly bound within the AB interface where it is subsequently hydrolysed. The ADP nucleotide product is then more loosely bound before it leaves which enables another ATP molecule to bind[161].
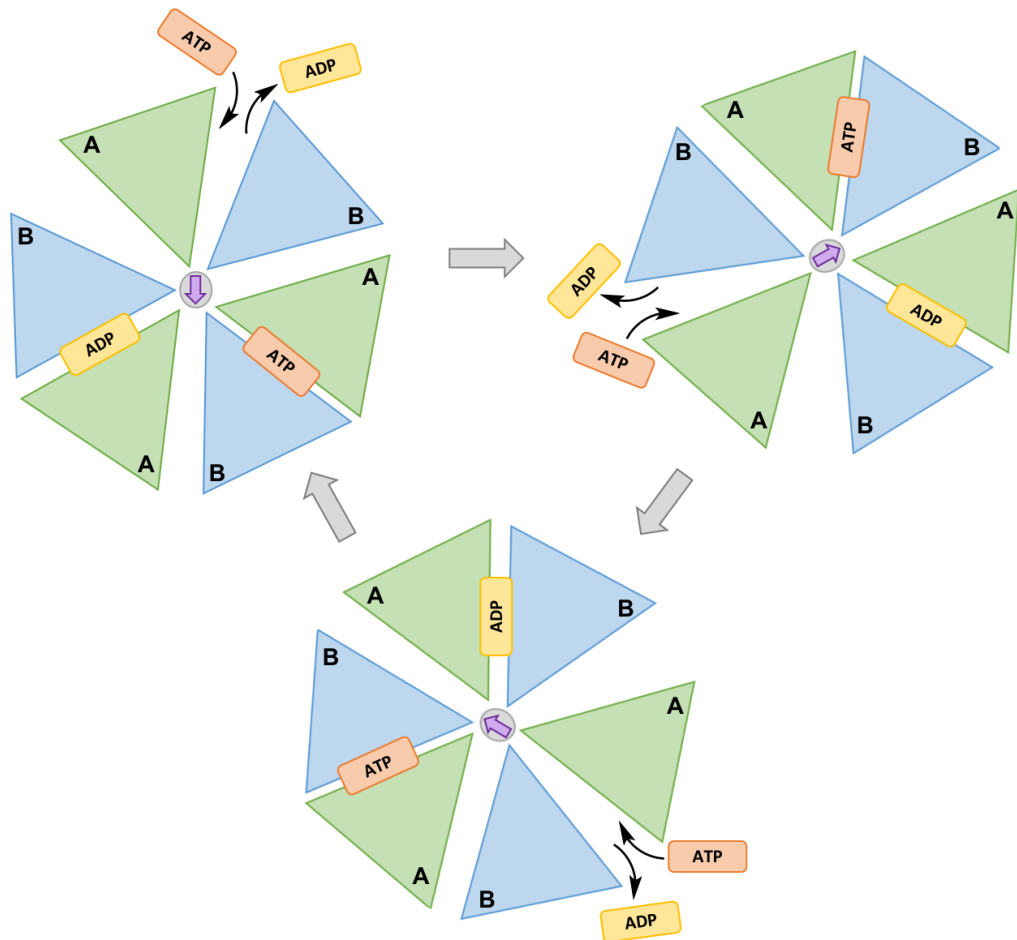
Figure 1.20: The mechanism of ATP hydrolysis. *The alternating AB subunits are coloured in green and blue, respectively. ATP binds into an empty AB interface where it becomes tightly bound allowing hydrolysis to occur. This rotates the hexameric ring by 120°. ADP is subsequently bound less tightly before leaving and allowing another ATP nucleotide to bind. Adapted from[161].*

The energy produced from the hydrolysis of ATP is then transferred along the central DF rotor and through to the V_0 subunit *d*. This subsequently drives rotation of the *c*-ring past the membrane-embedded subunit *a* which facilitates proton transfer across the membrane[161]. At the *c*-ring subunit *a* interface there are two half channels (cytoplasmic and luminal) which allow protons to be transferred across the

membrane. The protons enter from the cytoplasmic half channel, where they pass along charged residues on subunit *a* before reaching the essential glutamate residue. The deprotonated glutamate is stabilised via the formation of a salt bridge with a highly conserved Arginine residue from subunit *a*. After the glutamate has picked up the proton, rotation occurs which allows the proton to enter the hydrophobic lipid bilayer where it is transported round a full 360°. The proton is then transported across the membrane through the luminal half channel where it can cross the membrane and enter the cell plasma or organelle depending on where the V-ATPase is located[161]. An overview of this process is shown in Figure 1.21. This mechanism is supported by recent high-resolution structures of the V_0 domain which were obtained using cryo-EM[162–164].
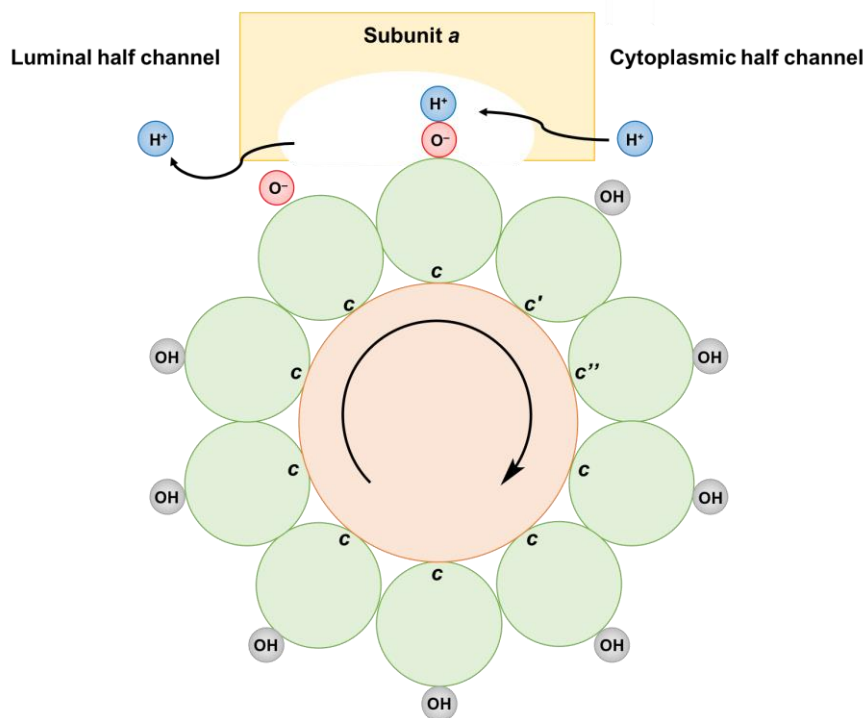
Figure 1.21: An overview of the proton pumping mechanism. *The c-subunits are shown in green and subunit-a coloured yellow. Protons (blue circles) enter via the cytoplasmic half channel and protonate a key glutamate residue. When protonated the c-ring rotates and completes a full 360° turn where the proton is deposited into the luminal half channel. Adapted from*[161].

1.6.3 Current V-ATPase inhibitors

The first inhibitors of V-ATPase were the plecomacrolides bafilomycin and concanamycin, which were discovered in the 1990's[165]. The compounds have binding affinities in the low nanomolar range and are selective V-ATPase inhibitors. Although the exact binding poses of how the ligands interact with the protein have not been structurally characterised, mutagenesis studies have shown that the compounds bind to the V_0 *c*-ring thereby disrupting proton translocation across the membrane[166]. Another V-ATPase specific, low-nanomolar inhibitor is the macrolactone archazolid A which was first isolated in 2003 from myxobacteria *Archangium gephyra* and *Cystobacter violaceus*[167]. Further mutagenesis studies have revealed that the inhibitor again interacts with the V_0 *c*-ring but at a different binding site to the plecomacrolides as only one of the residues involved in binding bafilomycin affected the sensitivity to archazolid[168]. The archazolid binding site involves the essential glutamate residue which is involved in transporting protons across the membrane. There are another family of V-ATPase inhibitors, the benzolactone enamides, which encompass salicyilhalamides, lobatamides, oximidines and apicularen[169]. The compounds all inhibit V-ATPase with nanomolar levels of activity and are predicted to bind to the V_0 *c*-ring at a different site to the plecomacrolides. Interestingly, the benzolactone enamides are highly selective for mammalian V-ATPases and do not inhibit fungi V-ATPase therefore they are thought to be attractive therapeutic agents[165]. A further class of inhibitors were developed after studying the SAR of bafilomycin. The most potent of the compounds was Indol0 which binds at the same site as bafilomycin against subunit *c*[170]. A final inhibitor of V-ATPase is dicyclohexylcarbodiimide (DCCD) which forms a covalent bond to the essential glutamine residue in the *c*-ring[171]. A problem with DCCD is that it is unselective as it also inhibits the F-ATPase. The structures of the inhibitors and the corresponding IC_{50} values are shown in Figure 1.22.
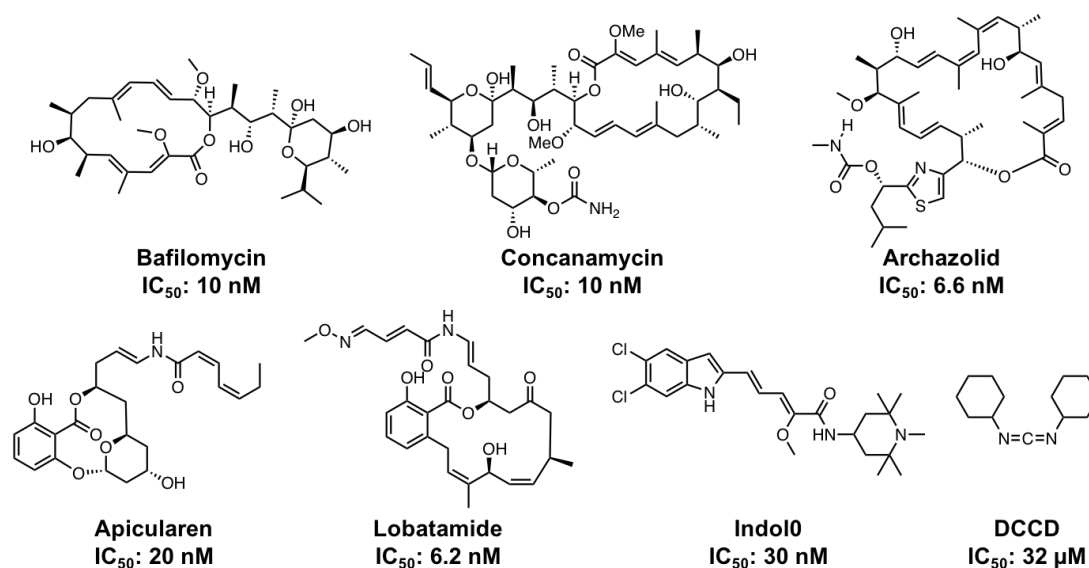
Figure 1.22: The structures of the current V-ATPase inhibitors and the corresponding IC₅₀ values. *All of the current inhibitors have binding affinities in the low nanomolar range and all of the compounds except for DCCD and Indol0 contain large macrocycle rings which are difficult to synthesise and expensive to buy. References for IC_{50} values: bafilomycin[172], concanamycin[173], archazolid[168], apicularen[174], lobatamide[169], indol0[170] and DCCD[175].*

With the exception of DCCD and Indol0 all of the current V-ATPase inhibitors are large macrocycles which are difficult to make and expensive to buy. Furthermore, the compounds are all unstable and degrade over time making them difficult to use when studying the system. Therefore, there is a need to develop new inhibitors which could be used as chemical probes when investigating the system or could be used in the treatment of diseases such as cancer and osteoporosis. One challenge in the design of new V-ATPase inhibitors is that despite the existence of highly selective V-ATPase inhibitors, the exact binding pose is not known which is hindering SBDD programs.

The work in Chapter 4 will focus on the identification of hit compounds to act upon V-ATPase and Chapter 5 will describe the design and synthesis of novel V-ATPase inhibitors with the goal of improving the chemical properties of the compounds, compared to the current macrocyclic-based inhibitors. The inhibitors have been designed to target yeast V-ATPase and have not been designed to act upon a specific isoform or disease therefore the inhibitors could act as chemical tools rather than a therapeutic agent to treat a particular disease. Moreover, Chapter 6 will

discuss the purification of V-ATPase and attempts to determine the cryo-EM structure of the complex in the presence of an inhibitor. The aim is then to determine whether the presence of the inhibitor has disrupted the dynamic nature of the complex by locking the protein into one conformation. This could improve the stability of the complex and thereby improve the resolution which will aid future SBDD programs.

2 Materials and methods

The Materials and Methods have been broken down into three distinct sections. The first is Electron Microscopy which provides background into the theory and principles of EM (Section 2.1). The methods used for Chapter 3, detailing the cytochrome bc_1 work, have been described in Section 2.2 and the V-ATPase work in Section 2.3; underpinning the work in Chapters 4-6. All of the chemistry experimental, describing how compounds were synthesised, is provided in Chapter 8.

2.1. Electron Microscopy

The pipeline for structure determination using cryo-EM can be broken down into three key stages; sample preparation, data collection and image processing. A general schematic of the steps involved is shown in Figure 2.1. The tasks which are performed at each of the three stages will be discussed in Sections 2.1.1, 2.1.2 and 2.1.3, respectively.
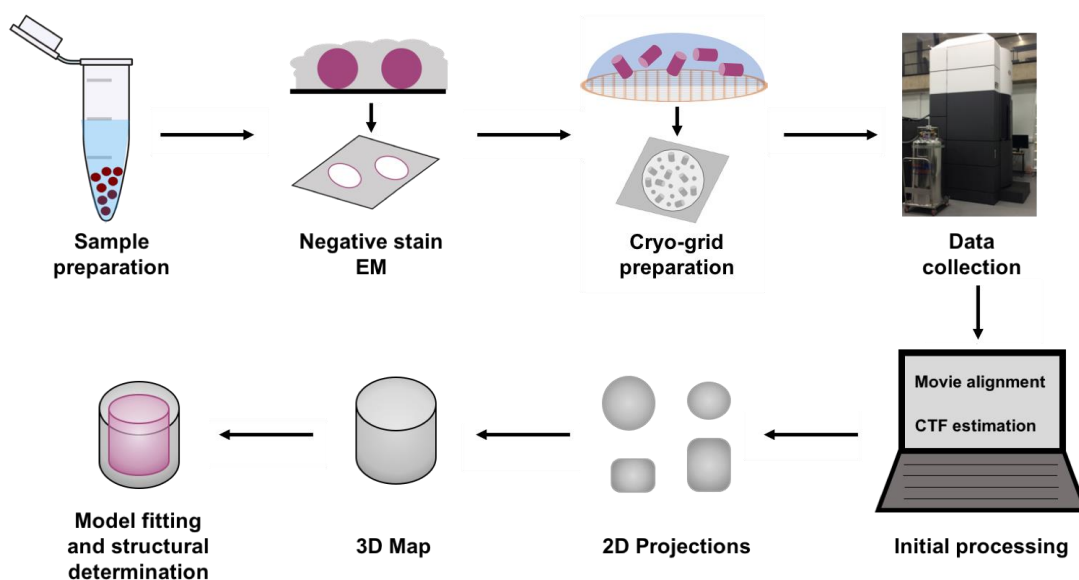
Figure 2.1: The pipeline from collecting sample to structural determination. *Initially the sample is analysed by negative stain EM before being taken to cryo-EM where the first stage is to optimise the cryo-grid conditions. After data collection, the individual movie frames are first motion-corrected and CTF estimation is performed. 2D classification of the particle stack results in numerous 2D projections of the sample which are recombined to produce a 3D map. The model is subsequently fit into the density, either by flexibly fitting an existing structure or if the resolution allows, by de novo building it into the density to generate the structure of the protein. In the schematic, the protein exists as a pink cylinder.*

2.1.1. Sample preparation

To facilitate visualisation of the specimen under the electron microscope, the sample is applied to grids which are 3 mm in size, contain a metal mesh (gold or copper) and are overlaid with a thin layer of amorphous carbon (or sometimes gold)[176]. This means that the surface of the grid is hydrophobic. Before the sample is applied to the gird, it is glow-discharged which results in ionised gas molecules being deposited onto the surface of the grid rendering it hydrophilic thus preventing the sample droplet from being repelled from the surface of the grid[177]. The protein of interest is made from low molecular weight atoms such as hydrogen, oxygen, carbon and nitrogen and is therefore a weak phase object[178]. One method of visualising the protein molecules under the electron microscope is by negative staining.

Negative staining involves coating the biological specimen with a thin layer of heavy metal atoms, such as uranium, molybdenum or tungsten[177]. The heavy metal atoms can absorb the electrons more readily than protein molecules which enables the protein to become visible against the background. The protein molecules appear white compared to the darker background hence why the protein has been negatively stained[179]. These heavy atoms are less prone to radiation damage which helps to preserve the protein from radiation damage caused by the electron beam. Moreover, small amounts of sample are used to prepare the negative stain grids with concentrations in the 10-100 μg/mL range commonly being used. However, artefacts such as flattening can occur which is a disadvantage of using this method. Through the stain grain size and limitations of the technique, the typical resolutions achieved are ~20-40 Å[179]. Yet, analysing the protein via negative stain can provide useful insights into the purity and heterogeneity of the sample before it is taken forward to cryo-EM. Moreover, analysis of negative stain data can provide unique biological insights into the mechanism of the protein as there are examples of inhibitor and antibody binding sites on therapeutic targets being revealed. This is exemplified by visualisation of PA1b binding to V-ATPase[155] and a neutralising antibody bound to the HCMV gH/gL/gO glycoprotein complex[180].

To facilitate high-resolution structure determination, cryo-electron microscopy, (cryo-EM) is used. This involves rapidly freezing the sample using a cryogen, normally liquid ethane, propane or a mixture of both, which freezes the sample in a thin layer of vitreous ice permitting its visualization in the high vacuum of the microscope and partly protecting the sample from radiation damage[181]. For cryo-EM

grids, the support film is usually made from copper or gold and contains holes which can either be regular in size and distribution (quantifoil grids) or contain an irregular arrangement of holes which vary in shape and size (lacey grids). Examples of the grid types are shown in Figure 2.2.
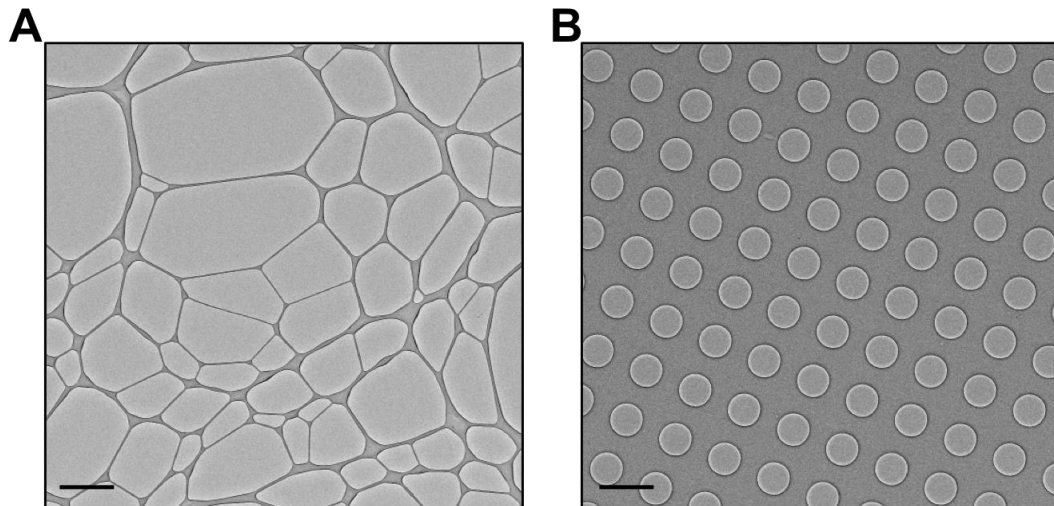
Figure 2.2: The types of cryo-grids used. A) A lacey carbon grid which contains an irregular array of holes. B) A quantifoil R1.2/1.3 grid which contains a regular distribution of evenly sized holes. Scale bars are 2 μm. Adapted from[182].

Preparing cryo-EM grids

Compared to the other areas of EM which have recently undergone rapid developments, such as image processing and camera advances, grid preparation remains largely under-developed. Grids are commonly prepared using the same approaches which were first developed in the 1980's by Jacque Dubochet[75,76]. These grids can be prepared manually or by using so-called 'first generation devices', such as the Leica or Vitrobot. The basic principle is that ~3 μL of sample is applied to a cryo-EM grid in a humidity-controlled chamber. The grid is subsequently blotted with filter paper, to remove excess liquid, before being plunge frozen into a cryogen thus suspending the protein in the thin layer of vitreous ice. The process has been summarised in Figure 2.3. By using these devices, the blotting force and blotting time can be controlled which in turn adjusts the thickness of the ice[183].
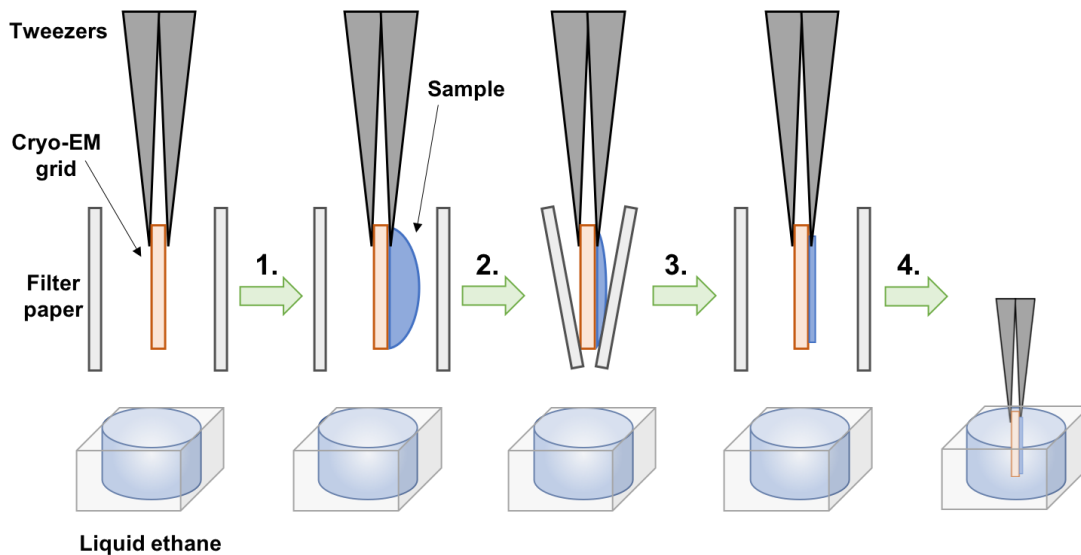
Figure 2.3: A schematic of how cryo-EM grids are made. *Initially the grid is held in a humidity-controlled chamber by a pair of tweezers above the liquid ethane. 1. The protein sample (3 μL) is applied to the cryo-EM grid. 2. The grid is blotted using filter paper to create a thin layer of the sample (3). 4. The grid is plunge frozen into the liquid ethane to vitrify the sample. The grid can now be loaded into the microscope or stored at liquid nitrogen temperatures. Adapted from[184].*

A disadvantage of the blotting approach is that because excess liquid is being blotted away, more than 99% of the sample is being wasted and not applied to the grids which represents a large waste for precious protein samples. Typically, blotting times are on the second timescale which can cause problems with the sample interacting at the air water interface. For an ideal cryo-EM grid; the protein would be evenly distributed, would reside in the holes in a wide range of orientations, have an even ice thickness and the protein molecules would occupy the centre of the ice layer without any interactions at the air-water interface (Figure 2.4A)[182,185]. However, this is not always the case. Common problems with grids include the particles adopting a preferred orientation within the ice (Figure 2.4B), the protein interacting favourably with the carbon support resulting in few particles within the holes (Figure 2.4C) or the protein being unevenly distributed within the holes (Figure 2.4D). The latter is common for membrane protein samples where a 'halo effect' is sometimes observed[182]. This is often caused by the centre of the ice being too thin to accommodate the protein causing them to be clumped towards the outside of the holes thus resulting in protein molecules overlapping with one another within the ice layer which is unsuitable for data processing. This can be overcome by using a

smaller hole size or by imaging in areas of thicker ice, although this can be detrimental to the contrast of the images obtained.
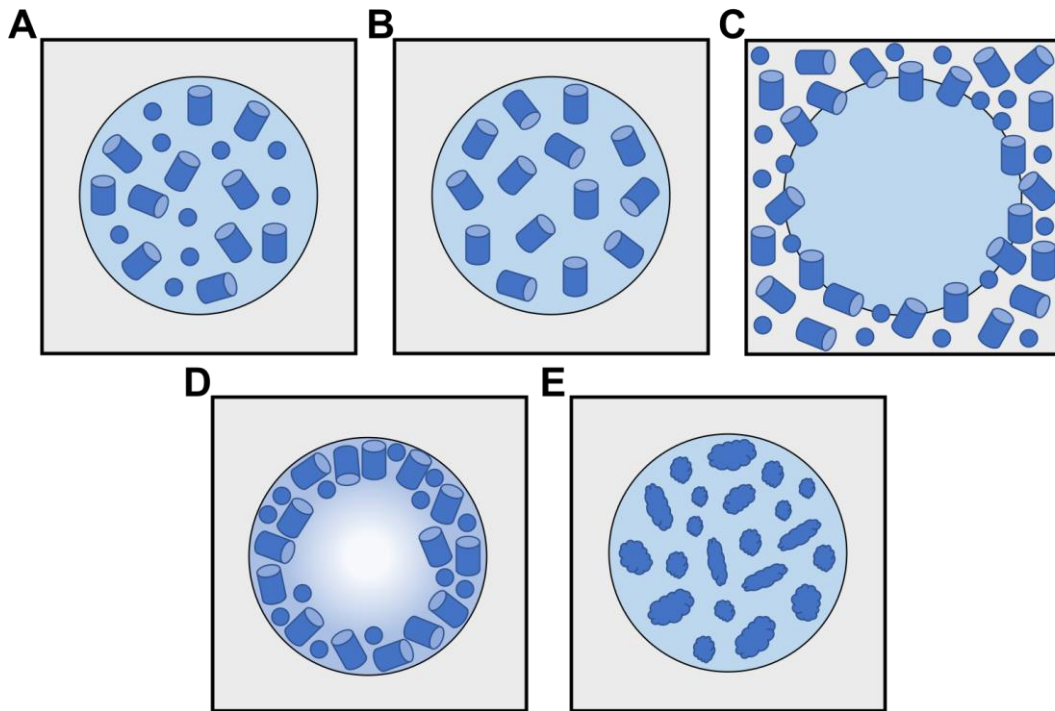
Figure 2.4: Particle distribution across the hole. A) An ideal distribution of particles (blue cylinders) across the holes where the protein is evenly spread across the hole and adopts different orientations. Non-ideal distributions: B) Protein adopts a preferred orientation within the ice. In this case the top view of the cylinder is missing. C) Protein has a high affinity for the carbon support layer and does not enter the hole. D) Example of the 'halo effect' where the protein is pushed towards the outside of the hole and the centre of the ice is very thin or not present. E) Protein is unfavourably interacting at the air water interface causing the protein to denature (non-uniform cloud shapes). Adapted from[182].

Moreover, another common problem which occurs is that the protein can interact with the air-water interface resulting in preferred orientation or protein degradation (Figure 2.4E)[186]. It has been shown that the protein can diffuse and interact with the air-water interface on a millisecond timescale[187]. Preventing the particles from interacting with and subsequently denaturing at the air water interface is the subject of many on-going studies[187].

After the EM grids have been prepared, either by negative staining or by plunge-freezing, the grid can subsequently be loaded into the electron microscope to allow visualisation of the specimen of interest. This will be described in Section 2.1.2.

2.1.2. The microscope

After the EM grids have been prepared they are inserted into an electron microscope to enable visualisation of the protein of interest. The microscope contains an electron source which emits electrons. The emitted electrons are subsequently directed through the column by a series of electromagnetic lenses until they ultimately reach an image detection system, which could be a screen or camera[188]. The column is held at high vacuum to prevent the electrons from interacting with any air inside of the column which would provide noise in the resulting images. A basic schematic of the microscope is shown in Figure 2.5.

Electron source

Condenser lens

Condenser aperture

Sample

Objective lens

Objective aperture

Projector lens

Detector
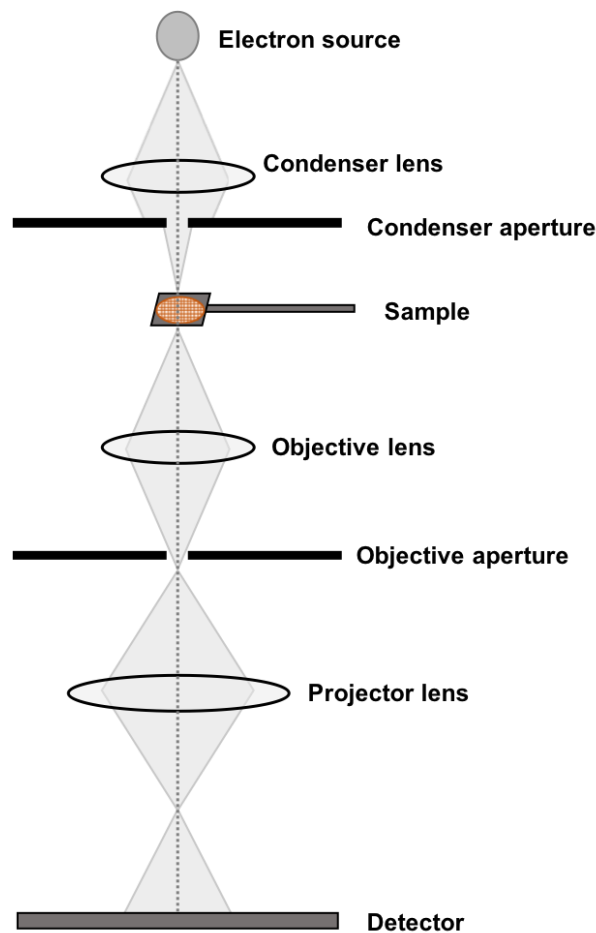
Figure 2.5: A schematic of the electron microscope. The dotted grey line represents the central axis along the column. The electron beam is directed through a series of electromagnetic lenses and apertures to create a magnified image of the sample. Adapted from[188].

Imaging the specimen

For high-resolution structure determination, field emission guns (FEGs) are used as the electron source and the electrons are accelerated through the column at a voltage of 300 kV. Although, high resolution structures have also been obtained from 200 kV microscopes[86]. Imaging with electrons allows high resolution structures to be determined as the short wavelengths do not impose a limit on the resolution which can be achieved (0.02 Å at 300 kV)[189]. After the electrons have been omitted from the gun, they pass through the condenser lens and reach the sample which is situated in the objective lens system. An objective aperture is placed after the sample which filters out all of the high scattering electrons which helps to reduce the noise in the images. The electrons then pass through the projector lens system before arriving at the detector.

Electrons interact with the sample in two main ways; they can be elastically scattered (without energy loss) or inelastically scattered (where some energy is transferred to the specimen)[190]. Electrons which have been elastically scattered are useful for the image formation whereas inelastically scattered electrons can have detrimental effects on the sample. For instance, when the electrons interact with the protein they can cause chemical bonds to break, atoms to become ionised or result in secondary electron scattering events. All of these effects can cause damage to the sample which is known as radiation damage[190]. Therefore, imaging the specimen with the correct electron dose becomes a balance. The dose needs to be high enough to produce contrast and see the protein but having too much dose would cause the sample to become irradiated and subsequently destroy the structure of the protein. Therefore, low total doses are typically used (~20 $e^-/Å^2$)[191]. It has been estimated that 3-4 of these inelastic scattering events occur for every elastically scattered event[69]. This radiation damage is one of the main limiting factors when imaging biological molecules using cryo-EM.

Another consideration which needs to be considered for cryo-EM is contrast. Contrast can be split up into amplitude and phase contrast[188]. For negative stain EM, suspending the proteins within a thin layer of heavy metals introduces contrast into the images as the electrons are scattered by the heavy metal atoms which allows the protein to be seen (as discussed in Section 2.1.1). This has introduced amplitude contrast into the images. However, for cryo-EM the buffers in which the protein is stable usually consist of the same low molecular weight atoms as the protein (C, H, N and O). This means that there is very little amplitude contrast within the images.

Furthermore, as biological samples are weak phase objects, very small amounts of phase contrast are introduced into the image[178]. Therefore, if images were taken at focus in the electron microscope, the protein would not be visible. This results in the ideal optics of the electron microscope being compromised as images are recorded with small amounts of defocus which introduces a phase shift and subsequently contrast into the images. In practical terms, adding defocus means focusing the objective lens to a few microns beyond the specimen[192]. Moreover, imaging specimens with defocus also becomes a balance because images taken closer to focus (low defocus) retain more of the high-resolution information. However, the particles are much harder to see than images taken with a higher defocus. Defocus ranges within data sets are typically between -1 and -3 μm. Figure 2.6 shows changes in contrast at different defocus values. Phase plates, such as the Volta Phase Plate (VPP), can also be used to introduce contrast into images and allows images to be recorded closer to focus[193–195].
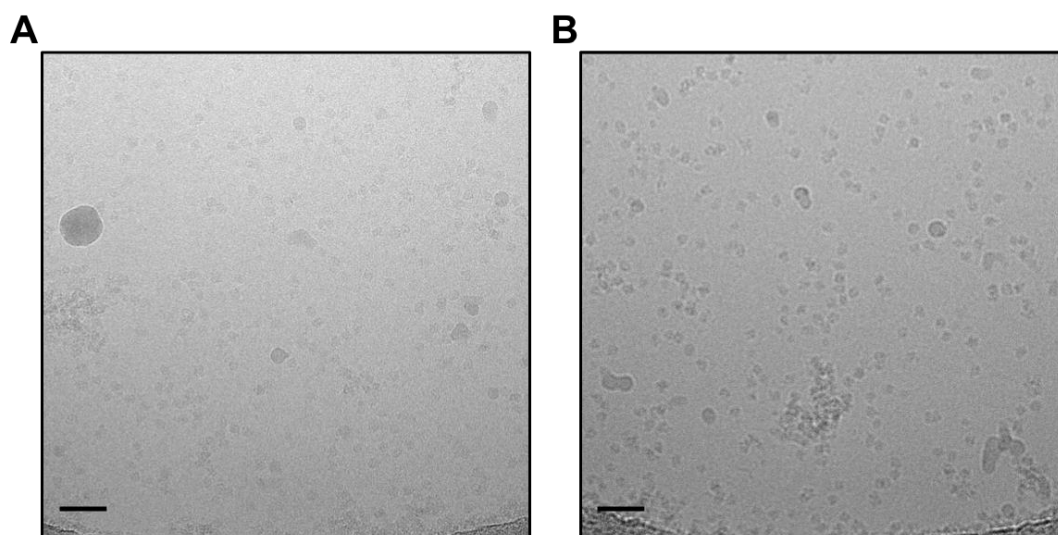
Figure 2.6: Micrographs with varying defocus. Examples micrographs of a 95 kDa membrane protein taken at two different defocus values; A) 1.1 μm, B) 4.3 μm. The higher defocus micrograph in B), the higher the contrast in the image allowing the particles to be easily visualised. Scale bars are 36 nm.

Direct electron detectors

Once the electrons have passed through the column, they reach the detection system. One of the major reasons behind the 'resolution revolution' in cryo-EM is the introduction of direct electron detectors which records individual electron events thus enabling the signal-to-noise ratio of the data to be improved resulting in higher

contrast within the images[196,197]. Moreover, instead of taking single images, due to their high frame rates, the detectors now record movies broken down into a number of individual frames which have only been exposed to a small proportion of the total dose in a single exposure. This is advantageous because during a typical exposure, radiation damage to the sample and/or mechanical drift in the microscope can cause blurring of the image. This has allowed the frames to be dose-weighted upon completion of the data set[198]. Therefore, the first and final few frames of the data set can be down-weighted as they have suffered from the largest amount of beam-induced motion and radiation damage, respectively. As the images are recorded as movies, a higher total dose (40-80 $e^-/Å^2$) can be used. This will be discussed in more detail in the image processing Section 2.1.3.

An ideal detector would not add any noise to the cryo-EM images and would have a detective quantum efficiency (DQE) of 1. The DQE is measured by the square ratio of the output signal-to-noise ratio (SNRo) over the input signal-to-noise ratio (SNRi) at a given spatial frequency[199,200]:

$$DQE = SNRo^2/SNRi^2$$

The detectors contain monolithic active pixel sensors (MAPs) which record individual electrons as they deposit energy whilst passing through a semiconductor membrane[201]. The membrane is typically ~10 μm in depth and the electrons can be back scattered meaning that they can pass through the detector multiple times, adding noise into the image. Therefore, detectors have been back-thinned which means that the depth of the membrane has been reduced as much as possible to reduce these interactions[201]. There are three main direct electrons detectors which are routinely used; Gatan K2 or K3, Falcon III and the Direct electron[202]. Detectors such as the Falcon III can be operated in two different modes; integrating and counting. Integrating mode integrates the signal of all of the electrons hitting the detector and has short exposure times of 1-2 seconds. Whereas, for counting mode, the individual electrons which are hitting the detector are recorded which means that the location of the incoming electron hitting the detector can be determined. As this is more sensitive than integrating mode the DQE is higher and more higher resolution information can be obtained[202]. However, counting mode detectors are dependent on having low dose rates which increases the exposure times of the images; typically, ~60 seconds for the Falcon III and ~8-10 seconds for the Gatan K2.

2.1.3. Data collection and image processing

After the microscope has recorded the images and the data set has been collected, the data needs to be processed in order to obtain a structure. This is usually carried out using the technique called single particle averaging (SPA)[183]. The principle behind the technique is that the protein is frozen in its native state in as many different orientations as possible. Individual protein particles are then picked, either by hand or using automated software, to generate a typical stack containing hundreds of thousands to millions of particles[181]. These particles are classified based upon their different orientations which means that the signal within the individual particles is averaged together to increase the signal-to-noise ratio. The classes are 2D projections of the sample which can be combined together to generate a 3D reconstruction of the protein. Postprocessing and polishing steps are subsequently completed in order to improve the quality of the map before a model is built into the density. The steps will be described below and an overall schematic of the process is shown in Figure 2.7 although the order of the steps and the number of times they are performed can vary.
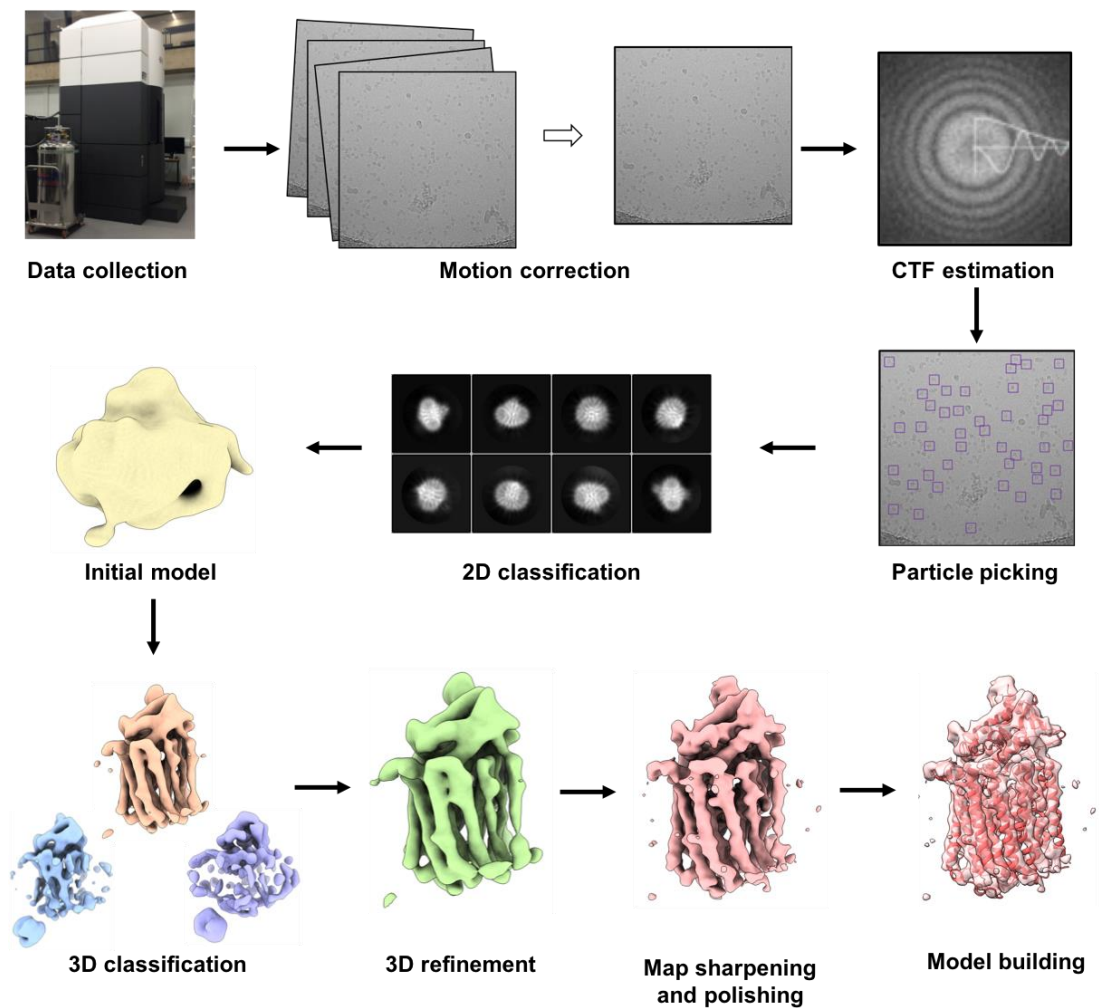
Figure 2.7: An overview of the cryo-EM processing pipeline. *Processing schematic exemplified for a 95 kDa membrane protein. Data were initially collected on a high-end microscope with images collected as movies. After the frames were aligned, the CTF estimation was performed subsequently enabling particles to be picked. The particles were classified based upon their orientation and were used to generate an initial model. 3D classification was performed to remove junk particles with the best class being refined. The map was post-processed to restore the high-resolution information before a model was built into the map.*

Data collection and motion correction

For most high-resolution data collections, typically 1,000-10,000's of images are recorded. The individual images are recorded as movies composed of individual frames and are termed micrographs. There are a variety of different software programs available which enable the micrographs to be collected automatically such as SerialEM[203], Leginon[204] and EPU (ThermoFisher). The user simply needs to select the areas of the grid where they would like to image and set the dose rates/defocus parameters for the data collection. After performing the direct alignments, the

microscope can then run unattended for the duration of the data collection (typically 48-72 hours).

Once the data collection has started, the micrographs can be processed in what is known as 'on-the-fly' processing[205]. On-the-fly processing has not only greatly enhanced the speed with which structures are determined but it also allows the quality of the data coming off the microscope to be monitored. Moreover, if there are any problems with the microscope then these can be quickly identified and solved to enable the user to collect the best possible data. The first step which is generally performed is the alignment of the individual movie frames which corrects for the beam-induced movement in the sample[206] by for example MotionCorr/MotionCor2[207]. To improve the accuracy of the motion correction the individual micrograph is generally split up into a 5x5 grid consisting of 25 squares. This gives more accurate results because the motion is not consistent across the whole micrograph (Figure 2.8). For instance, when the ice is thinner such as towards the centre of the hole, more beam-induced motion occurs than at the thicker edges of the holes which is exemplified in Figure 2.8B.
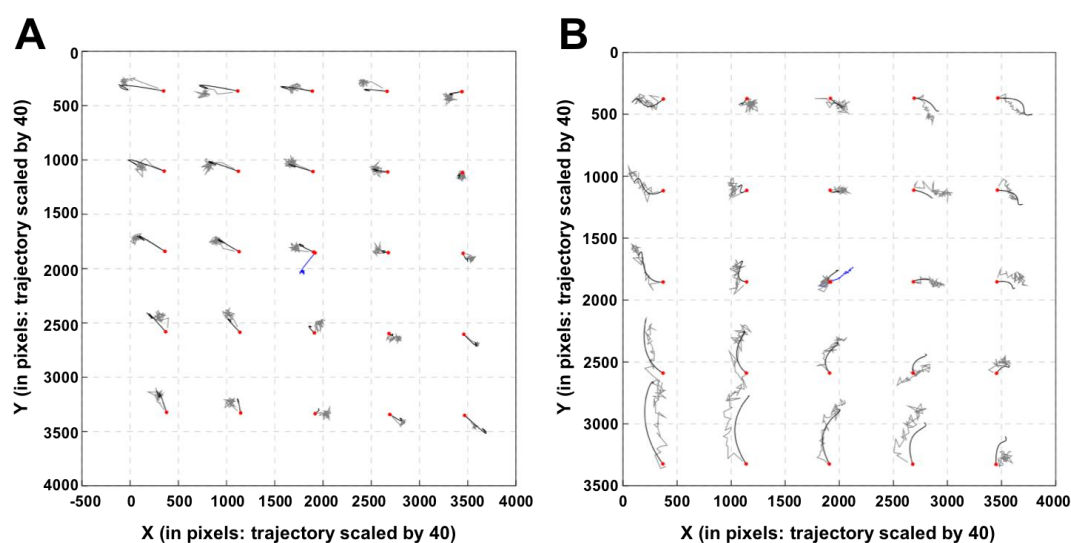
Figure 2.8: Whole-frame motion-correction. *An output from the MotionCor2 software packaged in RELION. The trajectory of the motion within the individual squares is shown by the grey lines. A) The motion is generally consistent across all of the micrograph compared to in B) where there is greater motion in the bottom-left corner. This could be due to a micrograph showing the 'halo-effect' where the ice is thin towards the centre of the hole.*

48

The general principle behind the motion correction is that the frames are aligned to one another to produce an image which is the overall sum of all of the frames[208]. This is completed by looking at the cross correlation between the frames and subsequently moving the frames so they become aligned. This can be done by using one frame as a reference or it can be an iterative process. This whole-frame approach can correct for large motions across an image but it does not consider the motion of the individual particles. This can be corrected for by particle polishing.

CTF estimation

In order to introduce contrast into the images, they are recorded with small amounts of defocus which introduce imperfections into the microscope. This information, and the amount of defocus applied, is held within the Contrast Transfer Function (CTF) of the microscope[188]. The information gleaned from the CTF estimation can subsequently be used to monitor the quality of the data collection[205]. Therefore, the CTF estimation is typically performed after the motion-correction stage of the processing and is often incorporated into many on-the-fly processing pipelines. Programs such as CTFFIND4[209], Gctf[210] and XMIPP[211] are commonly used to estimate the defocus in the micrographs and this is corrected for in subsequent processing steps.

The CTF is an oscillating function which measures the amplitude of an electron wave at a given resolution (spatial frequency). The 2D Fourier transform of the image allows the CTF to be observed as a sine wave which can be plotted through the different Thon rings in the power spectrum[188]. This is used to estimate the defocus in the micrograph. At higher defocus values, the frequency of the wave is increased which results in more information being lost due to the wave passing through zero more often. Applying defocus to the microscope shifts the position of these zero points therefore data is collected at a variety of defocus values[212]. This ensures that no information is lost after the data has been averaged together. As the curve is an oscillating function, the amplitude at some frequencies will be negative therefore the phases are flipped. This allows all of the sine waves to be summed together. A summary of the CTF estimation process is shown in Figure 2.9. The CTF is then corrected for in the classification/refinement steps of the processing[192].
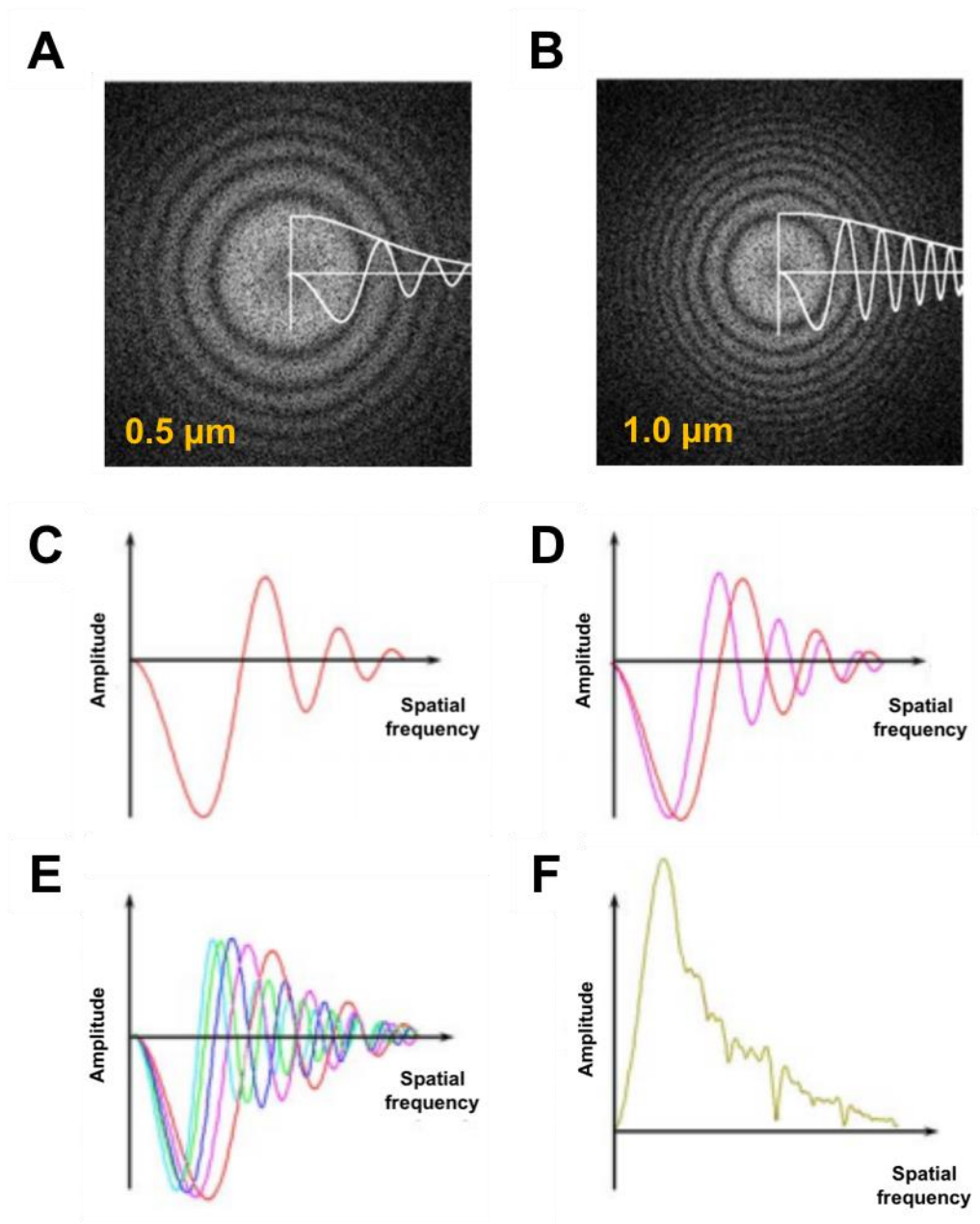
Figure 2.9: Estimating the CTF in micrographs. *The 2D power spectrum of micrographs with A) 0.5 μm defocus and B) 1 μm defocus applied. The sine wave can be fitted through the different Thon rings. C) An example CTF curve which shows that the amplitude has been plotted against the resolution. D) Two different CTF curves at two different defocuses which show that the curve is shifted. The pink curve is at a higher defocus than the red curve as the wave has a higher frequency. E) Example curves at a wide range of defocus values. This ensures that there are no resolutions with zero information when all of the data has been averaged. F) An example curve after the phases have been flipped and the sine waves have been summed together. Adapted from[188].*

Particle picking

   After motion correction and CTF estimation has been performed, individual particles need be picked and extracted from the micrographs[192]. This can be completed in a manual, semi-automatic or automated way. For manual picking the user goes through each micrograph and picks all of the particles by hand. Despite this being an accurate way of ensuring all of the particles are picked, it can be a very slow process, especially when the data set consists of greater than 1,000 images. One semi-automated approach which is commonly used is template matching. This involves the user manually picking ~2,000 particles by hand and classifying the resultant particles to generate 2D templates which can subsequently be used to pick the data after being low-pass filtered to ~30 Å. However, one problem with this approach is that the user may not select uncommon or difficult to see views which means that this view will be missed when all of the data is autopicked. Moreover, bias can also be introduced into the data if 2D projections of high-resolution structures are used to pick the data. This is commonly known as the 'Einstein from noise' problem[213] which is exemplified in Figure 2.10.
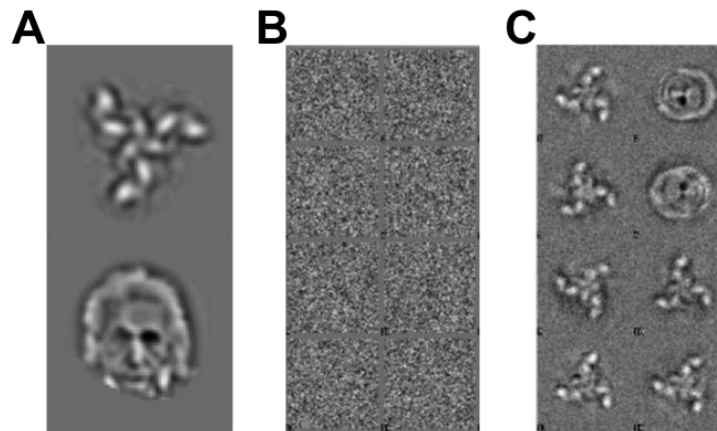
Figure 2.10: Einstein from noise. An example of introducing bias into the particle picking. A) The references which were used to pick the B) noisy images. C) The classes obtained from the data. Adapted from[213].

   There are fully automated programs which can be used to pick the micrographs such as RELION3.0 Laplacian of gaussian filter[214] and the swarm picking tool used in EMAN2[215]. Moreover, machine learning algorithms have been developed and implemented in programs such as crYOLO[216]. Having been trained on a range of different data sets, the user just needs to tell the program the particle diameter and

the general model is able to pick all of the particles within the micrograph without the user having to generate any 2D classes. Although, if the algorithm fails to pick particles the user would need to train the model so it acts specifically on the data set by manually picking ~20 micrographs.

Classification and alignment of particles

The picked particles are extracted from the micrographs into boxes which are ~1.5-2.0x the size of the longest diameter of the particle. Individual micrographs and particles have different signal-to-noise ratios and are therefore normalised which allows dark or light pixels, which would dominate the alignment of the particles, to be removed[217]. Individual micrographs can contain protein aggregates, carbon from the edges of the holes or ice contamination. During autopicking, 'particles' from these areas can be picked which do not represent any view of the protein of interest. These particles are subsequently removed during 2D classification which groups together particles of the same view thus increasing the signal-to-noise ratio.

Approaches to 2D classification are varied between different programs. They can use either references picked by the user, or a reference-free approach to align the particles[188]. One of the most commonly used programs to process cryo-EM data is RELION. RELION uses a reference-free maximum likelihood approach for aligning and classifying the particles[218]. Initially, the references used are random and the user only has to specify the number of classes. All of the particles are then aligned to all of the references in all possible orientations. Probabilities are then assigned based upon how likely it is that the particles contribute to each class which are based upon the translation, rotation and degree of matching to each reference[219]. The translated/rotated particles are then averaged together in order to form classes which act as references in the next iteration. This results in every particle contributing somewhat to every class. However, particles usually contribute strongly to one or two classes and negligibly to the others[219]. Particles which contribute equally to a large number of classes can also be identified and this is indicative of the particles not representing any view of the protein allowing the 'junk' particles to be removed from the data set.

Many data sets contain heterogeneous populations, with the protein occupying different conformational states. Therefore, it is often useful to perform 3D classifications to separate out the different states. The same maximum likelihood approach can be utilised for unsupervised 3D classifications[220]. An initial model of the

protein, filtered to low resolution (~30-60 Å) is provided and the user only needs to specify the number of classes to split the data into, which should represent the number of conformational states the protein resides in. Similar to 2D classification, the probabilities of the particles contributing to the different 3D models can be calculated and the models generated can be used as references for the next iteration. The classification would finish after the number of particles which change class stabilises[220]. The particles which make up each class is usually unambiguous and these particle subsets can subsequently be refined separately to obtain structures of the different conformational states. Moreover, the 3D classification could be performed around a certain domain or area of the protein and a focussed classification of this region could be performed[220].

Reconstructing a 3D map

     After 2D classes of the protein have been obtained the next stage is to generate a 3D map. The 2D classes are a 2D projection of the 3D object. Therefore, the 'projection-slice theorem' can be used to obtain a 3D map of the protein[221]. The theorem states that the Fourier Transform (FT) of the 2D image represents a central slice-through of the 3D object. A schematic of the theorem is shown in Figure 2.11.
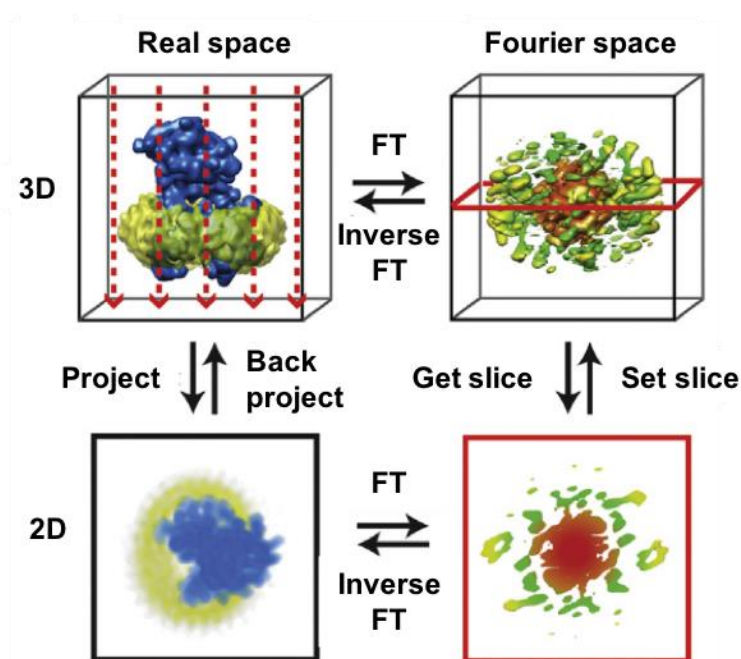
***Figure 2.11: A schematic of the Fourier transform theorem**. A 3D object can be back projected to give a 2D central slice of the 3D object. The Fourier transform of this central slice can then be combined in Fourier space. The inverse of this will give a 3D map. Adapted from[221]*

By having 2D classes of the protein in a variety of different orientations, the 3D structure can subsequently be calculated[221]. This underlies the projection-matching approach which is utilised in a number of programs. For this approach, the individual particle images are compared to back-projections of a user-provided 3D model in lots of different directions. The angular assignment for the back projections can then be used to calculate the 3D map of the protein[188]. This is done by placing the 2D Fourier transforms of the slices into the 3D transform and subsequently determining the 3D map by calculating the inverse transform. This process is performed iteratively because the 3D map will diverge from the initial model and more accurately resemble the 3D structure of the protein[221]. As there are large amounts of uncertainty in the angular assignment, a maximum likelihood approach can again be taken, as implemented in RELION. This allows a probability to be estimated for each of the angular assignments which in turn produces more reliable models[219].
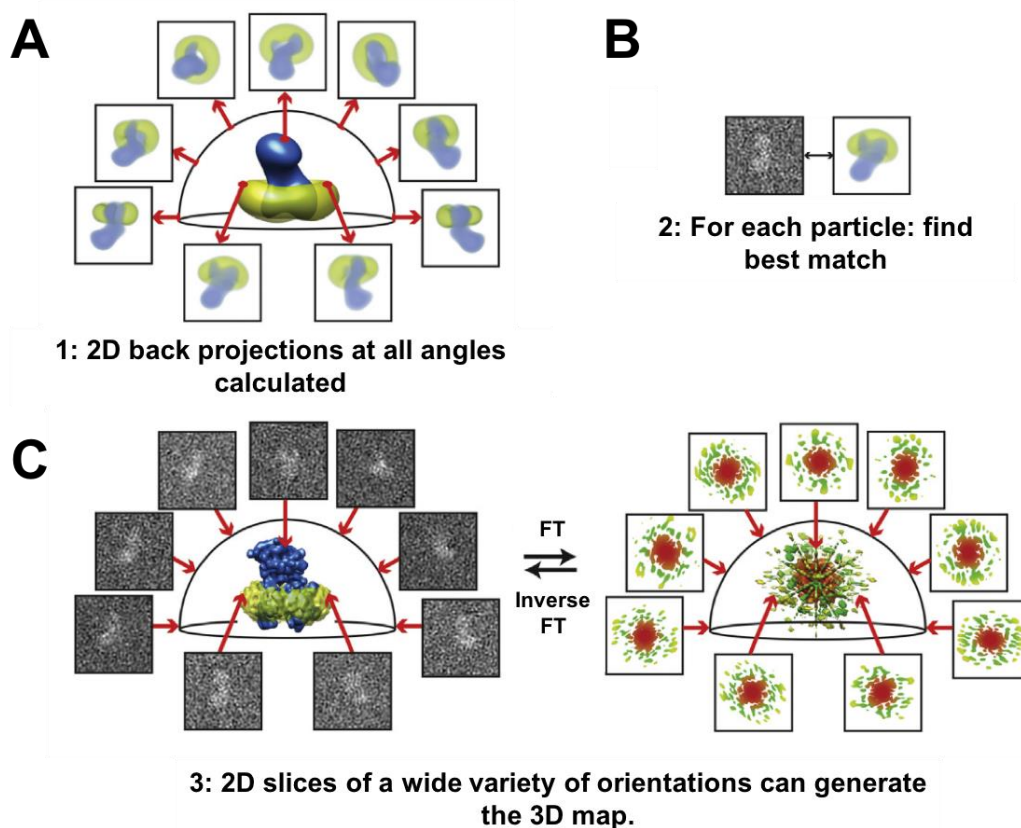
Figure 2.12: Schematic of the projection-matching theorem. *A) 2D back projections of the initial model are calculated in all directions. B) The particles/classes are compared to the back-projection to see which view gives the best match. C) All experimental images are then orientated relative to the 3D model. The projection-matching theory can then be applied to combine all of the projections in Fourier space which can then be inverse-transformed to give a 3D map. Adapted from[221]*

One limitation of the projection matching algorithm is that the user needs to provide an initial model which could again introduce model bias into the data[192]. Therefore, high resolution structures should not be used and any existing structures should be low-passed filtered to ~30-60 Å so only the vague globular shape of the protein is used. To prevent biasing the data, an *ab initio* initial model could be computationally calculated from the experimental data. This could be done by using a common-line method, random conical tilt or by using tomography[67]. In programs such as RELION and cryoSPARC[222], this is generally completed by using a stochastic gradient descent algorithm which allows accurate initial models to be obtained.

Another limitation of the projection-matching approach is that the refinement can get stuck in a local minimum and converge on the incorrect structure[221]. Therefore, it is important to ensure that the initial model does resemble the structure of the protein, despite being filtered to low resolution. After the refinements have finished, the maps can be analysed. For maps where the resolution is sufficient to visualise secondary structure (<9 Å for α-helices and <4 Å for β-sheets), crystal structures or homology models can be docked into the density. However, assessing whether the obtained map is correct can be difficult at low resolutions (~ >10 Å), especially if there is no existing structural information of the protein. Tilt-pairs can be collected and used to validate low resolution maps[223,224].

During the 3D refinements, the data is generally split into two independent subsets and the particles in each subset are refined separately. This allows the Fourier shell correlation (FSC) to be calculated between the two maps[225]. The FSC gives information on the resolution of the map and provides information on the signal-to-noise ratio at a given spatial frequency (resolution). Therefore, the resolution of cryo-EM maps is calculated by how well the two half maps correlate to one another at a given resolution, as shown in the FSC curve. This cut-off level has been intensely debated within the field. However, resolutions are generally reported using the 0.143 cut-off level; which is often termed the FSC gold-standard[226,227]. The problem with this approach though is that if the map is incorrect and is mostly made up of noise, so long as this noise is consistent between the two half maps, a high-resolution map could be reported. Therefore, it is often wise to analyse the maps and assess whether the quality of the map represents the reported resolution. For instance, if the reported resolution is less than 3.5 Å, then all secondary structure should be resolved and individual side chains should be visible for at least the bulky residues.

After the 3D map has been obtained, inspection of the map shows that it often does not contain the features which are expected at the reported resolution. This is due to the high-resolution features being lost as a result of radiation damage, imaging imperfections or errors introduced in the refinement process (eg in angular assignments)[228]. This loss of information can be modelled by a Gaussian amplitude decay of structure factors which is known as a B-factor. In order to restore the high-resolution features of the map, it needs to be 'sharpened'. One way to do this is to estimate the decay by comparison to a known theoretical scattering curve and to apply a weighting function which is used to avoid the amplification of noise and to prevent over-fitting[226]. The B-factor can then be implemented which enables the high-resolution features to be up-weighted so they can be visualised within the map. Automatic calculation of the B-factor has been included in programs such as EM-BFACTOR[228], which has subsequently been incorporated into the RELION pipeline. Other ways to improve the quality of the map are to perform per-particle CTF estimation and individual particle motion-correction (polishing). Both steps are implemented into the RELION3.0 pipeline and have improved the resolution and quality of maps[214]. The output from both jobs are shown in Figure 2.13 which shows that the CTF and motion of individual particles is not consistent within a micrograph.
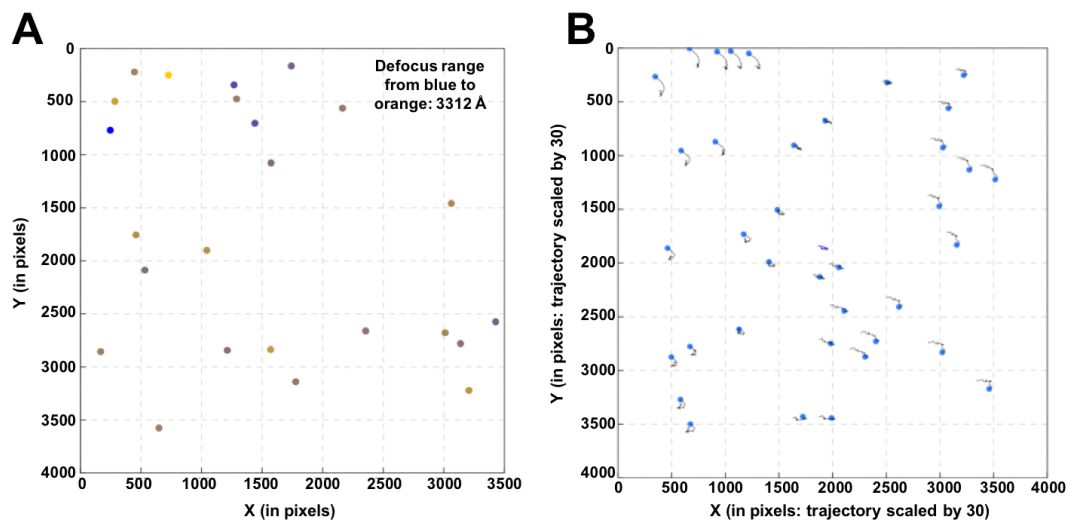
Figure 2.13: Per-particle CTF estimation and polishing. A) Particles within a micrograph are represented as circles. The colours represent different defocus values which show that the CTF of the individual particles is not consistent across a micrograph. B) Individual particle motions within a micrograph which show that the particles move randomly within the micrograph.

