STUDY OF THE YEAST VACUOLAR-TYPE ATPASE BY ELECTRON CRYOMICROSCOPY

by

Jianhua Zhao

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Medical Biophysics
University of Toronto

© Copyright 2016 by Jianhua Zhao
Abstract

Study of the yeast vacuolar-type ATPase by electron cryomicroscopy

Jianhua Zhao
Doctor of Philosophy
Graduate Department of Medical Biophysics
University of Toronto
2016

Vacuolar-type ATPases (V-ATPases) are rotary enzymes that use energy from hydrolysis of adenosine triphosphate (ATP) to pump protons across membranes and control the pH of many intracellular compartments. ATP hydrolysis in the soluble catalytic region of the enzyme is coupled to proton translocation through the membrane-bound region by rotation of a rotor subcomplex. Studies of ATP synthases, V-ATPases, and vacuolar/archaeal ATPases (V/A-ATPases) have suggested that flexibility is necessary for the catalytic mechanism of rotary ATPases but the structures of different rotational states have never been observed experimentally. V-ATPases also serve an important role in the immune system and are targeted by intracellular pathogens. *Legionella pneumophila*, the causative agent of Legionnaires’ Disease, secretes a protein effector called SidK that binds and inhibits V-ATPases, preventing acidification of lysosomes and promoting survival of the bacterium inside macrophages. The mechanism of V-ATPase inhibition by SidK is not known. The structure of V-ATPases can be analyzed by single particle electron cryomicroscopy (cryo-EM), an imaging technique involving the identification of tens to hundreds of thousands of individual proteins in micrographs. A partially-automated image selection program was developed to increase particle image identification throughput. A method was also developed to correct for magnification anisotropy that was identi-
fied in our images. Cryo-EM maps were obtained for structures of three rotational states of the V-ATPase from the yeast *Saccharomyces cerevisiae*. The three different maps reveal the conformational changes that occur to couple rotation in the soluble catalytic region to the symmetry-mismatched membrane-bound proton-translocating region. The structures of these states provide direct evidence that deformation during rotation enables the smooth transmission of power through rotary ATPases. To investigate the molecular mechanism of V-ATPase inhibition by SidK, we determined structures of the V-ATPase:SidK complex by cryo-EM. The cryo-EM maps reveal SidK to be an elongated protein composed of a structured region that interacts with the soluble catalytic region of V-ATPase and a flexible region of unknown function. The structures suggest a novel inhibition site for V-ATPases and provides evidence that the activity of these proton pumps is linked to the flexibility of the A-subunit.
Contents

1 Introduction

1.1 Rotary ATPases ....................................................... 1
  1.1.1 Three types are F-, V-, and V/A-ATPases ................. 1
  1.1.2 V-ATPases are the most complicated rotary ATPases ...... 4
  1.1.3 Rotor rotation couples ATP hydrolysis to proton transport . 8
  1.1.4 V-ATPases are regulated by mechanisms that are unclear ... 11

1.2 Electron cryomicroscopy ........................................... 13
  1.2.1 Purified samples are vitrified ................................ 14
  1.2.2 Low-exposure images are acquired to minimize specimen damage 16
  1.2.3 Images are sorted and combined into 3D maps ............. 19
  1.2.4 Cryo-EM reveals the structure of intact rotary ATPases ... 25

2 Template matching and classification system (TMaCS) .......... 29
  2.1 Author contributions ............................................. 29
  2.2 Acknowledgements ................................................. 30
  2.3 Introduction ....................................................... 30
  2.4 Methods ............................................................ 35
    2.4.1 Data and programming languages ......................... 35
    2.4.2 Pretreatment of images .................................... 35
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4.3</td>
<td>Template matching</td>
<td>36</td>
</tr>
<tr>
<td>2.4.4</td>
<td>Classification by machine learning</td>
<td>37</td>
</tr>
<tr>
<td>2.4.5</td>
<td>Ranking of images by projection matching</td>
<td>38</td>
</tr>
<tr>
<td>2.4.6</td>
<td>Computational performance</td>
<td>38</td>
</tr>
<tr>
<td>2.5</td>
<td>Results</td>
<td>38</td>
</tr>
<tr>
<td>2.5.1</td>
<td>Overview and definitions</td>
<td>38</td>
</tr>
<tr>
<td>2.5.2</td>
<td>Template matching</td>
<td>39</td>
</tr>
<tr>
<td>2.5.3</td>
<td>Machine learning with a support vector machine</td>
<td>44</td>
</tr>
<tr>
<td>2.5.4</td>
<td>Ranking of particle images by projection matching</td>
<td>50</td>
</tr>
<tr>
<td>2.6</td>
<td>Discussion</td>
<td>51</td>
</tr>
<tr>
<td>3</td>
<td>Magnification Anisotropy</td>
<td>55</td>
</tr>
<tr>
<td>3.1</td>
<td>Author contributions</td>
<td>55</td>
</tr>
<tr>
<td>3.2</td>
<td>Acknowledgements</td>
<td>56</td>
</tr>
<tr>
<td>3.3</td>
<td>Introduction</td>
<td>56</td>
</tr>
<tr>
<td>3.4</td>
<td>Methods and Results</td>
<td>60</td>
</tr>
<tr>
<td>3.5</td>
<td>Discussion</td>
<td>67</td>
</tr>
<tr>
<td>3.6</td>
<td>Software availability</td>
<td>68</td>
</tr>
<tr>
<td>4</td>
<td>Rotational states of the V-ATPase</td>
<td>69</td>
</tr>
<tr>
<td>4.1</td>
<td>Author contributions</td>
<td>69</td>
</tr>
<tr>
<td>4.2</td>
<td>Acknowledgements</td>
<td>69</td>
</tr>
<tr>
<td>4.3</td>
<td>Author information</td>
<td>70</td>
</tr>
<tr>
<td>4.4</td>
<td>Introduction</td>
<td>70</td>
</tr>
<tr>
<td>4.5</td>
<td>Methods</td>
<td>71</td>
</tr>
<tr>
<td>4.5.1</td>
<td>Yeast membrane purification</td>
<td>71</td>
</tr>
<tr>
<td>4.5.2</td>
<td>Purification of V-ATPase</td>
<td>72</td>
</tr>
</tbody>
</table>
5.5.2 SidK contains a putative α-helical domain and a flexible domain 105
5.5.3 SidK reduces the flexibility of the A-subunit in the ‘open’ con-
formation .......................................................... 114
5.5.4 SidK binds A-subunit in ‘tight’ conformation preferentially . . 114
5.5.5 SidK decreases the affinity of the V₁ subcomplex for ATP and
stabilizes V₁ ......................................................... 116
5.6 Discussion ....................................................... 120

6 Future Directions 122
6.1 V-ATPase mechanism ............................................ 122
   6.1.1 V-ATPase structure during ATP hydrolysis ............... 122
   6.1.2 Proton translocation ...................................... 123
   6.1.3 V₀ inhibition ............................................. 124
6.2 SidK mechanism .................................................. 124
   6.2.1 Atomic structure of SidK ................................ 124
   6.2.2 Cryo-EM of V-ATPase:SidK ............................ 125
   6.2.3 Function of the flexible region of SidK ................. 126
6.3 Cryo-EM methods ................................................ 126
   6.3.1 Magnification anisotropy ................................ 126
   6.3.2 Image classification and refinement .................... 127

A Crystallization of SidK 128
A.1 Acknowledgments .............................................. 128
A.2 Crystallization methods ...................................... 128
A.3 In situ proteolysis of SidK .................................. 129
A.4 Identification of elastase cleavage sites in SidK .......... 129
A.5 Crystallization of truncated SidK ........................... 132
A.6 Techniques to improve crystal quality ........................................ 137
A.7 Metal-mediated synthetic symmetrization .................................... 139
A.8 Mutation of cysteine residues .................................................. 140

Bibliography ................................................................................. 144
List of Tables

A.1 Purification and crystallization of SidK truncation constructs. 135
List of Figures

1.1 Localization of rotary ATPases in different organisms and membranes ............................................. 2
1.2 Structure of rotary ATPases ................................................................. 6
1.3 The binding-change mechanism and rotary catalysis ....................... 9
1.4 The two half-channel model of proton translocation .................. 12
1.5 Vitrified cryo-EM sample in thin layer of perforated carbon supported by a metal mesh grid ..................... 15
1.6 Regions of representative micrographs of V-ATPase recorded on the Gatan K2 Summit DDD and Kodak Electron Image Film SO-163 .......................................................... 18
1.7 2D and 3D classification of particle images ................................. 21
1.8 A 2D image is an approximate projection of a 3D density map 23
1.9 Fourier shell correlation (FSC) as a measure of map resolution 27
1.10 Previous cryo-EM map of the eukaryotic V-ATPase at ∼11Å resolution ...................................................... 28

2.1 Flowchart describing the steps in TMaCS ............................... 33
2.2 Pre-processing of micrographs and generation of template images for template matching 40
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3</td>
<td>Template matching search of micrographs</td>
<td>43</td>
</tr>
<tr>
<td>2.4</td>
<td>Graphical user interface (GUI) for training the support vector machine (SVM) classifier</td>
<td>45</td>
</tr>
<tr>
<td>2.5</td>
<td>Classification of candidate particle images from template matching of ten V/A-ATPase micrographs</td>
<td>48</td>
</tr>
<tr>
<td>3.1</td>
<td>Anisotropic magnification in images</td>
<td>59</td>
</tr>
<tr>
<td>3.2</td>
<td>Modeling the contrast transfer function (CTF)</td>
<td>62</td>
</tr>
<tr>
<td>3.3</td>
<td>Effect of anisotropic magnification on CTF parameters</td>
<td>63</td>
</tr>
<tr>
<td>4.1</td>
<td>Data collection of V-ATPase images</td>
<td>76</td>
</tr>
<tr>
<td>4.2</td>
<td>3D maps from rotational states</td>
<td>77</td>
</tr>
<tr>
<td>4.3</td>
<td>Rotational states of the V-ATPase</td>
<td>78</td>
</tr>
<tr>
<td>4.4</td>
<td>Map segmentation and molecular dynamics flexible fitting</td>
<td>80</td>
</tr>
<tr>
<td>4.5</td>
<td>The membrane-bound $V_O$ region</td>
<td>82</td>
</tr>
<tr>
<td>4.6</td>
<td>C-terminal domain of the proton-translocating a-subunit</td>
<td>84</td>
</tr>
<tr>
<td>4.7</td>
<td>Symmetry mismatch between the $V_1$ and $V_O$ regions</td>
<td>87</td>
</tr>
<tr>
<td>4.8</td>
<td>Flexibility of V-ATPase subunits</td>
<td>88</td>
</tr>
<tr>
<td>4.9</td>
<td>Flexibility in catalytic subunits and peripheral stalks</td>
<td>90</td>
</tr>
<tr>
<td>5.1</td>
<td>Protein purification and cryo-EM imaging</td>
<td>101</td>
</tr>
<tr>
<td>5.2</td>
<td>Image processing workflow and 3D classification scheme</td>
<td>103</td>
</tr>
<tr>
<td>5.3</td>
<td>Cryo-EM maps of V-ATPase:SidK</td>
<td>104</td>
</tr>
<tr>
<td>5.4</td>
<td>Local resolution, particle image distribution, and comparison of V-ATPase with and without SidK bound</td>
<td>107</td>
</tr>
<tr>
<td>5.5</td>
<td>Structure of the V-ATPase:SidK complex and SidK</td>
<td>108</td>
</tr>
<tr>
<td>5.6</td>
<td>Secondary structure prediction of SidK from <em>L. pneumophila</em></td>
<td>109</td>
</tr>
</tbody>
</table>
5.7 Multiple sequence alignment between SidK and other *Legionella* effectors ................................................. 112
5.8 SidK reduces the flexibility of the A-subunit in the ‘open’ conformation .......................................................... 113
5.9 SidK binds the A-subunit preferentially in the ‘tight’ conformation ............................................................. 115
5.10 SidK specifically inhibits V-ATPases, decreases the binding affinity of V-ATPase for nucleotide, and stabilizes the V$_1$ subcomplex ................................................................. 119

A.1 Limited proteolysis reveals stable SidK species when incubated with trypsin and elastase ............................... 130
A.2 Crystals from *in situ* proteolysis of SidK ................................................. 131
A.3 Identification of trypsin-digested peptides by mass spectrometry of solubilized SidK crystals ............................ 133
A.4 Crystallization of SidK10-543 ................................................................. 134
A.5 Crystallization of SidK constructs .......................................................... 136
A.6 Crystallization of lysine-methylated SidK10-543 ............. 138
A.7 Cross-linking of SidK proteins ............................................................. 141
A.8 Thermal stability of cysteine-to-serine SidK mutants ........... 142
A.9 Crystals of quadruple cysteine SidK mutant (C32S/C314S/C490S/C541S) 143
List of Movies

Movie 1. Cross-sections through the maps of the V-ATPase in three rotational states.

Movie 2. Interpolation between the three observed rotational states of the V-ATPase.

Movie 3. Exploded view of subunits when interpolating between the three observed rotational states of the V-ATPase.
List of acronyms

ABC, dimethylamine-borane complex
ADP, adenosine diphosphate
ATP, adenosine triphosphate
ATPase, adenosine triphosphatase
BLAST, basic local alignment search tool
CCD, charge-coupled device
CTF, contrast transfer function
cryo-EM, (single particle) electron microscopy
DDD, direct detector device
DTT, dithiothreitol
FSC, Fourier shell correlation
GUI, graphical user interface
IF1, inhibitory factor 1
MALDI-TOF MS, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry
MDFF, Molecular Dynamics Flexible Fitting
MILK, Machine Learning Toolkit
PA1b, pea albumin 1 subunit b
PCA, principle component analysis
PtpA, protein tyrosine phosphatase A
SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis
SNR, signal-to-noise ratio
SVD, singular value decomposition
SVM, support vector machine
TMaCS, Template Matching and Classification System

V-ATPase, vacuolar-type ATPase

V/A-ATPase, vacuolar/archael ATPase
Chapter 1

Introduction

1.1 Rotary ATPases

1.1.1 Three types are F-, V-, and V/A-ATPases

The universal energy currency of organisms is adenosine triphosphate (ATP), a molecule that stores chemical potential energy used to drive biological processes. ATP is produced primarily from harnessing of energy stored in an electrochemical ion gradient, a conversion that is catalyzed by a class of rotary ATPase enzymes called ATP synthases or F-ATPases (Figure 1.1) [23, 199, 187]. ATP synthases can also catalyze the reverse reaction and hydrolyze ATP to generate an electrochemical ion gradient. However, mechanisms exist to block this process in vivo [146, 34, 111], ensuring unidirectional operation of the enzyme. In eukaryotes, the reverse reaction of generating an ion gradient using ATP is catalyzed by another class of rotary ATPases called vacuolar-type ATPases (V-ATPases) (Figure 1.1B) [52]. A third class of rotary ATPases called vacuolar/archaeal ATPases (V/A-ATPases) exist in bacteria and archaea that resemble V-ATPases structurally but serve mainly to synthesize ATP (Figure
Figure 1.1: Localization of rotary ATPases in different organisms and membranes. ATP synthases are found in the plasma membrane of some bacteria and archaea (A), the inner mitochondrial membrane in eukaryotes (B), and the thylakoid membrane in the chloroplast of plants (B). V-ATPases localize to the membranes of numerous cell compartments and the plasma membrane of eukaryotes (B). V/A-ATPases are found in the plasma membranes of some bacteria and archaea (A).
ATP synthases make ATP

ATP synthases produce the majority of ATP required by most organisms for survival and are essential players in energy metabolism. In animals, ATP synthases are found in the inner membranes of mitochondria. Through the process of cellular respiration, the electron transport chain produces a proton gradient across the inner mitochondrial membrane with a higher concentration of protons in the intermembrane space than the mitochondrial matrix [67, 160]. The proton gradient is dissipated by ATP synthases to generate ATP through catalyzing the dehydration reaction between adenosine diphosphate (ADP) and inorganic phosphate (P$_i$) [1, 23]. In plants, ATP synthases are found in the thylakoid membranes of chloroplasts. A proton gradient across the thylakoid membrane is produced through photosynthesis [183, 66], which then drives ATP synthesis by ATP synthase. These rotary ATPases are also found in the cell membranes of some bacteria, where bacterial electron transport chains generate electrochemical gradients to drive ATP synthesis [7]. Some bacterial ATP synthases can also make use of sodium ions, rather than hydrogen ions, as an electrochemical energy source [96, 174].

V-ATPases acidify cellular compartments

Acidic cell compartments in eukaryotes are important for numerous biological functions [114, 52]. Compartments such as lysosomes, endosomes, and the Golgi are acidified by V-ATPases. V-ATPases use energy from ATP hydrolysis to pump protons across lipid membranes and against a concentration gradient, catalyzing the reverse reaction of ATP synthases. The acidic pH is required for many biological processes including protein glycosylation in the Golgi [36, 37], loading of neurotransmitters in
Chapter 1. Introduction

secretory vesicles [206, 85], protein trafficking in endosomes [75, 35, 185], and amino acid sensing in lysosomes [209, 17]. In some specialized cells, V-ATPases localize to the plasma membrane where they facilitate acidification of urine by renal tubule cells [84, 169], bone remodeling in osteoclasts [21, 56, 180], and sperm maturation in the epididymis [140]. Mutations in V-ATPases can result in renal tubule acidosis [84, 169] and osteopetrosis [56]. Furthermore, V-ATPases have been found recently in the plasma membrane of some tumor cells where they promote tumor invasion and metastasis [72, 28]. V-ATPases play crucial roles in diverse cellular and physiological processes, making them intriguing drug targets in many areas of health and disease.

V/A-ATPases serve mainly to make ATP

V/A-ATPases are found in the cell membranes of archaea and some bacteria. These rotary enzymes have a similar structure to V-ATPases from eukaryotes [120], suggesting that V/A-ATPases have a closer evolutionary relationship with V-ATPases than with ATP synthases. Although some V/A-ATPases are known to act as sodium pumps [22, 82], V/A-ATPases function mainly as ATP synthases to generate ATP in vivo [74, 39, 124]. V/A-ATPases contain fewer protein subunits and are smaller enzymes compared to V-ATPases. Due to their structural similarity to V-ATPases and simpler architectures, V/A-ATPases have been used as model enzymes to understand the eukaryotic V-ATPases. The rotary ATPases from Thermus thermophilus [76, 129, 94, 93] and Enterococcus hirae [123, 8] have been studied primarily.

1.1.2 V-ATPases are the most complicated rotary ATPases

Rotary ATPases are macromolecular complexes composed of many different protein subunits (Figure 1.2). These protein assemblies share similar architectures but
differ in subunit composition and structure. Prokaryotic ATP synthases and V/A-ATPases are the simplest rotary ATPases, consisting of the fewest and smallest subunits [199, 120]. On the other hand, eukaryotic V-ATPases are perhaps the most complex rotary ATPases: they have subunits not present in other rotary ATPases, are regulated by reversible dissociation of the enzyme assembly [177, 83], and are glycosylated [115, 88, 150]. The complexity of V-ATPases reflects the diversity of biological processes involving the enzymes, but has also hindered the study of these proton pumps. Work described in this thesis was focused on eukaryotic V-ATPases and the rest of this introduction will center around these enzymes.

**V-ATPases have a catalytic region, rotor, and stator**

Rotary ATPases suffer from a history of confusing nomenclature. For clarity, only the names of subunits from V-ATPases will be used here. V-ATPases have a subunit stoichiometry of $A_3B_3CDE_3FG_3Hac_xc'_yc'_zd e$ where $x$, $y$, and $z$ indicate unknown stoichiometries. Upper and lower case letters denote the soluble $V_1$ and membrane-embedded $V_O$ regions, respectively. The ‘catalytic’ region ($A_3B_3$) is composed of three $AB$ heterodimers arranged in a pseudo-symmetric trimer and contains the active sites of ATP hydrolysis/synthesis. The ‘rotor’ ($DFdc_xc'_yc'_zd e$) couples ATP catalysis to proton transport and contains the membrane-embedded ‘c-ring’ ($c_xc'_yc'_zd e$). The ‘stator’ ($CE_3G_3Hae$) contains the proton-translocating $a$-subunit and three peripheral stalks ($3 \times EG$) that act as a scaffold for the rest of the enzyme complex.

**The subunit composition of the c-ring varies**

The c-ring forms an important part of the proton translocation pathway by carrying protons into the lipid bilayer. It is an oligomer of $c$-subunits, with two and four $\alpha$-helices making up each $c$-subunit in ATP synthases and V-ATPases, respec-
Figure 1.2: **Structure of rotary ATPases.** V-ATPases, V/A-ATPases, and ATP synthases share similar overall architectures that include the catalytic region, rotor, and stator subcomplexes. The V-ATPase is comprised of a soluble $V_1$ region ($A_3B_3CDE_3FG_3H$) and a membrane-embedded $V_O$ region ($c_xc'_yc''_zd$ where $x$, $y$, and $z$ denote unknown stoichiometries).
Chapter 1. Introduction

tively [109, 176]. V/A-ATPases can have c-subunits composed of two or four α-helices [123, 178]. Different rotary ATPases have different c-subunit stoichiometries, which define the number of protons translocated for each ATP hydrolyzed. The c-rings of different rotary ATPases contain between 8 to 15 c-subunits [120, 118, 143], which means ~2.7 to 5 protons are translocated across the membrane per ATP hydrolyzed. Until recently, the number of c-subunits in the c-ring of the eukaryotic V-ATPase was unknown (Chapter 4) [201]. To complicate matters further, the c-ring of the eukaryotic V-ATPase is a hetero-oligomer of different c-, c′-, and c″-subunits. All three of these subunits are required for V-ATPase function [73].

