The Study of Cyanobacterial Bicarbonate Transporters

and Applications of Microcrystal Electron Diffraction

by

Guanhong Bu

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved June 2021 by the
Graduate Supervisory Committee:

Brent L. Nannenga, Chair
Po-Lin Chiu
Jeremy H. Mills
David R. Nielsen
Cesar I. Torres

ARIZONA STATE UNIVERSITY

August 2021

ABSTRACT

Cyanobacteria contribute to more than a quarter of the primary carbon fixation worldwide. They have evolved a CO_2 concentrating mechanism (CCM) to enhance photosynthesis because inorganic carbon species are limited in the aqueous environment. Bicarbonate transporters SbtA and BicA are active components of CCM, and the determination of their structures is important to investigate the bicarbonate transport mechanisms. *E. coli* was selected as the expression host for these bicarbonate transporters, and optimization of expression and protein purification conditions was performed. Single particle electron cryomicroscopy (cryo-EM) or protein crystallography was carried out for each transporter. In this work, SbtA, BicA and SbtB, a regulator protein of SbtA, were heterologously expressed in *E. coli* and purified for structural studies. SbtB was highly expressed and two different crystal structures of SbtB were resolved at 2.01 Å and 1.8 Å, showing a trimer and dimer in the asymmetric unit, respectively. The yields of SbtA and BicA after purification reached 0.1 ± 0.04 and 6.5 ± 1.0 mg per liter culture, respectively. Single particle analysis showed a trimeric conformation of purified SbtA and promising interaction between SbtA and SbtB, where the bound SbtB was also possibly trimeric. For some crystallization experiments of these transporters, lipidic cubic phase (LCP) was used. In the case of LCP, often times the crystals grown are generally too tiny to withstand radiation damage from the X-ray beam during an X-ray diffraction experiment. As an alternative approach for this research, the microcrystal electron diffraction (MicroED) method was applied to the LCP-laden crystals because it is a powerful cryo-EM method for high-resolution structure determination from protein microcrystals. The new technique is termed as LCP-MicroED, however, prior to applying LCP-MicroED to

i

the bicarbonate transporters, methods needed to be developed for LCP-MicroED. Therefore the model protein Proteinase K was used and its structure was determined to 2.0 Å by MicroED. Additionally, electron diffraction data from cholesterol and human A_{2A} adenosine receptor crystals were collected at 1.0 Å and 4.5 Å using LCP-MicroED, respectively. Other applications of MicroED to different samples are also discussed.

DEDICATION

I dedicate this work to my parents for their unconditional love and trust in me and their

supports for my educational and career development; to my late grandparents for their

love, trust and support; to my cousin for her precious advice to pursue this journey of

graduate study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Cyanobacterial CO_2 Concentrating Mechanism

Atmospheric CO_2 concentration is expected to increase from 0.03% to up to 0.06% by volume over the 21st century due to human activities, which can cause irreversibly long-time climate change [1]. Many researchers have been motivated on utilizing atmospheric CO_2 for other purposes including energy storage and chemical synthesis [2-4]. In nature, photoautotrophs (i.e. plants, algae and cyanobacteria) use inorganic carbon source (Ci) for synthesis of organic compounds and release oxygen via photosynthesis, which is essential for many living organisms including humans. This process includes light-dependent reactions which produces oxygen from water and light-independent reactions which converts CO_2 into the carbohydrates by carbon fixation [5, 6].

Cyanobacteria are a group of ancient, largest, most diverse phototrophic prokaryotes, and are assumed to be the first oxygenic photoautotrophs appeared on the Earth over 2.3 billion years ago [7, 8]. They contribute to nitrogen fixation and up to 30% of the worldwide primary photosynthetic productivity (i.e. conversion of around 25 billion tons of carbon dioxide) [9, 10]. However, as limited concentration of Ci species are available in aqueous environment (in the magnitude of μM), cyanobacteria have evolved several mechanisms to improve photosynthesis, and one of them is the CO_2 concentrating mechanism (CCM) [11, 12]. CCM actively transports Ci into the cells to enhance the carbon fixation around the ribulose 1,5-bisphosphate carboxylase-oxygenase (Rubisco) in the carboxysome as Ci are limited in their natural environment. The pool of Ci inside the cell can be accumulated by up to 1,000 fold than the environmental level, and is measured between 20 to 40 mM. As

shown in Figure 1-1, cyanobacterial CCM is largely attributed to 5 different Ci uptake systems: three bicarbonate transporters BCT1, SbtA and BicA, which are located on the plasma membrane, and two CO_2 uptake systems NDH-I$_3$ and NDH-I$_4$, which are located on the thylakoid membrane [13-17].
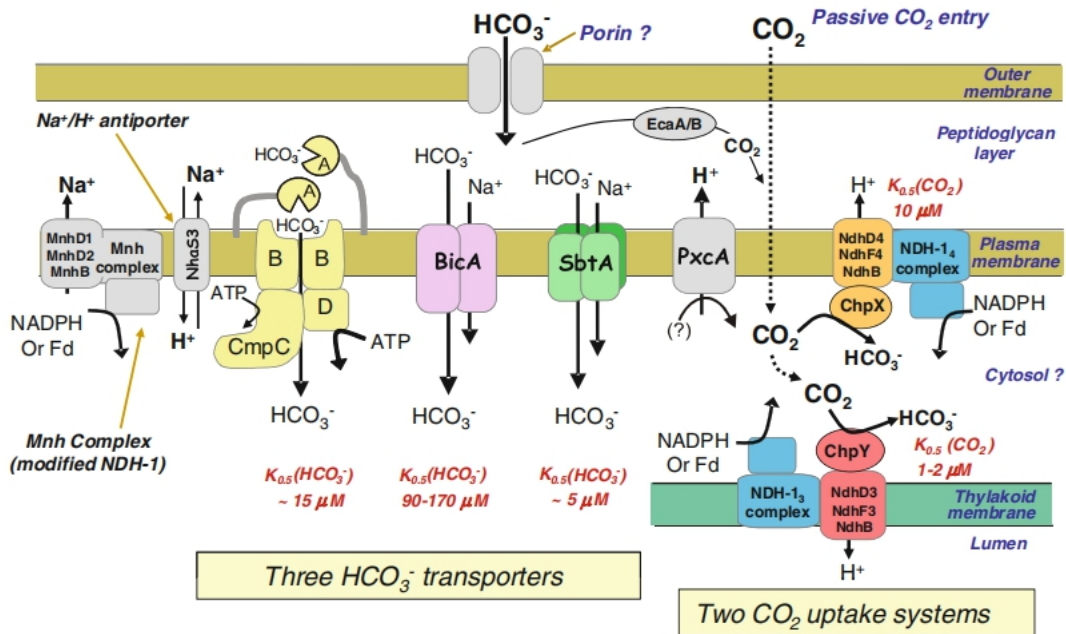
Figure 1-1: Scheme of the five known cyanobacterial Ci uptake systems in colors with some ancillary components in gray [16].

BCT1 is a member of ATP binding cassette (ABC) transporter family, and was identified in a freshwater cyanobacteria *Synechococcus sp.* strain PCC7942. It is encoded by the *cmpABCD* genes which are inducible under Ci limiting condition [18, 19]. BCT1 transports HCO_3^- with high affinity ($K_m = 15$ μM) and a low to medium flux rate [20]. It consists of 4 different protein subunits. CmpA is a periplasmic membrane protein which acts as the major HCO_3^- binding protein with high affinity ($K_d = 5$ μM). CmpB is a dimeric integral

membrane protein which forms the ion transport channel. CmpC is located on one cytoplasmic face of CmpB and contains 2 domains, where the N-terminal domain has ATP-binding motifs (similar to the ATP-binding domain in other ABC transporters) and the C-terminal domain is 29% identical to CmpA. CmpD is located on the other cytoplasmic face of CmpB and contains a ATP-binding motif which is 53% identical to N-terminal domain of CmpC. BCT1 is similar to another cyanobacterial ABC nitrate/nitrite transporter NRT1 which is composed of 4 protein subunits NrtA to NrtD. NrtC is involved in the inactivation of NRT1 when NH_4^+ is present. Likewise, CmpC could play a role in the dark inactivation of BCT1 transport activity possibly from its C terminal domain [16, 20].

SbtA was discovered in freshwater *Synechocystis sp.* PCC6803 as inactivation of BCT1 had little impact on the HCO_3^- transport activity, which indicated the existence of another HCO_3^- transporter. SbtA has no ATP-binding domains while it transport HCO_3^- driven by a Na^+ electrochemical gradient across the plasma membrane. The Na^+ concentration gradient is established by a Na^+/K^+ transporter NtpJ which is also essential for the growth of *Synechocystis sp.* PCC6803. 6 and 1 mM of external Na^+ is required for maintaining maximal and half-maximal HCO_3^- transport, respectively. Therefore, SbtA stands for Sodium-bicarbonate transporter A and belongs to the sodium solute symporter (SSS) family. SbtA is inducible under Ci-limiting conditions and transports HCO_3^- with high affinity (K_m = 16 µM) and low to medium flux rate (Vmax = 140 µmol·mg chl^{-1}·h^{-1}) [21, 22]. The scarcity of SbtA structure makes it difficult to investigate its molecular transport mechanism, however a study on SbtA topology was published and shows that SbtA has ten transmembrane helices (TMH) with 5+5 duplicated structure, where helices 1-5 and

6-10 share 23% sequence identity and 53% similarity, and this indicates that the pair adopt opposite or inverted orientation in the plasma membrane [23]. This 5+5 inverted structural symmetry motif is not uncommon for other SSS members, such as the bacterial Na^+/leucine symporter (LeuT) and bacterial Na^+/galactose symporter (vSGLT) [24-26]. Transporters with this internal two-fold pseudo-symmetry axis employ an alternating access mechanism, which means two halves of the TMHs alternately open the substrate binding gates at the periplasmic and cytoplasmic faces. However, SbtA is different from other SSS members that both N- and C-termini are periplasmic for SbtA and cytoplasmic for other SSS members. Topology map also shows that there is a cytoplasmic loop between TMH 5 and 6, which contains many positive residues, may be involved in the regulation of SbtA transport activity [23].

SbtA is assumed to be regulated by another protein SbtB, whose gene is located at the same operon as that encoding SbtA, in the dark conditions by binding to SbtA and suppressing HCO_3^- transport. This phenomenon has been confirmed in *Escherichia coli* (*E. Coli*) and several cyanobacteria strains including *Synechocystis* sp. PCC6803. SbtB is similar to the P_{II} signaling proteins, which form trimers and in some cases take part in regulating nitrogen metabolism in bacteria [27]. SbtA from *Synechocystis sp.* PCC6803 (SbtA-6803) is approximately 40 kDa in size and probably exists as a tetramer of 160 kDa measured from a blue-native PAGE experiment. However, 160 kDa can also be the result of complex size between a SbtA trimer and a SbtB trimer since trimeric SbtB from *Synechocystis sp.* PCC6803 (SbtB-6803) is about 36 kDa. Besides, SbtA-SbtB interaction is structurally similar to an ammonia transporter and its P_{II}-like inhibitor in *E. coli*, AmtB-GlnK, which forms a complex of trimer and trimer as shown in Figure 1-2 [27-29].

4

The ammonia transport activity of AmtB is blocked by insertion of an arginine containing T-loop from each protomer of the signaling protein GlnK [29]. Furthermore, SbtA was reported to be active in boron tolerance in several filamentous fungi, which suggests that SbtA plays a different role in other species [30].
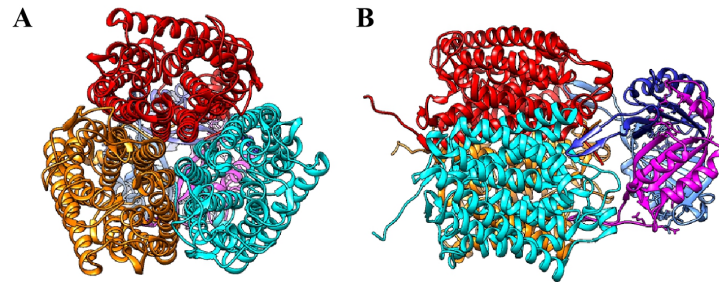
Figure 1-2: Ribbon representation of AmtB-GlnK model (PDB ID: 2NUU, http://www.rcsb.org) [29]. (A) Top view shows the trimer interface of ammonia channel AmtB. (B) Side view shows the complex binding interface where a loop of each signaling protein GlnK inserts into AmtB.

BicA was discovered in marine cyanobacteria *Synechococcus* PCC7002 because a mutant, which had SbtA insertion but no CmpABCD genes, showed response to HCO_3^- uptake at a higher Ci concentration. This phenomenon suggested that there was another HCO_3^- transporter with transport affinity lower than SbtA. A gene *sul3* encoding homologue to plant sulfate transporter was screened and identified as HCO_3^- transporter gene, and hereafter named as *bicA*. BicA from *Synechococcus* PCC7002 (BicA-7002) and *Synechocystis sp.* PCC6803 (BicA-6803) are both active transporters. They transport HCO_3^- with medium to low affinity (K_m = 38 μM for BicA-7002, and 171 μM for BicA-6803) and high flux rate (Vmax = 1333 μmol·mg chl^{-1}·h^{-1} for BicA-7002, and 1013 μmol·mg chl^{-1}·h^{-1} for BicA-6803) under their optimal conditions. Bicarbonate transport is

also driven by Na^+ gradient across the plasma membrane. 20 mM and 1.7 mM of extracellular Na^+ are required for BicA-7002 to maintain maximal and half-maximal transport activity, respectively [31].

BicA belongs to a large family of prokaryotic sulfate permease and eukaryotic solute carrier 26 transporter (SulP/SLC26). SulP/SLC26 proteins transport a large number of monovalent and divalent anions, such as sulfate, chloride, iodide, formate, oxalate, hydroxyl, bicarbonate, thiocyanate, etc. SLC26 transporters consist of 11 subfamilies (SLC26A1-SLC26A11) in mammalian, and act as electroneutral exchangers, ion channels, symporters and sometimes transport substrates in conjunction with ABC transporters. They play an important role in ion homeostasis and tissue development, and mutation or inactivation of SLC26 members is related to some rare genetic diseases. For example, an SLC26A1 member Sat-1 is expressed in hepatocytes and its malfunction leads to hepatotoxicity. An SLC26A2 member DTDST is located in chondrocytes and related to chondrodysplasias. An SLC26A4 member pendrin and SLC26A5 member prestin are widespread in cochlear cells and are associated with Pendred syndrome and deafness. An SLC26A8 member TAT1 is found in sperm cells and likely to be linked with male infertility. SulP/SLC26 members generally have N-terminal transmembrane domain (TMD) with 14 TMHs and C-terminal cytosolic sulfate transporter and anti-sigma factor antagonist (STAS) domain, while they may share as little as 23% sequence identity (e.g. BicA-7002 and DTDST). There are several highly conserved sites across SLC26 members, which lies in TMH 1-2, cytoplasmic loop between TMH 8 and TMH 9, and STAS [32-37].

Structural study of BicA-6803 was recently published by employing X-ray crystallography and electron cryo-microscopy (cryo-EM). A truncation of TMD (TM 14 was absent) crystallized in lipid cubic phase. Similar to other SLC26 proteins, the TMD was arranged as the 7+7 inverted repeats in which a pseudo-C2 symmetry axis between TMH 1-7 and 8-14 was parallel to the membrane plane, and divided into a core domain (TMH 1-4 and 8-11) and gate domain (TMH 5-7 and 12-14). The structure represented an inward-facing conformation with a HCO_3^- bound to a hydrophilic pocket which was formed by TMH 3, 8 and 10. The STAS domain crystallized as a stable homodimer and was required for dimerization of full-length BicA-6803 as revealed in a low resolution single particle cryo-EM map. Dimerization occurs in its physiological membrane distribution and is essential for HCO_3^- transport activity [38]. However, a high resolution structure of the full-length BicA has yet been available for investigating the mechanism of HCO_3^- transport and related conformations.
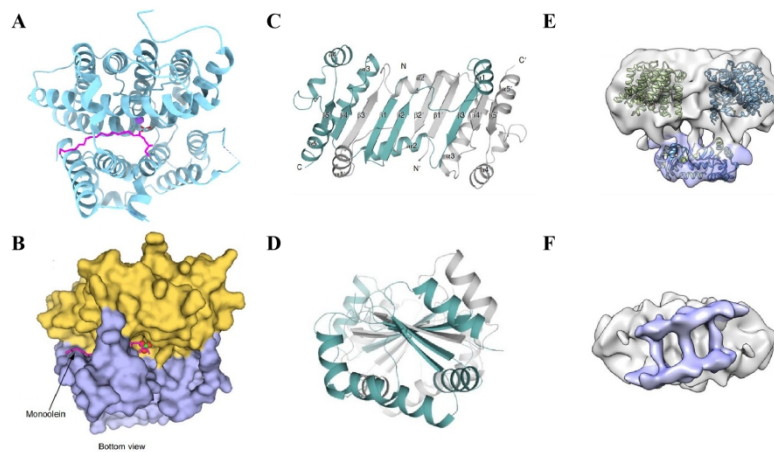
Figure 1-3: Published structural study of BicA-6803 [38]. (A) Ribbon view of TMD truncate showing the HCO_3^- binding site and a monoolein lipid in the hydrophilic pocket. (B) Detailed view of TMD truncate showing the arrangement of the gate (purple) and core (orange) domain. (C) Top view of the STAS domain showing the 2 STAS

monomers in cyan and gray. (D) Side view of the STAS domain. (E) Front view of the full-length BicA-6803 cryo-EM map with the TMD and STAS fitted. (F) Bottom view of the map showing the STAS dimer.

In recent years, several structures of eukaryotic SLC26A9 transporters have been published, and their atomic models are shown in Figure 1-4. A structure of mouse SLC26A9 transporter was solved in 2019 by single particle cryo-EM. It revealed a dimer conformation where two STAS domains acted as the primary dimer interface (Figure 1-4 A and B). The STAS domains were swapped and appeared as a knob at the center of particle. The TMD contributed to 11% of dimer interface and adopted a 7+7 inverted repeat architecture, where TMH 1-4 and 8-11 formed a convex core carrying a putative ion-binding site, and TMH 5-7 and 12-14 formed a concave gate [39]. A human SLC26A9 protein was determined at 2.6 Å in 2020 by single particle cryo-EM. The structure is very similar to the mouse SLC26A9 (Figure 1-4 C and D). It resolved multiple ion binding sites in the map and revealed a novel C-terminal sequence binding into the TMD which potentially acted as a gating modulator. This C-terminal sequence fold exhibited no obvious conformational change compared with the mouse SLC26A9, however it altered the surface charge [40].
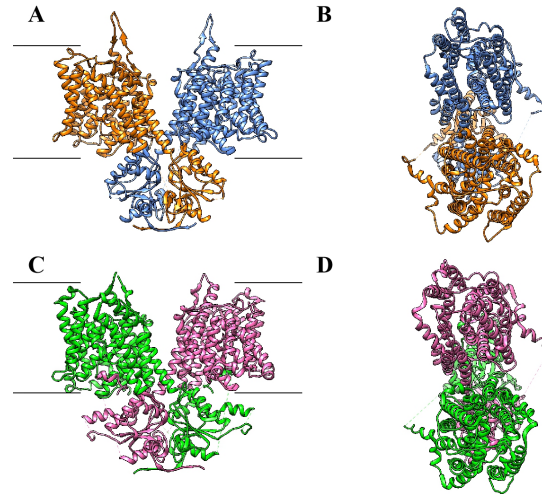
Figure 1-4: Ribbon representation of the eukaryotic SLC26A9 transporters [39, 40]. (A) Front view of the mouse SLC26A9 protein showing a clear dimer interface (Lines represent the surface of membrane bilayers). (B) Top view of the mouse SLC26A9 protein. (C) Front view of the human SLC26A9 protein showing a clear dimer interface (Lines represent the surface of membrane bilayers). (D) Top view of the human SLC26A9 protein.