Model building

After the final map has been obtained, the model can be fit into the map or built from scratch depending on the resolution and whether there are any existing structures available. For high resolutions where side chains are visible, the model can be built and refined into the density using programs which have traditionally been used by X-ray crystallographers. These include COOT[229], Phenix[230] and Refmac[231]. For more moderate resolution structures, where only secondary structure is visible, programs such as MDFF[232] and ISOLDE[233] can be used to flexibly fit a model into the density. Although care needs to be taken not to overfit or overinterpret the structure when the map is at these lower resolutions. For instance, commenting on side chain positions when there is no side chain density present in the map would be an example of over-interpreting the data. Moreover, for low resolution maps, large protein domains or existing crystal structures could be rigidly positioned into the cryo-EM density to determine the approximate location within a complex. Model fitting or building is the final step in the structure determination pipeline and allows the intricate structural details of the protein to be obtained.

2.2. Cytochrome bc_1

2.2.1. Preparation of negative stain grids

Carbon-coated copper grids were glow discharged for approximately 30 seconds prior to sample application using a Pelco glow discharge unit. 3.0 μL of 10.0 μg/mL protein was added, followed by 2 × 3.0 μL of 1% uranyl acetate solution which were blotted away after 60 seconds using filter paper. Images were taken on an FEI T12 microscope at 50,000x magnification.

2.2.2. Preparation of cryo-grids

Purified cytochrome bc_1 was buffer-exchanged in 25 mM Tris pH 7.5, 100 mM NaCl, 0.5 mM EDTA, 0.015% DDM or 0.01% LMNG by several dilutions in a centrifugal ultrafilter and adjusted to a concentration of 5 mg/mL. 3 μL aliquots of the sample were applied to Quantifoil Cu R1.2/1.3, 300 mesh holey carbon grids, which had been glow discharged for 30 seconds using a Pelco glow discharge unit. An FEI Vitribot Mark IV was used to blot the grids for 6 seconds (blot force 6) at 100% humidity and 4°C before plunging into liquid ethane. The grids were loaded into an FEI Titan Krios transmission electron microscope (University of Leeds Astbury Biostructure Laboratory) operating at 300 kV.

2.2.3. Cryo-EM data collection and image processing

For all inhibitor-bound bc_1 data sets, data were collected on a Titan Krios microscope (300 kV) fitted with a Falcon III direct electron detector which was operated in integrating mode. Automated data collection was carried out using EPU software, with a defocus range of -1 to -4 μm, and a magnification of 75,000 x which yielded a pixel size of 1.065 Å. Five data sets with four different inhibitors (MJM, GSK, SCR and JAG) bound have been collected. The total dose, exposure times and dose per frames varied slightly for each data set and is summarised in Table 2.1. The apo-bc_1 data set was collected on Krios 2 at the Astbury Biostructure Laboratory which was fitted with a Gatan K2 direct electron detector operating in counting mode. Data were automatically with a pixel size of 1.047 Å and a defocus range of -1 to -4 μm. The total dose was 44 e⁻/Å² over a 12 second exposure which was split into 20 frames. The apo, MJM, GSK, SCR and JAG bound bc_1 data sets were collected over three days resulting in 3,256, 5,903, 8,840, 7,893 and 5,356 micrographs respectively. For the JAG-bc_1 data-set, 5,356 micrographs were collected but a large proportion contained crystalline ice so these were manually removed leaving 2,960 micrographs.

	bc_1-apo	bc_1-MJM	bc_1-GSK	bc_1-SCR	bc_1-JAG
Detector	K2	Falcon III	Falcon III	Falcon III	Falcon III
Detector mode	Counting	Integrating	Integrating	Integrating	Integrating
Voltage (kV)	300	300	300	300	300
Pixel size (Å)	1.047	1.065	1.065	1.065	1.065
Defocus (μm)	-1 to -4	-1 to -4	-1 to -4	-1 to -4	-1 to -3.5
Total dose (e⁻/Å)	44	110	75	85	66
No. of frames	20	39	50	40	59
Exposure time (s)	12	2.0	1.5	2.0	1.5
Dose per frame	2.20	2.82	1.50	2.125	1.12
No. of micrographs	3,256	5,902	8,840	7,893	5,356
Total particle No.	260,201	78,306	466,865	629,258	439,009
Final particle No.	57,571	41,223	232,910	114,130	211,916
Resolution	4.4 Å	4.6 Å	4.1 Å	4.1 Å	3.3 Å

All of the processing was performed in RELION2.1 or RELION3.0[234] unless otherwise stated. For all data sets the initial drift and CTF correction was carried out using MotionCor2[235] and Gctf[210] respectively. All of the other processing steps will be described within the relevant sections in Chapter 3. Existing crystal structures were rigidly fit into all of the maps using UCSF Chimera[236], before MDFF[232] was used to flexibly fit the model into the map. Rosetta was subsequently used to perform model relaxation before finally being refined using PHENIX[230]. The maps were then inspected manually in COOT[229] and the model corrected for any errors in refinement and the placement of residues.

2.3. Vacuolar-ATPase

2.3.1. Virtual screening of chemical compounds

Virtual screening was performed using the chembridge library which contained 100,000 compounds. All docking was performed using GLIDE[38] in the Schrodinger software Maestro (version 10.3 or 11.3.016). Initially the compounds were screened

in HTVS mode, ranked according to their predicted binding affinities and the top scoring 10,000 compounds were docked in the higher precision XP mode. The top scoring ~100-200 compounds were visually inspected using Pymol to identify the binding interactions to the protein and to rule out compounds with steric clashes to the protein. The structure of the compound was analysed to determine its 'drug-like' properties, ie whether they adhere to the Lipinski rule of five and do not contain any frequent hitter motifs. The top scoring compounds which fit the criteria were then ordered from Chembridge (Hit2Lead) and were shipped as dry stocks in 1.0 mg or 5.0 mg quantities. The compounds were dissolved in DMSO at 200 mM concentration and stored at -20 °C.

2.3.2. Yeast growth conditions

The strain of *Saccharomyces cerevisiae,* W303-1B was used throughout this study. Initially the *Saccharomyces* was grown on YEPD agar plates at pH 5.5 supplemented with adenine (10 μg/mL). Throughout the project the yeast were grown aerobically in YEPD media (1% yeast extract, 2% special peptone, 1% MOPS, 1% MES, 2% glucose) at pH 7.5.

2.3.3. Preparation of yeast vacuolar membranes

Vacuolar membranes were isolated from *Saccharomyces* by a modification of the method by Uchida *et al* 1988[29]. Routinely, 10 x 1 L cultures were grown to late log phase (OD_{600} ~ 5) in a shaking incubator (200 rpm) at 30 °C and harvested by centrifugation at 4,500g for 10 minutes (JLA8.100 rotor). The cells were washed with water, followed by 1.1 M sorbitol with 1% glucose (pH 7.5) and recovered by centrifugation (3,000g, 4 °C, JLA10.5 rotor). The cell pellet was resuspended in 100 mL of 1.1 M sorbitol/litre of culture (buffered with 5 mM MOPS/MES, pH 7.0). 2-mercaptoethanol (0.5 mL) was added to increase efficiency of spheroplast formation. Yeast lytic enzyme (Zymolase 100T, amsbio) was added to a final concentration of 0.1 mg/mL and incubated at 30 °C for ~90 min with gentle shaking (30 rpm with manual swirling every 15-20 mins). The spheroplasts were recovered by centrifugation (3,000g, 5 mins, JLA10.5 rotor), washed with 1.1 M sorbitol/1% glucose and recovered by centrifugation (3,000g, 5 mins, JLA10.5 rotor). The pellet was resuspended in 80 mL Buffer A, containing 10 mM Mes/Tris, pH 6.9, 0.1 mM $MgCl_2$, 12% (w/v) Ficoll-400 with, 100 μL (Sigma) Protease inhibitor cocktail and homogenised using a loose-fitting Dounce homogeniser (8 strokes). The homogenate was centrifuged at 10,000g for 10 minutes at 4 °C using the JLA10.5

rotor to remove cell debris. The supernatant was transferred to Beckmann SW32Ti centrifuge tubes and overlaid with Buffer A. The tubes were centrifuged using the SW32Ti rotor at 52,000g (17,400 rpm, 55 mins, 4 °C). This produced a white float of crude vacuoles which were carefully removed using a spatula and re-suspended in 25 mL Buffer A, supplemented with 50 μL protease inhibitor cocktail. This suspension was further homogenised using the Dounce homogeniser (5 strokes) and transferred to fresh Beckmann SW32Ti tubes and overlaid with Buffer B (10 mM Mes/Tris, pH 6.9, 0.5 mM $MgCl_2$, 8% (w/v) Ficoll 400). These tubes were again centrifuged using the SW32Ti rotor at 52,000g (17,400 rpm) for 55 minutes at 4 °C. The white float of purified vacuoles was carefully removed using a spatula and resuspended in 16 mL of Buffer C (10 mM Mes/Tris, pH 6.9, 0.5 mM $MgCl_2$ and 25 mM KCl) which enabled the hypertonic lysis of vacuoles, leading to the formation of vacuolar membrane vesicles. The vesicles were recovered by centrifugation (100000g, 30 mins, 4 °C, Beckmann TLA110 rotor). The pellet was resuspended in 20 mM Tris/HCl (pH 7.5), 20 μL of protease inhibitor cocktail, 1 mM EDTA and 15% glycerol and flash frozen in liquid nitrogen. The membranes were stored at -80 °C.

2.3.4. Protein purification from the vacuolar membranes

The vacuolar membranes were washed with EDTA to remove contaminating proteins. All subsequent steps were carried out at 4 °C. The membranes were suspended in a solution of 10 mM Tris/HCl, 1 mM EDTA (pH 7.5) at a protein concentration of ~1 mg/mL and homogenised three times in a loose fitting Dounce homogeniser. The membranes were recovered by centrifugation at 100,000g (4 °C, 30 minutes, TLA110 rotor). This EDTA wash was performed three times in total. The EDTA-washed membranes were resuspended at ~5 mg/mL in solubilising buffer (10 mM Tris-HCl, 1 mM EDTA, 2 mM DTT, 0.5 mM PMSF, 10% glycerol, pH 7.5) supplemented with protease inhibitor cocktail (Sigma). To this suspension, 25% DDM (w/v) was added dropwise to a final weight ratio of detergent to protein at 10:1. The suspension was stirred at 4 °C for 30 minutes and then centrifuged at 100,000g for 30 minutes (4 °C, TLA110 rotor). The supernatant was collected as the solubilised fraction.

For glycerol gradients, aliquots (500 μL) of the supernatant were then layered on top of a 14 mL 20-50% glycerol gradient in a solution of 10 mM Tris/HCl (pH 7.5), 1 mM EDTA, 2 mM DTT, 0.5 mM PMSF, and 0.01% (w/v) DDM, and centrifuged at 180,000g (8 h, 4 °C, SW40Ti rotor). After centrifugation, fractions (500 μL) were

carefully taken. For size exclusion chromatography, the supernatant (500 μL aliquots) was loaded onto a Superose 6 10/300 column which was pre-equilibrated in eluting buffer (50 mM Tris, 150 mM NaCl, 1 mM EDTA, 2 mM DTT, 0.5 mM PMSF and 0.01% DDM). The flow rate of the column was 0.1 mL/min and the purification was carried out at 4 °C. 400 μL fractions were collected in 96 deep-well pates. Fractions from both purification methods were tested for activity using the ammonium molybdate assay. The concentration of protein in the fractions was determined using the nanodrop (DeNovix DS-11T) and/or BCA assay and if appropriate the fractions were concentrated by centrifugation using Generon X-Spinner 2.5 Pack (100 kDa cut-off) columns. The purity of the protein within the fractions was analysed by SDS-PAGE gels and negative stain EM. The fractions containing purified in-tact complex were used in inhibitor enzyme assays.

2.3.5. Determination of yeast vacuolar membrane activity

The V-ATPase activity was measured using an ammonium molybdate absorbance assay which measured the release of inorganic phosphate as ATP is hydrolysed. The phosphate is detected by the formation of a phosphomolybdate complex which is formed between phosphate and molybdate at low pH. It is subsequently reduced to molybdenum blue in the presence of ascorbic acid. The resulting absorbance can then be measured using the spectrophotometer. The assays were carried out in a final volume of 40 μL and were based on the procedure described in Chifflet et al.,1988[30]. The membranes/protein (~10-20 μg per well) were suspended in assay Buffer A (50 mM Mes, 50 mM Tris, 5 mM $MgSO_4$) and added to a 96-well microtiter plate. To start the reaction, 5 mM of Mg-ATP was added and the plate was incubated at 30 °C. At various time intervals (0, 10, 20 or 30 mins) the reaction was stopped by the addition of 12% SDS solution. The colour development reagent was prepared by mixing 6% ascorbate (w/v) in 1 M HCl with 1% ammonium molybdate (w/v) in equal volumes. This was added to each well and the colour was allowed to develop for 5 minutes at room temperature. A solution containing 2% sodium citrate (w/v), 2% sodium meta-arsenite (w/v) and 2% acetic acid (v/v) was added and the plate was incubated at 37 °C for 15 minutes. Once cool the absorbance at 600 nm was read on a spectrophotometer (FluoroStar Optima). The amount of phosphate released was determined via comparison to a standard curve of potassium dihydrogen orthophosphate.

2.3.6. Testing the compounds for activity against the vacuolar membranes or purified protein

The compound activity was determined using the ammonium molybdate assay as described above, with a few minor changes. The vacuolar membranes or purified protein were suspended in assay Buffer A with ~10-20 µg of protein used for each well within the 96-well plate. The inhibitors were added at the desired concentration to each well and the plate was incubated at 30 °C for 5 minutes. ATP (5 mM) was added to initiate the reaction to give a final volume of 40 µL. The plate was incubated at 30 °C for 30 minutes before the 12% SDS stop solution was added to all wells. The same procedure as described above was then carried out. For all inhibitor assays, there were controls of no inhibitor and no membranes which measured the ATP background in the assay. The activity of the compounds was determined by calculating the amount of phosphate released in the presence of a compound as the percentage of the amount of phosphate released when there was no compound present.

2.3.7. Determining the toxicity of compounds against HEK cells

HEK-293 cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% foetal bovine serum and 1% penicillin/streptomycin mix at 37 °C, 5% CO_2. The cells were seeded in 12-well plates and once the cells were ~50% confluent, the compounds were added at the desired concentrations. After 48 hours the cells were visually inspected and the observations were recorded.

2.3.8. Determining protein concentration using a BCA assay

The protein concentration in the vacuolar membranes was determined using a bicinchoninic acid (BCA) method which was carried out in 96-well microtiter plates. The protein was precipitated from the vacuolar membranes by adding 9 volumes of a 1:1 mixture of acetone and ethanol and incubating the solution at -20 °C for 30 minutes. The precipitate was collected by centrifugation (13,000 rpm, 5 minutes, Hettich M-24 microcentrifuge) before the acetone/ethanol mixture was removed. The protein was resuspended and solubilised in 2% SDS (w/v) solution. Bovine serum albumin (1 mg/mL) was used to generate a standard curve.

2.3.9. Preparation of V-ATPase negative stain grids

Carbon-coated copper grids were glow discharged for approximately 30 seconds prior to sample application using a Pelco glow discharge unit. 3.0 µL of ~50.0 µg/mL

protein was added, followed by 2 × 3.0 μL of 1% uranyl acetate solution which were blotted away after 60 seconds using filter paper. Images were taken on an FEI F20 microscope fitted with a Ceta camera and was operated at 50,000x magnification.

2.3.10.　　　Preparation of V-ATPase cryo-EM grids

3.0 μL aliquots of purified V-ATPase at a concentration of 2.5-3.5 mg/mL were applied to Quantifoil Cu R1.2/1.3, 300 mesh holey carbon grids, which had been glow discharged for 30 seconds using a Pelco glow discharge unit. An FEI Vitribot Mark IV was used to blot the grids for 6 seconds (blot force 6) at 100% humidity and 4°C before plunging into liquid ethane. The grids were loaded into an FEI Titan Krios transmission electron microscope (University of Leeds Astbury Biostructure Laboratory) operating at 300 kV.

2.3.11.　　　V-ATPase cryo-data collections

There have been four cryo-EM data collections of V-ATPase. All data sets were recorded at the Astbury Biostructure Laboratory on Titan Krios microscopes operating at 300 kV. A summary of the data collection parameters for each data set is shown in Table 2.2.

Table 2.2: A summary of the V-ATPase cryo-EM data collections.

	16-hour data collections		72-hour data collections	
	Apo V-ATPase	*LDC-2 bound V-ATPase*	*Apo V-ATPase*	*LDC-2 bound V-ATPase*
Detector	K2	Falcon III	K2	Falcon III
Detector mode	Counting	Integrating	Counting	Integrating
Voltage (kV)	300	300	300	300
Pixel size (Å)	1.07	1.065	1.07	1.065
Defocus (μm)	-1 to -4	-1 to -3	-1 to -3	-1 to -3
Total dose (e⁻/Å)	65	80	75	74.7
No. of frames	40	59	50	59
Exposure time (s)	10	1.5	10	1.5
Dose per frame	1.63	1.36	1.5	1.27
No. of micrographs	698	1,088	2,944	6,919
Total particle No.	20,847	9,205	77,867	133,059
Final particle No. (Resolution)	1,359 (8.7 Å) 2,098 (8.1 Å)	2,857 (9.1 Å)	3,362 (9.5 Å) 2,827 (10.2 Å) 3,762 (10.2 Å)	6,833 (10.4 Å) 4,738 (11.2 Å) 6,169 (10.4 Å) 7,243 (9.5 Å)

For all data sets, the motion-correction and CTF estimation were carried out on-the-fly using MotionCor2[207] and Gctf[210], respectively. Unless otherwise stated, all processing was carried out using RELION3.0[214]. For the 16-hour data collections, the micrographs were manually picked whereas crYOLO[216] was used to pick particles from the 72-hour data collections. The particles from each data set underwent iterative rounds of 2D and 3D classification. An initial model was built from the best 2D classes in each data set. Models were fit into the maps using MDFF[232] and UCSF chimera[236] was used to analyse the resulting structures. All other processing steps will be described in the text in Chapter 6.

3 The use of cryo-EM to study inhibitor binding to cytochrome bc_1

3.1 Introduction

Parasitic diseases, such as malaria and toxoplasmosis, are devastating to the developing world, with 435,000 deaths in 2017 caused by malaria alone[239]. There are a number of small molecule treatment options for malaria but resistance is emerging to all current treatments therefore there is an urgent need to develop new medicines. Cytochrome bc_1 is a validated anti-malarial drug target with the current treatment atovaquone acting upon this protein[240]. Despite being the subject of extensive X-ray crystallography programs, there is no structural information derived from the parasite protein. Therefore, the interactions between current inhibitors and the parasite protein remain unknown. Moreover, homology modelling of the *Plasmodium falciparum* (*P. falciparum*) enzyme has failed to predict the structure of a key loop in cytochrome *b* which encompasses the inhibitor binding site. This has resulted in SBDD programs being hindered by a lack of reliable parasitic models, resulting in compounds such as GSK932121 being developed which are toxic to the host, as discussed in Section 1.5.3[151]. The parasite protein cannot be over-expressed in traditional hosts so an alternative method of obtaining purified protein is needed which consists of extracting bc_1 from the native source organism. This will result in much smaller quantities of protein being obtained than the mammalian homologue proteins thus an alternative method for structural determination is needed. As cryo-EM typically uses less protein than X-ray crystallography, the technique could play an important role in determining the structure of parasite-derived bc_1. For this part of the project, the aim was to determine whether cryo-EM is an amenable technique to study inhibitor-binding to the bovine bc_1 enzyme.

3.2 Results

3.2.1 Sample preparation and negative stain analysis

Cytochrome bc_1 was purified from bovine heart tissue by Kangsa Amporndanai in the Antonyuk and Hasnain groups at the University of Liverpool. Briefly, the protein was purified from bovine heart mitochondria, solubilised in DDM or LMNG detergents and purified via an anion exchange column followed by size exclusion chromatography. The protein was subsequently concentrated to ~40 mg/mL and was

buffer exchanged before being used for EM. All of the protein samples discussed in this chapter, with the different inhibitors bound, were purified using this approach, kindly delivered to Leeds and used fresh without freezing the purified protein at any stage.

Bovine bc_1 was initially analysed by negative stain EM (as described in Section 2.1.1) to determine the sample purity and heterogeneity (Figure 3.1A). The resulting micrographs showed monodispersed particles with very little contamination in the sample. Subsequently, 101 micrographs were collected on an FEI T12 microscope at a magnification of 50,000x. The micrographs were initially picked in EMAN2[215] and were subsequently processed using Imagic-5[241] with the resulting 2D classes showing the expected shape of protein with the detergent micelle and intracellular regions of the protein clearly visible (Figure 3.1B). However, the classes showed that the protein adopted a strong preferred orientation on the carbon as only side views were seen therefore no 3D reconstructions were attempted. The negative stain results showed that the purity of the sample was very high and could be taken forward to cryo-EM.

A

18 nm

B

18 nm
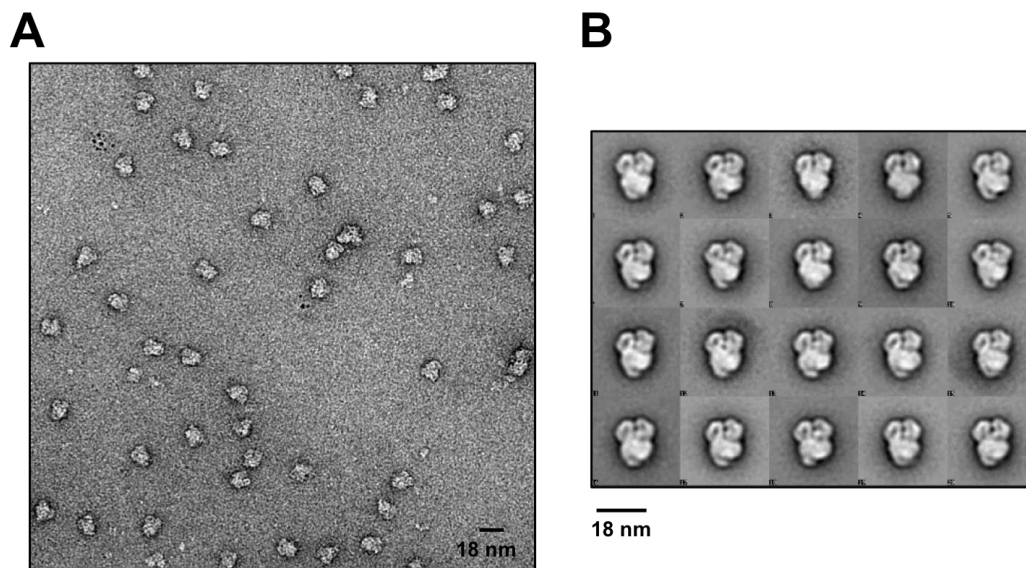
Figure 3.1: Negative stain analysis of bc₁. *A) A micrograph of bc₁ showing a high purity of the sample. B) 2D classes showing the globular shape of bc₁ but only side views of the protein.*

68

3.2.2 Initial cryo-EM data sets

To get an initial idea of the behaviour of bovine bc_1 on cryo-EM grids, the first attempt used lacey grids which contained an irregular distribution of holes. Grids were prepared at a concentration of ~1 mg/mL. The grids were screened on the Titan Krios microscope showing the majority of the particles were bound to the carbon layer therefore there were very few particles suspended in the ice. Furthermore, there appeared to be no preference between the different sizes and shapes of the holes on the lacey grid. To overcome this problem and to achieve a good particle distribution in ice, lacey grids which contained a thin layer of carbon support over the grid were prepared. However, this resulted in the presence of a predominant side view of the protein, even when a proportion of the micrographs were tilted at 20°. The rationale behind the tilted data set was to try and fill in some of the missing components in Fourier space which would improve the quality of the resulting 3D structure. A data set of 2,887 micrographs were collected resulting in a particle stack containing ~500,000 particles. After processing the data, both the 2D classes and the 3D reconstruction indicated that, despite tilting the specimen, there were still problems with preferential views (Figure 3.2A,B). This was highlighted by an angular distribution plot which shows the number of particles going into one particular view of the protein (Figure 3.2C). The resulting 3D reconstruction had a different overall architecture to existing crystal structures showing the map was incorrect. The presence of the radial spikes in the 2D classes also showed that the particles were being over-fit, indicating problems with the quality of the data. Therefore, it was clear that the cryo-grids needed to be optimised to result in grids which did not only contain the additional carbon layer but also had a good distribution of protein particles in a wide range of orientations within the ice.
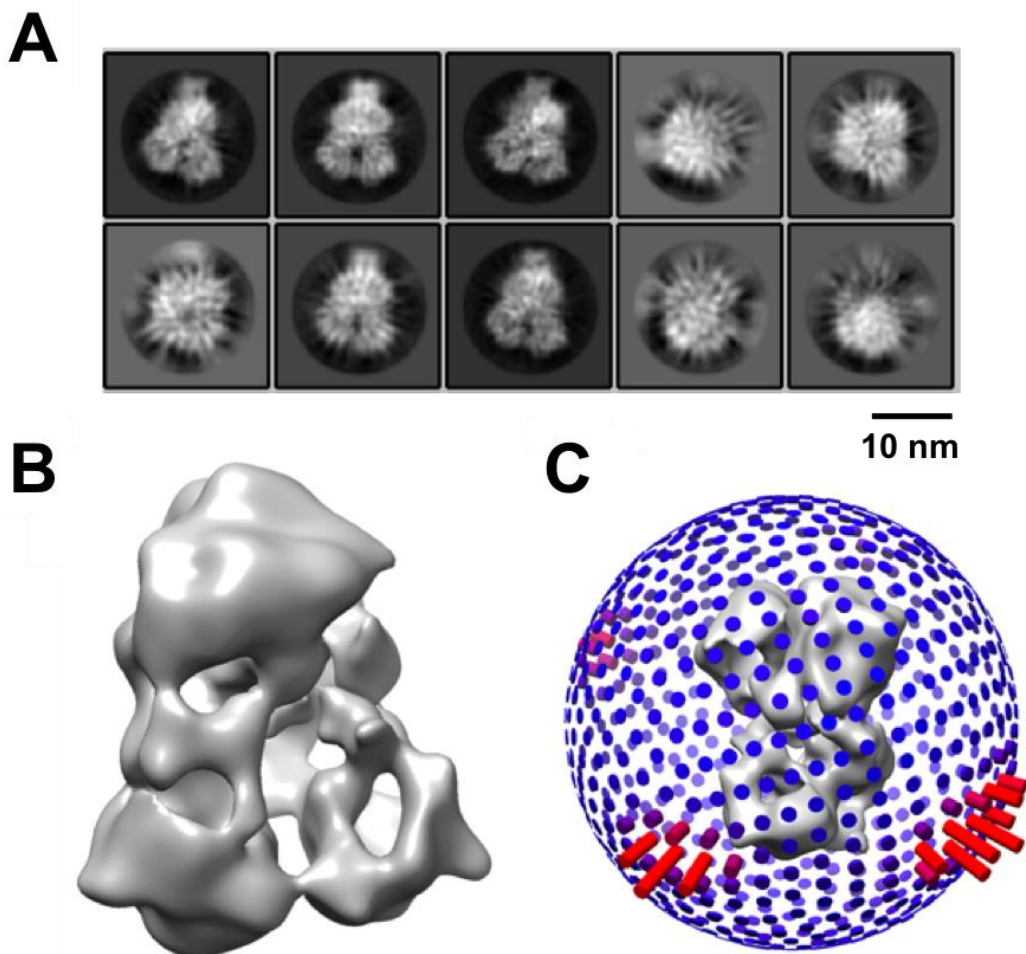
Figure 3.2: Analysis of the initial cryo-EM data set. *A) 2D classes which contain side views which show the correct shape of bc1 and include the detergent micelle. The classes also potentially show other views but these show features consistent with over-fitting. B) An 18 Å 3D reconstruction and C) the angular distribution which suggests there is one predominant view.*
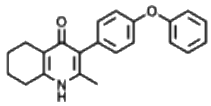
3.2.3 Cryo-grid optimisation

Due to the problems described in Section 3.2.2, the cryo-grid preparation of bc_1 required optimisation which aimed to remove the need for using carbon-backed grids thereby preventing the particle from adopting a favourable orientation on the grid. Initially, bovine bc_1 was applied to quantifoil R2.2 grids which have an even distribution of equally sized holes (2 μm) spaced at equal distance (2 μm) across the grid. It was found that high concentrations of sample (5 mg/mL) were needed to saturate the carbon enabling the protein molecules to become suspended in ice. However, the protein molecules were not evenly distributed throughout the ice but were instead aggregated around the edges of the holes. An example representative

micrograph is shown in Figure 3.3A. Towards the centre of the hole, the protein molecules became more distinguished and could be manually picked and used in 2D classification. As the protein was not distributed evenly within the ice, the data could not be auto-picked due to the neighbouring particle interactions and subsequent crowding. Therefore, the grids required further optimisation. It was also noticed that in the centre of the holes, the ice layer was either extremely thin or not present at all giving rise to a 'halo effect' of particle distribution which could be due to the interactions of the protein or buffer at the air-water interface[182]. This effect could also have played a part in pushing the protein towards the edges of the holes resulting in a poor distribution of protein within the ice.

Despite the non-ideal distribution of protein on the grid, data were collected on a Titan Krios microscope fitted with a Falcon III detector operating in integrating mode resulting in 5,902 micrographs being collected. A summary of the data collection parameters is shown in Methods Section 2.2, Table 2.1. The sample contained bovine bc_1 with an inhibitor, MJM170, bound which was designed to treat toxoplasmosis. The structure and binding affinities of MJM170 is shown in Table 3.1. The binding affinity data was provided by collaborators at the University of Liverpool.

Table 3.1: The structure and binding affinity for the inhibitor MJM170.

	Compound structure	Plasmodium bc_1 (IC_{50})	Bovine bc_1 (% inhibition)	
			0.1 ∝M	1 ∝M
MJM170		10 nM	31.3 ± 7.0	49.7 ± 3.1

After initial motion correction using MotionCorr2[207] and CTF estimation (GCtf)[210] were performed, the data were subsequently processed using RELION2.0[234]. Due to the poor distribution of protein within the ice, autopicking the data was unsuccessful which resulted in 78,306 particles being manually picked. After 2D and 3D classification, 41,223 particles remained which resulted in a map with a global resolution of 4.6 Å after post processing. Local resolution estimations showed the core of the complex, including the Q_i inhibitor binding site, was at a higher resolution (4.4 Å) than the global average. Examination of the Q_i inhibitor binding site showed

that there was density which could be attributed to the inhibitor, which is shown in Figure 3.3D. There is strong density for the pyridone head of the molecule and weak density for the second aromatic ring (Figure 3.3D). This could be due to rotation around the linking carbon-oxygen bond which causes the inhibitor to adopt different poses within the binding site. The side chains around the molecule are not well resolved at 4.4 Å so the resolution of the map would need to be improved to elucidate details on the side chain interactions with the inhibitor. As the particle distribution on the grid was not optimised, resulting in the particles being manually picked, there was a low number of particles going into the classification steps. The resolution of the map could be particle limited therefore the cryo-grids need further optimisation to allow a data set to be collected which contains a large number of particles (100,000's). A summary of the structural details for this data set is shown in Figure 3.3.
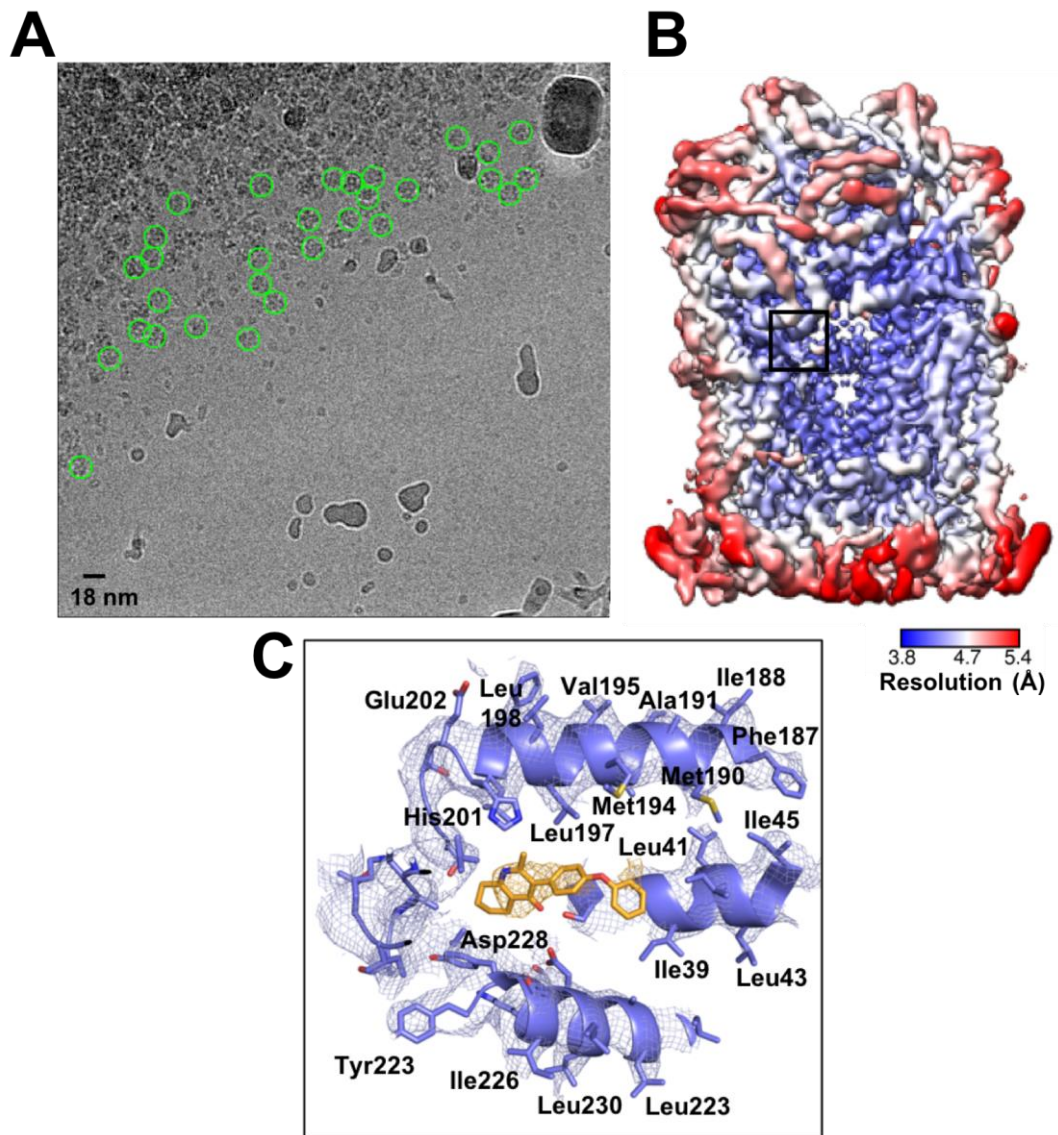
Figure 3.3: Overview of MJM-bc₁ data. *A) A representative micrograph which shows a poor particle distribution as they are clumped towards the edge of the holes. The green circles represent particles which were manually picked. B) A local resolution map of MJM-bc₁ showing the core of the complex is at ~4.2 Å. The black square indicates the position of the Q_i inhibitor binding site. C) The Q_i inhibitor binding site. MJM170 and the corresponding inhibitor density is shown in orange and the protein in purple. There is density for the head of the MJM compound but it is weaker at the tail of the molecule.*

The particle distribution was improved by using a different grid type which contained smaller sized holes; quantifoil R1.2/1.3. The smaller hole, (1.2 µm diameter rather than 2 µm) significantly reduced the amount of protein molecules clumped towards the edges of the holes thus improving the protein distribution (Figure 3.4). It was also noted that in the smaller hole, the 'halo effect' which resulted in the centre

of the hole being very thin or not present at all, was not as prevalent as in the R2.2 grids. Moreover, it was found that micrographs taken in areas of the grid with thicker ice had the better distribution of particles. The improved protein distribution enabled the micrographs to be auto-picked which, for large data sets, would significantly reduce the amount of time spent picking the micrographs whilst also increasing the total particle number. Notably, it was also found that high concentrations of protein (~5 mg/mL) were needed in order to achieve a good distribution of particles within the ice. The quantity of protein used and how it compares to X-ray crystallography is discussed in more detail in the discussion (Section 3.3).

Three more data sets were subsequently collected using these optimised cryo-grid conditions. Firstly, bc_1 with no inhibitor bound (apo) was collected, followed by two inhibitor-bound bc_1 data sets. Figure 3.4 shows a representative micrograph from all three data sets which show a monodisperse distribution of protein molecules within the ice thus allowing autopicking parameters to be established. The quality of the grids was reproducible between all of the data sets which allowed multiple inhibitors to be screened. The two inhibitors were chosen as they contained two different cores (pyridone and quinolone) which would determine whether the different cores influenced the binding pose.
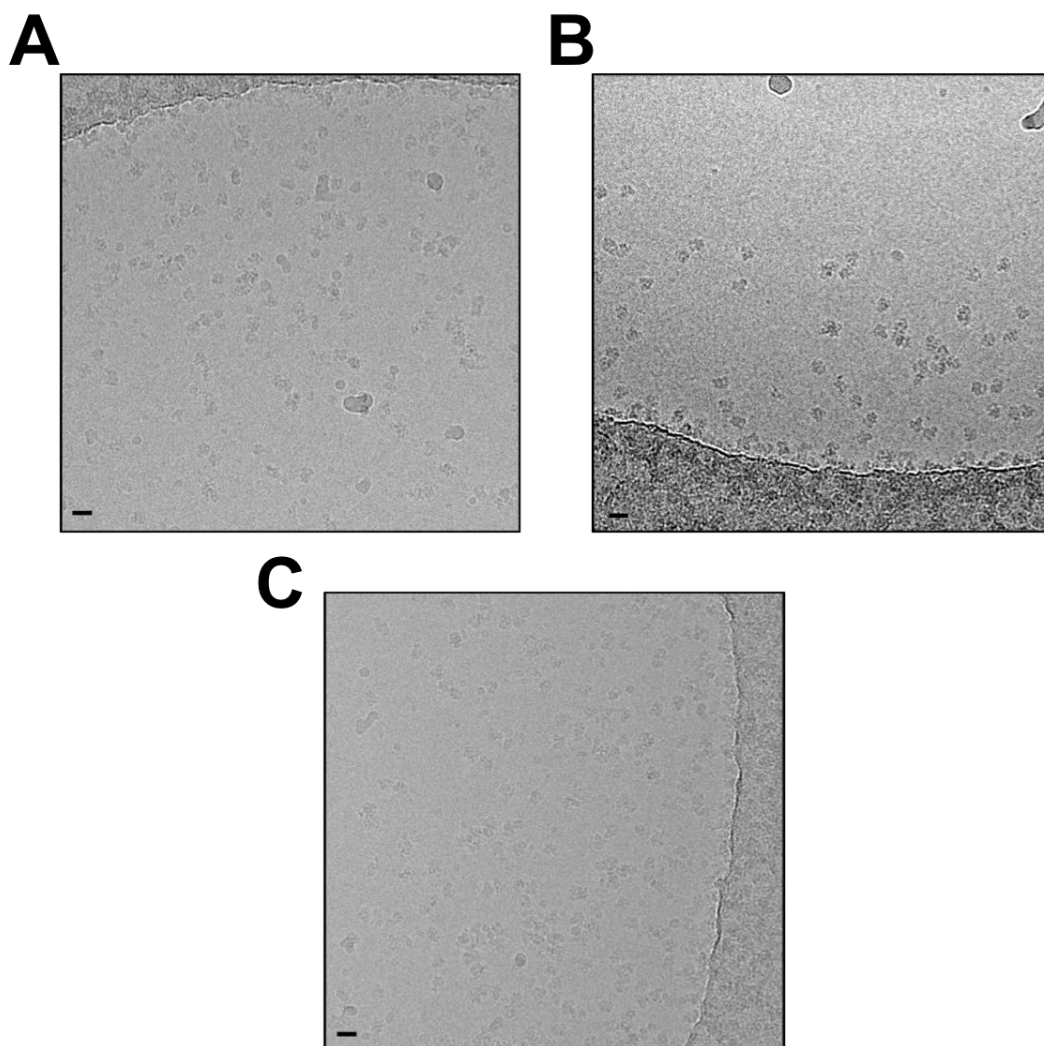
Figure 3.4: Representative cryo-EM micrographs. *The particles within the micrographs for the A) Apo, B) GSK and C) SCR data sets show a monodisperse distribution of particles within the micrographs for each sample. The scale bar is 18 nm.*
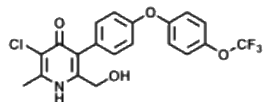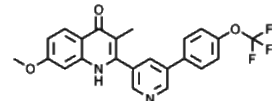
As the cryo-grids were optimised to achieve to a good particle distribution in the ice, it allowed a wider study to be conducted. This consisted of collecting data on bovine *bc₁* bound to three different small molecule inhibitors (GSK932121, SCR0911 and later JAG021) which contained either a pyridone or quinolone core. By having a reliable and reproducible way of making the grids, it allowed the structures to be determined relatively quickly thus enabling the differences in small molecule binding to be probed. For example, after the grids were made and loaded into the microscope a structure was typically obtained within a week of data collection, with additional time being spent on further rounds of classification and polishing steps. The speed of

obtaining the structures is very important for SBDD programs as X-ray crystallography has allowed multiple inhibitors, or even fragments, to be screened per day. Even with the developments to the on-the-fly data processing, the cryo-EM time-frame does not yet rival X-ray crystallography. Therefore, being able to make consistent and reliable cryo-EM grids, which has been achieved for bovine bc_1, is one hurdle to overcome to enable multiple inhibitor-bound structures to be determined in a short timeframe.

3.2.4 Cryo-EM Data collection

Cryo-EM has been used to study the structure of bovine cytochrome bc_1 in complex with two different families of inhibitors which have a pyridone (MJM170 and GSK932121[151]) or quinolone (SCR0911[242]) core, along with the un-liganded apo form. The GSK932121 (GSK) and SCR0911 (SCR) inhibitors were designed to treat malaria and the structure of the compounds and activity against malarial and bovine bc_1 are shown in Table 3.2. Both inhibitors are highly potent against their target plasmodium bc_1 with IC_{50} values of 7 nM and 14 nM, respectively. Experiments and data were performed in the Hasnain and Antonyuk groups at the University of Liverpool.

Table 3.2: The structure and activity of the anti-malarial inhibitors.

	Compound structure	Plasmodium bc_1 (IC_{50})	Bovine bc_1 (% inhibition)	
			0.1 ∝M	1 ∝M
GSK932121		7 nM	64.7 ± 3.8	81.7 ± 4.6
SCR0911		14 nM	9.2 ± 2.3	72.1 ± 4.9

Initially data were collected on purified bc_1 with no inhibitor bound (apo), on an FEI Titan Krios microscope fitted with a Gatan K2 direct electron detector. The rationale behind collecting the apo data set was to determine that there was no natural substrate bound in the active site which would make it difficult to distinguish between substrate or inhibitor molecules. In total, 3,256 micrographs were collected and corrected for drift and beam-induced movement using MotionCor2[235] before CTF estimation was performed using Gctf[210]. After auto-picking in RELION[234], ~260,000

particles underwent iterative rounds of 2D and 3D classification which removed any poorly aligned particles and clear heterogeneity, thus generating a particle set consisting of ~57,000 particles which was refined and post-processed to a global resolution of 4.4 Å. The local resolution calculated in RELION[234] showed the core of the complex, including cytochrome *b*, was at a higher resolution (4.2 Å) than the global average. However, the Rieske domain, which contains the iron-sulphur cluster, was resolved to ~5.8 Å.

Purified bovine *bc$_1$* complexed with two inhibitors (GSK and SCR), were imaged on a FEI Titan Krios microscope at 300 kV fitted with a Falcon III direct electron detector. For both data sets 8,840 (GSK) and 7,893 (SCR) micrographs were collected and processed as described above resulting in both maps having a global resolution of 4.1 Å as calculated by the gold standard Fourier shell correlation (FSC) = 0.143[243]. The data sets have been summarised in Table 3.3 and Figure 3.5 and the full data collection parameters are shown in Methods Table 2.1.

Table 3.3: An overview of the particle number and resolution achieved for the three data sets.

	Apo-*bc$_1$*	GSK-*bc$_1$*	SCR-*bc$_1$*
Detector	Gatan K2	Falcon III	Falcon III
No. of micrographs	3,256	8,840	7,903
Total particles	260,201	466,865	629,258
Particles in final reconstruction	57,571	232,910	114,130
Global resolution	4.4 Å	4.1 Å	4.1 Å

For the GSK-*bc$_1$* data set, ~466,000 particles were auto-picked and after 2D and 3D classification ~232,000 particles made up the final reconstruction. The 3D map for SCR-*bc$_1$* consisted of ~114,000 particles which had been selected by 2D and 3D classification from a total of ~630,000 particles. The resulting 2D classes showed a good distribution of views and structural detail, consistent with a high-quality data set. Further rounds of 2D and 3D classification were attempted but the map was not improved. Moreover, attempts were made to improve the resolution further using the

particle polishing algorithms implemented in RELION2.1. However, these were unsuccessful and both the resulting map and resolution were identical. The previously solved crystal structure for bc_1[151] (pdb 4D6U) was rigidly fit within the cryo-EM maps using Chimera[236] before dynamic flexible fitting in MDFF[232] to account for significant changes in secondary structure position between the EM and crystallography derived model. The models then underwent relaxation using Rosetta[244] before being refined using PHENIX[230]. Finally, the models were examined using COOT to manually inspect the residues and how they fit into the density, with incorrect rotamers being corrected for.
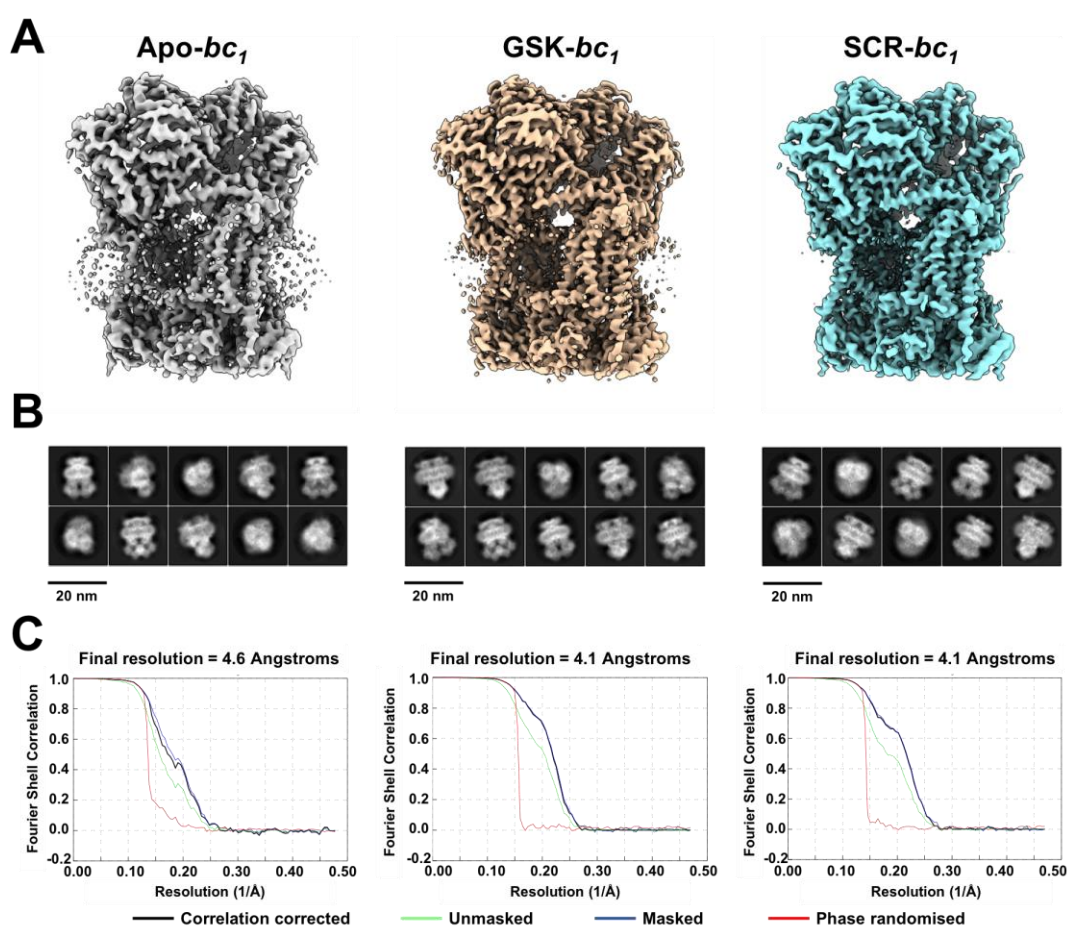
Figure 3.5: Summary of the cryo-EM data collections. *A) The cryo-EM maps for each of the data sets showing the overall structure of the protein is consistent in each map. B) Example 2D classes for each data set as generated by RELION which shows a wide range of orientations within the ice. C) The FSC curves for each data set to which were generated using the 0.143 gold-standard.*

3.2.5 Cryo-EM structure analysis

Cryo-EM data sets of cytochrome bc_1 bound without any inhibitor (apo) and with two inhibitors (GSK and SCR) were collected which achieved resolutions of 4.4 Å and 4.1 Å, respectively. After the maps were generated as described above, an existing crystal structure (pdb 4D6U) was rigidly fit into the density maps so the cryo-EM structure could be analysed. For all three cryo-EM maps of bovine bc_1, the overall architecture of the fitted structures was, as expected, consistent to published crystal structures. In the apo-bc_1 map the resolution of 4.6 Å, only allowed for the α-helices within the complex to be modelled into the density as the resolution was not sufficient to allow accurate side chain placement. In the GSK-bc_1 and SCR-bc_1 maps, the increased resolution allowed both the α-helices and β-sheets to be modelled into the density. Furthermore, in the best resolved regions in the map, the density allowed the larger, aliphatic/aromatic side chains to be fitted. This is exemplified in Figure 3.6 where the side chain density in the apo (grey) structure is not as strong in the α-helices as it is for the GSK and SCR inhibitor bound maps (gold and cyan, respectively) at a consistent sigma level (4σ).
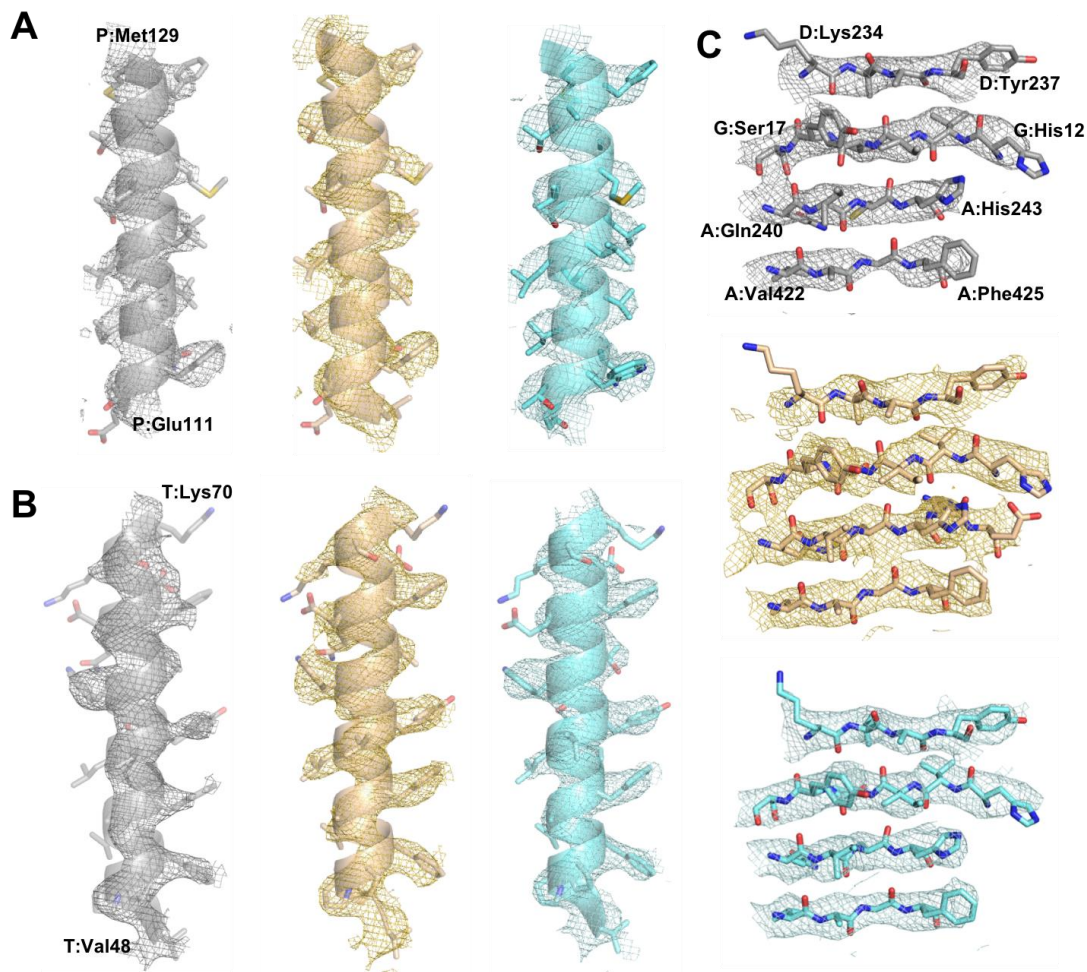
Figure 3.6: Representative cryo-EM density. *A comparison of the quality of each map (coloured grey, gold and cyan for the apo, GSK bound and SCR bound structures, respectively) with two representative transmembrane helices (A, B) and a β-strand within the soluble domain (C). The density was contoured at 4σ with the side chains better resolved in the two inhibitor bound structures.*

Within cryo-EM maps, a range of resolutions can be obtained within one structure. For instance, the core of proteins can be determined to a higher resolution than regions which are inherently flexible. Therefore, local resolution maps for all three data sets were calculated using RELION which are shown in Figure 3.7. In all three maps, the poorest resolved region occurred at the Rieske domain. For instance, in the GSK and SCR bound maps this domain was resolved to 5.1 Å and 5.3 Å respectively compared to the global resolution of 4.1 Å. The iron-sulphur cluster within this region is responsible for transferring electrons to cytochrome *c* in the electron transport chain. Previously it had been reported that this domain is mobile and can exist in different conformations[245,246]. The other heme groups (b_H at the Q_i site and b_L at the Q_o site) are the strongest features within the cryo-EM maps in all three

structures. However, the inability to resolve the iron–sulfur cluster, which should also be a strong feature, suggests that it is mobile and can adopt different positions, thereby lowering the resolution.
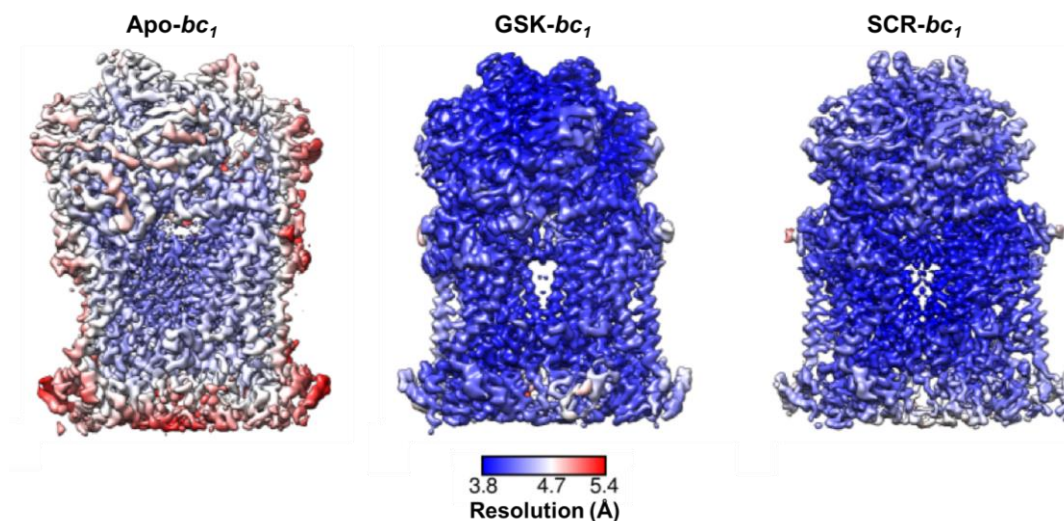
Figure 3.7: The local resolution for the three maps. *Local resolution maps coloured to the same scale apo-bc₁, GSK-bc₁ and SCR-bc₁ which show the core of the complex is at the highest resolution and the Rieske protein is the poorest resolved feature in the map.*

Attempts were made to trap the iron-sulphur cluster into different conformations using soft masks around this region however, these classifications and subsequent refinements were unsuccessful in trapping the complex into distinct states. These classifications were performed using RELION2.1 which did not contain the multibody refinement features implemented in the latest version of the software (RELION3.0). For a different cytochrome bc_1 data set, the multibody refinement was carried out and is discussed in more detail later in Section 3.2.8. The 3D classifications were conducted using both C1 and C2 symmetry to determine whether the two Rieske domains move independently during the catalytic cycle. Analysis of the maps showed that there was no evidence for this and the overall resolution of the map did not improve. This could suggest that rather than being in distinct conformations the Rieske protein is instead mobile. The flexibility of the Rieske domain is discussed in more detail in Section 3.2.7.

It was interesting to note that the final resolution of both inhibitor-bound structures was 4.1 Å. To assess whether this was a feature of particle number, the inhibitor data

sets were combined and selective masking of the individual monomer subunits was carried out. The total number of particles which went into the final 3D reconstructions for the three inhibitor data sets, including the particles from MJM-bc_1, were pooled resulting in ~388,000 particles being refined to 4.6 Å. After numerous 3D classifications to reduce the heterogeneity of the particle stack, the highest resolution map obtained after post-processing was 4.4 Å which consisted of 68,584 particles. Moreover, a soft mask which encompassed the cytochrome b subunit, and tightly covered the inhibitor binding site, was created and 3D classifications were conducted at this region. Examination of the resulting 3D maps showed that the classification was unsuccessful in separating out the different inhibitor bound complexes. The resolution did not improve with the increased particle number, suggesting that the final resolution may not be particle number limited but instead the isolated bc_1 complex may be flexible and contain low levels of heterogeneity which is not permitting the resolution to go below 4.0 Å. Additionally, it is worth noting that the GSK and SCR inhibitor-bound structures were solubilised in different detergents (DDM and LMNG respectively) which could be another reason as to why the resolution did not improve when all the particles were combined.

To determine whether the imposed two-fold symmetry was having an effect, the particle stack for the SCR-bc_1 was re-refined with no symmetry imposed to a resolution of 4.3 Å. Analysis of the map showed that there were no differences between the monomers in the map and that there were inhibitors bound at both Q_i sites. Moreover, a 3D mask of a monomer subunit was created and the results from the 3D classifications and refinements also highlighted that there were no differences detected at this resolution in the structures of the two monomers in the bc_1 complex suggesting that this is not affecting the resolution of the complex.

3.2.6 Q_i site analysis

After the cryo-EM maps had been generated and the models had been fit into the density, the Q_i site could be analysed to determine whether any density for the inhibitors was present. Analysis of the local resolution of the GSK and SCR inhibitor-bound maps showed the majority of the bc_1 core, was resolved to a higher resolution than the global average of 4.1 Å for both maps. There were slight differences in the local resolution of the catalytic cytochrome b protein in the two models. For instance, in the SCR-bc_1 map the core of the complex, containing the Q_i binding site, was resolved to ~3.8 Å whereas in the GSK bound map the Q_i site was at ~4.0 Å. The

differences in the resolution could be due to different detergents, DDM for GSK and LMNG for SCR, being used to solubilise the complex which could have an effect on the stability of the transmembrane domain. The Q_i site in all three maps was analysed to determine whether inhibitor density could be visualised (Figure 3.8).
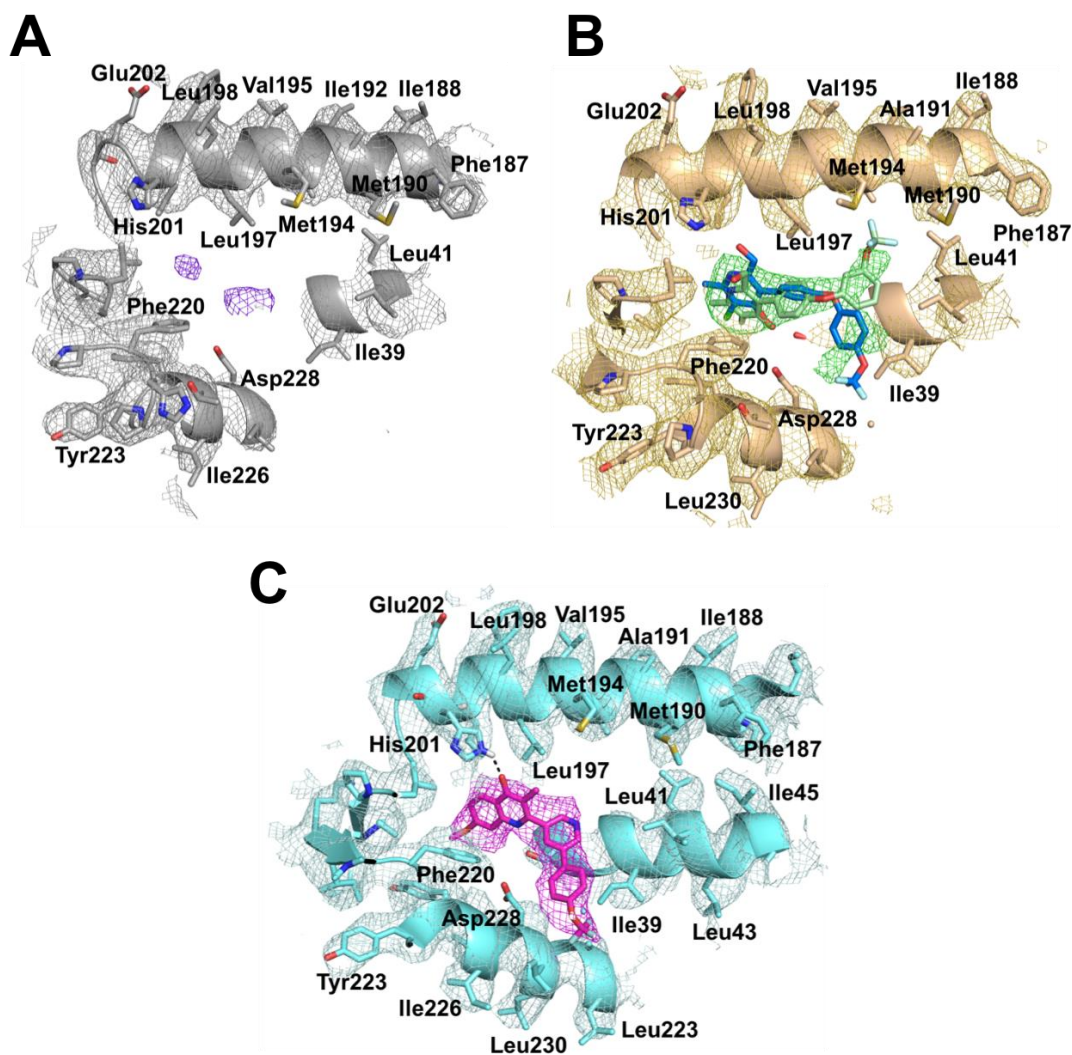
Figure 3.8: Analysis of the Q_i site in the three different cryo-EM maps*. A) The Q_i site in the apo-bc$_1$ cryo-EM map shows minimal density (purple mesh) which does not correspond to any side chains or the heme B_L group suggesting noise within the map or that natural substrate ubiquinone is bound. B) The Q_i site in the cryo-EM map of GSK-bc$_1$. The inhibitor density (coloured green) suggests there are two modes of inhibitor binding, accommodated by rotation around the oxygen bond. The binding pose shown in green agrees with the crystal structure with the trifluoromethoxy group pointing towards Met194. There is additional density which suggests the trifluoromethoxy phenyl group could be rotated and point toward Asp228 revealing an additional mode of binding (shown in blue). C) The Q_i site in the cryo-EM map of SCR-bc$_1$, with the inhibitor shown in pink, being located in strong density. The inhibitor is expected to make a H-bond contact to His201 and strongly fits the density. For all maps the density is contoured at 3σ.*

In the apo structure there is weak density at the Q_i site which could be a result of the natural substrate, ubiquinone, being bound as the complex was purified or just due to noise in the map in this region (Figure 3.8A). Due to the lower resolution of the structure, the side chains for the neighbouring amino acid residues could not be unambiguously modelled into the density. For the GSK and SCR inhibitor-bound structures, there was clear inhibitor density at the Q_i site which does not appear at the Q_o site indicating that these are selective Q_i site inhibitors. The GSK Q_i site is shown in Figure 3.8B and the SCR Q_i site is shown in Figure 3.8C. At 3σ contour level, the strongest inhibitor density occurred in the SCR-bc_1 map where the density for the compound was equivalent to the neighbouring side chains. The density suggests the ligand can adopt an unambiguous conformation which points out towards the surface of the protein with a potential H-bond to Ser35. The quinolone head of SCR was placed between residues His201 and Phe220, and the biaryl tail further extends into the hydrophobic region of Ile39 and Ala232.