The number of the peripheral stalks differ

The peripheral stalks hold the a-subunit stationary relative to the $V_1$ region, allowing rotation of the rotor against the a-subunit to facilitate proton translocation. The number of peripheral stalks in the stator differs between different rotary ATPases: there are one, two, and three in ATP synthases, V/A-ATPases, and V-ATPases, respectively [87, 119, 94, 155, 14]. The peripheral stalks of V-ATPases and V/A-ATPases are similar [97, 175, 132], but differ from the peripheral stalk of ATP synthases [44, 118]. The functional role of peripheral stalk stoichiometry remains unclear. It has been suggested that the peripheral stalks may play a role in the assembly, disassembly, and regulation of eukaryotic V-ATPases [132]. In support of this idea, cryo-EM structures of the intact eukaryotic complex show the three peripheral stalks in different conformations [119, 20], which indicates structural flexibility that could aid in the assembly and disassembly of the enzyme. Flexibility of the peripheral stalks has also been observed in high-resolution crystal structures of these subcomplexes [175, 132] and by mass spectrometry of a bacterial V/A-ATPase [205].
Rotary ATPases are flexible structurally

Studies of ATP synthases, V-ATPases, and V/A-ATPases have suggested that flexibility is important for the catalytic mechanism of rotary ATPases [128, 175, 205]. Deformation of the peripheral stalks has been proposed to accommodate the conformational changes that occur during catalysis [175, 205]. Furthermore, the rotor has been found to be even more flexible than the peripheral stalks, with the most flexible region located near the middle of the rotor [164, 189]. This flexibility may allow smooth power transfer between the catalytic $V_1$ and the membrane-embedded $V_O$ regions as discussed in Chapter 4.

1.1.3 Rotor rotation couples ATP hydrolysis to proton transport

The binding-change mechanism allows ATP hydrolysis

Many decades of study, primarily on ATP synthases, have unraveled key aspects of the mechanism of rotary ATPases [1, 191]. The conversion between ATP and ADP/P$_i$ is catalyzed through a process called the binding-change mechanism (Figure 1.3). Each AB heterodimer in the catalytic region contains a nucleotide-binding site at the interface between the two subunits. The subunits can adopt different conformations termed ‘tight’, ‘loose’, and ‘open’ with ATP, ADP and P$_i$, and no nucleotide bound, respectively. Binding of ATP to the AB heterodimer in the ‘open’ conformation causes the AB heterodimer to adopt the ‘tight’ conformation. ATP hydrolysis to ADP and P$_i$ then causes the AB heterodimer to adopt the ‘loose’ conformation. In the final step, ADP and P$_i$ are released and the AB heterodimer adopts the ‘open’ conformation. The AB heterodimers are coupled so that conformational changes...
Figure 1.3: **The binding-change mechanism and rotary catalysis.** Cross-sections through the catalytic region are depicted looking from the catalytic region towards the membrane-embedded region. The catalytic region of rotary ATPases is composed of three heterodimers of A- and B-subunits (yellow and red, respectively). Each heterodimer is in a different conformation termed ‘tight’, ‘loose’, and ‘open’ with ATP, ADP and P$_i$, and no nucleotide expected to be bound, respectively. During ATP hydrolysis, binding and release of nucleotide induces conformational changes in the heterodimers that drive rotation of the rotor (blue).
in one AB heterodimer induces conformational changes in neighboring A- and B-subunits, with each of the three AB heterodimers adopting a different conformation. This asymmetric distribution of conformational states in the three AB heterodimers can be induced upon nucleotide binding or when the catalytic region is in complex with the D-subunit of the rotor [8]. As a consequence, the catalytic regions of intact rotary ATPases are pseudo-symmetric.

Rotary catalysis drives proton transport

Conformational changes in the AB heterodimers are coupled to rotation of the D-subunit by a process called rotary catalysis [1, 128, 159, 191] (Figure 1.3). Structural studies indicate that the C-terminal α-helices in the A- and B-subunits ‘push’ on the D-subunit during ATP hydrolysis, and *vice versa* for ATP synthesis [1, 129, 8]. Computational analyses suggest that this push/pull mechanism acts through electrostatic or hydrophobic interactions [145, 121, 122]. The D-subunit rotates with respect to the catalytic region and causes the c-ring to rotate against the a-subunit, driving proton translocation across the membrane. Single molecule studies have demonstrated that one full rotation occurs through three 120° steps [128, 198, 76, 45, 124, 116], corresponding to the three catalytic AB heterodimers. In ATP synthases, each 120° step contains smaller substeps corresponding to binding, catalysis, and release of nucleotide [197, 193]. However, similar experiments with V/A-ATPases have not found smaller substeps [76, 58, 116], indicating differences in the rotary catalysis mechanism.

The two half-channel model describes proton translocation

Ion transport by rotary ATPases is not well understood, but is believed to occur via two half-channels in the a-subunit (Figure 1.4) [79]. A proton enters through one half-channel and moves to the middle of the lipid bilayer where it interacts with
the c-ring. The c-ring surface that contacts the membrane is hydrophobic, with the exception of conserved glutamate or aspartate residues located near the middle of the membrane bilayer. The acidic residue is protonated by the proton from the first half-channel, neutralizing the charge and allowing the c-ring to rotate into the hydrophobic membrane environment. Rotation of the c-ring allows another proton from a different c-subunit to exit via the second half-channel to the other side of the membrane. Consistent with this model, mutation of the conserved acidic residues in the c-ring prevents ion transport [50]. Recent work has provided insight into the structure of the a-subunit and defined possible locations of the proton channels (Chapter 4) [4, 201, 204].

1.1.4 V-ATPases are regulated by mechanisms that are unclear

\textbf{V}_1 \textbf{and} \textbf{V}_O \textbf{undergo reversible dissociation}

Under cellular stress, such as starvation, the eukaryotic V-ATPase undergoes reversible dissociation into the V$_1$ and V$_O$ regions [83, 177]. This process involves the H- and C-subunits [65, 11, 135, 102], which form a collar around the middle of the enzyme assembly between the V$_1$ and V$_O$ regions [119, 194, 63, 200]. Upon disassembly, the dissociated V$_1$ region is auto-inhibited by the H-subunit to prevent unproductive hydrolysis of ATP [135, 102, 77, 20] and ion transport in the V$_O$ region is inhibited [134, 40]. Reversible dissociation is a unique regulatory mechanism in V-ATPases that is not known to occur in ATP synthases or V/A-ATPases \textit{in vivo}, which could explain the increased size and complexity of V-ATPases compared to other rotary ATPases.
Figure 1.4: **The two half-channel model of proton translocation.** The α-subunit (*green*) contains two half-channels (dashed outline) that are offset. Each half-channel accesses a different side of the membrane. A proton (*gray sphere*) enters through one half-channel and neutralizes an acidic residue near the middle of the membrane bilayer. The c-ring (*magenta*) rotates, carrying the proton into the membrane and around to the second half-channel. The proton then exits through the second half-channel to the other side of the membrane.
V-ATPases interact with numerous proteins

V-ATPases interact with numerous proteins to carry out its biological functions: the RAVE complex is involved in V-ATPase assembly [168, 167, 166], ARNO and Arf6 interact with V-ATPases at endosomal membranes to regulate protein trafficking [75], and the Ragulator binds V-ATPase to recruit mTORC1 to lysosomes in amino acid sensing [209, 17]. In the pea plant, pea albumin 1 subunit b (PA1b) is a peptide that inhibits V-ATPases in insects and acts as a natural insecticide [31]. V-ATPases also have an important role in the immune system through the acidification of lysosomes and are targeted by bacterial effectors: SidK is secreted by *Legionella pneumophila* and inhibit V-ATPases in phagocytic white blood cells [196], protein tyrosine phosphatase A (PtpA) is secreted by *Mycobacterium tuberculosis* and binds V-ATPases to prevent phagosome fusion with lysosomes [195], and VopQ from *Vibrio parahaemolyticus* binds V-ATPases to prevent lysosomal acidification and fusion [173]. The mechanisms of binding and inhibition of V-ATPases by protein inhibitors are not well understood. Structural insight into the inhibition of V-ATPase by SidK is provided in Chapter 5.

1.2 Electron cryomicroscopy

Single particle electron cryomicroscopy (cryo-EM) is a structure determination method for the study of biological macromolecules. Cryo-EM involves imaging frozen specimens embedded in a thin layer of vitrified buffer using a transmission electron microscope. The images are combined computationally to calculate a 3D map of the protein complex. Through recent advances in technology and methodology, cryo-EM has become a powerful technique for the study of a wide variety of macromolecular complexes [170, 90, 27].
1.2.1 Purified samples are vitrified

Samples must be pure

The optimal specimen for high resolution cryo-EM is pure, meaning that it contains a single type of macromolecule. A pure sample is necessary because it is difficult to distinguish between different macromolecules of similar size in cryo-EM images. Furthermore, structure determination requires averaging of thousands of images of a macromolecule in the same conformation. Therefore, macromolecules that are flexible or that exist in many different conformations will resist structure determination to high resolution. To reduce structural heterogeneity, high-affinity ligands have been used to lock macromolecular complexes in specific conformations [100, 136]. This procedure reduces the complexity of the computational analysis required for structure determination, resulting in higher-resolution maps. If high-affinity ligands are not available, heterogeneous data can also be separated into homogeneous datasets by computational methods (section 1.2.3).

The specimen is embedded in a thin layer of vitreous ice

Many biological macromolecules function in an aqueous environment, making it important to study these macromolecules in a hydrated state. Imaging by electron microscopy exposes samples to high vacuum and dehydrates liquid samples, limiting early studies to dehydrated and stained specimens (negative-stain electron microscopy) [24]. Furthermore, imaging occurs usually over the order of seconds, necessitating a stable sample that does not move. These challenges were overcome by the vitrification of water containing biological macromolecules [48, 3]. Decades later, little has changed in the sample vitrification process. First, a few µl of sample is placed on a small metal grid coated with a layer of perforated carbon (Figure 1.5) [110].
Figure 1.5: Vitrified cryo-EM sample in thin layer of perforated carbon supported by a metal mesh grid. (A) Copper mesh grid (*gold*) acts as structural support for a thin layer of perforated carbon. Scale bar, 1 mm. (B) A perforated carbon film (*gray*) acts as structural support for vitrified sample (*blue*). Scale bar, 500 nm. (c) Macromolecular complexes oriented randomly in a thin film of vitrified buffer. Scale bars, 20 nm.
Excess sample is removed by blotting with filter paper. The grid is then plunged into a bath of liquid ethane or an ethane/propane mixture, freezing the sample in vitreous water/buffer. Samples are then stored in liquid nitrogen until imaging. Because image contrast is generated between the macromolecule and the buffer, minimal buffer surrounding the macromolecule (i.e. a thin sample) is crucial to optimize image contrast. Different macromolecules and buffers have different chemical properties and will behave differently during sample preparation. As a consequence, conditions for the preparation of optimal cryo-EM samples must be determined empirically.

1.2.2 Low-exposure images are acquired to minimize specimen damage

Cryo-EM uses transmission electron microscopes that were designed originally for material sciences but have been adapted for imaging frozen biological samples. The high-energy electrons used for imaging have small wavelengths, making the physical resolution limit well beyond that of atoms. In practice however, imperfections in the magnetic lenses and the damaging effects of the electrons on the sample limit the resolutions attainable by cryo-EM.

Images have low signal-to-noise ratios (SNRs)

Samples are maintained at near liquid nitrogen temperature in the microscope, minimizing sample dehydration and specimen movement. However, the high-energy electrons used in imaging induce specimen movement by breaking chemical bonds, damaging biological macromolecules in a process termed radiation damage [59, 15, 61]. To minimize radiation damage, cryo-EM micrographs are taken with a low exposure of electrons. As a consequence, the resulting images suffer from a low signal-to-noise
ratio (SNR), which is defined as

$$SNR = \frac{P_{signal}}{P_{noise}} = \frac{\theta_{signal}^2}{\theta_{noise}^2}$$  \hspace{1cm} (1.1)$$

where $P_{signal}$ is the power of the signal, $P_{noise}$ is the power of the noise, $\theta_{signal}^2$ is the variance in the signal (zero mean), and $\theta_{noise}^2$ is the variance in the noise (zero mean).

Typical samples imaged by cryo-EM scatter electrons weakly and do not absorb the incident electrons, resulting in low $P_{signal}$ and poor image contrast. To increase the image SNR, micrographs are taken out of focus so that the electron waves interfere at the imaging plane to increase $P_{signal}$. The resulting images have artifacts that must be removed by correction for the contrast transfer function (CTF) (section 1.2.3).

**Detectors have different SNRs**

Numerous imaging modalities are available for cryo-EM. Emulsion films are an analog detection method offering a good SNR at the cost of low throughput. On the other hand, electronic detectors such as charged couple devices (CCDs) convert the signal from electrons into light signals and offer higher throughput at the expense of lower SNR. The best imaging devices available currently are direct detector devices (DDDs) that record electrons directly [99]. DDDs offer higher SNRs compared to film while allowing high throughput (Figure 1.6). Some DDDs have high frame-rates and allow the recording of movies rather than single images. This camera feature enables the correction of specimen movement that occurs during movie acquisition by aligning and then averaging individual frames of a movie [99, 154], further increasing image quality.
Figure 1.6: Regions of representative micrographs of V-ATPase recorded on the Gatan K2 Summit DDD and Kodak Electron Image Film SO-163. Increasing the amount of underfocus (defocus) increases image SNR as shown by the K2 images with defocus values of 1, 2.5, and 4.5 µm. Comparison between K2 and film images with defocus 4.5 µm shows the superior SNR of DDD images compared to film. Example particle images of V-ATPase are indicated by the red boxes. Scale bars, 50 nm.
1.2.3 Images are sorted and combined into 3D maps

The CTF is estimated to correct image artifacts

Micrographs are taken underfocus to maximize the SNR, which induces undesirable artifacts in the micrographs that must be corrected. An underfocused image can be represented as a convolution between the in-focus image and the point-spread function of the microscope. The convolution theorem provides a convenient way to handle and correct out-of-focus image artifacts:

\[ \text{FT}(f * g) = \text{FT}(f) \cdot \text{FT}(g) \]  

where \( \text{FT} \) is the Fourier transform, \( f \) is the in-focus image, \( g \) is the point spread function, and \( f * g \) is the convolution between \( f \) and \( g \). The Fourier transform of the point-spread function (\( \text{FT}(g) \)) is called the contrast transfer function (CTF). The CTF is discussed in more detail in Chapter 3. The CTF is estimated from the power spectrum of out-of-focus images [117]. To correct image artifacts, one simple solution would be to divide \( \text{FT}(f) \cdot \text{FT}(g) \) by the CTF, \( \text{FT}(g) \). However, this solution is problematic when the image is corrupted by noise, creating large noise artifacts when the CTF is near zero. Furthermore, no information about the object being imaged is available at the specific frequencies when the CTF is zero. As a solution, images are taken at different defocus settings and are combined towards the end of image processing where CTF correction takes the form of a Wiener filter [137]. Analysis of the CTF parameters can also be useful for diagnosing problems encountered during image processing such as magnification anisotropy (Chapter 3).

Protein particles are identified from micrographs

Cryo-EM micrographs contain numerous individual macromolecules (termed ‘particles’) embedded in vitrified buffer. The particles must be identified and separated
from other image features in a process called particle image selection. The low SNR of micrographs makes the particle image selection task difficult for automated programs. As a consequence, the task was performed previously by hand in many cases. However, the low SNR in images necessitates the collection of tens or hundreds of thousands of particle images to achieve higher resolution, making manual selection approaches impractical. To overcome this bottleneck, a semi-automated particle image selection program was developed to increase data throughput (Chapter 2).

2D classification separates true and false particle images

Due to the low SNR in micrographs, particle image selection results in a mixture of desirable and undesirable images. Undesirable images include those of protein contamination, ice crystals, noise, and other non-specimen features that would lower the quality of cryo-EM maps. To remove undesirable images, 2D classification is performed to group and average the particle images into representative 2D classes (Figure 1.7). Algorithms to perform 2D classification include multivariate statistical analysis [54] and maximum likelihood methods [162, 161]. The 2D classes have a higher SNR than individual images and allow identification of desired macromolecules with higher fidelity. Particle images from desired 2D classes are selected for further processing.

3D classification separates structural heterogeneity

Macromolecular complexes can exist in equilibria between different assembly and conformational states. To identify the different structures that may be present in a dataset, 3D classification can be performed on the particle images (Figure 1.7) using maximum likelihood algorithms [162, 161, 107]. This procedure can separate images based on how well they represent specific 3D classes and allows the isolation
Figure 1.7: **2D and 3D classification of particle images.** Images are first sorted into 2D classes and desirable classes are identified. Green and red outlines indicate desirable and undesirable classes, respectively. Particle images that constitute desirable 2D classes are then sorted into 3D classes. Particle images from desirable 3D classes are selected for further analysis. Scale bars, 25 Å.
of more homogeneous datasets for structure determination to higher resolution. The development of DDDs for image acquisition has enhanced the performance of 3D classification programs, making the software an essential component of the cryo-EM toolbox.

**Particle images are combined into 3D maps**

Particle images are approximate 2D projections of macromolecules orientated randomly in vitrified buffer and are related to a 3D Coulomb potential density $P(x, y, z)$ by

$$I(x, y) \approx \int T(x, y, \phi, \theta, \psi) P(x, y, z) dz$$  \hspace{1cm} (1.3)

where $I(x, y)$ is the projected image and $T(x, y, \phi, \theta, \psi)$ is a transformation describing the orientation of the macromolecule (Figure 1.8). The relative orientations $(x, y, \phi, \theta, \psi)$ of the macromolecules are unknown and must be determined computationally. The orientation parameters can be determined by comparison of an image with different projections of a 3D density map using cross-correlation methods [64, 20] or maximum likelihood algorithms [162, 161]. A starting model is required and can be an existing density map of a similar macromolecule (as was done in work described in this thesis) or a model generated by more involved methods [42, 184, 149, 98, 13]. Once the orientation parameters are determined, 2D images can be combined into a 3D map using the Fourier slice theorem [43]. Thousands to hundreds of thousands of particle images must be analyzed because of their low SNR, resulting in redundancy in the 3D Fourier space coverage. Averaging $N$ particle images with the same orientation parameters decreases the noise by $\sqrt{N}$ and increases the SNR by $N$ in equation 1.1. The iterative process of calculating image orientation parameters and building density maps is called map refinement. Maps that progress to high resolu-
Figure 1.8: A 2D image is an approximate projection of a 3D density map. The relative orientation of each particle image is defined by two translation parameters (x,y) and three Euler angles (φ,θ,ψ). The orientation parameters describe a transformation of a 3D density map. An integration along the projection axis corresponds to the 2D image (right). A 3D density map is constructed by combining the particle images based on their relative orientations using the Fourier slice theorem [43]. During iterative refinement, a map from one round of refinement is used in the next round to improve the accuracy of the image orientation parameters. Scale bars, 25 Å.
tion (∼3 Å or better) can be used to build atomic models in software developed for X-ray crystallography. Otherwise, if available, existing high-resolution structures of individual components could be fitted (or ‘docked’) into lower-resolution maps of a larger assembly to gain functional insight into the macromolecular complex. It is often found that the structures of individual components change upon interaction with other components of an assembly. A popular method to dock high-resolution structures into cryo-EM maps involves flexible fitting of atomic models using molecular dynamics [182].

**FSC is a measure of map resolution**

The resolution of a cryo-EM map is a measure of the map’s SNR. How this quantity is calculated is still a matter of some debate, but one method called the Fourier shell correlation (FSC) is accepted generally. This method involves dividing the particle images into two random datasets and each dataset is refined independently to generate two density maps. The density maps are transformed into Fourier space and the correlation in Fourier components between the two transforms is calculated as a function of frequency. The frequency-dependent correlation in Fourier components between two maps is the FSC [53] and is given by

\[
FSC(k) = \frac{\sum_{i}^{n} F_1(k_i) \cdot F_2(k_i)^*}{\sqrt{\left( \sum_{i}^{n} |F_1(k_i)|^2 \right) \left( \sum_{i}^{n} |F_2(k_i)|^2 \right)}}
\]

(1.4)

where \( k \) is the spatial frequency, \( F_1(k) \) and \( F_2(k) \) are Fourier components from the Fourier transforms of two density maps, and \( n \) is the number of Fourier components in a particular frequency range (Figure 1.9). It can be shown using equation 1.1 that
the FSC is related to the SNR by the equation [138]

\[
\text{SNR} = \frac{\text{FSC}}{1 - \text{FSC}}.
\]  

(1.5)

Therefore, an FSC of 0.5 corresponds to a SNR of 1, which can be used to define the resolution of the maps. However, because the FSC is calculated from two halves of a dataset, an FSC of 0.5 would be an underestimate of the resolution of a map constructed from the entire dataset. It can be shown that an FSC of 0.143 calculated from two halves of a dataset is equivalent to an FSC of 0.5 calculated between a noiseless map and a map constructed from the entire dataset [153]. Therefore, the frequency at which the FSC falls below 0.143 for two maps constructed from two halves of a dataset is typically deemed the information limit and the inverse of this frequency is considered the resolution of the map constructed from the entire dataset. This procedure calculates an average resolution and provides a measure of the overall map variance. Some macromolecular complexes can have dynamic parts that result in regions of cryo-EM maps with lower resolution. Methods to estimate the resolution of local regions in maps exist but tend to be sensitive to noise [71, 89].