This work aims to investigate the structural mechanisms of BicA and SbtA transport activity, and the interaction between SbtA and SbtB. *E. coli* is selected as the host system for heterologous expression of SbtB, SbtA and BicA due to the low cost, fast growth, and capability for high protein yield [27, 38, 41]. As both SbtA and BicA are integral membrane proteins, much work on screening protein purification conditions and methods for structure determination will be in demand as membrane protein structures are generally difficult to be determined, especially at high resolution.

1.2 Structure Determination of Membrane Proteins

Membrane proteins represent 30% of proteins encoded and expressed in living organisms [42]. They are the gates for substrate exchange between outside and inside of cells. They are

related with many diseases which make them ideal targets for more than 50% of drugs on the market and their design [43]. Based on their nature and extent of interaction with membrane bilayers, they can be classified as peripheral, amphitropic, integral and lipid-anchored membrane proteins [44]. Peripheral proteins (e.g. Cytochrome c) are attached to the membrane surface by electrostatic interaction or hydrogen bonding with either the surface of integral proteins or the head groups of phospholipids. They hardly penetrate the membrane bilayers while exhibit some weak hydrophobic interaction [44-46]. Amphitropic proteins (e.g. Protein kinase C) are either water-soluble or weakly associated with the membrane surface because they can exchange between conformations through a combination of electrostatic and hydrophobic forces. Both forms co-exist and the membrane-bound conformation is reversibly converted by water-soluble globular conformation following phosphorylation, acylation or ligand binding to expose the membrane binding site [44, 47, 48]. Integral proteins are tightly and permanently bound to the membrane by integration into the membrane's hydrophobic interior and the most abundant type among membrane proteins. In brief, they are divided into endo, ecto or trans-membrane proteins [44]. Endo or ecto proteins are inserted into the membrane's hydrophobic interior from the inner surface or the outer surface and do not penetrate the entire membrane (e.g. Cytochrome b_5 [49]). Transmembrane proteins are inserted throughout the membrane bilayers with some segments exposed both outwardly and inwardly to the aqueous spaces. They are further extended into single transmembrane α-helix proteins (e.g. Glycophorin [50]), multiple transmembrane α-helical proteins (e.g. Bacteriorhodopsin [51]), and multiple transmembrane β-barrel proteins (e.g. OmpF [52]). Lipid-anchored proteins are linked to membrane lipids through formation of thioester or

amide, and many of them are also transmembrane proteins (e.g. Calcineurin B). The anchored lipids involve myristic acid, palmitic acid, prenylated hydrophobic acyl chains and glycosylphosphatidylinositols (GPIs) [44, 53].

To determine a high resolution protein structure sheds light on the mechanism of protein functions and provides a model for further applications including protein engineering and drug design, since most intramolecular and intermolecular interactions are visible at high resolution. However, only a small fraction (around 1%) of protein structures deposited into the protein data bank (PDB, https://www.rcsb.org/stats [54]) refers to membrane proteins. Membrane proteins are difficult to be expressed at good quantity in host cells as they are hydrophobic and usually toxic to the cells. To purify membrane proteins from the host cells, detergents are employed to extract them from the lipid bilayers and form protein-detergent complex to maintain soluble in aqueous environment. Detergent screening is crucial in membrane protein structural biology since the correct choice of detergent, mostly nonionic or zwitterionic detergents, can maintain the right secondary and higher-order structures of membrane proteins [55]. X-ray crystallography and single particle cryo-electron microscopy (cryo-EM) have been developed to determine the structures of membrane proteins.

X-ray crystallography is one of the major methods of membrane protein structure determination since it works for membrane proteins of any size and can yield to high resolution structures. A membrane protein of interest has to be crystallized to form well-ordered crystals, where individual molecules adopt one or a few identical orientations and closely pack to each other [56]. The workflow is summarized in Figure 1-5. A protein crystal is mounted in the X-ray beam, the electron cloud within the exposed crystal is

scattered, and the scattered rays are recorded on a detector on which many spots occur meeting the Bragg's law. The spots form a pattern in reciprocal space and are measured for determining the space group and unit cell parameters in real space. Structure factors are calculated by measuring the intensity and phase of the diffraction spots (the latter is lost during data collection while can be determined by other methods), and can be used for solving the electron density map by Fourier transform. The map is interpreted with a model of protein tertiary or quaternary structure [57]. A well-ordered and closely-packed protein crystal can be diffracted to high resolution (i.e. < 3 Å for membrane protein), which is the most difficult and time-consuming part of protein crystallography [56, 57]. The first high-resolution membrane protein structure was the protein subunits in the photosynthetic reaction center determined at 3 Å using X-ray crystallography [58]. Around 90% of membrane protein structures deposited in membrane protein data bank (MPDB, http://www.mpdb.tcd.ie/stat.asp [59]) have been determined by X-ray crystallography.
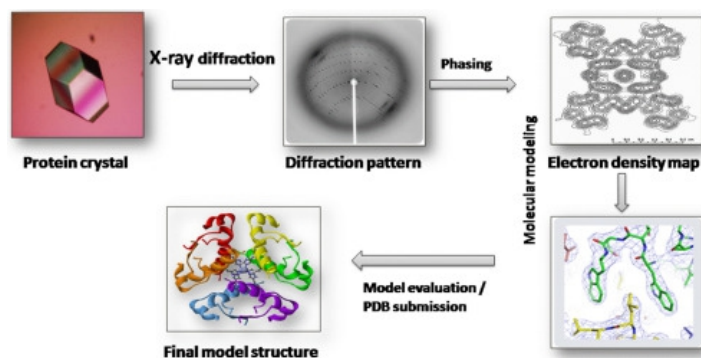
Figure 1-5: A general workflow of protein structure determination by X-ray crystallography. Gawas et al. (2019).

High quality membrane protein crystals are necessary for high resolution structure determination, and there are several approaches to grow them: vapor diffusion, counter-diffusion, lipidic cubic phase (LCP), bicelle, etc. Vapor diffusion and LCP are covered in this dissertation. Vapor diffusion is the most commonly used in both soluble and membrane protein crystallography. Purified membrane protein solution is mixed with the crystallization solution at designed ratios to produce the mixed droplets of 0.2 to 10 µL. The droplets are sealed in a chamber and separated from a large volume of crystallization solution at the bottom of chamber. The crystallization solution evaporates in the chamber and slowly concentrates the drop to force membrane proteins into supersaturation. Crystals may occur if nucleation sites form and grow (Figure 1-6 A) [60]. Crystallization in LCP is achieved by mixing membrane protein sample with a host monoacylglycerol lipid (e.g. monoolein) at a ratio of 2:3 (v:v) in order to form bicontinuous lipid cubic phase, and the phase is incubated against the crystallization solution in sealed glass sandwiches (Figure 1-6 B). This method provides a near-native environment for membrane proteins and has been successful in obtaining high quality crystals [61-63].

A   **B**
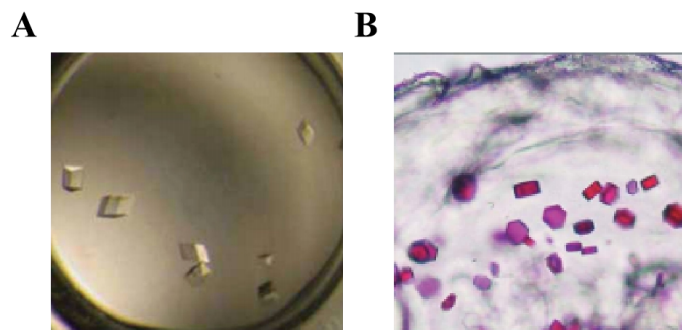
Figure 1-6: Examples of membrane protein crystals grown in (A) vapor diffusion [60] and (B) lipidic cubic phase [63].

As membrane proteins need to be concentrated to reach the state of supersaturation, followed by addition of precipitate which slows down the diffusion rate and forms the nucleation sites for crystal growth, the quality of membrane protein crystals are affected by factors including host environment, temperature, pH, salt, precipitate, additive, and so on. It usually begins with testing a number of precipitate solutions which are available from commercial kits, and once crystal hits are found, optimization such as gradients of salt, precipitates and pH or lower temperature will be performed. Most membrane proteins are crystallized in the presence of polyethylene glycol (PEG) of different molecular weight as precipitates. Crystallization are commonly set up at 293 K, although incubation at lower temperature like 277 K or 285 K is also common for sensitive proteins [57].

Once crystals are generated, they need to be mounted in the X-ray beam for diffraction experiments. The most commonly used X-ray source is synchrotron radiation (SR). SR is produced by the electrons which are accelerated to near the speed of light and confined to a giant near-circle ring by electromagnetic field. As electrons are forced into curved or bent motion, they emit energy as SR which contains a fraction of X-ray. SR is one of the most brilliant light source on earth and the most powerful tool for crystallography since a synchrotron facility offers many beamlines for various purposes including single wavelength diffraction, anomalous diffraction, Laue diffraction, etc. Membrane protein crystals are continuously rotated in the X-ray beam via a goniometer, and diffraction patterns from as many different orientations as possible can be collected for structure determination. However, radiation damage occurs as crystal is continuously exposed to X-ray beam. Cryocrystallography (i.e. crystals are cryo-cooled at liquid nitrogen) is one

of the solution to address radiation damage because the rates of free radical formation, chemical bond breakage and lattice spacing interruption are much lower at liquid nitrogen temperature than at ambient temperature [57]. Furthermore, microfocus SR beamlines are adopted to facilitate diffraction of membrane protein crystals as they are often small or a small fraction of the crystal can be diffracted to high resolution [64].

Cryo-EM utilizes a transmission electron microscope (TEM) operating at cryogenic temperature for macromolecular structure determination, including membrane proteins [65]. In cryo-EM, low dose of electron beam, whose wave nature was confirmed a century ago [66], and cryogenic temperature (i.e. 77 K of liquid nitrogen or 4 K of liquid helium) are crucial for biomolecule structure determination in order to reduce radiation damage and preserve physiological conditions [67, 68]. Biological samples are embedded in a thin layer of vitreous water solution to remain hydration within the samples and enhance contrast as visualization of biological samples are impacted by ice. Thus, biological samples must be applied onto TEM grids coated with holey film, blotted for a few seconds to generate thin hydration layers, and plunged into liquid ethane to vitrify the thin hydration layers. Biological samples maintain near-aqueous conditions in the holes [69].

Single particle cryo-EM has been popular in recent years for protein structure determination. Unlike the collection of diffraction patterns in X-ray crystallography, it is the electron micrographs or images of millions of individual biological macromolecules that are collected by a cryo-TEM. This giant number of particles are computationally classified and averaged by different 2D orientations to generate templates for initial 3D reconstruction map, the initial 3D map has been continually refined to a final map which can be interpreted by a structure model (Figure 1-7). Single particle cryo-EM bypasses

the time and effort spent producing crystals and allows fast discernment of molecules of different structural and functional states. However, protein particles below a size limit (typically 100 kDa) are difficult to be visualized and collected. Besides, membrane proteins purified in detergent solution tend to aggregate, thus it is necessary to screen the homogeneity of membrane protein molecules prior to cryo-EM [70-73]. Prior to data collection in a cryo-TEM, negative staining is very useful for fast screening of sample purity and homogeneity to optimize sample buffer recipe and detergent [74].
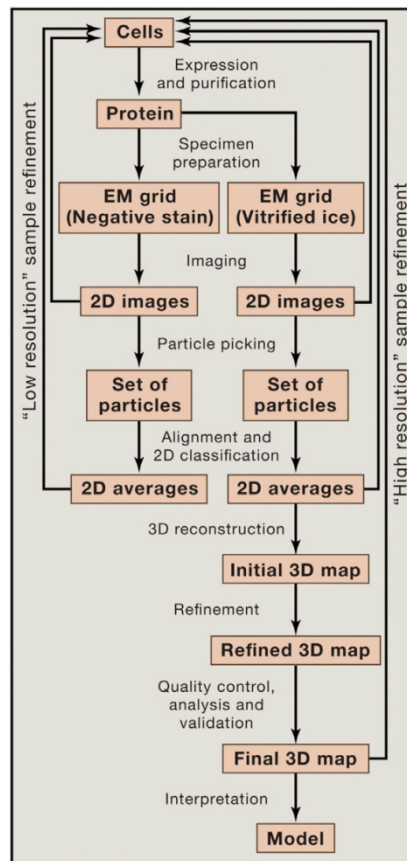
Figure 1-7: A general workflow of protein structure determination by single particle cryo-EM [70].

There are six types of host environment for membrane protein sample preparation in single particle cryo-EM: detergent, amphipols, liposome, bicelle, lipid nanodisc, and peptidisc. Detergent is the most common since membrane proteins are extracted from the biological membrane and purified with detergents [75]. Amphipols are amphiphilic polymers designed to bind to the transmembrane domain of membrane proteins via their multiple hydrophobic side chains with high affinity, and the resulting membrane protein-amphipol complexes are soluble in aqueous solution via the multiple hydrophilic side chains of amphipols. The stability of membrane proteins can be enhanced as the complexes remain soluble in detergent-free aqueous solution and no interaction between detergent monomers and membrane protein-detergent micelle complexes occurs, which usually causes membrane proteins to be dissociated from their higher oligomeric state or denatured [76, 77]. Liposome is a spherical vesicle of lipid bilayer where membrane proteins are incorporated. Membrane proteins maintain their natural conformation in lipid bilayer. However, liposomes are sometimes too large to achieve high number of liposomes in the holes, and therefore longer sample incubation time on the holey carbon grids and multiple sample application are necessary for grid preparation [78, 79]. Lipid nanodisc uses a nanoparticulate lipid bilayer disc assembled and trapped by two membrane scaffold proteins which are usually amphipathic helical proteins to stabilize the membrane proteins incorporated into the lipid bilayer disc [80, 81]. Peptidisc is structurally similar to nanodisc, however the two membrane scaffold proteins are replaced with many amphipathic peptides. Both membrane scaffold proteins and amphipathic peptides align in a double belt to support the shape of lipid bilayer disc [82, 83].

1.3 Microcrystal Electron Diffraction (MicroED)

Apart from X-ray crystallography and single particle cryo-EM, electron crystallography offers alternatives for membrane protein structural biology. Electron diffraction is conducted in a transmission electron microscope (TEM). The electrons produced by a TEM have de Broglie wavelength of 0.0251 Å at acceleration voltage of 200 kV and 0.0197 Å at 300 kV, thus they are precise probe for molecular structure. Besides, every elastic X-ray scattering event by a crystal to contribute to the a diffraction pattern is accompanied with 10 inelastic events that cause radiation damage, however electrons deposit energy into a crystal two to three orders of magnitude less than X-ray per elastic scattering event [84].

Electron crystallography in cryo-EM was developed in the mid 1970s from 2D crystals. 2D crystals are produced by crystallizing membrane proteins in lipid environment and extending on the surface of TEM grids as films. It is unique for 2D crystals to observe the membrane protein molecules from the images and the diffraction spots by processing fast Fourier transform of the images. Although it is difficult to produce well-ordered 2D crystals, and the electron dose rate applied to them leads to rapid destruction of crystals and limited number of diffraction frames, high resolution structures of membrane proteins have been determined by electron crystallography [86-89]. To mitigate radiation damage, the TEM runs at low dose setting throughout data collection and cryogenic temperature maintained by liquid nitrogen or liquid helium [57]. The structure of Bacteriorhodopsin was determined at 3 Å resolution over a decade ago [85], and around 8 structures are determined by electron crystallography.

However, structure determination by electron crystallography from 3D crystals had not been demonstrated until 2013 [89]. 3D electron crystallography utilizes a cryo-EM method to collect diffraction data from crystals whose sizes are up to several micron from 3 dimensions. Such method is termed as Microcrystal electron diffraction (MicroED). In traditional single crystal X-ray crystallography, large protein crystals (i.e. >10 μm) are required to withstand the radiation damage. In MicroED, electron interacts with materials stronger than X-ray, thus much smaller crystals (membrane protein crystals are often tiny) are enough to produce meaningful diffraction data and withstand the radiation damage. Although X-ray is scattered only by electron cloud, electrons are scattered by both electron cloud and nuclei, and hence information on charged atoms are more sensitive from MicroED. Furthermore, both X-ray and electron diffraction share many common features, so electron diffraction patterns can also be processed in the programs developed for single crystal X-ray crystallography, which skipped the need for developing algorithms for data processing [89-91]. In a nutshell, MicroED provides a good option for membrane protein structure determination.

The first protein structure determined by MicroED was lysozyme, using 90 series of still diffraction patterns from lysozyme micro-crystals and yielding to final resolution of 2.9 Å. High quality diffraction patterns were obtained from crystals 1-6 μm long and wide, and less than 1 μm thick [89]. Later, the continuous rotation MicroED method was applied to Lysozyme, Bovine liver catalase, Proteinase K, peptides, and so on [90-94]. The first membrane protein structure determined by MicroED was Ca^{2+}-ATPase. The map was refined to 3.4 Å and showed partial charges on residues glutamic acid, aspartic acid and bound calcium ions [95]. Later, 2.9 Å resolution structure of TGF-βm-TβRII complex was

determined by MicroED. The complex crystals were fragmented from needle clusters by vortexing with disruption glass beads prior to grid preparation [96]. A truncated wild-type NaK ion channel structure was solved at 2.5 Å by MicroED, where new densities of charged ions were identified surrounding the external selectivity filter. Combined with X-ray structures, a molecular movie of ion partition and conduction could be drawn [97].

MicroED is a part of cryo-EM methods, thus the sample preparation and data collection steps are similar to single particle method. In terms of sample preparation, droplets of microcrystals are applied onto the surface of holey TEM grids, followed by blotting the grid to make thin hydration layer and plunging into liquid ethane. With regard to data collection, grids are loaded into a cryo-TEM and the overall quality is screened under the low magnification imaging mode at low dose settings. Once a good-looking crystal is identified and centered in the beam, the microscope will be switched to the diffraction mode for data collection, unlike single particle data collection. All diffraction spots or reflections are recorded by sensitive scintillator-based complementary metal oxide semiconductor (CMOS) detector as the crystal is continuously rotated by the stage, which is similar to the goniometer applied in X-ray crystallography. Thanks to the CMOS detectors, the electron dose rate and total dose deposited on crystals could be less than $0.01 \ e^-/(Å^2 \cdot s)$ and $5 \ e^-/Å^2$ to collect meaningful diffraction data sets [98, 99]. Once data sets are collected, they can be processed in the programs developed for X-ray crystallography.