The density in the Q_i site for the GSK-bc_1 map was strong for the pyridone head group of the inhibitor at 3σ contour level but was weaker at the tail of the molecule. The density also suggested that there could be two different binding poses of the compound caused by rotation around the oxygen–carbon bond (Figure 3.8B). The different binding poses are discussed in more detail in Section 3.2.7. The Q_i sites of the two inhibitor-bound cryo-EM structures have been compared, showing a strong agreement in the secondary structure of the protein. At 4.1 Å resolution it is difficult to detect subtle changes in the positions of the amino-acid side chains which would result from inhibitor binding. However, gross changes in the side-chain position can be detected; for example, His201 shifts in position between the two maps (Figure 3.9). For SCR0911, His201 is well defined and is in a position consistent with hydrogen-bond formation with the inhibitor, whereas the density for GSK932121 suggests that there is no hydrogen-bond interaction with His201 and the density is more poorly defined. Comparisons of the Q_i site of apo bc_1 and inhibitor bound bovine bc_1 in the cryo-EM structures show that there is no difference in the position of the α-helices which surround the active site, consistent with the observations from crystal studies that show no gross structural changes accompany inhibitor binding.
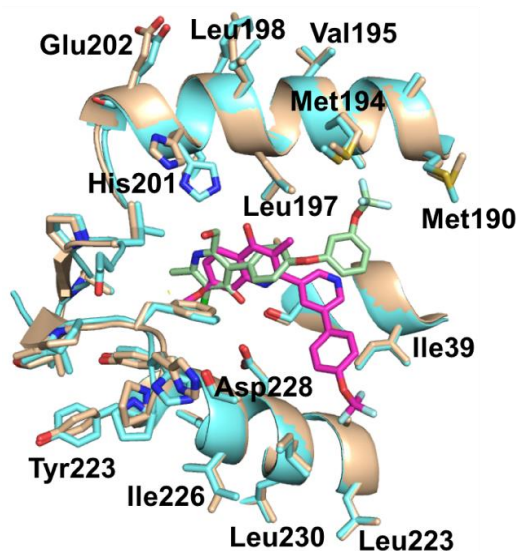
Figure 3.9: Comparing the Q_i site in the inhibitor-bound cryo-EM maps. *The EM structures of SCR-bc$_1$ (cyan) and GSK-bc$_1$ (gold) overlaid. The SCR inhibitor is coloured in green and the GSK inhibitor in magenta. Overlay of the two structures show there is no difference in the secondary structure positions in the two maps and only minor differences in the positions of the amino acid residues. The biggest difference occurs in His201 and that could form a H-bond to the SCR inhibitor but not the GSK compound.*

3.2.7 Crystal structure comparisons

In parallel to the cryo-EM work being carried out, Kangsa Amporndanai in the Antonyuk and Hasnain groups at the University of Liverpool also solved the structure of SCR0911-*bc$_1$* to 3.0 Å by X-ray crystallography. There is also an existing crystal structure of cytochrome *bc$_1$* with the GSK compound bound which was resolved to 4.1 Å[151]. Cytochrome *bc$_1$* has not previously been studied using cryo-EM. Therefore, to determine whether the absence of the crystal lattice has had an influence on the structure of the protein, the cryo-EM and X-ray derived models have been compared. In both the X-ray crystallography and cryo-EM structures compared here, Subunit 11 was not resolved within the maps. It is thought that the subunit was lost during protein purification. In existing crystal structures which contain the subunit, the purification protocols involved using an ammonium sulphate precipitation compared to the ion exchange and gel filtration approach used in this study. Unsurprisingly, the EM structures display a high agreement in the overall architecture of the protein in comparison to the X-ray structures. This is exemplified when comparing the two SCR bound structures which have similar Cα r.m.s.d. values (Figure 3.10A). There is very high agreement between the core proteins and transmembrane domains. However,

there are differences in the mobile Rieske protein in the soluble region, where the $C\alpha$ r.m.s.d. value is greater than 2 Å (Figure 3.10B,C).
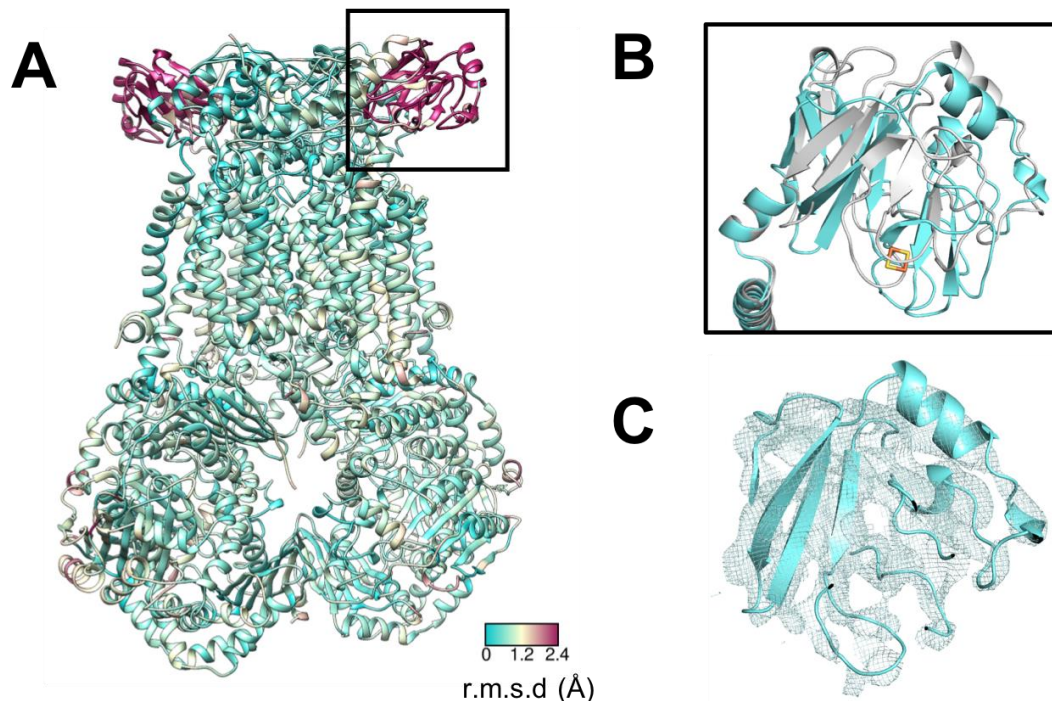
Figure 3.10: Global comparison between EM and X-ray derived SCR-bound structures. *A) The EM and X-ray SCR-bc1 structures have been overlaid and the crystal structure was coloured according to the calculated $C\alpha$ r.m.s.d. value (low coloured cyan and high maroon). The main difference (r.m.s.d. >2 Å) occurs at the Rieske domain (black square) which is known to be mobile during the catalytic cycle. B) An overlay of the Rieske domains in the EM (cyan) and X-ray (grey) derived SCR-bc1 structures. C) The EM derived map for the SCR-bc1 structure with the Rieske domain fitted shown in the same orientation as in (A) and (B).*

The lower resolution of the Rieske domain is consistent to the literature where there are crystal structures which show the Rieske protein is captured in different states in the apo, natural substrate bound and Q_o-inhibitor bound forms[145,245–247]. This is due to the Rieske domain being involved in the transfer of electrons from the Q_o site heme group in cytochrome *b* to the cytochrome *c_1* heme. One reason for the difference in positions of the Reiske domain between the EM and crystal structures might be due to the formation of crystal contacts which are made as the protein crystallises. In the cryo-EM structures, the crystal lattice has been removed so there is no influence on the position of this domain from the formation of protein crystals.

Therefore, the low resolution of this region in the EM structures indicate that the domain is flexible and does not adopt a defined state(s). The ability to better resolve this domain by crystallographic, means could therefore be explained by the crystal packing. Examination of the crystal contacts in COOT show that there are many neighbouring protein molecules surrounding the Rieske protein, which lock the domain into a particular conformation, thus improving the crystallographic density in this region (Figure 3.11). Whereas, in the EM structure, in the absence of a crystal lattice, the Rieske protein is not locked into a particular conformation and can reside in a more native conformation, resulting in the differences between the two models.
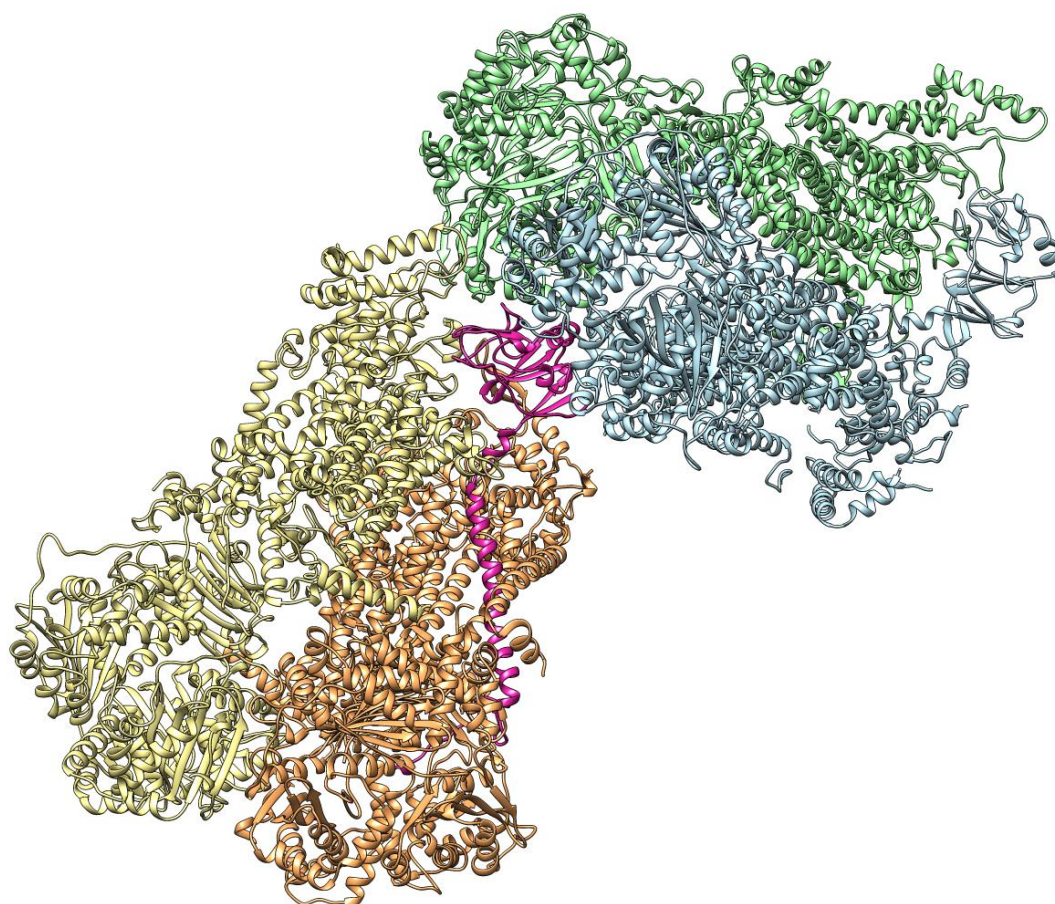
Figure 3.11: Examination of the crystal contacts. *The monomer of the crystal structure containing the SCR compound is shown in orange. The monomers coloured in blue, yellow and green represent the crystal contacts formed in the lattice. The Rieske domain is coloured in pink and shows that it is surrounded by crystal contacts thus explaining the differences between the EM and X-ray structures.*

Analysis of the Q_i site in the GSK-bc_1 cryo-EM map suggests that the inhibitor can bind in two different conformations (Figure 3.12A). Both binding modes have the pyridone head placed between heme b_H and His201 and have the diaryl tail group extending out of the hydrophobic channel away from the heme group. The difference in the binding conformations occurs at the 'tail' 4-trifluoromethoxyphenyl ring of the molecule, as there is rotation around the C-O bond which results in two different binding poses. Pyridone binding pose one has the trifluoromethoxy group from the second aromatic ring pointing towards Met194. The cryo-EM structure has also identified a second pyridone binding pose, which shows the aromatic ring faces the opposite direction with the trifluoromethoxy group pointing towards Asp228. In comparison, in the X-ray crystal structure only one ligand binding pose is identified. An overlay of the EM and X-ray Q_i sites for the GSK-bc_1 structure is shown in Figure 3.12B. The weak density for residue His201 at 3σ contour level in the map could be accounted for by the residue changing position depending on the binding pose of the inhibitor. Classifications around the inhibitor binding site using soft masks to distinguish the occupancy and binding pose of the ligands were attempted but were ultimately unsuccessful. This could be due to the size of the region which was being classified being too small to pull out different inhibitor binding conformations.

Analysis of the Q_i site for SCR-bc_1 shows that the ligand adopts a very similar position in the cryo-EM and X-ray derived models (Figure 3.12B). In both structures, it is proposed that the ligand can form hydrogen bonds to His201 and Ser35 on the neighbouring helix. The Q_i site amino acids adopt similar positions in both structures, with an average Cα r.m.s.d. value of 1.5 Å where there is a clear overlap of the α-helices. There are minor differences in the positions of the amino-acid side chains, such as Phe220; however, this could be owing to the difference in the resolutions of the two maps (4.1 Å for the EM map and 3.0 Å for the X-ray structure).
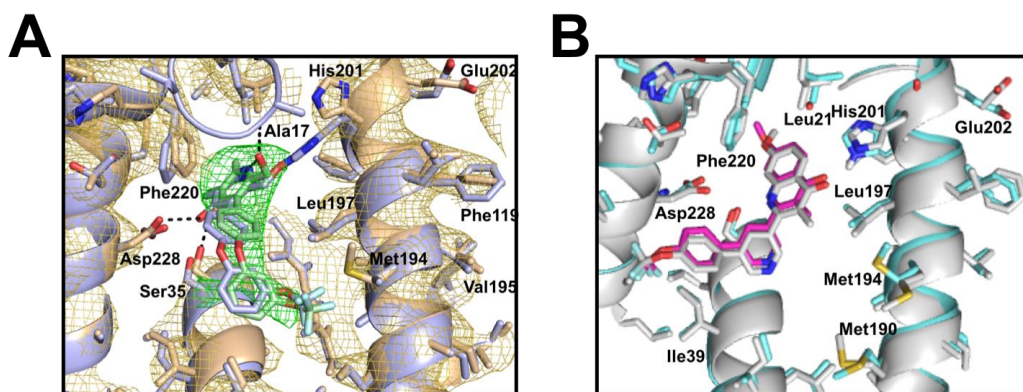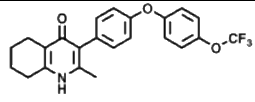
Figure 3.12: The inhibitor-bound Q_i sites derived from X-ray and EM structures. *A) An overlay of the GSK Q$_i$ site in the EM (gold) and X-ray (purple) models. The density in gold corresponds to the EM map that shows a difference in the position of His201 compared to the X-ray structure. B) An overlay of the SCR EM (cyan) and X-ray (grey) Q$_i$ sites. There are no significant differences between the two models in either side chain or secondary structure position.*

Overall, the direct comparisons between the EM and X-ray derived models for GSK and SCR-bound bc_1 have highlighted that there are few differences in the structures obtained using the two different techniques. This is particularly important when analysing the Q$_i$ inhibitor-binding site which shows how the compound is interacting with the protein as this is the information which is utilised in SBDD programs.

3.2.8 Determining the structure of bc_1 bound to JAG021

A final bovine bc_1 data set was collected as a separate study to SCR and GSK-bound bc_1 which have been discussed previously. For SCR and GSK-bound bc_1, the inhibitors were designed to treat malaria. For this new study, an anti-toxoplasmosis inhibitor, JAG021, which was synthesised by James Gordon in the Fishwick group at the University of Leeds, was bound. The compound and binding affinities to *P. falciparum* and bovine bc_1 are given in Table 3.4.

Table 3.4: The structure and binding affinities of JAG021.

	Compound structure	Plasmodium bc_1 (IC_{50})	Bovine bc_1 (% inhibition)	
			0.1 ∝M	1 ∝M
JAG021		40 nM	36.0 ± 4.6	62.8 ± 4.8

Grids were prepared using the same conditions (5 mg/mL concentration applied to Quantifoil R1.2/1.3 grids) as described in Section 3.2.3, which once again gave a monodisperse distribution of particles within the ice (Figure 3.13A). Similar to the previous structures, the data were collected on a Titan Krios microscope fitted with a Falcon III direct electron detector operating in integrating mode. A total of 5,356 micrographs were collected during the 72-hour data collection. Full details of the microscope parameters are shown in Table 2.1. However, the grid used for data collection contained large amounts of crystalline ice so the Fourier transforms for each micrograph were examined and the ones which contained a large feature corresponding to crystalline ice were manually removed (Figure 3.13B).

A

B
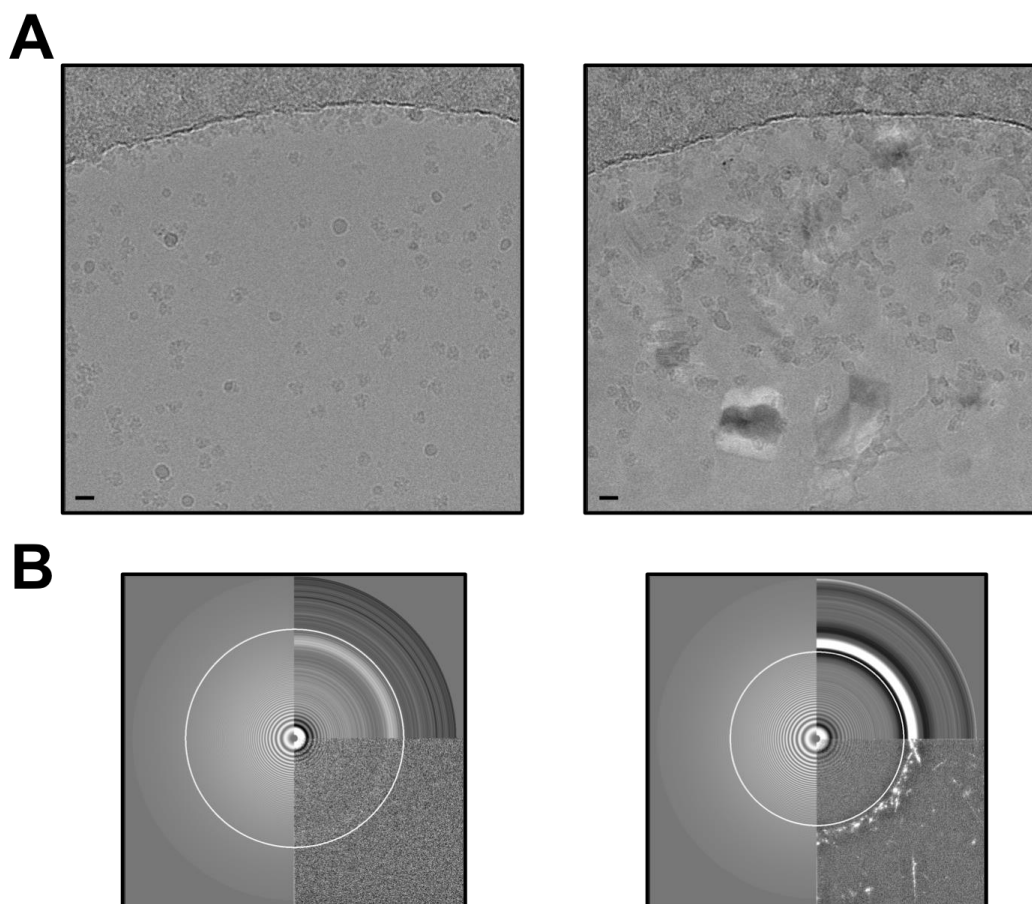
Figure 3.13: Quality of JAG-bc₁ micrographs. A) Example micrographs showing vitreous ice with a good particle distribution and a micrograph which contains crystalline ice. The particles are still visible yet the ice quality is poor. Scale bars are 20 nm. B) The corresponding Fourier Transforms with and without the crystalline ice feature. All of the Fourier transforms were inspected and those which contained this feature were removed.

The remaining 2,960 micrographs were processed in RELION2.1 using the same approach as described for the other data sets in Section 3.2.4. In total, 439,009 particles were autopicked and after classification 211,916 particles remained which were refined to a global resolution of 3.8 Å. By implementing the particle polishing approaches in RELION2.1 the resolution of the map was improved to 3.7 Å. RELION3.0 was subsequently released which introduced new per-particle CTF-refinement and polishing algorithms[214]. The particle stack which resulted in the highest resolution map was re-processed in RELION3.0 to take advantage of these features. By conducting three iterative rounds of per-particle CTF refinement and particle polishing, including per-particle astigmatism and beam tilt estimation, the global resolution of the map was improved to 3.3 Å after postprocessing.

Unsurprisingly, the overall architecture of the protein was consistent to what has been described previously. The density is consistent with the improved resolution of the map. As expected, the α-helices and β-sheets were well resolved and there was clear side chain density. A summary of the data set is shown in Figure 3.14.
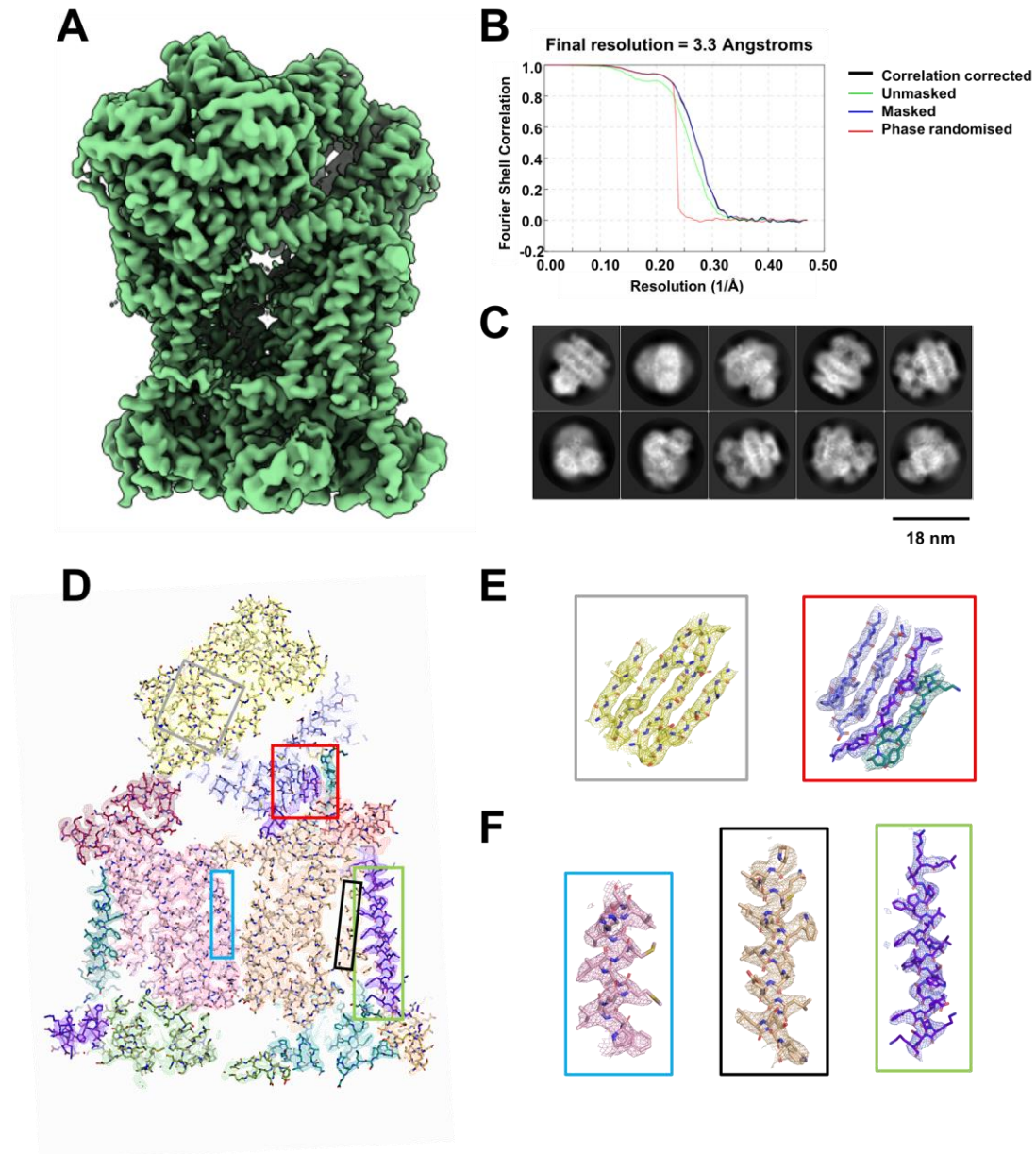
Figure 3.14: An overview of the JAG-bc₁ data set. A) The cryo-EM map of JAG-bc₁ which was resolved to 3.3 Å and clearly showed all of the secondary structure. B) The FSC curve for the data which was calculated using the 0.143 gold standard. C) 2D classes showing a wide variety of orientations within the ice. D-F) Example density for the structure showing E) β-sheets strands are separated and F) side chains are well resolved in the α-helices.

Analysis of the local resolution maps from both versions of RELION showed that the core of the complex was improved from 3.5 Å to 3.1 Å in RELION3.0 which highlights the improved quality of map (Figure 3.15). The Rieske domain was once again the most poorly resolved region of the map at ~ 5 Å.
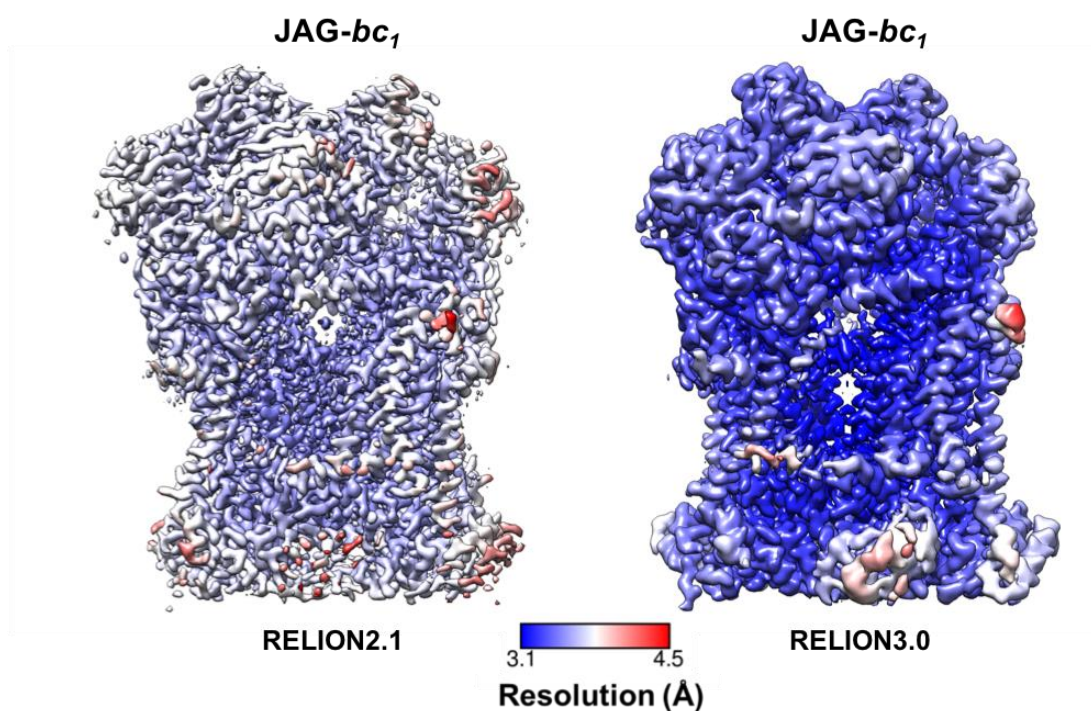
Figure 3.15: JAG-bc₁ local resolution comparison. *The local resolution maps have been calculated in RELION 2.1 and RELION 3.0. The core of the complex was improved to 3.1 Å from 3.5 Å using RELION 2.1.*

Collaborators at the University of Liverpool in the Antonyuk and Hasnain groups, also obtained a crystal structure of bovine bc_1 with JAG021 bound to 3.5 Å resolution. For the first time, the global average of the cryo-EM structure was higher than what has been achieved by using the traditional X-ray crystallography approaches. Overlaying the cryo-EM and X-ray models, showed that there were few differences in the Cα position in the majority of the structures with a Cα r.m.s.d. <2 Å. (Figure 3.16). Like before, the only significant changes appeared in the Rieske domain which is known to be flexible, whereas the core of the complex showed little variation.
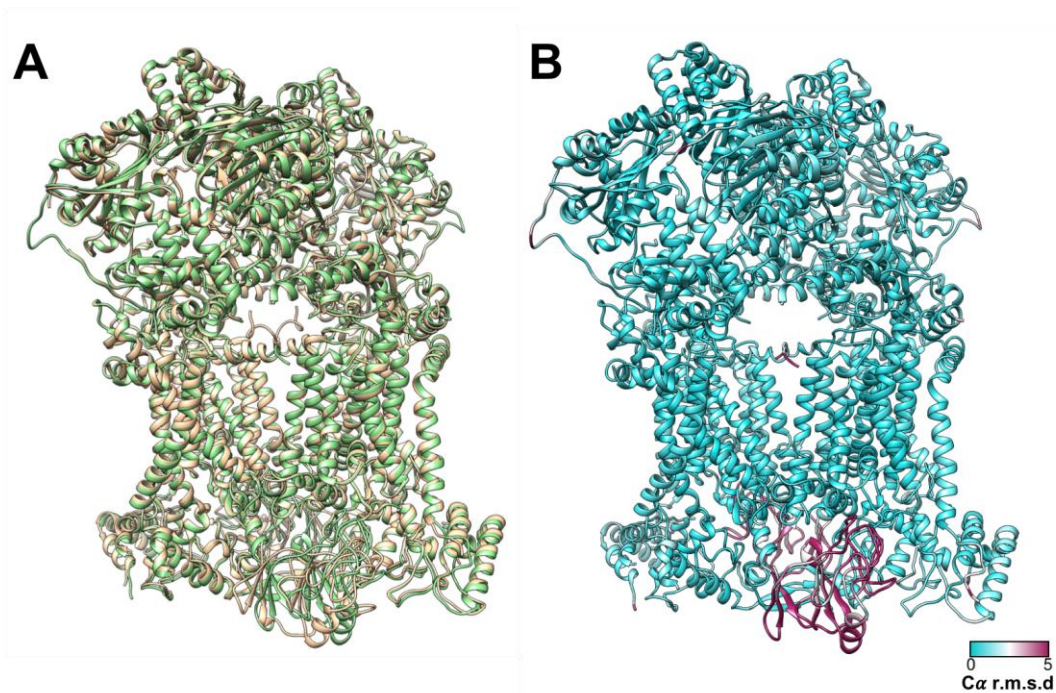
Figure 3.16: Cryo-EM and X-ray structures for JAG-bc₁. *A) An overlay of the EM (green) and X-ray (gold) structures showing a strong overlay in the overall structure of the protein. B) The EM structure coloured by Cα r.m.s.d. with the Rieske domain in Maroon representing the only large differences.*

Due to the multi-body refinement feature being implemented in RELION3.0, a mask was created around both Rieske domains in order to determine whether the domain could be trapped into different conformations[248]. The motion of the Rieske domain relative to the main body of *bc₁* was probed. The results confirmed that the Rieske domain is mobile and the two domains appeared to move asymmetrically to one another. An overview of this motion is shown in Figure 3.17 and shows that the domain can move from side-to-side. However, when the maps obtained were used as references for 3D classification with C1 symmetry imposed, the particles could not be separated into the different states. This could be due to the size of the Rieske domain (34 kDa) and the resulting movement not being large enough for the classification to align the particles into the different states.
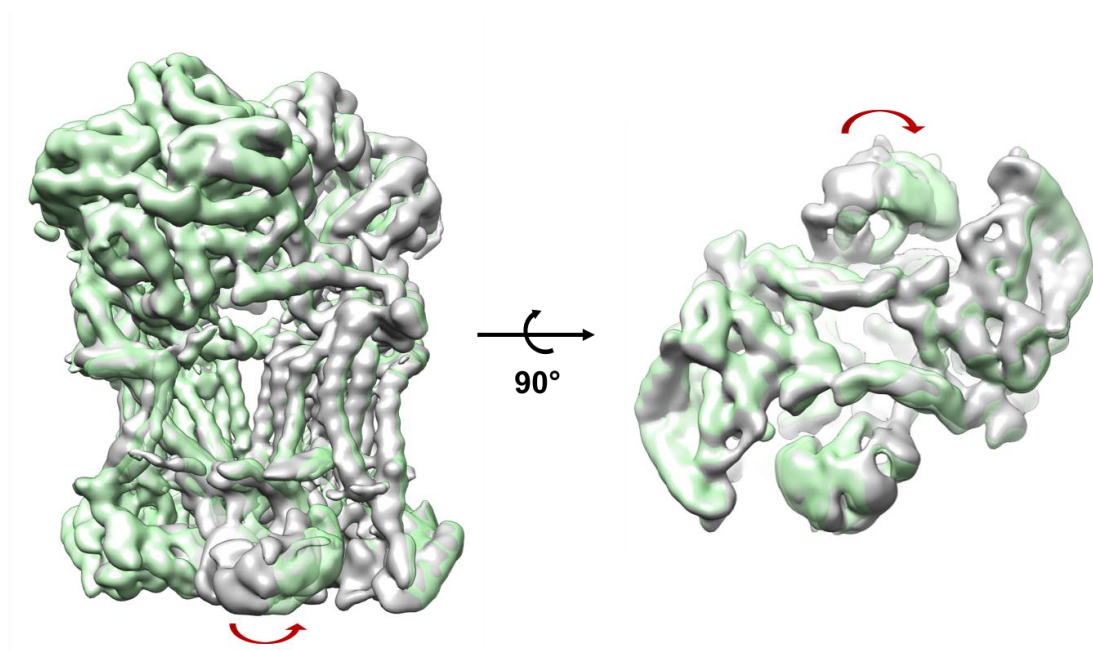
Figure 3.17: Probing the movement of the Rieske domain. *An example of one of the components produced from multi-body refinement. The first and last component is shown in grey and green respectively with the movement highlighted by the red arrow.*

The Q_i inhibitor binding site has also been analysed showing clear density for JAG021 (Figure 3.18A). At the same contour level (3σ), there is strong density for the neighbouring side chain atoms. Similar to the GSK-bc_1 map, the density for JAG021 suggests that the molecule could potentially adopt a second binding conformation with rotation around the C-O bond. Although this feature is not as prevalent as in the GSK-bc_1 structure, the density is strongest for the trifluoromethoxy group to be orientated towards Met194. As the extra density is not as strong as in the GSK-bc_1 map, it could be a result of noise being present in the inhibitor binding site and not due to a second conformation of the inhibitor. The inhibitor position within the density suggests that it could form H-bond contacts from the quinolone head group to His201, Asp228 and Ser35. The density for His201 is not as strong as neighbouring amino acids which suggests that this residue could change position depending on which binding conformation the inhibitor adopts. At the equivalent sigma levels (3σ), the density for the pyridone head group in the JAG-bc_1 map was not as strong as in the GSK-bc_1 map. This could be due to the compound showing higher selectivity to the parasite protein and subsequently binding more-weakly to the bovine protein. This is exemplified by JAG021 being less active than GSK932121 at both 0.1 μM and 1 μM concentrations when the compounds were tested for activity against the bovine

enzyme. This activity data was provided again by Kangsa Amporndanai in the Antonyuk and Hasnain groups at the University of Liverpool. For instance, at 0.1 μM, JAG021 inhibits the activity of the enzyme by 36% compared to 64% for the GSK compound (Table 3.2 and 3.4). As the GSK compound shows higher levels of activity against the bovine enzyme, it suggests that it binds to the protein with higher affinity therefore the occupancy of the ligand could be higher than the JAG compound. This could explain why the density for the JAG compound is not as strong as the GSK compound when the maps are contoured to the same sigma level, despite the map being at a greater resolution.

The Q_i site of the JAG021-bound bc_1 crystal structure shows that the inhibitor only adopts one binding pose, with the trifluoromethoxy group pointing towards Met194 (Figure 3.18B). Comparing the cryo-EM and X-ray derived models highlight that there are no changes in the secondary structure between the two models, although there are minor differences in the side chain positions for a few residues such as Met194 and Phe187.
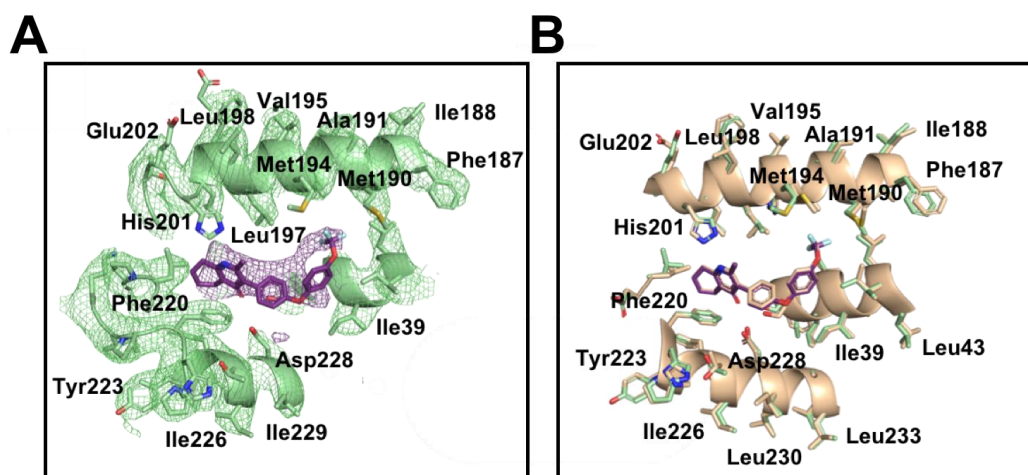
Figure 3.18: The JAG-bc₁ Qᵢ inhibitor binding site. *A) The protein is shown in green and the inhibitor in purple. The cryo-EM density is shown as a mesh and highlights that the inhibitor density is as strong as the neighbouring amino acid residues. B) An overlay of the EM (green) and X-ray (orange) Qᵢ sites. There are no differences in the secondary structure of the protein but slight differences in some amino acid positions such as Phe187 and Met194.*

The structure of bovine bc_1 bound to JAG021 was completed as a separate study to the GSK and SCR inhibitor bound structures. The resolution attained for JAG-

bc₁ was 3.3 Å which is higher than the 4.1 Å achieved for the previous inhibitor bound structures. It is thought that the increase in resolution is due to the developments in the image processing software and has not been influenced by the JAG021 compound. Notably, the introduction of RELION3.0, in particular the per-particle CTF refinement and polishing steps, made a big impact on the resolution. The previous data sets have not been reprocessed in RELION3.0 due to the raw frames being removed after the study was published and other work taking priority but it is expected that if the frames were available the resolutions could be improved to sub 4 Å. Consistently in all data sets, the EM and X-ray derived models have shown no significant differences in the important Q_i site and have only differed at the Rieske domain which is accounted for by the formation of crystal contacts. In all three cases, the inhibitor density was visualised which proves that cryo-EM can be used to probe inhibitor binding to *bc₁*.

3.3 Discussion

Diseases caused by parasite infections represent a great concern to global health and new therapeutics are urgently required. Cytochrome *bc₁* is a validated anti-parasitic drug target and is at an advanced stage of the drug discovery pipeline. One inhibitor, atovaquone, was successfully developed as a medicine to treat malaria which binds to the protein at the Q_o site. However, as resistance is emerging to this treatment new medicines are urgently needed. Current structure-based drug design programs are being hindered by a lack of structural information from the parasite organism with existing X-ray crystal structures being solved from mammalian homologues. Although this provides sufficient levels of structural details and in many cases shows how the inhibitor binds to the homologue protein, it does not show how the molecule interacts with the target parasite protein. Compounds have been developed which bind with high affinity to the parasite protein. However, the compounds also show high levels of toxicity to the host organism resulting in them not being taken forward to the clinic. Being able to obtain a structure of the parasite organism would not only show inhibitor binding to the target protein but it would also elucidate structural differences between the host and target species which could be utilised in the design of new therapeutics.

Cytochrome *bc₁* has been extracted from the source organism for many of the existing crystal structures. For instance, *bc₁* can be isolated from bovine hearts producing large quantities of material at high concentration (40 mg/mL) which

facilitates protein crystallisation. Obtaining equivalent amounts of parasite protein, by extracting from the native organism, would be extremely challenging and would require significant time and expense due to the size of the parasite organism. Therefore, an alternative method of structural determination is needed which will enable the resolution to be high enough to see inhibitor binding yet use less protein. One approach is to use cryo-EM as it requires micrograms of protein rather than the milligrams needed for X-ray crystallography. For this study, after optimisation of the cryo-EM grids, inhibitor-bound bc_1 grids were prepared at a concentration of 5 mg/mL which gave reproducibly consistent ice with a good distribution of protein within the holes. For each inhibitor, four cryo-EM grids were made resulting in 12 µL of sample being used, equating to 60 µg of protein. After the crystallisation conditions had also been optimised, one 24-well plate containing three 2 µL droplets in each well was prepared for each inhibitor sample. This required 144 µL of 40 mg/mL sample, resulting in 5.7 mg of protein being used. Cryo-EM therefore used ~100 times less protein than X-ray crystallography for each inhibitor which was studied.

By using bovine bc_1, work in this chapter has shown that cryo-EM is an amenable technique to determine high resolution structures of the system thereby representing a viable alternative to X-ray crystallography. This is exemplified by four structures of bovine bc_1 being determined to similar resolutions attained using X-ray crystallography. The overall architecture of the protein is similar to crystal structures of the complex as there is density for all but one of the different subunits (which was lost during the purification step). The only differences between the X-ray and EM derived structures occurs at the Rieske domain which is known to be mobile during the catalytic cycle and is locked into one particular conformation upon formation of the crystal lattice. An advantage of using cryo-EM to determine the structure is that the need for crystallisation is removed therefore the protein is not locked into any particular conformation due to crystal packing thus the structure can be determined in a more-native state. The poor resolution also suggests the domain is continuously moving and does not adopt particular conformations as these could not be trapped during 3D classification attempts.

Furthermore, the resolution achieved using cryo-EM has enabled inhibitor density at the Q_i site to be visualised in all three of the inhibitor-bound maps. Interestingly, in GSK-bc_1 the inhibitor density revealed two different binding poses which had previously not been identified by X-ray crystallography. The alternative

binding pose was not an artefact of the imposed C2 symmetry as the inhibitor does not sit on the symmetry axis and the density is the same in the C1 map. As the additional binding pose was not initially identified using crystallography, cryo-EM could provide unique insights into alternative modes of binding and the occupancy of inhibitors. Moreover, the resolution obtained for JAG-bc_1 achieved a resolution of 3.3 Å which was higher than the X-ray crystallography structure at 3.5 Å. At the inhibitor binding site, the local resolution was 3.1 Å. The density for the JAG021 compound was the weakest in all three of the maps as this compound is more potent against the parasite enzyme and therefore does not bind to the bovine model system as strongly. This is highlighted by JAG021 inhibiting the bovine enzyme less than GSK932121 when tested at the same concentrations (at 0.1 μM, JAG021 inhibits the activity by 36% compared to 64% for the GSK compound). JAG021 and GSK932121 both contain the same pyridone core and the EM density suggests that they could adopt two different binding conformations at the Q_i site. Therefore, obtaining structural details showing how these compounds interact with the parasite protein are vital to future SBDD programs.

The structural information gleaned from cryo-EM is equivalent to that provided from the X-ray crystal structures for the bovine enzyme. Therefore, if purified parasite bc_1 from either *P. falciparum* or *T. gondii* can be obtained, cryo-EM is an amenable technique to study the system. Whether the parasite protein behaves in the same way as the bovine test sample, not only on an EM grid but also during the image processing steps is yet to be determined. However, the work in this Chapter has shown that cryo-EM provides a viable alternative to X-ray crystallography to study the system and provide the essential information on how the compounds bind to the target protein. Future inhibitor bound studies of bc_1 will subsequently be driven using an EM approach rather than X-ray crystallography which will be underpinned by the work described here. This structural information will ultimately be utilised in future SBDD programs leading to the design of a new generation of inhibitors which could be used to treat the devastating parasitic diseases.

4   The use of virtual screening to identify V-ATPase inhibitors

4.1   Introduction

V-ATPase is a member of the rotary family of enzymes and has been implicated in a number of diseases such as cancer and osteoporosis. At present, there are no clinically relevant therapeutic agents which target V-ATPase. This could be due to the enzyme having many roles within the cell which include maintaining the pH in cellular organelles, protein processing and degradation, and the transport of small molecules and ions (see Section 1.6.1). This means that designing a therapeutic to knock-out just one of these functions is extremely challenging. However, the protein does exist in different isoforms within the body; with the several isoforms being specific to certain tissue types. The paucity of structural information on the different isoforms has hindered the design of specific isoform inhibitors. There are compounds which are highly potent and selective inhibitors to V-ATPase but none of these compounds are used in the clinic. Moreover, many of the inhibitors are macrocyclic rings which means that they are expensive to buy and difficult to synthesise which makes it challenging to study the system in more detail.

Therefore, the aim of this part of the project was to design and synthesise novel inhibitors against yeast V-ATPase using a structure-based drug design approach. The inhibitors would act as chemical tools to probe the system in more detail and would not be used to treat any disease at this stage. The initial 'hit' compounds could then be modified, synthesised and tested for biological activity with the aim of making a potent inhibitor which was not only more synthetically tractable but also easier to handle than the current macrocyclic V-ATPase inhibitors.

4.2   Results

4.2.1   Sequence similarity between different species

Compounds were initially designed against the yeast V-ATPase, which acts a model system. Selectivity for certain species can be engineered into the inhibitor scaffold based on sequence differences within the binding pocket. The sequence conservation was analysed for the full complex using consurf[249] (Figure 4.1A). Highly conserved residues are often important for function or maintaining protein stability and can present a lower mutagenesis propensity, therefore designing selective

compounds can be challenging. Analysis of the V-ATPase conservation revealed that the stators (subunits E and G), have the highest sequence variation and unsurprisingly, the ATP binding and catalytic site within the AB domain were highly conserved. Closer examination of the AB domain revealed there are three potential inhibitor binding sites (Figure 4.1B). The first site was the ATP binding site, whereby a compound would act as a competitive inhibitor to ATP. Designing selective inhibitors at this site would present a significant challenge but if successful locking the V-ATPase into one state could reduce the inherent flexibility of the complex thus enabling the resolution of the structure to improve. The other two sites ('top' and 'bottom') could act as allosteric binding sites and are positioned at the top and bottom of the channel between subunits A and B of the V_1 domain (Figure 4.1B). If inhibitors bind at these sites, it would theoretically prevent the subunits from closing thereby disrupting the catalytic cycle of the complex.
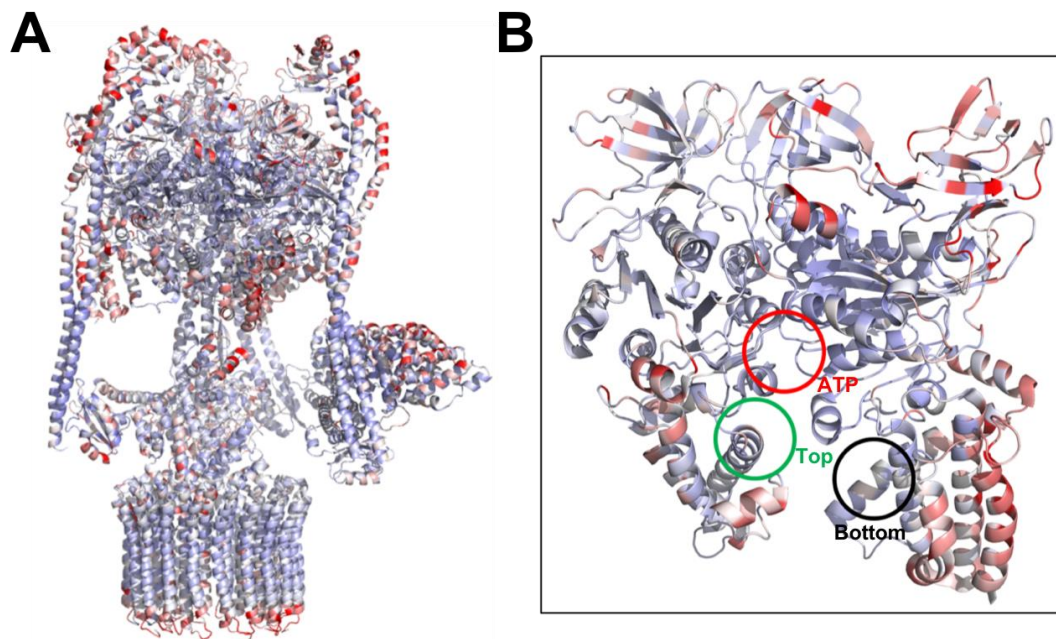
Figure 4.1: A consurf model of the V-ATPase. *A) Residues which are coloured in blue are highly conserved whereas residues in red indicate variation between species. B) A consurf model of the AB open subunit. The ATP binding site is circled in red and is highly conserved. The green circle indicates the 'top' allosteric binding site and the black circle shows the 'bottom' allosteric inhibitor binding site.*

4.2.2 Identification of inhibitor binding sites

In October 2016, Rubinstein *et al* published a cryo-EM structure of the yeast V_0 domain to 3.9 Å resolution[162]. Analysis of the structure revealed how the *c*-ring interacts with subunit *a,* and suggested a mechanism for proton translocation across the membrane. They were able to show that at the *c*-ring subunit-*a* interface, there is a cytoplasmic half channel which enables entry of the protons to the *c*-ring thus allowing them to be pumped across the membrane (Figure 4.2). This half channel exists as a cavity which is 15 Å by 20 Å in length and is expected to be filled with water molecules[162]. This site could represent a potential inhibitor binding site as inhibitor binding would not only prevent rotation of the *c*-ring but would also block protons from entering the half channel and being pumped across the membrane. Therefore, a vHTS screen was performed at this site (discussed in Section 4.2.3.).
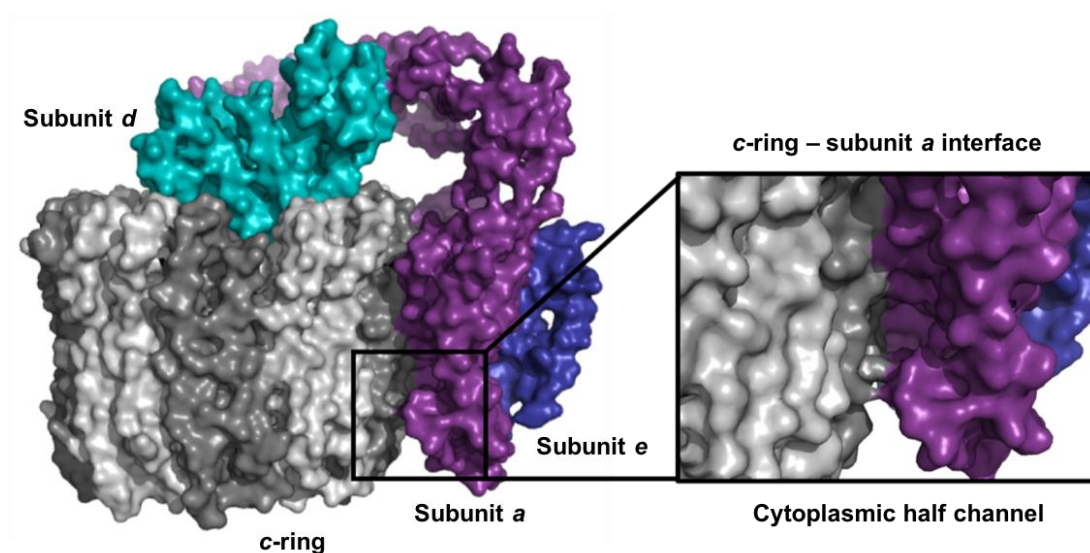
Figure 4.2: The 3.9 Å structure of the membrane embedded V_0 domain. *The alternating subunits of the* c*-ring are shown in different shades of grey. The a-subunit is shown in purple, the e subunit is shown in dark blue and the d subunit is shown in cyan. The box highlights the cytoplasmic half channel which could represent an inhibitor binding site.*

There are eight known inhibitors of the V-ATPase complex which are thought to bind to the *c*-ring and through mutagenesis the residues involved in binding bafilomycin, concanamycin and archazolid have been identified[168]. Bafilomycin inhibits *S. cerevisiae*, *N. crassa* and *M. sexta* V-ATPase[168]. The 3.9 Å map of V_0,

Phyre2[250] was used to thread the *N. crassa* and *M. sexta* sequences onto *S. cerevisiae*. The amino acid residues identified in bafilomycin binding are conserved between all three species, with the main variations at the top and bottom loop regions of the *c*-ring implying that these residues are not integral to the function of the complex. Utilising this information, bafilomycin was docked to *S. cerevisiae* to predict its mode of binding. AutoDock[40] produced the most viable docking pose which resulted in bafilomycin making three H-bonds to the receptor. The chembridge diversity library, consisting of 100,000 compounds, were overlaid using rapid overlay of chemical screening (ROCs) software[251] which resulted in five compounds being selected for testing which were predicted to mimic the proposed bafilomycin binding pose.
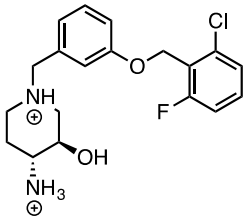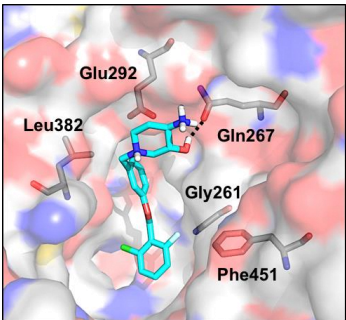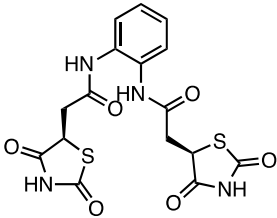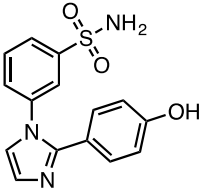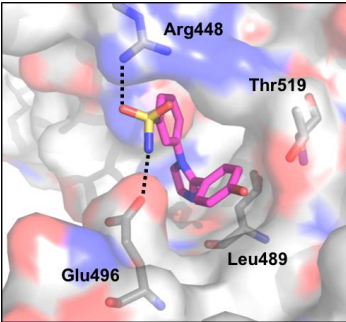
4.2.3 Virtual screens against the AB interface and V_0 site

In order to identify novel V-ATPase modulators, a virtual screen approach was performed on the identified potential binding pockets. These binding sites consisted of the ATP binding site, the allosteric 'top' and 'bottom' sites and the V_0 cytoplasmic half channels which were described in Sections 4.2.1 and 4.2.2, respectively. The cryo-EM structure of the yeast V-ATPase in state one was used to dock the compounds against the AB subunit whereas the cryo-EM structure of the V_0 domain was used to perform the *c*-ring docking. State one was chosen because this contained the highest proportion of particles in the cryo-EM data set so was assumed to be the most physiologically prevalent (see Section 1.6.2). vHTS screens at the three AB binding sites and the *c*-ring-subunit *a* interface were carried out.

All screens were conducted using the Maestro GLIDE docking software[38] with the Chembridge diversity library. Initially, all 100,000 compounds were docked in a high-throughput (HTVS) mode. The top scoring 10,000 compounds were subsequently re-screened using the higher precision (GLIDE XP) mode and the top 100 compounds were manually inspected to look for specific interactions to the target protein. In total, 15 compounds which targeted the AB domain and 10 compounds which targeted the V_0 domain were selected for testing. Example compounds and the predicted interactions to the binding site are displayed in Table 4.1. The compounds selected from the ATP binding site had the highest average docking score (~ -8.5), whereas the compounds predicted to bind to the allosteric 'top' and 'bottom' sites had scores of approximately -7 which suggests that the compounds are more likely to bind

at the ATP binding site than the allosteric inhibitor binding sites. Docking scores which are approximately -9 and -6 correspond to predicted binding affinities in the nanomolar and micromolar range, respectively. Therefore, scores in the range of -7 to -8.5 are likely to be hit compounds although due to inaccuracies with the docking this rarely correlates to actual binding affinities.

Table 4.1: Example compounds and the predicted interactions to the three sites which target the AB subunits.

Structure	Binding site	Docking Score	Interactions
	ATP	-8.198	
	'Top'	-7.416	
	'Bottom'	-7.746	

Initially the compounds were tested in a yeast cell-based assay (assays performed by an MBiol student, Jade Fowler). Yeast growth was measured in the absence and presence of inhibitors at 250 μM and 500 μM concentrations every hour to monitor 'hit' molecules by reducing or stagnating growth. The assay was performed at pH 7.5 and 5.5 as yeast are less sensitive at pH 7.5 to V-ATPase inhibitors which provided a control against general cell toxicity. All 25 compounds identified using vHTS screens and the five compounds predicted to mimic the bafilomycin binding pose were assayed. One 'hit' compound was identified which was predicted to bind to the V_0 c-ring-subunit a interface. The structure of the hit, and its predicted interaction to the protein are shown in Figure 4.3. The molecule has a carboxyl group which is predicted to form a salt bridge to Arg153. The hydrophobic 'tail' of the molecule fills a hydrophobic pocket within the complex and could form π-stacking interactions to Trp737 and Phe583. Closer examination of the binding site revealed that the compound could be modified to improve the binding affinity of the compound. For instance, small hydrophobic groups such as halogens or methyls could be inserted onto the aromatic rings in order to extend the molecule further into the hydrophobic cavity. Also, there are two residues, Ile116 and Ser122, which could be utilised to form extra H-bond interactions.
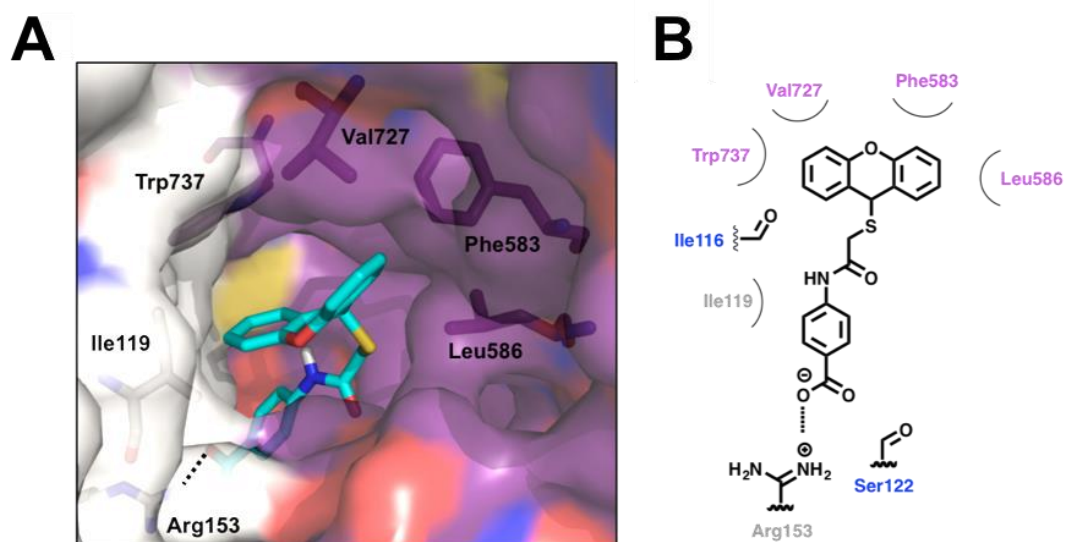
Figure 4.3: The structure and interactions of the hit compound*. A) How the compound binds at the c-ring (grey) – subunit a (purple) interface. B) A schematic of the interactions the compound makes to the complex. It is predicted to form an ion-ion interaction to Arg153 and to form hydrophobic interactions to Phe583 and Trp737. The backbone carbonyl groups from residues shown in blue (Ile116 and Ser122) could indicate potential H-bond contacts.*

However, the compound was shown to be highly toxic against HEK cells in a toxicity assay, therefore, it was not taken forward for further development. The structure of the compound resembles that of a typical detergent as it has the polar 'head' group and hydrophobic 'tail' so it could act by disrupting the cell membrane.

The low success rate of the virtual screens could be due to the modest resolution EM structures which were used to design the compounds. At 6.8 Å, the secondary structure information for the protein can clearly be seen however, there is little accurate side chain density. Therefore, the shape of the inhibitor binding sites may be different to what was used for the docking. Furthermore, vHTS screens traditionally have a low hit rate. Only 25 compounds identified through the vHTS screens were tested for activity, therefore there was a low probability of finding a hit. Furthermore, the five compounds which were predicted to mimic the bafilomycin binding pose had no effect upon the rate of yeast growth which suggested that they did not have an inhibitory effect on V-ATPase. Moreover, the activity of the compound was determined using a yeast cell-based assay therefore it was difficult to tell whether the compounds were having the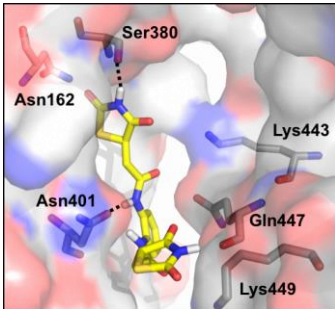 desired effect against the target protein. These problems could be overcome by using a higher resolution structure for docking (ie a crystal structure), selecting a larger number of compounds for testing and testing specifically against the V-ATPase enzyme rather than using a cell-based assay.

4.2.4 Virtual screens using V_1 crystal site

As described in Section 4.2.3, the initial vHTS screens did not identify any hit compounds therefore a new approach was needed. To increase the chances of finding a hit compound, additional vHTS screens were conducted using a 3.4 Å resolution crystal structure of the *E. hirae* V_1 domain[252]. This crystal structure shows the three different catalytic states of the AB domain (empty, loose and tight), two of which have ATP bound allowing the mode of ATP binding to be identified. The vHTS screens focused on finding an inhibitor which would act as a competitive inhibitor to ATP. vHTS was conducted on all three sites to identify compounds which would bind to either one or multiple ATP binding sites. The activity of the compounds were tested against yeast therefore the sequence similarity between *E. hirae* and *S. cerevisiae* was analysed. There was a 75% sequence identity around the ATP binding site with key residues involved in binding ATP, including Lys238, Arg262 and Phe425, conserved in both species (Figure 4.4).
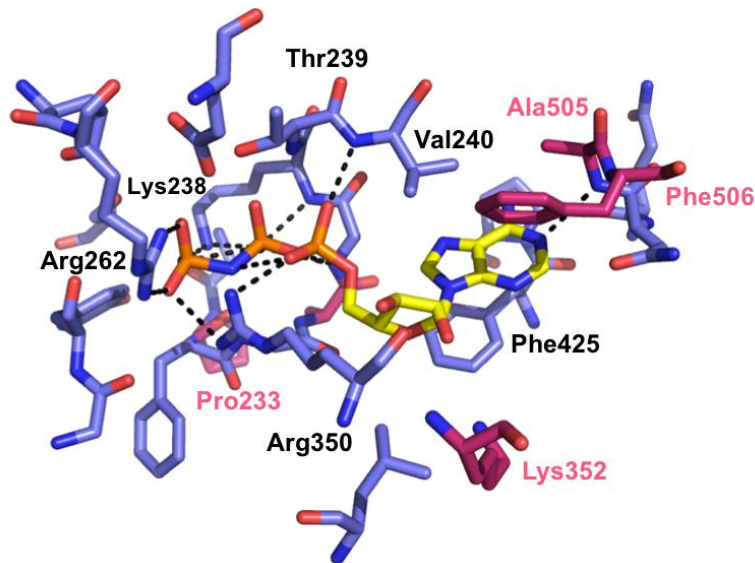
Figure 4.4: The sequence conservation of the ATP binding site between E. hirae and S. cerevisiae.
ATP is shown in yellow. Those residues coloured in blue are conserved whereas those coloured in pink are not conserved. The black lines indicate H-bonds which the ATP makes to the enzyme.

The AB domains at all three catalytic sites were overlaid in UCSF Chimera[236] and coloured by Cα r.m.s.d. values to look for differences in the ATP binding sites. For the tight and loose ATP binding sites, there is very little variation in the position of the amino acid resides. The only significant difference occurs at the two outer helices in chain A (Figure 4.5A,B). However, when the tight and empty site are compared, there are significant changes around the ATP binding site suggesting ATP no longer fits inside the cavity (Figure 4.5C,D).
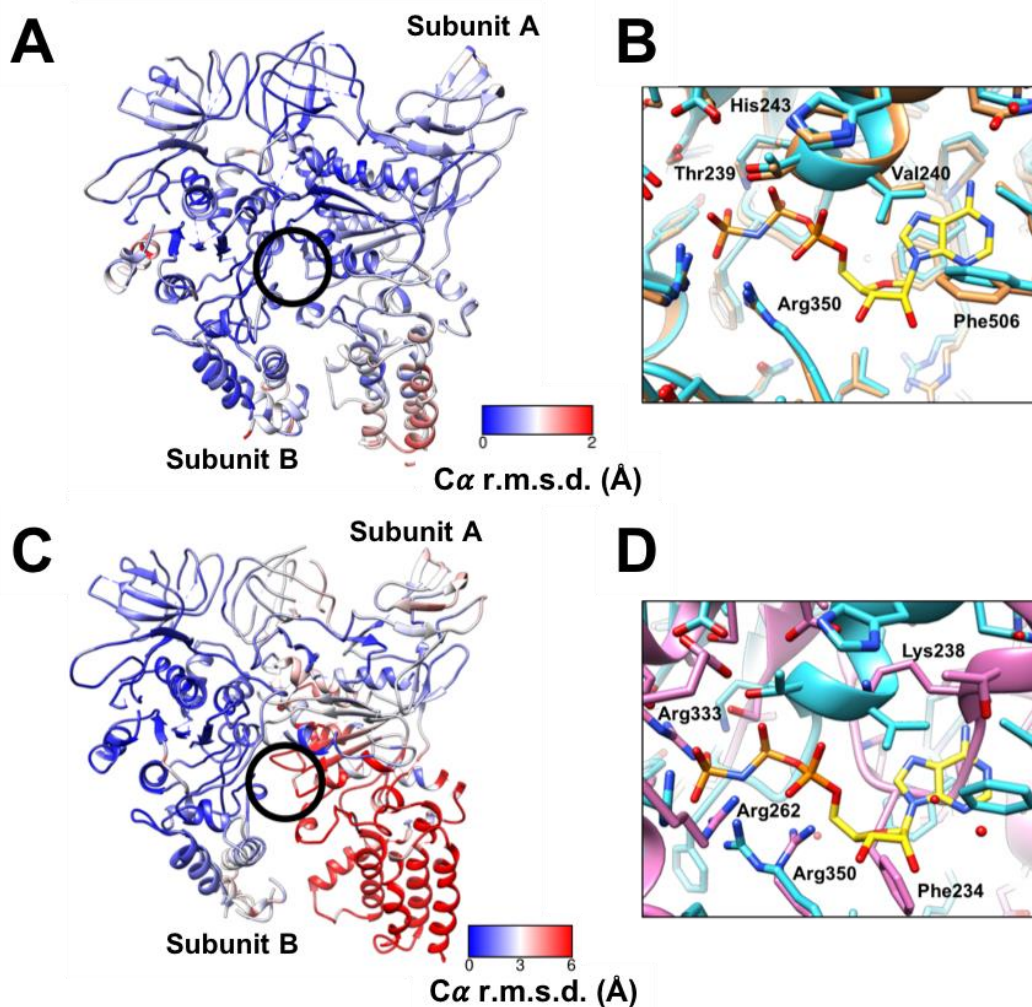
Figure 4.5*: **A comparison of the ATP binding site in the empty, loose and tight conformations.** A) The ATP binding site (black circle) in the tight and loose conformations have been overlaid and coloured according to Cα r.m.s.d. value. The only difference occurs in the outer helix which changes position in the tight and loose conformations. B) An overlay of the ATP binding site between the tight (cyan) and loose (orange) states. There is little change in the positions of the amino acid residues which surround the ATP molecule (yellow). C) A comparison of the tight and empty ATP binding sites. There are large differences in subunit a as the domain adopts the two different conformations. D) An overlay of the tight (cyan) and empty (pink) ATP binding sites highlighting that there is little overlap in the amino acid residues and ATP (yellow) no longer fits into the binding site.*

Virtual screening at all three sites was conducted in GLIDE using the Chembridge diversity library. Using the same approach described above (Section 4.2.3), 44 of the highest scoring compounds were chosen for testing. The docking scores of the compounds at the ATP binding site were high (~ -14) and comparable to that of ATP docked into the loose and tight sites resulting in scores of

-13.658 and -13.807, respectively. Interestingly, despite the high similarity in the position of the amino acids within the two ATP binding sites, only seven out of the top 50 compounds were predicted to bind to both sites. Figure 4.6 exemplifies the interactions which the compounds are predicted to form at all three of the inhibitor binding sites.
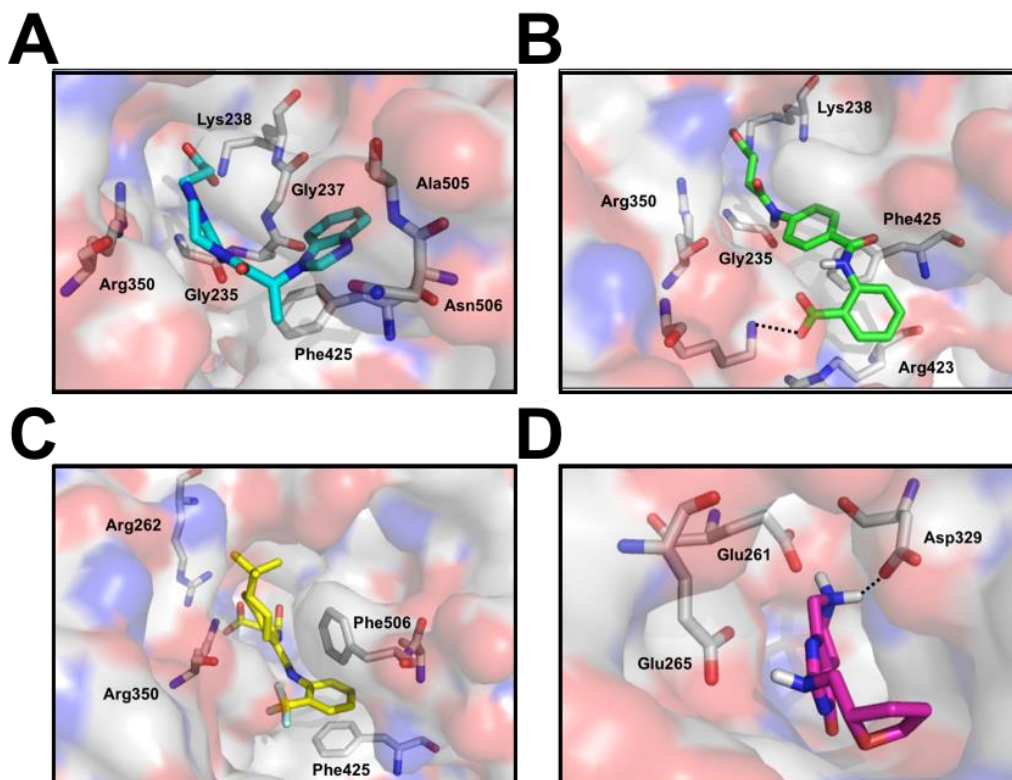
Figure 4.6: Example compounds selected from the vHTS screens against the crystal structure. A) A compound predicted to bind to both the loose and tight ATP binding sites with scores of -13.542 and -14.164, respectively. The compound could make an ion-ion interaction to Lys238 and a π-stack with Phe425. B) A compound docked into the loose ATP binding site (docking score -13.236) forming an ion-ion interaction with lys 238. C) A compound docked into the tight ATP binding site (docking score -15.168) which forms hydrophobic interactions to Phe506 and Phe425. D) A compound docked into the empty site with a score of -9.662. It can form ion-ion interactions to Asp329, Glu261 and Glu265.