1.2.4 Cryo-EM reveals the structure of intact rotary ATPases

Cryo-EM has been the primary method used to study the structure of intact rotary ATPases. Early cryo-EM and negative-stain electron microscopy studies of rotary ATPases provided clear evidence that different enzyme complexes had different peripheral stalk stoichiometries [194, 46, 155]. Later improvements in methodology and increases in the size of datasets allowed individual subunits in the complexes to be
resolved [95, 119, 94, 93, 20, 14]. The improved resolution facilitated accurate fitting of atomic structures into cryo-EM maps and provided insight into the subunit-subunit interactions in different complexes. Cryo-EM was used to map the architecture of rotary ATPases, providing insight into the membrane-embedded regions where there is a lack of high-resolution structures available [94, 93, 14]. Studies of the V/A-ATPase from *T. thermophilus* allowed subunits of the c-ring and a-subunit to be resolved, providing important insight into the mechanism of proton translocation [94, 93]. In ATP synthases, cryo-EM allowed visualization of the membrane-embedded subunits required for enzyme dimerization [14]. Cryo-EM structures of eukaryotic V-ATPases provided insight into how the H- and C-subunits facilitate enzyme assembly and inhibition [20, 119]. However, the structure of the membrane-embedded region of V-ATPases remains ill-defined (Figure 1.10). Structures of three rotational states of an intact V-ATPase are described in Chapter 5 [201], which reveal the secondary structure of the membrane-embedded region of these rotary ATPases and suggest that flexibility may be important for enzyme function. Separate studies have resolved the architectures of the a-subunits from two different ATP synthases [4, 204], which show similar features with the a-subunit of V-ATPases and suggest a conserved mechanism of proton transport. How the V-ATPase may be inhibited during bacterial infection is discussed in Chapter 5, where the structural basis for the interaction between V-ATPase and SidK is presented.
Figure 1.9: **Fourier shell correlation (FSC) as a measure of map resolution.**

Two density maps (gray surface renderings) are refined independently using separate datasets. The Fourier transforms of the two maps are compared by calculating the correlation between Fourier components at different frequencies to estimate the FSC. The high SNR at low frequencies results in a FSC near one, whereas low SNR at high frequencies results in a FSC around zero. The frequency corresponding to an FSC of 0.143 is taken to be $1/resolution$ [153].
Figure 1.10: Previous cryo-EM map of the eukaryotic V-ATPase at \( \sim 11\,\text{Å} \) resolution [20]. Densities for most of the subunits were identified and segmented. Subunits constituting the c-ring (magenta) and the D- and F-subunits (blue) of the rotor could not be distinguished and their densities were grouped. \( \alpha \)-helices were resolved in some subunits including the peripheral stalks (beige and purple arrowheads) and parts of the rotor (blue arrowheads). No separation of conformations were performed. As a result, the map is an average of the V-ATPase in different rotational states. Scale bars, 25 Å.
Chapter 2

Template matching and classification system (TMaCS)


2.1 Author contributions

Jianhua Zhao\(^1,2\) (JZ) wrote the computer programs and analyzed the data. Marcus A. Brubaker\(^3,4\) (MAB) contributed parts of the machine learning program. MAB and John L. Rubinstein\(^1,2,5\) (JLR) conceived and supervised the research. JZ, MAB, and JLR wrote the manuscript. JZ and JLR prepared the figures.

\(^1\)Molecular Structure and Function Program, The Hospital for Sick Children Research Institute
\(^2\)Department of Medical Biophysics, University of Toronto
\(^3\)Department of Computer Sciences, University of Toronto
\(^4\)Toyota Technological Institute at Chicago
\(^5\)Department of Biochemistry, University of Toronto
2.2 Acknowledgements

We thank Drs. Ryan Lilien and Peter Rosenthal for a critical reading of the manuscript. JZ was supported by a studentship from the Hospital for Sick Children. MAB was supported by a postdoctoral fellowship from the Natural Science and Engineering Research Council of Canada. JLR was supported by a New Investigator Award from the Canadian Institutes of Health Research and an Early Researcher Award from the Ontario Ministry of Research and Innovation. This research was supported by the Natural Science and Engineering Research Council Discovery Grant 401724-12. Images of the V/A-ATPase were acquired during research funded by Canadian Institutes of Health Research Grant MOP 81294.

2.3 Introduction

Initial maps of protein particles may be calculated with as few as several thousand particle images. In principle, it should be possible to calculate 3D maps with ∼3 Å resolution from fewer than 13,000 particle images [69]. However, larger numbers of particle images often allow for improved resolution in 3D maps [10] due to a variety of sources of image deterioration [68] and, for some complexes, the inability to determine particle orientations precisely [14, 70]. In some cases, near-atomic resolution maps of protein complexes have been generated after analyzing hundreds of thousands of particle images [100, 62, 12]. Collection of large numbers of particle images presents a bottleneck in the process of determining 3D structures. At the initial stages of map calculation, it is often beneficial to select particle images manually. This process gives the microscopist an opportunity to inspect closely the quality of images and become familiar with attributes of the protein complex under investigation. However, manual
particle image selection is a tedious task and the edifying effect of manual selection is outweighed by its disadvantages when tens or hundreds of thousands of particle images are needed.

Computers are often suited better to highly-repetitive tasks, such as particle image selection, than are humans. Several groups have implemented automated particle image selection algorithms with varying levels of success [207]. Many of these approaches have been reviewed elsewhere [92, 207] and a full review is beyond the scope of this thesis. Programs have included generative approaches, such as template matching, that attempt to identify particles in micrographs based on similarity to reference images [29, 152, 186]; unsupervised approaches that attempt to recognize particles in micrographs by characteristics that distinguishes them from the background and other objects [2, 188]; and discriminative methods that attempt to learn to distinguish particles from non-particles in a micrograph based on positive and negative example images [9, 92, 108, 130, 172].

In our laboratory, we use cryo-EM to study membrane protein complexes such as rotary ATPases [14, 20, 94, 95, 93, 155]. These complexes present unique challenges in cryo-EM for several reasons. The complexes can be relatively small by cryo-EM standards (600 to 900 kDa), which results in low signal-to-noise ratios (SNRs) in micrographs. The complexes are intrinsic membrane proteins and consequently require detergents for solubilization, which further reduces image contrast [156, 163]. Finally, although the rotary ATPases are asymmetric, they possess a pseudo-two-fold axis between their soluble $F_1/V_1/A_1$ regions and their membrane-bound $F_O/V_O/A_O$ regions, and consequently the middle of the protein particles are regions of low protein density rather than high protein density, as is typical for other protein complexes. The challenge in identifying good particle images from a micrograph means that two or more different people tasked with selecting particle images from a micrograph often do not
select the same set of images. These complications also make the computational problem of automated particle image selection more challenging. In order to automate the particle image selection process, we tested many of the available software packages but were not able to establish conditions that recovered particle images sufficiently well to satisfy our requirements. Previous data used as a test-set for particle image selection programs [207] were often higher contrast than rotary ATPase images, making it difficult to assess which programs might be useful for selecting particle images from our micrographs.

Rather than attempting to adapt an existing particle image selection program to the rotary ATPases images, we elected to build an algorithm that would be well suited to the problem. As a result, we have developed an approach to particle image selection for the situation where a preliminary 3D map already exists for the protein complex under investigation. The program uses the same file formats and conventions as the program Frealign [64] and our own image analysis programs [20, 93] allowing the program to be easily integrated into the map refinement pipeline. A flow chart of the steps performed by the program is shown in Figure 2.1. Template images are generated from the initial 3D map and used for exhaustive searching of micrographs by calculation of cross-correlation functions. A low cross-correlation threshold is used to identify a large set of candidate particle images, including many non-particle images, from each micrograph. A subset of these candidate particle images is labeled manually as particles or non-particles in a purpose-built graphical user interface (GUI) and these data are used to train a machine learning classifier employing a support vector machine (SVM) [38]. Machine learning algorithms offer a data-driven approach to discriminating between classes of images, as opposed to attempting to encode rules manually to distinguish a particle image from a non-particle image. Support vector machines implicitly find a hyperplane that separates the two classes of images in a
Figure 2.1: Flowchart describing the steps in TMaCS. Projections are generated from a reference 3D map and used to perform template matching with a masked micrograph. A classifier is trained iteratively in an interactive user interface with a subset of the candidate particle images selected from template matching. Once trained, the classifier is used to identify particle images in the rest of the dataset. Projection matching is then used to rank the selected particle images from most to least reliable.
high dimensional and potentially non-linear feature space. SVMs are efficient to train, fast to evaluate and, given sufficient data, capable of learning complex classifiers. The classifier is refined iteratively by interactive correction of its output within the GUI. The remaining candidate particle images are then labeled as particles or non-particles by the algorithm. In a final stage of the analysis, the particle images identified by machine learning are subjected to projection matching to obtain a correlation coefficient between the particle image and its best matching projection of the initial 3D map. The correlation coefficient from projection matching is used to rank the selected particle images from most reliable to least reliable. Due to the combination of template matching and machine learning for classification, we have named our program TMaCS (template matching and classification system).

Analysis of the behavior of TMaCS shows that it performs almost as well as a human particle image selector for rotary ATPase images, with several independent human selectors opting to discard fewer than 6% of the images selected by the program. By using this approach, datasets can be extended rapidly into much larger datasets with little need for repetitive work by the experimenter. New datasets can also be acquired automatically for samples that are closely related to an existing 3D map, such as a structure of a complex with a new co-factor or binding-partner attached. Therefore, the TMaCS approach will help an experienced microscopist generate large datasets of particle images that can be incorporated into 3D map calculation with minimal human intervention.
2.4 Methods

2.4.1 Data and programming languages

Images of keyhole limpet hemocyanin from *Megathura crenulata* (KLH) were acquired from the National Resource for Automated Molecular Microscopy (http://ami.scripps.edu/redmine/projects/ami/wiki/KLH_dataset_I) [208]. KLH has a molecular weight of \( \sim 8 \) MDa with dimensions of \( \sim 300 \times 300 \times 400 \) Å. Images of the V/A-ATPase from *Thermus thermophilus* were from data recorded to calculate a 9.7 Å map of the complex [93] (EMD-5335). V/A-ATPase is \( \sim 600 \) kDa with dimensions of \( \sim 100 \times 100 \times 200 \) Å. Image pretreatment and template matching were performed with new programs written in modern Fortran and compiled with the gfortran compiler from the GNU compiler project. The interactive particle-labeling program and GUI were implemented in Python and the machine learning algorithm was implemented in Python.

2.4.2 Pretreatment of images

Micrographs of KLH and V/A-ATPase were \( 2048 \times 2048 \) pixels and \( \sim 13,000 \times \sim 12,000 \) pixels with pixel sizes of 2.2 Å and 1.4 Å, respectively. Micrographs were downsized \( 4 \times 4 \) and \( 2 \times 2 \) for the KLH and V/A-ATPase datasets, respectively. For the V/A-ATPase images, micrographs were masked by setting regions of high and low local variance to the average micrograph pixel value [29]. Once masking parameters, which are adjustable in the program, were selected manually with a few micrographs, they could be applied without change to all micrographs in the dataset. This process involved setting the values of the pixels in \( 240 \times 240 \) pixel blocks to the micrograph average where the variance was below 0.02 and above 1.1 times the image
variance. This local variance thresholding was applied twice in order to better distinguish desirable and undesirable regions in micrographs. Low frequency gradients in the V/A-ATPase micrographs were removed by subtracting from each pixel the average value of a $500 \times 500$ pixel box centered on that pixel.

2.4.3 Template matching

Projections of 3D maps of KLH and V/A-ATPase were created with the program Genproj_Fspace (J. Rubinstein, unpublished software), which uses the Euler angle convention from Spider [54] and Frealign [64]. A template image of the KLH complex was generated by averaging the $\theta=90^\circ$, $\phi=0^\circ$ projection of the complex with the $\theta=90^\circ$, $\phi=36^\circ$ projection of the complex. In-plane rotations were generated with $\psi \in \{0, 180^\circ\}$ in steps of $10^\circ$. The same procedure used to generate template images for KLH was performed for the V/A-ATPase, except that the Euler angles used were $\theta=90^\circ$ and $\phi \in \{0, 360^\circ\}$ in steps of $10^\circ$ to create the initial template with $\psi \in \{0, 360^\circ\}$ in steps of $10^\circ$ to produce 36 in-plane rotated template images. Normalized cross-correlation functions between each template image and the pretreated micrographs were calculated by padding the template image with zeros to the size of the entire micrograph. Cross-correlation functions were calculated including spatial frequencies between 500 and 20 Å for the KLH micrographs and 300 and 20 Å for the V/A-ATPase micrographs. The highest correlation coefficient among all the templates for each position in the micrograph was recorded in a combined cross-correlation map. Micrograph positions with cross-correlation coefficients above 0.01 and 0.05 were windowed from the untreated micrographs in $272 \times 272$ pixel and $256 \times 256$ pixel boxes for KLH and V/A-ATPase, respectively, and used as candidate particle images. For suppression of high cross-correlation values near cross-correlation peaks
during particle windowing, suppression radii of 300 and 140 Å for KLH and V/A-ATPase were used, respectively.

2.4.4 Classification by machine learning

An initial 100 candidate V/A-ATPase particle images from the template-matching algorithm were labeled with the GUI to assign particle images and non-particle images. Images were bandpass filtered in Fourier space, masked using a circular mask with a radius of 60 pixels, downsized to $32 \times 32$ pixels, and the mean of the image was subtracted from each pixel in the image. Principal component analysis (PCA) [78] was performed by singular value decomposition (SVD) [144] using the linear algebra algorithm numpy.linalg.svd in NumPy [131]. PCA functions as a dimensionality reduction step when the number of training examples is smaller than the number of pixels in each image and whitens the data, removing correlations between pixels, when the number of training examples is greater than or equal to the number of pixels in each image. The PCA coefficients for each image were then used to train a binary support vector machine (SVM) classifier [38] with a radial basis function kernel using an implementation of the algorithm from the Machine Learning Toolkit (MILK) in Python (Luis Pedro Coelho, http://packages.python.org/milk/). Subsequently, sets of 100 candidate particle images were displayed to the user with a particle or non-particle label assigned by the learning algorithm. These labels were corrected manually until 1,018 candidate particle images and 5,182 non-particle images had been inspected. These approved labels were used to train the classifier for the classification of the remaining candidate particle images.
2.4.5 Ranking of images by projection matching

After completion of template matching and machine learning to select particle images, the output particle images were subjected to a final ranking by projection matching. V/A-ATPase images were downsized to produce $64 \times 64$ pixel images (5.6 Å/pixel) and the average perimeter pixel value was subtracted from each image. The best possible cross-correlation coefficients between each image and 144,000 different projections of a V/A-ATPase cryo-EM map [93] were determined by projection matching with the program Search_Fspace [20].

2.4.6 Computational performance

All steps in the program were performed using a home built workstation with an Intel Core i7 2.93 GHz processor and 3.9 GB RAM memory running the Ubuntu 11.04 operating system. Pretreatment and template matching required $\sim 1 \text{ sec/micrograph}$ and $\sim 10 \text{ min/micrograph}$ for the KLH and V/A-ATPase datasets, respectively. Training of a classifier with V/A-ATPase candidate particle images required $\sim 7 \text{ min}$ and classification of 7,408 test images required $\sim 10 \text{ min}$. Manual training of the SVM classifier can be expected to take an experienced microscopist approximately one day of work. Projection matching with Search_Fspace required $\sim 13 \text{ sec/particle image}$.

2.5 Results

2.5.1 Overview and definitions

The TMaCS particle image selection approach consists of the sequential application of three algorithms: a template matching algorithm, a machine learning algorithm, and a projection matching algorithm. For clarity, we define template images as
the images derived from the initial 3D map that are used by the template matching algorithm to search the micrographs and produce a combined cross-correlation function for each micrograph. The template matching algorithm generates many candidate particle images. The machine learning algorithm labels the candidate particle images from template matching as either particle images or non-particle images. Finally, the projection matching algorithm ranks the images labeled by machine learning as particle images in order of reliability.

2.5.2 Template matching

Two sets of micrographs were used to characterize the template matching image selection algorithm. The first set of micrographs was comprised of images of KLH recorded on a CCD at 120 kV [208]. In the KLH images (Figure 2.2A), regions corresponding to protein particles have relatively high SNRs, are easy to detect by eye, and appear fairly homogeneous. Side views of KLH particles in these images were manually labeled previously [208] in order to allow development of particle image selection algorithms. The second set of micrographs were from cryo-EM of detergent-solubilized V/A-ATPase particles recorded on photographic film at 200 kV and digitized with a densitometer [93]. Film images often contain features such as a text label, regions of the carbon support substrate, and defects from developing the film. A dramatic example of a film with these artifacts is shown in Figure 2.2Bi (artifacts indicated by arrows). This second set of micrographs is typical of the film data acquired in our laboratory for the structural analysis of F- [14, 95, 155], V- [20], and V/A-ATPases [94, 93]. DDDs [99] have allowed for the collection of higher quality images, but images still suffer from low SNRs and it is often unclear which densities correspond to acceptable particle images. A set of acceptable particle images in mi-
Figure 2.2: **Pre-processing of micrographs and generation of template images for template matching.** (A) An example micrograph from the KLH dataset shows the desired side views of KLH (rectangular views), undesired top views of the complex (circular views), and the tobacco mosaic virus particles used as a magnification standard (elongated rod shapes). (B) An example of an atypical low-quality micrograph from the V/A-ATPase dataset before (i) and after (ii) variance thresholding and local background subtraction. The variance thresholding removes micrograph text labels, regions of carbon support substrate, ice contamination, and film developing artifacts from micrographs, indicated by the arrows. (C) Template images derived from a 3D KLH cryo-EM map (http://ami.scripps.edu/redmine/projects/ami/wiki/KLH_dataset_I) used for template matching. (D) Templates images derived from a 3D cryo-EM map of V/A-ATPase [93]. Scale bars, 300 Å.
crographs collected on film were labeled previously during a project that led to the determination of the V/A-ATPase structure at subnanometer resolution [93].

Micrographs were masked as described in the Methods section to remove areas of carbon, large areas of ice contamination, film development defects, and text labels on the film (Figure 2.2Bii, artifacts indicated by arrows). Template images from the KLH map (Figure 2.2C) and V/A-ATPase map (Figure 2.2D) were generated from the map projections as described in the Materials and Methods section. A larger number of templates can be incorporated easily in the template matching procedure at the expense of increased computation time. The map projection process included multiplication of the Fourier transform of the projection with an estimated contrast transfer function (CTF) for the microscope to account for the effects of the microscope’s point-spread function [64]. Here, a constant underfocus of 3 \( \mu m \) was used for all images, but defocus values specific to each micrograph could be incorporated into the analysis. However, we expect that this change would not affect the accuracy of the template matching.

Cross-correlation functions were generated between the micrographs and each of the template images. For each micrograph, a combined cross-correlation function was calculated that gives, at every position, the maximum value from all of the template-specific cross-correlation functions at that position. Examples of these combined cross-correlation functions are shown in Figure 2.3Ai and Bi for KLH and V/A-ATPase, respectively. This approach to template-based particle image selection ranks the candidate particle images from each dataset by assigning each candidate particle image with a cross-correlation score. Candidate particle images were selected from the micrographs by extracting windows systematically in order of cross-correlation coefficient. After each extraction, nearby maxima were suppressed with a mask that had a diameter similar to the diameter of the protein complex. Candidate particle
images were extracted until either no more candidates with correlation above an arbitrary cutoff could be found or a specified maximum number of candidate particle images per micrograph were extracted. An example of the output from this step of the algorithm is shown in Figure 2.3. For the KLH data (Figure 2.3Ai), every candidate particle image extracted from a single micrograph is shown in order of decreasing correlation with the templates. For the V/A-ATPase (Figure 2.3Bi), every 400\textsuperscript{th} extracted candidate image from a single micrograph is shown in order of decreasing correlation with the templates. As can be seen from the figure, the KLH particle images correspond to side views of KLH particles in the earlier, higher correlation coefficient, candidate images. In comparison, the selected V/A-ATPase candidate images correspond primarily to features in the image that are not V/A-ATPase particles. In order to assess quantitatively the precision of this template-based approach against the previous manual labeling of particle images, precision-recall curves were plotted for datasets of 81 micrographs of KLH and a single V/A-ATPase micrograph. Precision-recall curves \cite{92} plot the ability of a program to classify a particle image correctly against the fraction of labeled particle images that are retrieved from the image. In these curves, precision is defined by

\[ \text{precision} = \frac{TP}{TP + FP} \]  

(2.1)

where TP is the number of true positives, or true particle images, selected, and FP is the number of false positives, or features that are selected that do not correspond to particles. Recall, the fraction of available true particle images retrieved from the micrographs, is defined by

\[ \text{recall} = \frac{TP}{TP + FN} \]  

(2.2)
Figure 2.3: **Template matching search of micrographs.** (Ai) A combined cross-correlation map from the KLH micrograph from Figure 2.2A with the KLH template images from Figure 2.2C shows several clear maxima corresponding to side views of the KLH particles. (ii) Candidate particle images selected from the micrograph are primarily side views of the complex (green bordered images) with a few undesirable views of the complex (red bordered images) extracted with smaller correlation coefficients. (iii) A precision-recall curve constructed against a list of particle images selected manually from 81 micrographs shows high precision with recall. (Bi) A combined cross-correlation map from the V/A-ATPase dataset shows many more peaks of nearly equivalent intensity than the KLH combined cross-correlation map. (ii) Representative output from the cross-correlation search of one micrograph showing every 400\textsuperscript{th} candidate particle image reveals that the majority of selected regions from the micrograph do not correspond to views of the V/A-ATPase complex. (iii) A precision-recall curve constructed against a list of V/A-ATPase particle images selected manually from a single micrograph shows poor precision with recall. Scale bars, 300 Å.
where FN is the number of false negatives, or particle images that the program fails to identify in the micrographs. Precision recall curves were generated from the template matching algorithm by varying the correlation threshold at which a candidate particle image was accepted as a particle. As can be seen from the precision-recall curves in Figure 2.3Aiii and Biii, the template-based method works significantly better for the KLH images than for the V/A-ATPases images. Based on these results, it is apparent that the template-based particle image selection method is sufficient for identifying KLH particle images from micrographs, as seen previously with other particle image selection programs used with this dataset [29, 151, 186], but it is not sufficient for selection of V/A-ATPase particle images.

2.5.3 Machine learning with a support vector machine

In order to better identify particle images from the micrographs of V/A-ATPase, a machine learning approach was implemented. The output of the template-based particle image selection program with a low cross-correlation threshold was used to generate 256,039 candidate particle images from 306 micrographs of V/A-ATPase. From these candidate particle images, a test set consisting of 7,408 candidate particle images was set aside for characterizing the algorithm. A GUI-driven program was developed to aid the use of the machine learning algorithm. A screenshot from the GUI is shown in Figure 2.4. With the GUI, 100 particle images drawn randomly from the training set are displayed on the computer screen and labeled manually by the user as particles (Figure 2.4, green boxes) or non-particles (Figure 2.4, red boxes). The images are masked, bandpass filtered in Fourier space, downsized, and normalized prior to further processing. The program performs PCA [78] so that each of the images is described by their principal component coefficients. The principal
Figure 2.4: **Graphical user interface (GUI) for training the support vector machine (SVM) classifier.** Stacks of candidate particle images are imported into the program and 100 images are drawn randomly from all of the candidates and displayed. These initial images are labeled manually by clicking images to turn red-bordered images (non-particles) to green-bordered images (particles), and vice-versa. Subsequent windows containing 100 candidate particle images are then displayed with red- or green-borders assigned automatically in order to continue training the classifier. Once a sufficient training set has been inspected, the remaining candidate particle images are labeled by the classifier.
component coefficients are used to train a classifier that employs a support vector machine (SVM) [38]. SVMs are kernel-based algorithms that maximize the distance from a decision boundary to the support vectors in a high dimensional feature space, while minimizing the error associated with the decision boundary for inseparable datasets. The classification score is the signed distance to the SVM decision boundary. Images with a classification score greater than zero are classified as particles, and those with a classification score less than zero are classified as non-particles.