A

Microcrystals
↓ Plunge freezing
EM grids
↓ Cryo-TEM
Imaging at low magnification ←
↓ Screening
Crystal at the beam center
↓ Diffraction mode
Still diffraction — Bad
↓ Good / Back to imaging mode
Maximal stage tilt range
↓ Diffraction mode
Continuous rotation data collection

B

Continuous rotation data sets
↓
Data reduction
↓
Data scaling
↓ High completeness
Better than 1.2 Å
Yes ↓   No →
Direct phasing   Molecular replacement
↓
Data refinement
↓ Interpretation
Structure model

Figure 1-8: A general workflow of MicroED (A) data collection and (B) data processing.

Lipidic cubic phase (LCP) or *in meso* crystallization of membrane proteins first appeared in 1996 [61]. In general, membrane proteins are purified from the biological membrane with detergent, reconstituted into monoacylglycerol (MAG) lipid bilayer to form cubic phase, and incubated against precipitants [100, 101]. LCP was developed based on the phase behavior of MAG/water system. For example, monoolein (9.9 MAG) adopts a cubic phase at the ratio of 3/2 (v/v) monoolein/water, where monoolein adopts highly packed and curved continuous bilayer structure, resembling saddles, and water forms intersecting channels at the core of "saddles". LCP serves as infinite reservoir of reconstituted proteins which arrange in random orientation within the monoolein bilayer. Once LCP is formed, precipitate is added to form local lamellar phase to facilitate protein-protein interaction and concentrate proteins to trigger nucleation and crystal growth [100-103]. Crystals grown in LCP resemble the granule crystalline embedded in thick toothpaste [100, 101, 104]. LCP is a powerful method in high resolution structural studies of membrane proteins as it offers a lipid bilayer environment [60, 100, 101, 105].

21

LCP is thermodynamically stable at 293 K by mixing two parts water phase (i.e. protein solution) and three parts monoolein. Other short-chain MAGs are also employed if the crystallization is set up at low temperature or protein contains large soluble domains [102, 106, 107]. In most cases, membrane proteins are crystallized against polymer-based precipitants (e.g. polyethylene glycol (PEG)) which cause the cubic phase to be spongified and form lamellae in which the bilayer interfacial curvature and its bending rigidity are lowered and the mesophase can absorb more aqueous solution [100, 101]. Besides, presence of additive lipids (e.g. cholesterol) in the mesophase facilitates crystallization of tough membrane proteins, especially G protein-coupled receptors (GPCRs) when crystallized against precipitant conditions involving PEG 400 [105].

Membrane proteins adopt crystal packing type I in LCP and type II in vapor diffusion. In type I, crystals are packed as stacked 2D crystals where hydrophobic surfaces are the driving force of crystal contact, thus lattice packing is close in the plane of membrane. Close and well-ordered crystal lattices are necessary for diffraction at high resolution. In type II, crystal contacts are established by hydrophilic surface, and hence membrane proteins with large soluble domain or associated with soluble proteins are easier to be crystallized. Besides, short-chain detergents need to be screened as not to cover too much of polar surfaces [108, 109].

Some membrane protein crystals grown in LCP are large enough for single crystal X-ray diffraction using synchrotron radiation, however most LCP-based crystals are tiny. Serial femtosecond crystallography (SFX) was employed for structure determination of LCP-based microcrystals, where the crystal laden LCP is directly delivered to the free electron laser X-ray beam via a specially designed high viscosity media injectors (also known as

LCP injectors) [110, 111]. Such technique, referred to as LCP-SFX, skips the effort spent on harvesting crystals and allows *in situ* data collection at room temperature which is close to the physiological conditions of most membrane proteins [101].

MicroED shows the potentials for structure determination of microcrystals grown in LCP. Most universities and research institutes are equipped with cryo-TEM, thus it is more available and accessible for users to collect MicroED data than waiting for beamtime assigned by synchrotron or free electron laser facilities. Besides, LCP-SFX requires over hundreds of thousands of crystals since each hit contributes to a snapshot of diffraction pattern, however, MicroED significantly reduces the number of hit crystals for structure determination (i.e. <10) with high completeness. The combined technique, hereafter designated as LCP-MicroED, is likely to open a new era for membrane protein structure determination, especially for GPCRs and other tough targets.

This work will demonstrate the first example of LCP-MicroED application. Since LCP is similar to toothpaste, it is required to reduce the viscosity of crystal-laden mesophase for TEM grid preparation. Proteinase K microcrystal embedded *in meso* was selected for testing the diffraction quality after lowering the phase viscosity because it was respectively screened at high resolution by MicroED [96, 112] and LCP [113]. Once the test was confirmed feasible, GPCRs crystallized *in meso* would be applied.

Although initially developed for protein structure determination, MicroED was recently applied in the field of small organic molecules. For a long time, structures of small organic molecules have been assigned by ultraviolet-visible spectroscopy [114] and infrared spectroscopy [115], mass spectrometry [116], nuclear magnetic resonance (NMR) spectroscopy [117], X-ray powder diffraction [118], single crystal X-ray diffraction (SXRD) [119], and so on.

23

Among them, SXRD is the only technique to determine the position and orientation of individual atoms and bonds within a given molecule with precision [120, 121]. Although SXRD requires crystal size to be at least 10 μm for synchrotron and 50 μm for home source, MicroED imposes no lower crystal size limit which bypasses the time for producing large crystals [122].

In 2018, two independent articles on the application of MicroED in small organic molecules were published at nearly the same time. One demonstrated the applicability of MicroED on natural products from chemical suppliers and pharmacies, purified samples by flash column chromatography but not crystallized, and heterogeneous samples (i.e. mixtures of compound) which are precluded by NMR and SXRD. A total of 11 different structures were determined at high resolution in minutes [121]. The other employed electron diffraction of a cold flu medicine which was hardly diffracted even by synchrotron X-ray source and of a novel methylene blue derivative $MBBF_4$ at 0.9 Å [122]. Later, MicroED proved successful in determining the absolute stereochemistry of several small organic molecules at high resolution [123-125], and identifying different enantiomers in the asymmetric unit [126].

This work will demonstrate some examples of application of MicroED structure determination to small organic molecules.

CHAPTER 2: CRYSTAL STRUCTURES OF THE CYANOBACTERIAL BICARBONATE TRANSPORT REGULATOR SBTB

2.1 Introduction

Cyanobacteria account for over a quarter of global carbon fixation [9]. To do so, cyanobacteria require the ability to transport inorganic carbon which, in the aqueous environments that they occupy, is found mostly in the form of HCO_3^-. While cyanobacteria are known to transport HCO_3^- into the cell through the coordinated function of multiple membrane-protein transporters, much remains to be understood regarding how these systems both function and are regulated.

One of the key cyanobacterial bicarbonate transporters SbtA is a Na^+-dependent bicarbonate transporter with low flux rate and high affinity [16]. SbtB, a soluble cytoplasmic protein with a similar folding to P_{II} signaling proteins, has been shown to regulate the transport activity of SbtA. Studies suggest that this occurs via the direct binding of SbtB to SbtA, which is controlled as a function of ATP/ADP levels within the cell and then blocks HCO_3^- transport by SbtA [27, 127, 128].

Recent crystal structures of SbtB originating from both *Cyanobium sp.* PCC 7001 [127] and *Synechocystis sp.* PCC 6803 (SbtB-6803) [129] show trimeric structure and canonical P_{II}-like folding. P_{II} proteins are widespread in nature and regulate a range of processes, most notably those involved in nitrogen metabolism [130]. In this work, two crystal structures of apo-form SbtB-6803 are presented. One was determined at 2.0 Å and contains a trimer in the asymmetric unit (ASU). The other structure contains a dimer in the ASU, rather than the native trimeric form. The structure was determined at a resolution of 1.8 Å, shows an

extended C-terminus, and contains an intermolecular disulfide bond between residues C94 from adjacent SbtB-6803 dimers. When viewed together, the symmetry related disulfide bonded dimers suggest a potential dimer of dimers arrangement within the crystal, in which one of the two-fold axis is crystallographic.

2.2 Materials and Methods

2.2.1 Expression Vector and Strain

The gene encoding full length SbtB-6803 (i.e., *slr1513* from *Synechocystis sp.* PCC 6803) was constructed with N-terminal 6x-His tag and tobacco etch virus (TEV) protease cleavage site (Integrated DNA Technologies), and fused into pBLN200 vector using Gibson assembly [131, 132] to give the plasmid pSbtB-200. The plasmid was confirmed by DNA sequencing, and transformed into *E. coli* strain BW25113 [133].

2.2.2 Protein Expression and Purification

Cells were grown at 37 °C in 1 L of Luria Broth (LB) media supplemented with 50 µg/mL kanamycin. Gene expression was induced when the cell density (OD600) reached approximately 0.5 by addition of 0.2% (w/v) L-arabinose, and expression was carried out for 3 h. Cell pellets were harvested at 4,000 x g and lysed by sonication in 50 mM Tris-HCl pH 8.0, 100 mM NaCl and 1 mM PMSF. The cell lysate was centrifuged at 10,000 x g to remove cell debris and unbroken cells. The supernatant was loaded into a HisTrap HP column (GE Healthcare) pre-equilibrated with 20 mM Tris-HCl pH 7.5 and 0.5 M NaCl. The column was then washed with 20 mM Tris-HCl pH 7.5, 0.5 M NaCl and 75 mM imidazole. His-tagged SbtB-6803 was eluted from the column in 20 mM Tris-HCl pH 7.5, 0.5 M NaCl and 500 mM imidazole. To remove the 6x-His tag, His-tagged TEV

protease was added at a ratio of 1 mg per 50 mg of His-tagged SbtB-6803. The reaction mixture was added to a 2.5 kDa cut-off dialysis membrane, and cleavage was performed as the sample was dialyzed overnight at 4 °C against 50 mM Tris-HCl pH 8.0 and 50 mM NaCl. Uncleaved SbtB-6803 and TEV protease were removed by passing the digested sample through a pre-equilibrated HisTrap HP column. The flow through containing cleaved SbtB-6803 was collected and dialyzed against 20 mM Tris-HCl pH 7.8, and further purified by running anion exchange chromatography with a HiTrap Q HP column (GE Healthcare) also pre-equilibrated with 20 mM Tris-HCl pH 7.8. The same pre-equilibration buffer was used along with a linear gradient of 1 M NaCl in 20 mM Tris-HCl pH 7.8 to elute SbtB-6803 for crystallization, which occurred at approximately 340 mM NaCl. SbtB-6803 was collected and SEC was performed in 20 mM Tris-HCl pH 7.5 and 150 mM NaCl on a Superdex 75 10/300 GL column (GE Healthcare).

2.2.3 Polyacrylamide Gel Electrophoresis (PAGE)

Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed on 4% stacking gels followed by 15% resolving gels. Protein solutions were mixed with equal volume of 62.5 mM Tris-HCl pH 6.8, 5 mM DTT, 25% glycerol, 0.1% bromophenol blue and 0.2% SDS. Native-PAGE was performed on 4-20% mini-protean TGX precast protein gels (Bio-Rad). Protein solutions were mixed with equal volume of 62.5 mM Tris-HCl pH 6.8, 5 mM DTT, 25% glycerol and 0.1% bromophenol blue. The protein bands were visualized by 0.1% Coomassie blue R-250.

2.2.4 Protein Crystallization

Purified SbtB-6803 was concentrated to 13-25 mg/mL and crystallization screening was set up in 96-well Swissci UVXPO MRC crystallization plates (Molecular Dimensions)

via sitting-drop vapor diffusion at 21 °C. Initial hits were found with a precipitant containing 0.2 M sodium citrate tribasic dihydrate, 0.1 M HEPES pH 7.5 and 20% (v/v) isopropanol. Optimized crystals were grown in 24-well Cryschem sitting drop crystallization plates (Hampton Research) by mixing 2.5 µL of protein and 2.5 µL of precipitant containing 0.2 M trisodium citrate, 0.1 M HEPES pH 7.2, 12% (v/v) isopropanol and 0.2 M NiCl$_2$. Quality crystals were cryoprotected by soaking using artificial mother liquor supplemented with additional 30% (v/v) glycerol for 10 minutes, and subsequently cryo-cooled in liquid nitrogen for data collection.

SbtB-6803 purified in the SEC buffer solution supplemented with 5 mM Tris(2-carboxyethyl)phosphine (TCEP) was concentrated to 13-27 mg/mL and crystallization screening was set up in 96-well Swissci UVXPO MRC crystallization plates (Molecular Dimensions) via sitting-drop vapor diffusion at 20 °C. Initial hits were found with a precipitant containing 0.1 M Tris pH 8.5 and 2 M ammonium phosphate monobasic. Optimized crystals were grown in 24-well Cryschem sitting drop crystallization plates (Hampton Research) by mixing 3 µL of protein and 1.5 µL of precipitant containing 0.1 M Tris pH 8.6 and 2.25 M ammonium phosphate monobasic. Quality crystals were cryoprotected by soaking in 0.1 M Tris pH 8.6 and 3 M ammonium phosphate monobasic, 5% (v/v) glycerol and 10 mM TCEP for overnight, and subsequently cryo-cooled in liquid nitrogen for data collection.

2.2.5 Data Collection and Processing

Diffraction data of SbtB-6803 were collected at the National Synchrotron Light Source II (NSLS-II) beamline AMX 17-ID-1 on an Eiger 9M detector at X-ray wavelength of

0.92009 Å for a total of 225 x 1° oscillation frames to produce a data set that was processed to 1.8 Å using HKL2000 [134].

Diffraction data of SbtB-6803 supplemented with TCEP were collected at the Stanford Synchrotron Radiation Lightsource (SSRL) beamline BL9-2 on a Pilatus 6M detector at X-ray wavelength of 0.9794 Å. The data set was processed to 2.01 Å using iMOSFLM [135].

2.2.6 Structure Determination and Refinement

The data of SbtB-6803 without and supplemented with TCEP were phased by molecular replacement using Phaser [136] with the coordinates of a previously determined homologous structure from *Anabena variabilis* (PDB entry 3dfe) and a published structure of SbtB-6803 (PDB entry 5o3s [129]), respectively. Data were refined using phenix.refine [137] within the PHENIX program suite [138]. The model building was manually performed using Coot [139]. Calculation of free R values for the structure was based upon 10% of the reflections. MolProbity [140] was used for Ramachandran analysis. Structural representation was performed in UCSF chimera [141]. The data of SbtB-6803 supplemented with TCEP were phased by molecular replacement.

2.3 Results and Discussions

Heterologous expression of SbtB-6803 in *E. coli* BW25113 and subsequent purification procedures produced 16.0 ± 3.0 mg His-tag cleaved SbtB-6803 per liter of culture for crystallization experiments. The protein yield was so high that no optimization of expression and purification conditions was needed.

SDS-PAGE of SbtB-6803 following SEC purification both without and with TCEP (Figure 2-1 A; middle lane and right lane) showed 3 bands at about 12 kDa, 24 kDa and

45 kDa which corresponded to the size of monomer, dimer and possibly trimer. Besides, native PAGE showed single band from samples both without and with TCEP (Figure 2-1 B; left lane and right lane), and both bands were at the same size. It proved that SbtB-6803 was naturally trimer and no oligomeric state change happened between a TCEP-containing and a TCEP-free environment. The odd trimer band position on SDS-PAGE was likely due to incomplete polypeptide chain unfolding and less binding of SDS molecules to the residues.
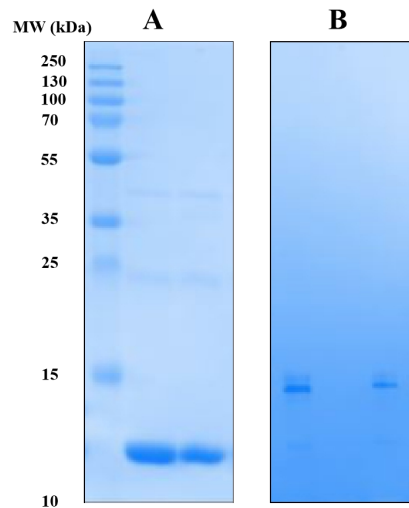
Figure 2-1: PAGE analysis of SbtB-6803. (A) SbtB-6803 purified without TCEP (middle lane) and with TCEP (right lane) during SEC purification was analyzed by SDS-PAGE followed by staining with Coomassie blue R-250. The resulting gel shows 3 bands from both lanes where the major bands are at around 12 kDa and faint bands are at around 24 kDa and 45 kDa. (B) SbtB-6803 purified without TCEP (left lane) and with TCEP (right lane) during SEC purification was analyzed by native PAGE followed by staining with Coomassie blue R-250. The resulting gel shows high level of pure and homogeneous SbtB-6803 from both lanes.

Initial screening experiments of SbtB-6803 in a non-reducing environment in sitting-drop 96-well plates produced several hits, with the largest crystals (Figure 2-2 A) coming from

a precipitant consisting of 0.2 M sodium citrate tribasic dihydrate, 0.1 M HEPES pH 7.5, 20% (v/v) isopropanol from the Crystal Screen HT kit (Hampton Research). Optimization of crystal size was performed using larger volumes, and a final condition with pH 7.2, 12% isopropanol, and supplemented with 10 mM $NiCl_2$ (Figure 2-2 B) was used for the growth of crystals to be used for diffraction experiments. Using these conditions, the crystals grew to a size of approximately 300 μm by 100 μm by 40 μm. Cryo-cooled SbtB-6803 crystals typically diffracted in the range of 2-3 Å, with the best crystal producing data which were processed to 1.8 Å (Figure 2-3 A, Table 2-1).