The 44 compounds identified were tested against yeast vacuolar membranes in an ammonium molybdate assay which measures the amount of phosphate released as ATP is hydrolysed. The results from the assay will be discussed in Section 4.2.5.

4.2.5 Screening vHTS compounds using membrane assays

The 44 compounds identified as described above, as well as the initial 22 compounds identified through vHTS screens were tested for activity. The compounds were predicted to bind to seven sites; the AB open interface from the EM map encompassing the ATP-binding, 'top' and 'bottom' allosteric sites (Section 4.2.1), the V_0 subunit a/c-ring interface (Section 4.2.3) and the X-ray derived V_1 structure showing the ATP-binding site in the empty, loose and tight conformations (Section 4.2.4). The compounds were tested against the vacuolar membranes which were obtained as described in Methods Section 2.3.3. The compound activity at 1 mM and 0.5 mM was measured to see if the compound had any effect on ATP turnover and if so whether there was a dose-dependent response. The activity of the compound was calculated as a percentage of the activity of the protein when there was no inhibitor present. Therefore, a high level of inhibition would correlate to the protein having a low percentage activity. The results from the initial assays are shown in Figure 4.7 and are coloured by predicted binding site.
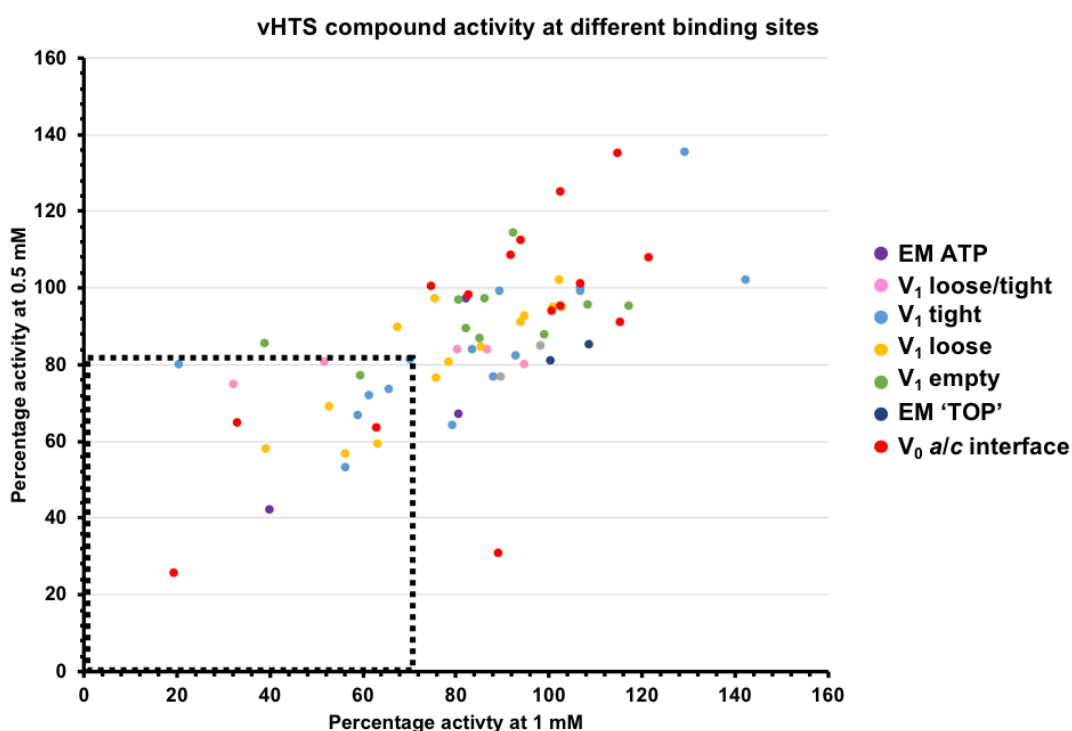
Figure 4.7: The activity of the vHTS compounds. *The results of the membrane assays from the 66 compounds identified through virtual screening. The activity of the compound at 1 mM is plotted against the activity of the compound at 0.5 mM. The colours indicate the binding site which they were designed against. The box highlights the 17 hit compounds which were predominantly designed from virtual screens using the ATP binding site in the crystal structure.*

It was decided that for the assaying, a 'hit' compound would be one that reduced the activity of the protein to <70% at 1 mM and <80% at 0.5 mM. Out of the 66 compounds tested, 17 compounds (25%) were hits. Unsurprisingly, the majority of the hit compounds (76%) were designed against the V_1 crystal structure where the higher resolution allowed for more accurate side chain placement. There were more compounds predicted to bind in the loose (4) and tight (6) sites than those predicted to bind at the empty site where only one compound was classified as a hit. This could be due to the empty site having a more open conformation to facilitate ATP binding therefore making it more difficult for small molecules to bind. Moreover, three of the hit compounds (18%) targeted the V_0 subunit *a/c*-ring interface which could again be influenced by the high-resolution structure (3.9 Å) used to perform the docking. There was only one hit compound found which was designed using the full complex EM map. Moreover, another reason for not identifying any hit compounds at the 'top' and 'bottom' allosteric sites is that they could be unsuitable for accommodating ligands or if the compounds do bind, they might not interfere with the function of the complex. The percentage inhibition of the protein in the presence of the hit compounds was calculated and the results are summarised in Figure 4.8.
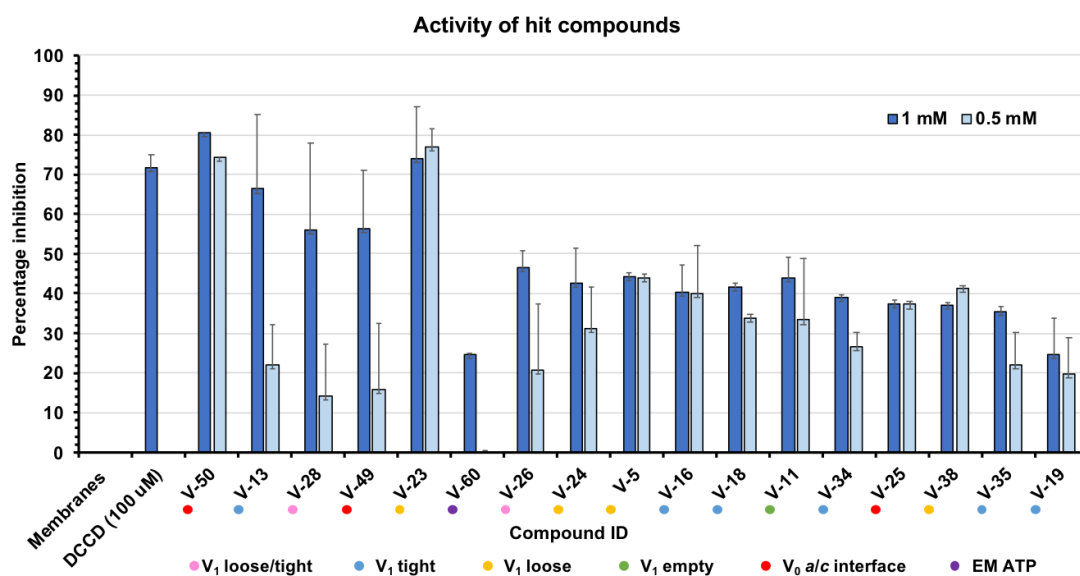
Figure 4.8: The assay results for the hit compounds. *The percentage inhibition of the 17 hit compounds at 1 mM (dark blue) and 0.5 mM (light blue) concentrations. The circles indicate the binding site which the compounds were docked against. 12 out of the 17 compounds show higher levels of inhibition at 1 mM compared to 0.5 mM indicating a dose-dependent response.*

The positive control in all assays was 100 μM DCCD (IC$_{50}$: 32 μM) as this inhibits V-ATPase and is cheaper and easier to handle than other V-ATPase inhibitors, such as bafilomycin and concanamycin where the IC$_{50}$ values are both ~10 nM. However, a disadvantage of using DCCD is that it does not fully inhibit the complex so there will always be background levels of activity. The results indicate that **V-50** and **V-18** were strong inhibitors at both concentrations however both compounds were highly coloured and interfered with the absorbance readout in the assay and were subsequently not taken forward. The structures of all of the hit compounds are shown in Figure 4.10.

The majority of the remaining compounds indicated a dose response where the amount of phosphate released over the 30-minute timescale was reduced when the compound was at 1 mM concentration compared to 0.5 mM. Therefore, the dose dependent response was probed further by testing the compounds for activity at seven different concentrations, varying between 20 mM and 250 μM. It is worth noting, that this assay was performed with only two technical repeats due to there being insufficient amounts of compound available so the error bars represent the standard deviation between two technical repeats. Due to the limited amounts of compound available the assay could not be repeated so it was an N=1 experiment. Therefore, accurate IC$_{50}$ values were not calculated and only the general trend was observed. The results from the assay are showed in Figure 4.9.
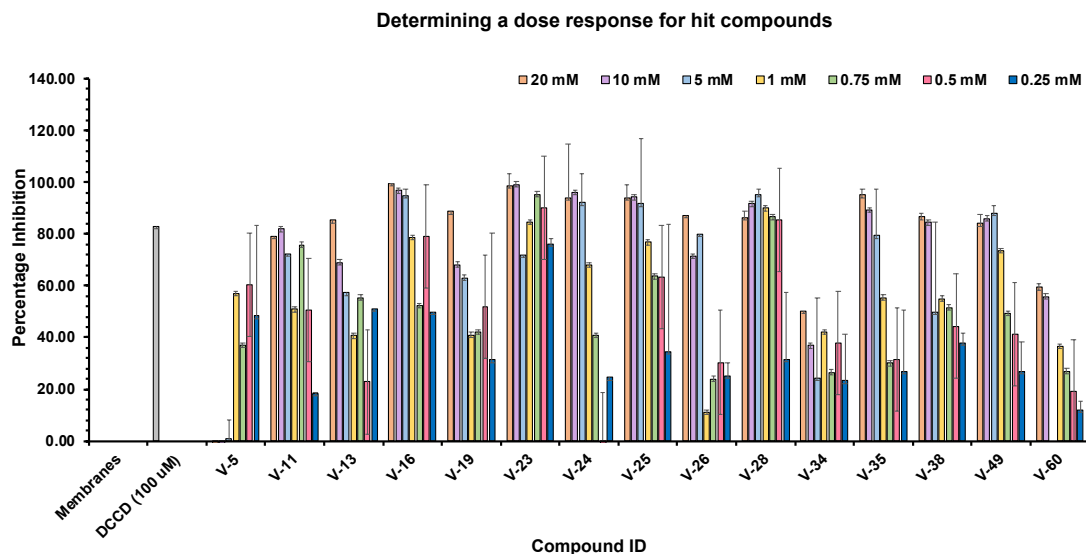
Figure 4.9: Determining a dose dependent response for the hit compounds. *The 15 hits were tested for activity at 7 different concentrations to establish whether the inhibitor reduced phosphate release in a dose dependent manner. The majority of the compounds show a strong dose dependent response with some of the compounds inhibiting V-ATPase at a higher level than the DCCD control at high concentrations.*

The results indicated some levels of a dose dependent-response for 9 out of the 15 compounds (60%). With **V-5**, **V-11**, **V-23**, **V-26**, **V-28** and **V-34** not showing a clear dose dependent relationship. For **V-5**, the compound precipitated out of the assay buffer at the high concentrations (>5 mM). **V-23** and **V-28** showed high levels of inhibition at almost all of the concentrations tested and **V-24**, **V-25**, **V-35**, **V-38**, **V-49** and **V-60** all showed a strong dose dependent relationship. However, **V-25** was the 'hit' compound which was previously identified as an inhibitor at the V_0 subunit *a/c*-ring interface. As described earlier in Section 4.2.3, this compound was shown to be highly toxic against HEK cells and was not taken any further. The structures of all of the hit compounds were analysed to see if there were any similarities in the structures of the hits at the site at which they were predicted to bind. The structures of the hit compounds are shown in Figure 4.10.

114
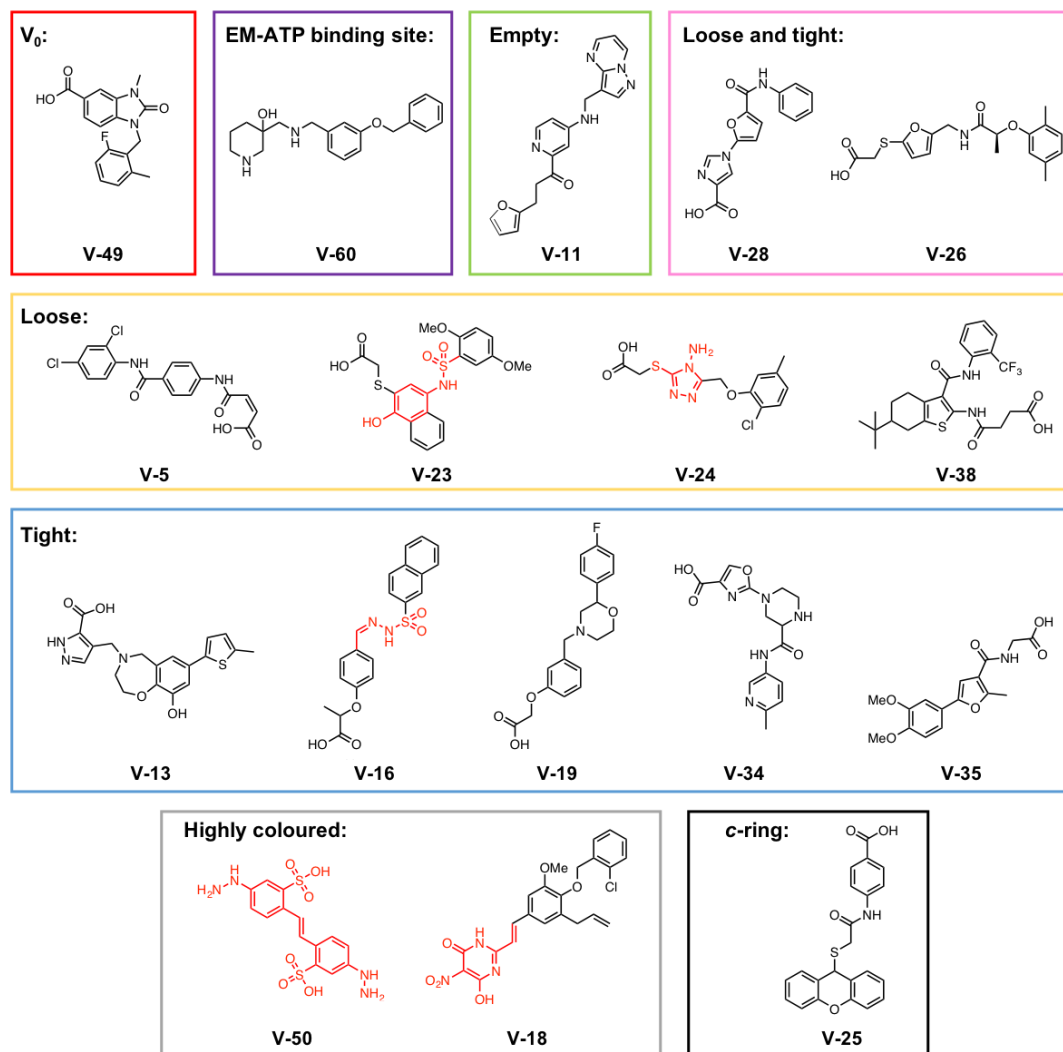
Figure 4.10: The chemical structure of the hit compounds. *The structures of the hits, grouped by predicted binding site, are shown. Structures which contain PAIN-like motifs are highlighted in red. All compounds predicted to bind at the ATP binding site which were docked using the crystal structure contain a carboxylic acid implying that this is key to binding.*

The structures were first analysed to identify any pan assay interference (PAIN) motifs which could have affected the assay results if the compounds were frequent hitters[253]. In Figure 4.10, the PAIN motifs are coloured in red in the affected compounds. **V-23** which showed the highest activity in the assays so far contains one such motif which could be influencing the assay results. The compound contains a hydroxyl group which is in the para position to a sulphonamide and can therefore act as a Michael acceptor resulting in promiscuous binding to other targets. Moreover, **V-24** and **V-16** also contained some PAIN motifs which could have affected the results from the assay. To determine whether the compounds were acting upon the V-

ATPase and were not general 'frequent hitters' the compounds were also tested against the purified V-ATPase enzyme (Section 4.2.7). Moreover, the compounds could be screened against other proteins, in the same assay setup, which the compounds are not predicted to bind to. If the compounds show signs of activity then it would be likely that they are PAINs and are promiscuous compounds.

It was noticed that all compounds predicted to bind to the ATP binding site in either the loose or tight conformations contained a carboxylic acid 'head' group at one end of the molecule which is unsurprising as this carboxylate is predicted to mimic the phosphate group in ATP. Moreover, the opposite 'tail' end of the molecules are generally hydrophobic and contain aromatic rings which could potentially mimic the π-stacking interactions which ATP forms with the protein. Due to these structural features being shared between hit compounds predicted to bind at an ATP binding site, a general common pharmacophore has been established. Interestingly, **V-11** was predicted to bind to the empty site and did not have this carboxylate group.

To further investigate and develop the hit compounds, an 'SAR-by-inventory' approach was taken to allow for structurally similar commercially available compounds to be purchased for testing. This approach allowed for derivatives of all of the hit compounds to be quickly screened thus identifying the best scaffold to be taken forward. This approach involved searching the Chembridge commercial database to find analogue compounds to the current hits which had a 2D and 3D similarity score greater than 80% to the original hit compound. There were variable numbers of analogues available for the hit compounds. For instance, **V-38** had ~90 analogue compounds but **V-11**, **V-23**, **V-49** and **V-60** had no analogues available. The analogue compounds were subsequently docked to the appropriate binding site using GLIDE and the docking score was used to select the analogue compounds for testing. Both compounds which were predicted to bind and compounds which should not bind were selected for testing. This enabled a crude SAR of the hits to be conducted. This is exemplified for **V-38** where eleven analogue compounds were selected for testing; the structures of which are shown in Figure 4.12B. Three of these compounds did not contain the carboxyl group so if the carboxyl group was essential for binding as the docking predicted, then the activity of these compounds should be reduced relative to **V-38**. For the other eight analogues, compounds were selected based upon the predicted binding score and they generally had different substituents on the aromatic and the cyclohexyl ring. The same approach was taken for **V-13**, **V-16**, **V-24** and **V-35** resulting in 22 compounds being purchased for testing.

4.2.6 Determining the activity of the analogue compounds

Following the 'SAR-by-inventory' approach, 35 compounds including the 12 original hits and the 22 analogue compounds, were re-purchased and tested for activity against the vacuolar membranes. The activity was determined at the high concentrations of 1 mM and 0.5 mM to determine whether there was any effect of the compound on the activity of the complex and the results from the assays are displayed in Figure 4.11. The results are shown as a scatter graph with the percentage activity of the protein at 1 mM plotted against the percentage activity at 0.5 mM relative to when there was no inhibitor present. The compounds have been coloured according to the structure of the compound with the original hit compound shown as a diamond and the analogues for that compound displayed in the same colour as circles.
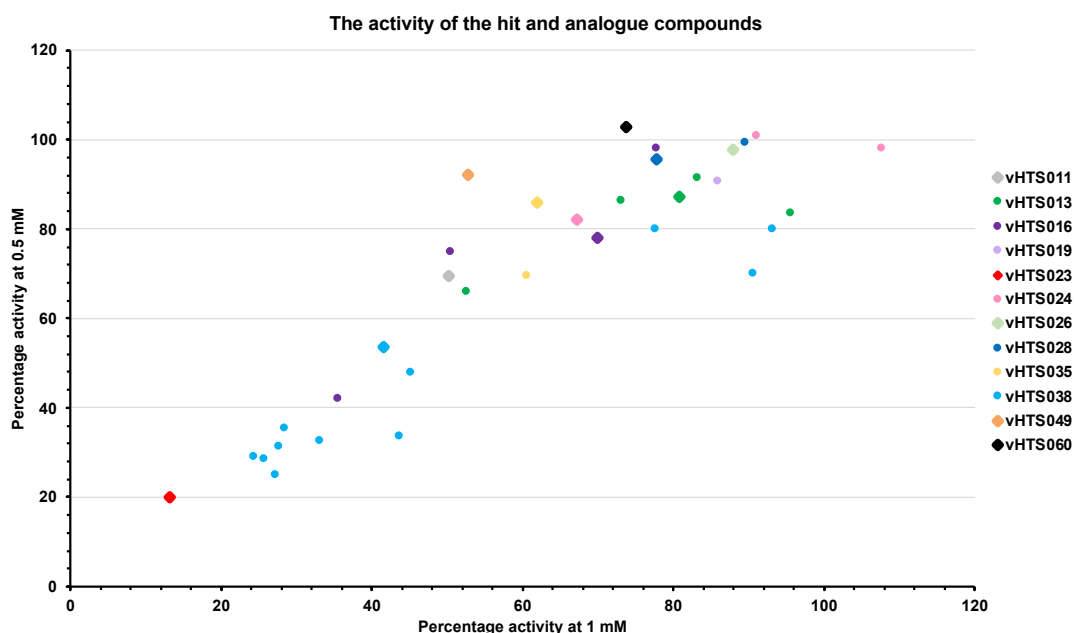
Figure 4.11: The activities of the analogue compounds at 1 mM and 0.5 mM. *The compounds are coloured based upon the scaffold of the compound. The original hits are shown as diamonds and any analogues for that compound are shown in the same colour as circles. The results show that V-38 and the analogues (pale blue) are very effective at reducing phosphate release.*

The results suggest that **V-38** and the associated analogues are highly active against the vacuolar membranes with eight analogues reducing the activity of the protein to <50% at 1 mM. Moreover, **V-23** still has the highest activity at both concentrations. The activity of some of the original hit compounds was reduced compared to when they were first tested. For instance, if the same criteria of defining

a hit compound is used (<70% protein activity at 1 mM and <80% at 0.5 mM concentrations) then only four of the twelve compounds would be classified as being hits. This could be due to the compounds degrading over time and subsequently losing affinity or due to different batches of membranes being used which could have different levels of background ATP-hydrolysing activity subsequently affecting the results. Due to the PAIN motifs present in **V-23** and **V-16**, and the fact that all of the analogue compounds had high levels of activity, the **V-38** scaffold was probed in more detail. Eleven analogue compounds based upon the **V-38** scaffold were purchased and tested for activity against the membrane. The assay results against the vacuolar membranes and the structures of all analogue compounds are shown in Figure 4.12.
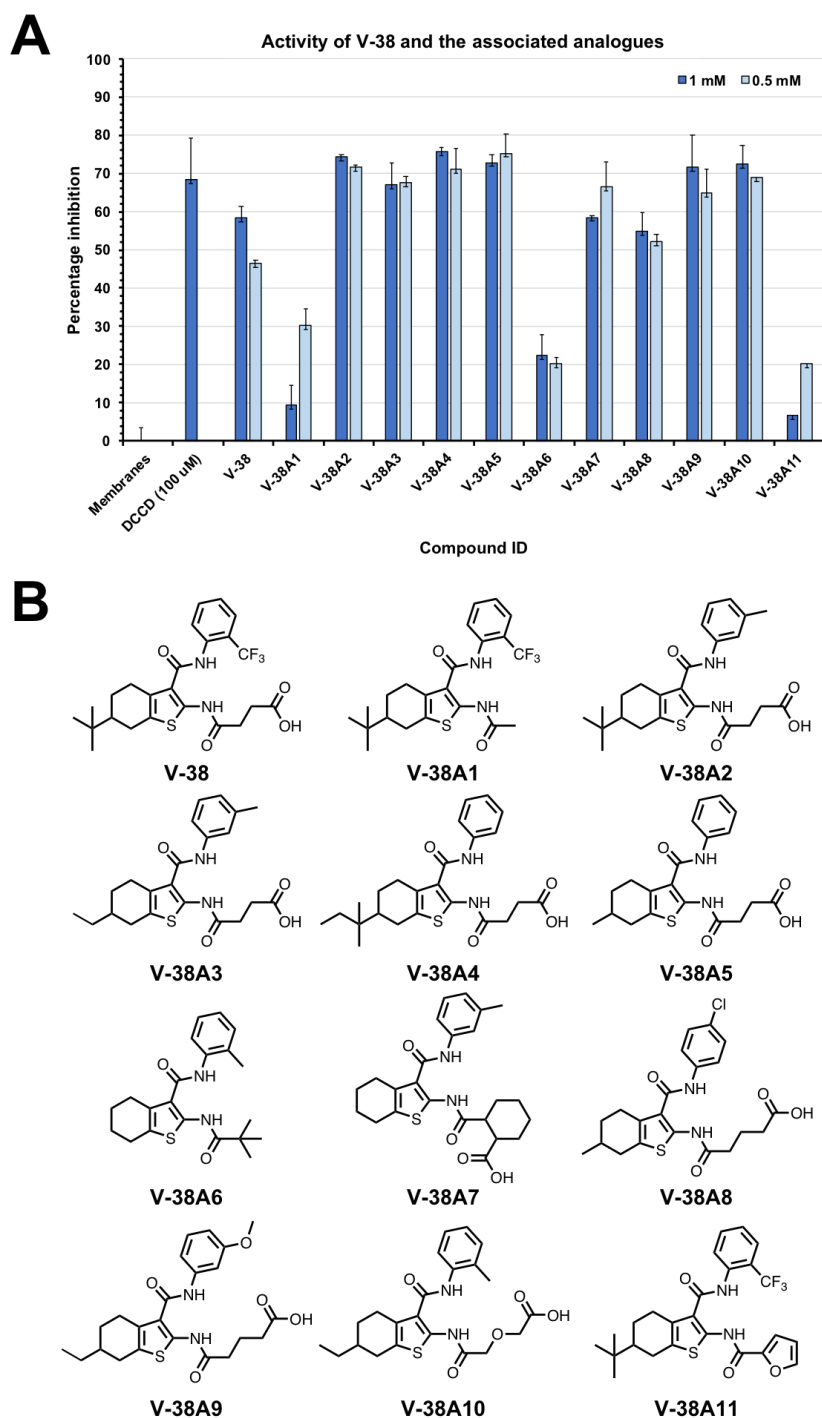
Figure 4.12: The activity and structures of V-38 analogues. *A) The activity of the analogue compounds. The dark blue bars indicate the percentage inhibition at 1 mM concentrations whereas the light blue bars show the percentage inhibition at 0.5 mM. For some of the compounds the percentage inhibition was higher than the DCCD control suggesting the compounds are more effective than DCCD. B) The structures of the analogue compounds. All of the compounds contain the same tetrahydrobenzothiophene core with three of the compounds (V-38A1, V-38A6 and V-38A11) not containing the carboxyl group which is expected to be important for binding. These compounds have a weaker effect on membrane activity suggesting that those without the carboxyl group are not as effective.*

Three of the compounds, **V-38A1**, **V-38A6** and **V-38A11** do not contain the carboxyl group and were therefore not predicted to bind to the complex. The assay results confirm that these compounds do not inhibit the complex as much as the compounds containing the carboxyl group. Therefore, they have a lower percentage inhibition in comparison to the original **V-38** hit compound. This suggests the carboxylate group is important for binding which would indicate that the compound is binding at an ATP binding site where it could mimic the phosphate group in ATP which in turn has helped to validate the predicted binding pose, generated using GLIDE.

The other eight analogue compounds which do contain the carboxyl group were still able to inhibit V-ATPase with similar or higher percentage inhibitions than the original **V-38** hit compound. Furthermore, for five of the compounds the inhibitory effect was greater than the 100 μM DCCD control. The results also show that a range of hydrophobic substituents can be tolerated on the phenyl group in all of the *ortho*, *meta* and *para* positions around the ring. The eight analogues which had an effect on membrane activity were tested at six concentrations ranging between 2 mM and 0.1 mM to establish whether there was a dose response. The assay was performed as an N=1 experiment due to there being insufficient amounts of compounds to repeat the assay. Therefore, the general trend of how the percentage inhibition varied at the different concentrations was observed and IC$_{50}$ values were not calculated. The results from the assay are shown in Figure 4.13.
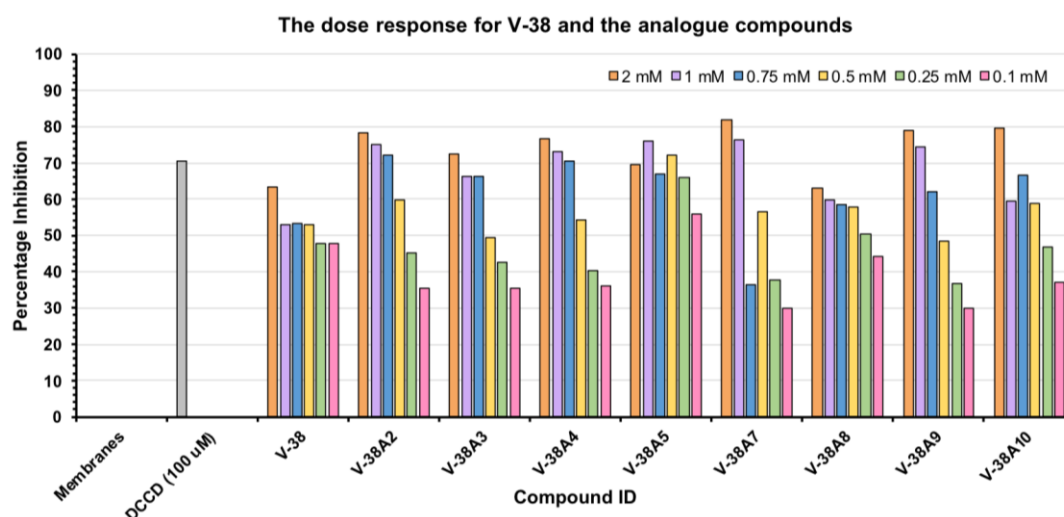
Figure 4.13: The dose-dependent response for V-38 and the analogue compounds. *For all of the compounds, as the concentration of inhibitor was increased, the percentage inhibition was also increased. This highlights the compounds are effective in a dose dependent manner.*

The results show that all of the analogues display a strong dose response with **V-38A5** having the highest percentage inhibition at 500 μM, 250 μM and 100 μM concentrations. Moreover, the approximate IC_{50}'s of the compounds could be estimated to be in the mid-micromolar range. However, for more accurate measurements, the IC_{50} values should be calculated against the purified enzyme where the background activity in the assay would be reduced. Also, the compounds would need to be tested in triplicate to allow accurate and reliable readings to be obtained.

V-38 and the analogue compounds contain a tetrahydrobenzothiophene core and are predicted to bind at an ATP binding site. Literature searches suggest that compounds containing the benzothiophene core can act as kinase inhibitors therefore there could be problems with toxicity if the compounds do not show high levels of selectivity toward V-ATPase[254,255]. To determine whether the hit compounds could have problems with off-target binding the compounds were tested for activity against the Aurora A kinase which catalyses the hydrolysis of ATP and was readily available in the department. Moreover, all of the hit compounds have also been added to HEK cells to determine whether they are toxic.

4.2.7 Testing the compounds for activity against purified V-ATPase

To overcome potential problems with screening against the vacuolar membranes such as the compounds showing off-target effects, the hit compounds and a selected number of analogue compounds were tested for activity against the purified protein. The purification of the protein is described in Chapter 6. The compounds were again tested at 1 mM and 0.5 mM concentrations and ~10 μg of protein was used per well. The results from the assay are shown in Figure 4.14.
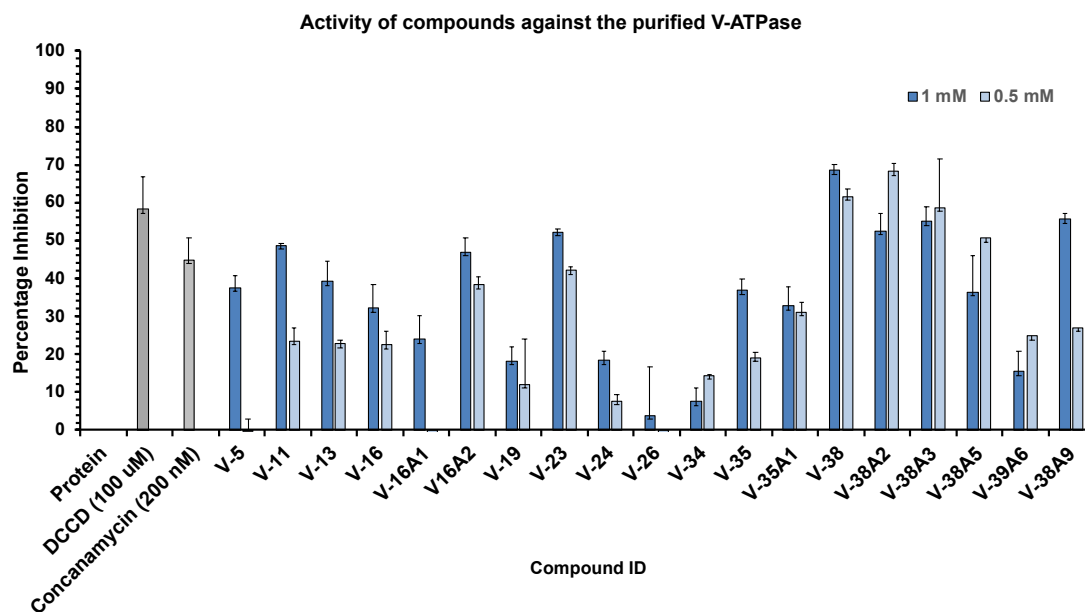
Figure 4.14: The percentage inhibition of the hit compounds against purified V-ATPase. *The hit compounds were tested at 1 mM (dark blue) and 0.5 mM (light blue) concentrations and the percentage inhibition of the protein compared to when there was no inhibitor present was calculated. The results show that V-38 and the analogues have the biggest effect on inhibiting the protein. Error bars represent standard deviation.*

The results show that **V-38** and the analogue compounds have the biggest effect on the amount of phosphate released by the enzyme; percentage inhibition ~70% at 1 mM concentration. Moreover, **V-38A6**, which lacks the carboxyl group, did not have a large inhibitory effect on the activity of the protein compared to the other **V-38** analogues. This again re-enforces how important the carboxyl group might be for binding and inhibiting the complex. In this assay both DCCD and concanamycin were used as positive controls at 100 μM and 200 nM, respectively. Whilst they do reduce the activity of the protein it is not as high as expected. For instance, concanamycin has an IC_{50} of ~10 nM so at 200 nM concentration it would be expected to completely reduce the activity of V-ATPase. One reason for this not being the case is that it could have degraded during storage as the compound is unstable. However, it could also indicate that the quality of the protein preparation needs to be improved if there are still background levels of ATP hydrolysis. The assay suggests that the compounds are acting upon the V-ATPase complex which means that the assays which used the vacuolar membranes provided a good indication on the compound's activity against the V-ATPase. IC_{50} values were not obtained for the

analogue compounds due to there being insufficient quantities of both protein and compound.

4.2.8 Determining the off-target effects against Aurora A kinase

As the majority of the hit compounds were designed against the ATP-binding site, there was a risk that the compounds would display high levels of off-target activity. To test whether the hit compounds are general ATP-binders, they have been tested in the ammonium molybdate assay against the Aurora A kinase which was kindly supplied by the Bayliss group at the University of Leeds. The results from the assay are shown in Figure 4.15.
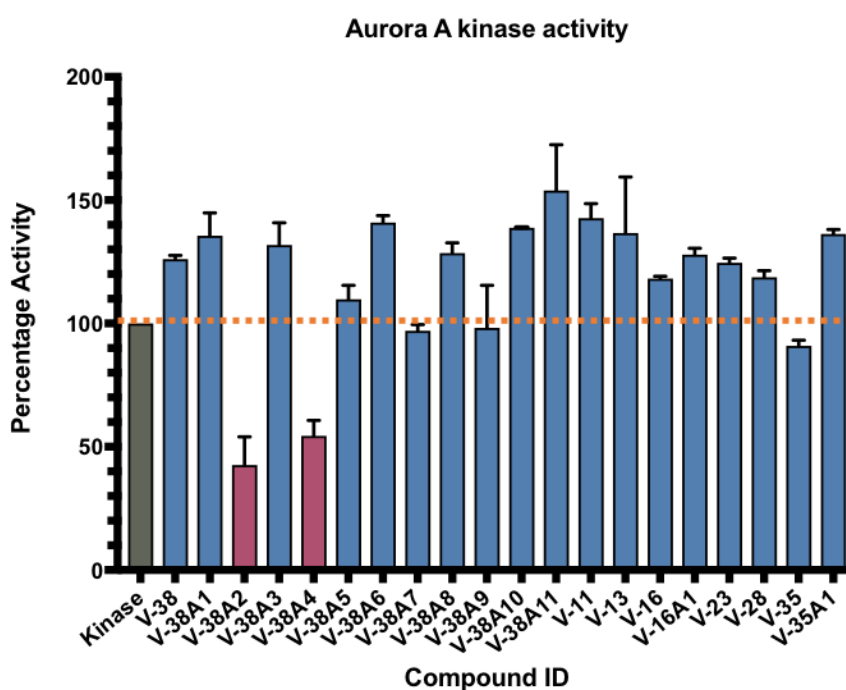
Figure 4.15: Determining whether the compounds have any effect on Aurora A kinase activity*. The hit and analogue compounds were tested against Aurora A kinase at 1 mM concentration. The results show that only V-38A2 and V-38A4 reduce activity to <80%.*

For the assays, 10 μg of Aurora A kinase was used and the compounds were tested at 1 mM concentration to determine whether there was any effect on the activity of the protein. The results show that the majority of compounds do not inhibit Aurora A kinase suggesting the compounds could show some levels of selectivity towards the ATP binding site of the V-ATPase. However, there are two compounds (**V-38A2** and **V-38A4**) which do reduce the activity of the protein to ~50%. These two

compounds are two of the **V-38** analogues which could indicate problems with the selectivity of the scaffold. **V-38A2** was shown to be active against Aurora A with the activity of the kinase reduced to ~50%. Therefore, the compound was docked into the ATP binding site of Aurora A using GLIDE in standard precision (SP) mode which produced a docking score of ~-5. ATP binds to the protein by forming five H-bonds to the protein and two of the carbonyls in the phosphate group are coordinated to the magnesium ion. The predicted binding pose of **V-38A2** suggests only one H-bond is formed between the compound and the protein to Lys162, which also interacts with the phosphate group in ATP. Moreover, there is a good overlap between the benzyl ring in **V-38A2** and the adenine ring in ATP showing that the two compounds can both form the same hydrophobic interactions. As **V-38A2** appears to mimic the binding pose of ATP, the inhibitory effect could be caused by the compound acting as a competitive inhibitor to ATP. A summary of the docking result is shown in Figure 4.16.
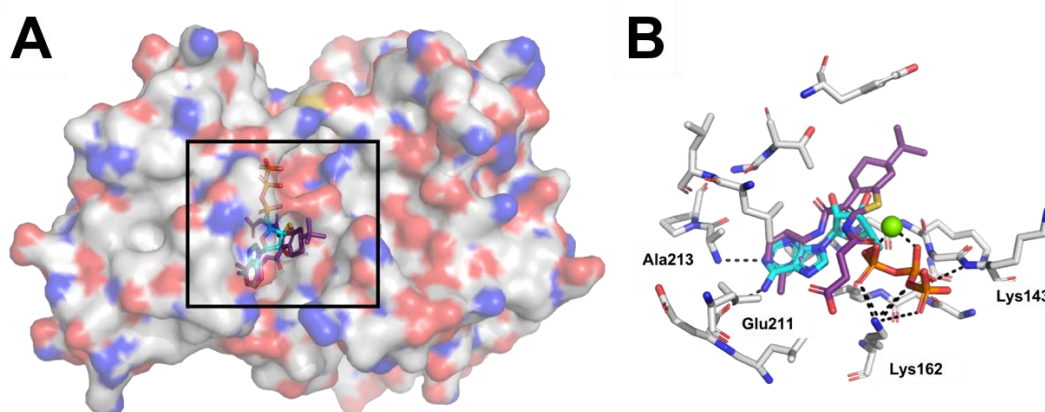
Figure 4.16: Docking V-38A2 into Aurora A kinase. A) The position of the ATP binding site within the kinase (pdb 5dy4)[26]. ATP is coloured by element and shown in cyan whereas V-38A2 is coloured in purple. The docking shows a good overlap in the benzyl rings of the molecules but the tri-phosphate group in ATP extends further into the cavity than V-38A2. B) The H-bonds which ATP (cyan) forms with the kinase are shown in black with the magnesium ion represented as a green sphere. V-38A2 only maintains one H-bond which ATP forms to the kinase to residue Lys162.

The binding of **V-38** and the analogues has been probed to determine whether the compounds could be modified to increase the selectivity of the compounds to V-ATPase and reduce the off-target effects. This is described in Chapter 5. As only one kinase has been tested, this is not representative of all ATP binding proteins. To determine whether the compounds are selective to V-ATPase a much wider panel of

kinases and ATP hydrolysing enzymes would need to be screened. Furthermore, the compounds should be tested against other members of the rotary ATPase family including the F-ATPase to determine if there is any selectivity towards V-ATPase.

4.2.9 Examining the toxicity of the compounds using HEK cells

The 'hit' compounds were shown to be active against purified V-ATPase (Section 4.2.7). In order to test the cytotoxicity of the compounds, the hits were added to adherent HEK-293 cells (~50% confluency) at 20 μM and 100 μM concentrations. The cells were left to grow for 48 hours and then observed to see whether the compounds had inhibited cell growth and/or changed morphology. This was determined by estimating how confluent the cells were and by monitoring any changes in the appearance. Characteristically, if a toxic compound results in cell death, the cells no longer adhere to the bottom of the plate but instead float in the media. Healthy cells are usually 100% confluent after 48 of hours of growing and still adhere to the surface of the plate. It should be noted that this is a crude assay to measure whether the compounds induce sudden cell death which would indicate that the compounds are highly toxic and having off-target effects. The cell observations after 48 hours are recorded in Table 4.2.

Table 4.2: The HEK cell assay observations. *The observations of the confluency and appearance of the cells 48 hours after 20 μM and 100 μM of compound are added.*

Compound ID	Appearance of cells	
	20 μM	100 μM
V-5	100% confluent. Healthy cells	90% confluent. Healthy cells
V-11	100% confluent. Healthy cells	100% confluent. Healthy cells
V-13	100% confluent. Healthy cells	100% confluent. Healthy cells
V-16	100% confluent. Healthy cells	100% confluent. Healthy cells
V-16A1	100% confluent. Healthy cells	100% confluent. Healthy cells
V-16A2	100% confluent. Healthy cells	100% confluent. Healthy cells
V-19	100% confluent. Healthy cells	70% confluent. Healthy cells
V-23	~70% confluency. Some cells floating on surface.	Compound showed moderate levels of toxicity. Cells were floating on surface, no longer adherent.
V-24	100% confluent. Healthy cells	100% confluent. Healthy cells
V-26	100% confluent. Healthy cells	100% confluent. Healthy cells
V-34	100% confluent. Healthy cells	100% confluent. Healthy cells
V-35	100% confluent. Healthy cells	85% confluent. Small amount of inhibited growth but healthy cells
V-35A1	100% confluent. Healthy cells	Cells ~50-60% confluent. Cells appear unhealthy
V-38	100% confluent. Healthy cells	Compound showed some levels of toxicity. ~20% of cells were floating on surface, no longer adherent.
V-38A2	100% confluent. Healthy cells	Compound showed some levels of toxicity. ~20% of cells were floating on surface, no longer adherent.
V-38A5	~80% confluency. No floating or clumped dead cells	Decreased growth, ~50% confluent. Cells appear unhealthy
V-49	100% confluent. Healthy cells	100% confluent. Healthy cells
V-60	100% confluent. Healthy cells	Cells ~50-60% confluent. Cells appear unhealthy

At 20 μM concentration there was little effect on cell growth for the majority of the compounds. Only **V-23** and **V-38A5** reduced the cell growth to approximately 70% and 80%, respectively. There was a bigger effect on cell growth at 100 μM

126

concentration as only ten of the compounds had no inhibitory effect on the cell growth. Moreover, for six of the compounds tested, including **V-23**, **V-38** and the analogues, there was a clear toxic effect as some of the cells were floating in the media and were no longer attached to the surface of the plate. By looking at the cells ~5 minutes after adding the compounds it was clear that this toxic effect built up over time and did not occur immediately. In all cases, after the compound had just been added to the cells, they all appeared to be healthy. The toxicity build-up could be caused by the compounds binding to other ATP hydrolysing proteins within the cell and preventing them from carrying out the function rather than killing the cells on contact. Moreover, another reason for the long-term cell death is that the compounds are acting upon human V-ATPase thereby causing toxicity to the cells.

A problem with these observations is that the results are qualitative, therefore to determine a more quantitative cell viability of the compounds an MTT assay could be carried out. The principle behind this assay is that MTT is a yellow water-soluble dye which is reduced to an insoluble purple formazan salt by NADPH-dependent oxidoreductase enzymes. The purple crystals can then be solubilised and the absorbance of the resulting purple colour can be measured spectrophotometrically at 570 nm. This would provide a quantitative analysis of whether the cells were affecting cell growth. If the compounds are toxic to the cells, then cell growth will be reduced which will lower the amount of the formazan salt produced thus reducing the absorbance.

4.2.10 Producing a hit compound

After conducting a virtual screening approach, and determining the toxicity and selectivity of the hit compounds, **V-38** was chosen to be the hit compound which would be taken forward to hit-to-lead optimisation. The compound contained a tetrahydrobenzothiophene core and showed the highest levels of inhibition when the hits were tested against the purified protein. Other hit compounds included **V-23** and **V-16** which had similar affinities to **V-38**. However, analysis of the chemical structure indicated that these compounds contain PAIN motifs and could therefore have promiscuous activities. Moreover, as these compounds would be difficult to synthesise, it was decided that **V-38** would be the hit compound taken forward for synthesis. Due to two of the **V-38** compounds showing some levels of inhibition against the Aurora A kinase enzyme, derivatives of the compound need to be designed to act upon V-ATPase which will improve the selectivity of the compounds

towards the V-ATPase enzyme. For this to occur, the compounds need to be synthesised. The design, synthesis and biological evaluation of analogue compounds will be described in Chapter 5.

4.3 Discussion

V-ATPase is large membrane protein complex which is implicated in a number of diseases. There are current inhibitors which act upon V-ATPase but they are toxic to the cell and therefore cannot be used to treat any diseases. Moreover, there is currently no structural information which shows how the inhibitors bind to the protein. Therefore, due to the lack of the structural information, SBDD programs have been hindered for the design of new therapeutic treatments.

In this study, a virtual screen (vHTS) approach was conducted at different potential inhibitor-binding sites on V-ATPase. Yeast V-ATPase was chosen as a model system because there was existing cryo-EM structural information from this species and there were purification protocols already in place within the group. As the vHTS approach had not been carried out on this enzyme before, the aim was to see if it was successful in identifying a hit compound against V-ATPase. Therefore, the compounds were not designed to treat a specific disease. Any hit compounds identified could be used as a chemical tool to probe the system and future work could involve modifying the compounds to treat a particular disease.

The vHTS approach involved screening a library of 100,000 compounds at seven potential inhibitor binding sites and then selecting the top scoring compounds for testing in an enzyme-based assay. In total, 66 compounds were initially purchased for testing. Due to the large number of compounds, they were initially screened against the vacuolar membranes and not purified protein. This was because the membranes were easier to obtain than the purified protein and as the membranes contain a naturally high abundance of V-ATPase, it would give a good indication that the compounds were interacting with the target protein.

The compounds were initially tested for activity at the high concentrations of 1 mM and 0.5 mM. The high concentrations were used to determine whether there was any inhibitory effect upon the enzyme. Hit compounds could then be taken forward to enable the dose responses to be obtained. The site which yielded the most hit compounds was the ATP binding site in the 'loose' or 'tight' conformations which meant that the compounds would act as competitive inhibitors to ATP. After

128

screening for toxicity against HEK cells and against a kinase to determine whether there was any selectivity, it was found that **V-38** was the hit compound to take forward. Initially, an SAR by inventory approach was taken, resulting in eleven compounds which had the same scaffold as **V-38** being purchased and tested for activity. This confirmed that the scaffold was active against V-ATPase and that the compounds did show some level of selectivity towards V-ATPase over the Aurora A kinase. The 'SAR-by-inventory' approach also implied that the predicted binding poses generated from the docking as compounds missing a key interaction to the protein did not inhibit the enzyme which suggests that the compounds are interacting at the ATP binding site. The next steps were to determine whether the compounds could be modified to improve the potency of the compounds and to obtain structural information showing how the compounds bind to the protein. This will be discussed further in Chapters 5 and 6, respectively.

5 Synthesis and biological evaluation of V-ATPase inhibitors

5.1 Introduction

Virtual screens were used to identify 'hit' compounds which were predicted to act upon the ATP binding site of V-ATPase. This work was described in detail in Chapter 4. After testing the hits for activity against the yeast vacuolar membranes, the compounds were tested for toxicity against HEK cells and selectivity using Aurora A kinase. Initially an 'SAR-by-inventory' approach was performed to test analogues of the hit compounds and establish whether certain functional groups within the molecule were important for binding. The most promising compound was **V-38**. **V-38** not only showed the highest levels of activity against the purified enzyme but the predicted binding site was also validated by testing compounds for activity which lacked a key carboxylate group. The compounds which did not contain this group, did not inhibit the protein, thus indicating the compound was interacting with the target. Therefore, the aim of the work described in this Chapter was to improve the potency and selectivity of **V-38**, which contained a tetrahydrobenzothiophene scaffold. This would be completed by analysing how the compound was predicted to interact with the target before utilising this structural information in the design, and subsequent synthesis, of novel compounds.

5.2 Results

5.2.1 Exploring V-38 binding to V-ATPase

Before analogues of **V-38** were synthesised, the predicted docking pose of the compound within the binding site was examined. **V-38**, and the analogue compounds, were predicted to bind to the loose ATP binding site of the V_1 domain. **V-38A9** had the highest docking score (-13.425) and a clear dose response in the membrane assay. The docking pose obtained using GLIDE (XP mode) was analysed and it suggested that the compound could bind to the protein by forming H-bonds to Thr239 and an ionic interaction to Arg262, mimicking the interactions made by ATP (Figure 5.1). There was also a π-stacking interaction to Phe425 which also mimics the ATP binding pose. Moreover, it was noted that there were amino acid residues in the binding site which were not being utilised for inhibitor binding such as Glu272 and Gln503. Therefore, a series of systematic modifications were made to the compound using Maestro with the aim of forming H-bond interactions to these residues. These

modified compounds were subsequently docked into the protein to determine whether the predicted binding affinity was improved. One of these modifications involved introducing two hydroxyl groups at the R_1 and R_2 positions (Figure 5.1B). The docking score was improved from -13.425 to -16.509. If the modified compounds could be synthesised and the binding affinity improved, then this could represent a way of making the compounds more selective towards V-ATPase thus overcoming the potential toxicity problems.
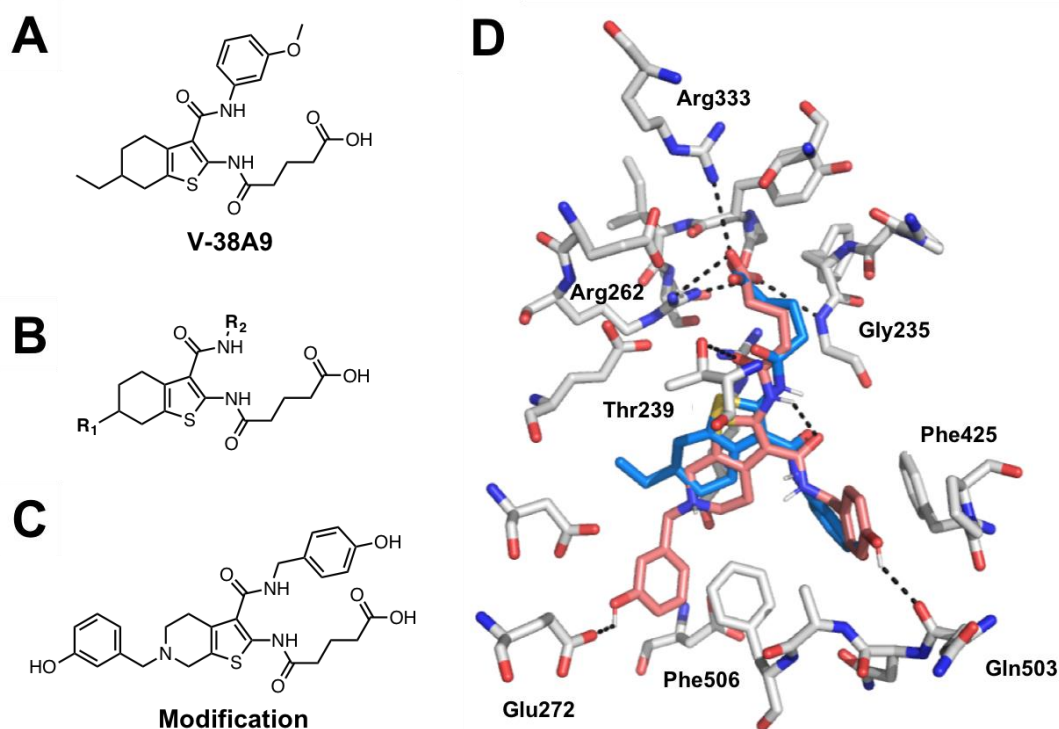
Figure 5.1: Modifications of V-38 based on the predicted binding pose. *A) The structure of V-38A9 which had the highest docking score of all the analogue compounds. B) The R_1 and R_2 positions which were modified to determine whether the predicted binding affinity could be improved. C) The modified compound with the highest predicted binding affinity. D) An overlay of V-38A9 (blue) and the modified compound (coral) in the ATP binding loose site. The modified compound is now able to form H-bonds with Glu272 and GLN503 which may improve selectivity of the compound towards V-ATPase.*

In order to see whether the modifications improved the potency of the compounds, they would need to be synthesised before being tested for their biological activity.

5.2.2 Determining the synthetic tractability of V-38

A synthetic route to the hit compound (**V-38**) was established to allow a series of derivatives to be made. Initially a retrosynthetic analysis was performed which showed that the compound contains a tetrahydrobenzothiophene core (**2**) (Figure 5.2). This core can be synthesised through a Gewald reaction[257,258] which would allow a range of functional groups to be introduced instead of the *tert*-butyl group. The analysis also showed that **V-38** could be further broken down into 2-trifluormethylaniline (**1**) and butanedioic acid (**3**) building blocks.
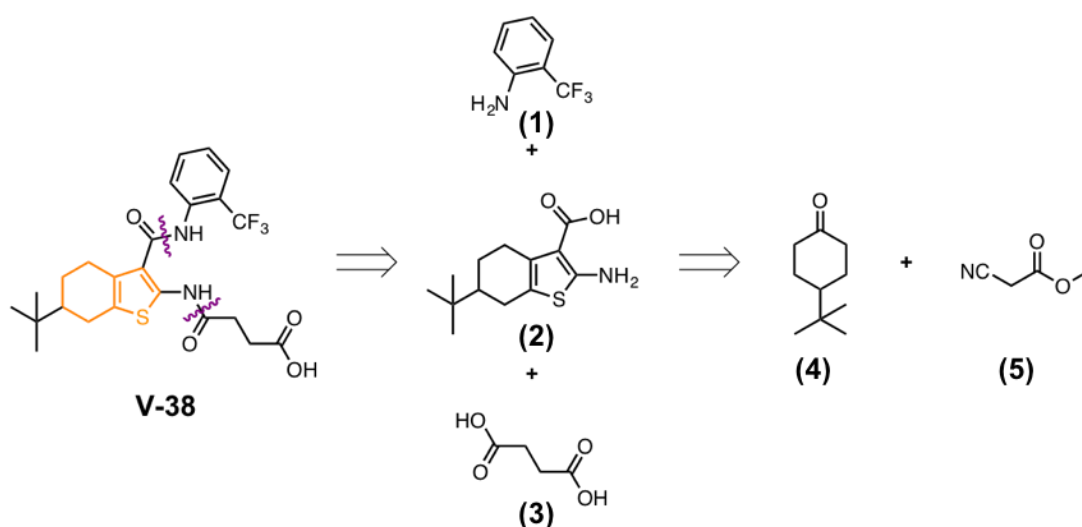
Figure 5.2: A retrosynthetic analysis of V-38. *The tetrahydrobenzothiophene core is shown in orange and the bonds to disconnect are shown in purple. V-38 can be broken down into three main building blocks, (1), (2) and (3), with the core (2) of the molecule broken down into a further two building blocks (4) and (5).*

There is literature precedent for the synthesis of similar compounds which contain the tetrahydrobenzothiophene core which is shown in Scheme 5.1[259]. The route consists of six key steps; synthesis of the core via the Gewald reaction, boc protection, ester hydrolysis, amide coupling, boc de-protection and a final amide coupling reaction to generate the final compound. In order to make derivative compounds, different functional groups can be introduced in the first step, during the Gewald reaction, during the amide coupling step by changing the aniline used and in the final amide coupling step which would determine whether changing the acid has any effect on the activity of the molecules. This route represented the starting point in the synthesis of the analogue compounds.
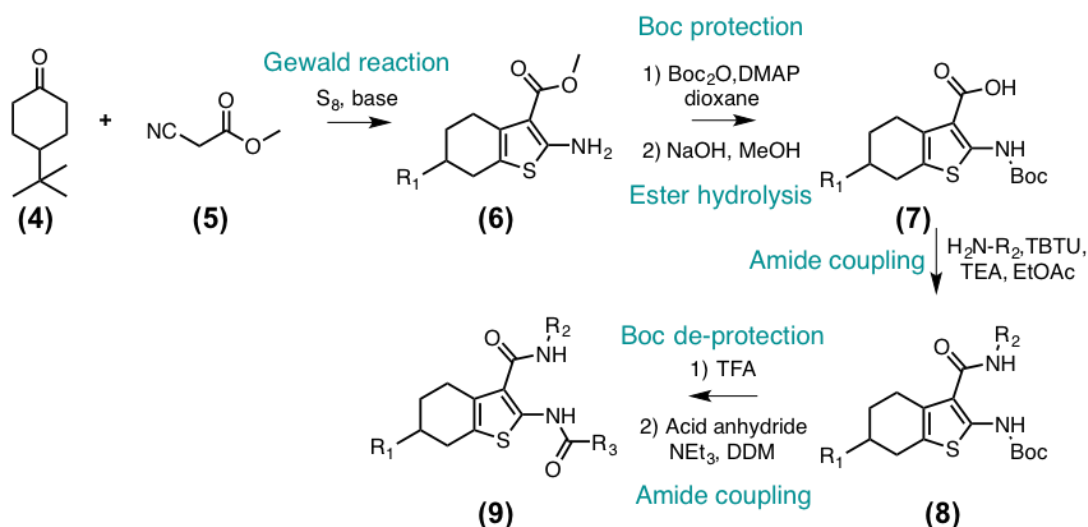
Scheme 5.1: A proposed route to the synthesis of the tetrahydrobenzothiophene core. The synthesis consists of six different reactions. The types of reaction are named in blue font and some of the reaction conditions are also stated.

5.2.3 Synthesising analogues of V-38

Five commercially available compounds were found that had different substituents at the 6-position on the tetrahydrobenzothiophene core. This would enable the effects of the different substituents on the binding affinity to be probed, whilst minimising the number of steps in the synthesis. Therefore, the first step (Gewald reaction) was removed from the synthesis. The commercially available compounds are shown in Figure 5.3.
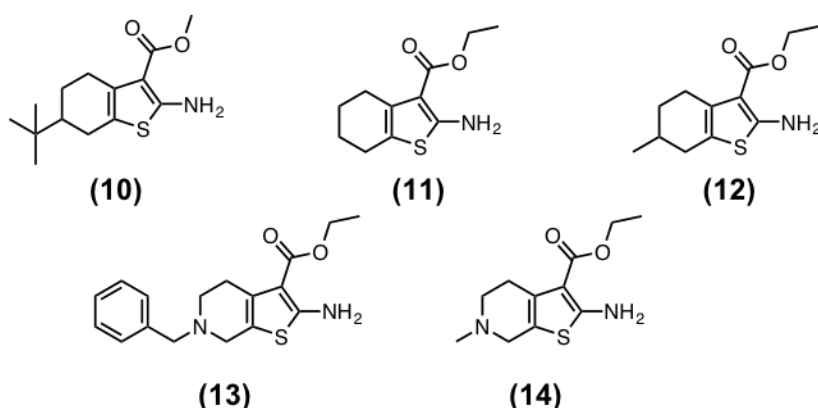
Figure 5.3: Commercially available tetrahydrothiophene cores: The five compounds which were purchased which have different functional groups at the 6-position on the ring. These include (10) tert-butyl group, (11) H-atom, (12) methyl group, (13) N-benzyl group and finally (14) N-Me group.

For two of the commercially available compounds, the 6-position substituents were a *tert*-butyl (**10**) or methyl (**12**) group. The stereochemistry of the *tert*-butyl or methyl substituents within these commercially available compounds was not specified by the vendors, therefore the compounds could exist as two enantiomers. The two enantiomers would be difficult to separate which could create problems for testing the final compounds. Ways to separate out these enantiomers could involve using chiral HPLC or introducing chiral auxiliary groups later in the synthesis. The other three compounds did not contain any chiral centres thereby removing this problem. For two of the compounds (**13** and **14**), the 6-position carbon was replaced by a nitrogen atom. The rationale behind this was that if the activity was maintained, then the amine could be exploited to add a wider range of functional groups at this 6-position thus enabling more compounds to be synthesised and a larger SAR to be established. The docking scores suggested that the protonated nitrogen would not have a detrimental effect on the activity of the compounds. This could be tested by directly comparing final compounds containing core **12** and core **14** where the only difference would be the atom at the 6-position. Moreover, the introduction of the benzyl moiety in core **13** would enable some of the docking modifications to be tested as this would add a bulkier group in this region which was predicted to improve the binding affinity, even without any substituents around the ring.