A subsequent set of 100 candidate particle images are then drawn randomly from the remaining candidate images and classified as particle images or non-particle images before being displayed by the GUI. The user corrects the classification manually where necessary, thereby providing more information for further training of the classifier. This process of PCA, training of the classifier, and correction of the classification is repeated until diminishing returns on training are observed, in this case with 1,018 images labeled as particle images and 5,182 images labeled as non-particle images being inspected or corrected by the user. The number of particle images required for optimal training of the classifier will likely depend on the nature of the images, with fewer training examples required for images that have high SNRs.

After training, the classifier was used to label the candidate particle images in the test set. Out of the 7,408 candidate particle images in the test set, 1,446 images were labeled by the classifier as particle images. Figure 2.5A shows every 10th image selected as a particle image by the machine learning algorithm. By default, candidate particle images that receive a positive score from the SVM are classified as particles while negative scores are classified as non-particles. However, precision recall curves can be generated from the SVM classifier by varying the default decision threshold of 0 to include fewer or more images as particles. A precision-recall curve for the SVM classification of particle images (Figure 2.5C) shows reasonable precision with recall
Figure 2.5: **Classification of candidate particle images from template matching of ten V/A-ATPase micrographs.** (A) Representative output from classification of images showing every 10th image that was labeled automatically as a particle, ranked in order of classification score. Most of the images correspond to typical V/A-ATPase particle images (labeled manually with green borders) while a small fraction of images still correspond to undesirable images (labeled manually with red borders). The first undesirable image to appear in this ranking is at position 981 out of 1,440 particle images. Scale bar, 300 Å. (B) The particle images were subjected to projection matching against the original 3D map and ranked according to their resulting cross-correlation score. In this ranking, the first undesirable image occurs at position 1,330 out of 1,440, suggesting that removal of the worst 8% of particle images in this ranking would eliminate almost all of the remaining undesirable images. Scale bar, 300 Å. (C) A precision-recall curve constructed against a list of particle images selected manually appears to indicate a lower precision than what is expected from the output in parts A and B, due probably to different particle images being selected in the manual particle image selection and automated selection. (D) Particle image selection precision was assessed in a subjective manner by having three experienced microscopists inspect the 1,440 images labeled as particles and either agree with or reject the label. The average interpolated precision was plotted against particle number in the ranked list. The curve shows high precision with particle number, consistent with what is observed from the output images in 4A and 4B. The error bars correspond to one standard deviation.
when compared to the dataset selected manually from the corresponding micrographs. In the curve, precision decreases slowly with recall until \(\sim 80\%\) recall where there is a sharp drop in precision. Inspection of the particle images associated with recall values of 80\% or higher revealed many particle images that were selected off center by the template matching algorithm. These off-center particle images tended to have particularly poor contrast, which likely led to correlation of the reference images with noise during template matching. It is probable that these particle images would also behave poorly during 3D map refinement by projection matching. Template matching with micrographs of KLH, which had high SNRs, did not lead to selection of off-center particle images.

Although better than the template-based approach shown in Figure 2.3Biii, the machine learning classification appears to have, at best, \(\sim 80\%\) precision in the precision-recall curve. However, inspection of the output particle images (e.g. Figure 2.5A) shows that a majority are particle images that an experienced human selector who has worked with rotary ATPase particles would designate as ‘good’ particle images. The reason for this apparent discrepancy is that, as described previously, even two experienced human microscopists will often not select the same particle images from a micrograph but will deem each other’s selections as appropriate. A more relevant statistic for an automated particle image selection algorithm in this situation is what fraction of the output images are acceptable to a human particle image selector. Here, precision retains the same definition as above but false positive particles are identified by inspection. To test this form of precision for the algorithm, we had three experienced rotary ATPase microscopists inspect the output of the machine learning algorithm and label undesirable, or false positive, images manually. Images could be labeled as undesirable for any reason, including crowding of particles, an apparent defect in the particle, or an apparent defect in the image. With this approach, the
number of false negative particles cannot be determined and recall is not known. The results of this analysis are shown in Figure 2.5D as the average precision as a function of particle number, where the particle images are ranked according to their output score from the classifier. As can be seen from Figure 2.5D, at most \(~6\%\) of particle images were labeled as undesirable, or false positives, by the average microscopist. Examples of undesirable particle images are shown with red borders in Figure 2.5A. Analysis of the accuracy of particle orientation determination [14, 70] has shown that 25 to 50 \%\) of rotary ATPases particle images have orientations that are determined incorrectly, but these errors affect only the appearance of the map in the worst cases [14]. Misaligned particle images and inclusion of undesirable images should affect the final 3D map in similar ways, suggesting that including 6 \%\) of undesirable images in a map should not affect the map significantly. The observation that the three different microscopists labeled different numbers of particle images as false positives (116, 36, and 114) illustrates the subjectivity of manual particle selection for these images. It is also worth noting that the average number of particle images identified by the machine learning algorithm (\(~140\) per micrograph) corresponds well with the average of \(~150\) particle images selected by a microscopist in the initial analysis of the data [93]. This number of particle images was obtained with the default SVM classifier score of 0 used as the decision threshold. The SVM=0 threshold corresponds to a precision of 65 \%\) with 70 \%\) recall as assessed by comparison with the dataset that was selected manually.

2.5.4 Ranking of particle images by projection matching

As can be seen from Figure 2.5A, the undesirable particle images in the dataset after classification, although accounting for at most 6 \%\) of the dataset, are spread
throughout the last 40% of the set when particle images having a classification score greater than zero are arranged according to their scores. In order to test if a better ranking could be obtained so that a fixed fraction of particle images that were ranked poorly could be discarded, we subjected the candidate particle images to a round of projection matching with Search_Fspace [20]. The projection matching utilized nearly 150,000 projections of the 3D map, a number of reference images that would result in slow processing speeds if used in the initial template matching step in TMaCS. Particle images were rearranged according to their correlation with an optimal matching projection of the reference 3D map. Again, the three experienced microscopists identified undesirable, or false positive, particle images in the dataset. The result of this analysis is shown in Figure 2.5B, where the undesirable particle images are shown with red borders. As can be seen from Figure 2.5B, ranking of particle images by their correlation against a matching projection of a 3D map is better able to discriminate desirable images from undesirable images as indicated by the clustering of the undesirable images towards the last 20% of the dataset. It is again interesting to note that the same experienced microscopist, when presented with the same particle images in a different sequence, often does not choose the same number as undesirable, further emphasizing the difficulty of this particle image selection task.

2.6 Discussion

The process we have described for partially-automated particle image selection lends itself easily to batch processing. Although some initial investment is required to establish a data processing pipeline, such as the generation of template images, determination of threshold values for cross-correlation, and training of a classifier,
these preparatory steps need only to be performed once. Subsequently, micrographs can be processed quickly to increase the size of the particle image dataset. As with manual particle image selection, there is some danger that the TMaCS approach will be limited by preconceived notions of what particles should look like and will consequently select only a subset of particle images, which may distort the final 3D map. Therefore, this approach does not remove the requirement for a reliable starting 3D map, generation of a diverse series of reference templates, and familiarity with the different views of the 3D complex when training the machine learning classifier.

Classification has been used previously to improve the quality of output from template matching particle selection algorithms [151]. A variety of machine learning algorithms have been employed similarly in the task of automated particle image selection [9, 92, 108, 130, 172]. Of these approaches, two groups attempted to select small asymmetric particles by cryo-EM that could be considered as challenging as the rotary ATPase particles used here. The other algorithms were tested on particle images that came from cryo-EM of large and/or symmetric particles or negative stain EM. Ogura and Sato [130] characterized their algorithm with a 200 kDa sodium channel while Arbelaez et al. [9] included images of the 900 kDa RNA polymerase II in their test data. Ogura and Sato [130] provided little quantitative evidence to support the effectiveness of their neural-network approach to particle selection. The approach of Arbelaez et al. [9] for selecting cryo-EM images of RNA polymerase II achieved moderate success with 50 % precision at 70 % recall.

Most of the earlier approaches attempt to use machine learning to identify particles directly from micrographs, with positive examples selected interactively [9, 108, 130, 172]. Negative examples were either selected interactively [108, 130], taken as all regions in a micrograph not selected by the user [172], or derived from the selected positive examples and their surrounding regions [9]. Overall, the approach we describe
here is most similar to that of Langlois et al. [92] in that we use one algorithm to generate candidate particle images and machine learning algorithm to classify them. As a consequence, both TMaCS and the approach of Langlois et al. [92] can provide both positive and realistic negative examples of particle images for optimal training of the classifier. Langlois et al. [92] used a modified Difference of Gaussians (DoG) picker to provide candidate particle images to an AffinityRank learning algorithm. Our approach differs in that the user interacts with the machine learning algorithm to improve the quality of the classification iteratively. With our program also, any other candidate particle image generating approach could be used to provide images to the machine learning algorithm. However, in our experience, the DoG picker approach [92, 188] tends to select individual F\textsubscript{1}/V\textsubscript{1}/A\textsubscript{1} regions or F\textsubscript{0}/V\textsubscript{0}/A\textsubscript{0} regions from rotary ATPases, rather than the centers of the particles, and is therefore not suited to this particular problem. This issue makes the template-matching algorithm a better choice for rotary ATPases, and presumably for other protein complexes that have their centre of mass in the middle of a low-density portion of the complex. Furthermore, a final projection-matching step in TMaCS provides a cross-correlation score that can be used to exclude the non-particle images that were labeled erroneously as particles by the learning algorithm. This procedure can increase the accuracy of particle image selection and speed up subsequent 3D map refinement.

The accuracy of TMaCS will depend on the quality of the micrographs and particle images. For micrographs with high SNRs, template matching is often sufficient for particle image selection, as we demonstrated with KLH. Bandpass filtering images can have a strong effect both for template matching and classification. Other program options, such as masking and downsizing images, affect computational speed but appear to have little effect on the accuracy of particle image selection. The TMaCS approach is also not expected to have particular sensitivity to inaccuracies in the
3D map used to generate template images for template matching. This tolerance is apparent from the way we produced template images by averaging 36 different side views of the V/A-ATPase.

Although TMaCS was developed for selecting rotary ATPase particle images, we predict that the program will be useful for many other specimens. Template matching and classification are both general techniques that can be adapted easily to a wide range of specimen images. The ability to separate the different steps in TMaCS further increases the flexibility of the system and facilitates its use for particle selection for other specimens. However, we expect that the components of the approach may become infeasible computationally with heterogeneous and impure samples. Template matching would require an impractical number of templates to identify particle images with structural heterogeneity in micrographs. Furthermore, classification by SVM would require a large number of training examples when it is required to distinguish between a protein with structural heterogeneity and an impurity of similar size in the sample.

In combination, machine learning algorithms for automated particle image selection are beginning to provide the tools necessary to obtain large cryo-EM datasets for high-resolution structure determination and analysis of heterogeneous protein structures. TMaCS provides one such route to obtaining a large dataset, performing almost as well as a human particle selector. The system has now been used with several different datasets in our laboratory and could be applied easily to other datasets. The program includes a user friendly GUI to facilitate the interactive training of the SVM classifier and all steps of the process are implemented easily into a pipeline for map refinement with Frealign and our own compatible programs.

The TMaCS software has been made available through the website:
http://www.sickkids.ca/research/rubinstein/
Chapter 3

Magnification Anisotropy

Published as a part of “Description and comparison of algorithms for correcting anisotropic magnification in cryo-EM images.” *Journal of Structural Biology* 192(2):209-15, 2015 [203]

3.1 Author contributions

Jianhua Zhao\(^1,2\) (JZ) noticed the systematic error in CTF parameters. Samir Ben-lekbir\(^1\) (SB) acquired the images of thallous chloride. John L. Rubinstein\(^1,2,4\) (JLR) analyzed the images of thallous chloride to measure the anisotropy of magnification, and wrote the programs to correct images for anisotropic magnification in real space and in Fourier space. JLR, JZ, and Marcus A. Brubaker\(^3\) (MAB) conceived of the approach to correct CTF parameters measured from images with anisotropic magnification and JZ wrote the program. JLR made the figures and JLR and JZ wrote the manuscript.

\(^1\)Molecular Structure and Function Program, The Hospital for Sick Children Research Institute
\(^2\)Department of Medical Biophysics, University of Toronto
\(^3\)Department of Computer Sciences, University of Toronto
\(^4\)Department of Biochemistry, University of Toronto
3.2 Acknowledgements

We are grateful to Alexis Rohou and Timothy Grant from Nikolaus Grigorieff’s laboratory, who quickly put us on the correct track when we noticed the apparent systematic astigmatism in plots of $\Delta z_1$ versus $\Delta z_2$. We thank Richard Henderson for critical comments on the manuscript. JZ was supported by a Doctoral Postgraduate Scholarship from the National Science and Engineering Council of Canada and a Mary Gertrude l’Anson Scholarship. JLR was supported by a Canada Research Chair. This work was funded by NSERC grant 401724-12 to JLR.

3.3 Introduction

Anisotropic magnification in electron microscope images of 2D crystals was first described more than 30 years ago [16] but has not been detected recently with most modern electron microscopes used under conditions typical for cryo-EM with film or CCD cameras. DDD cameras have revolutionized cryo-EM [90, 170]. Two DDD manufacturers, Gatan and Direct Electron, have produced cameras with pixel sizes between 5.0 and 6.5 $\mu$m, which is smaller than is typical with CCD cameras. Because DDD cameras are placed below the projection chamber of the microscope, images formed on the DDD have an additional magnification relative to photographic film used with the same microscope. As a consequence, electron microscopes used for high-resolution studies with a DDD may need to be set to a lower nominal magnification than was typical previously [99]. At these conditions, anisotropic magnification has been detected in several microscopes, representing a large fraction of the instruments where this issue has been investigated. The phenomenon was seen first in a modern electron microscope with a 300 kV FEI Titan Krios microscope used with a Gatan K2
Summit DDD [61, 60] and has been detected subsequently on other FEI microscopes, including at least a 300 kV Tecnai Polara and a 200 kV Tecnai TF20. With a 14 µm pixel size, the FEI Falcon series of DDDs does not require a low magnification setting and anisotropic magnification has not been detected with any microscope used in combination with this camera.

The consequence of anisotropic magnification in single particle cryo-EM is that images of molecules lying in different orientations on the specimen grid cannot be averaged coherently, limiting the resolution that can be obtained in 3D maps calculated from these data. For example, with 2 % magnification anisotropy, a particle that is 300 Å long would appear to be 300 Å long when lying in one orientation and 306 Å long when lying in another orientation. This effect will have more severe consequences for larger particles than for smaller particles. Anisotropic magnification can be detected by an elliptical appearance of the powder diffraction patterns calculated from images of a variety of specimens including polycrystalline gold, graphite, or thallous chloride. The observed anisotropy in magnification exists at the low magnification settings used to acquire images, but not at the higher magnifications used for correcting objective lens astigmatism. The cause of this effect has been proposed to be dirt in the microscope column that becomes charged and acts as an additional lens; in one 120 kV microscope the anisotropy was found to change slowly from 1.4 % to 2.5 % over the 7 years between 1983 and 1990 (Richard Henderson, personal communication). This observation, as well as our own measurements described below, suggest that anisotropic magnification is stable over the course of weeks or months, but should be measured periodically with any microscope. The dependence of magnification anisotropy on the nominal magnification of the microscope explains why the issue was not detected in current microscopes when higher nominal magnifications were typical.
In order to calculate high-resolution maps from cryo-EM images, the CTF of the microscope must be corrected. Most algorithms currently in use correct for the CTF during calculation of the 3D map. Introduction of anisotropic magnification after correction of objective lens astigmatism causes Thon rings from images to be elliptical, even when there is no objective lens astigmatism present. CTF parameters determined from elliptical Thon rings will therefore suggest the presence of astigmatism. A consistent ellipticity for Thon rings from images despite several attempts at correcting astigmatism at high magnification could indicate anisotropic microscope magnification. The problem may be detected similarly by the presence of systematic astigmatism in CTF parameters for datasets obtained over several EM sessions where different amounts of astigmatism are expected. An electron optical method for removing the effect has not yet been described.

To measure magnification anisotropy, movies of thallous chloride crystals were recorded in a FEI TF20 microscope operating at 200 kV and equipped with a Gatan K2 Summit direct detector (Figure 3.1A). The thallous chloride lattice has a spacing of 3.842 Å. Consequently, the power spectrum from each crystal is expected to show a peak at distance \( p \cdot N/3.842 \) Å pixels from the origin, where \( p \) is the pixel size in Ångstroms, and \( N \) is length in pixels along each edge of the image (Figure 3.1B). The average of power spectra from images of particles is expected to produce a ring with this distance as the radius (Figure 3.1C). However, the resulting average of power spectra had a slightly elliptical appearance. From 4096 × 4096 pixel averaged power spectra of images, each containing several thallous chloride particles, we recorded the lengths and angles of 209 vectors from the origin of the pattern to the diffraction ring. Figure 3.1D shows a plot of the lengths of the vectors as a function of the angles they make with the positive \( k_x \)-axis of the power spectrum. A sinusoidal function was fit to the plot, allowing us to determine the amplitude and direction of magnification.
Figure 3.1: Anisotropic magnification in images. (A) A thallous chloride particle in a $512 \times 512$ pixel image. Interatomic planes are apparent in the image. The scale bar corresponds to 100 Å. (B) The calculated power spectrum from the image in part A shows diffraction peaks. The expected distance from each peak to the centre of the image is $1.45 \, \text{Å} \cdot 512/3.842 \, \text{Å}$ pixels. (C) The average of 558 power spectra from thallous chloride particle images. The pattern is elliptical, not round, as seen by overlay of a perfect square that touches the pattern at its sides but not the top or bottom. (D) A plot of the distance of the pattern from the origin of the power spectrum versus the angle of each point from the $k_x$-axis in the power spectrum from the average of 99 full frame power spectra padded to $4096 \times 4096$ pixels.
anisotropy. We determined a 2% difference in magnification between the x- and y-axes, with the images magnifying more along the y-axis compared to the x-axis. Images were corrected for the magnification anisotropy by stretching the images in the x-axis to match the magnification in the y-axis. We then developed a method for recovering true CTF parameters from CTF parameters calculated from images with anisotropic magnification.

3.4 Methods and Results

For high-resolution maps to be calculated from cryo-EM data, images must be corrected for the effects of the contrast transfer function (CTF). CTF parameters are measured typically from the power spectra of images and used to correct for the CTF during calculation of the 3D map. A stretch or contraction of an image causes an analogous change in the image power spectrum that affects the CTF parameters determined from the power spectrum [190, 117]. The CTF is described by the equation:

\[ CTF(f) = -w_{\text{phase}} \sin(\chi(f)) - w_{\text{amp}} \cos(\chi(f)) \]  

(3.1)

where \( w_{\text{amp}} \) is the fractional amount of amplitude contrast (~0.07 for a cryo-EM image [181]). The parameter \( w_{\text{phase}} \), the fractional amount of phase contrast, is given by \( \sqrt{1 - w_{\text{amp}}^2} \), and

\[ \chi(f) = \frac{\pi \lambda}{f^2} \cdot \Delta z - \frac{\pi \lambda^3}{f^4} \cdot C_s \]  

(3.2)

where \( \lambda \) is the wavelength of electrons in Ångstroms, \( f \) is the frequency in the Fourier transform in Ångstroms\(^{-1} \), \( \Delta z \) is the defocus of the microscope in Ångstroms, and \( C_s \) is the spherical aberration of the objective lens in Ångstroms. For the resolutions of current interest in biological cryo-EM (up to \(~2\) Å) and with the defocuses used
typically (tens of thousands of Ångstroms), the behaviour of the CTF is due almost entirely to the first term in equation 3.2. The CTF is modeled by assuming a specific image pixel size \( p \) and then determining the value of \( \Delta z \). A change in the pixel size from \( p \) to \( p/a \) would change \( f \) to \( a \cdot f \) and equation 3.2 becomes

\[
\chi(a \cdot f) \approx \frac{\pi \lambda}{(a \cdot f)^2} \cdot \Delta z = \frac{\pi \lambda}{f^2} \cdot \frac{\Delta z}{a^2}.
\]  

(3.3)

As a consequence, stretching of the image or contraction of the Fourier transform by a factor \( a \) will modify the apparent defocus in the image by a factor of \( 1/a^2 \). This phenomenon can be observed in Figure 3.3A, which compares, for a 200 kV microscope with a \( C_s \) of 2 mm, a 1D CTF with \( \Delta z =10,000 \) Å (black line), a 1D CTF with \( \Delta z =10,000 \) Å but with the x-axis contracted by a factor of \( 1.1 \times \) (broken blue line), and a 1D CTF with \( \Delta z =10,000/1.1^2 \) Å= 8,264 Å (red line).