Initial screening experiments of SbtB-6803 in a reducing environment in sitting-drop 96-well plates produced several hits, with the largest crystals (Figure 2-2 C) coming from a precipitant consisting of 0.1 M Tris pH 8.5 and 2 M ammonia phosphate monobasic from the Crystal Screen HT kit (Hampton Research). Optimization of crystal size was performed using larger volumes, and a final condition with pH 8.6 and 2.25 M ammonia phosphate monobasic (Figure 2-2 D) was used for the growth of crystals to be used for diffraction experiments. Using these conditions, the crystals grew to a size of approximately 200 μm by 150 μm by 150 μm. Cryo-cooled crystals diffracted to 1.8 Å at the best (Figure 2-3 B, Table 2-1).
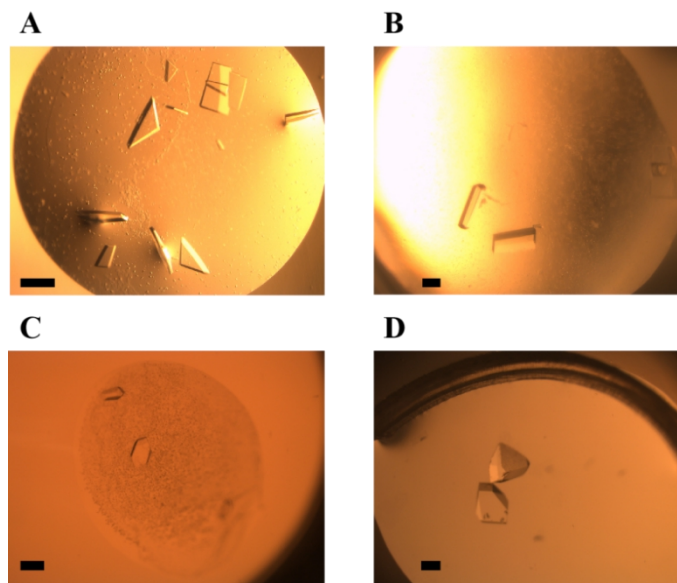
Figure 2-2: Images of SbtB-6803 crystals in a non-reducing environment grown in (A) 96-well sitting drop plates where crystals were trigonal plates, and optimized in (B) 24-well sitting drop plates where crystals were tetragonal rods. Crystals of SbtB-6803 in a reducing environment grown in (C) 96-well sitting drop plates where crystals were rods, and optimized in (D) 24-well sitting drop plates where crystals were diamond-like. Scale bar represents 50 μm.
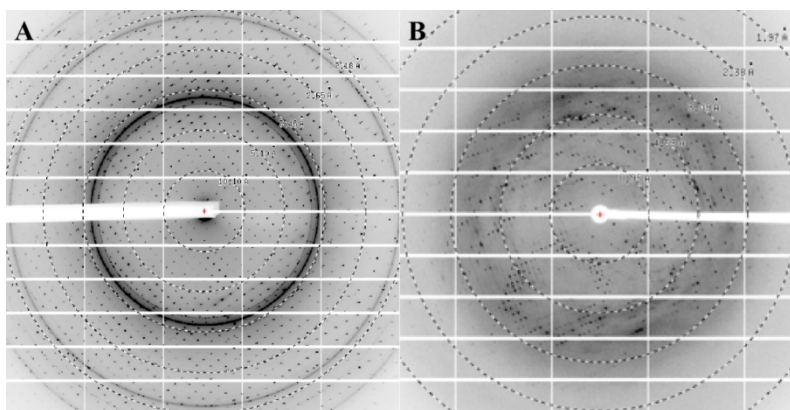
Figure 2-3: Representative diffraction patterns of SbtB-6803 (A) in the non-reducing environment at NSLS-II beamline AMX (17-ID-1) on an Eiger 9M detector at X-ray wavelength of 0.92009 Å, where diffraction spots are above 1.8 Å, and (B) in a reducing environment at SSRL beamline BL9-2 on a Pilatus 6M detector at X-ray wavelength of 0.9794 Å, where diffraction spots are above 2 Å.

Table 2-1: Diffraction statistics of SbtB-6803.

	Non-reducing	Reducing
Data processing		
Space group	$P4_32_12$	$P3_121$
Cell dimensions		
a, b, c (Å)	51.571, 51.571, 178.412	72.29, 72.29, 92.78
α, β, γ (°)	90, 90, 90	90, 90, 120
Resolution (Å)	50.0-1.80 (1.83-1.80)	62.6-2.01 (2.06-2.01)
R_{merge}	0.093 (0.448)	0.036 (1.272)
$CC_{1/2}$	0.987 (0.920)	1.000 (0.873)
I/σ_I	29.1 (4.3)	22.4 (2.0)
Completeness (%)	89.5 (96.2)	99.9 (100.0)
Multiplicity	7.9 (7.8)	9.9 (10.3)
Total reflections	165,824	189,957
Total unique reflections	20,957	19,180
Data refinement		
R_{work}/R_{free} (%)	18.64/20.50	25.48/28.76
RMSD bond (Å)	0.019	0.006
RMSD angle (°)	1.655	1.097
Ramachandran favored (%)	99.4	94.27

Values in parentheses refer to the highest resolution shell.

The X-ray diffraction data from rod crystals were processed to 1.8 Å in the space group P4$_3$2$_1$2 with the unit cell dimension of a=b=51.571 Å, c=178.412 Å, and α=β=γ=90° (Table 2-1). The diffraction data from diamond-like crystals were processed to 2.01 Å in the space group P3$_1$21 with the unit cell dimension of a=b=72.29 Å, c=92.78 Å, α=β=90° and γ=120° (Table 2-1).

The data were phased using a monomeric model derived from the structure of a putative P$_{II}$-like signaling protein from *Anabena variabilis* (PDB ID: 3dfe), which has 77.3% identity with SbtB-6803. After phasing and structure refinement, it was found that the asymmetric unit of this SbtB-6803 crystal form contained a dimer (Figure 2-4 A and B), where the dimeric interface between chain A and chain B consists of two anti-parallel beta sheets between of residues K30 to N36 from each monomer. Additionally, the final 12 residues of each SbtB monomer adopts an extended confirmation that is stabilized by interacting with the final 12 residues of another symmetry related SbtB dimer (Figure 2-4 C and D; Figure 2-5 A).
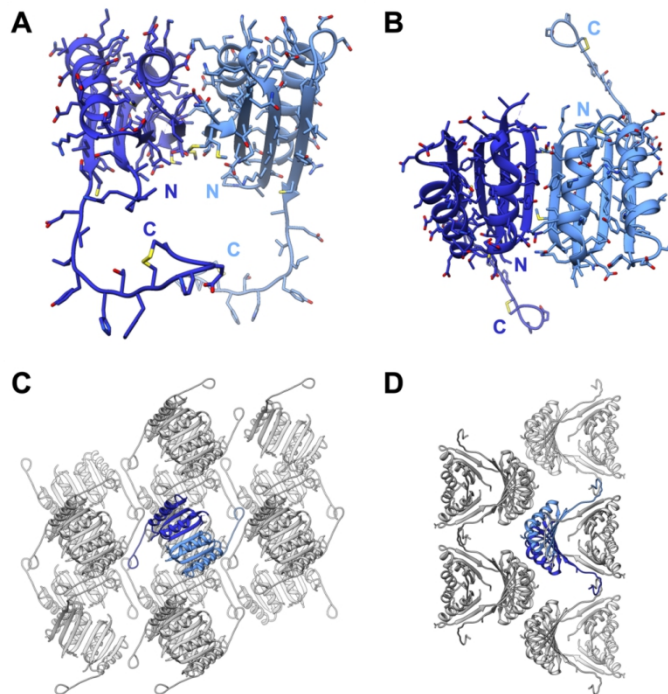
Figure 2-4: Arrangement of SbtB-6803 dimers within the crystal. (A-B) The asymmetric unit consists of two SbtB-6803 chains with an extended C-terminus. The dimer is colored two shades of blue, with the darker shade and lighter shade being chain A and B, respectively. When the asymmetric unit is viewed in the context of the entire crystal lattice it can be seen that the extended C-termini make contacts with adjoining dimers (C), and each dimer interacts closely with an adjacent dimer (D).

This new dimeric configuration was unexpected as native SbtB is a trimer and all recent structures of SbtB have either contained trimers within the asymmetric unit, or the trimer is formed through crystal symmetry [127, 129]. When analyzing the structure of this new crystal form of SbtB, it was further found that an intermolecular disulfide bond was formed between adjacent SbtB-6803 dimers (Figure 2-5 B). The disulfide bonds are formed between C94 residues, where C94 from chain A bonds with C94 from chain B of an adjacent asymmetric unit. Likewise, C94 from chain B forms a disulfide bond with C94 from chain A of the same adjacent dimer. These disulfide bonds and other interfacial

integrations between the SbtB-6803 dimers in adjoining asymmetric units leads to a tetrameric dimer of dimer organization (Figure 2-6).
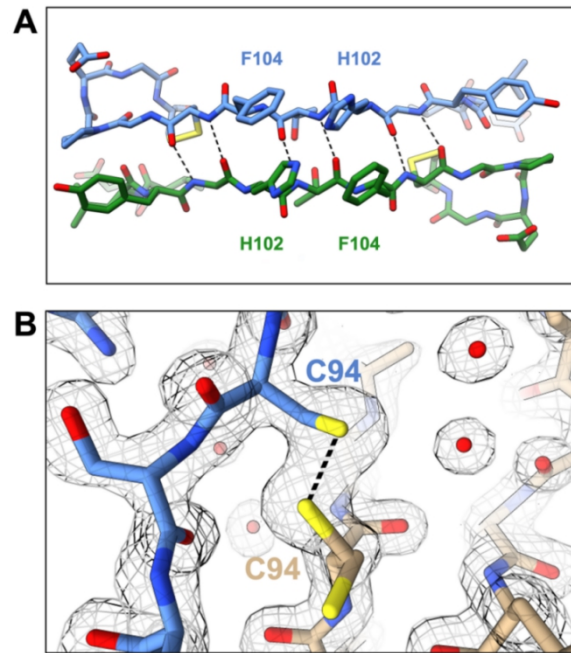
Figure 2-5: Detailed view of unique interactions of the tetragonal SbtB-6803 crystal form. (A) The C-termini, with a loop formed by intramolecular disulfide bonds, from two adjacent dimers interact through backbone hydrogen bonding. In this view, the blue model is chain B of our reference dimer while the green model is chain A from an adjacent dimer. Residues H102 and F104 are indicated in the image only as a reference, and not an indication of a special role for these residues in the formation of this interaction. (B) The intermolecular disulfide bonds are formed by residue C94 from two adjacent SbtB dimer interactions. In this panel, the blue model is again chain B from our reference dimer (same as in A) and the tan model is chain A from the nearby SbtB dimer (note that this chain A is different than in panel A). C94 from chain A is modeled with two conformations and occupancies were refined during refinement procedures. The main conformation of C94 takes part in the disulfide bond (76% occupancy) and the minor conformation is oriented away (24% occupancy). The 2Fo-Fc electron density map shown in panel B is contoured at 1.5σ and is displayed to show the density for the disulfide bond.
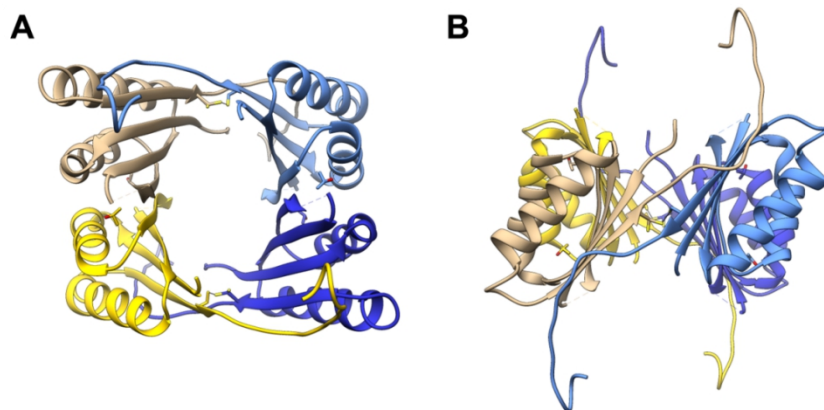
Figure 2-6: Views of the two SbtB-6803 dimers that are bonded by disulfides and form a tetrameric arrangement. The dark blue and light blue models are chain A and B, respectively, for the reference dimer. The tan and yellow models are chain A and B, respectively, of the adjacent disulfide bonded dimer.

From this study, it is not clear whether this disulfide bond formed after crystallization, and is simply a crystallization artifact, or if the disulfide bond and tetrameric organization is formed prior to crystallization which leads to this new crystal form. Because crystallization experiments were set up immediately after purification, and the final SEC trace for each purification showed a peak corresponding very well with the trimeric form of SbtB-6803, this new form is most likely a crystallization artifact. However, even if it is a non-native conformation of SbtB-6803, these findings indicate that C94 is susceptible to oxidization and disulfide bond formation. It is hypothesized that the regulation of SbtA-6803 by SbtB-6803 could be sensitive to redox conditions within the cell [127, 129]. The proposed redox-sensitive element of SbtB-6803, residues CGPEGC, is located at the extreme C-terminus of the protein, where C105 forms a disulfide bond with C110 to make a looped structure. This looped C-terminus can be seen in this structure as well (Fig 2-5 A). Based on the SbtB-6803 structure reported here, future functional investigations

of SbtB-6803 should also focus on the effect of C94 in addition to C105 and C110, when studying the role that oxidation and reduction of SbtB may play in regulating bicarbonate transport by SbtA.

Coordinates and structure factors for the tetragonal crystal form of SbtB-6803 have been deposited in the Protein Data Bank under accession code 6WUE.

The diffraction data of SbtB-6803 in reducing environment were phased using a published trimeric model of SbtB-6803. After phasing and structure refinement, a trimer was found in the asymmetric unit which shows a canonical P_{II}-like fold (Figure 2-7 A and B). The trimeric interface is composed of β strands which are organized in a coiled and closed form, and is maintained by the residues of T33-G39 and G89-V98. All three subunits in the refined structure are in same conformation. The C-terminal hairpin loop of each subunit formed by residues HTFCGPEGC is highly flexible, and the density can hardly be observed in the refined map (Figure 2-7 C). It is hypothesized that introduction of TCEP prevented the formation of disulfide bonds at the C-terminal loop and made it disordered. No disulfide bonds were found at the residue C94. As observed in the tetragonal crystal structure of SbtB-6803, the T-loop between residues K40 and E57 of each subunit is missing in the map.
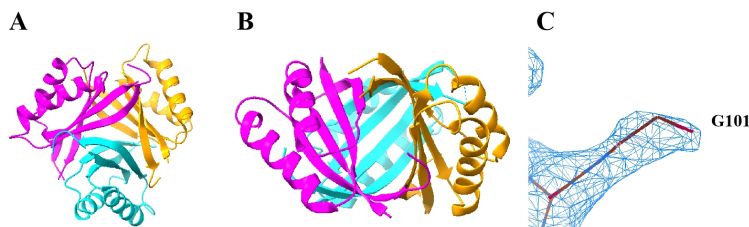
Figure 2-7: Structural representation of trimeric SbtB-6803. Overall architecture of SbtB-6803 from (A) top view and (B) side view shows a canonical P_{II}-like fold. (C) The 2Fo-

Fc electron density map of C-termini extends to G101, contoured at 1σ, while the hairpin loop formed by CGPEGC is disordered.

Since differences in the number of SbtB-6803 molecule in the asymmetric unit of yielded crystals were observed with and without reducing reagents, SEC was used for testing the protein molecular size when SbtB-6803 was incubated with various reducing or oxidizing reagents, including 2 mM TCEP, 2 mM CuCl₂, 2 mM reduced glutathione and 2 mM oxidized glutathione. As shown in Figure 2-8, all traces looked the same in terms of peak location, where each run had a single peak at around 36 kDa. Additionally, there was a small shoulder peak (about 25 kDa) from the control in which there was no reducing or oxidizing reagents. Besides, in other structural studies, SbtB-6803 was also determined as trimer with the C-terminal hairpin loop stabilized by the disulfide bonds. It demonstrates that SbtB-6803 is native in trimer, while the dimer of dimer form was likely to be a crystallographic artefact.
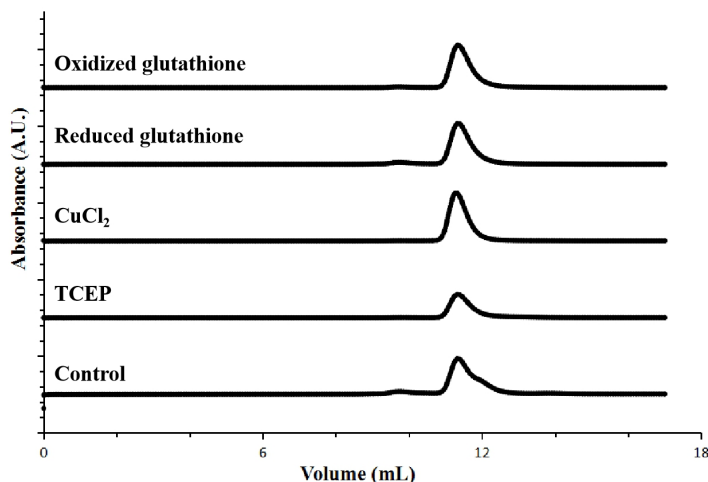
Figure 2-8: SEC characterization of SbtB-6803 incubated with various reducing or oxidizing agents: 2 mM TCEP, 2 mM CuCl₂, 2 mM reduced glutathione and 2 mM

oxidized glutathione. Each chromatogram aligned relative to the injection of the sample on the Superdex 75 Increase 10/300 GL column, which was set at 0 mL. SEC traces have not been normalized relative to each other in order to show the relative levels of the peak fractions from each run.

The monomeric model of SbtB-6803 acquired from both the crystal forms were compared in UCSF chimera (Figure 2-9 A) [141]. It was found that the secondary structures, such as α helices and β sheets, were similar and the major conformational changes appeared at the T-loop (Figure 2-9 B), N- and C-termini (Figure 2-9 C) in both crystal structures. Both N-termini (P4 to N6) and C-termini (C94 to C110) are more flexible as loops in the tetragonal crystal form than the trigonal crystal form. The conformational change at C94 is likely to be the reason for the disulfide bonds form at C94 between two monomer chains of SbtB-6803 from the adjacent asymmetric unit.
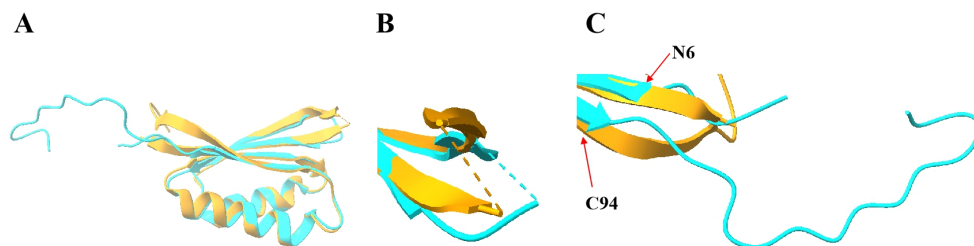
Figure 2-9: Comparison of SbtB-6803 monomeric models from the trigonal (orange) and tetragonal (cyan) crystal forms. (A) overlapped structural representations of monomeric SbtB-6803 from both the trigonal and tetragonal crystal forms, which shows that the major deviations appear at the T-loop, N-termini and C-termini. (B) Detailed view of the regions surrounding the flexible T-loop, where the T-loop are missing from both structures and conformational change occurs at the ends of T-loop. (C) Detailed view of the regions surrounding the N- and C-termini. In the trigonal crystal form of SbtB-6803, β sheets extend at the N-termini and close to C-termini, however in the tetragonal crystal form of SbtB-6803, both N-termini (P4 to N6) and C-termini (C94 to C110) are more flexible as loops, which may be why the disulfide bonds form at C94 between two monomer chains of SbtB-6803 from the adjacent asymmetric unit.