For all of the commercially available compounds, the initial boc-protection step of the synthesis was successfully performed with the products obtained in high yields (85-92%). The boc-protected material was purified using column chromatography before being taken forward in the synthesis. The second step was the ester hydrolysis which was also successful but gave more moderate yields (51-91%) of product. The lower yields could be due to the product not precipitating out of the reaction when the pH was adjusted to pH 5 in the reaction work-up. The optical rotation of compounds **15**, **16**, **19** and **20** were obtained to determine the enantiomeric ratio. For all four compounds the $[a]_D$ was ~0 thus suggesting a racemic mixture. However, it is assumed that the *tert*-butyl and methyl groups would reside in the equatorial position as this is the most energetically favourable conformation. A summary of the conditions used in the first two steps of the synthesis is shown in Scheme 5.2.
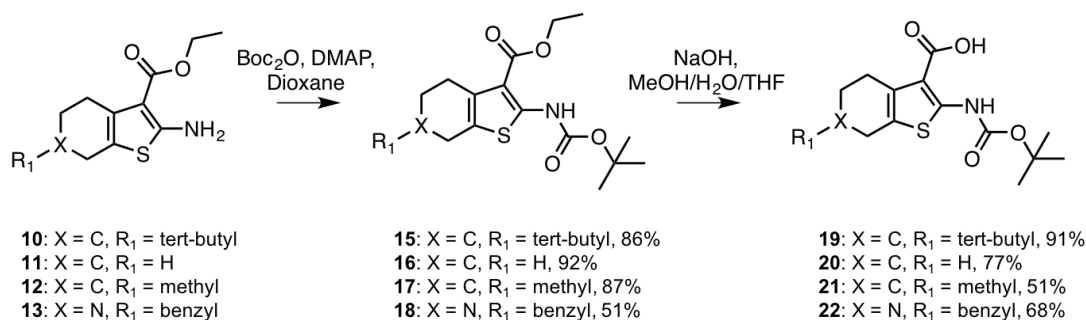
Scheme 5.2: A summary of the first two steps of the synthesis. *The first step was the boc protection which was completed in high yields and the second step was the ester hydrolysis reaction.*

The next step of the synthesis was the first amide coupling reaction. The acid which was used in the initial optimisation steps contained a *tert*-butyl group at the 6-position of the tetrahydrobenzothiophene core (R₁ group). Initially the amide coupling reagent used was EDC.HCl with DMF as the solvent. However, no product was isolated and it is thought that the material was lost during the work-up which involved extracting the product from the reaction mixture using ethyl acetate. The next conditions which were trialled used EDC.HCl as the coupling reagent, DCM as the solvent and the crude material was loaded directly onto the column for purification. This resulted in ~15 mg (~6% yield) of product being isolated and therefore needed further optimisation. The amide coupling reagent was subsequently changed to HCTU with EtOAc used as the solvent. Following purification using column chromatography, the yield of the reaction was improved (~120 mg, 52%) therefore there was enough material to be taken forward to the next steps. An overview of the different conditions trialled is shown in Scheme 5.3.
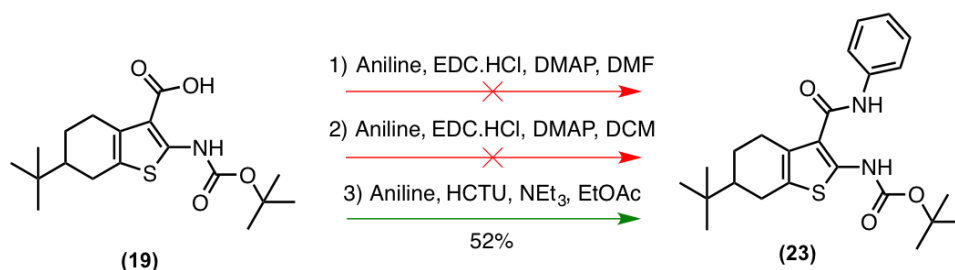
Scheme 5.3: An overview of the amide coupling conditions which were trialled. *The successful conditions used HCTU as the coupling reagent and ethyl acetate as the solvent.*

The successful amide coupling conditions were subsequently used for the different cores containing the different substituents at the R_1 position. Four amines were used; aniline, 4-methoxybenzylamine, m-toluidine and 4-chloroaniline, enabling the activity of the compounds with different R_2 substituents to be probed. The three anilines were chosen because the analogue compounds of **V-38** which were active also contained these groups. 4-Methoxybenzylamine was used to validate the docking which suggested that adding in an extra CH_2 group and a methoxy substituent on the ring would improve the potency of the compounds. In total, eight compounds were made which contained different combinations of the R_1 and R_2 positions achieving yields between 31-52% (Scheme 5.4). For the purification, the crude material was directly loaded onto the column without any washes or extractions as this was previously found to improve the yield of compound obtained.

19: X = C, R_1 = tert-butyl
20: X = C, R_1 = H
21: X = C, R_1 = methyl
22: X = N, R_1 = benzyl

23: X = C, R_1 = tert-butyl, R_2 = aniline, 52%
24: X = C, R_1 = H, R_2 = 4-chloroaniline, 40%
25: X = C, R_1 = H, R_2 = m-toluidine, 38%
26: X = C, R_1 = methyl, R_2 = aniline, 47%
27: X = C, R_1 = methyl, R_2 = 4-methoxybenzylamine, 31%
28: X = C, R_1 = methyl, R_2 = m-toluidine, 47%
29: X = N, R_1 = benzyl, R_2 = aniline, 37%
30: X = N, R_1 = benzyl, R_2 = 4-methoxybenzylamine, 31%
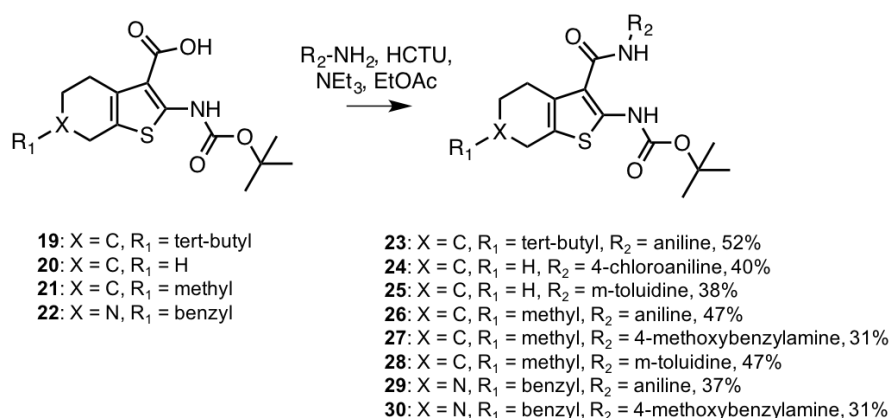
Scheme 5.4: An overview of the amide coupling reaction.

The next step in the synthesis was to remove the boc-protecting group in order to allow the amine to be involved in the final amide coupling step. The boc de-protection was carried out using HCl in dioxane. Initially, the product was neutralised, extracted into DCM and purified using column chromatography. However, material was lost during the purification step resulting in yields of ~50%. Therefore, it was decided that the product would be isolated as the chloride salt which subsequently gave quantitative amounts of product to be used in the final step of the reaction. The reaction was monitored via LCMS and once complete, the solvent was removed to give an oil. DDM was added to the oil and subsequently removed *in vacuo* in order to remove any traces of 1,4-dioxane which was present. This caused the oil to solidify

and this crude material was used in the final amide coupling step without the need for any further purification.

For the final step, succinic anhydride was used to avoid introducing further protection/deprotection steps to the synthesis as ultimately the final compounds need to contain the carboxylic acid group. At this stage, only succinic anhydride was used and the length or variety of the acid was not altered. This would determine whether the compounds retain their activity after synthesis rather than being purchased from commercial sources. The final step of the synthesis was initially performed using the neutralised boc-deprotected amine, succinic anhydride and DCM as the solvent. The reaction was heated at 40 °C and was monitored by LCMS. However, this resulted in very small amounts of protein being isolated (<5 mg, <2%) and needed further optimisation. After the boc-deprotection step was changed, the chloride salt was isolated instead of the neutralised material. Therefore, in the final amide coupling step, one equivalent of triethylamine was added to the reaction in order to allow the amine to react with the succinic anhydride. However, after heating the reaction at 40 °C for over 72 hours the reaction was still not complete by LCMS which showed starting material was still present even after adding in extra equivalents of the succinic anhydride. The solvent was removed and the product was purified via reverse phase chromatography but the yield was extremely poor (<2%) with only ~2 mg of product obtained.

It was suggested that the reaction is acid catalysed and the triethylamine was affecting the rate of reaction so the amide coupling was again trialled without the triethylamine added. The solvent was also changed to THF to allow the reaction to be heated to higher temperatures (65 °C rather than 40 °C), increasing the reaction rate and leading to more product being isolated (~20 %). A summary of the conditions trialled is shown in Scheme 5.5.
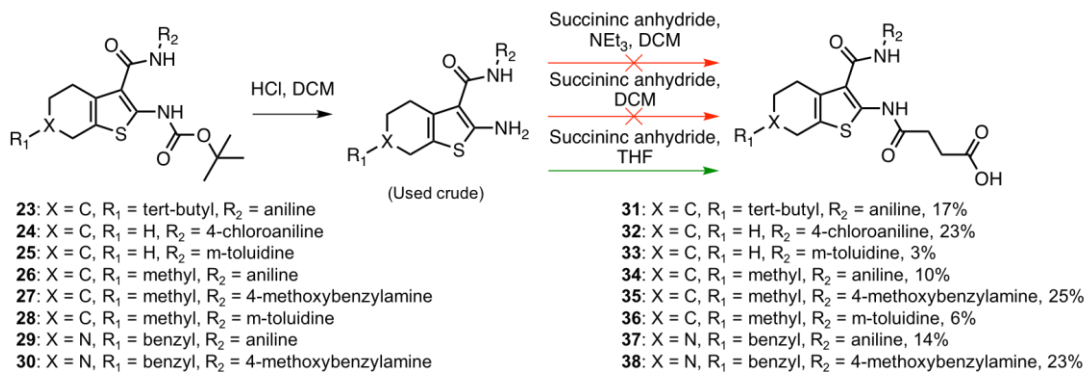
Scheme 5.5: An overview of the final steps of the synthesis. *The first step shows the condition used for the boc de-protection whereas the second step shows the conditions for the final amide coupling reaction.*

One of the reasons behind the low yield in the final step of the synthesis could be that the amine is a weak nucleophile as the electrons can be delocalised around the five-membered ring into the amide bond. Therefore, a method to improve the efficiency of the reaction would be to increase the reactivity of the electrophile. One way to do this would be to replace the succinic anhydride with the acid chloride. However, the acid chloride would need to be protected to prevent one molecule from reacting with two amines. Therefore, an additional deprotection step would have to be added to the synthetic route. This is shown in Scheme 5.6.
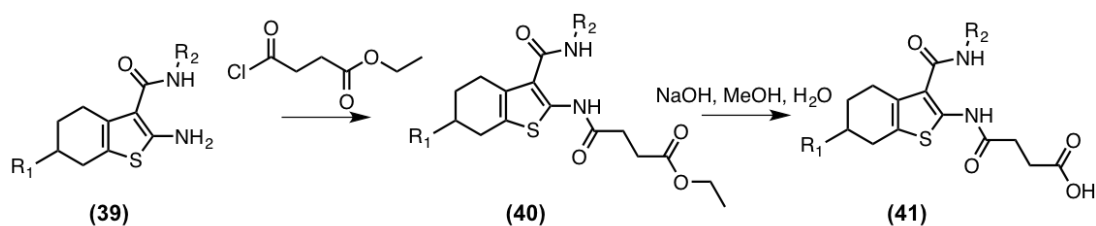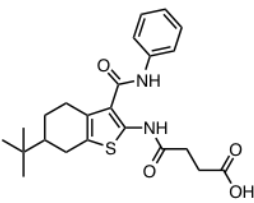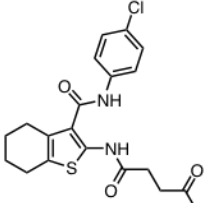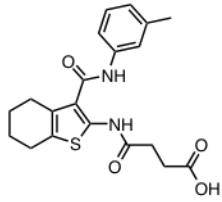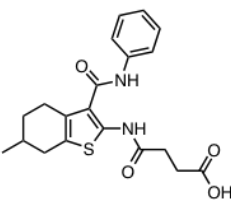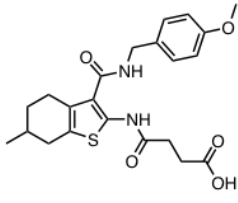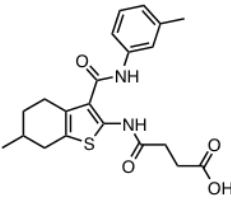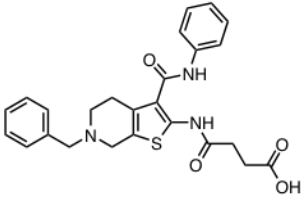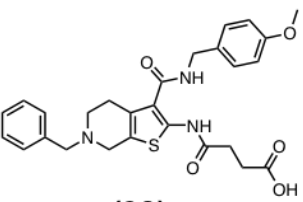
Scheme 5.6: An alternative route to the final compounds. *This would involve using an acid chloride as the electrophile rather than the succinic anhydride. Although, this would need to be hydrolysed in the final step of the synthesis.*

Despite the low yields (3-25%) of the final step, eight final compounds were made using succinic anhydride which were purified using reverse phase chromatography. Therefore, the new method of obtaining the final compound using the acid chloride was not attempted. All of the compounds contained the

139

tetrahydrobenzothiophene core but contained different R_1 and R_2 substituents which were based on the hit compounds identified in vHTS screens or docking different compounds into the binding site. The structure of the compounds and the rationale for their synthesis is shown in Table 5.1.

Table 5.1: Structure of the synthesised compounds, docking scores and the rationale for their synthesis.

Compound	Rationale	Compound	Rationale
(31) Docking score: -9.540	Closest compound to initial hit (**V-38**)	**(32)** Docking score: -7.899	To determine whether substituent at R_1 position is important for binding
(33) Docking score: -7.663	To determine whether substituent at R_1 position is important for binding	**(34)** Docking score: -8.757	Re-make of **V-38A5** to determine accurate IC_{50}
(35) Docking score: -8.279	To determine whether the 4-methoxybenzyl amine improved the binding affinity	**(36)** Docking score: -8.672	To determine the effects of the methyl groups at the R_1 and R_2 position
(37) Docking score: -9.019	To determine whether bulky substituents at 6-position were tolerated	**(38)** Docking score: -9.921	Closest compound to the docking modifications

The synthesis of compounds containing core **14** (Figure 5.3), where the 6-position carbon was replaced with an N-Me group, was unsuccessful in yielding any final compounds. Problems were encountered at the final stage of the synthesis. Even
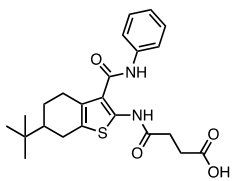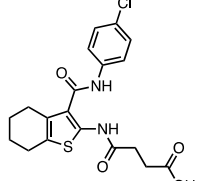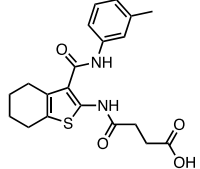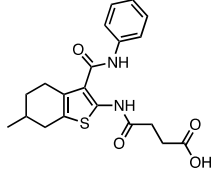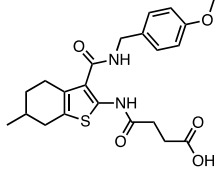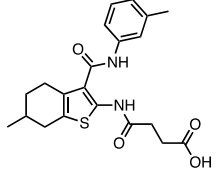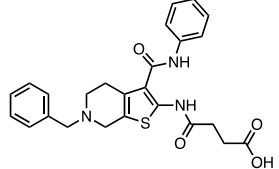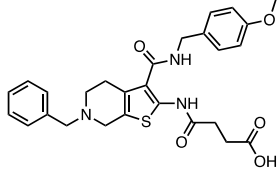
though the correct mass was determined by LCMS, the final product could not be isolated from the reverse-phase chromatography column. Material which was isolated did have the correct mass and was not the correct product by NMR as there was no signal for any aromatic hydrogens. Therefore, the direct comparison of final compounds which contained core **12** and **14** could not be tested.

The eight final compounds would need to be tested for their biological activity. This would determine whether the modifications, as suggested by the docking, had any effect on the potency of the compounds. The following section will discuss how the compounds were tested for activity and whether any SAR had been established.

5.2.4 Biological evaluation of the synthesised compounds

The eight compounds which were synthesised were tested against the purified V-ATPase enzyme in an ammonium molybdate ATPase assay. The IC_{50} values were calculated after determining the activity of the compounds at twelve concentrations between 200 nM and 1 mM. Table 5.2 shows the IC_{50} values for the synthesised compounds and the IC_{50} curves are shown in Appendix A.

Table 5.2: IC₅₀ values of the synthesised compounds. *Compounds were coloured therefore assay readout may have been affected.*

ID	Structure	IC$_{50}$ (μM)	ID	Structure	IC$_{50}$ (μM)
31		41	32		72
33		180	34		112
35		441	36		78
37		411*	38		72*

All of the eight compounds had IC$_{50}$ values in the micromolar range with four of the compounds having IC$_{50}$ values of less than 100 μM. Compound **31** showed the highest levels of activity with an IC$_{50}$ value of 41 μM. Compounds **32**, **36** and **38** all showed similar levels of activity at 72, 78 and 72 μM, respectively. Compounds **35** and **37** were the least potent compounds with IC$_{50}$ values greater than 400 μM and compounds **33** and **34** showed moderate levels of activity (IC$_{50}$ values of 180 and 112 μM, respectively). The IC$_{50}$ values of compounds **37** and **38** should be taken with caution due to the bright yellow appearance of the compounds which could have affected the absorbance readings of the assay. Despite the colour possibly affecting the assay readout, it was interesting to note that **38** was one of the top scoring compounds in the assay. Although missing a hydroxyl group on the benzyl ring, **38** was the compound which most closely resembled the compound with one of the highest predicted docking scores after a series of modifications were made to the original hit compound (**V-38**). The IC$_{50}$ value, of 72 μM, was the second highest scoring compound in the assay which suggests that inserting the N-benzyl group at

the 6-position did not prevent binding to the protein and thus was tolerated at the binding site. The presence of this 6-position nitrogen atom would allow a larger number of substituents to be added at this position which would enable a larger SAR study to be completed. The docking of this modified compound which predicts how the compound will bind to the protein is shown in Figure 5.4.
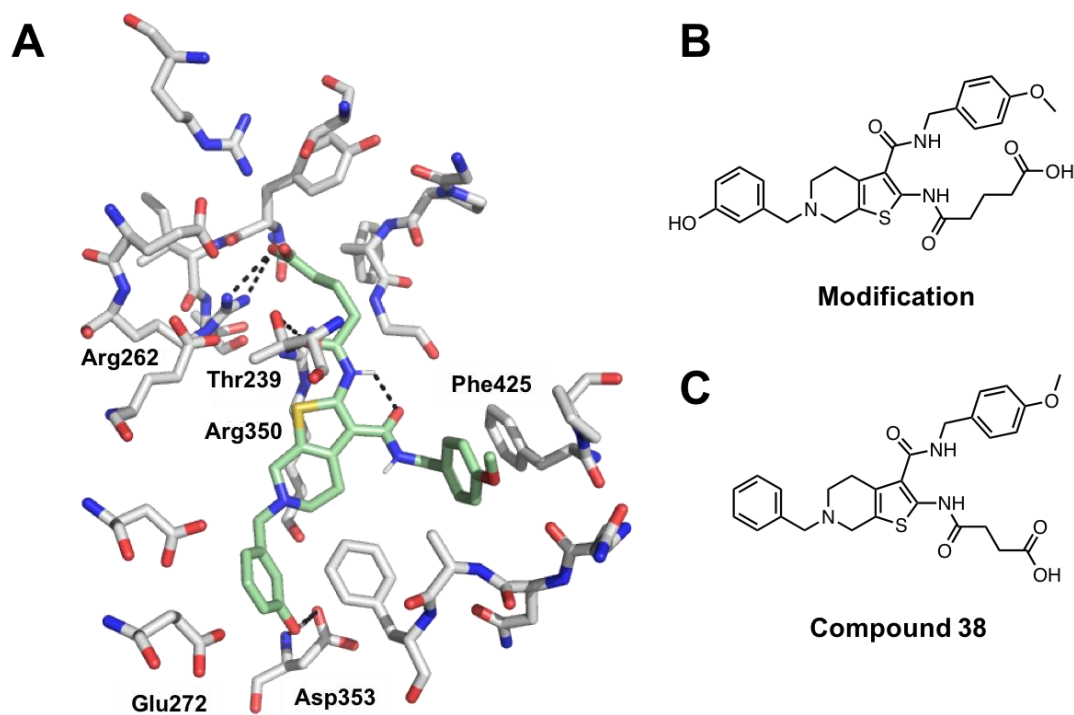
Figure 5.4: The predicted docking pose of the modified compound. *A) The predicted binding pose of the modified compound which is shown in green. The compound makes H-bonds to Arg262, Arg350, Thr239 and Asp353. B) The structure of the modified compound and C) compound 38.*

The addition of the hydroxyl group on the benzyl ring, as suggested by the docking, could improve the potency of the molecule even further by forming an extra H-bond to Asp353. **38** also had a 4-methoxybenzyl group at the R_2 position. Within the docking, the addition of the extra CH_2 group enabled a π-stacking interaction to be made with residue Phe425 which could have also improved the binding affinity of the compound. However, compound **35** which also contains the 4-methoxybenzyl group had the lowest IC_{50} therefore further compounds would need to be made to validate whether the presence of this group was important for binding.

Of the other three top scoring compounds (**31**, **32** and **36**) there were three different substituents at the R_1 position and different groups around the R_2 benzyl ring. At the R_1 position the functional groups were a tert-butyl group (**31**), H-atom (**32**) and a methyl group (**36**) suggesting that a range of groups can be tolerated at this 6-position with the compounds able to maintain activity against the enzyme by reducing ATP hydrolysis. Compound **31** had the highest IC_{50} at 41 μM which could indicate that the larger tert-butyl group is favoured at this R_1 position. The original hit compound **V-38** also contained the tert-butyl group at the 6-position and the docking pose for **V-38** is shown in Figure 5.5.
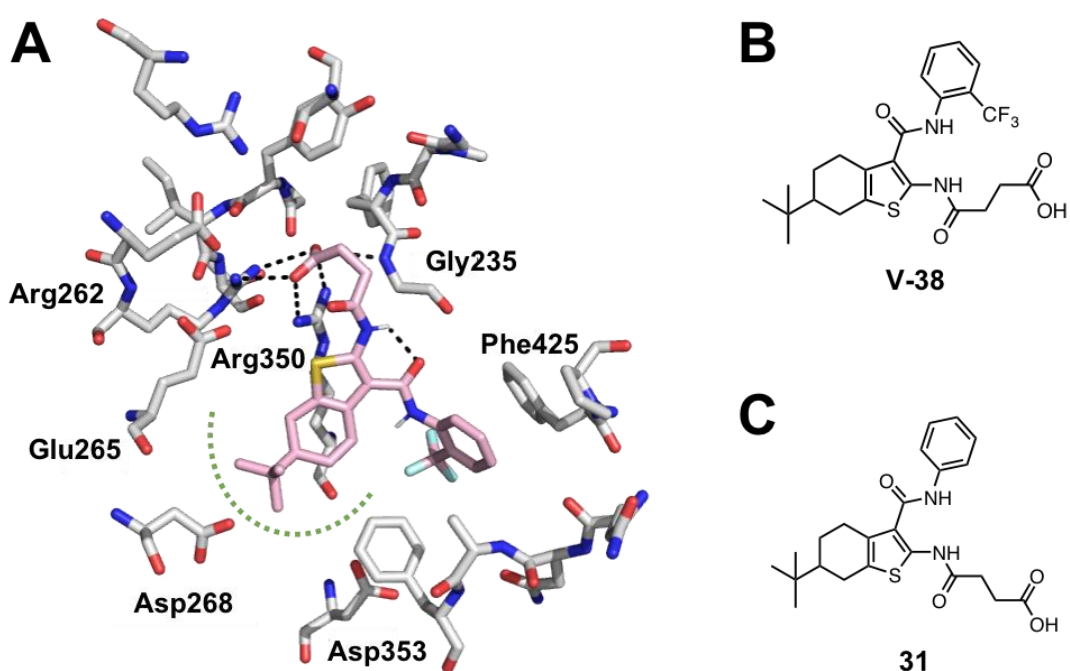
Figure 5.5: The docking pose of V-38. *A) **V-38** is shown in pink and the H-bond interactions it forms to the protein are shown as the black likes. The dashed green line represents extra space around the tert-butyl group which could indicate larger functional groups in this region would be tolerated at the binding site. The structures of B) V-38 and C) compound 31 are shown showing the only difference is the trifluoromethyl group which is present in V-38. This aromatic ring is predicted to form a π-stacking interaction with Phe425 but there is extra space in this region for more substituents to be added around the ring.*

Analysis of the docking for **V-38** in the binding site showed that there was lots of space around the tert-butyl group. This could be why the larger tert-butyl group had a slightly higher IC_{50} than compounds **32** and **36** as these would not fill the space as

well as compound **31.** The only difference between **V-38** and **31** is the presence of the trifluoromethoxy group on the R_2 phenyl ring so it was assumed that the binding poses of both compounds would be similar. Of the synthesised compounds, **31** was the only one which contained this tert-butyl group so more compounds would need to be made to allow a wider SAR study to be completed.

At the R_2 position for compounds **31**, **32** and **36**, there are differences in the substituents around the ring. For **31**, the phenyl ring does not contain any functional groups whereas for compound **32** there is a -chloro group at the 4-position and in **36** a -methyl group at the 3-position. Within the docking, the phenyl ring appears to be important for binding as it makes a π-stacking interaction to Phe425 which possibly mimics the π-stacking interaction which ATP makes with the protein.

Only eight compounds were synthesised which had varying groups at the R_1 and R_2 positions. This meant that only a small SAR study could be carried out. Due to the limited number of compounds, and the compounds containing similar functional groups, there are no clear SAR conclusions. A much bigger library of compounds, with a wider range of substituents, would need to be tested to enable this study to be completed. However, the compounds tested had IC_{50} values in the low micromolar range which represents a good starting point for future SBDD programs against V-ATPase.

5.2.5 Lead discovery centre V-ATPase inhibitors

During the optimisation of the synthesised compounds, three further V-ATPase inhibitors were provided by the Lead Discovery Centre (LDC) in Germany. Other than the compounds being highly potent, very little was known about the compounds including their structure and where they bound to V-ATPase. Therefore, if the protein could be purified, then the aim was to determine the cryo-EM structure of V-ATPase in the presence of the compounds to see whether there was any inhibitor density present. This would act as a double-blind study because without knowing either where the compound bound or the expected shape of the inhibitor density, there would be no bias when searching for the inhibitor density. For instance, if the compound was known to bind at a particular site then any additional density in this region would be attributed to the inhibitor and extra density present at other binding sites could be overlooked. By not knowing the location of the binding site, extra

146

density across all of the protein could be probed equally before determining whether this was due to the inhibitor.

Before the structure of the V-ATPase in the presence of the inhibitor could be determined, the compounds were tested for activity in the same ammonium molybdate assay which was used to calculate the IC_{50} values of the synthesised compounds. The compounds were tested at twelve different concentrations ranging between 50 pM and 1 μM. The molecular weight of all three compounds was estimated to be 300 kDa. Figure 5.6 shows the IC_{50} curves for the three compounds.
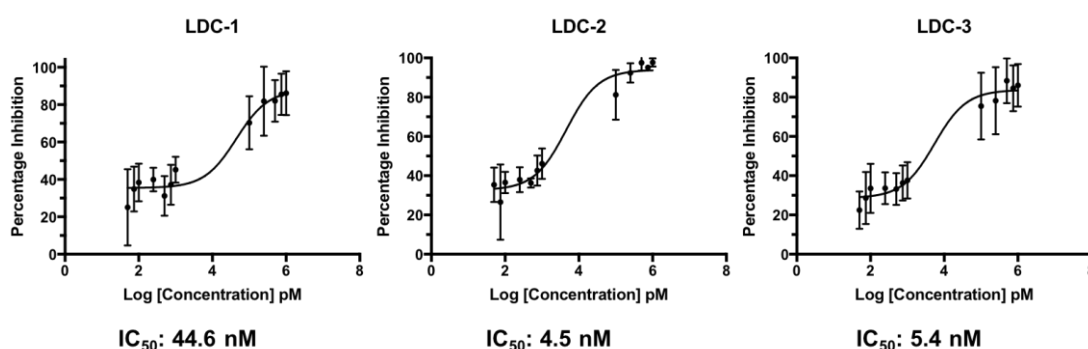
Figure 5.6: IC_{50} curves of the LDC compounds. *The IC_{50} curves of the three compounds provided by the LDC are shown. The compounds are highly potent as the IC_{50} values are in the low nanomolar range.*

All three of the LDC compounds had IC_{50} values in the nanomolar range with **LDC-2 and LDC-3** being highly potent with IC_{50}'s of 4.5 and 5.4 nM, respectively. As the structures of the compounds were unknown the SAR of the three compounds could not be established. It is also worth noting that the IC_{50} values are approximate as the molecular weight of all three compounds was estimated to be 300 kDa. The true IC_{50} values can be calculated when the actual molecular weight of the compounds is known. Nevertheless, all three compounds showed activity and were highly potent against V-ATPase so could therefore be taken forward to structural studies. The structural work will be discussed in detail in Chapter 6.

5.3 Discussion

Obtaining hit compounds to target V-ATPase using a structure-based drug design approach had not previously been reported in the literature. Chapter 4 described how virtual screening had been used to identify hit compounds which were predicted to interact at the ATP binding site within the V_1 domain. Hit compound **V-38** was shown to be the most promising, after determining the activity against purified protein and the toxicity using HEK cells. The next step was to try and improve the potency of the hit compounds which involved examining the predicted binding poses of **V-38** and the analogue compounds and designing new molecules which could make extra interactions with the protein, thus improving the binding affinity. This resulted in a further eight compounds, containing the same scaffold which consisted of the tetrahydrobenzothiophene core, being synthesised.

All of the eight compounds had IC_{50} values in the micromolar range with four compounds having IC_{50} values <100 μM. The most active compound had an IC_{50} of 41 μM (compound **31**). This is the first example of compounds which contain the tetrahydrobenzothiophene core having any inhibitory effect on V-ATPase and is one of the first vHTS drug discovery programs to be performed against this target. After one round of chemical synthesis, the IC_{50}'s of the compounds were in the mid micromolar range which is extremely promising and represents a good starting point for future drug discovery programs.

During the project, three compounds were also obtained from collaborators at the Lead Discovery Centre (LDC). They had previously designed potent compounds to act upon V-ATPase but did not have any structural information which showed how they bound. The activity of all three of the LDC compounds was much higher than the synthesised compounds. This is exemplified by the IC_{50} values. For instance, the most potent synthesised compound, **31**, had an IC_{50} of 40 μM compared to 4.5 nM for **LDC-2** which is therefore 10,000 times more potent. Improving the potency of the synthesised compounds to similar levels to those of the LDC compounds would likely take many more rounds of chemical synthesis and biological evaluation. This could be aided by obtaining structural information which showed how the compounds bound to V-ATPase rather than relying upon the *in silico* predicted binding pose. Visualising the inhibitor within the binding site would enable the interactions it made to the protein to be elucidated. Moreover, examining this structural information would allow any amino acids which were not involved in binding to be determined and the compounds

could subsequently be modified to utilise these extra interactions and improve the binding affinity of the compounds. Therefore, it was decided that obtaining this structural information would become the focus of the project rather than conducting further rounds of chemical synthesis at this stage. If the high-resolution structural information could not be obtained then a moderate resolution structure of V-ATPase in the presence of an inhibitor could still be useful for the design of new therapeutic agents.

V-ATPase is highly dynamic and can exist in three major catalytic conformations. One of the major aims of the project was to determine whether the presence of the inhibitor would have any effect on the overall structure of V-ATPase. The rationale was that an inhibitor could lock the protein into one particular conformational state, reduce the dynamic nature of the complex, thereby improving the stability of the complex and ultimately improving the resolution. If the resolution of the complex was still not good enough to visualise the inhibitor binding, then it would be interesting to see whether the compound was able to lock the complex into one particular catalytic conformation. If this was the case, then future SBDD programs could use this structural information to design compounds which would target this physiologically relevant conformation.

As the LDC compounds were much more potent than the synthesised compounds, the LDC compounds would initially be used in the structural studies of the full intact complex. This is because the higher potency means that less compound would need to be added to the protein to facilitate binding which would not hinder the contrast in the cryo-EM micrographs. Furthermore, due to the higher potency of the compounds it would be more likely that any effect of the compound could be determined within the structure. Chapter 6 will discuss how the protein was purified and the work which has been completed on obtaining the cryo-EM structure of inhibitor bound V-ATPase to elucidate whether inhibitor binding has had any effect on the structure of the complex.

6 Purification and structural studies of V-ATPase using cryo-EM

6.1 Introduction

The rotary family of ATPases, encompassing both V-ATPase and the F-ATP synthase, have been studied using cryo-EM. For F-ATP Synthase, one of the most recent cryo-EM structures (published in 2019) has shown that the complex exists as a dimer and occupies 13 different rotary states[87]. All of the states were resolved to resolutions of ~3.0 Å. In comparison, the highest resolution structure of the full V-ATPase complex is 6.9 Å which was published in 2015[160]. Therefore, with the recent improvements in detectors and image processing algorithms, the resolution of the full complex could not only be improved but more catalytic states could also be identified.

The work presented in Chapters 4 and 5 describe the design and synthesis of novel V-ATPase inhibitors. In order to improve the potency and selectivity of the compounds, and to continue the SBDD pipeline, structural information which shows how the compounds bind to the target site was sought. Furthermore, three compounds (**LDC-1**, **LDC-2** and **LDC-3**) which showed very high levels of potency, were also provided by the Lead Discovery Centre. Due to the high binding affinities of these compounds, it was decided that these compounds would first be used in the cryo-EM structural studies of V-ATPase, as it is more likely that the compound density can be visualised due to the high potency (Chapter 5). Once the pipelines and protocols for high resolution structure determination have been established, the structure of V-ATPase in complex with the synthesised compounds can subsequently be obtained. Therefore, the aim of the work presented in Chapter 6 is to firstly obtain purified V-ATPase and then to use cryo-EM to determine the structure of the complex with an inhibitor bound. The structural information gleaned could subsequently be analysed to determine whether the inhibitor has improved the resolution of the complex or locked it into a particular conformational state, thus guiding the design of new inhibitors to act upon V-ATPase.

6.2 Results

6.2.1 Optimising the purification of V-ATPase

Previous work within the Muench laboratory has purified V-ATPase from yeast without the need for introducing any affinity tags. This purification is based on a

protocol published in 1988 by Uchida *et al*[37] which uses glycerol gradients to isolate the purified protein. Full details of the purification are described in Chapter 2 (Section 2.3.4). The purification relies upon V-ATPase being the only ~1 MDa sized protein in the vacuolar membranes and having a high natural abundance (~25-50% of total vacuolar membrane protein). Due to the protein being successfully purified using this method by previous members of the group, this purification protocol was used. The purification was initially attempted using glycerol gradients to isolate the protein. The solubilised fraction was layered upon a 20-50% glycerol gradient which was centrifuged at 100,000g for 8 hours overnight. 750 μL fractions were collected and an SDS-PAGE gel was used to determine which fractions contained protein. Fractions with protein were then examined using negative stain EM to assess the quality of the protein obtained. Figure 6.1 shows an SDS-page gel and a representative negative stain image for two separate purifications attempts.
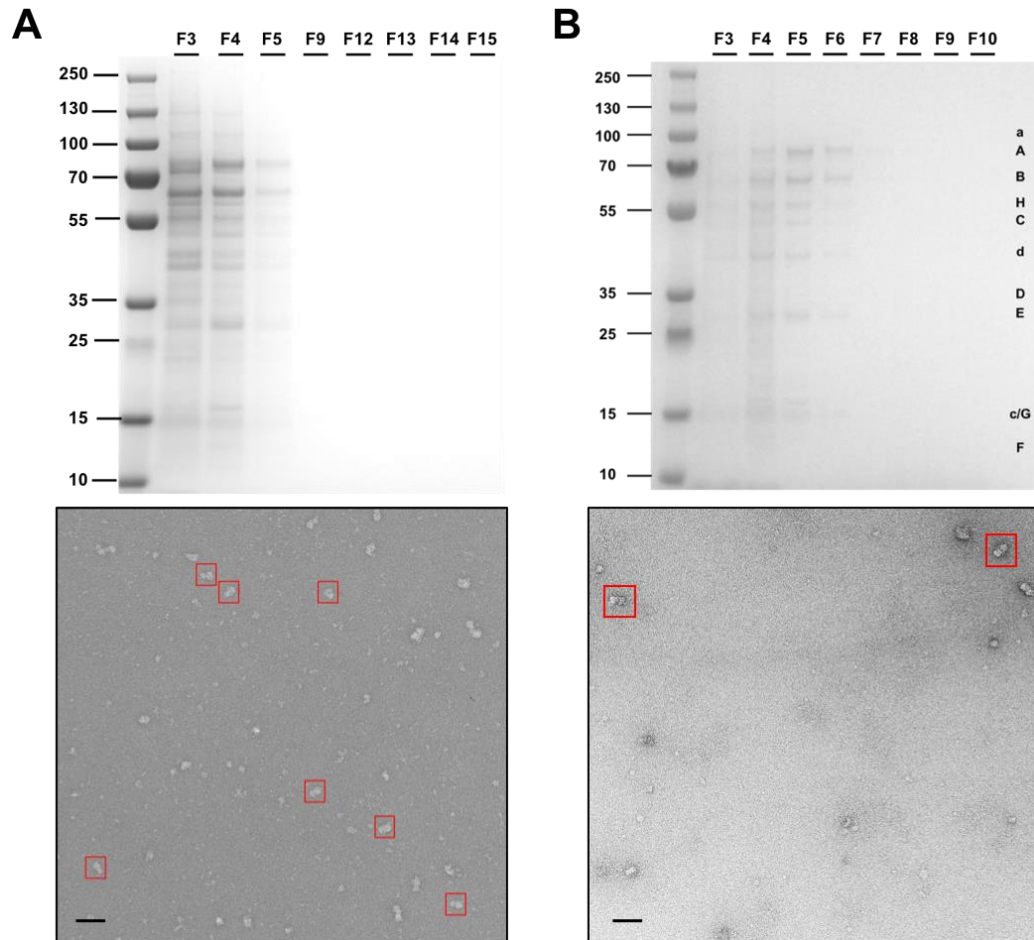
Figure 6.1: Purification using glycerol gradients. *A and B) show an SDS page gel and a micrograph from two separate purifications which used the glycerol gradient to isolate the protein. The amount of protein obtained was low and the negative stain showed that there were smaller proteins present showing that optimisation was needed. The red boxes in the micrograph show the locations of particles. Scale bars 50 nm.*

For the first purification, the SDS-PAGE gel showed protein was present in Fractions 3 and 4 (Figure 6.1A). When V-ATPase runs on a gel it breaks down into the individual subunits so multiple bands corresponding to the different subunits are expected. However, the gel shows extra bands which could be due to the protein degrading or due to the glycerol gradient not separating the solubilised fraction efficiently. The negative stain of Fraction 4 shows some particles which resemble V-ATPase but also smaller proteins indicating the sample is impure or degraded.

For the second purification attempt, Fractions 4-6 contained protein as the correct sized bands were present on the SDS-PAGE gel (Figure 6.1B). For this

purification, the centrifugation time for the glycerol gradient was increased to 12 hours from 8 hours thus allowing a better separation of V-ATPase from other proteins. The negative stain of Fraction 5 also indicates a cleaner sample. However, the amount of protein obtained was very low, with concentrations of ~0.1 mg/ml after concentrating the protein which equates to ~20 μg of total protein thus explaining why there were very few particles per micrograph. Therefore, the purification needed to be optimised in order to obtain enough protein to facilitate cryo-EM structure determination.

During the purification optimisation process, it was found that purifying the protein without freezing the cell pellet or vacuolar membranes at any stage of the protocol resulted in a higher yield of protein being obtained. This meant that the protein was purified within ~12 hours of the yeast cells being harvested which reduced the chances of the protein degrading. Moreover, the homogenisation step to break open the spheroplasts was also optimised. It was found that using a loose-fitting Dounce homogeniser instead of a tight-fitting homogeniser and changing the force and number of strokes were all important factors in governing the amount of vacuolar membranes isolated. After several rounds of optimisation, it was found that ~8 strokes with the loose-fitting homogeniser gave the optimum protein sample.

Purifying V-ATPase using glycerol gradients was not only unsuccessful in yielding enough protein for structural studies but the high glycerol concentrations within the sample would hinder later cryo-EM studies due to the large amount of background noise which would affect the contrast in the images. Therefore, size exclusion chromatography presented an alternative method of isolating V-ATPase and was subsequently attempted. The solubilised fraction was loaded onto a Superose 6 10/300 column which had been equilibrated using the same buffer which was used for the glycerol gradient, except it only contained 5% glycerol (10 mM Tris, 1 mM EDTA, 2 mM DTT, 0.5 mM PMSF, 5% glycerol and 0.01% DDM). The flow rate of the column was also optimised. Flow rates of 0.4 mL/min, which were initially trialled, did not effectively separate out the complex (Figure 6.2A). Under these conditions the 'cleanest' fraction resulted in ~110 μg of protein being obtained. Reducing the flow rate to 0.1 mL/min resulted in a cleaner sample and more of the in-tact complex being obtained, ~170 μg in the purest fraction (Figure 6.2B). This is highlighted in the gel filtration trace with the peaks being more clearly resolved than in the 0.1 mL/min trace. Figure 6.2 shows both the gel filtration trace, a negative stain micrograph and example 2D classes for the two purifications.
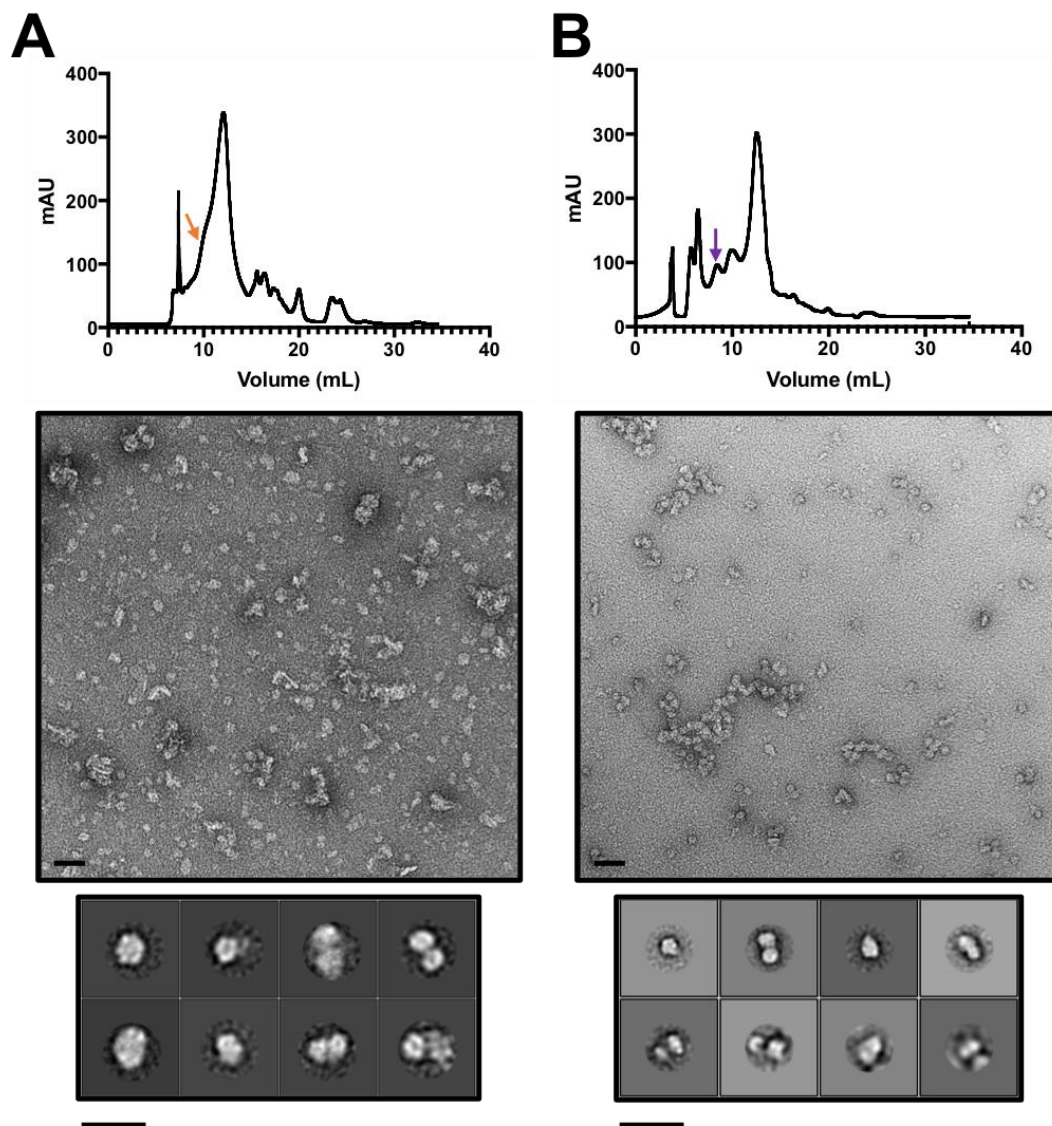
Figure 6.2: Purifications using size exclusion chromatography. *A size exclusion trace, example micrograph and 2D classes from two purifications using size exclusion chromatography. The arrows on the size exclusion trace indicate where V-ATPase is eluted which corresponds to the fractions which were used to make the negative stain grids. For A) the flow rate of the column was 0.4 mL/min. The negative stain micrograph and trace show that V-ATPase has not been separated efficiently. For B) the flow rate was reduced to 0.1 mL/min. The peaks on the size exclusion trace are now more defined and the micrographs show clear particles, albeit being slightly aggregated. The 2D classes for both purifications do not clearly show V-ATPase particles. The scale bar for the micrographs is 50 nm.*

For both purifications, negative stain images were collected and processed using RELION. Some of the resulting 2D classes, particularly for the 0.1 mL/min purification, vaguely resembled V-ATPase with the V_1 and V_0 domains visible. However, the quality of the classes was not sufficient to visualise any detail of the

individual subunits. Moreover, the negative stain at the lower flow rate showed that the protein was aggregated and therefore, the purification required further optimisation. The eluting buffer was subsequently optimised in order to obtain a higher quality protein. The buffers which were trialled are shown in Table 6.1.

Table 6.1: The different eluting buffers which were trialled during the purification optimisation.

	Eluting buffer composition	Comments
1	10 mM Tris, 1 mM EDTA, 2 mM DTT, 0.5 mM PMSF, 5% glycerol and 0.01% DDM	Buffer from original protocol
2	10 mM Tris, 1 mM EDTA, 2 mM DTT, 0.5 mM PMSF, and 0.01% DDM	Glycerol removed to improve contrast in cryo-EM
3	50 mM Tris, 150 mM NaCl and 0.01% DDM	Buffer from literature where size exclusion chromatography was used to purify V-ATPase
4	50 mM Tris, 150 mM NaCl, 1 mM EDTA, 2 mM DTT, 0.5 mM PMSF and 0.01% DDM	Optimised buffer composition to improve contrast and reduce aggregation

The first buffer trialled (**1**) was the original buffer used in the protocol and contained 5% glycerol. The presence of the glycerol reduced the contrast in the cryo-EM micrographs which will be discussed in Section 6.2.2 so would need to be removed from the buffer. Buffer **2** removed the glycerol but kept all of the other components the same but the protein was still aggregated. Therefore, literature searches were conducted to see what buffer was used in published V-ATPase structures. Buffer **3** was the same as the one which was used in the paper which showed that V-ATPase existed in three rotational states[160]. However, the protein looked partially degraded when it was examined by negative stain EM. The optimal eluting buffer was buffer **4** which was composed of 50 mM Tris, 150 mM NaCl, 1 mM EDTA, 2 mM DTT, 0.5 mM PMSF and 0.01% DDM. This buffer contained no glycerol so would not hinder the later cryo-EM studies. Moreover, the buffer contained 150 mM NaCl, which was used in published V-ATPase structures, and the negative stain showed that the aggregation was reduced. This buffer composition also resulted in the highest yields of protein obtained, which was ~1 mg after the best fractions had been pooled and concentrated. Example negative stain micrographs from this eluting buffer are shown in Figure 6.3 and are consistent between different purification attempts.
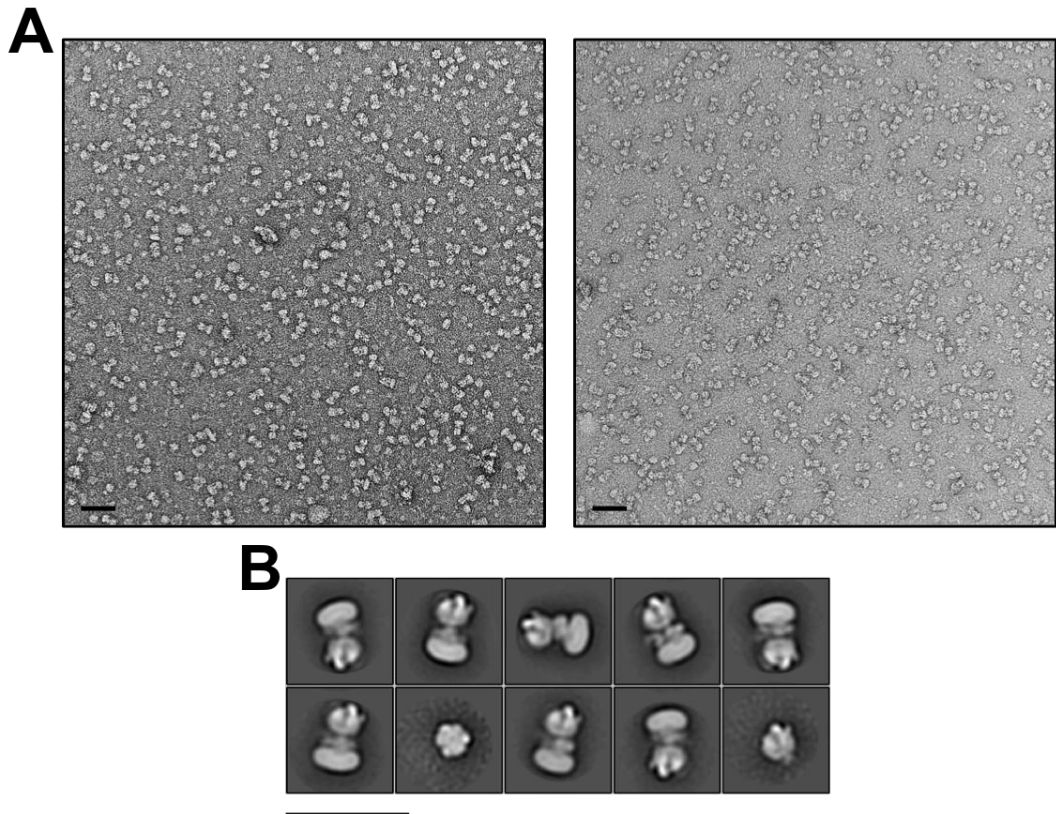
Figure 6.3: Negative stain micrographs of V-ATPase isolated from size exclusion chromatography. *A) Two example micrographs of V-ATPase isolated using eluting buffer 4 from two separate purification attempts. The micrographs show a good distribution of intact protein particles indicating a high quality sample. B) Example 2D classes from one of the data sets. The classes clearly resemble V-ATPase and show in detail the two separate V_1 and V_0 domains. Scale bars 50 nm.*

The negative stain micrographs showed a monodisperse distribution of mostly in-tact V-ATPase with no aggregation, thus indicating a high-quality sample. Therefore, a small data set of 120 micrographs were collected on a F20 microscope fitted with a FEI Ceta CCD camera. The micrographs were auto-picked using the Laplacian-of-Gaussian feature implemented in RELION3.0 resulting in ~60,000 particles being picked. After two rounds of classification, the resulting 2D classes from ~30,000 particles showed the in-tact V-ATPase complex with the V_1 and V_0 domains clearly defined. There were also two classes which showed the V_1 domain on its own indicating that some of the complex could have degraded. Moreover, the C, H and *a* subunits which reside between the two domains were also visible highlighting that the sample could be taken forward to cryo-EM.

6.2.2 V-ATPase cryo-grid optimisation

To continue with the structural studies, and allow a high-resolution structure to be determined, the protein was taken forward to cryo-EM. The first step was to optimise the cryo-grids which occurred in parallel to the purification optimisation. Ideally the protein would have a good distribution within the ice and reside in a variety of different orientations enabling the 3D reconstruction to be determined. Before making cryo-grids fractions from the size exclusion column, which contained the best protein, were pooled and concentrated. The concentration of protein obtained varied depending on the purification but was typically between ~1-3.5 mg/mL.

Cryo-grids were prepared using the Vitrobot Mark IV device (Section 2.3.10) with a range of different grid types and protein concentrations (Figure 6.4). The grids were subsequently screened using the Titan Krios microscopes. Previous work in the Muench group had involved determining the structure of *Manducca sexta* V-ATPase using cryo-EM[260]. Within this study, data were collected on carbon-coated lacey grids with a protein concentration of ~1 mg/mL. Grids were firstly prepared using carbon-coated lacey grids. However, the protein had poor good distribution within the holes and was clumped towards the carbon edges which was unsuitable for data collection (Figure 6.4A). Quantifoil 1.2/1.3 grids were subsequently prepared with a protein concentration of ~1 mg/mL with the buffer either containing 5% glycerol or no glycerol at all. The contrast within the micrograph in the presence of 5% glycerol was very poor as it was very difficult to distinguish any particles within the ice (Figure 6.4B). In the absence of glycerol, protein molecules were visible but were again clumped towards the edges of the holes (Figure 6.4C). This suggested the concentration needed to be increased to improve the distribution of protein in the ice.

In subsequent purifications, the amount of protein obtained was improved which resulting in the concentration increasing to 2.0-2.5 mg/mL, as measured by a nanodrop. As the protein had displayed a high affinity for the carbon support layer, which is common for many membrane proteins, gold ultrafoil 1.2/1.3 grids were tested. The theory was that the protein would have a lower affinity to the gold support than the carbon support so more protein molecules would reside within the ice. As shown in Figure 6.4D, there were more protein molecules within the ice but they were still clumped towards the edges of the holes and were unsuitable for data collection. Quantifoil 1.2/1.3 grids were prepared with protein concentration of 2.5 mg/mL. The increased concentration enabled the protein to saturate the carbon support layer, allowing more protein molecules to reside within the holes. Interestingly, within the

same grid there was a varying distribution of protein in holes with different ice thicknesses. In thin ice, the protein was pushed out from the centre and was close to the carbon, somewhat resembling the 'halo-effect' (Figure 6.4E). Although some protein molecules could be picked. In comparison, within the thicker ice there was a much better distribution of protein molecules which would enable more particles to be picked per micrograph in subsequent data collections (Figure 6.4F).
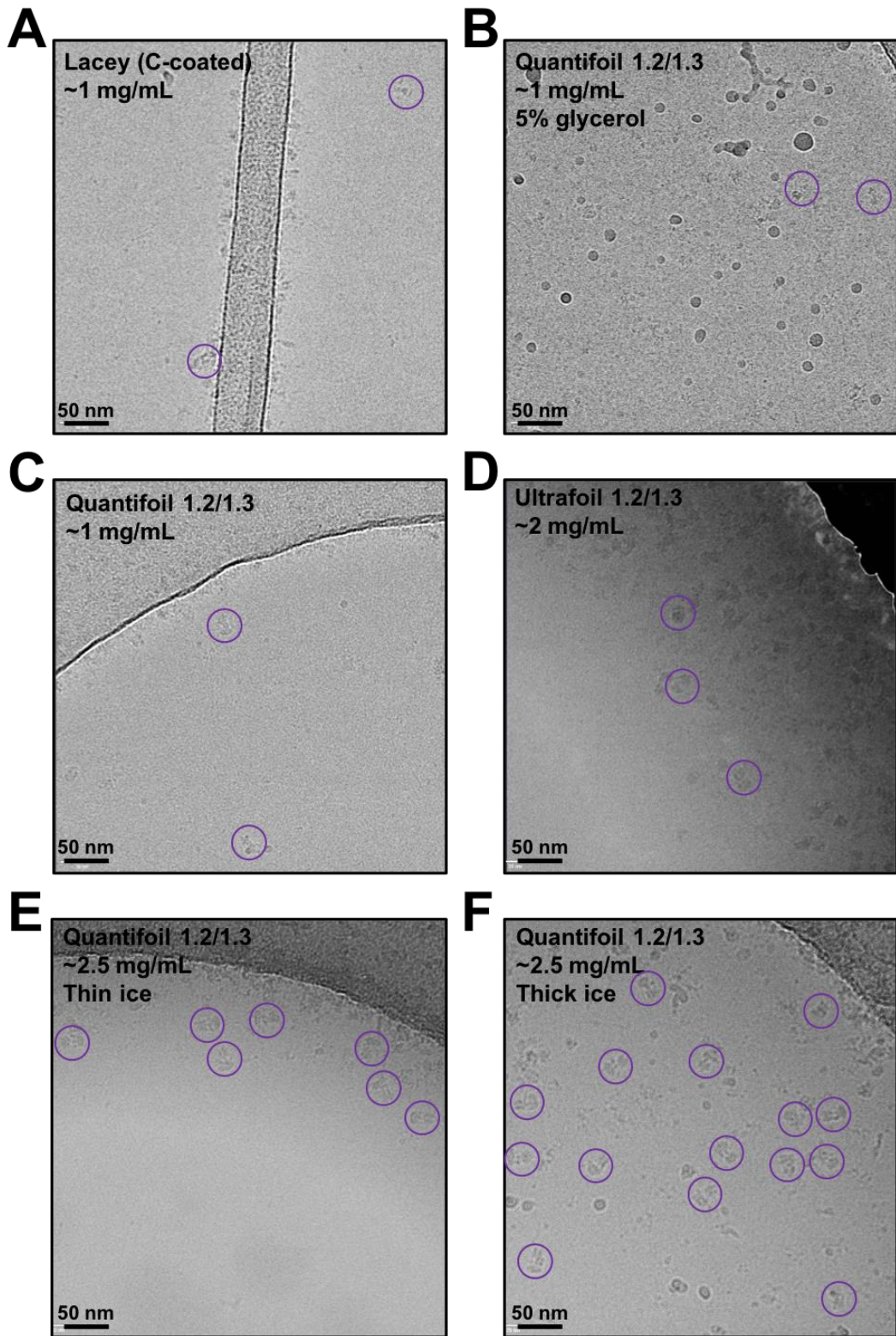
Figure 6.4: V-ATPase cryo-grid optimisation. *A-F) A range of different grid types, protein concentrations and ice thicknesses were screened. F) shows the best distribution of protein molecules within the ice. The purple circles show the position of protein molecules within the micrograph.*

After optimisation of the cryo-grids, it was found that applying high concentrations of protein (~2.5 mg/mL) to quantifoil 1.2/1.3 grids resulted in the best distribution of protein within the ice. These grid conditions were suitable for data collection especially within the thicker areas of the grid. V-ATPase protein molecules can be seen which indicates that the protein remains intact during the grid preparation and is not interacting unfavourably at the air-water interface. Two preliminary data sets were collected in the absence and presence of a small-molecule inhibitor.

6.2.3 Preliminary cryo-EM data sets

After optimisation of the cryo-grids, two preliminary overnight (~15 hour) data sets of apo and inhibitor-bound V-ATPase were collected resulting in 698 and 1,088 micrographs being collected, respectively. Images of apo V-ATPase were recorded on a K2 detector whereas inhibitor-bound V-ATPase was collected on the Falcon III, operated in integrating mode. The concentration of the protein applied to both grids was 2.5 mg/mL and there was a good distribution of protein molecules within areas containing the thick ice on the grid. For the inhibitor bound, **LDC-2** was used at a final concentration of ~200 nM which was ~40 times the IC_{50} of the compound. The micrographs were subsequently motion-corrected using MotionCor2[207] and the CTF was estimated using Gctf[210]. All subsequent processing was performed in RELION3.0[214]. The resulting micrographs were manually picked as RELION's autopicking algorithm failed to pick all of the particles or picked too much noise. The subsequent particle stacks, 20,847 particles for apo V-ATPase and 9,205 particles for **LDC-2** bound V-ATPase, underwent rounds of 2D classification and the resulting 2D classes showed that the protein adopted a wide range of orientations within the ice. The classes clearly showed the intact complex and also contained some secondary structure details. Example micrographs and 2D classes are shown in Figure 6.5.
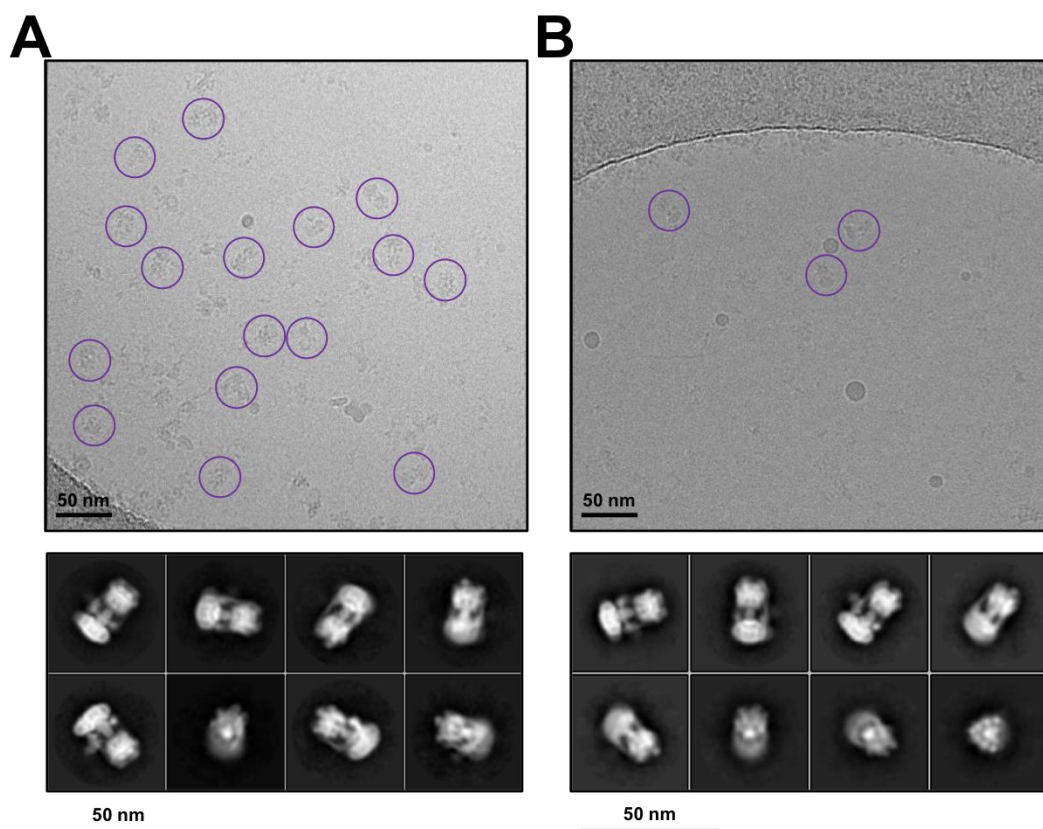
Figure 6.5: Preliminary cryo-EM data sets of V-ATPase. *An example micrograph and 2D classes of A) apo V-ATPase and B) LDC-2 inhibitor-bound V-ATPase. There is a good distribution of particles within the apo-V-ATPase data set but much fewer particles within the inhibitor-bound V-ATPase micrograph. This could be due to the protein being diluted upon the addition of the compound. For both data sets the classes show that the protein adopts a range of orientations within the ice and there is no problem with preferred orientation.*

The particles resulting in the best 2D classes were then subjected to 3D classification using an initial model which was generated for both data sets. For the apo data set, two different conformational states of V-ATPase were identified with resolutions of 8.7 Å and 8.1 Å from 1,359 and 2,098 particles, respectively. The local resolution was higher for both states (~6 Å). For **LDC-2** bound V-ATPase only one state was identified and this achieved a resolution of 9.1 Å from 2,857 particles. For all of the maps obtained, the resolution of the V_1 domain was higher than the global average. For the two different states in the apo V-ATPase structure, the resolution within this region was sufficient to visualise the secondary structure of the V_1 domain. The different maps showed the AB domain in the different catalytic states. This is exemplified by the AB dimer located above the H-subunit which was in the tight and open conformations in State 1 and 2, respectively. In comparison, the V_0 domain was

always poorly resolved and the resolution was not good enough to visualise secondary structure in this region. The local resolution maps are shown in Figure 6.6.
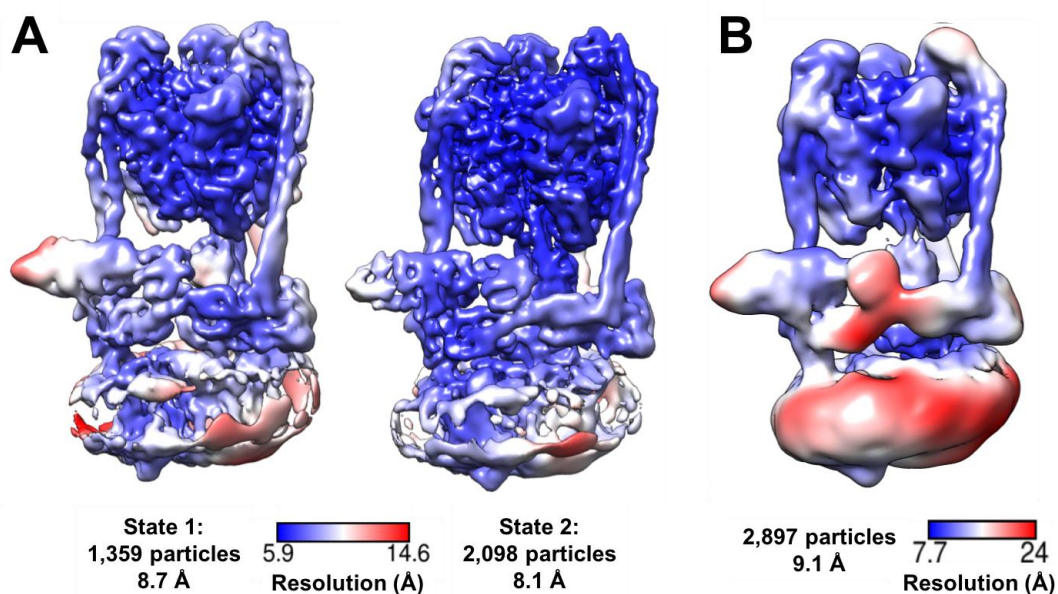
Figure 6.6: Local resolution maps of the preliminary data collections*. A) The local resolution of both state 1 and state 2 from apo V-ATPase which was coloured on the same resolution scale. The V₁ domain is at a higher resolution than the global average and was good enough to see the secondary information of the protein. B) The local resolution of LDC-2 bound V-ATPase in the one state which was identified. The highest local resolution was again in the V₁ domain.*

It was interesting to note that for the inhibitor-bound data set, only one class was identified whilst for the apo data set, two classes were identified through 3D classification. This could be due to the smaller number of particles which went into the 3D classification for the inhibitor bound data set which was 2,857 particles compared to 3,457 particles for the apo data set. Whilst the low number of particles makes it difficult to draw clear conclusions, the reduced number of states identified could be indicative of the inhibitor trapping the complex into one particular state thereby limiting the number of states the complex resides in. A much larger data set would need to be collected to confirm this with many more particles (10,000's-100,000's) needed in the classification steps.

The initial preliminary data sets were promising as sub-nm resolution was unexpectedly obtained from very low numbers of particles (~2,000). Therefore, larger data sets would need to be collected in order to, not only improve the resolution, but

also to gain insights into whether the compound is having an effect on the conformational states of the complex.