For a 2D image, two defocus values, \( \Delta z_1 \) and \( \Delta z_2 \), and an angle of astigmatism, \( \phi_{ast} \), are needed to describe the CTF (Figure 3.2). The angle of astigmatism \( \phi_{ast} \) is defined as the angle between the semi-axis of the Thon ring with defocus \( \Delta z_1 \) and the \( k_z \)-axis of the Fourier transform. The parameters \( w_{phase}, w_{amp}, \lambda, \) and \( C_s \) are microscope specific while \( \Delta z_1, \Delta z_2, \) and \( \phi_{ast} \) can change with each image. Equations 3.1 and 3.2 may be used to describe a 2D CTF if one defines \( \Delta z \) in equation 3.2 as the effective defocus in the direction \( \phi \) as given by:

\[
\Delta z = \frac{\Delta z_1 + \Delta z_2 + \cos(2 \cdot (\phi - \phi_{ast})) \cdot (\Delta z_1 - \Delta z_2)}{2}.
\]  

(3.4)

During analysis of a dataset of cryo-EM images acquired on a FEI Company 200 kV F20 microscope equipped with a Gatan K2 Summit DDD, we plotted the measured
Figure 3.2: **Modeling the contrast transfer function (CTF).** (A) The CTF is described in 2D as a distortion of a unit circle to form an ellipse. The semi-axes of the ellipse correspond to two defocus values ($Dz_1$ and $Dz_2$). The angle of astigmatism ($\phi_{\text{ast}}$) describes the angle from the $k_x$-axis to the $Dz_1$ semi-axis. (B) Power spectrum from a micrograph (*right half*) and square of the modeled CTF (*left half*) showing oscillations in the CTF. Signal in the power spectrum fades at high frequencies due to fewer scattering events and incoherence in the microscope.
Figure 3.3: **Effect of anisotropic magnification on CTF parameters.** (A) The effects of stretching or contracting a Fourier transform on the CTF can be corrected to high precision by adjusting the defocus parameter. (B) A plot of $\Delta z_1$ vs $\Delta z_2$ for CTF parameters determined from images that suffer from magnification anisotropy show an apparent systematic astigmatism, with $\Delta z_1$ being consistently larger (or smaller) than $\Delta z_2$. Note that the ratio of $\Delta z_1$ to $\Delta z_2$ can be reversed by redefining which defocus corresponds to which parameter and adding $\pi/2$ to $\phi_{ast}$. (C) The apparent systematic astigmatism induced by anisotropic magnification can be removed by correcting CTF parameters as described in the text.
contrast transfer function defocus parameters $\Delta z_1$ versus $\Delta z_2$ (Figure 3.3B). Note that the choice of which defocus to define as $\Delta z_1$ and which to define as $\Delta z_2$ is arbitrary and can be reversed by changing $\phi_{ast}$ by $\pi/2$. The plot showed an apparent systematic astigmatism: for randomly introduced astigmatism, one would expect points in the plot to fall on the line $\Delta z_1 \approx \Delta z_2$. In contrast, the points fall on a line that has a slope that indicated a constant astigmatism of approximately 2%, consistent with the measured magnification anisotropy of the microscope. Plotting $\Delta z_1$ versus $\Delta z_2$ and looking for deviation from the line $\Delta z_1 = \Delta z_2$ with the form $\Delta z_1 \approx a\Delta z_2$ or $\Delta z_1 \approx \frac{1}{a}\Delta z_2$ presents a straightforward way of detecting anisotropic magnification in a microscope. However, if the anisotropic magnification factor $a$ is small compared to the variance of the ratio of $\Delta z_1/\Delta z_2$ due to errors in adjusting the microscope objective lens astigmatism, it will not be possible to detect anisotropic magnification in this way.

Correction of images for anisotropic magnification before measuring CTF parameters removes this systematic deviation from the $\Delta z_1 = \Delta z_2$ line. However, it is also possible to correct CTF parameters that have been measured from images with anisotropic magnification. It is advantageous to measure CTF parameters from images with magnification anisotropy and then correct the parameters for three reasons. First, correcting whole images and movies, rather than individual particle images, could be expensive computationally. Second, storing corrected whole images and movies places significant demand on data storage resources, which may be strained already by DDD data collection. Finally, correcting images for anisotropic magnification can introduce image artefacts that may decrease the accuracy with which CTF parameters can be calculated. Recovery of true parameters from images with a known amount of magnification anisotropy relies on representing the CTF as a distortion of the unit circle in Fourier space, $k_x^2 + k_y^2 = 1$, to an ellipse (Figure 3.2A). The unit
circle may be represented in matrix notation as

\[ k^Tk = 1 \]  \hspace{1cm} (3.5)

where \( k^T = \begin{bmatrix} k_x & k_y \end{bmatrix} \) and \( k = \begin{bmatrix} k_x \\ k_y \end{bmatrix} \). As seen in Figure 3.2A, an ellipse can be used to represent the CTF parameters \( \Delta z_1, \Delta z_2, \) and \( \phi_{ast} \). Defining \( k = E_{CTF} \ell \) where \( E_{CTF} = R_{CTF}S_{CTF}R_{CTF}^T \) with

\[
R_{CTF} = \begin{bmatrix}
\cos \phi_{ast} & -\sin \phi_{ast} \\
\sin \phi_{ast} & \cos \phi_{ast}
\end{bmatrix}
\]

and \( S_{CTF} = \begin{bmatrix}
\frac{1}{\sqrt{\Delta z_1}} & 0 \\
0 & \frac{1}{\sqrt{\Delta z_2}}
\end{bmatrix} \)

and substituting into equation 3.5 gives the equation of the ellipse as

\[
\ell^TE_{CTF}^TE_{CTF}\ell = 1.
\]  \hspace{1cm} (3.6)

\( R_{CTF}^T \) describes a rotation by the angle of astigmatism \( \phi_{ast} \) from the \( k_x \)-axis, \( S_{CTF} \) describes a stretch or compression along the \( k_x \)- and \( k_y \)-axes by \( \Delta z_1 \) and \( \Delta z_2 \), and \( R_{CTF} \) rotates the ellipse axes back to their correct angles. The magnitudes of the two defocus values (\( \Delta z_1 \) and \( \Delta z_2 \)) can be recovered by determining the two eigenvectors of \( E_{CTF}^TE_{CTF} \). The angle of astigmatism \( \phi_{ast} \) can be calculated as the angle that the eigenvector representing \( \Delta z_1 \) makes with the \( k_x \)-axis. By definition, the eigenvectors of a matrix \( Y \) obey the equation \( Yx = \lambda x \). That is, upon multiplication by \( Y \) the eigenvectors change only their magnitude, not their direction. Eigenvectors of a 2 × 2 matrix can be obtained conveniently using algorithms in standard numerical analysis libraries such as \textit{LAPACK}, \textit{NumPy}, or in \textit{MATLAB}. As described earlier,
Chapter 3. Magnification Anisotropy

magnification anisotropy further distorts the ellipse by the transformation

$$E_{ani} = R_{ani} S_{ani} R_{ani}^T$$

(3.7)

where \( R_{ani} \) describes a rotation by the angle of magnification anisotropy \( \theta_{ani} \) from the \( k_x \)-axis, and \( S_{ani} \) describes a stretch or compression along the \( k_x \)-axis by the magnitude of magnification anisotropy \( a \). Defining \( \ell = E_{ani} m \) and substituting this definition into equation 3.6 gives the equation of the new ellipse as

$$m^T E_{ani}^T E_{ani}^T CTF E_{ani} m = 1.$$  

(3.8)

The eigenvectors of \( E_{ani}^T CTF E_{ani} \) correspond to the apparent CTF parameter values, \( \Delta z'_1, \Delta z'_2, \) and \( \phi'_ast \), measured from the power spectra of images where there is anisotropic magnification. From these values and values of \( a \) and \( \theta_{ani} \) measured from a powder diffraction pattern (Figure 3.1), it is possible to calculate the true values of \( \Delta z_1, \Delta z_2, \) and \( \phi_{ast} \) from the eigenvectors of

$$E_{CTF}^T CTF E_{CTF} = E_{ani}^T E_{CTF}^T E_{CTF} E_{ani}^{-1}$$

(3.9)

where \( E_{CTF} = R_{CTF}^T S_{CTF} R_{CTF} \) and \( E_{ani}^{-1} = R_{ani}^T S_{ani} R_{ani} \). The matrices \( R_{CTF}, R_{CTF}', S_{CTF}, \) and \( S_{CTF}' \) are the same as \( R_{CTF}, R_{CTF}^T, \) and \( S_{CTF}, \) but use the apparent CTF parameters measured from distorted images rather than the true CTF parameters, and

$$S_{ani}^{-1} = \begin{bmatrix} 1/a & 0 \\ 0 & 1 \end{bmatrix}.$$

This approach for correcting CTF parameters measured from images with magnification anisotropy was implemented in a standalone program that operates on Relion...
.star files [162]. As can be seen in Figure 3.3C, the method brings points in a plot of $\Delta z_1$ versus $\Delta z_2$ back to the $\Delta z_1 = \Delta z_2$ line. The method is equivalent to correcting an image for magnification anisotropy and then measuring CTF parameters from its power spectrum.

### 3.5 Discussion

Plotting of $\Delta z_1$ versus $\Delta z_2$ from a dataset of micrographs can reveal the presence of anisotropic magnification in the microscope (Figure 3.3B). If all of the images in a dataset were obtained from a single EM session where the objective lens stigmator was adjusted once, it would not be unusual to find the points of this plot fall off the line $\Delta z_1 = \Delta z_2$. However, if the deviation of points from the line is due to anisotropic magnification, the positions of the points should have the form $\Delta z_1 \approx a^2 \Delta z_2$ or $\Delta z_1 \approx \frac{1}{a^2} \Delta z_2$. Deviation due to a constant astigmatism should result in point positions with the form $\Delta z_1 \approx \Delta z_2 + x$, where $x$ is the amount of astigmatism. Once anisotropic magnification has been detected, the amount and extent of anisotropy can be measured precisely using a diffraction standard such as the thallous chloride crystals. The interpolation scheme proposed [203] is not entirely free of artefacts. Consequently, it would be better to correct anisotropic magnification with improved electron optics, rather than correct the effects of anisotropic magnification computationally. If correction by interpolation must be performed, the best approach would be to incorporate the correction into the 3D map calculation software in order to avoid performing interpolation twice with the same images.
3.6 Software availability

All of the original software described above is available from https://sites.google.com/site/rubinsteingroup/direct-detector-distortion.
Chapter 4

Rotational states of the V-ATPase

Published as “Cryo-EM observation of rotational states in a eukaryotic V-ATPase”


4.1 Author contributions

Samir Benlekbir¹ (SB) and John L. Rubinstein¹,²,³ (JLR) initiated the project. Jianhua Zhao¹,² (JZ) and SB collected images and performed pre-processing steps. JZ performed the image analysis. JZ and JLR interpreted the data, prepared figures, and wrote the manuscript. JLR and JZ contributed new computer algorithms used in image analysis.

4.2 Acknowledgements

We thank Drs. P. Rosenthal, R. Henderson, V. Kanelis, and L. Kay for comments on the manuscript. JZ was supported by a Doctoral Postgraduate Scholarship from

¹Molecular Structure and Function Program, The Hospital for Sick Children Research Institute
²Department of Medical Biophysics, University of Toronto
³Department of Biochemistry, University of Toronto
the National Science and Engineering Council of Canada and a Mary Gertrude lAnson Scholarship. JLR is the Canada Research Chair in Electron Cryomicroscopy. This work was supported by operating grant MOP 81294 from the Canadian Institutes of Health Research.

4.3 Author information

Cryo-EM maps have been deposited in the Electron Microscopy Data Bank with accession numbers EMD-6284, EMD-6285, and EMD-6286. Atomic models have been deposited in the Protein Data Bank with accession numbers 3J9T, 3J9U, and 3J9V. The authors declare no competing financial interests.

4.4 Introduction

Studies of ATP synthases, V-ATPases, and bacterial/archaeal V/A-ATPases have suggested that flexibility is necessary for the catalytic mechanism of rotary ATPases [133, 175, 205], but the structures of different rotational states have never been observed experimentally. Here we used cryo-EM to obtain structures for three rotational states of the V-ATPase from the yeast Saccharomyces cerevisiae. The resulting series of structures revealed the stoichiometry of the c-ring and the long, highly-tilted transmembrane α-helices in the a-subunit that interact with the c-ring. The three different maps show the conformational changes that occur to couple rotation in the soluble catalytic region to the symmetry-mismatched membrane-bound region. Almost all of the subunits of the enzyme undergo significant conformational changes during the transition between these three rotational states. The structures of these states provide direct evidence that deformation during rotation enables the smooth
transmission of power through rotary ATPases.

4.5 Methods

4.5.1 Yeast membrane purification

Yeast membranes were purified from \textit{S. cerevisiae} as described previously \cite{20} with slight modifications. Yeast cultures were grown at 30 °C for all steps. \textit{S. cerevisiae} strain SABY31 was grown on YPD agar plates (1 % [w/v] yeast extract, 2 % [w/v] peptone, 2 % [w/v] glucose, 1.5 % [w/v] agar). Large colonies were used to inoculate 100 ml of YPD media (1 % [w/v] yeast extract, 2 % [w/v] peptone, 2 % [w/v] glucose) and the culture was grown overnight with shaking (225 RPM). The saturated yeast culture was added to 10 l of YPD media supplemented with 0.3 % [v/v] Antifoam 204 (Sigma Life Sciences) and grown overnight in a 11 l BioFlo fermenter (New Brunswick Scientific). The rest of the steps were performed at 4 °C. The cells were harvested by centrifugation at 4,000 g for 10 min in a tabletop centrifuge and resuspended in lysis buffer (140 mM NaCl, 3 mM KCl, 10 mM Na$_2$HPO$_4$, 2 mM KH$_2$PO$_4$, 8 % [w/v] sucrose, 2 % [w/v] sorbitol, 2 % [w/v] glucose, 5 mM aminocaproic acid, 5 mM benzamidine, 5 mM ethylenediaminetetraacetic acid (EDTA), 0.001 % [w/v] phenylmethanesulfonyl fluoride [PMSF], pH 7.4). Cells were lysed using a BeadBeater (BioSpec Products) with 0.5 mm glass beads. Cell debris was removed by centrifugation at 3,000 g in a tabletop centrifuge. Cell membranes were separated from the soluble fraction by ultracentrifugation at 42,000 g. Cell membranes were resuspended in lysis buffer, aliquoted, and stored at -80°C.
4.5.2 Purification of V-ATPase

V-ATPase was purified from *S. cerevisiae* as described previously [20] with slight modifications. Steps were performed at 4 °C. Yeast membranes were solubilized in 1 % [w/v] n-dodecyl β-D-maltopyranoside (DDM) and insoluble material was removed by ultracentrifugation at 42,000 g. The solubilized V-ATPase was bound to 1 ml of anti-FLAG M2 agarose beads (Sigma Life Sciences) via 3×FLAG tags on the C termini of the A-subunits. The beads were washed with six bed volumes of DTBS buffer (50 mM Tris-HCl, 150 mM NaCl, 0.02 % [w/v] DDM, pH 7.4) and V-ATPase was eluted with 1.5 bed volumes of DTBS buffer containing 150 μg/ml 3×FLAG peptide.

4.5.3 Sample preparation

Grid nanofabrication

Holey carbon grids were prepared as described previously [110]. Three drops of a 1 % [w/v] Formvar/chloroform solution was applied to a glass slide and a holey-patterned polydimethylsiloxane (PDMS) stamp was pressed onto the glass slide with 2 lb of pressure for 1 min. The resulting holey Formvar film was floated onto water and 400 square mesh copper/rhodium grids (Electron Microscopy Sciences) were placed on top of the film, copper side up. The film and grids were removed from the water by touching and lifting with a large piece of wax backing paper from parafilm and placed on a filter paper (Whatman) to dry. Some parts of the holey pattern may be covered with a thin film of residual Formvar, resulting in partially-filled holes (‘pseudo-holes’). To remove pseudo-holes, the grids were baked at 120 °C for 1.5 h. The grids were coated with carbon using a carbon evaporator (Edwards) and the Formvar was removed by washing with chloroform.
Sample vitrification

4 µl of purified V-ATPase at ∼10 mg/ml was applied to a holey carbon grid that was glow discharged in air for 2 min. Excess sample was removed by blotting for 20 to 23 s in a FEI Vitrobot Mark III operating at 4 °C and 100 % humidity. A blotting offset of 5 mm was used. The grid was plunge-frozen in a 3:7 ethane:propane mixture maintained at near liquid nitrogen temperature. The samples were stored in liquid nitrogen until imaging.

4.5.4 Imaging

All imaging was conducted on an FEI Tecnai F20 electron microscope operating at 200 kV. Prior to each imaging session, microscope alignments were performed to correct for gun tilt, condensor aperture centering, condensor lens astigmatism, objective aperture centering, objective lens astigmatism, and coma. The microscope was focused on the carbon support layer in the sample at 100,000× magnification and defocused by 1-7 µm prior to imaging. Images and movies were collected with a Gatan K2 Summit direct detector device operating in counting mode at 34,483× magnification, resulting in an approximate pixel size of 1.45 Å. 15 s exposures were recorded at 2 frames/s and ∼2.4 e−/Å²/s.

4.5.5 Image processing

Movie processing and particle image extraction

Movie frames were aligned with alignframes_lmbfgs [154] and averaged with shiftframes. Averaged frames were used for contrast transfer function (CTF) determination with CTFFIND3 [117]. Coordinates for 200 to 250 particle-like features per
image, some of which corresponded to true particle images, were selected by template matching with TMaCS (Chapter 2) [202]. These coordinates were then used to extract candidate particle images (320×320 pixels) with individual particle motion correction using the alignparts_lmbfgs algorithm [154]. A measured ∼2 % anisotropy in the magnification of the microscope was corrected in particle images by interpolation in Fourier space (Chapter 3) [203], to produce a calibrated pixel size of 1.45 Å. CTF parameters were corrected to account for the effects of this anisotropy [203].

Image classification and refinement

All classification and refinement steps were performed in Relion [162, 161] using the SciNet cluster [103] and the SickKids High Performance Facility. Candidate particle images were subjected to 2D classification using 100 classes and batches of 50,000 to 100,000 images. Particle images in classes with averages that resembled projections of the V-ATPase were selected for 3D classification. 3D classification was performed with 2 to 10 classes and a previous map of the V-ATPase as the starting model [20]. Particle images in 3D classes that resembled the V-ATPase and contained high-resolution features were selected for further analysis. 3D refinement was performed using a circular mask and select 3D classes as starting models. Final maps were filtered and sharpened using relion_postprocess and a tight mask around the protein complex.

Map analysis and model building

Local resolution was estimated with ResMap [89] using two half maps refined independently. 3D maps were segmented with UCSF Chimera [139, 141]. Homology models were calculated either with Phyre2 [86] or with HHpred [171] and MODELLER [158]. Atomic models from the Protein Data Bank (PDB) and homology models
were docked into EM maps by Molecular Dynamics Flexible Fitting (MDFF) [182]. Rotations of subunits between states were measured in UCSF Chimera and movies were generated with UCSF Chimera.

4.6 Results and Discussion

4.6.1 Cryo-EM maps of three rotational states of V-ATPase

To investigate the rotational states of the eukaryotic V-ATPase, we isolated the complex from S. cerevisiae in detergent and imaged it by cryo-EM (Figure 4.1). Classification of V-ATPase images enabled the identification of three 3D classes that gave maps at resolutions of 6.9 Å (47 % of images), 7.6 Å (36 % of images) and 8.3 Å (17 % of images) (Figure 4.2a and b), with 106,445 particle images contributing to the three maps. The class with the largest population corresponds most closely to an earlier cryo-EM map of the S. cerevisiae V-ATPase where conformational separation was not performed [20], but differs from the state identified for the Manduca sexta V-ATPase [150]. Rotational state 3 is the most different from the other two and is also the class with the fewest particle images. The unequal distribution of particle images in the three classes suggests that the V-ATPase relaxes to three unequal states in the absence of nucleotide. Cross-sections through the maps show that α-helices are well resolved at these resolutions (Figure 4.3; Movie 1).

Map resolution does not appear to be homogenous within each 3D map, suggesting further conformational heterogeneity within each class. However, additional classification strategies involving focusing classification on the region of interest [6] did not produce meaningful classes. The resolution appears better in the soluble V$_1$ region of the complex (Figures 4.3a-i and 4.2c) than in the membrane-bound V$_O$
Figure 4.1: **Data collection of V-ATPase images.** (A) A representative micrograph with examples of V-ATPase particle images circled in red. These particle images were selected from the 200 candidate particle images identified automatically from template matching in \textit{TMaCS} [202]. (B) Tracking of particle and other image feature trajectories with the \textit{alignparts_lmbfgs} algorithm [154]. Trajectories are exaggerated by a factor of 5 to allow visualization.
Figure 4.2: **3D maps from rotational states.** (A) Surface rendered views of the three 3D maps are shown. Scale bars, 25 Å. (B) Fourier shell correlation curves after a gold standard refinement of the three maps are shown. The resolutions measured from these curves at FSC=0.143 are the same as the resolution measured after correcting for masking effects by high-resolution noise substitution calculations [30]. (C) Local resolution estimation shows that features in the V₁ region are better resolved than in the V₀ region.
Figure 4.3: **Rotational states of the V-ATPase.** (A) Cross sections through a 3D map of rotational state 1 of the V-ATPase V₁ region (i) and Vₒ region (ii) show that α-helices can be resolved. The best resolution in the Vₒ region is indicated with the pink arrow. (B) The three different rotational states each allow docking of atomic models of subunits. The experimental map is shown in grey. The blue ‘*’ indicates the central rotor. Scale bar, 25 Å.
Figure 4.4: **Map segmentation and molecular dynamics flexible fitting.** Different subunits are shown fitted into their corresponding map densities in rotational state 1, 2, and 3, including AB pair 3 (A), the N-terminal domain of the a-subunit (B), the central rotor DFd subcomplex (C), subunit C (D), and peripheral stalk 1 (E). Scale bar, 25 Å.
region (Figures 4.3a-ii and 4.2c). In the V_{0} region, resolution is best where the a subunit interacts with the c-ring (Figure 4.3a-ii, pink arrow), suggesting a rigid interaction within an otherwise dynamic complex. Crystal structures and homology models for V-ATPase subunits were docked into the three different maps and their conformations were refined by molecular dynamics flexible fitting [89] (Figure 4.3b; Figure 4.4). The most striking structural difference between the three maps is the position of the central rotor, consisting of subunits D, F, d, and the c-ring (blue * in Figure 4.3b).