2.4 Conclusions

SbtB-6803 was highly expressed in *E. coli* BW25113, thus there has been no need for optimization of expression and purification conditions. From the sitting drop vapor diffusion experiments, the tetragonal crystal forms are obtained and diffracted to 1.8 Å during single crystal X-ray diffraction. This crystal form contains a dimer of SbtB-6803 in the asymmetric unit, which is not a canonical form of P_{II}-like signaling proteins. Disulfide bonds form at the C-terminal hairpin loop between C105 and C110, along with C94 and another C94 from adjacent asymmetric unit. Therefore, SbtB-6803 is arranged as dimer of dimer in this crystal form. In addition, we have hypothesized that the non-canonical packing of SbtB-6803 is a result of disulfide bonds and introduced reducing agent TCEP at 5 mM later during the crystallography experiments. From the sitting drop vapor diffusion experiments, the trigonal crystal forms are obtained and diffracted to 2.01 Å during single crystal X-ray diffraction. This crystal form contains a trimer of SbtB-6803 in the asymmetric unit, which a canonical form of P_{II}-like signaling proteins. The structure of trimeric SbtB-6803 is very similar to the published structure by other groups [129], except that the C-terminal loop is too flexible to be seen in the map. When comparing the monomeric structure of SbtB-6803 from both crystal forms, it is found that significant conformational change between them occurs at the N- and C-termini, where the termini are less rigid in the tetragonal crystal form. The opening of C94 is likely to promote the formation of intermolecular disulfide bonds and the dimer of dimer spacial arrangement.

CHAPTER 3: STUDIES OF THE CYANOBACTERIAL BICARBONATE TRANSPORTER SBTA AND ITS INTERACTION WITH SBTB

3.1 Introduction

Bicarbonate transporter SbtA is an important component of cyanobacterial CO_2 concentrating mechanism. It is a Na^+-dependent bicarbonate transporter with low flux rate and high affinity [16]. However, little is known on its fold and structure. A precious study on membrane topology of SbtA from *Synechocystis sp.* PCC 6803 (SbtA-6803) shows that SbtA-6803 is inserted into the plasma membrane with 10 transmembrane helices (TMHs) and adopts a 5+5 inverted repeats. The termini of SbtA are located in the periplasm, while the termini of other membrane proteins with 5+5 inverted repeats are located in the cytoplasm [23-26]. These findings make SbtA unique and intriguing subject for structural and functional analysis.

Initial study suggested that SbtA existed as a tetramer on a native gel [28], however another study on interaction between SbtA and SbtB postulated that the tetramer band could be misled by the complex of SbtA trimer and SbtB trimer, as the size of SbtB trimer is nearly the same as an SbtA monomer [27]. SbtB is a P_{II}-like signaling protein, and another P_{II} homolog protein GlnK trimer binds to an *E. coli* ammonia channel AmtB trimer, and hence it is hypothesized that SbtB binds to SbtA in a similar way [27, 142]. In the crystal structure of AmtB-GlnK complex, an arginine-containing T-loop of each ADP-bound GlnK monomer is inserted to the AmtB monomer, and blocked the ammonia transport via AmtB [142]. Other studies showed that the association of SbtA and SbtB from *Cyanobium*

sp. PCC7001 and *Synechococcus elongatus* PCC7942 is depended on the relative adenosyl nucleotide ratio, and formed under high level of ADP or AMP over ATP [127, 128]. However, in order to facilitate fundamental studies of SbtA function and future applications, robust expression and purification procedures are first needed. With the systematic development of methodologies for enhanced heterologous expression and purification of SbtA from *Synechocystis sp.* PCC 6803 (SbtA-6803) in *E. coli*, the yields of purified protein reached 0.1 ± 0.04 mg per liter of culture. Besides, this dissertation shows that SbtA-6803 is purified as trimer and provides preliminary results on the SbtA-SbtB complex.

3.2 Materials and Methods

3.2.1 Expression Vector and Strains

The gene encoding full length SbtA-6803 (i.e., *slr1512* from *Synechocystis sp.* PCC 6803) was constructed with N-terminal 10x-His tag and thrombin cleavage site (Integrated DNA Technologies), and fused into pET-16b vector using Gibson assembly [132] to give the plasmid pSbtA-16N. The plasmid was confirmed by DNA sequencing, and transformed into following *E. coli* strains: BL21 (DE3), BL21 (DE3) pLysS, KTD212, C41 (DE3), Rosetta2 (DE3), C43 (DE3) and C43 (DE3) Rosetta2.

3.2.2 Screening Whole Cell Expression by Dot Blot

Overnight cultures were prepared in Luria broth (LB) with 50 µg/mL carbenicillin at 37 °C (34 µg/mL of chloramphenicol was also used for BL21 (DE3) pLysS, Rosetta2 (DE3) and C43 (DE3) Rosetta2), and used to inoculate 50 mL of fresh auto-induction media which is the same overnight culture media supplemented with 0.5% glycerol, 0.5

g/L glucose, and 1 mM IPTG [143]. Cultures were grown overnight at 37, 30 or 18 °C. Following overnight expression, the optical cell density of the culture was measured at 600 nm (OD600). Cell pellets from 1 mL of each culture were harvested at 4,000 x g and re-suspended with a buffer solution, ensuring that cell density of each sample was normalized based on OD600. Re-suspension buffer consisted of 62.5 mM Tris-HCl pH 6.8, 5 mM DTT, 25% glycerol, 1 mM PMSF and 2% SDS. Samples were applied to a nitrocellulose membrane which was then blocked with 5% milk in TBST and incubated with 6x-His tag monoclonal antibody (Invitrogen) followed by anti-mouse IgG AP-linked antibody (Cell Signaling Technology) in 5% milk in TBST. The blots were visualized with BCIP/NBT substrate (Promega). Blots were quantified by measuring the intensities digitally scanned by a Gel Doc EZ system (Bio Rad) and normalized based on the maximum intensities measured for each set of experiments.

3.2.3 Protein Purification and Detergent Screening

The best strain harboring pSbtA6803-16N were grown in 4 L of auto-induction media at the optimal temperature. Following overnight growth, cells were harvested and re-suspended in 50 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA and 1 mM PMSF, and then lysed by sonication (QSonica). Unbroken cells and cell debris were removed by centrifuging at 10,000 x g. The membrane pellet was harvested by centrifuging at 105,000 x g, washed with 20 mM Tris-HCl pH 8.0, 1 M NaCl, 20 mM KCl, and 10 mM MgCl$_2$ and then harvested at 105,000 x g. The pellet was solubilized in 20 mM Tris-HCl pH 7.5, 1 M NaCl, 10% glycerol, and detergent for 3 h. Solubilization buffer was supplemented with one of the following detergents: 2% DM, 1% DDM or 1% LMNG. Following solubilization, all buffers with detergent contained 0.2% DM, 0.02% DDM or

0.01% LMNG. Unsolubilized membranes were removed by centrifuging at 105,000 x g, and then solubilized SbtA in the supernatant was initially purified by immobilized metal-ion affinity chromatography (IMAC) using a HisTrap HP column (GE Healthcare). After extensive washing with 20 mM Tris-HCl pH 7.5, 1 M NaCl, 10% glycerol, 75 mM imidazole, and detergent, protein was eluted in 20 mM Tris-HCl pH 7.5, 1 M NaCl, 10% glycerol, 500 mM imidazole, and detergent. Protein was later subjected into a Superdex 200 Increase 10/300 GL column (GE Healthcare) pre-equilibrated in the same buffer solution. SDS-PAGE was carried out to assess degree of purification after each step while ImageJ [144] was used for quantitative analysis of the gels. Protein concentrations were determined on a Nanodrop (Thermo Scientific) using absorbance at 280 nm.

To cleave the His tag, instead of protein elution from HisTrap HP column after extensive column wash, the column was washed with 20 mM Tris-HCl pH 7.8, 150 mM NaCl, 10% glycerol, 3 mM $CaCl_2$ and detergent and injected with 10 units of thrombin in the same buffer solution per milligram of His-tagged proteins. The column was placed at room temperature for overnight with gentle rotation. Cleaved sample was obtained from the HisTrap HP column flow-through fractions by loading 10 mM HEPES pH 8, 200 mM NaCl and the corresponding detergent, and subjected into a Superdex 200 Increase 10/300 GL column (GE Healthcare) pre-equilibrated in the same buffer solution. The peak fraction eluted at around the right size was collected.

SEC purified sample in detergent could also be switched to amphipols by incubating the samples with amphipol A8-35 (Anatrace Products, LLC) in the mass ratio of 1/5 for 30 min at 4 °C, followed by incubation with 10 mg bio-beads for 2 h at 4 °C to remove excess amount of detergent. Sample was then subjected into a Superdex 200 Increase

10/300 GL column (GE Healthcare) pre-equilibrated in 10 mM HEPES pH 8 and 200 mM NaCl. The peak fraction eluted at around the right size was collected.

3.2.4 Circular Dichroism (CD) Spectroscopy

300 µL of sample (around 0.9 mg/mL) purified with the best detergent was placed into a 1 mm quartz cuvette (Starna Cells; Atascadero, CA) in SEC buffer. Multi-wavelength CD measurements were taken every 2 nm from 262 to 190 nm. All measurements were taken on a Jasco J-815 CD Spectrometer (Jasco Inc., Easton, MD) and processed with the Jasco Spectra Manager V1.54.03 software package.

3.2.5 Preparation of SbtA-SbtB Complex

The tag-cleaved SbtA-6803 was mixed with excess amounts of purified tag-cleaved SbtB-6803 (described in chapter 2). The mixture was diluted in 10 mM HEPES pH 8, 200 mM NaCl, 0.5 mM $CaCl_2$, 1 mM adenosine nucleotide and detergent (if used), incubated for 1 h at room temperature, and subjected into a Superdex 200 Increase 10/300 GL column (GE Healthcare) pre-equilibrated in the same buffer.

3.2.6 SDS-PAGE

SDS-PAGE was performed on 4% stacking gels followed by 15% resolving gels. Protein solutions were mixed with equal volume of 62.5 mM Tris-HCl pH 6.8, 5 mM DTT, 25% glycerol, 0.1% bromophenol blue and 0.2% SDS at room temperature for at least a day. The protein bands were visualized by 0.1% Coomassie blue R-250.

3.2.7 Membrane Protein Crystallization

Purified samples using the most suitable detergent were concentrated to 10-25 mg/mL using a Vivaspin protein concentrator (GE Healthcare) with 100,000 molecular weight cut-off, and mixed with molten monoolein at a ratio of 2/3 (v/v) in a dual-syringe mixer

until a transparent lipidic cubic phase (LCP) was formed. LCP was disperse onto a glass sandwich set as a bolus of 40 nL and covered with 1 uL of crystallization solution using a Crystal Gryphon robot (Art Robbins Instruments). The glass sandwich sets were incubated at 20 °C and loaded into a Rock Imager (Formulatrix) for monitoring crystal growth.

3.2.8 Single Particle Electron Microscopy

For negative-stain EM analysis, EM grids were prepared by applying 2 µL of cleaved sample onto FCF300-Cu grids glow discharged for 1 min on carbon side. Grids were washed twice with 10 mM HEPES pH 8 and 200 mM NaCl and stained in 0.75% uranyl formate. EM micrographs were recorded on a TVIPS TemCam-F416 camera (Gatan) with a Tecnai F20 electron microscope (Thermo Fisher) operated at 200 kV. Images were processed in cisTEM [145].

For cryo-EM analysis, Quantifoil R 2/1 Cu 300-mesh grids (Electron Microscopy Sciences) were pre-treated by either soaking in the SEC buffer for 1 h (for detergent-purified sample) or glow discharging for 15 s on both sides (for amphipols-stabilized sample). 6 µL of either detergent-purified or amphipols-stabilized sample was applied onto the grids. Grids were blotted for 3 or 6 s in a Vitribot Mark IV (Thermo Fisher Scientific) at room temperature and 100% humidity, then plunged into liquid ethane and stored in liquid nitrogen. Grids were imaged on a K2 direct electron detector (Gatan) with a Titan Krios electron microscope (Thermo Fisher) operated at 300 kV. Movies were recorded by SerialEM [146] at a nominal magnification of 22,500 x, which was calibrated to a pixel size of 1.03 Å. Movies were motion corrected and aligned as images using

cryosparc [147]. Contrast transfer function (CTF) estimation, 2D classification, ab-initio 3D reconstruction and 3D refinement were processed in cisTEM [145].

3.3 Results and Discussions

Initial expression efforts performed using in BL21 (DE3) at 37 °C failed to yield high levels of SbtA-6803 (data not shown). Therefore, to identify suitable expression strains and conditions for yielding increased levels of SbtA-6803, several small-scale expression tests in LB were next performed employing a range of different host strains and temperatures. Specific host strains of interest were selected based on their use in previous studies for enhancing membrane protein over-expression [148]. Comparing relative levels of produced His-tagged SbtA-6803, overall expression yields were found to be low for most conditions, and high in cells harboring pLysS at 30 °C (relative to 37 and 18 °C). This indicates that pLysS is determinantal to higher expression level of SbtA-6803 among all the strains selected for tolerance towards toxic protein over-expression (C41 (DE3), C43 (DE3) and KTD212). The gene encoding SbtA-6803 was not codon optimized in this study, and Rosetta2 (DE3) cells showed a substantial improvement in SbtA-6803 expression than BL21 (DE3) at 37 °C, indicating that codon bias might play a role in the expression levels of SbtA-6803 in *E. coli*. Overall, the strain that produced the highest levels of SbtA-6803 was BL21 (DE3) pLysS (Figure 3-1).
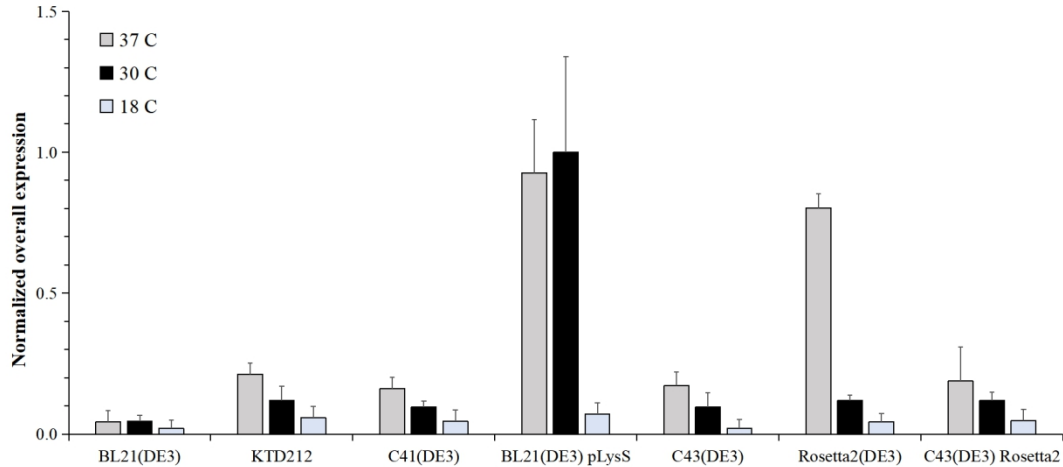
Figure 3-1: Screening expression strains and temperature for production of SbtA-6803. Levels of SbtA-6803 in whole cell samples were assayed by dot blot (n=3) and normalized relative to the highest expression condition.

Following the identification of conditions supporting SbtA-6803 production, we identified the best detergent for improving membrane solubilization. Three different detergents (DM, DDM and LMNG) were screened for this purpose. BL21 (DE3) pLysS cells harbouring pSbtA6803-16N were harvested and lysed before membranes were then collected by high-speed centrifugation. Isolated membranes were split into three fractions, each individually treated with one of four detergents of interest in subsequent solubilization and IMAC purification steps. The amount of solubilized His-tagged SbtA-6803 could be estimated by the areas of elution peaks, since the His tag is long enough to be exposed to the resin. As shown in Figure 3-2 A, DM failed to extract enough SbtA-6803 from membrane fractions, which was shown by an absorbance plateau following protein elution using buffer supplemented with 500 mM imidazole. DDM was found to support the extraction of SbtA-6803, however LMNG could extract the highest amount of

SbtA-6803 and would be the best detergent for SbtA-6803 solubilization from membrane fractions.

To determine the optimal detergent for SbtA-6803 purification, SEC was performed on each sample (Figure 3-2 B) following initial purification by IMAC. A single peak at around 300 kDa occurred at the profile of purification with LMNG, which seemed to be much larger than expected 120 kDa. However, it is normal for membrane proteins to behave as larger proteins during SEC purification owing to the additional size from detergent micelles. Purification with DDM showed a primary peak at around 300 kDa, while there were two other tiny peaks at around 125 kDa and 70 kDa, which could represent the dimer and monomer. SEC purification with DM was performed by incubating the IMAC-purified SbtA-6803 using LMNG with the SEC buffer supplemented with 0.2% DM for 30 min prior to purification by SEC since DM failed to extract SbtA-6803 from the membrane fractions, and the SEC trace showed 3 peaks at around 300 kDa, 100 kDa and 40 kDa, likely standing for trimer, dimer and monomer. Therefore, LMNG also supported the purification of intact SbtA-6803 and was selected as the most suitable detergent.
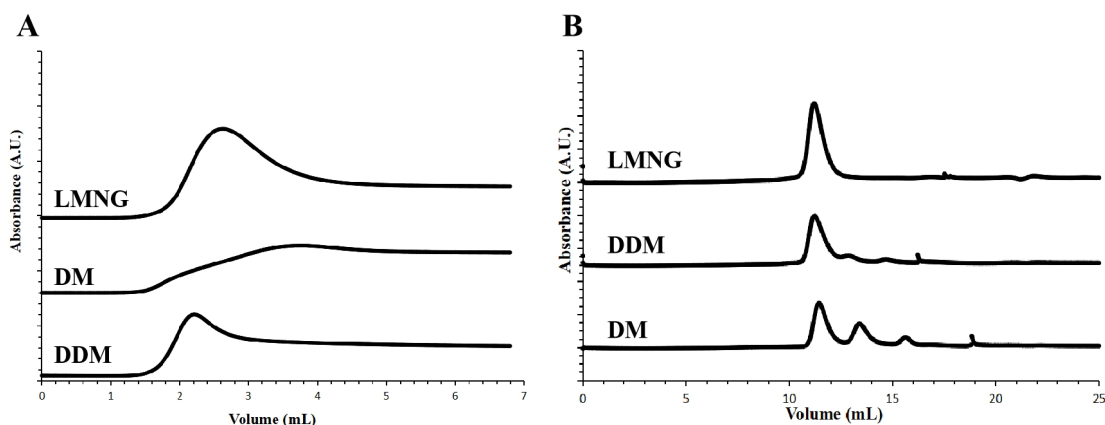
Figure 3-2: Chromatography characterization of detergent solubilized and purified SbtA-6803 from (A) Ni^{2+}-NTA elution fractions where each chromatogram aligned relative to the beginning of the protein elution on the HisTrap HP column, which was set at 0 mL; and from (B) SEC where each chromatogram aligned relative to the injection of the protein solution on the Superdex 200 Increase 10/300 GL column, which was set at 0 mL. Traces have not been normalized relative to each other in order to show the relative levels of the peak fractions from each detergent.