6.2.4 Cryo-EM of inhibitor-bound V-ATPase

A 72-hour cryo-EM data set of **LDC-2** inhibitor-bound V-ATPase was collected with the aim being to increase the number of particles, thereby improving the resolution of the complex. Cryo-grids were prepared using the Vitrobot as described in the Methods Section 2.3.10. Prior to making the grids, the sample was incubated with 200 nM **LDC-2** compound for ~1 hour prior to freezing. Data were recorded on an FEI Titan Krios microscope equipped with a Falcon III direct electron detector (integrating mode) operating at 300 kV. From the 72-hour data collection, 6,916 micrographs were collected and the motion correction and CTF estimation was performed on-the-fly using MotionCor2[207] and Gctf[210], respectively. The preliminary data sets had both been manually picked due to the autopicking algorithm within RELION3.0[214] failing to adequately pick all of the particles. Due to the large number of micrographs in this data set, manually picking was not considered to be a viable option. Therefore, crYOLO was used to pick the data.

crYOLO (You Only Look Once) is a deep learning object detection system which relies upon machine learning to accurately pick particles from micrographs[216]. A general model is available which has been trained on many data sets. The general model was used on a smaller subset of 576 inhibitor-bound V-ATPase micrographs where the particle diameter was specified as 250 Å. This resulted in ~21,000 particles being picked which were 2D classified in RELION3.0. The classes showed that many of the 'particles' picked were in fact junk which did not represent V-ATPase (Figure 6.7A). Moreover, the expected side views were not centred around the full complex but were instead centred over the V_1 or V_0 domains which would subsequently hinder the data processing. Therefore, the particle picking model was trained on the V-ATPase data set. By manually picking 22 micrographs, the general model was updated. After using the trained picking model, ~10,000 particles from the 576-micrograph subset were auto-picked. The 2D classes showed that the particles were centred around the full complex and less 'junk' particles were picked (Figure 6.7B).
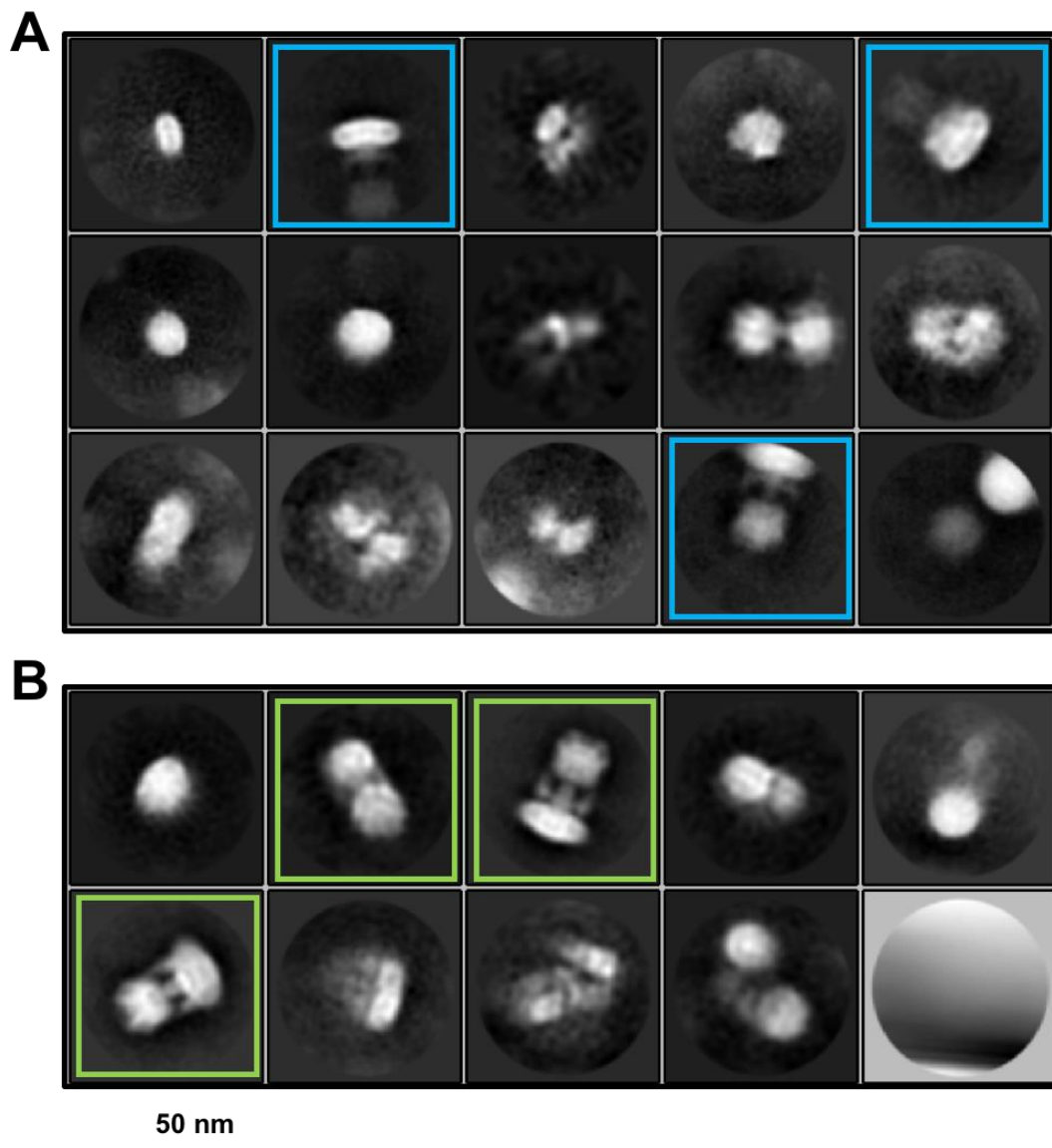
Figure 6.7: Optimising the picking of particles using crYOLO. *A) The 2D classes from the general model which had not been trained on the V-ATPase data set. The blue boxes highlight the side views which were not centred on the full complex. B) The 2D classes after the model had been trained. The green boxes show the side views which are now centred around the full V-ATPase complex.*

All 6,916 micrographs were picked using the trained model to give a particle stack containing ~133,000 particles which were extracted into a 400-pixel box and down-sampled to 200 pixels. After 2D classification, ~43,000 particles remained which allowed an initial model to be generated. The 3D refinement achieved a resolution of 8.9 Å after sharpening. Moreover, continuing the refinement from the last iteration, using a soft mask and solvent flattened FSCs, also helped to improve the resolution to 8.7 Å after post-processing (Figure 6.8).
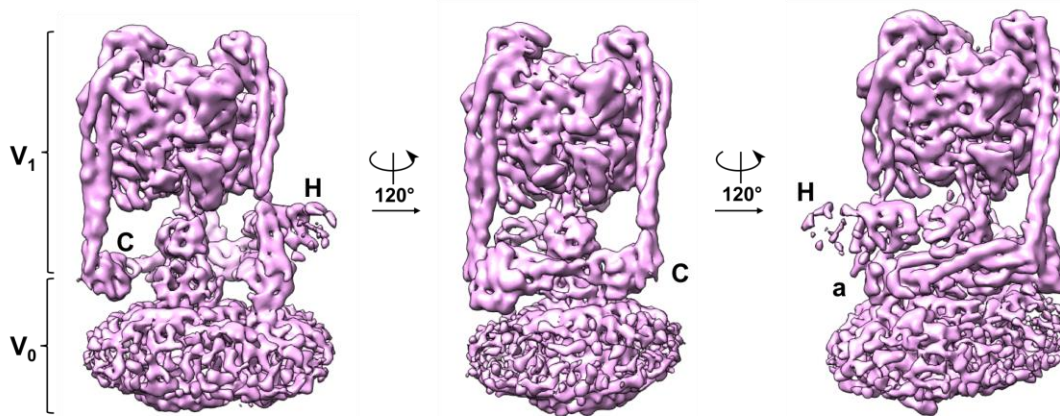
Figure 6.8: Cryo-EM map of LDC-2 bound V-ATPase. *The map is shown in three different orientations which have been rotated through 120° along the central axis of the protein. The density for the V_1 domain is strong with the α-helical secondary structure becoming visible. The density for the H subunit is weaker and the V_0 c-ring density does not show any secondary structure information.*

As shown in Figure 6.8, the density within the V_1 domain is sufficient to visualise the α-helices in the secondary structure of the protein particularly within the AB hexameric ring, C and *a* subunits, and the stators. In comparison, the density for the V_0 domain, which includes the *c*-ring, is not as well resolved and it is challenging to place the individual helices inside of the map. What was interesting to note was that the density for the H-subunit was weak compared to the C and *a*-subunits which help to link the V_1 and V_0 domains together. Moreover, it appeared that the open catalytic AB interface resided above the H and *a*-subunits but it was challenging to identify the loose and tight catalytic states. This is because the density for the different states would have been averaged together during the 3D refinement as the data had not yet been classified. In order to identify further states, the particles underwent two rounds of 3D classification, resulting in six classes (Figure 6.9).
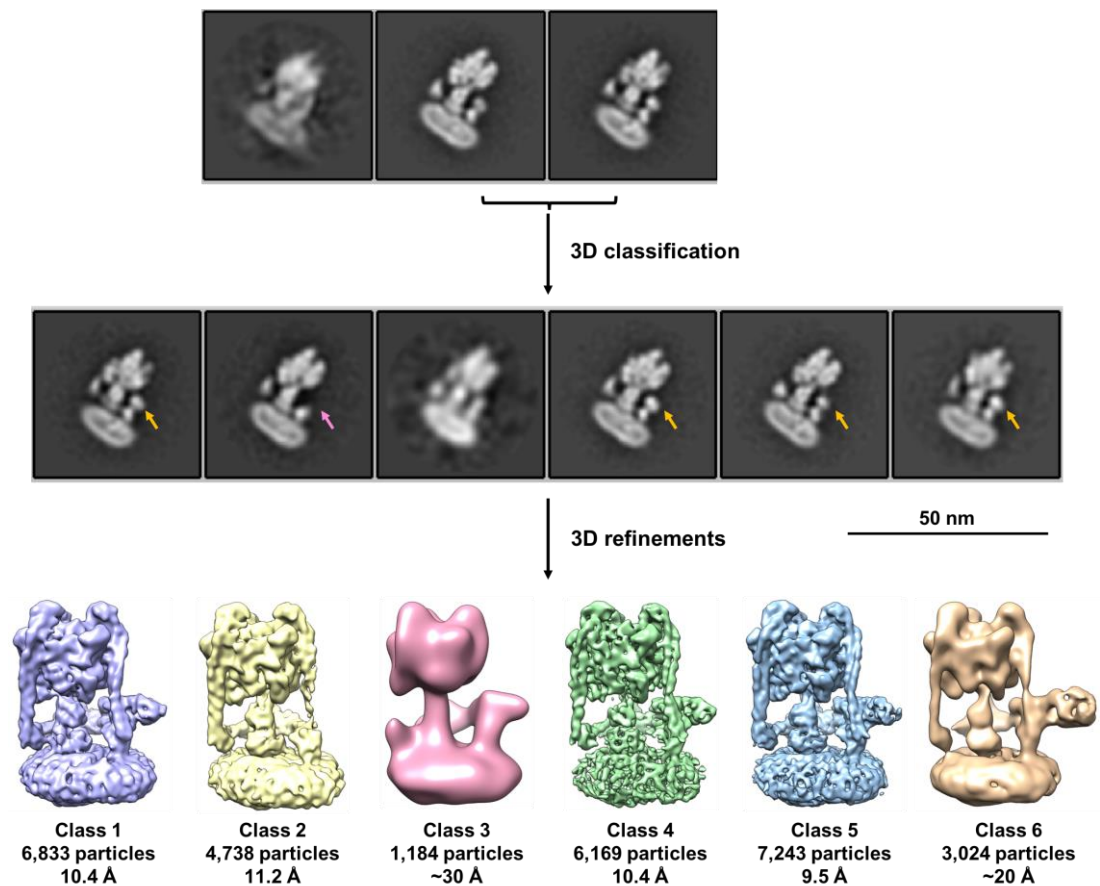
***Figure 6.9: The processing pipeline for LDC-2 bound V-ATPase**. Two rounds of 3D classification were performed. The two 'good' classes from the first round underwent a second round of classification where the data was split into 6 classes. The individual slice-throughs for each class are shown. The arrows point to the density for the H-subunit which was missing in Class 2.*

Following 3D classification, the particles which made up each class were refined. Two of the classes (Class 3 and Class 6) were poorly resolved which could be due to them having degraded or poorly aligned particles. Classes 1, 4 and 5 all contained greater than 6,000 particles and achieved resolutions of 10.4 Å, 10.4 Å and 9.5 Å, respectively. It was interesting to note that Class 2, which contained 4,738 particles and was resolved to 11.2 Å, was missing the H-subunit. No V-ATPase structures have been reported with the H-subunit missing in the literature or EMDB which is surprising given that this class represented ~19% of the particles (of the four 'good' classes). 2D classes of Classes 1 and 2 were generated which showed that the particles adopted a wide range of orientations within the ice (Figure 6.10).
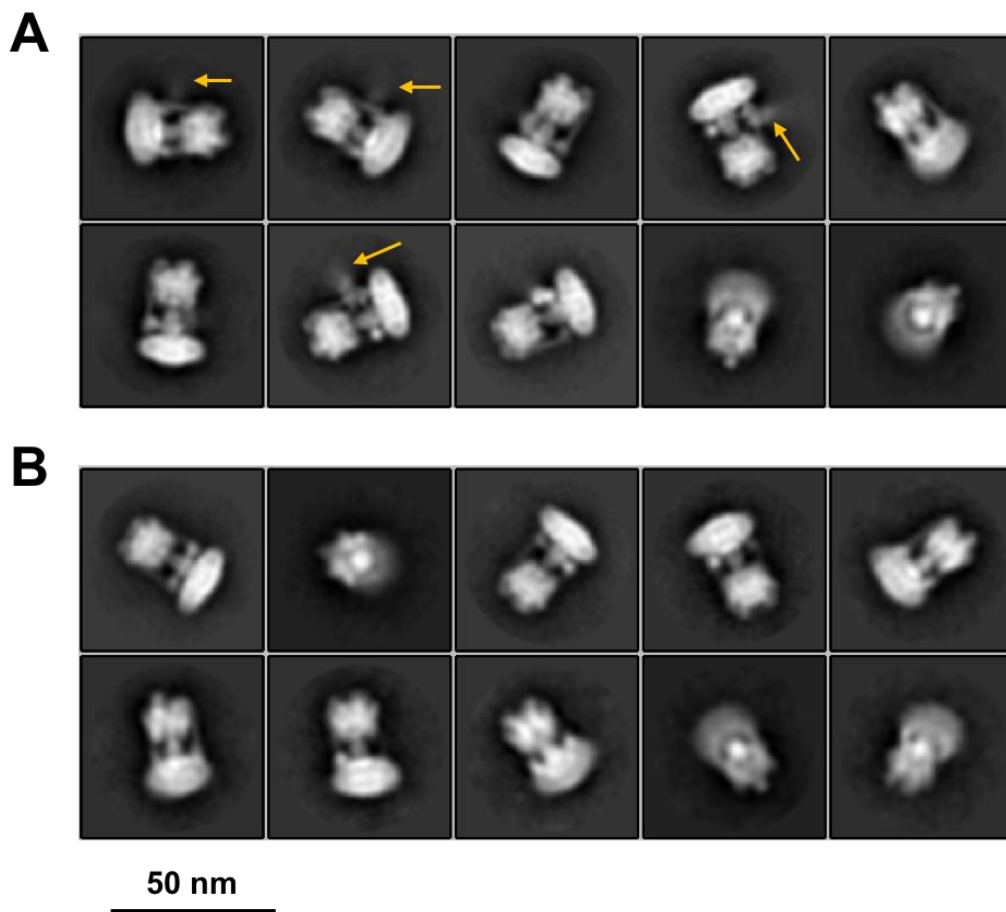
Figure 6.10: 2D classification of two of the different states. *The 2D classes for A) State 1 and B) State 2 are shown. The orange arrows indicate point to visible density for the H-subunit which is only present in Class 1. Class 2 has no visible density for the H subunit within the 2D classes which is consistent with the 3D map.*

Density for the H-subunit was visible in the 2D classes from Class 1 but was missing from the 2D classes generated from Class 2. The reason for the H-subunit not being present could be due to the complex degrading/breaking apart. However, another reason could be due to the small molecule, **LDC-2**, interacting with the protein and causing it to dissociate (see discussion 6.3). Similar to the preliminary data sets, it was noted that the resolution of the V_1 domain appeared to be higher than the global average for all of the classes identified. Therefore, attempts were made to improve the resolution of this region by creating a soft mask around the V_1 domain and then continuing the refinement using this mask to undergo a focussed refinement. In all cases, the resolution of the maps was slightly improved with some secondary structure visible. An overview of the maps before and after focussed refinement around the V_1 domain are shown in Figure 6.11.
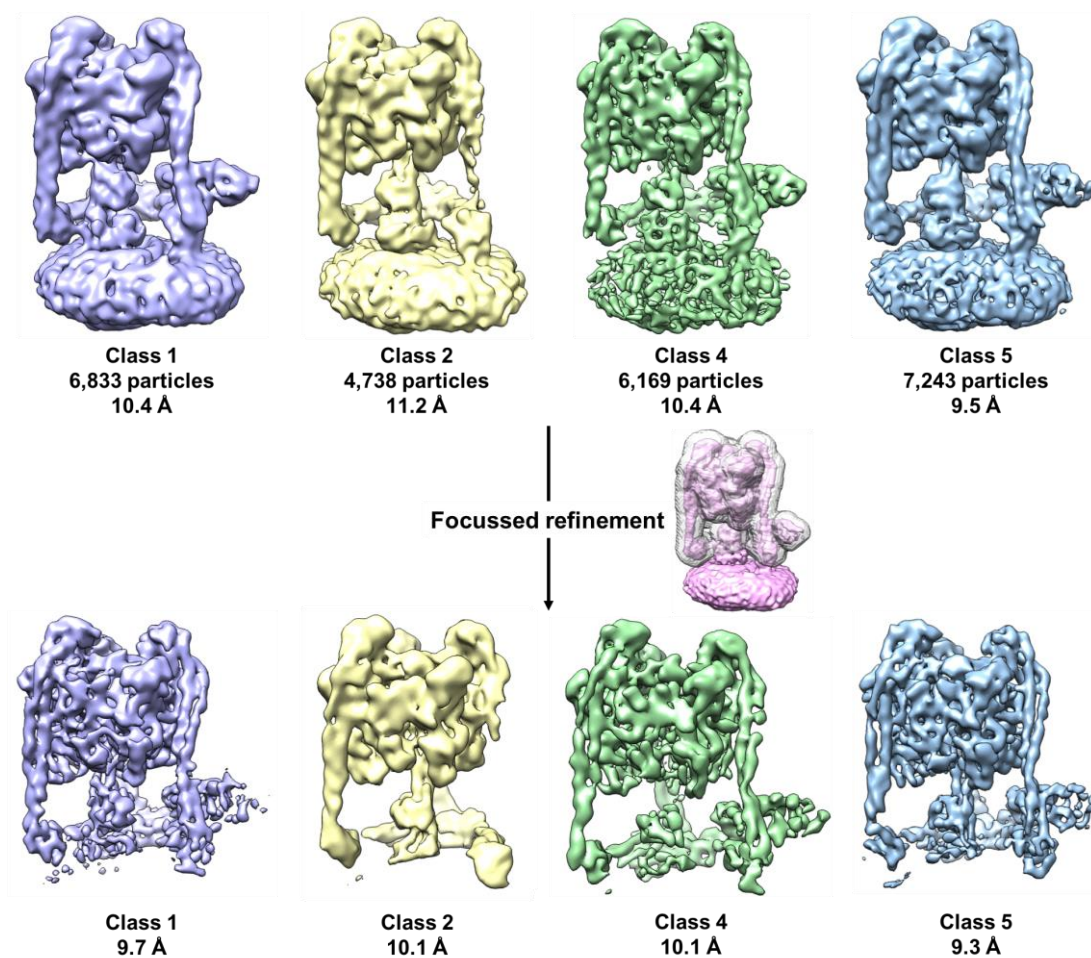
Figure 6.11: Focussed refinement of the V₁ domain. *The maps and resolution for the four classes are shown before and after the focussed refinement. The pink map next to the focussed refinement arrow is surrounded by the transparent mask which was used to improve the resolution. In all cases the resolution of the V₁ domain was improved after the focussed refinement.*

For Classes 1, 4 and 5, the resolution permitted some of the α-helices to be resolved within the map. However, the resolution was not sufficient to resolve β-sheets or contain any side chain density. The maps could subsequently be compared to the published cryo-EM structures which showed that the complex resides in three catalytic conformational states[160]. The catalytic states show that the AB subunits can reside in the open, loose and tight states at different positions within the hexameric ring in each conformation[160]. This is summarised in Figure 6.12.
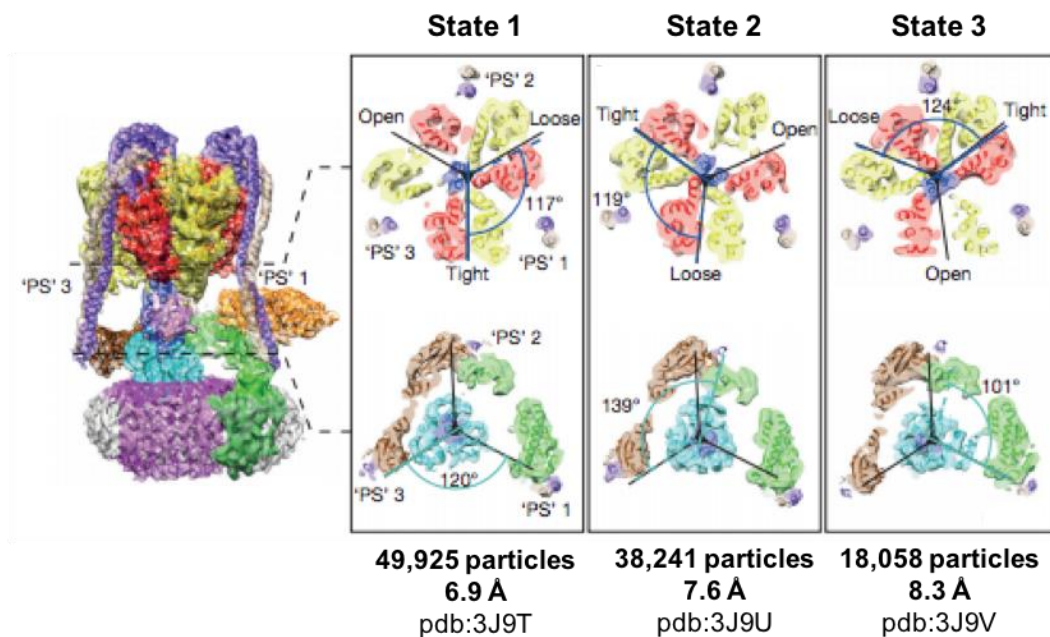
Figure 6.12: Conformational states of V-ATPase. *The different states of V-ATPase which were published in 2016. The slice throughs of each state highlight the position of the AB domains. The number of particles and resolution obtained is also stated. Adapted from*[160]*.*

After processing the data, the result is an EM density map. The map can then be used to fit in models which shows the positions of the α-helices, amino acids etc. To compare the EM maps with the published structures, the models (pdb files) were fit into the map and inspected to see which of the published states they best represented. For clarity, 'State' describes the published models which show the three different conformational states of the protein. Whereas, 'Class' describes the classes obtained from the experimental data within this study, summarised in Table 6.2.

Table 6.2: Conformational state analysis of LDC-2 bound V-ATPase.

	Above H-subunit	Above C-subunit	Above *c*-ring	Comparison to published state
Class 1	Loose	Open	Tight	State 1
Class 2	Open	Tight	Loose	State 2
Class 4	Open	Tight	Loose	State 2
Class 5	Open	Tight	Loose	State 2

170

It was found that Class 1 was similar to the published State 1 whereas Classes 2, 4 and 5 were most similar to the published State 2. In State 1, the open conformation of the AB domains resides above the C-subunit, the loose conformation above the H/*a*-subunits and the tight conformation above the *c*-ring. In the published structure, this was the most populated state which contained 47% of the particles (49,925). In comparison, Class 1 contained only 27% of the particles (6,833) which make up the four classes. The other classes (Class 2, 4 and 5) were most similar to the published State 2, where the open site of the AB domain was directly above the H/*a* subunits. As mentioned previously, Class 2 did not contain any density for the H-subunit which represented a novel conformation of V-ATPase. The conformations and position of the AB domains within Class 2 are consistent with the published State 2. Interestingly, the distribution of the conformational states was shifted compared to the published study and none of the classes resembled State 3. For all classes, the published model for the V_1 domain was flexibly fit into the map using MDFF[232]. To enable accurate comparisons between the published and experimental structures, the models were overlaid and coloured by Cα r.m.s.d. (Figure 6.13).
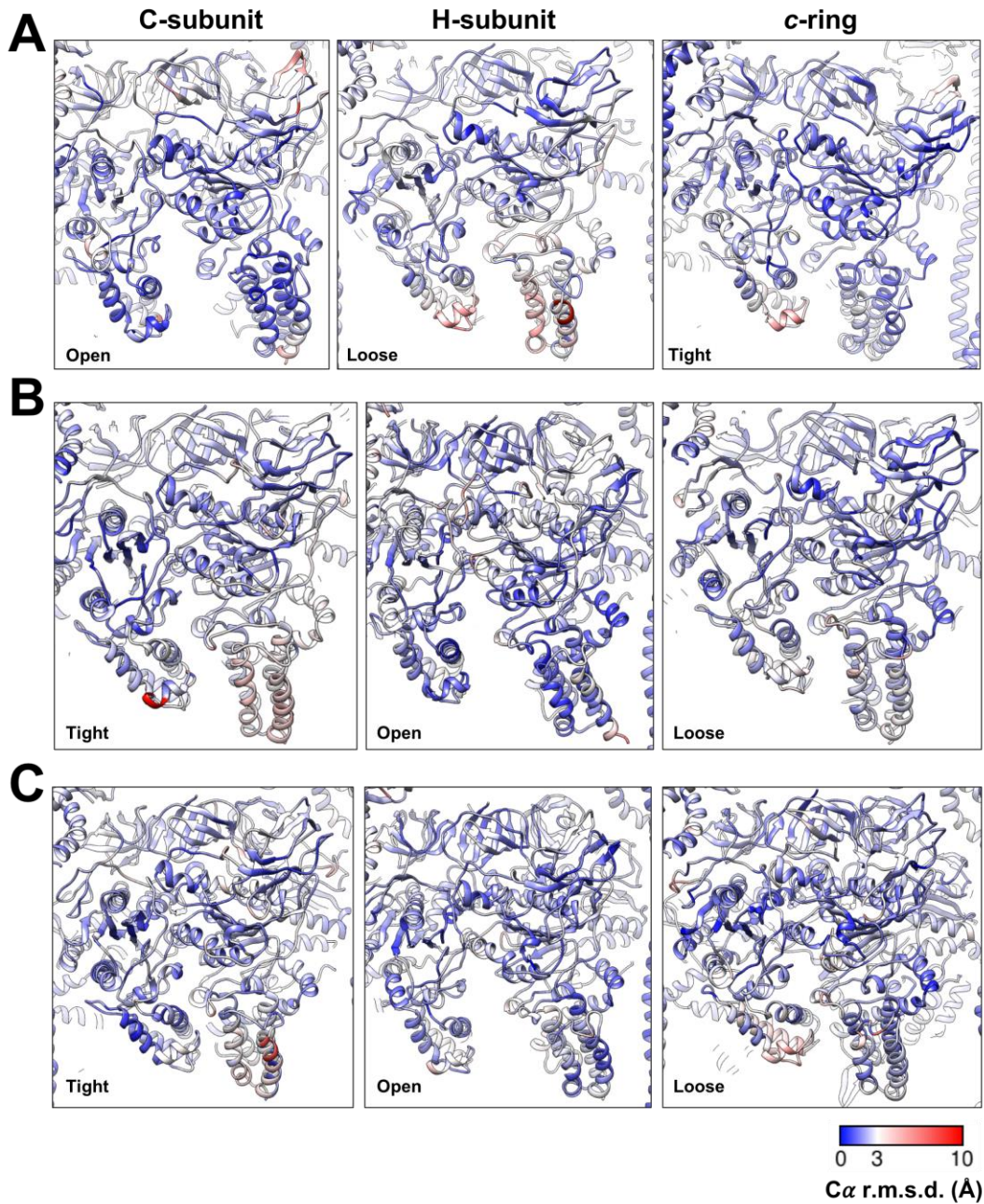
Figure 6.13: Comparison of the classes to published models. *The Cα r.m.s.d. colourings for the AB domains in A) Class1, B) Class 4 and C) Class 5. The C-subunit, H-subunit and c-ring indicate the subunits which the AB domains are located directly above. Generally, there is a good overlap between the two models (<3 Å). In the Class 1 loose state and Class 4 tight state there are differences in the positions of the outer helices on the B-subunit. However, this could be due to resolution differences between the maps.*

There was generally a good overlap of the Cα atoms (<3 Å) between the fitted and published models within the AB subunits. However, for the Class 1 loose and

Class 4 tight AB conformation, there were differences in the positions of the outer helices on subunit B. This is exemplified by the Cα r.m.s.d. being ~5 Å in both cases. The reason behind this could be due to the resolution of the published structures being higher. This lower resolution of all of the classes experimentally obtained could hinder the flexible fitting, thus suggesting that there are differences in the position of the helices. Another reason could be that **LDC-2** could be having an effect on the conformation/position of the AB domains which has caused there to be a movement in the helices. However, due to the particle number and subsequent resolution being low, this conclusion cannot accurately be drawn without more data being collected. Class 2 also resembled the published State 2 and the model was flexibly fit into the map using MDFF and the resulting model was coloured by Cα r.m.s.d. (Figure 6.14).

A C-subunit

Tight

B *c*-ring

Loose

C H-subunit

Open

D H-subunit

0   3   10
Cα r.m.s.d. (Å)
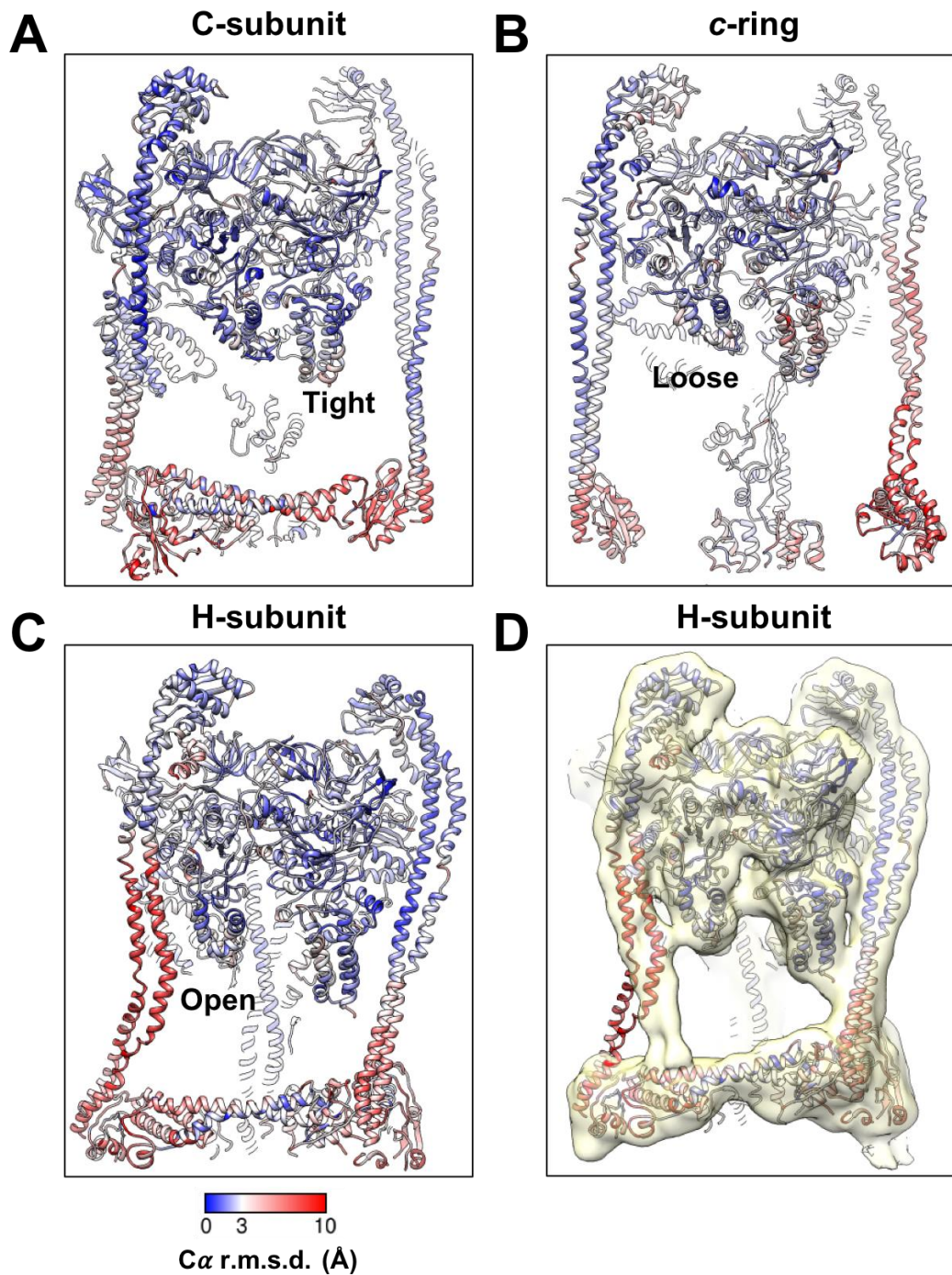
Figure 6.14: Comparison of Class 2 with the published State 2. The published State 2 model was flexibly fit into the Class 2 map using MDFF. A-C) The two models were then overlaid and coloured by Cα r.m.s.d. The red indicates that there is a greater than 3 Å difference in the position of the Cα atoms. This is particularly prevalent in the C-subunit, a-subunit and stator domains. D) The model shown within the map (yellow). The density for the stators has clearly changed relative to the published structures which suggest that these undergo a significant conformational change upon H-subunit dissociation.

174

For Class 2, there is once again a good agreement in the overlap of the Cα atoms within the AB domains. However, there are larger differences in the more peripheral subunits such as the C-subunit, *a*-subunit and the stators which have Cα r.m.s.d. values greater than 5 Å. However, it is worth noting that due to the resolution of this map, it is difficult to visualise the density for α-helices which would have affected the flexible fitting. The stator subunits which connect the V$_1$ domain with subunit *a* (and usually subunit H) have undergone a large change in position from the published models. Moreover, the extent of this movement was too large to be accounted for in the flexible fitting using MDFF. This large movement is unsurprising as the absence of the H-subunit has clearly resulted in the stators and *a*-subunits undergoing a conformational change. Unfortunately, the resolution in this region was not high enough to visualise the secondary structure information so how the stator now interacts with the *a*-subunit cannot be determined. There were also large differences in the C-subunit and stator interactions which could suggest that all of these subunits undergo a conformational change upon the dissociation of the H-subunit. The resolution of the map was not sufficient to determine whether the presence of the inhibitor caused the H-subunit to dissociate so further investigation was needed. Initially, this was probed by analysing negative stain grids of apo and **LDC-2** bound V-ATPase which will be described in Section 6.2.5.

6.2.5 Negative stain analysis of inhibitor-bound V-ATPase

To assess whether the missing H-subunit in the inhibitor-bound cryo-EM data set was a feature of inhibitor binding, negative stain grids of the complex in the apo and **LDC-2** bound form were examined. The rationale behind this was to determine whether incubation of the compound with the protein for ~90 minutes prior to making the negative stain grids resulted in dissociation of the H-subunit or whether the complex was degrading over time. Thereby, 50 micrographs of both apo and **LDC-2** bound V-ATPase were collected on the F20 microscope fitted with a FEI Ceta CCD camera at a magnification of 50,000x. The micrographs were manually picked resulting in 6,819 and 7,254 particles for the apo and **LDC-2** bound data sets, respectively. The particle stacks subsequently underwent iterative rounds of 2D classification to leave 4,255 apo particles and 4,640 inhibitor-bound particles. Example micrographs and 2D classes of the side views for both data sets are shown in Figure 6.15.
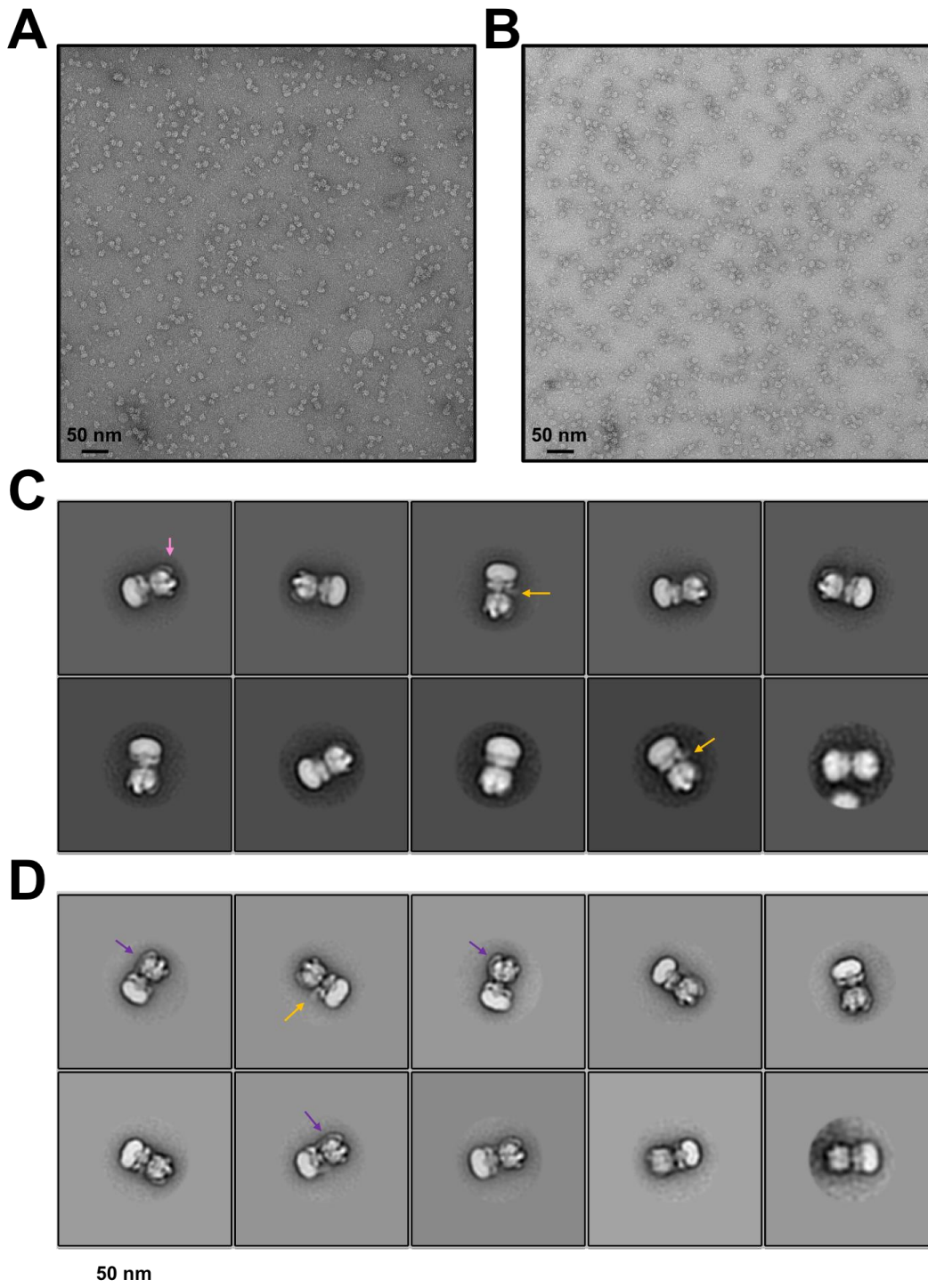
Figure 6.15: Negative stain of apo and inhibitor-bound V-ATPase. Example micrographs of V-ATPase after the protein was concentrated in A) the absence of inhibitor and B) after incubation of inhibitor for ~90 minutes prior to making negative stain grids. The micrographs clearly show V-ATPase and there is no obvious difference in the amount of protein degradation. C) 2D classes for apo V-ATPase. The pink arrow points to the stator and the orange arrow point to the H-subunit which is seen in two of the classes. D) 2D classes for LDC-2 bound V-ATPase. The classes are detailed enough to see the stators (purple arrows) and the orange arrow points to the H-subunit which is visible in one class.

The micrographs clearly showed intact V-ATPase particles within both data sets (Figure 6.15A,B). However, under both conditions there were smaller particles present which could represent the complex dissociated into the V_1 and V_0 domains. It appeared that there were no significant visible changes between the amount of dissociation in the absence and presence of inhibitor just by looking at the raw micrographs alone. The resulting 2D classes also showed high levels of details for the side views within both data sets. This is exemplified by the stators being visible in some of the classes in both of the data sets as shown in Figure 6.15C,D. The stator was more prevalent in the inhibitor-bound classes due to the lighter stain which enabled more structural information to be gleaned.

Further examination of the classes showed that in the apo data set two of the classes had visible detail for the H-subunit compared to only one class in the **LDC-2** bound data set. This could suggest that the H-subunit had dissociated more in the presence of the inhibitor. However, 3D reconstructions did not show any density for the H-subunit for either data set so the negative stain data was inconclusive in determining whether the inhibitor was interacting with the H-subunit and causing it to dissociate. Comparing the apo and inhibitor-bound data sets does indicate that the dissociation of the H-subunit does not result in V-ATPase dissociating any further. This is exemplified by there being no visibly different levels of dissociation in the two data sets although this would need to be quantified further.

It is also worth noting that these negative stain grids were prepared using a different protein sample to the one which was used to collect the cryo-EM **LDC-2** data set which was discussed in Section 6.2.4. However, if the compound was having an effect on the dissociation of the H-subunit then this should be consistent between protein purifications. As the negative stain data was inconclusive, a large data set of apo V-ATPase was collected, which will be described in Section 6.2.6.

6.2.6 Cryo-EM of apo V-ATPase

A 72-hour data collection of apo V-ATPase was collected to determine whether there was any evidence of V-ATPase undergoing any conformational changes in the presence of the inhibitor, including the dissociation of the H-subunit. The apo V-ATPase grids were prepared at the same time as the **LDC-2** grids thus allowing a direct comparison on the effect of the inhibitor to be made. Data were collected on a Titan Krios fitted with a K2 summit direct electron detector and energy filter operating

at 300 kV. 2,944 micrographs were recorded. After the initial motion-correction and CTF estimation were performed, crYOLO was used to pick the micrographs using the model which had been trained on the inhibitor bound data set (Section 6.2.4). This resulted in a particle stack containing 77,867 particles which were subjected to 2D classification. The best classes underwent iterative rounds of 2D and 3D classification resulting in three classes which contained 3,362, 2,827 and 3,762 particles, respectively. The resolutions obtained were 9.5 Å for Class 1 and 10.2 Å for Classes 2 and 4. An overview of the processing and maps obtained is summarised in Figure 6.16.
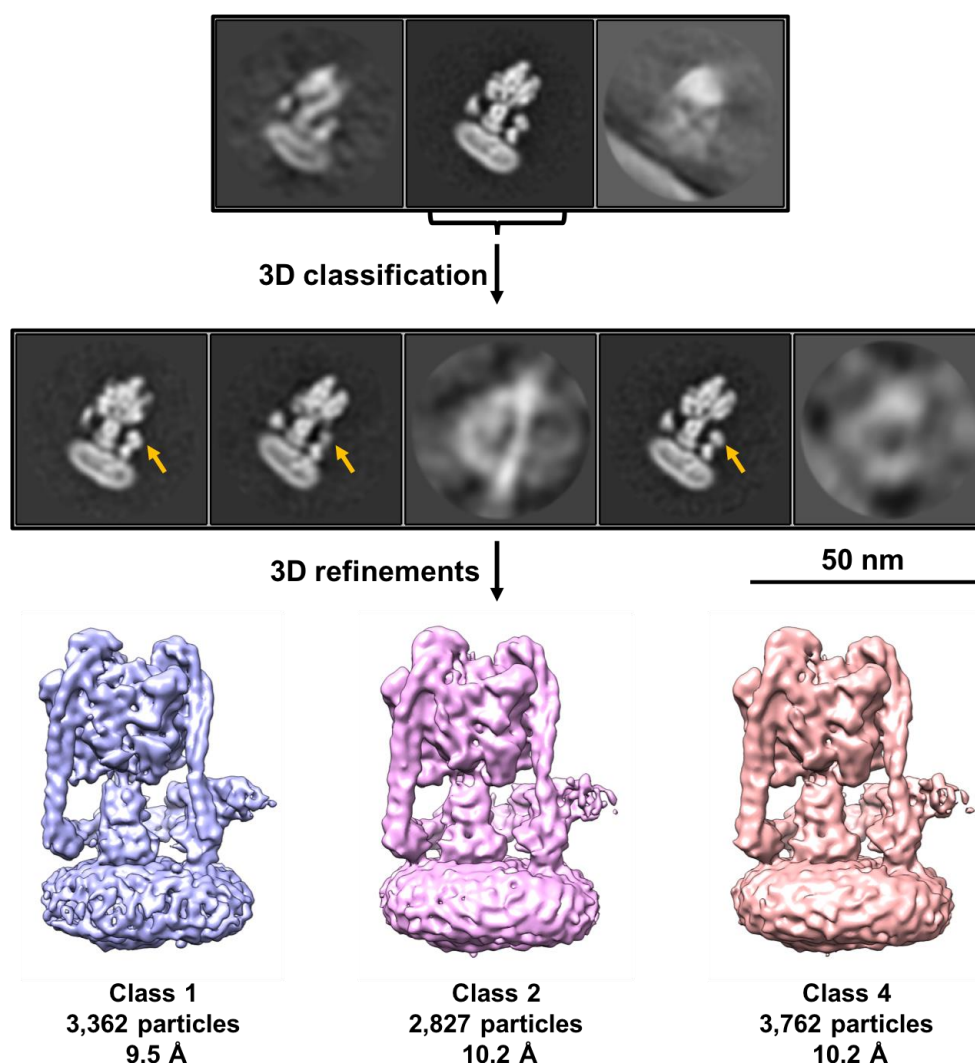
Figure 6.16: Processing pipeline of apo V-ATPase. The particles within the apo data set underwent two rounds of 3D classification. The slice-throughs for each of the classes are shown. The orange arrows indicate that there is density for the H-subunit in all of the classes, although this does appear to be weaker in Class 2. The maps, particle numbers and subsequent resolution of each class are also highlighted.

After classifying the data, three classes were obtained; Class 1, 2 and 4. Classes 3 and 5 represented junk particles which were not taken forward. In all three maps the V_1 domain was resolved to higher resolution than the V_0 domain which is consistent to the **LDC-2** inhibitor-bound data set. Within the V_1 domain, the resolution was once again sufficient to visualise the secondary structure which allowed comparisons to be made to the published structures. The positions of the AB domains relative to the other V_1 subunits, and therefore which of the published states the classes are most similar to, is summarised in Table 6.3.

Table 6.3: Conformational state analysis of apo V-ATPase.

	Above H-subunit	Above C-subunit	Above c-ring	Comparison to published state
Class 1	Loose	Open	Tight	State 1
Class 2	Open	Tight	Loose	State 2
Class 4	Open	Tight	Loose	State 2

For the apo-data set Class 1 resembled published State 1 which contained 34% of the particles (3,362). In comparison, Classes 2 and 4 were most similar to State 2, consisting of 28% and 38% of the data, respectively. Interestingly there were no classes which represented the published State 3 which is likely due to the low particle number in the data set. As there was no inhibitor present within the data set, full comparisons to the published models were not conducted because the protein is not expected to undergo any conformational changes to what is known in the literature.

One of the main reasons for collecting the apo data set was to determine whether there were any classes which did not contain the H-subunit. The maps, which are shown in Figure 6.16, all contain density for the H-subunit. However, close examination of the slice-throughs for Class 2 show that the density for the H-subunit is weaker in comparison to Class 1 and 4 (indicated with the orange arrows in Figure 6.16). This could represent a partially dissociated class, where the H-subunit is becoming dissociated from the complex. However, only ~9,900 particles went into the 3D classification. If the complex which does not contain the H-subunit represents only a very small proportion of the data, then the very low particle number might mean that the class could not be identified through 3D classification. Another possibility

could be that the presence of the inhibitor is causing the H-subunit to dissociate or to dissociate at a faster rate than in the absence of the compound. This would result in there being no classes in the apo structure which do not contain the H-subunit. In order to have a definitive answer on why the H-subunit is missing, a much larger data set would need to be collected, with ~100,000 particles going into the 3D classification. This would also hopefully improve the resolution of the complex so more detailed structural insights could be gained. A much larger data set, would also theoretically allow the resolution of the complex to be improved. This would allow any differences between the apo and inhibitor-bound maps to be elucidated as the resolution is currently not sufficient to determine whether there has been any conformational change upon inhibitor binding.

6.3 Discussion

Within this chapter, V-ATPase was purified from the native source organism (yeast, *saccharomyces cerevisiae*) without introducing any affinity tags onto the protein. The original protocol was published in the 1980's and although it has previously been successful in isolating V-ATPase, it required optimisation for single particle cryo-EM. The original protocol used glycerol gradients to separate out the complex following detergent solubilisation. However, this not only yielded very low quantities of protein but also would have been unsuitable for cryo-EM due to the large amounts of glycerol hindering the contrast within the micrographs, subsequently making it difficult to distinguish particles. It was found that using size exclusion chromatography to separate out the complex, along with changing the buffer composition to remove the glycerol in the eluting buffer, enabled high-quality purified protein to be isolated.

The optimisation of the protocol was not trivial and represented a major challenge during the PhD project, taking several months to achieve. Changing small things within the purification often had a big impact on the quality and/or amount of protein obtained. This was exemplified in the homogenisation step. A delicate balance between needing to use enough force to break open the spheroplasts, combined with not using excessive force resulting in premature lysis of the vacuoles, needed to be achieved. Moreover, in many purifications of other proteins, the cell pellet can be frozen after harvesting the cells or isolating the membranes (for membrane proteins). However, it was found that freezing the material at any stage of the V-ATPase protocol would hinder the amount/quality of protein obtained. Therefore, the full

protein purification needed to be performed on the same day that the cells were harvested. The rate of yeast growth, so the cells could be harvested early in the morning on the protein purification day, also required optimisation. However, after multiple rounds of optimisation pure V-ATPase was successfully purified and used to conduct inhibitor binding assays and for structural determination (including grid optimisation).

Previous work in the Muench group had used carbon-coated quantifoil grids for determining the structure of *Manducca Sexta* V-ATPase[260]. One of the limitations of using carbon-coated cryo-EM grids is that it can hinder the contrast within the images and can sometimes induce a preferred orientation of the protein. Within this study, a range of grid types were used, including carbon/gold films, and it was found that quantifoil R1.2/1.3 grids gave the best distribution of protein within the ice. By using a high concentration of protein (~2.5-3.5 mg/mL), the carbon support film was saturated with protein which allowed the protein to become suspended in the ice within the holes; thereby removing the need for carbon-coated grids. The preliminary overnight data sets of **LDC-2** and apo V-ATPase showed that the protein adopted a range of orientations within the ice and achieved sub-nm resolution from low particle numbers. Due to the promising resolution from the small data sets, larger (72 hour) data sets were collected for both inhibitor-bound and apo V-ATPase which would allow the particle number to be increased.

There are existing structures of V-ATPase which have been obtained using cryo-EM showing that the protein resides in three catalytic conformations[160]. For this study, ~100,000 particles were used to generate the three different states; each containing ~49,000, ~38,000 and ~18,000 particles, and achieving a resolution of 6.9 Å, 7.6 Å and 8.3 Å, respectively. Within this Chapter, two 72-hour data sets were collected of **LDC-2** bound and apo V-ATPase. However, only ~29,000 and ~10,000 particles went into the 3D classifications, respectively, which is less than a third of the total particle number used in the published study. For the inhibitor-bound data set four different classes were obtained which achieved resolutions in the range of 9.5-11.2 Å. The most populated class contained 7,243 particles (9.5 Å). For the apo data set three different classes were obtained, containing ~3,000 particles each. The highest resolved class was also 9.5 Å. The particle numbers within the 72-hour data sets were still low in comparison to the published study. For instance, the published State 1 only achieved a resolution of 6.9 Å, yet contained ~50,000 particles suggesting that the resolution of the complex is ultimately limited by the inherent flexibility of the complex.

It was interesting to note, that despite having approximately three times as many particles as the preliminary overnight data sets, the resolution did not improve with the higher particle number, resulting from the 72-hour data collections. This could be due to the quality of the protein sample. For instance, in the preliminary data sets the protein could have been more stable than the protein used to make the grids for the larger data sets. Unfortunately, for the preliminary data sets, the cryo-grids did contain enough areas of the grid with a good particle distribution, thus preventing a large data set to be collected. One of the limitations of working with V-ATPase is that the protein cannot be frozen as it easily degrades and loses activity upon freezing. Therefore, fresh protein was prepared each time new cryo-grids were prepared, resulting in possible discrepancies between batches of protein obtained.

The resolutions of both the inhibitor-bound and apo V-ATPase maps were similar (~9.5 Å) and the resolution was not sufficient to determine whether there were any conformational changes resulting from inhibitor binding. Moreover, for both data sets only classes which corresponded to the published State 1 and State 2 were identified. The published State 3 represents only 18% of the total particle number and is therefore the least populated class. Due to the small number of particles within the two data sets, State 3 might represent only a very small number of particles so this state was not obtained during 3D classification. One interesting observation in the inhibitor bound data set was the identification of a unique class which has never previously been reported. This class contained ~20% of the particles and did not contain any density for the H-subunit. Due to the low resolution (11.2 Å), it is not known whether this class is a direct result of an interaction with the **LDC-2** compound or due to the V-ATPase degrading over time and losing this subunit. Although, the equivalent class was not identified in the apo data set. This would require further investigation, and a much larger data set (>100,000 particles) to be collected which would enable the density for the small molecule inhibitor to be visualised. The high-resolution structures of the F-ATP synthase highlight that it is possible to obtain high resolution structures of these large rotary machines. Therefore, with the improvements in image processing and microscope technology it should be possible to achieve these high-resolutions for V-ATPase.

The synthesised compounds, which were discussed in Chapter 5, were designed to act upon the ATP binding site which resides in the V_1 domain of V-ATPase. The highest resolution structure of the full V-ATPase complex is ~6 Å which would not be high enough to visualise inhibitor binding. Moreover, it was found that

greater than 100,000 particles would likely be required to obtain a high-resolution structure of the full complex. One way to get around this would be to dissociate V-ATPase into the V_1 and V_0 domains. The structure of one of the compounds bound to the V_1 domain could then be obtained using cryo-EM which would represent the first cryo-EM structure of the V_1 domain to be published.

Within this study, obtaining purified V_1 protein was attempted but was ultimately unsuccessful. Purified V-ATPase was incubated with 5 mM ATP in the presence of $MgCl_2$, firstly at 30 °C (10 minutes) and then at 4 °C (1 hour). The mixture was then loaded onto a Superose 6 15/30 GL column to attempt to separate out the V_1 and V_0 domains. Unfortunately, attempts at carrying out this protocol were unsuccessful and no V_1 protein was obtained. Problems could have arisen due to the length of time which the protein was incubated with ATP for, the concentration of ATP/$MgCl_2$ used or due to inefficient separation or low quantities of protein used in the gel filtration. Therefore, obtaining a structure of inhibitor-bound V_1 was ultimately not successful and due to time restrictions (and wanting to focus on obtaining the structure of the full complex), the protocol was not optimised further but it could form the basis of future structural studies. If the resolution of a cryo-EM V_1 structure was not sufficient to visualise inhibitor density, then the V_1 domain could be crystallised in the presence of the inhibitor. There are existing crystal structures of this domain at ~3 Å resolution which should allow inhibitor density to be visualised. This high-resolution structural information would then aid the design of further compounds to create a highly potent V-ATPase inhibitor.

A high-resolution structure of V-ATPase which allowed inhibitor-binding to be visualised was ultimately not achieved within the course of the project. However, the purification of V-ATPase was optimised to result in a high-quality protein being obtained. Cryo-EM data sets were also collected which enabled sub-nm resolution structures of apo and inhibitor bound maps to be obtained which were of sufficient quality to visualise secondary structure information. Moreover, a unique class was identified in the inhibitor bound data set which has never previously been reported which consisted of V-ATPase missing the H-subunit. Furthermore, the distribution of the states which were obtained were different to the published study which could also be a feature of the interaction with the inhibitor. Although the exact biological relevance of this has not yet been determined, it represents an exciting prospect for future work. The protein purification and grid optimisation protocols represent a starting point for this work to be continued in the future.

7   Concluding statement and outlook

7.1   Discussion

The overall aim of the PhD project was to determine the role which cryo-EM can play in drug discovery. In 2016, when the PhD project was started, cryo-EM was only beginning to be thought of as a tool to be utilised in drug discovery programs. At the time, pharmaceutical companies were only starting to think about whether the technology should be invested in. Therefore, none of the major UK pharmaceutical companies had any of their own infrastructure required to use cryo-EM in their drug discovery programs. However, in 2016 the Cambridge cryo-EM consortium which allows five pharma companies (Astex, AstraZeneca, GSK, Heptares and UCB) access to high end Titan Krios microscopes, was formed. Over the past three years, due to improvements in the resolution, technology and image processing pipelines available, all of the major pharmaceutical companies have now invested in cryo-EM. This is exemplified by the majority of the consortium members having cryo-grid preparation devices and screening microscopes in-house, which highlights the significant investment made by the pharmaceutical industry and also highlights the rapid advances in the field within the past three years.

Before commencing the project, there were few examples of cryo-EM structures with small molecule inhibitors bound, with the most well-known example being the 2.2 Å structure of β-galactosidase[115]. One of the aims within the PhD project was to determine whether it was possible to obtain high-resolution structures of therapeutically relevant membrane proteins with inhibitors bound. To this end, two systems were studied which were at different stages of the drug discovery process; cytochrome bc_1 and V-ATPase. Although cytochrome bc_1 had previously been crystallised, there were no cryo-EM structures of the system. Therefore, bovine bc_1, with inhibitors bound, was studied using cryo-EM, with the aim of seeing whether the resolutions attained allowed inhibitor density to be visualised. This work resulted in five bc_1 structures being determined, four of which had small-molecule inhibitors bound, with resolutions ranging from 3.3-4.6 Å. For three of the inhibitor-bound structures, the resolution was sufficient to visualise density at the inhibitor-binding site thus allowing the inhibitor to be modelled into the map. For one of the structures, GSK-bc_1, the density showed that the compound could adopt two different positions within the binding site which had not previously been seen by X-ray crystallography, suggesting that cryo-EM could reveal novel insights into the occupancy of ligand

binding. Moreover, comparisons between existing crystal structures and the new cryo-EM structures, showed that the models obtained were consistent between the two techniques, which validates using an EM approach to obtain the structural information. This work was published in IUCrJ in 2018, with one of the structures forming part of a larger study into the development of a novel scaffold to treat Toxoplasmosis which it is hoped will be published in 2020.

Cytochrome bc_1 is a validated anti-malarial drug target, yet there is no structural information of the parasite protein thus hindering the design of new therapeutic agents. One of the problems preventing the structural determination of the parasite protein is the need to use milligrams of protein to facilitate protein crystallisation. As cryo-EM uses significantly less protein than X-ray crystallography, cryo-EM can be used as an alternative structural determination tool. The work described in Chapter 3 showed that it is possible to obtain high resolution structures of bc_1, so the protocols and pipelines developed for the bovine enzyme can be applied to the parasite protein in the future. Work is ongoing at the University of Liverpool, in the Hasnain and Antonyuk groups, to extract the parasite protein from malarial parasites or by using the CRISPR-CAS9 system to engineer tags onto the parasite protein to aid the purification. If this can be achieved then the work presented in this thesis, can provide the foundations for the structure determination of the parasite enzyme. Subsequently, if high resolution structures can be obtained, with inhibitors bound, this structural information can be utilised in the design of new therapeutic agents allowing highly potent and selective compounds to be made.

The second system studied was the V-ATPase. To date there are no examples of small molecules designed against V-ATPase which used a SBDD approach. The current V-ATPase inhibitors are macrocyclic rings which makes them difficult to handle and expensive to buy. Therefore, the aim was to initially use virtual screening to design inhibitors for yeast V-ATPase which could then be used as a chemical tool to study the system. Moreover, the original aim was to use cryo-EM structures to help design the compounds so the role of cryo-EM in drug discovery programs could be further enhanced. However, it was soon apparent that the resolution of the cryo-EM structures was not sufficient to aid the design of novel compounds. Therefore, the vHTS was repeated using higher resolution crystal structures. After one round of vHTS and an 'SAR-by-inventory' approach, hit compounds were identified. The hit compounds also showed some basic SAR, as it was found that the carboxyl group was important for binding, as compounds without this group did not inhibit V-ATPase.

This suggested that the compounds were in-fact binding at the target ATP-binding site suggesting that the docking poses generated are reliable. The most promising hit compound (**V-38**) was modified resulting in a small library of compounds being synthesised; four of these compounds had <100 μM IC$_{50}$ values (Compounds **31**, **32, 36** and **38**). This represents the first example of potent inhibitors being designed through a SBDD approach to act upon V-ATPase. A manuscript of this work is currently being prepared and will form a strong basis for future medicinal chemistry programs.

V-ATPase was also successfully purified from the native yeast organism during the PhD project and cryo grids were subsequently optimised resulting in apo and inhibitor-bound data sets being collected. Resolutions below 10 Å were achieved from low number of particles, which allowed the secondary structure of the protein to be visualised and compared to published structures. Interestingly, one novel state of the complex was found, which had not previously been reported, showing V-ATPase without the H-subunit in the inhibitor-bound data set. However, the resolution and number of particles obtained, were not sufficient to determine whether the dissociation of the H-subunit was a direct result of inhibitor binding or due to the protein degrading. The reasons behind the H-subunit dissociating could also have important implications to the regulation of the complex which should be investigated further. Moreover, the distribution of particles which make up each of the three classes of V-ATPase was shifted compared to the literature. Again, it is unclear whether this was due to the presence of the inhibitor. The work outlined in Chapter 6, including the optimisation of the protein purification and cryo-grid conditions, would form the basis of future structural studies which would allow the effect of the inhibitor to be probed in more detail. Furthermore, if a high-resolution structure can be obtained then this can aid the design of new inhibitors. These compounds could subsequently be modified to act as a pharmaceutical agent in the treatment of certain diseases such as toxoplasmosis or osteopetrosis. Currently, without high resolution structural information of the full complex, the design of new inhibitors is challenging. Therefore, obtaining this structural information is vital for the design of new compounds which could act as therapeutic agents in the future.

Drugs can bind to a range of different proteins, of all shapes and sizes. Therefore, to be a useful tool in SBDD, cryo-EM needs to be able to determine high resolution structures of a variety of proteins. Membrane proteins make up ~60% of all drug targets yet make up only ~3% of structures deposited into the protein data bank[261].

This is due to membrane proteins being notoriously difficult to crystallise. Cryo-EM removes the need for protein crystallisation so can play an important role in the structural characterisation of challenging drug targets. Outside of the work described in this thesis, a number of side projects were undertaken with internal and external collaborators. This has enabled multiple membrane proteins to be studied using cryo-EM. These side projects have resulted in two publications which are detailed in Appendix B. The work was not included within the thesis because the cryo-grid optimisation, data collection and image processing were carried out in Leeds by myself but the protein purification and most of the subsequent structure interpretation was completed by collaborators.

Cryo-EM structures were determined for three unique systems; AcrB, human TRPC5 and quinol-dependent Nitric Oxide Reductase (qNOR). In collaboration with the Postis group at Leeds Beckett University, Salmonella AcrB was solubilised using styrene maleic acid (SMA), which directly removed the protein from the membrane. This is advantageous as the native lipid interactions are maintained. The resolution achieved was ~4.5 Å which allowed the bulky side chains to become visible. This exemplifies how powerful cryo-EM can be when studying membrane proteins, not only in their native lipid environments but also without the restrictions of the crystal lattice, as Salmonella AcrB had previously evaded structural characterisation. Another collaboration with the Bon group at the University of Leeds enabled the high-resolution structure of human TRPC5 to be determined to 3.0 Å. The protein had a small molecule bound, which emphasises the ability to achieve high-resolution structures of therapeutically relevant proteins. This exemplifies how cryo-EM can aid SBDD programs which target membrane proteins allowing high-quality and reliable models to be obtained which can guide the design of new molecules.

At the start of the PhD project, the size limit for cryo-EM studies was ~200 kDa which would hinder SBDD programs as lots of drug targets have lower molecular weights. For instance, GPCRs are typically <150 kDa. In collaboration with the Hasnain and Antonyuk groups at the University of Liverpool, the cryo-EM structure of qNOR was determined in a dimeric and monomeric state which had molecular weights of 190 and 95 kDa, respectively. The resolutions attained were 3.3 Å for the dimer and 5.0 Å for the monomer, thus allowing side chains to be built into the map and visualisation of the secondary structure, respectively. Moreover, the dimeric state of qNOR had never previously been reported, despite numerous crystal structures being solved. This exemplifies how useful it can be to determine the structure of

proteins in their native state, away from the restrictions of the crystal lattice. The study also shows that the size of proteins which can be studied using cryo-EM is constantly reducing, thus allowing the structures of a broad range of important drug targets to be elucidated.

Over the course of the three-year PhD project, cryo-EM structures of five unique membrane proteins were determined, which had a size range of 95 kDa to 1 MDa. The resolutions attained in all maps allowed the secondary structure to be visualised and for cytochrome bc_1 and TRPC5 the resolution not only allowed for side chains to be fitted into the density, but it also enabled visualisation of how small molecules bind to the target. SBDD programs rely on using models of the protein which accurately model the position of side chains within the structure, thus guiding the design of new compounds. As shown in the examples described above, cryo-EM is now capable of producing maps of sufficient quality to obtain this vital structural information and can therefore be thought of as a powerful tool in SBDD programs, particularly for membrane proteins which have traditionally been difficult to crystallise. An overview of the maps obtained and resolutions achieved is shown in Figure 7.1.