### 4.6.2 Stoichiometry of the c-ring

The proton-carrying c-ring of the *S. cerevisiae* V-ATPase is a hetero-oligomer of c-, c′-, and c′′-subunits, each possessing a single conserved glutamate residue (Glu137, Glu145, Glu108, respectively) that can bind and transport protons during catalysis [73]. The c- and c′-subunits have four transmembrane α-helices each, with an additional N-terminal α-helix for the c′′-subunit that may not be membrane bound and is not necessary for function [127]. The cryo-EM maps show the c-ring to consist of an inner ring and an outer ring of transmembrane α-helices (Figure 4.3a-ii, Figure 4.5a and b magenta). The ring appears to have 20-fold symmetry (Figure 4.3a-ii), suggesting that the extra N-terminal α-helix of the c′′-subunit may protrude into the ring, as seen in a recent crystal structure of a bacterial heteromeric c-ring [112]. The structure of the c-ring in the cryo-EM map accommodates a total of 10 c-, c′-, or c′′-subunits (Figure 4.5a and b) with each subunit contributing two α-helices to the outer ring and two α-helices to the inner ring. The modeled c-subunits fit the density with their N and C termini facing the luminal side of the membrane [126]. The presence of 10 subunits in the c-ring is inconsistent with an earlier prediction of a 4:1:1
Figure 4.5: The membrane-bound $V_o$ region. (A) The c-ring segment from the cryo-EM map accommodates 10 c-, c$'$-, or c$''$-subunits (magenta). (B) The c-ring sits adjacent to the a-subunit (green). (C) The a subunit contains two long and highly-tilted $\alpha$-helices (red arrows). The conserved Glu residues of the c-subunits lie between the two tilted $\alpha$-helices of the a-subunit. In the eukaryotic V-ATPase c-ring, every other outer $\alpha$-helix has a Glu residue (blue circles). Gray lines indicate the approximate membrane position. (D) The tilted $\alpha$-helices of the a-subunit (green) contact 3 different $\alpha$-helices of the c-ring (magenta). Scale bars, 25 Å.
stoichiometry for subunits c, c', and c'' that would require a total of 6 or 12 subunits in the ring [142, 51]. A result of the existence of 10 subunits in the c-ring is that complete rotation of the c-ring driven by hydrolysis of 3 ATP molecules would deliver 10 protons across the lipid bilayer. Consequently, the c-ring stoichiometry suggests the ATP:H\(^+\) ratio is 3:10 for the *S. cerevisiae* V-ATPase, which is the same ratio as the F-type ATP synthase in *S. cerevisiae* mitochondria [176]. The ATP:H\(^+\) ratio was not known previously for any eukaryotic V-ATPase. With a Gibbs free energy for ATP hydrolysis of 57 kJ/mol at 30 °C, this ATP:H\(^+\) ratio limits the maximum pH gradient or voltage established by V-ATPase across a vacuolar membrane in *S. cerevisiae* to 3.0 units or 180 mV [125].

### 4.6.3 Structure of the a-subunit

The three maps provide the highest-resolution insight available into the structure of the membrane-bound portion of the a-subunit, which contains the channels that conduct protons to and from the c-ring. The a-subunit has been predicted to have eight transmembrane \(\alpha\)-helices [179]. The density corresponding to the a-subunit has a complex fold and appears to have at least eight transmembrane \(\alpha\)-helices and a number of well-defined structural elements above and below the expected position of the lipid bilayer. Strikingly, there are two highly-tilted \(\alpha\)-helices from the a-subunit that span the lipid bilayer where the a-subunit is in contact with the c-ring (Figure 4.5c, red arrows). These \(\alpha\)-helices contact a group of three \(\alpha\)-helices from the c-ring (Figure 4.5c and d). In F-type ATP synthases and V/A-ATPases, each outer \(\alpha\)-helix of the c-ring has a conserved proton-carrying Glu or Asp residue. In the eukaryotic V-ATPase, every other \(\alpha\)-helix on the surface of the c-ring has the conserved Glu residue. At the present resolution, it is not possible to distinguish the outer \(\alpha\)-helices
Figure 4.6: C-terminal domain of the proton-translocating a-subunit. (A) The membrane-bound C-terminal domain of the a-subunit appears similar in all three rotational states. (B) The C-terminal domain of subunit I, the *Thermus thermophilus* equivalent of the a-subunit, at 9.7 Å resolution is consistent with the structure of the a-subunit from *S. cerevisiae* (i). However, the transmembrane α-helical densities identified in that map (ii) are inconsistent with the current maps.
with the conserved Glu residues from the outer $\alpha$-helices lacking the Glu residues. Therefore, it is possible to fit the c-ring in the EM map in two different ways: one where a single Glu residue contacts the a-subunit and one where two different Glu residues contact the a-subunit. This latter arrangement would place two different c-subunits (Figure 4.5c, circled in blue) in different chemical environments, enabling one c-subunit to exchange protons with the cytoplasm while the other exchanges protons with the organelle lumen. This arrangement could create two half-channels for proton translocation [80] and was seen with the *T. thermophilus* enzyme [93]. The $\alpha$-helices in the a-subunit are much easier to identify in the present maps at 6.9-8.3 Å resolution than in the 9.7 Å resolution map of the *T. thermophilus* V/A-ATPase determined earlier [93]. The $\alpha$-helices identified here in all three states (Figure 4.6a) are consistent with the density from the earlier map of the bacterial enzyme (Figure 4.6b-i), but are inconsistent with the locations of $\alpha$-helices proposed previously based on that density (Figure 4.6b-ii).

### 4.6.4 Rotation of the central rotor between states

With three catalytic nucleotide-binding sites, the $V_1$ region is expected to operate as a 3-step motor while the 10 titratable Glu residues of the c-ring suggest that the $V_O$ region functions as a 10-step motor. This 3:10 ratio produces a symmetry mismatch between the $V_1$ and $V_O$ regions. The new maps demonstrate how symmetry mismatch can be tolerated in rotary ATPases. The rotational position of the D-subunit in the central rotor was measured relative to the $A_3B_3$ hexamer (Figure 4.7 upper, blue density and line), which itself twists within the rest of the enzyme between the three different states (Figure 4.7, black lines). The rotation of the D-subunit corresponds to steps of 117° from state 3 to 1, 119° from state 1 to 2, and 124° from state 2 to
3, all of which are in good agreement with the 120° steps expected for the three-fold symmetric V₁ region. In the V₀ region, rotation of one c-subunit against the a-subunit requires a 36° rotation of the c-ring (360°/10). The rotational position of the d-subunit and c-ring relative to the a-subunit could be measured precisely in rotational states 1 and 2 due to the resolution in the membrane region of these maps. In rotational state 3, the rotation of the c-ring against the a-subunit could not be measured with the same confidence. For the transition from state 1 to 2, we measured a 139° rotation of the d-subunit and c-ring relative to the transmembrane portion of the a-subunit (Figure 4.7 lower, cyan density and line). This rotation from state 1 to 2 matches the 144° rotation (36×4) expected from a 4/10ths rotation of the c-ring and would be expected to deliver 4 protons across the lipid bilayer. The transitions from state 2 to 3 and 3 to 1 were measured at ∼101° and ∼120°, respectively. Together, these rotations match ∼216° of rotation (36°×6), and are likely due to 108° (36°×3) of rotation for each transition. As a consequence, it appears that transitions from state 1 to 2, 2 to 3, and 3 to 1 transport 4, 3, and 3 protons, respectively. The observed rotational states reveal the conformations of the enzyme in vitro after the available ATP in solution has been hydrolyzed. Therefore, it is possible that the conformations of the enzyme when hydrolyzing ATP rapidly could be different than the conformations observed here. For example, when hydrolyzing ATP rapidly, the stepping motions of the enzyme may average to 3.3 protons transported for each ATP hydrolysis event. It is also possible that some slip occurs during proton translocation and less than 1 proton is transported for each c-subunit on average.
Figure 4.7: **Symmetry mismatch between the V₁ and V₀ regions.** *Upper,* the central rotor D subunit (dark blue) makes \( \sim 120^\circ \) steps (dashed blue line to solid blue line) from state 3 to 1 (A), 1 to 2 (B), and 2 to 3 (C) relative to the rest of the V₁ region, which itself undergoes a slight twist relative to the V₀ region (black lines). *Lower,* in the V₀ region, the d-subunit and c-ring undergo rotations of \( \sim 120^\circ \), \( \sim 139^\circ \), and \( \sim 101^\circ \) from state 3 to 1 (A), 1 to 2 (B), and 2 to 3 (C), respectively (dashed cyan line to solid cyan line). Scale bar, 25 Å.
Figure 4.8: **Flexibility of V-ATPase subunits.** (A) Comparison of AB pair 2 from the rotational states (state 1 in colour, state 2 and 3 in grey) shows that the pair progresses through ‘open’, ‘tight’, and ‘loose’ conformations. (B-E) Overlay of structures from the rotational states illustrates that during rotation the d-subunit wobbles relative to the D- and F-subunits in the central rotor (B), the peripheral stalks bend (C), the N-terminal domain of the a-subunit pivots towards the central rotor (D), and the ‘head’ domain of the C subunit twists relative to the ‘neck’ (E). These movements are most apparent in Movie 3. Scale bar, 25 Å.
Chapter 4. Rotational states of the V-ATPase
Figure 4.9: **Flexibility in catalytic subunits and peripheral stalks.** (A-C), Each AB pair in the A$_3$B$_3$ hexamer goes through the ‘open’, ‘loose’, and ‘tight’ conformations as the enzyme passes between the three rotational states. (D-F), Overlay of all three ‘open’, all three ‘loose’, and all three ‘tight’ structures shows that the conformations are similar for each AB pair. (G-I), Each of the three EG peripheral stalk structures undergoes similar bending motions between the three rotational states. Scale bar, 25 Å.
4.6.5 Conformational changes in A- and B-subunits

Overlaying the different conformations of various subunits in the complex suggests some of the structural changes that may occur during rotary catalysis (Figure 4.8). The conformational changes are illustrated dramatically by interpolating between the three rotational state structures (Movie 2). The different conformations of the catalytic subunits of rotary ATPases have been reported from mitochondrial and bacterial F$_1$-ATPases [1, 33] and bacterial V/A$_1$-ATPases [8, 129] but never before for a eukaryotic V-type enzyme and never before within an intact rotary ATPase. Different from crystal structures of isolated V/A-ATPase or F$_1$-ATPase subcomplexes, the availability of structures of the intact enzyme in different rotational states enables comparison of ‘open’, ‘tight’, and ‘loose’ conformations of the AB pairs when the rotor is in different positions. The observed conformations of the A$_3$B$_3$ hexamer (Figure 4.8a, Figure 4.9a-c) reveal a bend of the C-terminal domain of the A-subunit that closely resembles the conformational changes seen in the E. hirae V/A$_1$-ATPase [8] and mammalian mitochondrial F$_1$-ATPase [1] rather than the near rigid movement of subunits in the T. thermophilus V/A$_1$-ATPase [129]. The equivalent ‘open’, ‘tight’, and ‘loose’ conformations from different AB pairs can be overlaid with near perfect fidelity (Figure 4.9d-f). The protein samples used to prepare cryo-EM grids were not supplemented with nucleotide and therefore the nucleotide occupancy of the different catalytic sites is unknown. As a consequence, it is possible that the observed states correspond to the intrinsic asymmetry seen in AB pairs and αβ pairs of the E. hirae and S. cerevisiae V/A$_1$- and F$_1$-ATPase crystal structures lacking nucleotide, which resemble closely the conformations of the enzymes with bound nucleotide [8, 81].
4.6.6 V-ATPase subunits are flexible

Rotary ATPases have been proposed to have an elastic coupling between their catalytic and membrane-bound regions to smooth the transmission of power between ATP hydrolysis or synthesis and proton translocation [133, 175, 205]. This need for elastic coupling is exacerbated by the 3:10 symmetry mismatch of the V₁ and Vₒ regions: the enzyme must deform to allow the rotor to be in the correct position simultaneously in the catalytic V₁ region and the membrane-bound Vₒ region. Earlier studies have suggested that the central rotor of the Escherichia coli ATP synthase is the compliant element in that enzyme, while the peripheral stalk remains rigid [189]. The current structures indicate that the extended α-helical part of the central rotor D-subunit, equivalent to the F-type ATP synthase γ subunit, remains rigid during rotation while the part of the D-subunit in contact with the d-subunit bends (Figure 4.8b, dark blue, Movies 2 and 3). Further, the orientation of the d-subunit changes with respect to the D- and F-subunits, wobbling in order to accommodate distortion of the enzyme during rotation (Figure 4.8b, cyan; Movies 2 and 3). The catalytic A- and B-subunits push against the E- and G-subunits of the peripheral stalks during rotation. Due to this movement of the A- and B-subunits, the E- and G-subunits of the peripheral stalks undergo a bending motion along their elongated coiled-coil region, reminiscent of the action of a cantilever spring (Figures 4.8c and 4.8g-i, Movies 2 and 3). The N-terminal domain of the a-subunit swings parallel to the membrane, moving to and away from the rotation axis of the rotor like the arm of a record player. The bend in the a-subunit occurs at the narrow interface between its N- and C-terminal domains (Figure 4.8d, Movies 2 and 3). In comparison, the head domain of the C-subunit twists like a torsion spring at the neck domain and thereby maintains its connections between peripheral stalks 2 and 3 (Figure 4.8e, Movies 2 and 3). The
existence of these conformational changes is not obvious from inspection of crystal structures of the individual subunits. However, when visualized as a movie, each subunit appears to have evolved to carry out these motions. Overall, the structures presented here show that in the V-ATPase, both the rotor and stator part of the engine undergo coordinated conformational changes. This combination of flexibility and rigidity may explain the high efficiency of the rotary ATPase macromolecular machines.
Chapter 5

Inhibition of V-ATPase by SidK

5.1 Author contributions

Stephanie A. Bueler\(^1\) (SAB), Zhao-Quin Luo\(^4\) (ZQL), and John L. Rubinstein\(^1,2,5\) (JLR) initiated the project. Jianhua Zhao\(^1,2\) (JZ) and JLR designed the study and JZ performed the experiments. Claudia Alvarez\(^3\) (CA) and Voula Kanelis\(^3\) (VK) guided the fluorescence titration work. JZ and JLR analyzed and interpreted the data. JZ and JLR wrote the manuscript and JZ prepared the figures.

5.2 Acknowledgements

We thank Michael Runswick and John E. Walker (MRC Mitochondrial Biology Unit, Cambridge) for samples of the bovine ATP synthase. This research was supported by operating grant MOP 81294 from the Canadian Institutes of Health Re-

\(^1\)Molecular Structure and Function Program, The Hospital for Sick Children Research Institute
\(^2\)Department of Medical Biophysics, University of Toronto
\(^3\)Department of Chemistry, University of Toronto
\(^4\)Department of Microbiology, Purdue University
\(^5\)Department of Biochemistry, University of Toronto
search (CIHR), awarded to JLR. JZ was funded by a Doctoral Postgraduate Scholarship from the Natural Sciences and Engineering Research Council of Canada (NSERC) and a Mary Gertrude l’Anson Scholarship from the University of Toronto.

5.3 Introduction

Structures of the eukaryotic V-ATPase in different rotational states were presented in Chapter 4, suggesting that flexibility of the protein complex may be important for its mechanism [201]. As a consequence, reducing the flexibility of the V-ATPase may inhibit the rotary ATPase, but evidence supporting this hypothesis is lacking. In the immune system, V-ATPase activity is modulated by host-pathogen interactions that cause infection and disease. Phagocytic white blood cells, such as macrophages, kill invading pathogens within intracellular compartments called phagolysosomes [105]. Phagolysosomes are acidified by V-ATPases, which in turn activates digestive enzymes important for the breakdown of macromolecules and microorganisms. To avoid the harsh environment of phagolysosomes, intracellular pathogens secrete effectors that prevent phagolysosome acidification or formation [196, 173, 195]. *Legionella pneumophila*, the causative agent of Legionnaires’ Disease [113, 55], secretes a protein called SidK that binds and inhibits V-ATPases [196]. The inhibition of V-ATPase promotes survival of bacteria within phagocytic white blood cells, leading to disease. The mechanism of V-ATPase inhibition by SidK is unknown. Here we determined structures of different V-ATPase:SidK complexes, revealing how SidK interacts with the $V_1$ region of the V-ATPase, the structure of SidK, and the effect of SidK on the flexibility and stability of the V-ATPase.
Chapter 5. Inhibition of V-ATPase by SidK

5.4 Materials and Methods

5.4.1 Protein purification

V-ATPase and V1 were purified from Saccharomyces cerevisiae as described previously in Chapter 4 [20] via a 3×FLAG tag on the A-subunits. The SidK gene was cloned into a pET28 plasmid containing an N-terminal 6×Histidine tag followed by a tobacco etch virus (TEV) cleavage site. BL21 Codon+ cells containing the SidK plasmid were grown at 37 °C with vigorous shaking (225 RPM) in 1-4 l of LB media supplemented with 0.4 % (w/v) glucose and 50 mg/l kanamycin. At OD600 ≈0.7, protein expression was induced with 1 mM isopropyl-D-1-thiogalactopyranoside (IPTG) and cells were grown overnight at 16 °C. All subsequent steps were performed at 4 °C. Cells were harvested by centrifugation at 5,000 g and lysed by sonication in TBS buffer (50 mM Tris-HCl, 0.3 M NaCl, pH 7.4) containing 0.001 % (w/v) phenylmethanesulfonylfluoride (PMSF). Cell lysate was centrifuged at 38,000 g and the supernatant was loaded onto a HisTrap Ni-NTA column (GE Healthcare). The HisTrap column was washed in TBS buffer containing 25 mM imidazole and SidK was eluted in TBS buffer containing 0.3 M imidazole. SidK was mixed with TEV protease in a 1:100 TEV:SidK ratio and dialyzed overnight against 2 l of TBS buffer containing 1 mM dithiothreitol (DTT). Cleaved protein was dialyzed against 2×1 l of TBS and loaded onto a HisTrap column. The HisTrap column was washed with TBS buffer containing 25 mM imidazole and the flowthrough and wash were collected. Fractions were pooled and exchanged into ion exchange buffer (50 mM Tris-HCl, pH 7.4, and 1 mM ethylenediaminetetraacetic acid [EDTA]) by concentration and dilution in a centrifuge concentrating device (EMD Millipore). SidK was loaded onto a HiTrap Q anion exchange column (GE Healthcare) and eluted with a gradient of 0 to 200 mM NaCl. Fractions containing SidK were pooled and exchanged into TBS buffer
containing 5 mM DTT by concentration and dilution in a centrifuge concentrating device (EMD Millipore).

5.4.2 V-ATPase:SidK preparation

To purify fully-bound V-ATPase:SidK assemblies, SidK from after the second His-Trap column was added to detergent-solubilized yeast vacuolar membranes and the protein complex was purified as described in section 4.5.2 [20]. To purify substoichiometric V-ATPase:SidK assemblies, SidK from after TEV cleavage (which may contain cleaved and uncleaved SidK) was added to detergent-solubilized yeast vacuolar membranes and the protein complex was purified as described previously (section 4.5.2) except with twice the amount of washing.

5.4.3 Purification of V₁

Steps were performed at 4 °C. The supernatant from the ultracentrifugation step in section 4.5.1 was loaded onto a column containing anti-FLAG M2 agarose beads (Sigma Life Sciences). V₁ was purified as described in section 4.5.2 except TBS buffer (50 mM Tris-HCl, 150 mM NaCl, pH 7.4) was used instead of DTBS buffer.

5.4.4 ATPase activity assay

ATPase activity assays were performed as described previously [20] in assay buffer (50 mM Tris-HCl, pH 8, 0.05 % [w/v] n-dodecyl-D-maltoside [DDM], 3 mM MgCl₂, 1 mM DTT, 0.2 mM NADH, 10 U pyruvate kinase, 25 U L-lactic dehydrogenase, 1 mM phosphoenolpyruvate, and 2 mM ATP). Proteins were incubated on ice until assayed. 160 µl reactions were performed at room temperature using a 96-well plate in a Spectramax M2 UV/visible light plate reader (Molecular Devices). Four replicates
were performed for enzyme with SidK or enzyme with buffer conditions using the same stock of freshly purified V-ATPase or frozen ATP synthase. The final concentrations of enzyme and SidK were $\sim 1$ nM and $\sim 30$ nM, respectively. The concentration of bafilomycin used was 6 $\mu$M.

### 5.4.5 V₁ stability assay

$220 \mu l$ of V₁ at $\sim 5 \mu$M was incubated at room temperature in reaction buffer (50 mM Tris-HCl, pH 7.4, 0.15 M NaCl, 2 mM DTT) with or without $\sim 50 \mu$M SidK. At time zero, 100 $\mu l$ of each sample was analyzed using a Superdex 200 10/300 GL size exclusion column (GE Healthcare). The samples were incubated for 26 h and then 100 $\mu l$ of each sample was analyzed again by size exclusion chromatography.

### 5.4.6 Fluorescence titration

Experiments were performed in a Quantamaster QM-80 spectrofluorometer (Pho- ton Technology International) operating at an excitation wavelength of 465 nm with a 4 nm slit width and emission was recorded from 485 to 600 nm with a slit width of 10 nm. Temperature was controlled by a Peltier-unit maintained at 10.0 °C. Reactions (0.5 ml) containing 0.25 $\mu$M V₁ with or without 2.5 $\mu$M SidK were carried out in assay buffer (25 mM Tris-HCl, pH 7.9, 0.15 M NaCl, 1 mM MgCl₂, and 0.5 mM DTT). For each titration reading, 20 $\mu l$ of the reaction sample was removed and replaced with 10 $\mu l$ of trinitrophenyl-ATP (TNP-ATP) at different concentrations and 10 $\mu l$ of 0.5 $\mu$M V₁ in 2× assay buffer with or without 5.0 $\mu$M SidK. Fluorescence curves were corrected by baseline subtraction using the curve for [TNP-ATP]=0. The increase in fluorescence was modelled in GNUplot (www.gnuplot.info) with the fit command
using the equation [104]

\[
I = \frac{[ES]}{[ET]} \cdot I_0 + m([S_T] - [ES])
\] (5.1)

where \(I\) is the fluorescence intensity, \(I_0\) is the fluorescence intensity for all the enzyme bound to substrate, \([ET]\) is the total enzyme concentration, \([S_T]\) is the total substrate concentration, \(m\) is the increase in fluorescence per unit increase in free substrate, and \([ES]\) is the concentration of enzyme bound to substrate given by:

\[
[ES] = 0.5([ET] + [S_T] + K_d - \sqrt{([ET] + [S_T] + K_d)^2 - 4[ET][S_T]})
\] (5.2)

where \(K_d\) is the equilibrium dissociation constant given by \(K_d = [E][S]/[ES]\). Five repetitions were performed for each condition (with and without SidK) using the same batch of freshly purified V\(_1\) subcomplex.