SDS-PAGE result of SbtA-6803 following SEC purification with detergent LMNG or amphipol A8-35 (Figure 3-3 A) each showed a band at around 39 kDa, which is a bit smaller than the expected size of 42 kDa, however such deviation is acceptable for membrane proteins [149, 150]. By this approach, final yields of SbtA-6803 purified with LMNG reached 0.1 ± 0.04 mg per liter of culture; a substantial improvement for SbtA-6803 over initial expression efforts using BL21 (DE3) which yielded undetectable level. Besides, negative-stain EM images of SbtA purified with LMNG (Figure 3-3 B) or stabilized with A8-35 (Figure 3-3 C) showed well-dispersed and homogeneous particles, with the average particle dimension measured at approximately 120 Å or 80 Å by ImageJ [144], respectively. The difference in the average particle dimension between purification with LMNG and stabilization with A8-35 could be an artefact, while it was also likely that LMNG was not stripped from the particles during EM grid preparation since due to

51

its extremely low critical micelle concentration. The negative-stain EM data were processed in cisTEM [145]. After running 2D classification without imposing symmetry, class averages of SbtA-6803 purified with LMNG (Figure 3-3 D) or stabilized with A8-35 (Figure 3-3 E) predominantly presented the shape of triangles, indicating the trimeric conformation. Nevertheless, particles adopted preferred orientation on the carbon film and the side view was missing, it was difficult to build initial low-resolution 3D maps to make templates for cryo-EM study. In light of these results, SbtA-6803 was purified as trimer using both detergent and amphipol, and hence should be trimer in the native form.
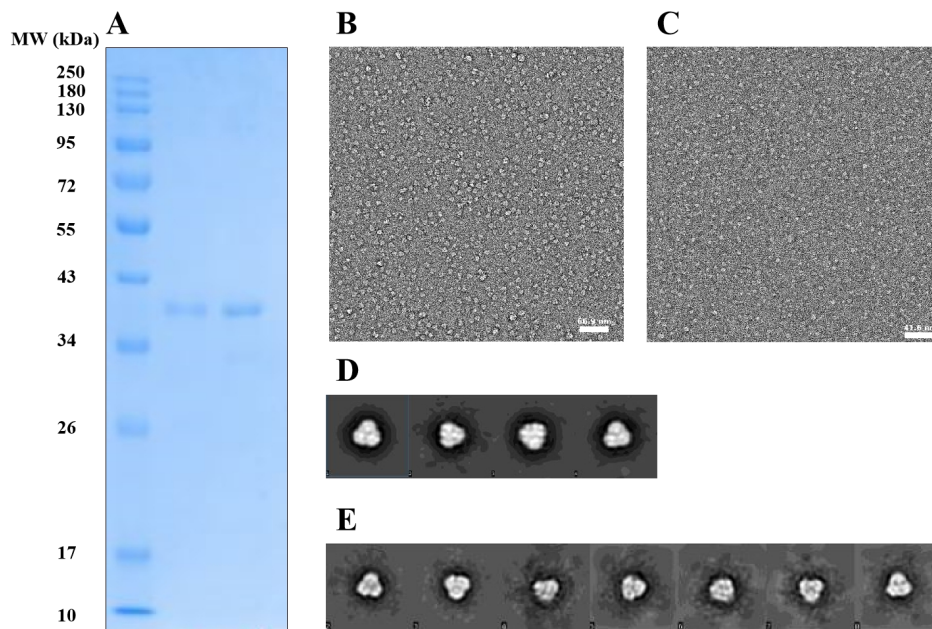
Figure 3-3: The purity and homogeneity of SbtA-6803. (A) SbtA-6803 purified with LMNG (middle lane) or stabilized with A8-35 (right lane) following SEC purification was analyzed by SDS-PAGE followed by staining with Coomassie blue R-250. The resulting gel shows a band at around 39 kDa from each detergent purification. Representative negative-stain EM image of SbtA-6803 purified with LMNG (B) or stabilized with A8-35 (C) following SEC purification shows the average particle dimension of 120 Å or 80 Å, respectively. Representative 2D class averages of negative-stain EM data of SbtA-6803 purified with LMNG (D) or stabilized with A8-35 (E)

showed that SbtA-6803 particles were predominantly triangular, matching the shape of trimer.

Additionally, CD spectroscopy was used to assess the secondary structure of SbtA-6803 purified using LMNG (Figure 3-4). The resulting CD spectrum shows a broad absorption peak between 210 and 220 nm, and hence indicates high level of α-helical content for the purified protein, which is expected as SbtA-6803 is an integral membrane protein composed mostly of α helices.
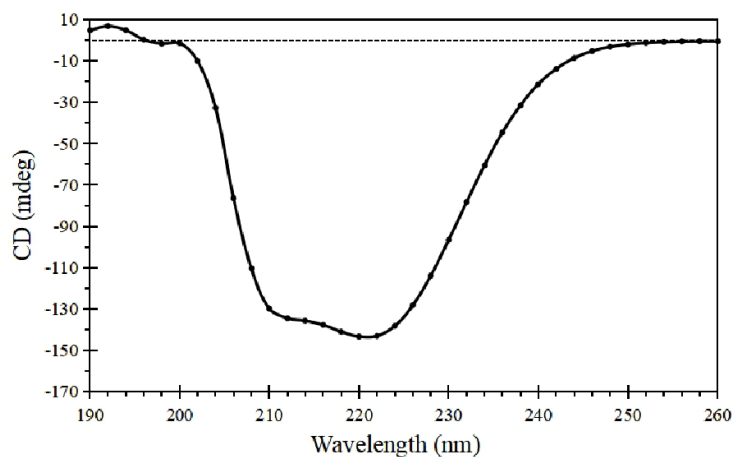
Figure 3-4: SbtA-6803 immediately following IMAC and SEC purification was analyzed by CD spectroscopy. The data indicate that purified SbtA-6803 is predominantly α-helical.

Purified SbtA-6803 following SEC purification was initially used for crystallography after examining its quality by negative-stain. For LCP crystallization, hit condition for crystals (Figure 3-5) was found at the crystallization solution containing 100 mM Tris-HCl pH 8.5, 100 mM MgCl$_2$, 100 mM NaCl, 10 mM L-glutathione reduced and 30% (v/v) PEG 400, although they were too tiny to be clearly visualized. Optimized screening of protein concentration and crystallization solution recipe failed to produce larger crystals.

These microcrystals were useless for the common single crystal X-ray diffraction in a synchrotron facility since they are too small to withstand the radiation damage from the X-ray beam and support complete high-resolution diffraction data collection. Therefore, alternative approaches, by either applying a different technique to diffracting these microcrystals or single particle cryo-EM, needed to be used for its structure determination.
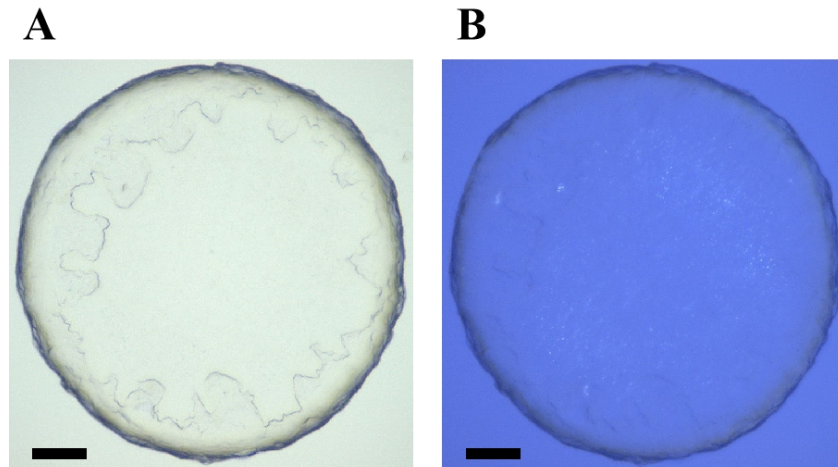
Figure 3-5: Potential SbtA-6803 crystals grown in LCP. (A) Representative LCP visualized with the normal light. (B) the same LCP visualized with the cross-polarizing light shows the signals of microcrystals (tiny white dots). The scale bars represent 100 μm.

Single particle analysis was applied then. During the first round of screening in the cryo-TEM using the sample at 1.3 mg/mL, only a small number of particles could be visualized per recorded image of the holes, whereas more particles were aggregated on the carbon films. To optimize the grid preparation, the grids were soaked in the SEC buffer supplemented with 0.001% (i.e. 1x the critical micelle concentration) LMNG for 30 min prior to sample application. Hereby, many particles stayed in the holes (Figure 3-

6 A) because the carbon films were partially saturated by the detergent molecules. A total of 1,756 images were collected and 256,734 particles were picked by cisTEM [145] for running 2D classification. As shown in the class averages (Figure 3-6 B), both top and bottom views of SbtA-6803 were captured and trimeric conformation was obviously seen. However, the center of particles was found to be hallow, which is abnormal for transporter proteins and indicates that particles were expanded. After a few other optimizations without significant improvement on the particle quality, amphipol A8-35 was used for stabilizing purified SbtA-6803 and cryo-EM study. A number of particles were observed in the holes of EM grids (Figure 3-6 C) pre-treated with the standard glow-discharging rather than incubating with detergent. A total of 3,455 images were collected and 1,778,627 particles were picked by cisTEM [145] for running 2D classification. As shown in the class averages (Figure 3-6 D), particles look more compact, though the alignment is not better compared with the data from LMNG purified sample. This may be explained the monomers were rotated and shifted during replacing LMNG with A8-35 since particles were expanded in the LMNG micelles. Signals predicting the secondary structure were absent from neither class averages, suggesting that the particles were not well aligned. As a result, the final maps of both data after running ab-initio 3D recontruction and 3D refinement showed that the highest resolution only reached approximately 10 Å, determined based on the Fourier shell correlation (FSC) value at 0.143.
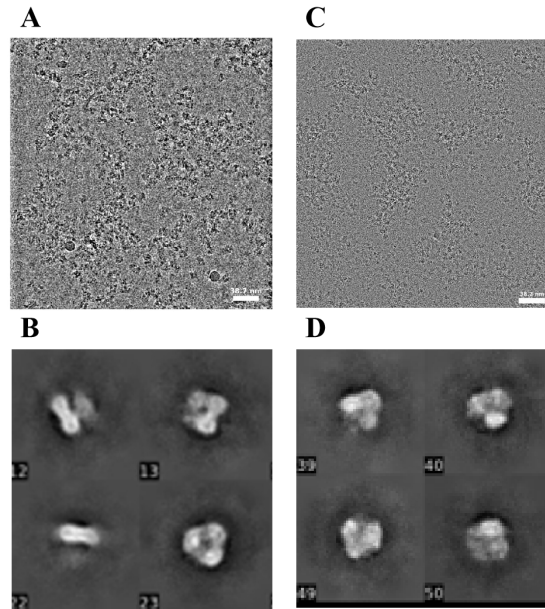
Figure 3-6: Single particle cryo-EM analysis of SbtA-6803. Representative image of plunge-frozen SbtA-6803 particles either (A) purified with LMNG or (C) stabilized with A8-35 collected from the K2 direct electron detector. Representative class averages of SbtA-6803 either (B) purified with LMNG or (D) stabilized with A8-35.

Because SbtA-6803 is like a flat disk, it was difficult to align the particles at the side views, perpendicular to or intersecting with the membrane plane. If SbtB can bind to SbtA either at the top or bottom, the typical way for a signaling protein regulating the activity of a transporter protein. It is hypothesized that particles displaying the side views can be better aligned during processing and spacial arrangement of SbtA-6803 trimer can be stabilized. From chapter 2, two different crystal structures of SbtB-6803 were determined, and offered references for data processing and structural analysis. Besides, a hypothesis was carried out that the C-terminal cysteine can play a role in the formation of SbtA-SbtB complex. In addition, SbtB-6803 shares similar fold to P_{II} signaling proteins, which bind to their receptors with the assistance of adenosyl nucleotides [127, 128]. To test which factor is important for the complex formation, broad screening was performed by

SEC purification and SDS-PAGE, and purified SbtA-6803 and SbtB-6803 were mixed and incubated with each one of the adenosyl nucleotides and with or without the addition of TCEP for 1 h prior to loading into the Superdex 200 SEC column. The affinity tags were cleaved from both proteins and excess amount of SbtB-6803 was introduced to facilitate the binding. SEC peaks from each experiment were collected and concentrated for SDS-PAGE. To make a comparison, SbtA-6803 and SbtB-6803 were separately loaded into the same SEC column for controls. As shown in the gel images (Figure 3-7), SbtB bands were visualized from the conditions involving ADP without TCEP, AMP without TCEP, and AMP with TCEP. Other conditions such as no adenosyl nucleotides included, supplemented with ATP, supplemented with cAMP, or supplemented with ADP and TCEP failed to produce any complex of SbtA-SbtB. These results match other studies of the interaction between SbtA and SbtB from other organisms or *Synechocystis sp.* PCC 6803 [127, 128]. A few bands other than standing for the monomer of SbtA-6803 and SbtB-6803 also appeared on the gel, which could be the results of overloaded thrombin during cleaving the affinity tag of SbtA-6803. Thrombin is known to cleave the amino acid residue sequence between arginine (R) and glycine (G), if overloaded, and there is a sequence of RG in one of the transmembrane helices of SbtA-6803. However, the amount of thrombin used for proteolysis can be optimized in order to maintain the desired specific sequence cleavage. In light of these findings, including ADP without TCEP was chosen as the experimental condition for generating the SbtA-SbtB complex for single particle analysis, based on an assumption that the higher ADP/ATP during the dark cycle promotes binding of SbtB to SbtA and hence blocks the bicarbonate transport activity of SbtA.
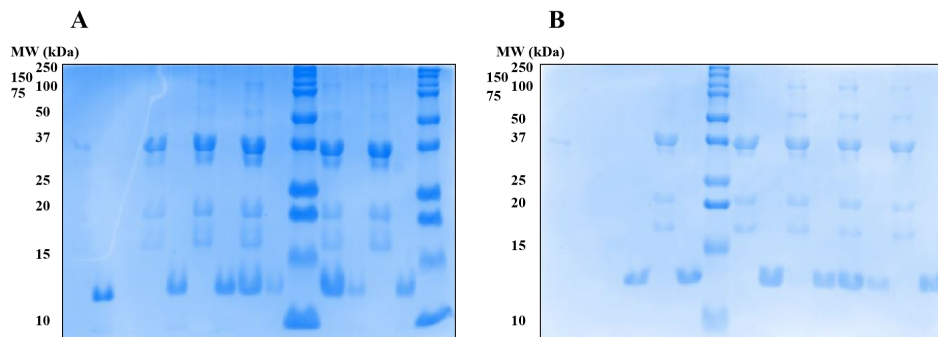
Figure 3-7: The gel images of testing SbtA-SbtB conditions following SEC purification. Purified SbtA-6803 and SbtB-6803 were mixed with or without the addition of TCEP or adenosyl nucleotides 1 h prior to SEC. (A) Samples without introducing TCEP, from left to right on the gel: SbtA peak from control, SbtB peak from control, blank, potential SbtA-SbtB peak from experiment without adding nucleotides, SbtB peak from experiment without adding nucleotides, potential SbtA-SbtB peak from experiment with adding ATP, SbtB peak from experiment with adding ATP, potential SbtA-SbtB peak from experiment with adding ADP, SbtB peak from experiment with adding ADP, molecular weight marker, potential SbtA-SbtB peak from experiment with adding AMP, SbtB peak from experiment with adding AMP, potential SbtA-SbtB peak from experiment with adding cAMP, SbtB peak from experiment with adding cAMP, molecular weight marker; (B) Samples incubated with TCEP, from left to right on the gel: SbtA peak from control, blank, SbtB peak from control, potential SbtA-SbtB peak from experiment without adding nucleotides, SbtB peak from experiment without adding nucleotides, molecular weight marker, potential SbtA-SbtB peak from experiment with adding ATP, SbtB peak from experiment with adding ATP, potential SbtA-SbtB peak from experiment with adding ADP, SbtB peak from experiment with adding ADP, potential SbtA-SbtB peak from experiment with adding AMP, SbtB peak from experiment with adding AMP, potential SbtA-SbtB peak from experiment with adding cAMP, SbtB peak from experiment with adding cAMP.

SEC purification was performed to collect the complex sample of LMNG-purified SbtA-6803 and SbtB-6803, and the collected sample was concentrated to 1.3 mg/mL using a Vivaspin protein concentrator (GE Healthcare) with 10,000 molecular weight cut-off in order to not lose the SbtB-6803 during protein concentrating. Few particles were observed in the holes during screening in the cryo-TEM (Figure 3-8 A), so SbtA-6803 was stabilized with amphipol A8-35 prior to incubating with SbtB-6803 followed by SEC.

The collected sample after SEC purification was also concentrated to 1.3 mg/mL for grid preparation, and many particles stayed in the holes during screening in the cryo-TEM (Figure 3-8 B). A total of 5,445 movies were collected for a large data set, and 3,741,335 particles were picked for the first round of 2D classification. Side views showing well aligned secondary structure information were obtained from some 2D class averages, indicating good alignment from this view, although the top and bottom views were not clear and the number of particles were relatively small (Figure 3-8 C). From the side views, it is clear to see two different features with the dimensions measured at approximately 65 x 40 Å and 40 x 25 Å for the big and small feature, respectively. The dimensions of small subunit are close to the size of SbtB-6803 trimer. These findings show the preliminary interaction between SbtA-6803 and SbtB-6803, which are both assumed to be trimeric conformation. However, only 23,668 particles were used for ab-initio 3D reconstruction and 3D refinement after filtering the junk particles, and the resulting map was resolved to approximately 12 Å, which is insufficient to visualize the details of side chain positions and most secondary structures.
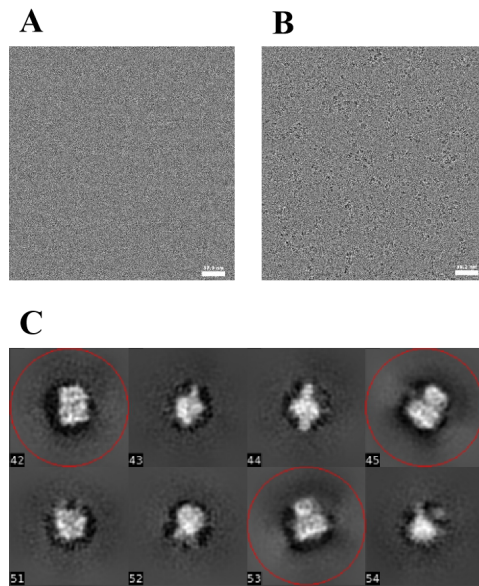
Figure 3-8: Single particle analysis of SbtA-SbtB following SEC purification. (A) Representative image of samples purified in LMNG where little particles could be seen in the hole. (B) Representative image of samples stabilized in A8-35 where lots of particles could be seen in the hole. (C) Representative 2D class averages of data from (B) showing two features with well aligned secondary structure information from the side views (highlighted in red circles). The dimensions of big and small feature were measured at approximately 65 x 40 Å and 40 x 25 Å, respectively, close to the size of predicted SbtA-6803 and SbtB-6803 trimers.