Vacuolar-ATPase	Cytochrome bc_1	TRPC5	AcrB	qNOR (dimer and monomer)	
1 MDa	480 kDa	440 kDa	360 kDa	190 kDa	95 kDa
8.1 Å	3.3 Å	3.0 Å	4.5 Å	3.3 Å	5.0 Å
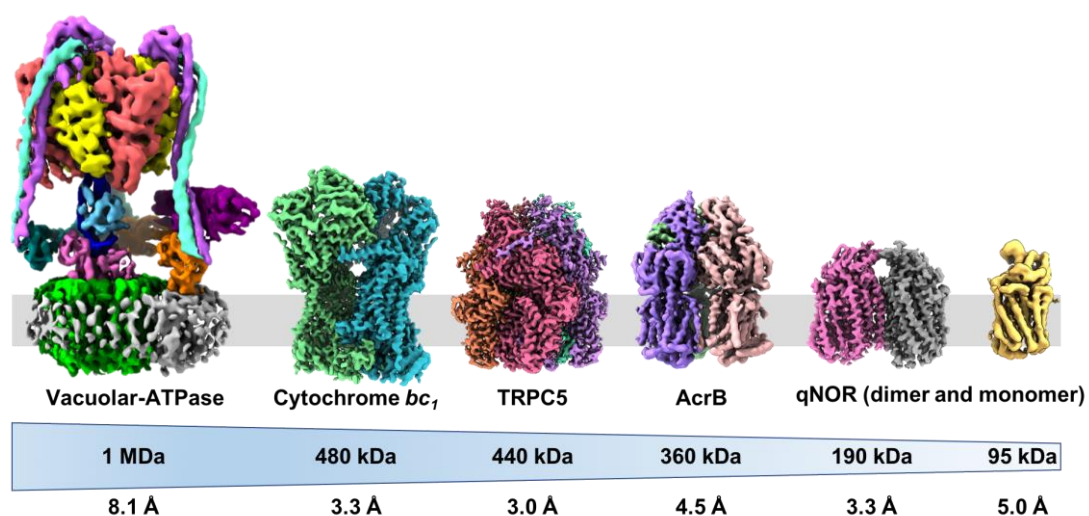
Figure 7.1: An overview of all of the systems studied during the PhD project. *The structure of five unique membrane proteins have been determined to resolutions high enough to visualise secondary structural information. For bc_1, TRPC5 and qNOR, resolutions below 3.5 Å were obtained which allowed the side chains to be accurately modelled into the map. These structural details can be utilised in the design of new therapeutic agents thus proving that cryo-EM can play a role in drug discovery programs.*

7.2 Outlook

At present, X-ray crystallography still remains the gold-standard for SBDD programs due to the high-resolution of the structures obtained, throughput of ligands which can be screened and speed of structure determination. The resolution attained (often below ~2 Å) allows the inhibitor to be accurately modelled into the density and enables interactions with water or metal ions to be visualised. However, cryo-EM is becoming a powerful tool in structure determination, as shown by the recent 'resolution revolution' which has resulted in a huge rise of EM maps being deposited into the EMDB. Moreover, the structure of important therapeutic targets such as GPCRs, ABC-transporters, LAT1 and HERG have now been determined to high-resolution, some with small molecule inhibitors bound. These structural insights can be utilised in the design of new therapeutics thus benefiting drug discovery programs. The cryo-EM field is rapidly evolving, with developments in hardware (detectors), processing algorithms and grid preparation constantly being made. If the improvements continue at such a rapid pace, then cryo-EM will play an even bigger role in drug discovery programs thus allowing a broad range of new medicines to be developed.

8 Experimental

General Information and Instrumentation

All reagents were obtained from commercial suppliers and were used without purification. Reactions were carried out under nitrogen with solvents which were obtained dry from commercial suppliers. Analytical TLC was carried out using silica coated aluminium plates and spots were visualised using UV radiation unless otherwise stated. Silica gel (230-400 mesh, Merck) was used during flash chromatography and solvents were removed under vacuum using a Buchi rotatory evaporator at diaphragm pump pressure. Reverse phase chromatography (Biotage) was performed using an Isolera Four EXP with Spektra.

^{1}H and ^{13}C NMR was carried out on a Bruker DPX300 Fourier transform spectrometer or a Bruker Avance 500 using an internal deuterium lock. The chemical shifts were reported in parts per million (ppm) and were reported downfield from a TMS reference peak at 0 ppm. Abbreviations have been used when reporting the NMR data, ie s = singlet, d = doublet, t = triplet, m = multiplet etc. Proton and carbon assignments have been carried out using COSY and HSQC spectra analysis. Assignments for the tetrahydrobenzothiophene cores are based on the numbering shown in Figure 8.1.
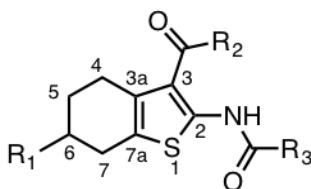
Figure 8.1: The numbering for the NMR assignments for the tetrahydrobenzothiophene core.

Infrared (IR) spectra were recorded on a Perkin Elmer Spectrum One FT-IR spectrophotometer using OPUS software and the vibrational frequencies were recorded in wavenumbers (cm^{-1}). High Resolution Mass Spectra (HRMS) was performed using a Bruker MaXis Impact Time of Flight spectrometer which uses electrospray ionization. HPLC analysis was carried out on a Dionex HPLC system using a Thermo Electron Corporation Hyperprep HS C18 column (8 μm, 250 × 4.6 mm) and diode array as a detector. Unless otherwise stated a water and acetonitrile (5-95%) gradient was used as the solvent. Liquid Chromatography Mass Spectrometry (LCMS) was performed on a Bruker Daltronics or Bruker HCT-Ultra running a gradient of increasing acetonitrile in water. Compounds were detected on

a Bruker mass spectrum analyser. Specific rotation measurements were recorded using a Schmidt and Haensch Polartronic H532 polarimeter, using a 50 mm cell and the Sodium D line (589 nm). $[\alpha]_D$ are reported in units of 10^{-1} deg dm^2 g^{-1}.

methyl 2-{[(tert-butoxy)carbonyl]amino}-6-tert-butyl-4,5,6,7-tetrahydro-1-benzothiophene-3-carboxylate (15)
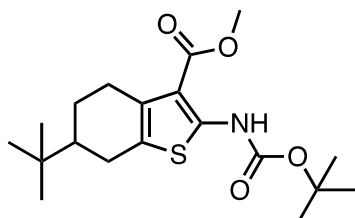
2-amino-6-tertbutyl-4,5,6,7-tetrahydrobenzothiophene-3-carboxylic acid methyl ester (500 mg, 1.87 mmol) was added to 1,4-dioxane (20 mL), DMAP (23 mg, 0.187 mmol) and di-tertbutyl decarbonate (860 mg, 3.93 mmol). The reaction mixture was heated at 80 °C overnight before hydrazine hydrate (0.17 mL, 5.6 mmol) was added. The reaction was stirred at 40 °C for 1.5 hours. Once cool, the solvent was removed *in vacuo* and the crude material was purified using column chromatography (19:1 petrol/EtOAc) to yield the title compound as a white solid (586 mg, 86%); **R$_f$**: 0.91 (EtOAc); **δ$_H$** (CDCl$_3$): 10.18 (s, 1H, N-H), 3.76 (s, 3H, O-CH$_3$), 2.91 (dd, 1H, J = 17, 6.5 Hz, 4-H), 2.57 (dd 1H, J = 15.5, 5 Hz, 7-H), 2.50-2.42 (m, 1H, 4-H), 2.32-2.26 (m, 1H, 7-H), 1.94-1.92 (m, 1H, 5-H), 1.45 (s, 9H, O-C(CH$_3$)$_3$), 1.42-1.35 (m, 1H, 6-H), 1.20 (qd, 1H, J = 12.5, 5 Hz, 5-H), 0.86 (s, 9H, C(CH$_3$)$_3$); **δ$_C$** (CDCl$_3$): 166.76 (CO-OMe), 152.19 (C2), 150.16 (NH-CO), 130.98 (C7a), 125.75 (C3a), 109.56 (C3), 81.80 (O-C(CH$_3$)$_3$), 51.21 (O-CH$_3$), 45.01 (C6), 32.43 (C(CH$_3$)$_3$), 28.24 (O-C(CH$_3$)$_3$), 27.52 (C7), 27.30 (C(CH$_3$)$_3$), 25.73 (C4), 24.41 (C5); **IR (v$_{max}$/cm^{-1});** 3352, 2981, 2945, 2851, 2802, 1711 and 1681; **HPLC** (MeCN): 100%; **HRMS** (ES): Found [M+H]$^+$ 368.1885, requires *[MH]* 368.1817; [a]$_D$ 0.00 (0.33, MeOH).

ethyl 2-{[(tert-butoxy)carbonyl]amino}-4,5,6,7-tetrahydro-1-benzothiophene-3-carboxylate (16)
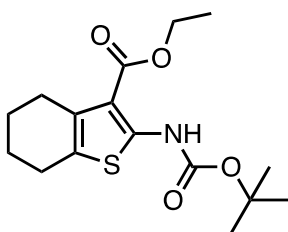
Ethyl-2-amino-4,5,6,7-tetrahydrobenzothiophene-3-carboxylate (1.0 g, 4.44 mmol) was added to 1,4-dioxane (30 mL), DMAP (54 mg, 0.44 mmol) and di-tertbutyl decarbonate (2.03 g, 9.32 mmol). The reaction mixture was heated at 80 °C overnight before hydrazine hydrate (0.41 mL, 13.3 mmol) was added. The reaction was stirred at 40 °C for 1.5 hours. Once cool, the solvent was removed *in vacuo* and the crude material was purified using column chromatography (19:1 petrol/EtOAc) to yield the title compound as a white solid (1.33 g, 92%); **R$_f$**: 0.56 (19:1 Petrol-EtOAc); **δ$_H$** (CDCl$_3$): 10.25 (s, 1H, N-H), 4.23 (q, 2H, J = 7.1 Hz, CH$_2$CH$_3$), 2.67 (td, 2H, J = 6.0, 1.9 Hz, 4-H), 2.54 (td, 2H,

J = 5.7, 5.0, 2.7 Hz, 7-H), 1.74-1.67 (m, 4H, 7-H, 5-H), 1.45 (s, 9H, (C$\underline{H}_3$)$_3$), 1.29 (t, 3H, J = 7.1 Hz, CH$_2$C$\underline{H}_3$); **δ$_C$** (CDCl$_3$): 166.39 ($\underline{C}$O-OEt), 152.21 (NH-$\underline{C}$O), 149.88 (C2), 131.04 (C7a), 125.11 (C3a), 110.07 (C3), 81.76 ($\underline{C}$-(CH$_3$)$_3$) 60.22 ($\underline{C}$H$_2$CH$_3$), 28.20 (C-($\underline{C}$H$_3$)$_3$), 26.46 (C7), 24.32 (C4), 23.02 (C6), 22.87 (C5), 14.34 (CH$_2$-$\underline{C}$H$_3$); **IR (v$_{max}$/cm^{-1});** 3275, 2979, 2941, 2860, 1716 and 1655; **HPLC** (acetonitrile): 100%; **HRMS** (ES): Found [M+Na]$^+$ 348.1238, requires *[MNa]* 348.1245.

ethyl 2-{[(tert-butoxy)carbonyl]amino}-6-methyl-4,5,6,7-tetrahydro-1-benzothiophene-3-carboxylate (17)
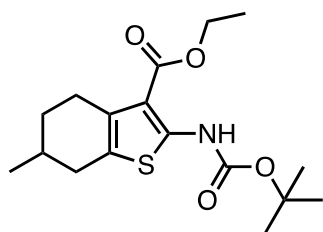
Ethyl-2-amino-6-methyl-4,5,6,7-tetrahydrobenzothiophene-3-carboxylate (1.0 g, 4.18 mmol) was added to 1,4-dioxane (30 mL), DMAP (51 mg, 0.418 mmol) and di-tertbutyl decarbonate (1.90 g, 8.77 mmol). The reaction mixture was heated at 80 °C overnight before hydrazine hydrate (0.39 mL, 12.5 mmol) was added. The reaction was stirred at 40 °C for 1.5 hours. Once cool, the solvent was removed *in vacuo* and the crude material was purified using column chromatography (19:1 petrol/EtOAc) to yield the title compound as a yellow solid (1.21 g, 87%); **R$_f$:** 0.31 (19:1 Petrol-EtOAc); **δ$_H$** (CDCl$_3$): 10.33 (s, 1H, N-H), 4.32, (q, 2H, J = 7 Hz, C$\underline{H}_2$-CH$_3$), 2.93 (dt, 1H, J = 17, 2 Hz, 4-H), 2.71-2.62 (m, 2H, 4-H, 7-H), 2.28-2.22 (m, 1H, 7-H), 1.91-1.86 (m, 2H, 6-H, 5-H), 1.54 (s, 9H, (C$\underline{H}_3$)$_3$), 1.42-1.34 (m, 1H, 5-H), 1.39 (t, 3H, J = 7 Hz, CH$_2$-C$\underline{H}_3$), 1.07 (d, 3H, J = 6.5 Hz, C-C$\underline{H}_3$); **δ$_C$** (CDCl$_3$): 166.38 ($\underline{C}$O-OEt), 152.21 (NH-$\underline{C}$O), 149.98 (C2), 130.72 (C7a), 124.77 (C3a), 109.94 (C3), 81.77 ($\underline{C}$-(CH$_3$)$_3$), 60.22 ($\underline{C}$H$_2$-CH$_3$), 32.40 (C7), 31.16 (C5), 29.27 (C6), 28.24 (O-C($\underline{C}$H$_3$)$_3$), 26.27 (C4), 21.43 (C-$\underline{C}$H$_3$), 14.34 (CH$_2$-$\underline{C}$H$_3$); **IR (v$_{max}$/cm^{-1});** 3247, 2977, 2953, 2921, 2871, 1715 and 1659; **HPLC** (DMSO): 100%; **HRMS** (ES): Found [M+Na]$^+$ 362.1396, requires *[MNa]* 362.1402; [a]$_D$ 0.01 (0.32, MeOH).

ethyl 6-benzyl-2-{[(tert-butoxy)carbonyl]amino}-4H,5H,6H,7H-thieno[2,3-c]pyridine-3-carboxylate (18)
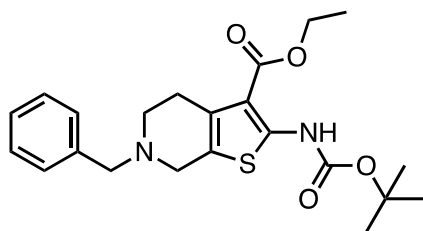
Ethyl-2-amino-6-benzyl-4,5,6,7-tetrahydrothiopheno[2,3-c]-pyridine-3-carboxylate (1.0 g, 3.16 mmol) was added to 1,4-dioxane (30 mL), DMAP (51 mg, 0.418 mmol) and di-tertbutyl decarbonate (1.44 g, 6.63 mmol). The reaction mixture was heated at 80 °C overnight before hydrazine hydrate (0.24 mL, 9.48 mmol) was added. The reaction was stirred at 40 °C for 1.5 hours.

Once cool, the solvent was removed *in vacuo* and the crude material was purified using column chromatography (1:1 petrol/EtOAc) to yield the title compound as a yellow solid (659 mg, 51%); **R$_f$**: 0.61 (1:1 Petrol-EtOAc); **δ$_H$** (CDCl$_3$): 10.30 (s, 1H, N-H), 7.39 (dd, 2H, *J* = 7, 1.5 Hz, 3'-H), 7.36 (dt, 2H, *J* = 7.5, 1.5 Hz, 2'-H), 7.31 (dt, 1H, *J* = 7, 1.5 Hz, 4'-H), 4.31 (q, 2H, *J* = 7 Hz, C$\underline{H}_2$-CH$_3$), 3.71 (s, 2H, N-C$\underline{H}_2$), 3.57 (s, 2H, 7-H), 2.87 (t, 2H, *J* = 7 Hz, 4-H), 2.77 (t, 2H, *J* = 6 Hz, 5-H), 1.54 (s, 9H, C-(C$\underline{H}_3$)$_3$), 1.37 (t, 3H, *J* = 7 Hz, CH$_2$-C$\underline{H}_3$); **δ$_C$** (CDCl$_3$): 166.16 ($\underline{C}$O-OEt), 152.12 (C2), 150.39 (NH-$\underline{C}$O), 138.22 (C1'), 129.71 (C7a), 129.13 (C3'), 128.36 (C2'), 127.23 (C4'), 122.71 (C3a), 109.66 (C3), 81.95 ($\underline{C}$-(CH$_3$)$_3$), 61.93 (N-$\underline{C}$H$_2$), 60.31 ($\underline{C}$H$_2$-CH$_3$), 51.36 (C7), 50.02 (C5), 28.23 (C-($\underline{C}$H$_3$)$_3$), 26.86 (C4), 14.30 (CH$_2$-$\underline{C}$H$_3$); **IR (v$_{max}$/cm^{-1});** 3253, 3071, 2977, 2949, 2932, 2901, 1722 and 1663; **HRMS** (ES): Found [M+H]$^+$ 417.1861, requires *[MH]* 417.1770. HPLC could not be obtained due to machine fault.

2-{[(tert-butoxy)carbonyl]amino}-6-tert-butyl-4,5,6,7-tetrahydro-1-benzothiophene-3-carboxylic acid (19)
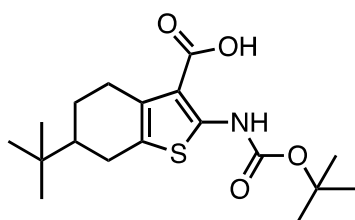
15 (500 mg, 1.36 mmol) was hydrolysed using NaOH (270 mg, 6.75 mmol) dissolved in solvent mixture containing MeOH/H$_2$O/THF in a 2:2:1 ratio (20 mL) by heating at 80 °C for 3 hours. Once cool, the mixture was neutralised using 2M HCl. The resulting precipitate was filtered and dried overnight to yield the title compound as a yellow powder (440 mg, 91%); **R$_f$**: 0.91 (EtOAc); **δ$_H$** (DMSO): 13.10 (s, 1H, O-H), 10.48 (s, 1H, N-H), 3.03 (dd, 1H, *J* = 17.5, 4.5 Hz, 4-H), 2.68 (dd, 1H, *J* = 16, 4 Hz, 7-H), 2.52-2.48 (m, 1H, 4-H), 2.39 (t, 1H, *J* = 16, 7-H), 2.00 (dd, 1H, *J* = 13, 4 Hz, 5-H), 1.55 (s, 9H, O-C(C$\underline{H}_3$)$_3$), 1.51-1.42 (m, 1H, 6-H), 1.25 (qd, 1H, *J* = 12.5, 5 Hz, 5-H), 0.97 (s, 9H, C(C$\underline{H}_3$)$_3$); **δ$_C$** (DMSO): 167.65 ($\underline{C}$OOH), 151.66 (NH-$\underline{C}$O), 148.93 (C2), 131.67 (C7a), 125.51 (C3a), 110.69 (C3), 82.16 (O-$\underline{C}$(CH$_3$)$_3$), 44.88 (C6), 32.64 ($\underline{C}$(CH$_3$)$_3$), 28.26 (O-C($\underline{C}$H$_3$)$_3$), 27.59 (C($\underline{C}$H$_3$)$_3$), 27.45 (C7), 25.71 (C4), 24.37 (C5); **IR (v$_{max}$/cm^{-1});** 3292, 2948, 1725 and 1647; **HPLC** (DMSO): 100%; **HRMS** (ES): Found [M-H]$^-$ 352.1578, requires *[M-H]* 352.1572; [a]$_D$ -0.01 (0.33, MeOH).

2-{[(tert-butoxy)carbonyl]amino}-4,5,6,7-tetrahydro-1-benzothiophene-3-carboxylic acid (20)
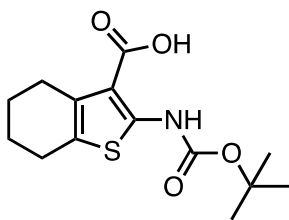
16 (1.30 g, 4.0 mmol) was hydrolysed using NaOH (720 mg, 10.0 mmol) dissolved in solvent mixture containing MeOH/H_2O/THF in a 2:2:1 ratio (40 mL) by heating at 80 °C for 3 hours. Once cool, the mixture was neutralised using 2M HCl. The resulting precipitate was filtered and dried overnight to yield the title compound as a white solid (907 mg, 77%). R_f: 0.18 (6:1 Petrol-EtOAc); δ_H (DMSO): 13.05 (s, 1H, O-H), 10.43 (s, 1H, N-H), 2.68 (t, 2H, J = 5.7, 4-H), 2.57 (t, 2H, J = 5.7 Hz, 7-H), 1.75-1.67 (m, 4H, 6-H, 5-H), 1.49 (s, 9H, (C$\underline{H}_3$)$_3$); δ_C (DMSO): 167.67 ($\underline{C}$OOH), 151.68 (NH-$\underline{C}$O), 148.75 (C2), 131.68 (C7a), 124.93 (C3a), 111.03 (C3), 82.16 ($\underline{C}$-(CH$_3$)$_3$), 28.25 (($\underline{C}$H$_3$)$_3$), 26.38 (C4), 24.18 (C7), 23.03 (C6), 22.71 (C5); **IR (v_{max}/cm^{-1});** 3280, 2980, 2927, 2848, 1722 and 1642; **HPLC** (acetonitrile): 100%; **HRMS** (ES): Found [M-H]$^-$ 296.0963, requires *[M-H]* 296.1035.

2-{[(tert-butoxy)carbonyl]amino}-6-methyl-4,5,6,7-tetrahydro-1-benzothiophene-3-carboxylic acid (21)

17 (1.20 g, 3.53 mmol) was hydrolysed using NaOH (630 mg, 15.9 mmol) dissolved in solvent mixture containing MeOH/H_2O/THF in a 2:2:1 ratio (40 mL) by heating at 80 °C for 3 hours. Once cool, the mixture was neutralised using 2M HCl. The resulting precipitate was filtered and dried overnight to yield the title compound as a white solid (567 mg, 51%); δ_H (DMSO): 10.49 (s, 1H, N-H), 2.86 (d, 1H, J = 16.5 Hz, 4-H), 2.66 (dd, 1H, J = 16, 4.5 Hz, 7-H), 2.60-2.56 (m, 1H, 4-H), 2.17 (dd, 1H, J = 15.5, 9.5 Hz, 7-H), 1.81-1.80 (m, 2H, 6-H, 5-H), 1.48 (s, 9H, (C$\underline{H}_3$)$_3$), 1.33-1.25 (m, 1H, 5-H), 1.01 (d, 3H, J = 6.5 Hz, C$\underline{H}_3$); δ_C (DMSO): 167.70 ($\underline{C}$OOH), 151.69 (NH-$\underline{C}$O), 148.68 (C2), 131.40 (C7a), 124.47 (C3a), 111.13 (C3), 82.09 (O-$\underline{C}$(CH$_3$)$_3$), 32.25 (C7), 31.00 (C5), 29.26 (C6), 28.25 (C-($\underline{C}$H$_3$)$_3$), 26.20 (C4), 21.70 (C-$\underline{C}$H$_3$); **IR (v_{max}/cm^{-1});** 3277, 2981, 2949, 2911, 2866, 2815, 1721 and 1644; **HPLC** (DMSO): 100%; **HRMS** (ES): Found [M-H]$^-$ 310.1119, requires *[M-H]* 310.1191; [a]$_D$ 0.00 (0.92, MeOH).

6-benzyl-2-{[(tert-butoxy)carbonyl]amino}-4H,5H,6H,7H-thieno[2,3-c]pyridine-3-carboxylic acid (22)
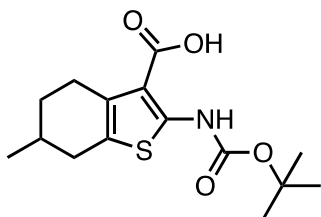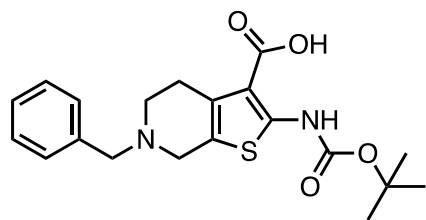

18 (600 mg, 1.44 mmol) was hydrolysed using NaOH (260 mg, 6.48 mmol) dissolved in solvent mixture containing MeOH/H$_2$O/THF in a 2:2:1 ratio (20 mL) by heating at 80 °C for 3 hours. The organic solvent was removed *in vacuo* and the resulting precipitate was filtered and dried overnight to yield the title compound as a yellow solid (378 mg, 68%). **R$_f$**: 0.22 (EtOAc); **δ$_H$** (DMSO): 13.27, (1H, s, N-H), 7.34, (s, 2H, 3'-H), 7.33 (d, 2H, *J* = 1 Hz, 2'-H), 7.28-7.25 (m, 1H, 4'-H), 3.62 (s, 2H, N-CH$_2$), 3.40 (s, 2H, 7-H), 2.83 (t, 2H, *J* = 5.5 Hz, 4-H), 2.61 (t, 2H, *J* = 5.5 Hz, 5-H), 1.45 (s, 9H, C-(CH$_3$)$_3$); **δ$_C$** (DMSO): 169.03 (COOH), 143.31 (NH-CO), 139.07 (C1'), 132.34 (C2), 129.21 (C3'), 128.68 (C2'), 127.41 (C4'), 119.97 (C7a), 119.07 (C3a), 119.17 (C3), 79.86 (C-(CH$_3$)$_3$), 61.80 (N-CH$_2$), 51.61 (C7), 50.63 (C5), 28.53 (C-(CH$_3$)$_3$), 27.35 (C4); **IR (ν$_{max}$/cm^{-1});** 3489, 3355, 2980, 2943, 2801, 1710 and 1685; **HPLC** (DMSO): 100%; **HRMS** (ES): [M-H]$^-$ Found 387.1375, requires *[M-H]*387.1457.

tert-butyl N-[6-tert-butyl-3-(phenylcarbamoyl)-4,5,6,7-tetrahydro-1-benzothiophen-2-yl]carbamate (23)
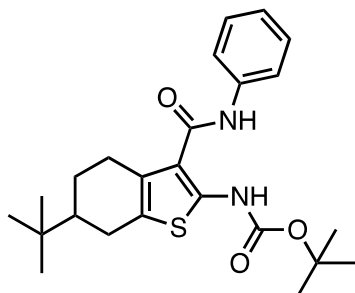
To a solution of **19** (200 mg, 0.56 mmol) and triethylamine (0.15 mL, 1.13 mmol) in EtOAc (20 mL) was added HCTU (234 mg, 0.56 mmol) and the mixture was stirred at RT for 10 minutes. Aniline hydrochloride (147 mg, 1.13 mmol) was added and the reaction was stirred at RT for a further 5 hours. The mixture was washed with water (10 mL) and extracted with EtOAc (3 x 10 mL). The organic layers were combined, dried (MgSO$_4$) and solvent removed *in vacuo*. The crude material was purified by column chromatography using petrol/EtOAc (8:1) to yield the title compound as a glassy yellow solid (125 mg, 52%). **R$_f$**: 0.91 (EtOAc); **δ$_H$** (CDCl$_3$): 10.74 (s, 1H, NH), 7.46 (d, 2H, *J* = 7.5 Hz, 2'-H), 7.30 (t, 2H, *J* = 7.5 Hz, 3'-H), 7.08 (t, 1H, *J* = 7.5 Hz, 4'-H), 2.93 (dd, 1H, *J* = 14.5, 5 Hz, 7-H), 2.71-2.68 (m, 1H, 7-H), 2.64 (dd, 1H, *J* = 16, 5 Hz, 4-H), 2.38 (t, 1H, *J* = 13 Hz, 4-H), 2.06 (dq, 1H, *J* = 11, 2 Hz, 5-H),1.51-1.47 (m, 1H, 6-H), 1.44 (s, 9H, O-C(CH$_3$)$_3$), 1.38-1.24 (m, 1H, 5-H), 0.90 (s, 9H, C(CH$_3$)$_3$); **δ$_C$** (CDCl$_3$): 164.43 (CO-NH-Ar), 152.46 (NH-CO), 148.61 (C2), 137.42 (C1'), 129.13 (C2'), 127.37 (C7a), 126.69 (C3a), 124.68 (C4'), 120.77 (C3'), 112.44 (C3), 81.30 (O-C(CH$_3$)$_3$) 44.81 (C6), 28.27 (O-C(CH$_3$)$_3$), 28.07 (C7), 27.26 (C-

($\underline{C}H_3)_3$), 25.87 (C4), 24.71 (C5); **HPLC** (DMSO): 75%; **HRMS** (ES): Found [M+Na]$^+$ 451.2028, requires [MNa] 451.2124. Product was used before IR could be obtained.

tert-butyl N-{3-[(4-chlorophenyl)carbamoyl]-4,5,6,7-tetrahydro-1-benzothiophen-2-yl}carbamate (24)
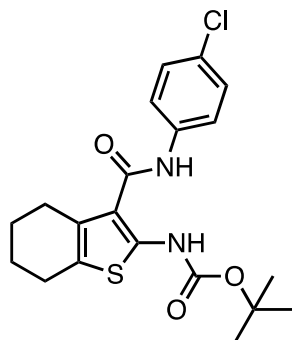
To a solution of **20** (200 mg, 0.67 mmol) and triethylamine (0.19 mL, 1.35 mmol) in DCM (20 mL) was added HCTU (278 mg, 0.67 mmol) and the mixture was stirred at RT for 10 minutes. To this was added 4-chloroaniline (171 mg, 1.35 mmol) and the reaction was stirred at RT overnight. The solvent was removed *in vacuo* and the crude material was purified by column chromatography using petrol/EtOAc (3:1) to yield the title compound as a white solid (110 mg, 40%); **R$_f$**: 0.49 (3:1 petrol/EtOAc); **δ$_H$** (CDCl$_3$): 10.76, (s, 1H, N-H), 7.56 (s, 1H, N-H), 7.51 (dt, 2H, J = 6.5, 2 Hz, 2'-H), 7.34 (dt, 2H, J = 6.5, 2 Hz, 3'-H), 2.84 (t, 2H, J = 4 Hz, 4-H), 2.71 (t, 2H, J = 4 Hz, 7-H), 1.92-1.87 (m, 4H, 6-H, 5-H), 1.53 (s, 9H, (C$\underline{H}_3)_3$); **δ$_C$** (CDCl$_3$): 164.36 ($\underline{C}$O-NH-Ar), 152.45 (NH-$\underline{C}$O), 148.71 (C2), 136.04 (C1'), 129.65 (C4'), 129.13 (C3'), 126.81 (C7a), 126.52 (C3a), 121.89 (C2'), 112.59 (C3), 81.74 ($\underline{C}$-(CH$_3$)$_3$), 28.25 (($\underline{C}H_3)_3$), 26.92 (C4), 24.30 (C7), 23.02 (C6), 22.65 (C5); **IR (v$_{max}$/cm^{-1}):** 3458, 3199, 2972, 2931, 2840, 1707 and 1632; **HPLC** (DMSO): 97%; **HRMS** (ES): Found [M+H]$^+$ 407.1188, requires *[MH]* 407.1118.

tert-butyl N-{3-[(3-methylphenyl)carbamoyl]-4,5,6,7-tetrahydro-1-benzothiophen-2-yl}carbamate (25)
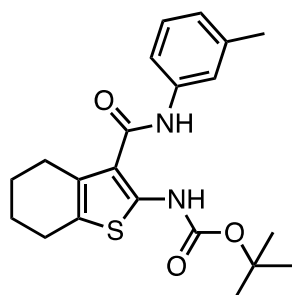
To a solution of **20** (250 mg, 0.84 mmol) and triethylamine (0.23 mL, 1.68 mmol) in DCM (20 mL) was added HCTU (348 mg, 0.84 mmol) and the mixture was stirred at RT for 10 minutes. To this was added m-toluidine (0.14 mL, 1.26 mmol) and the reaction was stirred at RT overnight. The solvent was removed *in vacuo* and the crude material was purified by column chromatography using petrol/EtOAc (3:1) to yield the title compound as a white solid (123 mg, 38%); **R$_f$**: 0.35 (3:1 petrol/EtOAc); **δ$_H$** (CDCl$_3$): 10.81 (s, 1H, N-H), 7.52 (s, 1H, N-H), 7.38 (s, 1H, 2'-H), 7.33 (d, 1H, J = 8 Hz, 6'-H), 7.26 (d, 1H, J = 8 Hz, 5'-H), 6.99 (d, 1H, J = 7.5 Hz, 4'-H), 2.84 (t, 2H, J = 4 Hz, 4-H), 2.71 (t, 2H, J = 4 Hz, 7-H), 2.40 (s, 3H, Ar-C$\underline{H}_3$), 1.92-1.88 (m, 4H, 6-H, 5-H), 1.53 (s, 9H, (C$\underline{H}_3)_3$); **δ$_C$** (CDCl$_3$): 164.42 ($\underline{C}$O-NHAr), 152.49 (NH-$\underline{C}$O), 148.25 (C2), 139.10 (C1'), 137.31 (C3'), 128.95 (C5'), 126.73 (C7a), 126.62 (C3a), 125.52 (C4'), 121.38

(C2'), 117.87 (C6'), 112.94 (C3), 81.58 ($\underline{C}$-(CH$_3$)$_3$), 28.26 (C-($\underline{C}$H$_3$)$_3$), 26.91 (C4), 24.32 (C7), 23.05 (C6), 22.70 (C5), 21.52 (Ar-$\underline{C}$H$_3$); **IR (v$_{max}$/cm^{-1}):** 3435, 3243, 2974, 2937, 2865, 1714 and 1621; **HRMS** (ES): Found [M+Na]$^+$ 409.1558, requires *[MNa]* 409.1562. HPLC could not obtained due to machine fault.

## tert-butyl N-[6-methyl-3-(phenylcarbamoyl)-4,5,6,7-tetrahydro-1-benzothiophen-2-yl]carbamate (26)

To a solution of **21** (210 mg, 0.67 mmol) and triethylamine (0.19 mL, 1.34 mmol) in DCM (20 mL) was added HCTU (280 mg, 0.67 mmol) and the mixture was stirred at RT for 10 minutes. Aniline hydrochloride (175 mg, 1.34 mmol) was added and the reaction was stirred at RT overnight. The solvent was removed *in vacuo* and the crude material was purified by column chromatography (1:1 petrol/EtOAc) to yield the title compound as a glassy yellow solid (123 mg, 47%); **R$_f$**: 0.55 (1:1 Petrol-EtOAc); **δ$_H$** (CDCl$_3$): 9.66 (s, 1H, N-H), 7.95 (d, 1H, *J* = 9 Hz, N-H), 7.45 (dd, 2H, *J* = 8.2, 6 Hz, 2'-H), 7.36 (dd, 2H, *J* = 8.9, 2 Hz, 3'-H), 7.08 (t, 1H, *J* = 7.6 Hz, 4'-H), 3.07 (dt, 1H, *J* = 17, 2.35 Hz, 4-H), 2.87-2.80 (m, 2H, 7-H, 4-H), 2.68 (dt, 2H, *J* = 16.7, 5.2 Hz, 7-H, 6-H), 2.26-2.20 (m, 2H, 5-H), 1.50 (s, 9H, C-(C$\underline{H}$$_3$)$_3$), 1.43 (d, 3H, *J* = 7.8 Hz, C-C$\underline{H}$$_3$); **δ$_C$** (CDCl$_3$): 162.1 ($\underline{C}$O-NH-Ar), 156.2 (NH-$\underline{C}$O), 151.6 (C2), 137.4 (C7a), 135.5 (C1'), 129.6 (C3a), 129.1 (C3'), 126.4 (C3), 126.3 (C2'), 124.7 (C4'), 81.6 (O-$\underline{C}$(CH$_3$)$_3$), 32.4 (C7), 31.1 (C5), 29.1 (C6), 28.2 (C-($\underline{C}$H$_3$)), 26.7 (C4), 21.3 ($\underline{C}$H$_3$); **IR (v$_{max}$/cm^{-1});** 3090, 2952, 2925, 2161, 2099, 1736 and 1717; **HRMS** (ES): Found [M+Na]$^+$ 409.1553, requires *[MNa]* 409.1562. HPLC could not obtained due to machine fault.

tert-butyl N-(3-{[(4-methoxyphenyl)methyl]carbamoyl}-6-methyl-4,5,6,7-tetrahydro-1-benzothiophen-2-yl)carbamate (27)
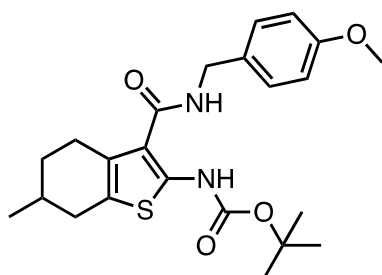
To a solution of **21** (210 mg, 0.67 mmol) and triethylamine (0.19 mL, 1.34 mmol) in DCM (20 mL) was added HCTU (280 mg, 0.67 mmol) and the mixture was stirred at RT for 10 minutes. 4-methoxybenzylamine (0.17 mL, 1.34 mmol) was added and the reaction was stirred at RT overnight. The solvent was removed *in vacuo* and the crude material was purified by column chromatography using petrol/EtOAc (1:1) to yield the title compound as a white solid (100 mg, 31%); **R$_f$**: 0.58 (1:1 Petrol-EtOAc); **δ$_H$** (CDCl$_3$): 11.06 (s, 1H, N-H), 7.26 (dd, 2H, *J* = 7, 3 Hz, 2'H), 6.91 (dd, 2H, *J* = 6.5, 2 Hz, 3'-H), 6.12 (s, 1H, N-H), 4.55 (t, 2H,

J = 6 Hz, NH-C$\underline{H}_2$), 3.83 (s, 3H, O-C$\underline{H}_3$), 2.75-2.71 (m, 2H, 4-H, 7-H), 2.62-2.60 (m, 1H, 4-H), 2.28 (dq, 1H, J = 6.5, 2Hz, 7-H), 1.92-1.88 (m, 2H, 5-H, 6-H), 1.53 (s, 9H, C-(C$\underline{H}_3$)$_3$), 1.45-1.41 (m, 1H, 5-H); δ_C (CDCl$_3$): 166.12 ($\underline{C}$O-NH-CH$_2$), 159.08 (C4'), 152.49 (NH-$\underline{C}$O), 147.75 (C2), 130.18 (C7a), 128.90 (C2'), 126.71 (C1'), 125.88 (C3a), 114.24 (C3'), 112.11 (C3), 81.36 ($\underline{C}$-(CH$_3$)$_3$), 55.33 (O-$\underline{C}$H$_3$), 43.00 (NH-$\underline{C}$H$_2$), 32.42 (C7), 31.09 (C5), 28.97 (C6), 28.28 (C-($\underline{C}$H$_3$)$_3$), 26.60 (C4), 21.26 (C-$\underline{C}$H$_3$); **IR (v$_{max}$/cm^{-1});** 3481, 3167, 2976, 2959, 2931, 2873, 2833, 1700 and 1614; **HPLC** (DMSO): 100%; **HRMS** (ES): Found [M-H]$^-$ 429.1848, requires *[M-H]* 429.1926.

tert-butyl N-{6-methyl-3-[(3-methylphenyl)carbamoyl]-4,5,6,7-tetrahydro-1-benzothiophen-2-yl}carbamate (28)
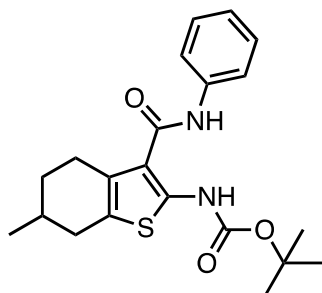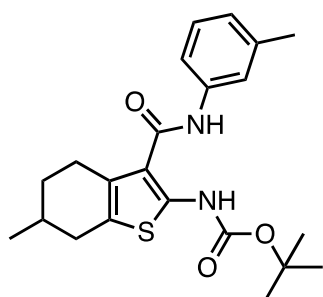

To a solution of **21** (250 mg, 0.80 mmol) and triethylamine (0.22 mL, 1.60 mmol) in DCM (20 mL) was added HCTU (332 mg, 0.80 mmol) and the mixture was stirred at RT for 10 minutes. To this was added m-toluidine (0.18 mL, 1.20 mmol) and the reaction was stirred at RT overnight. The solvent was removed *in vacuo* and the crude material was purified by column chromatography using petrol/EtOAc (3:1) to yield the title compound as a white solid (112 mg, 35%); **R$_f$**: 0.37 (3:1 petrol/EtOAc); δ_H (CDCl$_3$): 10.83 (s, 1H, N-H), 7.53 (s, 1H, N-H), 7.39 (s, 1H, 2'-H), 7.33 (d, 1H, J = 8 Hz, 6'-H), 7.26 (d, 1H, J = 8 Hz, 5'-H), 6.99 (d, 1H, J = 7.5 Hz, 4'-H), 2.94 (dt, 1H, J = 14, 3 Hz, 4-H), 2.84-2.81 (m, 2H, 4-H, 7-H), 2.40 (s, 3H, Ar-C$\underline{H}_3$), 2.35-2.30 (m, 1H, 7-H), 2.03-1.96 (m, 2H, 6-H, 5-H), 1.53 (s, 9H, (C$\underline{H}_3$)$_3$), 1.51-1.49 (m, 1H, 5-H), 1.12 (d, 3H, J = 6.5 Hz, C-C$\underline{H}_3$); δ_C (CDCl$_3$): 164.44 ($\underline{C}$O-NH-Ar), 148.41 (C2), 139.10 (C1'), 137.31 (C3'), 128.95 (C5'), 126.36 (C7a), 125.52 (C4'), 125.26 (C3a), 121.40 (C2'), 117.88 (C6'), 112.74 (C3), 81.58 ($\underline{C}$-(CH$_3$)$_3$), 32.45 (C7), 31.16 (C5), 29.05 (C6), 28.26 (C-($\underline{C}$H$_3$)$_3$), 26.71 (C4), 21.52 (Ar-$\underline{C}$H$_3$), 21.29 (C-$\underline{C}$H$_3$), $\underline{C}$O-NH was unresolved; **IR (v$_{max}$/cm^{-1}):** 3446, 3207, 2973, 2949, 2922, 2869, 2841, 1709 and 1627; **HPLC** (DMSO): 100%; **HRMS** (ES): Found [M+H]$^+$ 401.1892, requires *[MH]* 401.1821.

tert-butyl N-[6-benzyl-3-(phenylcarbamoyl)-4H,5H,6H,7H-thieno[2,3-c]pyridin-2-yl]carbamate (29)
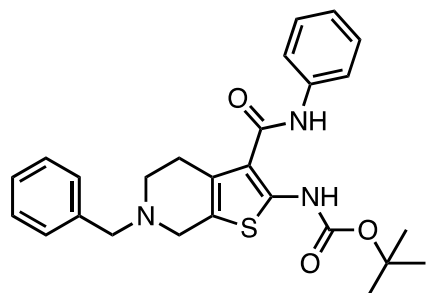
To a solution of **22** (200 mg, 0.51 mmol) and triethylamine (0.14 mL, 1.03 mmol) in DCM (20 mL) was added HCTU (214 mg, 0.51 mmol) and the mixture was stirred at RT for 30 minutes. Aniline hydrochloride (133 mg, 1.03 mmol) was added and the reaction was stirred at RT overnight. The solvent was removed *in vacuo* and the crude material was purified by column chromatography (2:1 Petrol-EtOAc) to yield the title compound as a glassy yellow solid (90 mg, 37%); **R$_f$**: 0.84 (EtOAc); **δ$_H$** (CDCl$_3$): 10.79 (s, 1H, N-H), 7.51 (dd, 2H, J = 1.5, 9 Hz, 2''-H), 7.43-7.37 (m, 2H, 2'-H,), 7.32 (dt, 1H, J = 7, 1.5 Hz, 4'-H) 7.20-7.16 (m, 4H, 3''-H, 3'-H), 6.78 (t, 1H, J = 7.5 Hz, 4''-H), 3.75 (s, 2H, N-C$\underline{H}_2$), 3.65 (s, 2H, 7-H), 2.96 (t, 2H, J = 5 Hz, 4-H), 2.87 (t, 2H, J = 5 Hz, 5-H), 1.53 (s, 9H, O-C(C$\underline{H}_3$)$_3$); **δ$_C$** (CDCl$_3$): 164.20 ($\underline{C}$O-NH-Ar), 148.97 (NH-$\underline{C}$O), 146.35 (C2), 137.78 (C1'), 137.29 (C1''), 129.15 (C3'), 128.48 (C2'), 127.46 (C4'), 125.45 (C7a), 124.77 (C3''), 124.28 (C3a), 120.74 (C2''), 115.12 (C4''), 112.33 (C3), 81.79 (O-$\underline{C}$(CH$_3$)$_3$), 61.97 (N-$\underline{C}$H$_2$), 51.64 (C7), 49.72 (C5), 28.25 (C-($\underline{C}$H$_3$)$_3$), 27.25 (C4); **IR (v$_{max}$/cm^{-1});** 3271, 2975, 2929, 2806, 1712 and 1634; **HPLC** (DMSO): 74%; **HRMS** (ES): Found [M+Na]$^+$ 486.1816, requires *[MNa]* 486.1827.

tert-butyl N-(6-benzyl-3-{[(4-methoxyphenyl)methyl]carbamoyl}-4H,5H,6H,7H-thieno[2,3-c]pyridin-2-yl)carbamate (30)
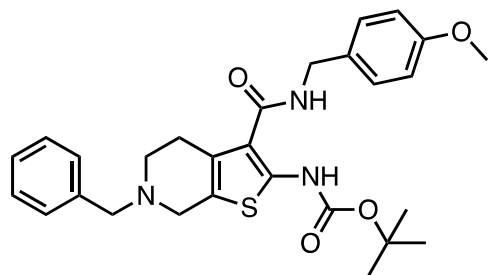
To a solution of **22** (200 mg, 0.51 mmol) and triethylamine (0.14 mL, 1.03 mmol) in DCM (20 mL) was added HCTU (214 mg, 0.51 mmol) and the mixture was stirred at RT for 30 minutes. 4-methoxybenzylamine (0.19 mL, 1.03 mmol) was added and the reaction was stirred at RT overnight. The solvent was removed *in vacuo* and the crude material was purified by column chromatography (1:1 Petrol-EtOAc) to yield the title compound as a yellow solid (80 mg, 31%); **R$_f$**: 0.21 (EtOAc); **δ$_H$** (CDCl$_3$): 11.00 (s, 1H, N-H), 7.33 (q, 4H, J = 8 Hz, 2'-H, 3'-H), 7.28 (t, 1H, J = 5 Hz, 4'-H), 7.21 (d, 2H, J = 8.5, 3 Hz, 2''-H), 6.87 (dt, 2H, J = 8.5, 3 Hz, 3''-H), 5.99 (s, 1H, N-H), 4.50 (d, 2H, J = 5.5 Hz, NH-C$\underline{H}_2$), 3.80 (s, 3H, O-C$\underline{H}_3$), 3.67 (s, 2H, N-C$\underline{H}_2$), 3.56 (s, 2H, 7-H), 2.74 (t, 2H, J = 5 Hz, 4-H), 2.70 (t, 2H, J = 5 Hz, 5-H); **δ$_C$** (CDCl$_3$): 165.83 ($\underline{C}$O-NH-CH$_2$Ar), 159.11 (C4''), 152.44 (NH-$\underline{C}$O), 148.23 (C2), 137.88 (C1''), 130.00 (C1'), 129.10

(C3''), 128.89 (C2'), 128.42 (C3'), 127.37 (C4'), 125.79 (C7a), 123.79 (C3a), 114.27 (C2''), 111.79 (C3), 81.55 (C-(CH$_3$)$_3$), 61.89 (N-CH$_2$), 55.32 (O-CH$_3$), 51.63 (C7), 49.63 (C5), 43.02 (NH-CH$_2$), 28.27 (C-(CH$_3$)$_3$), 27.11 (C4); **IR (v$_{max}$/cm^{-1}):** 3218, 2976, 2930, 2835, 1710 and 1613; **HRMS** (ES): Found [M+H]$^+$ 508.2246, requires *[MH]* 508.2192. HPLC could not be obtained due to machine fault.

3-{[6-tert-butyl-3-(phenylcarbamoyl)-4,5,6,7-tetrahydro-1-benzothiophen-2-yl]carbamoyl}propanoic acid (31)
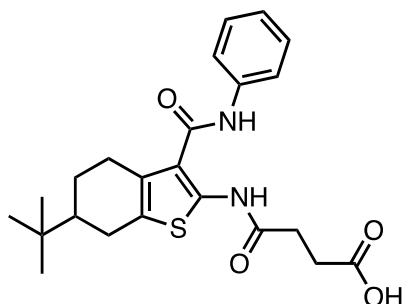
The boc-protected material **23** (110mg, 0.26 mmol) was stirred in a 1:1 ratio of HCl in dioxane (1 mL) and DCM (1 mL) for 8 hours. The deprotected material was isolated by removing the solvent *in vacuo* which was used crude without further purification. The material (40 mg, 0.12 mmol) and succinic anhydride (15 mg, 0.15 mmol) were dissolved in anhydrous DCM (5 mL) and the reaction mixture was heated at reflux for 3 hours. The solvent was removed *in vacuo* and the crude material was purified via column chromatography (EtOAc) to yield the title compound as a brown solid (8.7 mg, 16%). **R$_f$:** 0.25 (EtOAc); **δ$_H$** (CDCl$_3$): 11.86 (s, 1H, N-H), 7.56 (s, 1H, N-H), 7.45 (d, 2H *J* = 8 Hz, 2'-H), 7.31 (t, 2H, *J* = 8 Hz, 3'-H) 7.10 (t, 1H, *J* = 8 Hz, 4'-H), 2.94 (dd, 1H, *J* = 15, 5 Hz, 7-H), 2.71-2.69 (m, 3H, 7-H, CH$_2$-COOH), 2.65 (d, 1H, *J* = 5.5 Hz, 4-H), 2.38 (t, 1H, *J* = 13 Hz, 4-H), 2.07 (dd, 1H, *J* = 13, 5 Hz, 5-H), 1.50-1.43 (m, 1H, 6-H), 1.38-1.32 (m, 1H, 5-H), 1.18 (s, 2H, CO-CH$_2$), 0.89 (9H, s, (CH$_3$)$_3$); **δ$_C$** (CDCl$_3$): 176.38 (COOH), 168.55 (CO-CH$_2$), 164.54 (NH-CO), 146.10 (C1'), 137.12 (C7a), 129.21 (C2'), 126.42 (C3a), 125.00 (C4'), 120.96 (C3'), 113.90 (C3), 32.47 (C(CH$_3$)$_3$), 31.01 (CH$_2$COOH), 30.2 (C6), 28.86 (CO-CH$_2$), 27.94 (C7), 27.25 (C(CH$_3$)$_3$), 25.93 (C4), 24.68 (C5); **IR (v$_{max}$/cm^{-1});** 3282, 3043, 2958, 2925, 2853, 1713 and 1664; **HPLC** (DMSO): 82%; **HRMS** (ES): Found [M-H]$^-$ 427.1689, requires *[M-H]* 427.1770.

3-({3-[(4-chlorophenyl)carbamoyl]-4,5,6,7-tetrahydro-1-benzothiophen-2-yl}carbamoyl)propanoic acid (32)
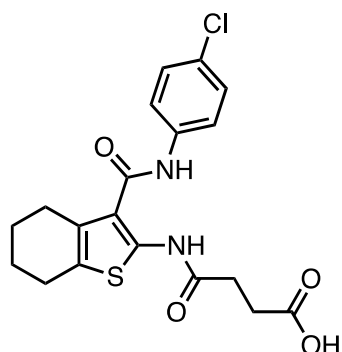
The boc-protected material **24** (100 mg, 0.26 mmol) was stirred in a 1:1 ratio of HCl in dioxane (1.5 mL) and DCM (1.5 mL) for 8 hours. The deprotected material was isolated by removing the solvent *in vacuo* which was used crude without further purification. The material (80 mg, 0.26 mmol) and succinic anhydride (52 mg, 0.52 mmol) were stirred in anhydrous THF (5 mL) overnight at 65 °C. The solvent was removed *in vacuo* and the crude material was purified via reverse phase chromatography (H$_2$O-MeCN gradient) to yield the title compound as a yellow solid (24.7 mg, 23%). **R$_f$**: 0.54 (EtOAc); **δ$_H$** (MeOD): 7.55 (dd, 2H, *J* = 7, 2 Hz, 2'-H), 7.23 (dd, 2H, *J* = 7, 2 Hz, 3'-H), 2.84 (s, 2H, 4-H), 2.65-2.55 (m, 6H, 7-H, CO-C$\underline{H}_2$, C$\underline{H}_2$-COOH), 1.80-1.70 (m, 4H, 6-H, 5-H); **δ$_C$** (MeOD): 174.71 (COOH), 172.41 (NH-$\underline{C}$O-CH$_2$), 164.58 ($\underline{C}$O-NH), 149.16 (C2), 137.09 (C1'), 129.23 (C4'), 128.87 (C7a), 128.36 (C3'), 128.10 (C3a), 121.72 (C2'), 119.28 (C3), 30.34 ($\underline{C}$H$_2$-COOH), 28.41 (CO-$\underline{C}$H$_2$), 24.87 (C4), 23.81 (C7), 22.78 (C6), 22.56 (C5); **IR (v$_{max}$/cm^{-1}):** 3328, 3009, 2927, 2655, 1721, 1687 and 1618; **HPLC** (DMSO): 97%; **HRMS** (ES): Found [M-H]$^-$ 405.0676, requires *[M-H]* 405.0754.

3-({3-[(3-methylphenyl)carbamoyl]-4,5,6,7-tetrahydro-1-benzothiophen-2-yl}carbamoyl)propanoic acid (33)
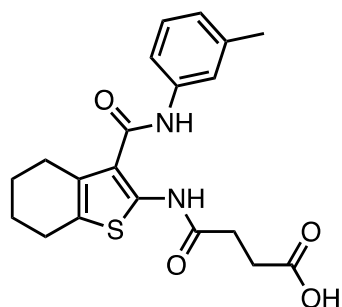
The boc-protected material **25** (110 mg, 0.28 mmol) was stirred in a 1:1 ratio of HCl in dioxane (1.5 mL) and DCM (1.5 mL) for 14 hours. The deprotected material was isolated by removing the solvent *in vacuo* which was used crude without further purification. The material (90 mg, 0.31 mmol) and succinic anhydride (63 mg, 0.63 mmol) were stirred in anhydrous THF (4 mL) overnight at 65 °C. The solvent was removed *in vacuo* and the crude material was purified via reverse phase chromatography (H$_2$O-MeCN gradient) to yield the title compound as a yellow solid (6 mg, 3%). **R$_f$**: 0.46 (EtOAc); **δ$_H$** (DMSO): 12.16 (s, 1H, N-H), 10.63 (s, 1H, N-H), 9.72 (s, 1H, O-H), 7.58 (s, 1H, 2'-H), 7.48 (d, 1H, *J* = 7.5 Hz, 6'-H), 7.20 (t, 1H, *J* = 7.5 Hz, 5'-H), 6.90 (d, 1H, *J* = 7.5 Hz, 4'-H), 2.63-2.62 (m, 6H, 7-H, 4-H, C$\underline{H}_2$-COOH), 2.5 (s, 2H, CO-C$\underline{H}_2$), 2.30 (s, 3H, Ar-C$\underline{H}_3$), 1.77 (d, 2H, *J* = 4 Hz, 6-H), 1.71 (d, 2H, *J* = 4 Hz, 5-H); **δ$_C$** (DMSO): 174.06 ($\underline{C}$OOH), 169.50 ($\underline{C}$O-NHAr), 163.55 (NH-$\underline{C}$O), 139.51 (C2), 138.49 (C1'), 138.14 (C3'), 130.01 (C7a),

128.84 (C5'), 126.72 (C3a), 124.59 (C4'), 120.91 (C2'), 120.41 (C2), 117.59 (C6'), 30.69 (C$\underline{\text{H}}_2$-COOH), 29.26 (CO-C$\underline{\text{H}}_2$), 24.98 (C4), 24.18 (C7), 23.21 (C6), 22.78 (C5), 21.69 (Ar-C$\underline{\text{H}}_3$); **IR (v$_{max}$/cm^{-1}):** 3321, 3284, 3041, 2921, 2846, 1704, 1687 and 1613; **HPLC** (DMSO): 93%; **HRMS** (ES): Found [M-H]$^-$ 385.1222, requires *[M-H]* 385.1300.

### 3-{[6-methyl-3-(phenylcarbamoyl)-4,5,6,7-tetrahydro-1-benzothiophen-2-yl]carbamoyl}propanoic acid (34)
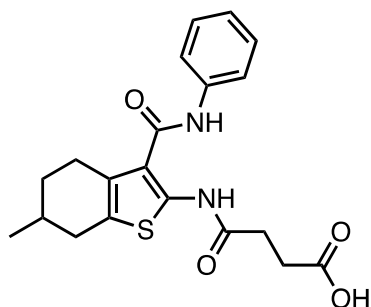
The boc-protected material **26** (120 mg, 0.31 mmol) was stirred in a 1:1 ratio of HCl in dioxane (1.5 mL) and DCM (1.5 mL) for 8 hours. The deprotected material was isolated by removing the solvent *in vacuo* which was used crude without further purification. The material (80 mg, 0.28 mmol) and succinic anhydride (34 mg, 0.36 mmol) were stirred in anhydrous THF (4 mL) overnight at 65 °C. The solvent was removed *in vacuo* and the crude material was purified via reverse phase chromatography (H$_2$O-MeCN gradient) to yield the title compound as a white solid (12.3 mg, 10%). **R$_f$:** 0.51 (EtOAc); **δ$_H$** (DMSO): 12.16 (s, 1H, O-H), 10.64 (s, 1H, N-H), 9.81 (s, 1H, N-H), 7.70 (dd, 2H, *J* = 9, 1 Hz, 2'-H), 7.32 (t, 2H, *J* = 7.5 Hz, 3'-H), 7.08 (dt, 1H, *J* = 7.5, 1 Hz, 4'-H), 2.73 (dd, 1H, *J* = 16, 5 Hz, 7-H), 2.65-2.64 (m, 2H, 4-H), 2.63 (t, 2H, *J* = 7, C$\underline{\text{H}}_2$-COOH), 2.53 (t, 2H, *J* = 2, CO-CH$_2$), 2.23 (dq, 1H, *J* = 10, 2 Hz, 7-H), 1.88-1.80 (m, 2H, 6-H, 5-H), 1.37-1.28 (m, 1H, 5-H), 1.05 (d, 3H, *J* = 2.5 Hz, C-C$\underline{\text{H}}_3$); **δ$_C$** (DMSO): 174.05 (C$\underline{\text{O}}$OH), 169.49 (C$\underline{\text{O}}$-NH-Ar), 163.61 (NH-C$\underline{\text{O}}$), 139.61 (C1'), 138.55 (C2), 129.71 (C7a), 128.99 (C3'), 126.44 (C3a), 123.90 (C4'), 120.42 (C2'), 120.22 (C3), 32.31 (C7), 31.02 (C5), 30.67 (C$\underline{\text{H}}_2$-COOH), 29.49 (C6), 29.24 (CO-C$\underline{\text{H}}_2$), 24.72 (C4), 21.77 (C-C$\underline{\text{H}}_3$); **IR (v$_{max}$/cm^{-1});** 3307, 2923, 2840, 1706, 1687 and 1613; **HPLC** (DMSO): 93%; **HRMS** (ES): Found [M-H]$^-$ 385.1222, requires *[M-H]* 385.1300.

3-[(3-{[(4-methoxyphenyl)methyl]carbamoyl}-6-methyl-4,5,6,7-tetrahydro-1-benzothiophen-2-yl)carbamoyl]propanoic acid (35)
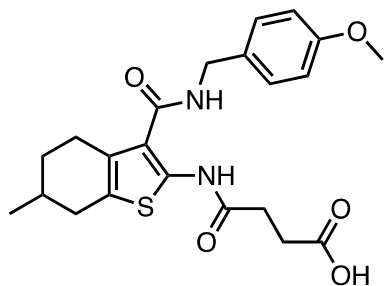
The boc-protected material **27** (80 mg, 0.19 mmol) was stirred in a 1:1 ratio of HCl in dioxane (1.5 mL) and DCM (1.5 mL) for 8 hours. The deprotected material was isolated by removing the solvent *in vacuo* which was used crude without further purification. The material (95 mg, 0.29 mmol) and succinic anhydride (35 mg, 0.35 mmol) were stirred in anhydrous THF (3 mL)

overnight at 65 °C. The solvent was removed *in vacuo* and the crude material was purified via reverse phase chromatography (H$_2$O-MeCN gradient) to yield the title compound as a white solid (5 mg, 25%). **R$_f$**: 0.54 (EtOAc); **δ$_H$** (DMSO): 11.12 (s, 1H, O-H), 8.14 (s, 1H, N-H), 7.96 (s, 1H, N-H), 7.26 (d, 2H, *J* = 8.5 Hz, 2'-H), 6.89 (d, 2H, *J* = 8.5 H, 3'-H), 4.38 (dd, 2H, *J* = 15, 6 Hz, NH-C$\underline{H}_2$), 2.70 (d, 1H, *J* = 5 Hz, 7-H), 2.68-2.66 (m, 2H, 4-H), 2.62 (t, 2H, *J* = 7 Hz, C$\underline{H}_2$-COOH), 2.53 (s, 2H, CO-C$\underline{H}_2$), 2.21 (dq, 1H, *J* = 9.5, 1.5 Hz, 7-H), 1.83-1.79 (m, 2H, 6-H, 5-H), 1.37–1.27 (m, 1H, 5-H), 1.02 (d, 3H, *J* = 6.5 Hz, C-C$\underline{H}_3$); **δ$_C$** (DMSO): 174.02 ($\underline{C}$OOH), 169.20 ($\underline{C}$O-NH-CH$_2$), 165.35 (NH-$\underline{C}$O), 163.59 (C4'), 158.62 (C2), 140.77 (C1'), 131.79 (C7a), 129.05 (C2'), 128.98 (C3a), 117.68 (C3), 114.15 (C3'), 55.50 (O-$\underline{C}$H$_3$), 42.22 ($\underline{C}$H$_2$-NH), 32.31 (C7), 30.96 (C5), 30.94 ($\underline{C}$H$_2$-COOH), 29.20 (C6), 29.17 (CO-$\underline{C}$H$_2$), 25.14 (C4), 21.62 (C-C$\underline{H}_3$); **IR (v$_{max}$/cm^{-1});** 3308, 3009, 2924, 2924, 2837, 1723, 1683 and 1612; **HPLC** (DMSO): 96%; **HRMS** (ES): Found [M-H]$^-$ 429.1488, requires *[M-H]* 429.1562.

### 3-({6-methyl-3-[(3-methylphenyl)carbamoyl]-4,5,6,7-tetrahydro-1-benzothiophen-2-yl}carbamoyl)propanoic acid (36)
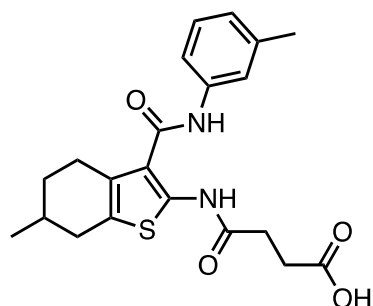
The boc-protected material **28** (100 mg, 0.25 mmol) was stirred in a 1:1 ratio of HCl in dioxane (1.5 mL) and DCM (1.5 mL) for 14 hours. The deprotected material was isolated by removing the solvent *in vacuo* which was used crude without further purification. The material (70 mg, 0.23 mmol) and succinic anhydride (47 mg, 0.47 mmol) were stirred in anhydrous THF (4 mL) overnight at 65 °C. The solvent was removed *in vacuo* and the crude material was purified via reverse phase chromatography (H$_2$O-MeCN gradient) to yield the title compound as a white solid (5 mg, 6%). **R$_f$**: 0.43 (EtOAc); **δ$_H$** (DMSO): 12.22 (s, 1H, O-H), 10.69 (s, 1H, N-H), 9.78 (s, 1H, N-H), 7.63 (s, 1H, 2'-H), 7.54 (d, 1H, *J* = 7.5 Hz, 6'-H), 7.26 (t, 1H, *J* = 8 Hz, 5'-H), 6.96 (d, 1H, *J* = 7.5 Hz, 4'-H), 2.78 (dd, 1H, *J* = 16, 5 Hz, 7-H), 2.64-2.62 (m, 4H, 4-H, C$\underline{H}_2$-COOH), 2.50 (s, 2H, CO-C$\underline{H}_2$), 2.36 (s, 3H, Ar-C$\underline{H}_3$), 2.33-2.26 (m, 1H, 7-H), 1.93-1.85 (m, 2H, 6-H, 5-H), 1.42-1.33 (m, 1H, 5-H), 2.97 (d, 3H, *J* = 6.5 Hz, C-C$\underline{H}_3$); **δ$_C$** (DMSO): 174.05 (COOH), 169.49 (NH-$\underline{C}$O-CH$_2$), 163.59 ($\underline{C}$O-NH-Ar), 139.51 (C2), 138.66 (C1'), 138.14 (C3'), 129.68 (C7a), 128.84 (C5'), 126.42 (C3a), 124.59 (C4'), 120.92 (C2'), 120.20 (C3), 117.60 (C6'), 32.31 (C7), 31.02 (C5), 30.68 ($\underline{C}$H$_2$-COOH), 29.47 (C6), 29.26 (CO-$\underline{C}$H$_2$), 24.73 (C4), 21.77 (Ar-C$\underline{H}_3$), 21.70 (C-C$\underline{H}_3$); **IR (v$_{max}$/cm^{-1}):** 3273, 2956, 2920, 2842, 1706, 1687 and 1611; **HPLC** (DMSO): 93%; **HRMS** (ES): Found [M-H]$^-$ 399.1380, requires *[M-H]* 399.4930.

## 3-{[6-benzyl-3-(phenylcarbamoyl)-4H,5H,6H,7H-thieno[2,3-c]pyridin-2-yl]carbamoyl}propanoic acid (37)

The boc-protected material **29** (85 mg, 0.18 mmol) was stirred in a 1:1 ratio of HCl in dioxane (1.5 mL) and DCM (1.5 mL) for 8 hours. The deprotected material was isolated by removing the solvent *in vacuo* which was used crude without further purification. The material (65 mg, 0.18 mmol) and triethylamine (0.08 mL, 0.54 mmol) were stirred in anhydrous THF (4 mL) at RT for 10 minutes. To this was added succinic anhydride (54 mg, 0.54 mmol) and the reaction was stirred at RT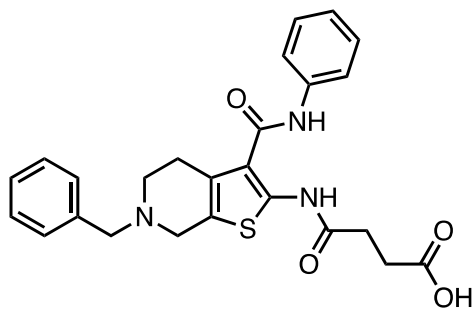 overnight. The solvent was removed *in vacuo* and the crude material was purified via reverse phase chromatography (H$_2$O-MeCN gradient) to yield the title compound as a yellow solid (12 mg, 14%). **R$_f$**: 0.41 (EtOAc); **δ$_H$** (DMSO): 12.17 (s, 1H, O-H), 10.73 (s, 1H, N-H), 9.78 (s, 1H, N-H), 7.70 (d, 2H, *J* = 4.5 Hz, 2''-H), 7.38-7.27 (m, 7H, 2'-H, 3'-H, 4'-H, 3''-H), 7.08 (t, 1H, *J* = 4.5 Hz, 4''-H), 3.69 (s, 2H, N-CH$_2$), 3.53 (s, 2H, 7-H), 2.76 (t, 2H, *J* = 6 Hz, 4-H), 2.70 (t, 2H, *J* = 6 Hz, 5-H), 2.65 (2H, s, CH$_2$-COOH), 2.51 (2H, s, CO-CH$_2$); **δ$_C$** (DMSO): 174.04 (COOH), 169.56 (CO-NH-Ar), 163.43 (NH-CO), 139.78 (C1'), 139.49 (C1''), 138.83 (C2), 129.25 (C3''), 129.00 (C2'), 128.76 (C3'), 128.66 (C7a), 127.54 (C4'), 124.64 (C3a), 123.98 (C4''), 120.52 (C2''), 119.32 (C3), 61.43 (N-CH$_2$), 51.30 (C7), 49.99 (C5), 30.71 (CH$_2$-COOH), 29.21 (CO-CH$_2$), 25.43 (C4); **IR (v$_{max}$/cm^{-1}):** 3405, 3258, 3119, 2914, 1718, 1584 and 1638; **HPLC** (DMSO): 97%; **HRMS** (ES): Found [M-H]$^-$ 462.1487, requires *[M-H]* 462.1566.