### 5.4.7 Cryo-EM imaging and image processing

Imaging and image analysis procedures were similar to those used in the study of V-ATPase (Chapter 4). 3 \(\mu\)l of 10 mg/ml V-ATPase:SidK sample was applied to nanofabricated holey carbon grids [110] that were glow discharged in air for 2 min. Excess sample was blotted away and the grid was plunge-frozen into a liquid propane-ethane mixture using a Vitrobot Mark III (FEI company). Vitrified samples were imaged with a FEI Tecnai TF20 electron microscope operating at 200 kV and 34,483\(\times\) magnification, resulting in an approximate pixel size of 1.45 Å. Images were collected with a Gatan K2 Summit direct detector device operating in counting mode. 15 s movies were recorded at 2 frames/s and \(\sim2.4\) e\(^-\)/Å\(^2\)/s. Movie frames were aligned using \texttt{alignframes.lmbfgs} [154] and averaged with \texttt{shiftframes}. 
CTF parameters from the averaged images were calculated with *CTFFIND3* [117] and corrected for magnification anisotropy using *star_fixmaganiso* [203]. Candidate particle image coordinates were identified automatically in the averaged images using *TMaCS* [202] and 2D projections of an existing map of the V-ATPase that was low-pass filtered to 20 Å as templates [20]. Candidate particle images were extracted from the raw movies and corrected for local drift with *alignparts_lmbfgs* [154]. The aligned and averaged particle images were corrected for magnification anisotropy with *correctmaganisotropy_fspace_list* [203] and processed by 2D and 3D classification and 3D refinement in *Relion* [162, 161]. Maps were visualized and segmented in *UCSF Chimera* [139]. Atomic models of the V-ATPase (PDB 3J9T, 3J9U, 3J9V) were fitted into the density maps using *UCSF Chimera* and *Molecular Dynamics Flexible Fitting (MDFF)* [182].

### 5.4.8 Bioinformatics

The secondary structure of SidK was predicted with the online server *JPred* [47]. Proteins with amino acid sequences similar to SidK were identified with the *Basic Local Alignment Search Tool (BLAST)* [5]. Sequences were aligned using *Clustal Omega* [165].

### 5.5 Results

#### 5.5.1 SidK interacts with the $V_1$ region of V-ATPase

V-ATPase was purified from *S. cerevisiae* and combined with SidK that was expressed heterologously in *E. coli* (Figure 5.1A). The resulting V-ATPase:SidK complexes were imaged by cryo-EM (Figure 5.1B-C). Classification of the particle images
Figure 5.1: **Protein purification and cryo-EM imaging.** (A) Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis showing binding of SidK to V-ATPase and the V$_1$ subcomplex. (B) Micrograph of V-ATPase:SidK complexes embedded in vitrified buffer. Examples of V-ATPase:SidK assemblies are indicated (red boxes). Scale bar, 30 nm. (C) Trajectories of individual image features from micrograph in (B) calculated using alignparts_lmbfgs [154]. (D) Representative class averages of V-ATPase:SidK complexes from 2D classification. Scale bar, 25 Å.
Chapter 5. Inhibition of V-ATPase by SidK

A 1,528 movies ('fully-bound')
   Template matching
   305,600 candidate particle images
   2D classification
   98,183 particle images
   3D classification

B 3,065 movies ('substoichiometric')
   Template matching
   739,350 candidate particle images
   2D classification
   221,328 particle images
   3D classification

C

Occupancy of (SidK_{x1}, SidK_{x2}, SidK_{x3})

Rotational state 1
- 5,861 images (90:108:2)
- 9,313 images (108:13:89)
- 18,832 images (71:14:4)
- 8,305 images (0:40:85)

Rotational state 2
- 7,893 images (64:108:31)
- 5,360 images (103:61:110)
- 2,709 images (22:116:125)
- 2,438 images (100:5:1)
- 9,751 images (10:10:77)

Rotational state 3
- 6,382 images (105:91:11)
- 5,915 images (92:23:102)
- 2,938 images (35:114:100)
- 7,357 images (15:91:1)
- 5,735 images (91:1:9)
- 5,774 images (14:17:88)
Figure 5.2: Image processing workflow and 3D classification scheme. Computational processing for images of fully-bound (A) and substoichiometric (B) V-ATPase:SidK complexes. Three distinct rotational states of V-ATPase were identified (states 1-3). Further classification allowed separation of V-ATPase:SidK assemblies in different SidK binding configurations (C). The numbers in brackets (x:y:z) correspond to the average density values of SidK relative to V-ATPase with x, y, and z denoting the density of SidK bound to A_1, A_2, and A_3, respectively, where the V-ATPase density is 100 and background is zero. Some classes show a mix of complexes with different SidK stoichiometries. Scale bar, 25 Å.
Chapter 5. Inhibition of V-ATPase by SidK

Figure 5.3: Cryo-EM maps of V-ATPase:SidK. (A) Cryo-EM maps of V-ATPase:SidK complexes with angular distributions of image orientations. Long red bars represent higher numbers of images while short blue bars represent lower numbers of images. Scale bars, 25 Å. (B) Fourier shell correlation (FSC) curves from gold standard refinement for masked maps. Based on the 0.143 FSC cutoff, the overall resolutions of the ‘fully-bound’ maps for states 1, 2, and 3 are 6.8, 7.9, and 7.6 Å, respectively, and the ‘substoichiometric’ maps for states 1, 2, and 3 are 6.8, 7.7, and 7.7 Å, respectively. These resolutions were the same when calculated from the ‘corrected’ FSCs in Relion [162, 161].
(Figures 5.1D and 5.2A) and refinement of density maps allowed identification of three distinct rotational states of the V-ATPase: state 1 (49% of images), state 2 (23% of images), and state 3 (28% of images). The maps had good orientational sampling (Figure 5.3A, ‘Fully-bound’) and overall resolutions of 6.8, 7.9, and 7.6 Å for states 1, 2, and 3 respectively (Figure 5.3B). This distribution of particle images differs from that of the V-ATPase without SidK bound, which had 47, 36, and 17% of images in state 1, 2, and 3, respectively (Chapter 4) [201]. It is possible that the presence of SidK may be affecting the performance of the classification algorithm, resulting in differences in particle image distributions between the different rotational states. To obtain a more objective comparison between the distributions of rotational states for the V-ATPase with and without SidK, 3D classification was performed with a mask around the V<sub>1</sub> region to focus image sorting based on the V<sub>1</sub> region alone. Classification with a mask around the V<sub>1</sub> region showed that SidK binding increases and decreases the relative proportion of V-ATPase complexes adopting states 1 and 2, respectively (Figure 5.4B). The V-ATPase:SidK maps show three SidK binding sites located in the V<sub>1</sub> region (Figure 5.5A, pink arrowheads). SidK interacts mainly with the N-terminal region of the A-subunits, supporting previous biochemical analysis of SidK binding partners [196]. Comparison of V-ATPase:SidK and V-ATPase [201] structures revealed no major conformational changes in the V-ATPase with SidK binding (Figure 5.4C-E).

### 5.5.2 SidK contains a putative α-helical domain and a flexible domain

The cryo-EM maps of V-ATPase:SidK offer the first available insight into the structure of SidK (Figure 5.5B). The protein is elongated and appears to be com-
Figure 5.4: Local resolution, particle image distribution, and comparison of V-ATPase with and without SidK bound. (A) Local resolution estimates for maps of V-ATPase and V-ATPase:SidK. The flexible region of the SidK proteins give low resolution estimates. Similarly, the C-terminal region of the A-subunit in the ‘open’ conformation (A\textsubscript{open}) is at lower resolution in the V-ATPase map compared to the V-ATPase:SidK map, indicating flexibility in this region. (B) Classification of images with a mask around the V\textsubscript{1} region shows that SidK binding increases and decreases the proportion of V-ATPase complexes adopting states 1 and 2, respectively. Three separate datasets were processed independently. Error bars represent one standard deviation. (C-E) Overlay of pseudo-atomic models of V-ATPase fitted into density maps of V-ATPase (\textit{blue}) and V-ATPase:SidK (\textit{pink}). Density for the a-subunit is shown as the transparent gray surface. No major conformational changes in the V-ATPase are observed. Scale bars, 25 Å.
Figure 5.5: **Structure of the V-ATPase:SidK complex and SidK.** (A) Three SidK proteins (gray surface) are bound to the V-ATPase in the soluble catalytic V\textsubscript{1} region. The SidK density is well-defined in the region where it binds the V-ATPase (pink arrowheads) and becomes blurred and has lower intensity values in the region closer to the membrane-embedded part of the complex (pink dashed circles). (B) SidK is composed of a putative α-helical region (pink arrowhead) and a flexible region (pink dashed circle). For clarity, the surface rendering of SidK is shown at a higher density threshold for the putative α-helical region (Thres. 0.04) than for the flexible region (Thres. 0.015). SidK interacts mainly with the A-subunit through a large contact area (yellow surface) composed of three α-helices (dark blue arrowheads). A structural feature resembling a 'hook' penetrates the interface between an A-subunit and an adjacent B-subunit. Scale bar, 25 Å.
Figure 5.6: **Secondary structure prediction of SidK from* L. pneumophila.**

The amino acid sequence (top row) is aligned with the predicted secondary structure element (bottom row). 'H' and 'E' represent α-helix and β-sheet, respectively. The protein is predicted to be mostly α-helical.
posed of at least two domains: a structured region that binds the V-ATPase and a flexible region that does not appear to interact with the enzyme assembly. The region of SidK that binds the V-ATPase contains putative α-helical densities. This observation is consistent with the prediction that SidK is composed mainly of α-helices (Figure 5.6). N-terminal truncations affected SidK function more significantly than C-terminal truncations [196] (Figure 5.10A), suggesting that the V-ATPase binding domain of SidK corresponds likely to the N-terminal region of the protein. This region contains a structural feature resembling a ‘hook’ that penetrates the non-catalytic interface between two AB heterodimers. Extensive contacts between SidK and the A-subunits occur through three putative α-helices in SidK (Figure 5.5B, dark blue arrowheads). The density corresponding to the region of SidK that is not involved with V-ATPase binding appears blurred and has lower intensity values, suggesting that the two SidK domains are connected by a flexible linker. This flexible domain is likely the C-terminal region of SidK (see above) and truncating a significant portion of this region did not affect its activity significantly (Figure 5.10A). These structural and biochemical data indicate that the C-terminal region of SidK is not involved in V-ATPase inhibition. The function of the flexible domain remains unclear. Multiple sequence alignment of SidK with L. pneumophila effectors secreted by the Dot/Icm type IV secretion system [32] show conserved residues in the C-terminal regions of the proteins (Figure 5.7), suggesting that part of the flexible domain may be involved in protein secretion. A C-terminal secretion signal is a consistent feature in Legionella effectors, but the exact signal sequences vary [148].
Figure 5.7: **Multiple sequence alignment between SidK and other *Legionella* effectors.** SidKLp is SidK from *L. pneumophila* (NCBI WP_010946703.1). hypoLM is a putative homolog of SidK from *Legionella moravica* (NCBI WP_028384931.1). LegC5 (NCBI WP_014843838.1), doticmLp1(NCBI WP_040147294.1), doticmLp2 (NCBI WP_014842287.1), and doticmLfl (GenBank CEG57703.1) are effectors secreted by the *Legionella* Dot/Icm secretion system. Completely conserved (*red*), highly conserved (*green*), and somewhat conserved (*blue*) residues are indicated. Conserved residues are concentrated in the C-terminal part of the amino acid sequences.
Figure 5.8: **SidK reduces the flexibility of the A-subunit in the ‘open’ conformation.** (A) Slices through cryo-EM maps of V-ATPase and V-ATPase:SidK in state 3. The maps were low-pass filtered to 10 Å. Density corresponding to the C-terminal region of the A-subunit (A$_{CT}$) in the ‘open’ conformation is blurred and has lower intensity compared to other parts of the V$_1$ region (‘-SidK’). Density for the ‘open’ A$_{CT}$ is comparable to other subunits in the V-ATPase:SidK map (‘+SidK’). (B) Comparison between segmented A-subunit densities in the ‘open’ conformation from V-ATPase (gray surface) and V-ATPase:SidK (gray mesh) cryo-EM maps. A$_{CT}$ is better defined with SidK bound. Scale bar, 25 Å. (C) Relative density values measured between A$_{CT}$ and the N-terminal region of the A-subunit (A$_{NT}$) in the ‘open’ conformation. There is an increase in the relative density of A$_{CT}$ upon SidK binding in all rotational states. Three separate datasets were used for the analysis. Error bars represent one standard deviation. Scale bars, 25 Å.
5.5.3 SidK reduces the flexibility of the A-subunit in the ‘open’ conformation

Subunit flexibility serves an important role in the catalytic mechanisms of the V-ATPase and ATP synthase rotary enzymes [201, 204]. Examination of previous cryo-EM maps of the V-ATPase without SidK bound reveals that the density corresponding to the C-terminal region of the A-subunit in the ‘open’ conformation is blurred and has lower intensity values relative to other A-subunits (Figure 5.8), indicating flexibility in this protein. In comparison, the C-terminal region of the ‘open’ A-subunit in the V-ATPase:SidK cryo-EM map is well defined, indicating a more rigid structure when SidK is bound. The ‘open’ A-subunit is poised to bind nucleotide and flexibility in this subunit may be important for its function. Furthermore, calculation of the local resolution of the density maps shows higher relative resolution estimates for the ‘open’ A-subunit C-terminal region in V-ATPase:SidK maps compared to maps of V-ATPase alone (Fig. 5.4A). These data indicate that SidK binding leads to reduced flexibility of the A-subunit in the ‘open’ conformation, which could interfere with the conformational changes required for rotary catalysis.

5.5.4 SidK binds A-subunit in ‘tight’ conformation preferentially

Three SidK binding sites in the V-ATPase with three rotational states suggests 24 possible configurations: three arrangements with three SidK bound, nine arrangements with two SidK bound, nine arrangements with one SidK bound, and three arrangements with no SidK bound. To investigate whether SidK binds cooperatively to the A-subunits such that binding to one A-subunit in a particular conformation increases binding of SidK to another A-subunit in a different conformation, V-
Figure 5.9: **SidK preferentially binds the A-subunit in the ‘tight’ conformation.** (A) Slices through the cryo-EM maps of V-ATPase:SidK with substoichiometrically-bound SidK. Density corresponding to SidK (blue, orange, pink triangles) were compared to the density of the A-subunit in the ‘tight’ conformation (yellow triangle). The average density values of SidK (blue, orange, pink numbers), measured relative to the average density values of the A-subunit in the ‘tight’ conformation (yellow numbers), were highest in SidK proteins bound to the ‘tight’ A-subunit for all rotational states. Scale bar, 25 Å. (B) Quantification of results from (A) indicate that SidK binds the A-subunit preferentially in the ‘tight’ conformation. Three separate datasets were used for the analysis. Error bars represent one standard deviation.
ATPase:SidK complexes with substoichiometrically-bound SidK were analyzed. Cryo-EM imaging and analysis of the substoichiometric V-ATPase:SidK assemblies allowed computational separation of different complexes (Figure 5.2B-C). Due to image noise and the finite dataset size, the computational sorting procedure resulted in density maps representing a mix of different structures in some cases. Almost all of the V-ATPase:SidK binding configurations were identified, suggesting that SidK does not bind in a cooperative manner to the three A-subunits of the V-ATPase. To examine whether SidK has a preference for binding any particular A-subunit of the V-ATPase, images of substoichiometric V-ATPase:SidK complexes were separated based on the rotational state of the V-ATPase and combined to build a map for each of the three states (Figures 5.2B and 5.3). Occupancies of the SidK proteins were estimated in the maps by comparing the average density values of well-defined α-helices in SidK with those of the A-subunit in the ‘tight’ conformation (Figure 5.9). The occupancy was highest for SidK bound to the ‘tight’ A-subunit in all V-ATPase rotational states. This preference of SidK for binding to the A-subunit in the ‘tight’ conformation could bias the A-subunit to remain in this conformation, reducing rotary catalysis in the V-ATPase.

5.5.5 SidK decreases the affinity of the $V_1$ subcomplex for ATP and stabilizes $V_1$

The binding of SidK to the A-subunit and the effect of SidK on the flexibility of this subunit suggests SidK may affect the binding of nucleotide to the catalytic AB heterodimers. To investigate the effect of SidK on the affinity of V-ATPase for ATP, the soluble $V_1$ subcomplex was purified from $S. cerevisiae$. As expected, the purified $V_1$ subcomplex had no detectable ATPase activity. However, the subcomplex was
still able to bind SidK (Figure 5.1A) and nucleotide (Figure 5.10C-D), suggesting that the catalytic $A_3B_3$ hexamer of the $V_1$ subcomplex remains in a conformation similar to the intact V-ATPase assembly and the auto-inhibition of the $V_1$ region does not interfere with accessibility of the active site. The fluorescent ATP analog trinitrophenyl-ATP (TNP-ATP) was titrated into a solution containing the purified $V_1$ assembly and fluorescence was measured (Figure 5.10C). The increase in fluorescence upon TNP-ATP binding was modeled to calculate the equilibrium dissociation constant $K_d$ (Figure 5.10D). The calculated $K_d$ of $V_1$:SidK/TNP-ATP (280 +/- 60 nM) was almost two-fold higher than the $K_d$ of $V_1$/TNP-ATP (160 +/- 50 nM), indicating that SidK decreases the affinity of the V-ATPase soluble catalytic region for TNP-ATP. Considering the ATP concentration in cells is on the order of a few mM [19], the change in ATP affinity in V-ATPase upon SidK binding will likely not affect the binding of ATP to V-ATPase \textit{in vivo}. However, the decrease in ATP affinity suggests that SidK is interfering with rotary catalysis in the V-ATPase.

The apparent increase in rigidity of the A-subunit with SidK binding suggested that SidK may be affecting the stability of the V-ATPase catalytic region. To investigate the effect of SidK on the stability of the $V_1$ subcomplex, the $V_1$ assembly was incubated with and without SidK for 26 hours at room temperature. Analysis by size exclusion chromatography showed a ~20 % decrease in $V_1$ subcomplexes and an increase in the population of lower molecular weight species in the absence of SidK (Figure 5.10E), indicating that the protein assemblies are dissociating over time. Treatment with SidK prevents the decrease in $V_1$ assemblies, indicating that SidK is able to stabilize the $V_1$ subcomplex.
Chapter 5. Inhibition of V-ATPase by SidK

![Graphs and charts illustrating the inhibition of V-ATPase by SidK.](image-url)
Figure 5.10: **SidK inhibits specifically V-ATPases, decreases the binding affinity of V-ATPase for nucleotide, and stabilizes the V₁ subcomplex.**

(A) ATPase activity assays with full-length SidK (‘SidK1-569’) decreased the ATPase activity of V-ATPase (‘V₁V₀’) by ~40 %. SidK with a C-terminal truncation (‘SidK10-414’) still inhibited V-ATPase activity. Bafilomycin (‘Baf.’) abolished V-ATPase activity. (B) SidK did not affect the ATPase activity of bovine ATP synthase. (C) Relative fluorescence values and fitted curves from titration of TNP-ATP into a solution containing V₁ with or without SidK. (D) Dissociation constant (Kₐ) values calculated from data in (C) show increased Kₐ with SidK treatment, indicating that SidK binding decreases the affinity of the V₁ subcomplex for ATP. (E) A₂₈₀ absorbance profiles from size exclusion chromatography show that the V₁ subcomplex dissociates over time and that SidK can prevent V₁ dissociation.
5.6 Discussion

Structural flexibility appears to be an important feature of rotary ATPases [201, 204] but it remains unclear how this flexibility is related to the activity of these enzymes. Comparison between the structures of the V-ATPase and V-ATPase:SidK complexes indicates that SidK binding leads to reduced flexibility of the A-subunit in the ‘open’ conformation, providing a possible explanation for the inhibition of V-ATPase by SidK. The reduced flexibility of the A-subunit suggests a model for V-ATPase inhibition that involves trapping of V-ATPase in a particular conformation to prevent rotary catalysis. Several lines of evidence support this model. First, conformational changes in the A-subunit are important for rotary catalysis (Chapter 4) [201] and reduced flexibility in the A-subunit may hinder those conformational changes. Second, SidK binds preferentially to the A-subunit in the ‘tight’ conformation, showing conformational selectivity. Third, SidK prevents dissociation of the V₁ subcomplex, indicating that SidK has a stabilizing effect on the catalytic region of V-ATPase. SidK also decreases the affinity of the V₁ region for TNP-ATP, further supporting a model where SidK interferes with rotary catalysis in V-ATPase.