3.4 Conclusions

As with many membrane proteins, low levels of expression can limit the structural and functional study of SbtA-6803. Although broad expression conditions have been tested in *E. coli*, the final yield of purified sample using the most suitable detergent reached 0.1 ± 0.04 mg per liter of BL21 (DE3) pLysS cell culture, which is sufficient for single particle EM and LCP crystallization but not for vapor diffusion method. SEC elution profiles, meanwhile, confirm that LMNG helps to keep SbtA-6803 in a stable form. Besides, the SEC elution time of SbtA-6803 along with results from negative-stain EM indicates that the herein developed protocol is capable of yielding stable, homotrimeric complexes.

60

Besides, Rosetta2 (DE3) strain has also shown improvement in the final yield of SbtA-6803 at 37 °C. Therefore, in the future work, Rosetta2 (DE3) pLysS strain can be beneficial in further increasing the yield. Besides, among the multiple experiments about generating the SbtA-SbtB complex with or without the addition of various adenosyl nucleotides and TCEP, three conditions have been identified and the condition involving ADP without TCEP has been selected for experiments.

Single particle cryo-EM was used for both SbtA-6803 and SbtA-SbtB complex. Although SbtA-6803 is too small to be properly aligned for data processing, SbtA-SbtB is well aligned during 2D classification and the features of SbtA-6803 and potential trimeric SbtB-6803 shown with the secondary structure information are clear on the class averages. However, due to the small number of particles after filtering out the bad particles, the resolution of the final map is too low to investigate the molecular details. Therefore, more data need to be collected.

Microcrystals were grown in LCP, although the are too small to withstand the radiation damage by the X-ray beam during the X-ray diffraction data collection in a synchrotron facility. Optimization in crystallization conditions failed to produce large crystals. For the future work, screening different host lipid, incubation temperature, additive lipids to be doped with the host monoolein can be tested, along with applying alternative diffraction technique for microcrystals. MicroED is a powerful tool for diffraction experiments on microcrystals and nanocrystals, and has the potentials in diffracting the SbtA-6803 microcrystals grown in LCP at high resolution. However, an example of applying MicroED to the microcrystals grown in LCP had not been carried out during that time.

Therefore, it will be illustrated in chapter 5 to offer a full picture of the first demonstration of MicroED structure determination of LCP-laden crystals.

CHAPTER 4: EXPRESSION, PURIFICATION AND STRUCTURAL STUDIES OF THE CYANOBACTERIAL BICARBONATE TRANSPORTER BICA

4.1 Introduction

Cyanobacterial bicarbonate transporter BicA is an active component of cyanobacterial CO_2 concentrating mechanism. It is a Na^+-dependent bicarbonate transporter with high flux rate and low affinity, and belongs to a large family of eukaryotic solute carrier 26 transporter and prokaryotic sulfate permease (SulP/SLC26) [16]. Significant biotechnological interest currently exists with respect to employing BicA and/or enhancing its inherent attributes as a means to increase rates of HCO_3^- uptake in different host systems. For example, over-expression of BicA has been shown to enhance the growth of both *Synechococcus sp.* PCC 7002 [151] and *Synechocystis sp.* PCC 6803 [152]; suggesting this initial transport step can represent a bottleneck under certain conditions. Others, meanwhile, have explored the potential expressing cyanobacterial bicarbonate transporters in plants as a means to improve carbon uptake and productivity [152-155]. Finally, several different BicA homologs have been reported to display different inherent HCO_3^- transport behaviors, in particular with respect to both affinity and flux (e.g., K_m values of about 74, 220, and 353 µM have been reported for BicA from *Synechococcus sp.* WH8102, *Synechococcus sp.* PCC 7002, and *Synechocystis sp.* PCC 6803, respectively [16]); suggesting the possibility that key molecular features could be identified that influence the inherent transport behaviors of BicA. Further maturation of HCO_3^- uptake by BicA as an engineering strategy and as a target for various biotechnological applications would be greatly supported via thoroughly elucidating the structural features

and molecular mechanisms that it employs for functional HCO_3^- transport. However, in order to facilitate fundamental studies of BicA function and future applications, robust expression and purification procedures are first needed. With the systematic development of methodologies for enhanced heterologous expression and purification of BicA from *Synechocystis sp.* PCC 6803 (BicA-6803) in *E. coli*, the yields of purified protein reached 6.5 ± 1.0 mg per liter of culture.

4.2 Materials and Methods

4.2.1 Expression Vector and Strains

The full length gene encoding BicA-6803 (i.e., *sll0384* from *Synechocystis sp.* PCC 6803) with an N-terminal 10x-His tag followed by thrombin cleavage site was cloned into pET-16b using Gibson assembly [132] to create pBicA6803-16N. The resulting vector was confirmed by DNA sequencing and then transformed into each of the following *E. coli* strains: BL21 (DE3), BL21 (DE3) pLysS, KTD212, C41 (DE3), Rosetta2 (DE3), C43 (DE3) and C43 (DE3) Rosetta2.

4.2.2 Screening Whole Cell Expression by Dot Blot

Overnight cultures were prepared in Luria broth (LB) with 50 μg/mL carbenicillin at 37 °C (34 μg/mL of antibiotic chloramphenicol was also used for BL21 (DE3) pLysS and C43 (DE3) Rosetta2), and used to inoculate 50 mL of fresh auto-induction media which is the same overnight culture media supplemented with 0.5% glycerol, 0.5 g/L glucose, and 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) [143]. Cultures were grown overnight at 37, 30 or 18 °C. Following overnight expression, the optical cell density of the culture was measured at 600 nm (OD600). Cell pellets from 1 mL of each culture

64

were harvested at 4,000 x g and re-suspended with a buffer solution, ensuring that cell density of each sample was normalized based on OD600. Re-suspension buffer consisted of 62.5 mM Tris-HCl pH 6.8, 5 mM DTT, 25% glycerol, 1 mM PMSF and 2% SDS. Samples were applied to a nitrocellulose membrane which was then blocked with 5% milk in Tris-buffered saline pH 7.5 with 0.1% Tween 20 (TBST) and incubated with 6x-His tag monoclonal antibody (Invitrogen) followed by anti-mouse IgG AP-linked antibody (Cell Signaling Technology) in 5% milk in TBST. The blots were visualized with 5-bromo-4-chloro-3-indolyl-phosphate/nitro blue tetrazolium (BCIP/NBT) substrate (Promega). Blots were quantified by measuring the intensities digitally scanned by a Gel Doc EZ system (Bio Rad) and normalized based on the maximum intensities measured for each set of experiments.

4.2.3 Protein Purification and Detergent Screening

The best strain harboring pBicA6803-16N was grown in 1 L of auto-induction media at the optimal temperature. Following overnight growth, cells were harvested and re-suspended in 50 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA and 1 mM PMSF, and then lysed by sonication for 5 min at a magnitude of 40% on the ice (QSonica). Unbroken cells and cell debris were removed by centrifuging at 10,000 x g. The membrane pellet was harvested from supernatants by centrifuging at 105,000 x g, washed twice with 20 mM Tris-HCl pH 8.0, 1 M NaCl, 20 mM KCl, and 10 mM $MgCl_2$ and then harvested at 105,000 x g. The pellet was solubilized in 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM TCEP, 10% glycerol, and detergent for 3 h at 4 °C. Solubilization buffer was supplemented with one of the following detergents: 2% n-decyl-β-D-Maltopyranoside (DM), 1% n-dodecyl-β-D-Maltopyranoside (DDM), or 1% Lauryl

Maltose Neopentyl Glycol (LMNG). All detergents were purchased from Anatrace Products, LLC. Following solubilization, all buffers with detergent contained 0.2% DM, 0.02% DDM, or 0.01% LMNG. Unsolubilized membranes were removed by centrifuging at 105,000 x g, and then solubilized BicA in the supernatant was initially purified by immobilized metal-ion affinity chromatography (IMAC) using a HisTrap HP column (GE Healthcare). After extensive washing with 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 10% glycerol, 1 mM TCEP, 75 mM imidazole, and detergent, protein was eluted in 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 10% glycerol, 1 mM TCEP, 500 mM imidazole, and detergent. The elution fraction was pooled, concentrated at 500 x g on a Vivaspin centrifugal concentrator (GE Healthcare), and separated by size exclusion chromatography (SEC) on a Superdex 200 Increase 10/300 GL column (GE Healthcare) pre-equilibrated in 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 10% glycerol, 2 mM TCEP, 1 mM $NaHCO_3$ and detergent. SDS-PAGE was carried out to assess degree of purification after each step while ImageJ [144] was used for quantitative analysis of the gels. Protein concentrations were determined on a Nanodrop (Thermo Scientific) using absorbance at 280 nm. The detergent used for further process was selected from which produced the highest single peak by SEC and the purest protein band on the gel.

To cleave the His tag, instead of protein elution from HisTrap HP column after extensive column wash, the column was washed with 20 mM Tris-HCl pH 7.8, 150 mM NaCl, 10% glycerol, 3 mM $CaCl_2$ and detergent and injected with 10 units of thrombin in the same buffer solution per milligram of His-tagged proteins. The column was placed at room temperature for overnight with gentle rotation. Cleaved sample was obtained from the HisTrap HP column flow-through fractions by loading 10 mM HEPES pH 7.5, 200 mM

NaCl and the corresponding detergent, and subjected into a Superdex 200 Increase 10/300 GL column (GE Healthcare) pre-equilibrated in the same buffer solution. The peak fraction eluted at around the right size was collected.

SEC purified sample in detergent could also be switched to amphipols by incubating the samples with amphipol PMAL-C8 (Anatrace Products, LLC) in the mass ratio of 1/5 for 30 min at 4 °C, followed by incubation with 10 mg bio-beads for 2 h at 4 °C to remove excess amount of detergent. Sample was then subjected into a Superdex 200 Increase 10/300 GL column (GE Healthcare) pre-equilibrated in 10 mM HEPES pH 7.5 and 200 mM NaCl. The peak fraction eluted at around the right size was collected.

4.2.4 SDS-PAGE

SDS-PAGE was performed on 4% stacking gels followed by 10.5% resolving gels. Protein solutions were mixed with equal volume of 62.5 mM Tris-HCl pH 6.8, 5 mM DTT, 25% glycerol, 0.1% bromophenol blue and 0.2% SDS at room temperature for at least a day. The protein bands were visualized by 0.1% Coomassie blue R-250.

4.2.5 CD Spectroscopy

300 µL of sample (around 10 µM) purified with the best detergent was placed into a 1 mm quartz cuvette (Starna Cells; Atascadero, CA) in SEC buffer. Multi-wavelength CD measurements were taken every 1 nm from 260 to 200 nm. All measurements were taken on a Jasco J-815 CD Spectrometer (Jasco Inc., Easton, MD) and processed with the Jasco Spectra Manager V1.54.03 software package.

4.2.6 Negative-Stain Electron Microscopy

For negative-stain EM analysis, EM grids were prepared by applying 2 µL of cleaved sample onto FCF300-Cu grids glow discharged for 1 min on carbon side. Grids were

washed twice with 10 mM HEPES pH 7.5 and 200 mM NaCl and stained in 0.75% uranyl formate. EM micrographs were recorded on a TVIPS TemCam-F416 camera (Gatan) with a Tecnai F20 electron microscope (Thermo Fisher) operated at 200 kV. Images were processed in cisTEM [145].

4.2.7 Membrane Protein Crystallization

For vapor diffusion methods, purified samples in the most suitable detergent were concentrated to 3-5 mg/mL using a Vivaspin protein concentrator (GE Healthcare) with 100,000 molecular weight cut-off. Initial crystallization trials were set up in 96-well Swissci UVXPO MRC crystallization plates (Molecular Dimensions) via sitting-drop vapor diffusion at 20 °C. Crystal hits of BicA-6803 occurred in the reservoir solution of 0.1 M NaCl, 0.1 M MES pH 6.5 and 36% (v/v) PEG 300. Crystal optimization was set up in 24-well Cryschem sitting drop crystallization plates (Hampton Research). BicA-6803 crystals were achieved by mixing 2 parts BicA-6803 solution with 1 part precipitant solution of 0.14-0.16 M NaCl, 0.1 M MES pH 6.5, 32-34% (v/v) PEG 300, 1 mM NaHCO$_3$ and 4.5 mM TRIPAO at 20 °C. Diffraction data were collected at the Advanced Photon Source (APS) microfocus beamline 23-ID-D on a Pilatus3 6M PIXEL detector at an X-ray wavelength of 1.033 Å for a total of 900 × 0.2° oscillation frames.

For LCP crystallization, purified samples were concentrated to 10-30 mg/mL using a Vivaspin protein concentrator (GE Healthcare) with 100,000 molecular weight cut-off, and mixed with molten monoolein at a ratio of 2/3 (v/v) in a dual-syringe mixer until a transparent lipidic cubic phase (LCP) was formed. LCP was disperse onto a glass sandwich set as a bolus of 40 nL and covered with 1 uL of crystallization solution using a Crystal Gryphon robot (Art Robbins Instruments). The glass sandwich sets were

incubated at 20 ℃ and loaded into a Rock Imager (Formulatrix) for monitoring crystal growth.

4.3 Results and Discussions

Initial over-expression efforts performed using in BL21 (DE3) at 37 °C failed to yield high levels of BicA-6803 (data not shown). Therefore, to identify suitable expression strains and conditions for yielding increased levels of BicA-6803, several small-scale expression tests were next performed employing a range of different host strains and temperatures. Specific host strains of interest were selected based on their use in previous studies for enhancing membrane protein over-expression [148]. Overall levels of BicA-6803 produced were analyzed by adding lysed samples from each induced culture to blotting paper and conducting dot blot experiments to detect His-tagged BicA-6803. Comparing relative levels of produced BicA-6803, overall expression yields were found to be highest at 37 °C (relative to 30 and 18 °C; except for BL21 (DE3)). Expression using the *Δtig* BL21 (DE3) variant, KTD212, showed minor improvements in yields relative to BL21 (DE3) suggesting that BicA-6803 overexpression may upset the native membrane protein folding pathway [131, 156]. The gene encoding BicA-6803 was not codon optimized in this study, and Rosetta2 (DE3) cells showed a substantial improvement in BicA-6803 overexpression, indicating that codon bias may play a role in the expression levels of BicA-6803 in *E. coli*. When comparing all the strains, it was found that higher-level expression of BicA-6803 is seen in strains that have reduced levels of expression or that were selected for tolerance towards toxic protein overexpression (C41 (DE3), C43 (DE3) and BL21 (DE3) pLysS), with C43 (DE3) performing the best. This indicates that high-

levels of overexpression is determinantal to the overall yields of BicA-6803. Overall, the

strain that produced the highest levels of BicA-6803 was C43 (DE3) Rosetta2,

demonstrating that the improvements seen in Rosetta2 (DE3) and C43 (DE3) could be

synergistically combined to enhance expression levels beyond the individual strains alone

(Figure 4-1).
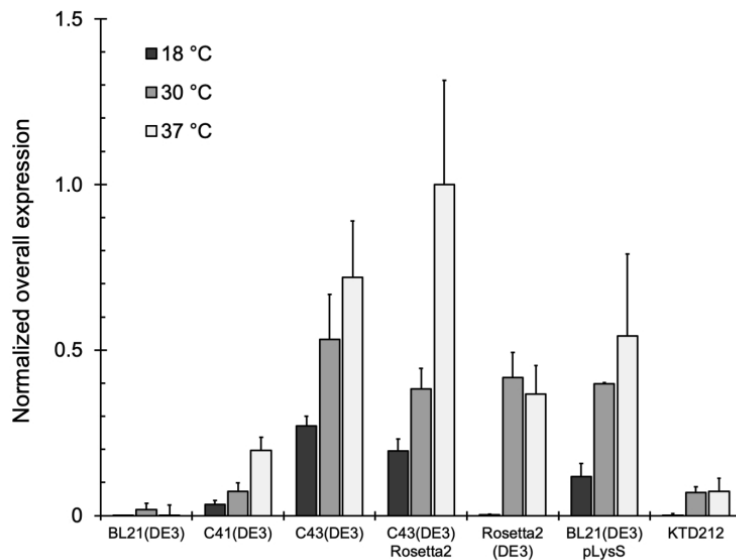
Figure 4-1. Screening expression strains and temperature for production of BicA-6803. Levels of BicA-6803 in whole cell samples were assayed by dot blot (n=3) and normalized relative to the highest expression condition.

Following the identification of conditions supporting high BicA-6803 production, we

next turned to identifying the best detergent for improving membrane solubilization and

BicA-6803 stability. Three different detergents (DM, DDM and LMNG) were screened

for this purpose. C43 (DE3) Rosetta2 cells expressing BicA-6803 were harvested and

lysed before membranes were then collected by high-speed centrifugation. Isolated

membranes were split into three fractions, each individually treated with one of three

detergents of interest in subsequent solubilization and purification steps. Ni^{2+}-NTA purification was used to identify the detergents suitable for membrane solubilization by measuring the areas of protein elution peaks, since the His tag is long enough to be exposed to the resin.. All three detergents were found to support the extraction of sufficient quantities of BicA-6803, and DDM and DM worked better than LMNG (Figure 4-2 A). Thus, to determine the optimal detergent for BicA-6803 purification, SEC was performed on each solubilized sample (Figure 4-2 B) following initial purification by IMAC. DM solubilized BicA-6803 migrated as a series of peaks, indicating that it produces an heterogeneous mixture of BicA-6803 oligomers, with the major peak occurring where monomeric BicA-6803 would be expected to elute. Due to these factors, DM was deemed unsuitable for BicA-6803 purification. In contrast, when using DDM or LMNG as detergent, SEC profiles showed much more homogenous distributions. Purification with DDM showed a large, single peak by SEC whereas LMNG showed a much-reduced major peak whose area 63% less than that obtained using DDM. Additionally, the major peak for LMNG purified BicA-6803 showed a subtle leading shoulder which was absent with DDM purified BicA-6803. In light of these findings, DDM was selected as the most suitable detergent for purification of BicA-6803.
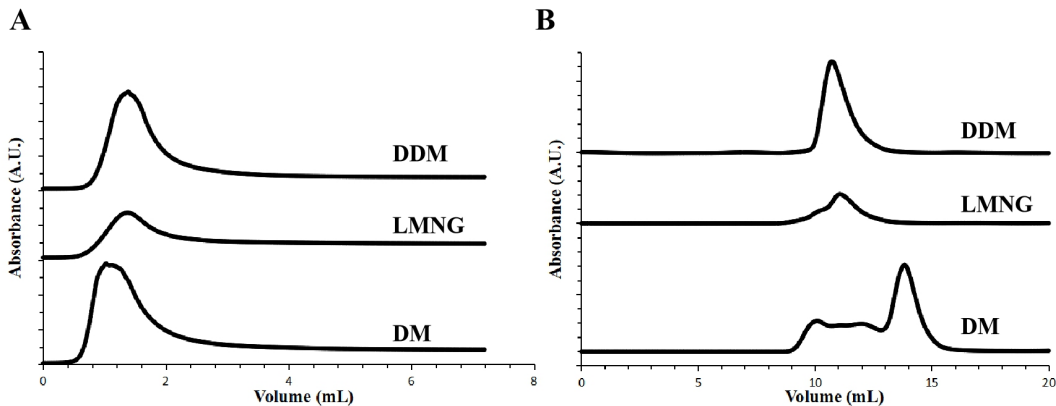
Figure 4-2: Chromatography characterization of detergent solubilized and purified BicA-6803 from (A) Ni^{2+}-NTA elution fractions where each chromatogram aligned relative to the beginning of the protein elution on the HisTrap HP column, which was set at 0 mL; and from (B) SEC where each chromatogram aligned relative to the injection of the protein solution on the Superdex 200 Increase 10/300 GL column, which was set at 0 mL. Traces have not been normalized relative to each other in order to show the relative levels of the peak fractions from each detergent.