3-[(6-benzyl-3-{[(4-methoxyphenyl)methyl]carbamoyl}-4H,5H,6H,7H-thieno[2,3-c]pyridin-2-yl)carbamoyl]propanoic acid (38)
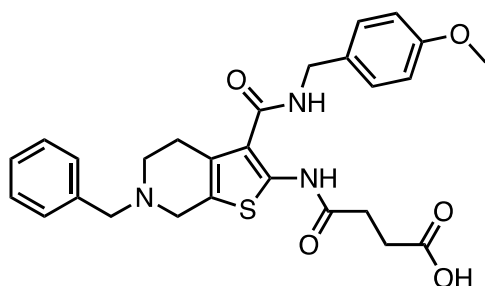

The boc-protected material **30** (80 mg, 0.16 mmol) was stirred in a 1:1 ratio of HCl in dioxane (1.5 mL) and DCM (1.5 mL) for 8 hours. The deprotected material was isolated by removing the solvent *in vacuo* which was used crude without further purification. The material (70 mg, 0.17 mmol) and succinic anhydride (34 mg, 0.34 mmol) were stirred in anhydrous THF (3 mL) overnight at 65 °C. The solvent was removed *in vacuo* and the crude material was purified via reverse phase chromatography (H$_2$O-MeCN gradient) to yield the title compound as a yellow solid

(19.8 mg, 23%). **R$_f$**: 0.57 (EtOAc); **δ$_H$** (DMSO): 8.16 (s, 1H, N-H), 7.34 (d, 4H, J = 4 Hz, 2'-H, 3'-H), 7.27 (t, 1H, J = 4 Hz, 4'-H), 7.21 (d, 2H, J = 8.5 Hz, 2''-H), 6.86 (d, 2H, J = 8.5 Hz, 3''-H), 6.79 (s, 1H, N-H), 4.31 (d, 2H, J = 6 Hz, NH-C$\underline{H}_2$), 3.72 (s, 3H, O-CH$_3$), 3.63 (s, 2H, N-C$\underline{H}_2$), 3.34 (s, 2H, 7-H), 2.71 (t, 2H, J = 6 Hz, 4-H), 2.68 (t, 2H, J = 5 Hz, 5-H), 2.66-2.63 (m, 2H, C$\underline{H}_2$-COOH), 2.42 (s, 2H, CO-C$\underline{H}_2$); **δ$_C$** (DMSO): 174.06 (COOH), 173.97 (CO-NH-CH$_2$Ar), 165.15 (NH-$\underline{C}$O), 158.62 (C4''), 143.73 (C1''), 141.66 (C1'), 131.75 (C2), 129.25 (C2''), 129.03 (C2'), 128.74 (C3'), 127.96 (C7a), 127.55 (C4'), 124.46 (C3a), 116.91 (C3), 114.15 (C3''), 61.20 (N-$\underline{C}$H$_2$), 55.50 (O-$\underline{C}$H$_3$), 51.35 (C7), 49.74 (C5), 42.45 (NH-$\underline{C}$H$_2$), 31.17 ($\underline{C}$H$_2$-COOH), 29.27 (CO-CH$_2$), 25.81 (C4); **HPLC** (DMSO): 95%; **HRMS** (ES): Found [M-H]$^-$ 506.1763, requires *[M-H]* 506.1828. Compound degraded before IR was obtained.

9   Appendix

9.1   Appendix A

IC$_{50}$ curves for the synthesised compounds described in Section 5.2.4.
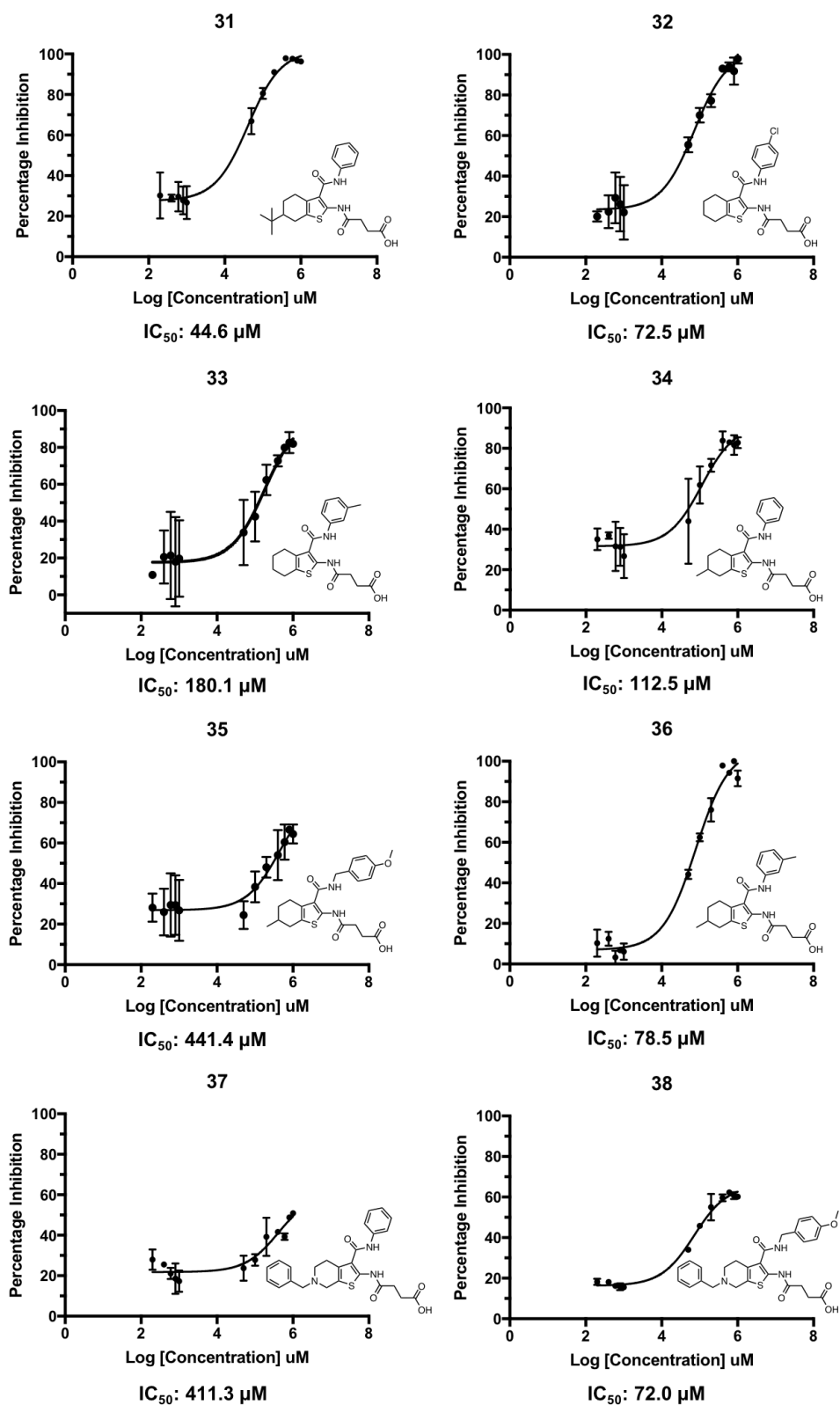
Figure 9.1: The curves used to calculate IC$_{50}$ values for the synthesised compounds.

9.2 Appendix B

The papers which directly contributed to the thesis are:

Chapter 1:

- **Johnson, R. M.**, Gordon, J. A., Rawson, S., McPhillie, M. J., Fishwick, C. W. G., Muench, S. P**.** The growing role of electron microscopy in anti-parasitic drug discovery. *Current medicinal chemistry*. 25 (39), 5279-5290 (2018).

- Rawson, S., McPhillie, M. J., **Johnson, R. M.**, Fishwick, C. W. G. & Muench, S. P. The potential use of single-particle electron microscopy as a tool for structure-based inhibitor design research papers*. Acta Crystallogr. Sect. D,* D73, 1–7 (2017).

- Johnson, R. M., Higgins, A. J., and Muench, S. P. *Emerging role of electron microscopy in drug discovery*. Trends in Biochemical Sciences, August 2019. *In press*.

Chapter 2:

- Drulyte, I., **Johnson, R. M**., Hesketh, E. L., Hurdiss, D. L., Scarff, C. A., Porav, S. A., Ranson, N. A., Muench, S. P. & Thompson, R. F. Approaches to altering particle distributions in cryo-electron microscopy sample preparation. *Acta Crystallogr. Sect. D Struct. Biol.* 74, 1–12 (2018).

Chapter 3:

- Amporndanai, K*., **Johnson, R. M***., O'Neill, P. M., Fishwick, C. W. G., Jamson, A. H., Rawson, S., Muench, S. P., Samar Hasnain, S. & Antonyuk, S. V. X-ray and cryo-EM structures of inhibitor-bound cytochrome bc_1 complexes for structure-based drug discovery. *IUCrJ.* 5, 200–210 (2018)

Discussion:

- Gopalasingam, C. C., **Johnson, R. M.**, Chiduza, G. N., Tosha T., Yamamoto M., Shiro, Y., Antonyuk, S. V., Muench, S. P., and Hasnain, S.S. Dimeric structures of quinol-dependent Nitric Oxide Reductases (qNOR) revealed by cryo-Electron Microscopy. Science Advances, 5 (8), eaax1803 (2019).

Publication which did not contribute to work in the thesis:

- Chiduza, G. N., **Johnson, R. M.**, Wright, G. S. A., Antonyuk, S. V., Muench, S. P., and Hasnain, S.S. LAT1 (SLC7A5) and CD98hc (SLC3A2) complex dynamics revealed by single-particle cryo-EM. *Acta Crystallogr. Sect. D Struct. Biol.* 75 (7), 660-669 (2019).

10 References

1. Ator, M. A., Mallamo, J. P. & Williams, M. Overview of Drug Discovery and Development. *Curr. Protoc. Pharmacol.* **Chapter 9,** 1–26 (2006).
2. Van Norman, G. A. Drugs, Devices, and the FDA: Part 1: An Overview of Approval Processes for Drugs. *JACC Basic to Transl. Sci.* **1,** 170–179 (2016).
3. Hughes, J. P., Rees, S. S., Kalindjian, S. B. & Philpott, K. L. Principles of early drug discovery. *Br. J. Pharmacol.* **162,** 1239–1249 (2011).
4. Maynard, J. A., Lindquist, N. C., Sutherland, J. N., Lesuffleur, A., Warrington, A. E., Rodriguez, M. & Oh, S.-H. Next generation SPR technology of membrane-bound proteins for ligand screening and biomarker discovery. *Biotechnol. J.* **4,** 1542–1558 (2009).
5. Recht, M. I., Nienaber, V. & Torres, F. E. Fragment-Based Screening for Enzyme Inhibitors Using Calorimetry. *Methods Enzymol.* **567,** 47–69 (2016).
6. Thorne, N., Auld, D. S. & Inglese, J. Apparent Activity in HTS: Origins of Compound-Dependent Assay Interference. *Curr. Opin. Chem. Biol.* **14,** 315–324 (2010).
7. Du, X., Li, Y., Xia, Y. L., Ai, S. M., Liang, J., Sang, P., Ji, X. L. & Liu, S. Q. Insights into protein–ligand interactions: Mechanisms, models, and methods. *Int. J. Mol. Sci.* **17,** 1–34 (2016).
8. Lipinski, C. A. Overview of Hit to Lead: The Medicinal Chemist's Role from HTS Retest to Lead Optimization Hand Off. in *Lead-Seeking Approaches. Topics in Medicinal Chemistry* 1–24 (2009).
9. Lin, J., Sahakian, D. C., de Morais, S. M. F., Xu, J. J., Polzer, R. J. & Winter, S. M. The role of absorption, distribution, metabolism, excretion and toxicity in drug discovery. *Curr. Top. Med. Chem.* **3,** 1125–54 (2003).
10. Hefti, F. F. Requirements for a lead compound to become a clinical candidate. *BMC Neurosci.* **9,** 1–7 (2008).
11. Umscheid, C., Margolis, D. & Grossman, C. Key Concepts of Clinical Trials: A Narrative Review. *Postgr. Med* **123,** 194–204 (2012).
12. Dowden, H. & Munro, J. Trends in clinical success rates and therapeutic focus. *Nat. Rev. Drug Discov.* **18,** 495–496 (2019).
13. Westbrook, J. D. & Burley, S. K. How Structural Biologists and the Protein Data Bank Contributed to Recent FDA New Drug Approvals. *Structure* **27,** 211–217 (2018).
14. Pereira, D. A. & Williams, J. A. Review: Historical Perspectives in Pharmacology. Origin and evolution of high throughput screening. *Br. J. Pharmacol.* **152,** 53–61 (2007).
15. 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., Schopfer, U. & Sittampalam, S. G. Impact of high-throughput screening in biomedical research. *Nat. Rev. Drug Discov.* **10,** 188–195 (2011).
16. MacArthur, R. D. & Novak, R. M. Maraviroc: The First of a New Class of Antiretroviral Agents. *Clin. Infect. Dis.* **47,** 236–241 (2008).
17. Temesgen, Z. & Feinberg, J. Tipranavir: A New Option for the Treatment of Drug-Resistant HIV Infection. *Clin. Infect. Dis.* **45,** 761–769 (2007).
18. Pham, D. Q., Nogid, A. & Plakogiannis, R. Sitagliptin: A novel agent for the management of type 2 diabetes mellitus. *Am. J. Heal. Pharm.* **65,** 521–531 (2008).
19. Paul, S. M., Mytelka, D. S., Dunwiddie, C. T., Persinger, C. C., Munos, B. H., Lindborg, S. R. & Schacht, A. L. How to impove R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discov.* **9,** 203–214 (2010).
20. Bakken, G. A., Bell, A. S., Boehm, M., Everett, J. R., Gonzales, R., Hepworth, D., Klug-mcleod, J. L., Lanfear, J., Loesel, J., Mathias, J. & Wood, T. P.

Shaping a Screening File for Maximal Lead Discovery Efficiency and Effectiveness: Elimination of Molecular Redundancy. *J. Chem. Inf. Model.* **52,** 2937–2949 (2012).

21. Mcgovern, S. L., Caselli, E., Grigorieff, N. & Shoichet, B. K. A Common Mechanism Underlying Promiscuous Inhibitors from Virtual and High-Throughput Screening. *J. Med. Chem.* **45,** 1712–1722 (2002).

22. Irwin, J. J., Duan, D., Torosyan, H., Doak, A. K., Ziebart, K. T., Sterling, T., Tumanian, G. & Shoichet, B. K. An Aggregation Advisor for Ligand Discovery. *J. Med. Chem.* **58,** 7076–7087 (2015).

23. Crisman, T. J., Jenkins, J. L., Parker, C. N., Hill, W. A. G., Bender, A., Deng, Z., Nettles, J. H., Davies, J. W. & Glick, M. "Plate Cherry Picking": A Novel Semi-Sequential Screening Paradigm for Cheaper, Faster, Information-Rich Compound Selection. *J. Biomol. Screen.* **12,** 320–327 (2007).

24. Zhu, H., Klug-mcleod, J. & Bakken, G. A. A Heuristic Algorithm for Plate Selection That Maximizes Compound Diversity. *Croat. Chem. Acta* **86,** 435–441 (2013).

25. Nissink, J. W. M., Schmitt, S., Blackburn, S. & Peters, S. Stratified High-Throughput Screening Sets Enable Flexible Screening Strategies from a Single Plated Collection. *J. Biomol. Screen.* **19,** 369–378 (2014).

26. Price, A. J., Howard, S. & Cons, B. D. Fragment-based drug discovery and its application to challenging drug targets. *Essays Biochem.* **61,** 475–484 (2017).

27. Jhoti, H., Cleasby, A., Verdonk, M. & Williams, G. Fragment-based screening using X-ray crystallography and NMR spectroscopy. *Curr. Opin. Chem. Biol.* **11,** 485–493 (2007).

28. Lamoree, B. & Hubbard, R. E. Current perspectives in fragment-based lead discovery (FBLD). *Essays Biochem.* **61,** 453–464 (2017).

29. Collins, P. M., Douangamath, A., Talon, R., Dias, A., Brandao-Neto, J., Krojer, T. & von Delft, F. *Achieving a Good Crystal System for Crystallographic X-Ray Fragment Screening. Methods in Enzymology* **610,** (2018).

30. Pearce, N. M., Krojer, T., Bradley, A. R., Collins, P., Nowak, R. P., Talon, R., Marsden, B. D., Kelm, S., Shi, J., Deane, C. M. & Von Delft, F. A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density. *Nat. Commun.* **8,** 24–29 (2017).

31. Velvadapu, V., Farmer, B. T. & Reitz, A. B. Fragment-Based Drug Discovery. in *The Practice of Medicinal Chemistry (Fourth edition)* 161–180 (2015). doi:10.1016/B978-0-12-417205-0.00007-9

32. Bollag, G., Tsai, J., Zhang, J., Zhang, C., Ibrahim, P., Nolop, K. & Hirth, P. Vemurafenib: the first drug approved for BRAF-mutant cancer. *Nat. Rev. Drug Discov.* **11,** 873–886 (2012).

33. Souers, A. J., Leverson, J. D., Boghaert, E. R., Ackler, S. L., Catron, N. D., Chen, J., Dayton, B. D., Ding, H., Enschede, S. H., Fairbrother, W. J., Huang, D. C. S., Hymowitz, S. G., Jin, S., Khaw, S. L., Kovar, P. J., Lam, L. T., Lee, J., Maecker, H. L., Marsh, K. C., *et al.* ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. *Nat. Med.* **19,** 202–208 (2013).

34. Rao, V. S. & Srinivas, K. Modern drug discovery process : An in silico approach. *J. Bioinforma. Seq. Anal.* **2,** 89–94 (2011).

35. van Montfort, R. L. M. & Workman, P. Structure-based drug design: aiming for a perfect fit. *Essays Biochem.* **61,** 431–437 (2017).

36. Sterling, T. & Irwin, J. J. ZINC 15 − Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **55,** 2324–2337 (2015).

37. Rognan, D. Proteome-scale docking: myth and reality. *Drug Discov. Today Technol.* **10,** e403–e409 (2013).

38. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E.,

Francis, P. & Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **47,** 1739–1749 (2004).

39.   Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. & Taylor, R. D. Improved Protein–Ligand Docking Using GOLD. *Proteins* **52,** 609–623 (2003).

40.   Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. & Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **30,** 2785–2791 (2010).

41.   Irwin, J. J. & Shoichet, B. K. Docking Screens for Novel Ligands Conferring New Biology. *J. Med. Chem.* **59,** 4103–4120 (2016).

42.   Gills, J. J., LoPiccolo, J., Tsurutani, J., Shoemaker, R. H., Best, C. J. M., Abu-Asab, M. S., Borojerdi, J., Warfel, N. A., Gardner, E. R., Danish, M., Hollander, M. C., Kawabata, S., Tsokos, M., Figg, W. D., Steeg, P. S. & Dennis, P. A. Nelfinavir, a lead HIV protease inhibitor, is a broad-spectrum, anticancer agent that induces endoplasmic reticulum stress, autophagy, and apoptosis in vitro and invivo. *Clin. Cancer Res.* **13,** 5183–5194 (2007).

43.   Noble, S. & Goa, K. L. Amprenavir: A Review of its Clinical Potential in Patients with HIV Infection. *Drugs* **60,** 1383–1410 (2000).

44.   Mangum, E. M. & Graham, K. K. Lopinavir-Ritonavir: A New Protease Inhibitor. *Pharmacotherapy* **21,** 1352–1363 (2001).

45.   Suryanarayanan, V., Panwar, U. & Chandra, I. De Novo Design of Ligands Using Computational Methods. in *Computational Drug Discovery and Design* **1762,** 71–86 (2018).

46.   Gillet, V., Johnson, A. P., Mata, P., Sike, S. & Williams, P. SPROUT: A program for structure generation. *J. Comput. Aided. Mol. Des.* **7,** 127–153 (1993).

47.   Böhm, H.-J. The computer program LUDI: A new method for the de novo design of enzyme inhibitors. *J. Comput. Aided. Mol. Des.* **6,** 61–78 (1992).

48.   Agarwal, A. K., Johnson, A. P. & Fishwick, C. W. G. Synthesis of de novo designed small-molecule inhibitors of bacterial RNA polymerase. *Tetrahedron* **64,** 10049–10054 (2008).

49.   Consortium, S. G., Structural, B., Consortium, G., Innovation, F., Structural, I., Structural, N., Consortium, G., Protein, O., Facility, P., Production, P. S. & Initiative, P. Protein production and purification. *Nat. Methods* **5,** 135–147 (2008).

50.   Baradaran, R., Berrisford, J. M., Minhas, G. S. & Sazanov, L. A. Crystal structure of the entire respiratory complex I. *Nature* **494,** 443–448 (2013).

51.   Xia, D., Yu, C., Kim, H., Xia, J., Kachurin, A. M., Zhang, L., Yu, L. & Deisenhofer, J. Crystal Structure of the Cytochrome bc1 Complex from Bovine Heart Mitochondria. *Science (80-.).* **277,** 60–66 (2016).

52.   Menz, R. I., Walker, J. E. & Leslie, A. G. W. Structure of Bovine Mitochondrial F1-ATPase with Nucleotide Bound to All Three Catalytic Sites: Implications for the Mechanism of Rotary Catalysis. *Cell* **106,** 331–341 (2001).

53.   Weber, J. L. Analysis of sequences from the extremely A + T-rich genome of Plasmodium falciparum. *Gene* **52,** 103–109 (1987).

54.   Baca, A. M. & Hol, W. G. J. Overcoming codon bias: A method for high-level overexpression of Plasmodium and other AT-rich parasite genes in Escherichia coli. *Int. J. Parasitol.* **30,** 113–118 (2000).

55.   Lackner, D. H., Carré, A., Guzzardo, P. M., Banning, C., Mangena, R., Henley, T., Oberndorfer, S., Gapp, B. V., Nijman, S. M. B., Brummelkamp, T. R. & Bürckstümmer, T. A generic strategy for CRISPR-Cas9-mediated gene tagging. *Nat. Commun.* **6,** 4–10 (2015).

56.   Ho, C. M., Beck, J. R., Lai, M., Cui, Y., Goldberg, D. E., Egea, P. F. & Zhou, Z. H. Malaria parasite translocon structure and mechanism of effector export. *Nature* **561,** 70–75 (2018).

57.   Ferreira, L. G., Santos, R. N., Oliva, G. & Andricopulo, A. D. Molecular Docking

and Structure-Based Drug Design Strategies. *Molecules* **20,** 13384–13421 (2015).

58. Shi, Y. A Glimpse of Structural Biology through X-Ray Crystallography. *Cell* **159,** 995–1014 (2014).

59. Davis, A. M., Teague, S. J. & Kleywegt, G. J. Application and limitations of x-ray crystallographic data in structure-based ligand and drug design. *Angew. Chemie (International ed.)* **42,** 2718–2736 (2003).

60. Carpenter, E. P., Beis, K., Cameron, A. D. & Iwata, S. Overcoming the challenges of membrane protein crystallography. *Curr. Opin. Struct. Biol.* **18,** 581–586 (2008).

61. Acharya, K. R. & Lloyd, M. D. The advantages and limitations of protein crystal structures. *Trends Pharmacol. Sci.* **26,** 10–14 (2005).

62. Pellecchia, M., Bertini, I., Cowburn, D., Dalvit, C., Giralt, E., Jahnke, W., James, T. L., Homans, S. W., Kessler, H., Luchinat, C., Meyer, B., Oschkinat, H., Peng, J., Schwalbe, H. & Siegal, G. Perspectives on NMR in drug discovery: a technique comes of age. *Nat. Rev. Drug Discov.* **7,** 738–744 (2008).

63. Sugiki, T., Furuita, K., Fujiwara, T. & Kojima, C. Current NMR Techniques for Structure-Based Drug Discovery. *Molecules* **23,** 1–27 (2018).

64. Kleckner, I. R. & Foster, M. P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochim. Biophys. Acta* **1814,** 942–968 (2012).

65. Harner, M. J., Frank, A. O. & Fesik, S. W. Fragment-Based Drug Discovery Using NMR Spectroscopy. *J. Biomol. NMR* **56,** 65–75 (2014).

66. Cala, O., Guillière, F. & Krimm, I. NMR-based analysis of protein-ligand interactions. *Anal. Bioanal. Chem.* **406,** 943–956 (2014).

67. Cheng, Y., Grigorieff, N., Penczek, P. A. & Walz, T. A Primer to Single-Particle Cryo-Electron Microscopy. *Cell* **161,** 438–449 (2015).

68. Scapin, G., Potter, C. S. & Carragher, B. Cryo-EM for Small Molecules Discovery, Design, Understanding, and Application. *Cell Chem. Biol.* **25,** 1–8 (2018).

69. Henderson, R. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q. Rev. Biophys.* **28,** 171–193 (1995).

70. Khoshouei, M., Radjainia, M., Baumeister, W. & Danev, R. Cryo-EM structure of haemoglobin at 3.2 Å determined with the Volta phase plate. *Nat. Commun.* **8,** 1–6 (2017).

71. Fan, X., Wang, J., Zhang, X., Yang, Z., Zhang, J.-C., Zhao, L., Peng, H.-L., Lei, J. & Wang, H.-W. Single particle cryo-EM reconstruction of 52 kDa streptavidin at 3.2 Angstrom resolution. *Nat. Commun.* **10,** 1–11 (2019).

72. Kuehlbrandt, W. The Resolution Revolution. *Science (80-.).* **343,** 1443–1444 (2014).

73. Rawson, S., Mcphillie, M. J., Johnson, R. M., Fishwick, C. W. G. & Muench, S. P. The potential use of single-particle electron microscopy as a tool for structure-based inhibitor design research papers. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **D73,** 1–7 (2017).

74. Renaud, J. P., Chari, A., Ciferri, C., Liu, W. T., Rémigy, H. W., Stark, H. & Wiesmann, C. Cryo-EM in drug discovery: Achievements, limitations and prospects. *Nat. Rev. Drug Discov.* **17,** 471–492 (2018).

75. Dubochet, J. & McDowall, A. W. Vitrification of Pure Water for Electron Microscopy. *J. Microsc.* **124,** 3–4 (1981).

76. Dubochet, J., Adrian, M., Chang, J.-J., Lepault, J. & McDowall, A. W. Cryoelectron Microscopy of Vitrified Specimens. in *Cryotechniques in Biological Electron Microscopy* (eds. Steinbrecht, R. A. & Zierold, K.) 114–131 (Springer Berlin Heidelberg, 1987). doi:10.1007/978-3-642-72815-0_5

77. Bai, X., McMullan, G. & Scheres, S. How cryo-EM is revolutionizing structural

biology. *Trends Biochem. Sci.* **40,** 49–57 (2015).

78. Zhang, X., Jin, L., Fang, Q., Hui, W. H. & Zhou, H. Z. 3.3 Å Cryo-EM Structure of a Nonenveloped Virus Reveals a Priming Mechanism for Cell Entry. *Cell* **141,** 472–483 (2010).

79. Rawson, S., Davies, S., Lippiat, J. D. & Muench, S. P. The changing landscape of membrane protein structural biology through developments in electron microscopy. *Mol. Membr. Biol.* **33,** 12–22 (2016).

80. Fischer, N., Neumann, P., Konevega, A. L., Bock, L. V, Ficner, R., Rodnina, M. V & Stark, H. Structure of the E. coli ribosome-EF-Tu complex at <3 Å resolution by Cs-corrected cryo-EM. *Nature* **520,** 567–570 (2015).

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

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

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

84. Liang, Y. L., Khoshouei, M., Radjainia, M., Zhang, Y., Glukhova, A., Tarrasch, J., Thal, D. M., Furness, S. G. B., Christopoulos, G., Coudrat, T., Danev, R., Baumeister, W., Miller, L. J., Christopoulos, A., Kobilka, B. K., Wootten, D., Skiniotis, G. & Sexton, P. M. Phase-plate cryo-EM structure of a class B GPCR-G-protein complex. *Nature* **546,** 118–123 (2017).

85. García-nafría, J., Nehmé, R., Edwards, P. C. & Tate, C. G. Cryo-EM structure of the serotonin 5-HT1B receptor coupled to heterotrimeric Go. *Nature* **558,** 620–623 (2018).

86. Wu, M. & Lander, G. C. High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM. *Nat. Commun.* **10,** 1–9 (2019).

87. Murphy, B. J., Klusch, N., Langer, J., Mills, D. J., Yildiz, Ö. & Kühlbrandt, W. Rotary substates of mitochondrial ATP synthase reveal the basis of flexible F1-Fo coupling. *Science (80-.).* **364,** (2019).

88. Conley, M. J., McElwee, M., Azmi, L., Gabrielsen, M., Byron, O., Goodfellow, I. G. & Bhella, D. Calicivirus VP2 forms a portal-like assembly following receptor engagement. *Nature* **565,** 377–381 (2019).

89. Baggen, J., Hurdiss, D. L., Zocher, G., Mistry, N., Roberts, R. W., Slager, J. J., Guo, H., van Vliet, A. L. W., Wahedi, M., Benschop, K., Duizer, E., de Haan, C. A. M., de Vries, E., Casasnovas, J. M., de Groot, R. J., Arnberg, N., Stehle, T., Ranson, N. A., Thibaut, H. J., *et al.* Role of enhanced receptor engagement in the evolution of a pandemic acute hemorrhagic conjunctivitis virus. *Proc. Natl. Acad. Sci.* **115,** 397–402 (2018).

90. Hesketh, E. L., Saunders, K., Fisher, C., Potze, J., Stanley, J., Lomonossoff, G. P. & Ranson, N. A. The 3.3 Å structure of a plant geminivirus using cryo-EM. *Nat. Commun.* **9,** 2369 (2018).

91. Liu, Y.-T., Jih, J., Dai, X., Bi, G.-Q. & Zhou, Z. H. Cryo-EM structures of herpes simplex virus type 1 portal vertex and packaged genome. *Nature* **570,** 257–261 (2019).

92. Wong, W., Bai, X., Brown, A., Fernandez, I. S., Hanssen, E., Condron, M., Tan, Y. H., Baum, J. & Scheres, S. H. W. Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. *Elife* **3,** 1–20 (2014).

93. Kaledhonkar, S., Fu, Z., Caban, K., Li, W., Chen, B., Sun, M., Gonzalez, R. L. & Frank, J. Late steps in bacterial translation initiation visualized using time-resolved cryo-EM. *Nature* **570,** 400–404 (2019).

94. Zhang, J., Ma, J., Liu, D., Qin, S., Sun, S., Zhao, J. & Sui, S. F. Structure of phycobilisome from the red alga Griffithsia pacifica. *Nature* **551,** 57–63 (2017).

95. Fitzpatrick, A. W. P., Falcon, B., He, S., Murzin, A. G., Murshudov, G.,

Garringer, H. J., Crowther, R. A., Ghetti, B., Goedert, M. & Scheres, S. H. W. Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **547,** 185–190 (2017).

96. Falcon, B., Zivanov, J., Zhang, W., Murzin, A. G., Garringer, H. J., Vidal, R., Crowther, R. A., Newell, K. L., Ghetti, B., Goedert, M. & Scheres, S. H. W. Novel tau filament fold in chronic traumatic encephalopathy encloses hydrophobic molecules. *Nature* **568,** 420–423 (2019).

97. Li, Y., Zhao, C., Luo, F., Liu, Z., Gui, X., Luo, Z., Zhang, X., Li, D., Liu, C. & Li, X. Amyloid fibril structure of α-synuclein determined by cryo-electron microscopy. *Cell Res.* **28,** 897–903 (2018).

98. Li, B., Ge, P., Murray, K. A., Sheth, P., Zhang, M., Nair, G., Sawaya, M. R., Shin, W. S., Boyer, D. R., Ye, S., Eisenberg, D. S., Zhou, Z. H. & Jiang, L. Cryo-EM of full-length α-synuclein reveals fibril polymorphs with a common structural kernel. *Nat. Commun.* **9,** 3609 (2018).

99. Yin, H. & Flynn, A. D. Drugging Membrane Protein Interactions. *Annu. Rev. Biomed. Eng.* **18,** 51–76 (2016).

100. Caffrey, M. Membrane protein crystallization. *J. Struct. Biol.* **142,** 108–132 (2003).

101. Hite, R. K., Yuan, P., Li, Z., Hsuing, Y., Walz, T. & MacKinnon, R. Cryo-electronmicroscopy structure of the Slo2.2 Na1-activated K1 channel. *Nature* **527,** 198–203 (2015).

102. Hite, R. K. & Mackinnon, R. Structural Titration of Slo2.2, a Na+-Dependent K+ Channel. *Cell* **168,** 390–399 (2017).

103. Clairfeuille, T., Cloake, A., Infield, D. T., Llongueras, J. P., Arthur, C. P., Li, Z. R., Jian, Y., Martin-Eauclaire, M. F., Bougis, P. E., Ciferri, C., Ahern, C. A., Bosmans, F., Hackos, D. H., Rohou, A. & Payandeh, J. Structural basis of a-scorpion toxin action on Nav channels. *Science (80-.).* **363,** 1–14 (2019).

104. Matthies, D., Bae, C., Toombes, G. E., Fox, T., Bartesaghi, A., Subramaniam, S. & Swartz, K. J. Single-particle cryo-EM structure of a voltage-activated potassium channel in lipid nanodiscs. *Elife* **7,** 1–18 (2018).

105. Manolaridis, I., Jackson, S. M., Taylor, N. M. I., Kowal, J., Stahlberg, H. & Locher, K. P. Cryo-EM structures of a human ABCG2 mutant trapped in ATP-bound and substrate-bound states. *Nature* **563,** 426–430 (2018).

106. Johnson, Z. L. & Chen, J. ATP Binding Enables Substrate Release from Multidrug Resistance Protein 1. *Cell* **172,** 81-89.e10 (2018).

107. Yan, R., Zhao, X., Lei, J. & Zhou, Q. Structure of the human LAT1–4F2hc heteromeric amino acid transporter complex. *Nature* **568,** 127–130 (2019).

108. Bayburt, T. H., Sligar, S. G. & Biology, C. Membrane Protein Assembly into Nanodiscs. *FEBS Lett.* **584,** 1721–1727 (2016).

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

110. Parmar, M., Rawson, S., Scarff, C. A., Goldman, A., Dafforn, T. R., Muench, S. P. & Postis, V. L. G. Using a SMALP platform to determine a sub-nm single particle cryo-EM membrane protein structure. *Biochim. Biophys. Acta - Biomembr.* **1860,** 378–383 (2018).

111. Sun, C., Benlekbir, S., Venkatakrishnan, P., Wang, Y., Hong, S., Hosler, J., Tajkhorshid, E., Rubinstein, J. L. & Gennis, R. B. Structure of the alternative complex III in a supercomplex. *Nature* **557,** 123–126 (2018).

112. Qiu, W., Fu, Z., Xu, G. G., Grassucci, R. A., Zhang, Y., Frank, J., Hendrickson, W. A. & Guo, Y. Structure and activity of lipid bilayer within a membrane-protein transporter. *Proc. Natl. Acad. Sci.* **115,** 12985–12990 (2018).

113. Ceska, T., Chung, C., Cooke, R., Phillips, C. & Williams, P. A. Cryo-EM in drug discovery. *Biochem. Soc. Trans.* **47,** 281–293 (2019).

114. 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. & Subramaniam, S. Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. *Cell* **165,** 1698–1707 (2016).

115. Bartesaghi, A., Merk, A., Banerjee, S., Matthies, D., Wu, X., Milne, J. L. S. & Subramaniam, S. 2.2 Å resolution cryo-EM structure of β-galactosidase in complex with a cell-permeant inhibitor. *Science (80-.).* **348,** 1147–1151 (2015).

116. Rawson, S., Bisson, C., Hurdiss, D. L., Fazal, A., McPhillie, M. J., Sedelnikova, S. E., Baker, P. J., Rice, D. W. & Muench, S. P. Elucidating the structural basis for differing enzyme inhibitor potency by cryo-EM. *Proc. Natl. Acad. Sci.* **115,** 1795–1800 (2018).

117. García-Nafría, J., Lee, Y., Bai, X., Carpenter, B. & Tate, C. G. Cryo-EM structure of the adenosine A2A receptor coupled to an engineered heterotrimeric G protein. *Elife* **7,** e35946 (2018).

118. Crofts, A. R. The cytochrome bc1 complex: function in the context of structure. *Annu. Rev. Physiol.* **66,** 689–733 (2004).

119. Althoff, T., Mills, D. J., Popot, J. & Ku, W. Arrangement of electron transport chain components in bovine mitochondrial. *EMBO J.* **30,** 4652–4664 (2011).

120. Mitchell, P. Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature* **191,** 144–148 (1961).

121. Sazanov, L. A. A giant molecular proton pump: structure and mechanism of respiratory complex I. *Nat. Rev. Mol. Cell Biol.* **16,** 375–388 (2015).

122. Zhu, J., Vinothkumar, K. R. & Hirst, J. Structure of mammalian respiratory complex I. *Nature* **536,** 354–358 (2017).

123. Blaza, J. N., Vinothkumar, K. R. & Hirst, J. Structure of the Deactive State of Mammalian Respiratory Complex I. *Structure* **26,** 312-319.e3 (2018).

124. Parey, K., Brandt, U., Xie, H., Mills, D. J., Siegmund, K., Vonck, J. & Ku, W. Cryo-EM structure of respiratory complex I at work. *Elife* **7,** e39213 (2018).

125. Agip, A. A., Blaza, J. N., Bridges, H. R., Viscomi, C., Rawson, S., Muench, S. P. & Hirst, J. Cryo-EM structures of complex I from mouse heart mitochondria in two biochemically defined states. *Nat. Struct. Mol. Biol.* **25,** 548–556 (2018).

126. Dudkina, N. V, Kou, R., Peters, K., Braun, H. & Boekema, E. J. Structure and function of mitochondrial supercomplexes. *Biochim. Biophys. Acta - Bioenerg.* **1797,** 664–670 (2010).

127. Wu, M., Gu, J., Guo, R., Huang, Y. & Yang, M. Structure of Mammalian Respiratory Supercomplex I1III2IV. *Cell* **167,** 1598–1609 (2016).

128. Letts, J. A., Fiedorczuk, K. & Sazanov, L. A. The architecture of respiratory supercomplexes. *Nature* **537,** 644–648 (2016).

129. Guo, R., Zong, S., Wu, M., Gu, J. & Yang, M. Architecture of Human Mitochondrial Respiratory Megacomplex I2III2IV2. *Cell* **170,** 1247-1257.e12 (2017).

130. Wiseman, B., Nitharwal, R. G., Fedotovskaya, O., Schäfer, J., Guo, H., Kuang, Q., Benlekbir, S., Sjöstrand, D., Ädelroth, P., Rubinstein, J. L., Brzezinski, P. & Högbom, M. Structure of a functional obligate complex III2IV2 respiratory supercomplex from Mycobacterium smegmatis. *Nat. Struct. Mol. Biol.* **25,** 1128–1136 (2018).

131. Sousa, J. S., Calisto, F., Langer, J. D., Mills, D. J., Refojo, P. N., Teixeira, M., Kühlbrandt, W., Vonck, J. & Pereira, M. M. Structural basis for energy transduction by respiratory alternative complex III. *Nat. Commun.* **9,** 1–10 (2018).

132. Xia, D., Yu, C. a, Kim, H., Xia, J. Z., Kachurin, a M., Zhang, L., Yu, L. & Deisenhofer, J. Crystal structure of the cytochrome bc1 complex from bovine heart mitochondria. *Science (80-.).* **277,** 60–66 (1997).

133. Hao, G. F., Wang, F., Li, H., Zhu, X. L., Yang, W. C., Huang, L. S., Wu, J. W., Berry, E. A. & Yang, G. F. Computational discovery of picomolar Qo site

inhibitors of cytochrome bc 1 complex. *J. Am. Chem. Soc.* **134,** 11168–11176 (2012).

134. Solmaz, S. R. N. & Hunte, C. Structure of complex III with bound cytochrome C in reduced state and definition of a minimal core interface for electron transfer. *J. Biol. Chem.* **283,** 17542–17549 (2008).

135. Huang, L. S., Cobessi, D., Tung, E. Y. & Berry, E. A. Binding of the respiratory chain inhibitor antimycin to the mitochondrial bc1 complex: A new crystal structure reveals an altered intramolecular hydrogen-bonding pattern. *J. Mol. Biol.* **351,** 573–597 (2005).

136. Esser, L., Elberry, M., Zhou, F., Yu, C. A., Yu, L. & Xia, D. Inhibitor-complexed structures of the cytochrome bc1 from the photosynthetic bacterium Rhodobacter sphaeroides. *J. Biol. Chem.* **283,** 2846–2857 (2008).

137. Hunte, C., Koepke, J., Lange, C., Roßmanith, T. & Michel, H. Structure at 2.3 Å resolution of the cytochrome bc1 complex from the yeast Saccharomyces cerevisiae co-crystallized with an antibody Fv fragment. *Structure* **8,** 669–684 (2000).

138. Berry, E. A., Huang, L.-S., Saechao, L. K., Pon, N. G., Valkova-Valchanova, M. & Daldal, F. X-Ray structure of *Rhodobacter capsulatus* cytochrome bc_1: comparison with its mitochondrial and chloroplast counterparts. *Photosynth. Res.* **81,** 251–275 (2004).

139. Yang, X. & Trumpower, B. L. Purification of a three-subunit ubiquinol-cytochrome c oxidoreductase complex from Paracoccus denitrificans. *J. Biol. Chem.* **261,** 12282–12289 (1986).

140. Schägger, H., Link, T. A., Engel, W. D. & von Jagow, G. Isolation of the eleven protein subunits of the bc1 complex from beef heart. *Methods Enzymol.* **126,** 224–37 (1986).

141. Kessl, J. J., Lange, B. B., Merbitz-zahradnik, T., Zwicker, K., Hill, P., Meunier, B., Pa, H., Hunte, C., Meshnick, S. & Trumpower, B. L. Molecular Basis for Atovaquone Binding to the Cytochrome bc1 Complex. *J. Biol. Chem.* **278,** 31312–31318 (2003).

142. Brandt, U. & Trumpower, B. The protonmotive Q cycle in mitochondria and bacteria. *Crit. Rev. Biochem. Mol. Biol.* **29,** 165–197 (1994).

143. Crofts, A. R., Lhee, S., Crofts, S. B., Cheng, J. & Rose, S. Proton pumping in the bc1 complex: A new gating mechanism that prevents short circuits. *Biochim. Biophys. Acta - Bioenerg.* **1757,** 1019–1034 (2006).

144. Fisher, N. & Meunier, B. Re-examination of inhibitor resistance conferred by Qo-site mutations in cytochrome b using yeast as a model system. *Pest Manag. Sci.* **61,** 973–978 (2005).

145. Xia, D., Esser, L., Tang, W.-K., Zhou, F., Zhou, Y., Yu, L. & Yu, C.-A. Structural Analysis of Cytochrome bc1 Complexes: Implications to the Mechanism of Function. *Biochim. Biophys. Acta - Biomembr.* **1827,** 1278–1294 (2014).

146. Ellis, J. E. Coenzyme Q homologs in parasitic protozoa as targets for chemotherapeutic attack. *Parasitol. Today* **10,** 296–301 (1994).

147. Nixon, G. L., Pidathala, C., Shone, A. E., Antoine, T., Fisher, N., O'Neill, P. M., Ward, S. & Biagini, G. Targeting the mitochondrial electron transport chain of Plasmodium falciparum: new strategies towards the development of improved antimalarials for the elimination era. *Future Med. Chem.* **5,** 1573–91 (2013).

148. Doggett, J. S., Nilsen, A., Forquer, I., Wegmann, K. W., Jones-Brando, L., Yolken, R. H., Bordón, C., Charman, S. A., Katneni, K., Schultz, T., Burrows, J. N., Hinrichs, D. J., Meunier, B., Carruthers, V. B. & Riscoe, M. K. Endochin-like quinolones are highly efficacious against acute and latent experimental toxoplasmosis. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 15936–41 (2012).

149. McPhillie, M., Zhou, Y., El Bissati, K., Dubey, J., Lorenzi, H., Capper, M., Lukens, A. K., Hickman, M., Muench, S., Verma, S. K., Weber, C. R., Wheeler, K., Gordon, J., Sanders, J., Moulton, H., Wang, K., Kim, T.-K., He, Y., Santos,

T., *et al.* New paradigms for understanding and step changes in treating active and chronic, persistent apicomplexan infections. *Sci. Rep.* **6,** 29179 (2016).

150. Painter, H. J., Morrisey, J. M., Mather, M. W. & Vaidya, A. B. Specific role of mitochondrial electron transport in blood-stage Plasmodium falciparum. *Nature* **446,** 88–91 (2007).

151. Capper, M. J., O'Neill, P. M., Fisher, N., Strange, R. W., Moss, D., Ward, S. A., Berry, N. G., Lawrenson, A. S., Hasnain, S. S., Biagini, G. A. & Svetlana V. Antonyuk. Antimalarial 4(1H)-pyridones bind to the Qi site of cytochrome bc1. *PNAS* **112,** 755–760 (2015).

152. Birth, D., Kao, W.-C. & Hunte, C. Structural analysis of atovaquone-inhibited cytochrome bc1 complex reveals the molecular basis of antimalarial drug action. *Nat. Commun.* **5,** 4029 (2014).

153. Srivastava, I. K. & Vaidya, A. B. A mechanism for the synergistic antimalarial action of atovaquone and proguanil. *Antimicrob. Agents Chemother.* **43,** 1334–1339 (1999).

154. Jiménez-Díaz, M. B., Mulet, T., Viera, S., Gómez, V., Garuti, H., Ibáñez, J., Alvarez-Doval, A., Shultz, L. D., Martínez, A., Gargallo-Viola, D. & Angulo-Barturen, I. Improved murine model of malaria using Plasmodium falciparum competent strains and non-myelodepleted NOD-scid IL2Rγnull mice engrafted with human erythrocytes. *Antimicrob. Agents Chemother.* **53,** 4533–4536 (2009).

155. Muench, S. P., Rawson, S., Eyraud, V., Delmas, A. F., 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. *J. Biol. Chem.* **289,** 16399–16408 (2014).

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

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

158. Toei, M., Saum, R. & Forgac, M. Regulation and isoform function of the V-ATPases. *Biochemistry* **49,** 4715–4723 (2010).

159. Beyenbach, K. W. & Wieczorek, H. The V-type H+ATPase: molecular structure and function, physiological roles and regulation. *J. Exp. Biol.* **209,** 577–589 (2006).

160. Zhao, J., Benlekbir, S. & Rubinstein, J. L. Electron cryomicroscopy observation of rotational states in a eukaryotic V-ATPase. *Nature* **521,** 241–246 (2015).

161. Oot, R. A., Couoh-Cardel, S., Sharma, S., Stam, N. J. & Wilkens, S. Breaking up and making up: The secret life of the vacuolar H+ -ATPase. *Protein Sci.* **26,** 896–909 (2017).

162. Mazhab-jafari, M. T., Rohou, A., Schmidt, C., Bueler, S. A., Benlekbir, S., Robinson, C. V & Rubinstein, J. L. Atomic model for the membrane-embedded VO motor of a eukaryotic V-ATPase. *Nat. Publ. Gr.* **539,** 118–122 (2016).

163. Roh, S.-H., Stam, N. J., Hryc, C. F., Couoh-Cardel, S., Pintilie, G., Chiu, W. & Wilkens, S. The 3.5-Å CryoEM Structure of Nanodisc-Reconstituted Yeast Vacuolar ATPase Vo Proton Channel. *Mol. Cell* **69,** 993–1004 (2018).

164. Vasanthakumar, T., Bueler, S. A., Wu, D., Beilsten-Edmands, V., Robinson, C. V & Rubinstein, J. L. Structural comparison of the vacuolar and Golgi V-ATPases from Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci.* **116,** 7272–7277 (2019).

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

166. Bowman, E. J., Graham, L. A., Stevens, T. H. & Bowman, B. J. The bafilomycin/concanamycin binding site in subunit c of the V-ATPases from Neurospora crassa and Saccharomyces cerevisiae. *J. Biol. Chem.* **279,** 33131–33138 (2004).

167. Sasse, F., Steinmetz, H., HÖfle, G. & Reichenbach, H. Archazolids, New Cytotoxic Macrolactones from Archangium gephyra (Myxobacteria) Production, Isolation, Physico-chemical and Biological Properties. *J. Antibiot. (Tokyo).* **56,** 520–525 (2003).

168. Bockelmann, S., Menche, D., Rudolph, S., Bender, T., Grond, S., Zezschwitz, P. Von, Muench, S. P., Wieczorek, H., Huss, M. & Universita, E. K. Archazolid A Binds to the Equatorial Region of the c-Ring of the Vacuolar H+ -ATPase. *J. Biol. Chem.* **285,** 38304–38314 (2010).

169. Boyd, M. R., Farina, C., Belfiore, P., Gagliardi, S., Kim, J. W., Hayakawa, Y., Beutler, J. a, McKee, T. C., Bowman, B. J. & Bowman, E. J. Discovery of a novel antitumor benzolactone enamide class that selectively inhibits mammalian vacuolar-type (H+)-atpases. *J. Pharmacol. Exp. Ther.* **297,** 114–20 (2001).

170. Páli, T., Whyteside, G., Dixon, N., Kee, T. P., Ball, S., Harrison, M. A., Findlay, J. B. C., Finbow, M. E. & Marsh, D. Interaction of inhibitors of the vacuolar H+-ATPase with the transmembrane V0-sector. *Biochemistry* **43,** 12297–12305 (2004).

171. Mizutani, K., Yamamoto, M., Suzuki, K., Yamato, I., Kakinuma, Y., Shirouzu, M., Walker, J. E., Yokoyama, S., Iwata, S. & Murata, T. Structure of the rotor ring modified with N,N'-dicyclohexylcarbodiimide of the Na+-transporting vacuolar ATPase. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 13474–13479 (2011).

172. Ohkuma, S., Shimizu, S., Noto, M., Sai, Y., Kuninori Kinoshita & Hiro-Omi Tamura. Inhibition of Cell Growth by Bafilomycin A1 , a Selective Inhibitor of Vacuolar H + -ATPase. *Vitr. cell Dev. Biol.* **29,** 862–866 (1993).

173. Huss, M., Ingenhorst, G., König, S., Gaßel, M., Dröse, S., Zeeck, A., Altendorf, K. & Wieczorek, H. Concanamycin A, the specific inhibitor of V-ATPases, binds to the VO subunit c. *J. Biol. Chem.* **277,** 40544–40548 (2002).

174. Huss, M., Sasse, F., Kunze, B., Jansen, R., Steinmetz, H., Ingenhorst, G., Zeeck, A. & Wieczorek, H. Archazolid and apicularen: Novel specific V-ATPase inhibitors. *BMC Biochem.* **6,** 1–10 (2005).

175. Sundquist, K. Characterization of ATP-dependent proton transport in medullary bone-derived microsomes. *Bone Miner.* 17–29 (1993).

176. Sgro, G. G. & Costa, T. R. D. Cryo-EM Grid Preparation of Membrane Protein Samples for Single Particle Analysis. *Front. Mol. Biosci.* **5,** 1–8 (2018).

177. Passmore, L. A. & Russo, C. J. Specimen preparation for high-resolution cryo-EM. *Methods Enzymol.* **579,** 51–86 (2016).

178. Glaeser, R. M. Methods for imaging weak-phase objects in electron microscopy. *Rev. Sci. Instrum.* **84,** 1–17 (2013).

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

180. Ciferri, C., Chandramouli, S., Leitner, A., Donnarumma, D., Cianfrocco, M. A., Gerrein, R., Friedrich, K., Aggarwal, Y., Palladino, G., Aebersold, R., Norais, N., Settembre, E. C. & Carfi, A. Antigenic Characterization of the HCMV gH/gL/gO and Pentamer Cell Entry Complexes Reveals Binding Sites for Potently Neutralizing Human Antibodies. *PLoS Pathog.* **11,** 1–20 (2015).

181. Milne, J. L. S., Borgnia, M. J., Bartesaghi, A., Tran, E. E. H., Earl, L. A., Schauder, D. M., Lengyel, J., Pierson, J., Patwardhan, A. & Subramaniam, S. Cryo-electron microscopy – a primer for the non-microscopist. *FEBS J.* **280,** 28–45 (2013).

182. Drulyte, I., Johnson, R. M., Hesketh, E. L., Hurdiss, D. L., Scarff, C. A., Porav, S. A., Ranson, N. A., Muench, S. P. & Thompson, R. F. Approaches to altering particle distributions in cryo-electron microscopy sample preparation. *Acta Crystallogr. Sect. D Struct. Biol.* **74,** 1–12 (2018).

183. Lyumkis, D. Challenges and opportunities in cryo-EM single-particle analysis.

J. Biol. Chem. **294,** 5181–5197 (2019).

184. Murata, K. & Wolf, M. Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. *Biochim. Biophys. Acta - Gen. Subj.* **1862,** 324–334 (2018).

185. Noble, A. J., Dandey, V. P., Wei, H., Brasch, J., Chase, J., Acharya, P., Tan, Y. Z., Zhang, Z., Kim, L. Y., Scapin, G., Rapp, M., Eng, E. T., Rice, W. J., Cheng, A., Negro, C. J., Shapiro, L., Kwong, P. D., Jeruzalmi, D., Georges, A. des, *et al.* Routine single particle CryoEM sample and grid characterization by tomography. *Elife* **7,** e34257 (2018).

186. Noble, A. J., Wei, H., Dandey, V. P., Zhang, Z., Tan, Y. Z., Potter, C. S. & Carragher, B. Reducing effects of particle adsorption to the air–water interface in cryo-EM. *Nat. Methods* **15,** 793–795 (2018).

187. Glaeser, R. M. Proteins, Interfaces, and Cryo-EM Grids. *Curr. Opin. Colloid Interface Sci.* **34,** 1–8 (2018).

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

189. Milne, J. L. S., Borgnia, M. J., Bartesaghi, A., Tran, E. E. H., Earl, L. A., Schauder, D. M., Lengyel, J., Pierson, J., Patwardhan, A. & Subramaniam, S. Cryo-electron microscopy: A primer for the non-microscopist. *Febs J.* **280,** 28–45 (2013).

190. Glaeser, R. M. Specimen Behavior in the Electron Beam. in *Methods in Enzymology* **579,** 19–50 (Elsevier Inc., 2016).

191. Cheng, Y., Grigorieff, N., Penczek, P. A. & Walz, T. A Primer to Single-Particle Cryo-Electron Microscopy. *Cell* **161,** 438–449 (2015).

192. Sigworth, F. J. Principles of cryo-EM single-particle image processing. *Reprod. Syst. Sex. Disord.* **65,** 57–67 (2016).

193. Danev, R. & Nagayama, K. *Phase Plates for Transmission Electron Microscopy. Methods in Enzymology* **481,** (Elsevier Masson SAS, 2010).

194. Danev, R. & Baumeister, W. Cryo-EM single particle analysis with the Volta phase plate. *Elife* **5,** e13046 (2016).

195. Danev, R., Tegunov, D. & Baumeister, W. Using the volta phase plate with defocus for cryo-em single particle analysis. *Elife* **6,** e23006 (2017).

196. Bammes, B. E., Rochat, R. H., Jakana, J., Chen, D. & Chiu, W. Direct electron detection yields cryo-EM reconstructions at resolutions beyond 3 / 4 Nyquist frequency. *J. Struct. Biol.* **177,** 589–601 (2012).

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

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

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

200. Ruskin, R. S., Yu, Z. & Grigorieff, N. Quantitative characterization of electron detectors for transmission electron microscopy. *J. Struct. Biol.* **184,** 385–393 (2013).

201. McMullan, G., Turchetta, R. & Faruqi, A. R. Single event imaging for electron microscopy using MAPS detectors. *J. Instrum.* **6,** C04001 (2011).

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

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

204. Carragher, B., Kisseberth, N., Kriegman, D., Milligan, R. A., Potter, C. S., Pulokas, J. & Reilein, A. Leginon: An automated system for acquisition of

images from vitreous ice specimens. *J. Struct. Biol.* **132,** 33–45 (2000).

205. Thompson, R. F., Iadanza, M. G., Hesketh, E. L., Rawson, S. & Ranson, N. A. Collection, pre-processing and on-the-fly analysis of data for high-resolution, single-particle cryo-electron microscopy. *Nat. Protoc.* **14,** 100–118 (2019).

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

207. Zheng, S. Q., Palovcak, E., Armache, J.-P., Verba, K. A., Cheng, Y. & Agard, D. A. MotionCor2 : anisotropic correction of beam-induced motion for improved cryo-electron microscopy Automatic tracing of ultra-volumes of neuronal images. *Nat. Methods* **14,** 331–332 (2017).

208. Zivanov, J., Nakane, T. & Scheres, S. H. W. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6,** 5–17 (2019).

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

210. Zhang, K. Gctf : Real-time CTF determination and correction. *J. Struct. Biol.* **193,** 1–12 (2016).

211. Scheres, S. H. W., Núñez-Ramírez, R., Sorzano, C. O. S., Carazo, J. M. & Marabini, R. Image processing for electron microscopy single-particle analysis using XMIPP. *Nat. Protoc.* **3,** 977–990 (2008).

212. Jeong, H.-S., Park, H.-N., Kim, J.-G. & Hyun, J.-K. Critical importance of the correction of contrast transfer function for transmission electron microscopy-mediated structural biology. *J. Anal. Sci. Technol.* **4,** 1–7 (2013).

213. Henderson, R. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proc. Natl. Acad. Sci.* **110,** 18037–18041 (2013).

214. Zivanov, J., Kimanius, D., Nakane, T., Forsberg, B. O., Hagen, W. J., Lindahl, E. & Scheres, S. H. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7,** e42166 (2018).

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

216. Wagner, T., Merino, F., Stabrin, M., Moriya, T., Antoni, C., Apelbaum, A., Hagel, P., Sitsel, O., Raisch, T., Prumbaum, D., Quentin, D., Roderer, D., Tacke, S., Siebolds, B., Schubert, E., Shaikh, T. R., Lill, P., Gatsogiannis, C. & Raunser, S. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2,** 1–13 (2019).

217. Guo, F. & Jiang, W. Single Particle Cryo-electron Microscopy and 3-D Reconstruction of Viruses. *Methods Mol. Biol.* **1117,** 401–443 (2014).

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

219. Scheres, S. H. W. Maximum-likelihood methods in cryo-EM. PartII: application to experimantal data. *Methods Enzym.* **6879,** 295–320 (2010).

220. Scheres, S. H. W. *Processing of Structurally Heterogeneous Cryo-EM Data in RELION. Methods in Enzymology* **579,** (Elsevier Inc., 2016).

221. Nogales, E. & Scheres, S. H. W. Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity. *Mol. Cell* **58,** 677–689 (2015).

222. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14,** 290–296 (2017).

223. Henderson, R., Chen, S., Chen, J. Z., Grigorieff, N., Passmore, L. A., Ciccarelli, L., Rubinstein, J. L., Crowther, R. A., Stewart, P. L. & Rosenthal, P. B. Tilt-pair analysis of images from a range of different specimens in single-particle electron cryomicroscopy. *J. Mol. Biol.* **413,** 1028–1046 (2011).

224. Russo, C. J. & Passmore, L. A. Robust evaluation of 3D electron cryomicroscopy data using tilt-pairs. *J. Struct. Biol.* **187,** 112–118 (2014).
225. Scheres, S. H. W. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* **9,** 853–854 (2012).
226. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333,** 721–745 (2003).
227. Henderson, R., Sali, A., Baker, M. L., Carragher, B., Devkota, B., Downing, K. H., Egelman, E. H., Feng, Z., Frank, J., Grigorieff, N., Jiang, W., Ludtke, S. J., Medalia, O., Penczek, P. A., Rosenthal, P. B., Rossmann, M. G., Schmid, M. F., Schröder, G. F., Steven, A. C., *et al.* Outcome of the first electron microscopy validation task force meeting. *Structure* **20,** 205–214 (2012).
228. Fernández, J. J., Luque, D., Castón, J. R. & Carrascosa, J. L. Sharpening high resolution information in single particle electron cryomicroscopy. *J. Struct. Biol.* **164,** 170–175 (2008).
229. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60,** 2126–2132 (2004).
230. Adams, P. D., Pavel, V., Chen, V. B., Ian, W., Echols, N., Moriarty, N. W., Read, R. J., Richardson, D. C., Jane, S. & Thomas, C. PHENIX : a comprehensive Python-based system for macromolecular structure solution research papers. *Acta Crystallogr. Sect. D* **66,** 213–221 (2010).
231. Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67,** 355–367 (2011).
232. Trabuco, Leonardo G., Villa, Elizabeth., Schreiner, Eduard., Harrison, Christopher B., Schulten, K. Molecular Dynamics Flexible Fitting: A practical guide to combine cryo-electron microscopy and X-ray crystallography. *Methods* **49,** 174–180 (2010).
233. Croll, T. I. ISOLDE: A physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. Sect. D Struct. Biol.* **74,** 519–530 (2018).
234. Scheres, S. H. W. RELION : Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180,** 519–530 (2012).
235. MotionCor2 : anisotropic correction of beam-induced motion for improved cryo-electron microscopy Automatic tracing of ultra-volumes of neuronal images. *Nat. Publ. Gr.* **14,** 331–332 (2017).
236. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. UCSF Chimera — A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).
237. Uchida, E., Ohsumi, Y. & Anraku, Y. Purification of Yeast Vacuolar Membrane H+-ATPase and Enzymological Discrimination of Three ATP-Driven Proton Pumps in Saccharomyces cerevisiae. *Methods Enzymol.* **157,** 544–562 (1988).
238. Chifflet, S., Torriglia, A., Chiesa, R. & Tolosa, S. A method for the determination of inorganic phosphate in the presence of labile organic phosphate and high concentrations of protein: Application to lens ATPases. *Anal. Biochem.* **168,** 1–4 (1988).
239. WHO. *World Malaria Report 2018.* (2018).
240. Baggish, A. L. & Hill, D. R. Antiparasitic Agent Atovaquone. *Antimicrob. Agents Chemother.* **46,** 1163–1173 (2002).
241. Heel, M. Van, Harauz, G. & Orlova, E. V. A New Generation of the IMAGIC Image Processing System. *J. Struct. Biol.* **116,** 17–24 (1996).
242. Charoensutthivarakul, S., David Hong, W., Leung, S. C., Gibbons, P. D., Bedingfield, P. T. P., Nixon, G. L., Lawrenson, A. S., Berry, N. G., Ward, S. A.,

Biagini, G. A. & O'Neill, P. M. 2-Pyridylquinoloneantimalarials with improved antimalarial activity and physicochemical properties. *Med. Chem. Commun.* **6,** 1252–1259 (2015).

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

244. Kaufmann, K. W., Lemmon, G. H., Deluca, S. L., Sheehan, J. H. & Meiler, J. Practically Useful : What the ROSETTA Protein Modeling Suite Can Do for You. *Biochemistry* **49,** 2987–2998 (2010).

245. Iwata, S. Complete Structure of the 11-Subunit Bovine Mitochondrial Cytochrome bc1 Complex. *Science (80-.).* **281,** 64–71 (1998).

246. Gao, X., Wen, X., Esser, L., Quinn, B., Yu, L., Yu, C. & Xia, D. Structural Basis for the Quinone Reduction in the bc1 Complex: A Comparative Analysis of Crystal Structures of Mitochondrial Cytochrome bc1 with Bound Substrate and Inhibitors at the Qi Site. *Biochemistry* **42,** 9067–9080 (2003).

247. Esser, L., Quinn, B., Li, Y. F., Zhang, M., Elberry, M., Yu, L., Yu, C. A. & Xia, D. Crystallographic studies of quinol oxidation site inhibitors: A modified classification of inhibitors for the cytochrome bc1 complex. *J. Mol. Biol.* **341,** 281–302 (2004).

248. Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S. H. Characterisation of molecular motions in cryo-EM single-particle data by multi- body refinement in RELION. *Elife* **7,** e36861 (2018).

249. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-tal, N. ConSurf 2010 : calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **38,** 529–533 (2010).

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

251. Hawkins, P. C. D., Skillman, A. G. & Nicholls, A. Comparison of Shape-Matching and Docking as Virtual Screening Tools. *J. Med. Chem.* **50,** 74–82 (2007).

252. Arai, S., Saijo, S., Suzuki, K., Mizutani, K., Kakinuma, Y., Ishizuka-katsura, Y., Ohsawa, N., Terada, T., Shirouzu, M., Yokoyama, S., Iwata, S., Yamato, I. & Murata, T. Rotation mechanismof Enterococcus hirae V1-ATPase based on asymmetric crystal structures Satoshi. *Nature* **493,** 703–707 (2013).

253. Baell, J. & Walters, M. A. Chemistry: Chemical con artists foil drug discovery. *Nature* **513,** 481–483 (2014).

254. Bagley, M. C., Dwyer, J. E., Molina, M. D. B., Rand, A. W., Rand, H. L. & Tomkinson, N. C. O. Microwave-assisted synthesis of 3-aminobenzo[b]thiophene scaffolds for the preparation of kinase inhibitors. *Org. Biomol. Chem.* **13,** 6814–6824 (2015).

255. Scherr, N., Honnappa, S., Kunz, G., Mueller, P., Jayachandran, R., Winkler, F., Pieters, J. & Steinmetz, M. O. Structural basis for the specific inhibition of protein kinase G, a virulence factor of Mycobacterium tuberculosis. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 12151–6 (2007).

256. Janeček, M., Rossmann, M., Sharma, P., Emery, A., Huggins, D. J., Stockwell, S. R., Stokes, J. E., Tan, Y. S., Almeida, E. G., Hardwick, B., Narvaez, A. J., Hyvönen, M., Spring, D. R., McKenzie, G. J. & Venkitaraman, A. R. Allosteric modulation of AURKA kinase activity by a small-molecule inhibitor of its protein-protein interaction with TPX2. *Sci. Rep.* **6,** 1–12 (2016).

257. Gewald, K. Zur Reaktion von a-0x0-mercaptanen:mit Nitrilen. *Angew. Chemie (International ed.)* **73,** 114 (1961).

258. Puterová, Z., Krutošíková, A. & Végh, D. Gewald reaction: synthesis, properties and applications of substituted 2- aminothiophenes. *ARKIVOC* **(i),** 209–246 (2010).

259. The Dang, H., Chorell, E., Uvell, H., Pinkner, J. S., Hultgren, S. J. & Almqvist,

F. Syntheses and biological evaluation of 2-amino-3-acyl-tetrahydrobenzothiophene derivatives; Antibacterial agents with antivirulence activity. *Org. Biomol. Chem.* **12,** 1942–1956 (2014).

260. 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 Reveals New Mechanistic Insights. *Structure* **23,** 461–471 (2015).

261. Arinaminpathy, Y., Khurana, E., Engelman, D. M. & Gerstein, M. B. Computational analysis of membrane proteins : the largest class of drug targets. *Drug Discov. Today* **14,** 1130–1135 (2009).