Obstruction of rotary catalysis is known to occur in enzyme assemblies that are similar to V-ATPases, including inhibition of ATP hydrolysis in ATP synthases by inhibitory factor 1 (IF1) [26] and the ε-subunit in bacteria [33]. However unlike IF1 and the ε-subunits that interact with the C-terminal regions of the catalytic subunits, SidK binds to an interface on the A-subunit in the N-terminal region, far from the active site. This binding interface includes a non-conserved region in the A-subunit called the non-homologous region that is present in V-ATPases and V/A-ATPases but not in ATP synthases. Binding to this non-homologous region may explain in part the specificity of SidK for V-ATPases and suggests a novel site of V-ATPase
inhibition that was not known previously. V-ATPases are likely inhibited by other intracellular pathogens during infection and this study provides insight into how this inhibition might occur.
Chapter 6

Future Directions

6.1 V-ATPase mechanism

6.1.1 V-ATPase structure during ATP hydrolysis

The structural studies of the V-ATPase described in Chapters 4 and 5 were not performed in the presence of nucleotide in the sample buffers. As a consequence, the structures of the V-ATPase described in this thesis represent the protein assemblies in their ground-state and inactive conformations. It is possible that the structure or conformations of the V-ATPase differs under ATP hydrolysis conditions. In particular, it is unclear how the enzyme accommodates the strain produced from the opposing torques generated by ATP hydrolysis in the $V_1$ region and proton pumping against an electrochemical gradient in the $V_O$ region. To investigate the structure of the active enzyme in the absence of an electrochemical gradient, cryo-EM samples of V-ATPase could be prepared with added nucleotide. One foreseeable limitation may be the high concentrations of nucleotide required in the sample buffer, which could add significant background noise to cryo-EM images and hinder structure determina-
tion. One solution would be to cross-link the c-ring to the a-subunit as was done in previous studies [164, 189], which may allow visualization of a strained enzyme under lower ATP concentrations. Another solution would be to reconstitute the V-ATPase in liposomes and image the membrane-embedded V-ATPase. This system would offer the capability to study the V-ATPase in different lipids and under proton-pumping conditions against an electrochemical gradient. A limitation of this particular approach may be the lower number of useful particle images that would be available per micrograph, reducing data throughput.

6.1.2 Proton translocation

There is no high-resolution structure of the a-subunit, limiting our understanding of the structural basis for proton translocation. In particular, the path of protons through the a-subunit and the mechanism of proton selectivity remain unknown. An atomic structure of the a-subunit would answer many important questions concerning the mechanism of proton translocation. However, structure determination of the a-subunit to high resolution may prove difficult due to the flexibility of this complicated membrane protein (Chapter 4). Cryo-EM analysis of the $V_O$ region may provide a possible solution (section 6.1.3). An alternative approach was proposed recently where a structural model of the a-subunit in the bovine ATP synthase was derived based on cryo-EM structures and evolutionary covariance analysis [204]. Through a similar analysis, structural models of the a-subunit from the $T. thermophilus$ V/A-ATPase and $S. cerevisiae$ V-ATPase have been generated (Schep, Zhao, and Rubinstein, submitted). These models provide important insight into the architecture of the membrane-embedded a-subunit of rotary ATPases. Although these models do not provide information on the position of amino acid side chains, they are consis-
tent with numerous biochemical studies conducted previously on specific a-subunit residues. Furthermore, they provide a foundation for future studies to probe the structure and function of this important V-ATPase subunit.

### 6.1.3 $V_O$ inhibition

Upon dissociation of the $V_1$ region from the $V_O$ region, proton translocation by the $V_O$ region is inhibited by an unknown mechanism. A recent low-resolution cryo-EM map of the $V_O$ region shows that inhibition of the $V_O$ region involves interaction between the d-subunit and the N-terminal region of the a-subunit [40]. However, the resolution of the map did not allow identification of the amino acid residues involved in the interaction. Ongoing work in the Rubinstein laboratory aims to elucidate the atomic structure of the $V_O$ region by making use of recent advances in cryo-EM imaging technology and methodology. Flexibility in the protein complexes, which can hinder structure determination to high resolution, may be removed by binding of small molecule inhibitors that lock the $V_O$ region into more stable conformations. These studies will provide important insight into the inhibition of the $V_O$ region and may reveal the mechanism of proton translocation across this membrane-embedded complex.

### 6.2 SidK mechanism

#### 6.2.1 Atomic structure of SidK

There is no high resolution structure of SidK, limiting the understanding of how SidK binds and inhibits the V-ATPase. The size of SidK (~65 kDa) and its high solubility and purification yield may make it amenable to structure determination by
X-ray crystallography. Attempts to crystallize SidK were successful, but crystals that
diffract to high resolution were not obtained (Appendix A). The flexibility of SidK
may be hindering the formation of a good crystal lattice. Therefore, crystallizing the
individual domains of SidK may prove more fruitful. Another strategy is to study
a homologue of SidK. One putative homologue of SidK is known, sequenced from
Legionella moravica with \(\sim 34\%\) sequence identity. This putative SidK homologue
may be amenable to crystallization and structure determination by X-ray crystallog-
raphy. An atomic structure of the putative homologue would provide insight into
the structure of SidK and the amino acid residues important for the V-ATPase:SidK
interaction.

\subsection*{6.2.2 Cryo-EM of V-ATPase:SidK}

Structure determination of the V-ATPase:SidK complex by cryo-EM may have
benefited from the extra mass contributed by the SidK proteins bound to the V-
ATPase. On the other hand, the flexible region of SidK could have hindered im-
age classification and refinement algorithms. Structure determination of the V-
ATPase:SidK complex may be improved by studying the protein assembly with a
truncated form of SidK where the flexible region is removed. Initial attempts at study-
ing the structure of V-ATPase in complex with a truncated form of SidK (SidK10-530)
were unsuccessful due to dissociation of the mutant SidK from the V-ATPase during
sample preparation. It is possible that SidK10-530 is a weak-binding SidK mutant
and that a stronger-binding construct may be more amenable for cryo-EM analysis.
An alternative solution may be to cross-link SidK to V-ATPase to prevent dissoci-
ation. A higher resolution map of the V-ATPase:SidK complex will facilitate accurate
fitting of atomic structures into the map. An atomic structure describing the V-
ATPase:SidK interaction would allow identification of the amino acid residues that are important for V-ATPase binding and inhibition by SidK.

### 6.2.3 Function of the flexible region of SidK

The flexible region of SidK does not interact with the V-ATPase (Chapter 5) and its function remains unclear. This region is predicted to constitute a C-terminal domain of the protein and may be involved in the secretion of SidK by the Dot/Icm secretion system of *L. pneumophila*. To investigate the role of the C-terminal region of SidK in protein secretion, SidK mutants with and without the C-terminal region can be assayed for secretion by *L. pneumophila*. Another possible role of the flexible region of SidK may be to recruit other *L. pneumophila* effectors or host proteins to the lysosomal membrane or V-ATPase to facilitate infection. To investigate this possibility, pull-down experiments could be performed on cell lysates using SidK as bait to isolate proteins that interact with SidK. Mass spectrometry analysis can then be performed to identify SidK interactors.

### 6.3 Cryo-EM methods

#### 6.3.1 Magnification anisotropy

The methods to correct magnification anisotropy in images described in Chapter 3 involves stretching or compressing the images, which can induce interpolation artifacts. This interpolation is additional to the interpolation performed during map construction. To minimize interpolation artifacts, magnification anisotropy in images should be corrected during map construction so that only one interpolation step is performed in total. Another solution may be to correct for magnification
anisotropy in the microscope. The cause of magnification anisotropy is currently unclear. Anisotropic magnification has been reported for images taken at lower magnifications (e.g. \( \sim 30,000 \times \)) but not at higher magnifications (e.g. \( \sim 70,000 \times \)), suggesting that the anisotropy may be dependent on the magnification setting. Investigation into magnification anisotropy under different imaging conditions may provide insight into the cause of the image distortions and reveal possible solutions to correct for the effects in the microscope.

### 6.3.2 Image classification and refinement

Recent implementation of maximum likelihood classification algorithms [162, 161, 106] have improved the separation of heterogeneous datasets into more homogeneous groups, improving the ability to isolate macromolecular structures in different conformations [49, 201] and facilitating structure determination by cryo-EM to high resolution [100, 18]. However, these algorithms are slow and can be inconsistent, making image classification and refinement a significant bottleneck in cryo-EM. New algorithms for cryo-EM have been developed recently that are based on stochastic methods [147, 25]. These algorithms are faster and allow structure determination of macromolecules without the need for initial models. Development of new algorithms to improve the efficiency of image analysis will increase cryo-EM throughput significantly.
Appendix A

Crystallization of SidK

A.1 Acknowledgments

This study was conducted in collaboration with Dustin J. Little and P. Lynne Howell at The Hospital for Sick Children Research Institute. Dustin J. Little and P. Lynne Howell guided the study and Jianhua Zhao performed the experiments. June Duong conducted some crystallization experiments with truncated SidK. Dustin J. Little analyzed the quality of crystals by X-ray diffraction. Hui Guo made the cysteine mutants and Ramesh B. Vanama performed the thermal shift assays.

A.2 Crystallization methods

Initial crystallization screening was performed with sparse-matrix screens purchased from Hampton Research and Microlytic. Fine grid screening of conditions was performed with reagents purchased from Hampton Research, Sigma-Aldrich, and BioShop. Experiments were performed by the hanging-drop method in 48-well plates and incubated at 20 °C. SidK was initially in TBS buffer (50 mM Tris-HCl, 300 mM...
NaCl, 5 mM dithiothreitol (DTT), pH 7.4) with concentrations ranging from \(\sim4\) to 20 mg/ml.

A.3 In situ proteolysis of SidK

Crystallization trials of full-length SidK were unsuccessful. To identify a more compact protein construct that may be amenable to crystallization, limited proteolysis of SidK was performed using the Proti-Ace kit (Hampton Research). 100 \(\mu\)l of SidK at 1 mg/ml was incubated separately with trypsin, chymotrypsin, papain, subtilisin, elastase, or endoproteinase Glu-C in 1:100 protease:protein ratios for 1 h at 37 °C. SDS-PAGE analysis of the reactions (Figure A.1) showed that SidK treated with trypsin and elastase had stable polypeptides with lengths that were shorter than wildtype, suggesting that these polypeptides may form stable cores that may be amenable to crystallization. Crystallization screening of SidK incubated with trypsin and elastase in situ revealed conditions where SidK crystallized in the presence of elastase (Figure A.2A-B). The crystals diffracted X-rays to \(\sim8\) Å (Figure A.2C-D).

A.4 Identification of elastase cleavage sites in SidK

A mixture of cleaved and uncleaved SidK proteins could hinder crystallization. In order to improve the homogeneity of the crystallization sample, the elastase cleavage site in SidK was determined by in situ proteolysis crystallization followed by mass spectrometry [101]. Crystals from in situ crystallization of SidK in the presence of elastase (Figure A.2) were collected and pelleted in a microfuge. The crystals were washed in crystallization buffer and solubilized in 3yes SDS-PAGE sample buffer (150 mM Tris-HCl, pH 6.8, 6 % SDS, 30 % glycerol, 10 % \(\beta\)-mercaptoethanol, 30 mM
Figure A.1: **Limited proteolysis reveals stable and smaller SidK species when incubated with trypsin and elastase.** 100 µl experiments were performed using the Proti-Ace kit from Hampton Research. The concentration of SidK was 1 mg/ml and proteases were added in 1:100 protease:protein ratios.
Figure A.2: Crystals from *in situ* proteolysis of SidK. (A) Incubation of SidK at 2.5 mg/ml with 1:100 elastase in buffer containing 8 % PEG8000, 20 mM KH$_2$PO$_4$, and 10 % glycerol. Crystals grew overnight. (B) Larger SidK crystals grew in optimized conditions (6.5 % PEG8000, 20 mM KH$_2$PO$_4$, and 9.5 % glycerol). Scale bar: 50 µm. (C) X-ray diffraction pattern of a SidK crystal from (B) in an arbitrary orientation and (D) the crystal rotated 90°. The crystals diffract to ∼8 Å. A high concentration of peaks along one axis (blue arrow) indicates that one axis of the crystal unit cell is long relative to the other axes.
Appendix A. Crystallization of SidK

EDTA, 0.06 % bromophenol blue). The solubilized crystals were analyzed by SDS-PAGE followed by in-gel tryptic digestion matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) (Figure A.3). Based on the mass spectrometry results and the knowledge that elastase cleaves polypeptides at small hydrophobic amino acids, it was hypothesized that the crystallized form of SidK had 9 and 30 amino acids truncated from the N- and C-terminal regions, respectively. A mutant construct of SidK missing these amino acids (SidK10-543) was created and crystallized in conditions similar to wildtype SidK incubated with elastase (Figure A.4A-B). Conditions that induced SidK10-543 crystallization in different forms were also identified (Figure A.4C-D). However, crystals that diffracted X-rays to better than 8 Å were not obtained.

A.5 Crystallization of truncated SidK

In order to probe different SidK constructs that may be more amenable to crystallization, truncated forms of SidK were created based on secondary structure predictions (Chapter 5, Figure 5.6). Table A.1 summarizes SidK constructs that were generated and crystallized (Figure A.5). An N-terminal truncation of 10-20 amino acids was necessary for SidK crystallization and C-terminal truncations of more than ~160 amino acids did not result in a soluble SidK construct that could be purified. None of the crystals obtained diffracted to high resolution.
Figure A.3: **Identification of trypsin-digested peptides by mass spectrometry of solubilized SidK crystals.** Peptides highlighted in yellow were identified by MALDI-TOF MS with residues in green representing possible modified residues in those peptides. Gaps in the middle of the sequence could be caused by the inability of the mass spectrometer to identify short peptides. Peptides from the N- and C-terminal regions of SidK were not identified and may have been cleaved by elastase.
Figure A.4: Crystallization of SidK10-543. (A) 5 % (w/v) PEG8000, 20 mM KH$_2$PO$_4$, and 14 % glycerol. (B) 10 % (w/v) PEG3350 and 0.1 M KI. (C) 750 mM (NH$_4$)$_2$SO$_4$, pH 7.3. (D) 825 mM Li$_2$SO$_4$, pH 7.1. Crystals in (A) and (B) resemble crystals from *in situ* proteolysis crystallization of SidK with elastase. All crystals that were screened did not diffract X-rays to better than 8 Å. Scale bar, 50 µm.
Table A.1: **Purification and crystallization of SidK truncation constructs.**

<table>
<thead>
<tr>
<th>Construct</th>
<th>Soluble</th>
<th>Crystallized</th>
<th>Construct</th>
<th>Soluble</th>
<th>Crystallized</th>
</tr>
</thead>
<tbody>
<tr>
<td>SidK1-573</td>
<td>yes</td>
<td></td>
<td>SidK1-543</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>SidK10-550</td>
<td>yes</td>
<td>yes</td>
<td>SidK10-543</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>SidK10-540</td>
<td>yes</td>
<td>yes</td>
<td>SidK10-535</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>SidK10-530</td>
<td>yes</td>
<td>yes</td>
<td>SidK10-525</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>SidK10-520</td>
<td>yes</td>
<td></td>
<td>SidK10-515</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>SidK10-510</td>
<td>yes</td>
<td></td>
<td>SidK10-504</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>SidK10-479</td>
<td>yes</td>
<td></td>
<td>SidK10-444</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>SidK10-414</td>
<td>yes</td>
<td></td>
<td>SidK10-395</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SidK10-364</td>
<td></td>
<td></td>
<td>SidK15-543</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>SidK15-530</td>
<td>yes</td>
<td>yes</td>
<td>SidK15-504</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>SidK15-479</td>
<td>yes</td>
<td></td>
<td>SidK15-444</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>SidK15-414</td>
<td>yes</td>
<td></td>
<td>SidK20-543</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>SidK20-530</td>
<td>yes</td>
<td>yes</td>
<td>SidK15-504</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>SidK20-479</td>
<td>yes</td>
<td></td>
<td>SidK20-444</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>SidK20-414</td>
<td>yes</td>
<td></td>
<td>SidK25-543</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>SidK25-530</td>
<td>yes</td>
<td></td>
<td>SidK25-504</td>
<td>yes</td>
<td></td>
</tr>
</tbody>
</table>
Figure A.5: Crystallization of SidK constructs. (A) SidK10-530 in 550 mM sodium malonate, 50 mM HEPES:NaOH, 0.25 % (v/v) Jeffamine ED-2001, and pH 7.0. (B) SidK10-525 in 15 % PEG400, 0.1 M magnesium chloride, 50 mM Tris:HCl, and pH 8.5. (C) SidK10-504 in 0.6 M DL-malic acid, 50 mM Bis-Tris propane:HCl, and pH 7.0. Scale bar, 50 μm.
A.6 Techniques to improve crystal quality

Crystal dehydration

Dehydration of crystals has led to improved X-ray diffraction in some cases [157]. 1 µl of 50 % PEG8000 was added to 2 µl hanging drops containing crystals of SidK10-543 in crystallization buffer (5 % [w/v] PEG8000, 20 mM potassium phosphate monobasic, and 14 % glycerol) and incubated overnight. The crystals were harvested, frozen in liquid nitrogen, and examined by X-ray diffraction. The crystals did not show improved X-ray diffraction from the dehydration procedure.

Slowing down crystallization

Most of the SidK crystals that were obtained grew overnight, which may not have allowed sufficient time for the formation of a well-ordered crystal lattice. To slow the rate of crystallization, SidK10-543 was crystallized at 4 °C over the period of 3-4 days. However, the crystals obtained were small and did not appear different from those obtained at 20 °C.

Lysine methylation

The methylation of lysines has been demonstrated to improve protein crystallization in some cases [192]. 5 ml of SidK10-543 at 1 mg/ml was mixed with 100 µl of 1 M dimethylamine-borane complex (ABC) and 200 µl of 1 M formaldehyde. The solution was incubated for 2 h at 4 °C. An additional 100 µl of 1 M ABC and 200 µl of 1 M formaldehyde was added to the reaction mixture and the solution was incubated for 2 h at 4 °C. An additional 50 µl of 1 M ABC was added to the reaction mixture and the solution was incubated overnight at 4 °C. The reaction mixture was concentrated using a centrifuge concentrating device and analyzed by size exclusion
Figure A.6: Crystallization of lysine-methylated SidK10-543. Buffer contained 12.5 % PEG3350, 0.1 M magnesium chloride, and 50 mM Bis-Tris propane:HCl (pH 6.5). Crystals were flexible and resembled limp noodles. Scale bar: 50 µm.
chromatography. Fractions corresponding to monomeric SidK were pooled and subjected to crystallization trials. Crystals were obtained (Figure A.6) but attempts at harvesting the crystals proved difficult due to flexibility of the crystals. This flexibility was thought to indicate a poorly-ordered crystal lattice and lysine methylation was not pursued further.

### A.7 Metal-mediated synthetic symmetrization

Crystal formation requires protein-protein interactions that stabilize the crystal lattice. Metal-mediated synthetic symmetrization is a method of inducing protein contacts artificially through the coordination of metal ions [91], which may promote crystal formation. The method involves incorporating two histidine residues in surface-exposed α-helical regions such that they are four amino acids apart, resulting in two histidine residues that are in similar orientations and are able to coordinate a divalent cation with two histidines from another protein. SidK is expected to be composed primarily of α-helices (Chapter 5, Figure 5.6) and contains numerous lysine residues, which tend to be surface-exposed. These attributes make metal-mediated symmetrization a viable method for improving SidK crystal formation. Four SidK mutants were made: K180H (paired with H184), K236H/K240H, E399H/E403H, and E496H/K500H. All four mutants were soluble and were purified. SidK has a tendency to form cysteine-cysteine cross-links and requires reducing agent to remain monomeric (Figure A.7). However, the proteins became unstable and precipitated upon exchange into buffer containing tris(2-carboxyethyl)phosphine (TCEP), a reducing agent compatible with divalent cations used in metal-mediated crystallization.
A.8 Mutation of cysteine residues

SidK has five cysteines and is able to cross-link into oligomeric forms (Figure A.7), which can hinder crystallization. DTT was included in crystallization buffers to maintain monomeric SidK, but this reducing agent reduces divalent cations required for crystallization methods such as metal-mediated crystallization. On the other hand, TCEP does not reduce divalent cations but destabilizes SidK (section A.7). In order to remove the requirement for reducing agent, the five cysteines in SidK were mutated to serines: C32S, C74S, C314S, C490S, and C541S. The individual cysteine mutants were stable and were purified, but the cysteine-free mutant (C32S/C74S/C314S/C490S/C541S) was insoluble and could not be purified. Thermal shift assays were performed as described previously [57] on the individual cysteine mutants (Figure A.8). The mutation C74S destabilized the SidK protein more than the other mutations, suggesting that C74 may be located in the hydrophobic core of SidK. However, the quadruple cysteine mutant (C32S/C314S/C490S/C541S) was able to form cross-links, indicating that C74 is surface-exposed. These data indicate that C74 may be partially buried in the hydrophobic core of SidK but remain solvent accessible. To account for the possible hydrophobic environment that may surround C74, this cysteine residue was mutated to valine. A cysteine-free SidK mutant (C32S/C74V/C314S/C490S/C541S) was soluble and was purified. Sparse-matrix screening of the cysteine-free mutant did not result in crystals. The quadruple cysteine mutant (C32S/C314S/C490S/C541S) was crystallized (Figure A.9) but the crystals did not diffract to high resolution.
Figure A.7: Cross-linking of SidK proteins. SDS-PAGE analysis of fractions from size exclusion chromatography under non-reducing conditions shows higher molecular weight bands due to cross-linked SidK (left). With addition of reducing agent (DTT), only bands corresponding to monomeric SidK are observed (right).
Figure A.8: Thermal stability of cysteine-to-serine SidK mutants. Thermal shift assays were performed using Sypro orange in a real-time PCR machine. The inflection points of the curves are indicated by the dashed lines. The inflection points for most of the cysteine mutants were close to 50 °C except for C74S where the inflection point was closer to 40 °C.
Figure A.9: Crystals of quadruple cysteine SidK mutant (C32S/C314S/C490S/C541S). Buffer contained 0.5 M sodium/potassium phosphate at pH 6.9. Scale bar: 50 μm.
Bibliography


[44] Veronica Kane Dickson, Jocelyn A Silvester, Ian M Fearnley, Andrew G W...


[100] Maofu Liao, Erhu Cao, David Julius, and Yifan Cheng. Structure of the TRPV1


[166] A M Smardon, H I Diab, M Tarsio, T T Diakov, N D Nasab, R W West, and


BIBLIOGRAPHY


[180] Takao Toyomura, Yoshiko Murata, Akitsugu Yamamoto, Toshihiko Oka, Ge-Hong Sun-Wada, Yoh Wada, and Masamitsu Futai. From lysosomes to the


[202] J Zhao, M A Brubaker, and J L Rubinstein. TMaCS: a hybrid template match-


mTORC1 Senses Lysosomal Amino Acids Through an Inside-Out Mechanism That Requires the Vacuolar H\textsuperscript{+}-ATPase, 2011.