Large scale production of BicA-6803 was lastly investigated using C43 (DE3) Rosetta2 as the expression strain, along with solubilization and purification of produced BicA-6803 in the presence of DDM. By this approach, final yields of purified BicA-6803 reached 6.5 ± 1.0 mg per liter of culture; a substantial improvement over initial expression efforts using BL21 (DE3) which yielded undetectable levels of BicA-6803. SDS-PAGE was used to assess the final purity of the BicA-6803 product obtained via this method (Figure 4-3). Following membrane extraction and purification by IMAC and SEC, the final purity of BicA-6803 was over 95% (based on quantitative density analysis of the Coomassie stained gels following SEC). Although SDS-PAGE analysis further revealed that purified BicA-6803 was gel shifted and migrated to a position that does not correspond with its predicted molecular weight, deviations such as this are not uncommon for membrane proteins [149, 150].
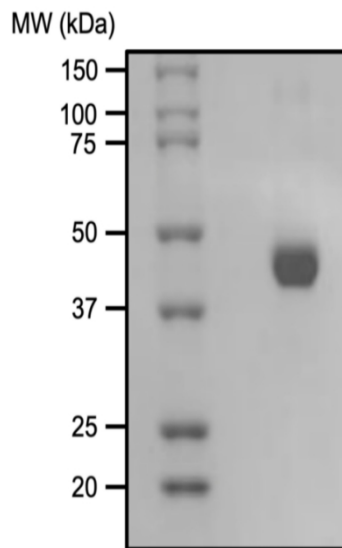
Figure 4-3: BicA-6803 immediately following IMAC and SEC purification was analyzed by SDS-PAGE followed by staining with Coomassie blue R-250. The resulting gel shows a high level of well-purified BicA-6803.

Additionally, CD spectroscopy was used to assess the secondary structure of the DDM solubilized and purified BicA-6803 (Figure 4-4). The resulting CD spectrum shows a broad absorption peak between 210 and 220 nm, and hence indicates high level of α-helical content for the purified protein, which is expected as BicA-6803 is an integral membrane protein composed mostly of α-helices.
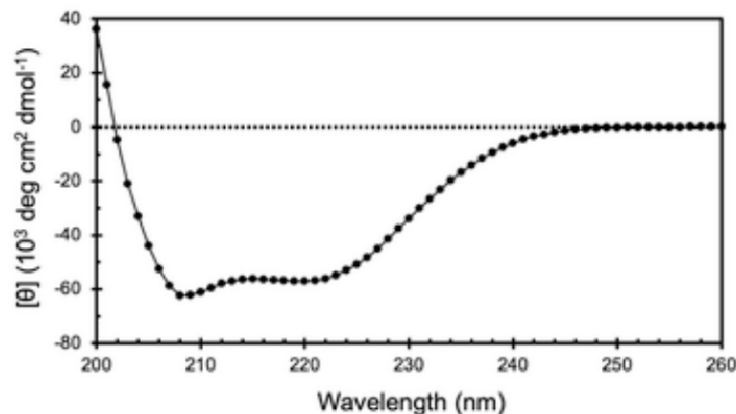
Figure 4-4: BicA-6803 immediately following IMAC and SEC purification was analyzed by CD spectroscopy. The data indicate that purified BicA-6803 is predominantly α-helical.

To further test the purity and homogeneity of purified BicA-6803, negative-stain EM was used to assess the size and homogeneity of the final purified product. As seen in Figure 4-5, well-dispersed and homogeneous particles were attained, with the longest dimension measuring approximately 100-120 Å. These dimensions are consistent with a recently published cryo-EM structure of the dimeric murine SLC26A9 anion transporter [39] which, like as BicA-6803, also belongs to the SulP/SLC26 family of transporters. These findings further indicate that BicA-6803 is produced and, according to the protocols developed here, purified as a dimer.

74
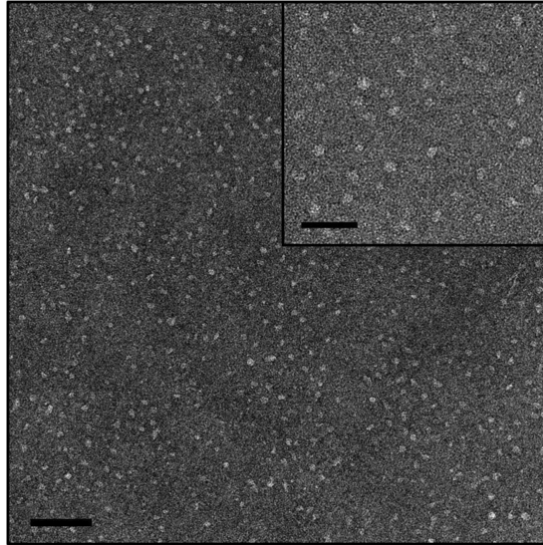
Figure 4-5: Negative-stain EM analysis of BicA-6803. Representative image and inset (top right: showing a closer view) of negatively stained BicA-6803 reveals relatively homogeneous single particles on the surface of the grid. The longest dimension of the particles measures approximately 10-12 nm, matching what would be expected from a homodimer. Scale bar represents 100 nm and 50 nm in the main figure and inset, respectively.

Membrane protein crystallization was applied to tag-cleaved BicA-6803 because crystallography has so far been the most common technique for high-resolution structure determination of biological macromolecules, especially for membrane proteins. Crystals were found at the crystallization solution of 0.1 M NaCl, 0.1 M MES pH 6.5 and 36% (v/v) PEG 300 using sitting drop vapor diffusion at 20 °C (Figure 4-6 A and B). The average dimensions of these crystals were measured as 30 x 10 μm. Crystals were optimized with screening crystallization solution in the range of 0.14-0.16 M NaCl, 0.1 M MES pH 6.5, 32-34% (v/v) PEG 300, 1 mM $NaHCO_3$ and 4.5 mM TRIPAO (Figure 4-6 C). In addition, tiny crystal hits were found at the crystallization solution of 0.1 M Tris-HCl pH 8.0, 0.23 M $CaCl_2$, 1.5% (w/v) 1,5-diaminopentane dihydrochloride and 38% (v/v) PEG 300 using LCP crystallization at 20 °C (Figure 4-6 D and E). They failed to

reach larger size during optimization of protein concentration and crystallization solution recipe, which limits the determination of its structure using the common synchrotron single crystal X-ray diffraction.
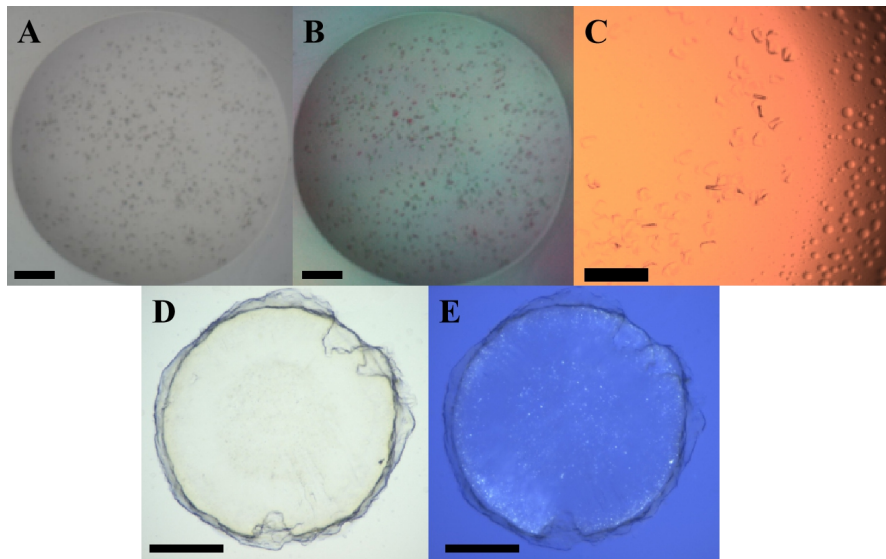
Figure 4-6: BicA-6803 crystals obtained from vapor diffusion and LCP crystallization. Crystals grown in vapor diffusion visualized with (A) normal light and (B) cross-polarizing light using a Rigaku imaging system (Rigaku); and optimized crystals visualized with normal light (C) using Leica imaging system (Leica). Crystals grown in LCP visualized with (D) normal light and (E) cross-polarizing light using a Rock Imager (Formulatrix). Scale bars represent 100 μm.

During the X-ray diffraction experiment, a diffraction pattern was obtained from a plate-shaped crystal which diffracted to approximately 8 Å (Figure 4-7). The crystal was too fragile to be well-diffracted in continuous oscillation frames, and the number of spots in each frame was small, and hence no indexing solution was determined for further data processing.
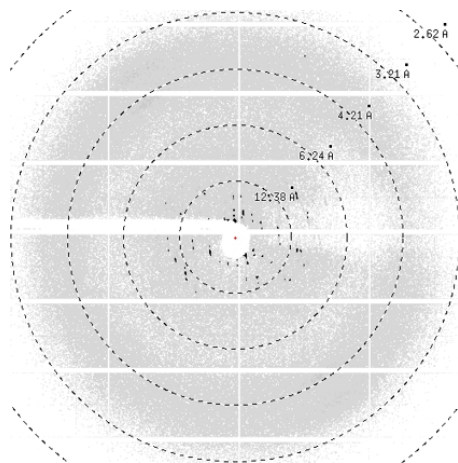
Figure 4-7: Representative diffraction pattern of BicA-6803 grown by vapor diffusion, collected at GM/CA APS beamline 23-ID-D on a Pilatus3 6M PIXEL detector at an X-ray wavelength of 1.03 Å. Diffraction spots are above 8 Å.

4.4 Conclusions

As with many membrane proteins, low levels of expression can limit the structural and functional study of BicA-6803. We have optimized the expression of BicA-6803 in *E. coli* and screened the purification methods, and approximately 6.5 ± 0.1 mg of purified BicA-6803 using DDM detergent can be obtained from one liter of C43 (DE3) Rosetta2 cell culture, which is relatively high for membrane protein expression. SEC elution profiles confirm that DDM helps to keep BicA-6803 in a stable form. Besides, the SEC elution time of BicA-6803 along with results from negative-stain EM and comparisons with other, full-length SulP/SLC26 family transporters indicate that the herein developed protocol is capable of yielding stable, homodimeric complexes.

The relatively high yield of purified BicA-6803 is appropriate for membrane protein crystallography. Although crystals were grown by vapor diffusion method, they diffracted to approximately 8 Å, which is of low resolution to solve the structure but shows that the crystals were packed with proteins. The low resolution is the result of

loose molecular packing in the crystal lattices since the length of DDM or the volume of DDM micelle is big. Close and tight crystal packing is a prerequisite for high-resolution diffraction as the side chains and chemical bonds are generally well ordered. During membrane protein crystallization by vapor diffusion method, the crystal contacts are primarily created by the hydrophilic domains. The molecular size of STAS domain (around 17 kDa) is much smaller compared with the hydrophobic membrane domain (around 42 kDa), therefore a close packing was likely to be inhibited by the volume of membrane domain and detergent. In order to optimize the crystal quality by vapor diffusion, detergent with low size can be screened for the crystallization set-up. Lauryldimethylamine-N-Oxide (LDAO) is a suitable candidate because it has the same length of alkyl chain as DDM while the hydrophilic head group consists of a small amine oxide instead of the large maltoside moiety in DDM [60]. It is assumed that replacing DDM with LDAO maintains the dimer conformation of BicA-6803, and facilitates close crystal packing. High-resolution crystal structure of membrane protein has been determined using LDAO as the detergent [29].

In addition, microcrystals were obtained in LCP, although the are too small to withstand the radiation damage by the X-ray beam during the X-ray diffraction data collection in a synchrotron facility. Optimization in crystallization conditions failed to produce large crystals. For the future work, screening different host lipid, incubation temperature, additive lipids to be doped with the host monoolein can be tested, along with applying alternative diffraction technique for microcrystals. MicroED can be applied to diffracting the BicA-6803 microcrystals grown in LCP, once the feasibility of applying MicroED to microcrystals grown in LCP can be verified. Therefore, it will be illustrated in chapter 5

to offer a full picture of the first demonstration of MicroED structure determination of LCP-laden crystals.

CHAPTER 5: HIGH-RESOLUTION STRUCTURE DETERMINATION OF MICROCRYSTALS EMBEDDED IN LIPID MESOPHASE BY MICROED

5.1 Introduction

In the chapters 3 and 4, crystallization of cyanobacterial bicarbonate transporters in lipid cubic phase (LCP) was applied in order to produce high quality crystals of SbtA or BicA for diffraction experiments, although the crystals were very small. Since the first high-resolution structure of Bacteriorhodopsin was determined using LCP or lipid mesophase (*in meso*) crystallization in 1997 [61, 157], over 120 unique high quality membrane protein crystals have been obtained in LCP [158-160]. However, membrane protein crystals grown in the LCP are often too tiny to withstand the radiation damage from the X-ray beam during synchrotron X-ray diffraction data collection. Serial femtosecond crystallography (SFX), which utilizes ultra-fast and high-brightness X-ray pulses generated by an X-ray free electron laser (XFEL) to capture diffraction patterns from tiny membrane protein crystals grown *in meso* before they are destroyed by radiation damage, has recently been employed with great success [111]. Nevertheless, the LCP-SFX technique is extremely time and resource consuming and is not widely accessible with only a few operational XFEL facilities in the world. Serial crystallography for LCP-microcrystals has been applied to synchrotron radiation [113, 161]. However, they both require a large number of microcrystals to capture complete diffraction data sets for structure determination. Thus, to improve the study of important integral membrane proteins and uncover new high-resolution details in a more high-throughput fashion, alternative methods for structure determination need to be developed for the microcrystals grown in LCP.

MicroED is a recently developed method to collect electron diffraction patterns from sub-micron sized three-dimensional crystals of biological macromolecules in the electron microscope operating at cryogenic temperature [89-99]. Microcrystals are deposited on EM grids followed by blotting the grids for thin windows, since electrons cannot penetrate thick samples. Then the samples on the EM grids are vitrified in liquid ethane and kept at the cryogenic temperature maintained by liquid nitrogen in order to ensure sample hydration exposed under the high vacuum of cryo-TEM. The continuous rotation MicroED data collection methods resemble the methods to collect single crystal X-ray diffraction data, resulting in a series of electron diffraction patterns that can be processed using the software developed for single crystal X-ray diffraction [91]. Since the initial implementation of MicroED, there have been further efforts in improving this method [98]. In this work, our goal is to combine MicroED data collection with LCP crystallization (LCP-MicroED) to determine structures from microcrystals within the lipid mesophase.

To begin with, we treated the LCP with different reagents to reduce its viscosity, because toothpaste-like LCP can hardly be evenly deposited on the EM grids and blotted thin enough to be screened in the electron beam. Two different strategies were identified for significantly lowering the viscosity of LCP: phase conversion using 2-methyl-2,4-pentanediol (MPD) or treatment with lipase. Next, we produced microcrystals of a model protein Proteinase K, reconstitute them into LCP, and used these strategies for screening diffraction quality in a cryo-TEM. Proteinase K was selected as a model protein since it is useful by both MicroED [96, 112] and LCP-based method [113]. Finally, both strategies were successful in determining the structure of Proteinase K at 2.0 Å resolution using MicroED.

In addition, we used the sample treatment strategies mentioned above for the microcrystals grown within the LCP matrix. Cholesterol was used to demonstrate the structure determination of small molecules by LCP-MicroED. After phase conversion using MPD, diffraction data of cholesterol were collected and a low-completeness (~30%) map was generated to around 1 Å resolution. A human G protein-coupled receptor adenosine A_{2A} receptor ($A_{2A}AR$) was used for elucidating LCP-MicroED of macromolecules. Diffraction to around 4.5 Å was observed after treatment with MPD.

5.2 Materials and Methods

5.2.1 Preparation of Microcrystals

Microcrystals of Proteinase K were prepared according to a previous protocol [161]. Briefly, equal volumes of Proteinase K (Sigma) solution at 40 mg/mL in 20 mM MES pH 6.5 and a crystallization solution composed of 0.1 M MES pH 6.5, 0.5 M $NaNO_3$ and 0.1 M $CaCl_2$ were mixed and crystallized in batch method. Microcrystals were grown after 20 min incubation at 20 °C [161]. Microcrystals were pelleted by low-speed centrifugation (500 x g) for 5 min, re-suspended in the crystallization solution, and then re-constituted into LCP. Reconstitution was realized by mixing microcrystal solution with molten monoolein host lipid at a ratio of 2:3 (v:v) using a Hamilton dual-syringe mixer until a homogeneous and transparent lipid mesophase was formed [162].

Cholesterol (Anatrace) was mixed with monoolein at a ratio of 3:7 (w/w) and co-dissolved in chloroform. The solvent was removed by evaporation under a stream of inert gas, and the lipid mixture was dried at high vacuum and room temperature for 24 h. LCP was formed by mixing the lipid mixture with water at a ratio of 3:2 (v/v) using a

Hamilton dual-syringe mixer, and injected into a Hamilton syringe containing a precipitant solution composed of 0.1 M sodium citrate pH 5.0, 40 mM NaSCN and 28% (v/v) PEG 400. Microcrystals formed after incubation at room temperature for 18 h.

$A_{2A}AR$ was prepared as described before [163-165]. In brief, the intracellular loop 3 of $A_{2A}AR$ was constructed with a fusion protein BRIL, cloned into the pFastBac1 vector and transformed into *E. coli* One Shot™ TOP10 competent cells. The recombinant bacmid DNA was prepared by transformation into *E. coli* MAX Efficiency™ DH10Bac competent cells, followed by transfection of Sf9 insect cells to generate the baculoviruses. The baculoviruses were not introduced into the Sf9 cell culture for infection until the cell density reached 2×10^6 cells per mL. Protein expression in Sf9 cells lasted for 48 h at 27 °C. $A_{2A}AR$ was purified after harvesting cells by centrifugation, concentrated to 25 mg/mL, and mixed with host lipids to generate LCP. The LCP consisted of 40% (w/w) protein solution, 6% (w/w) cholesterol and 54% (w/w) monoolein, and was injected into Hamilton syringes containing the same precipitant solution as cholesterol crystallization. Microcrystals of $A_{2A}AR$ were grown after incubation at 20 °C for 24 h.

5.2.2 Conversion of LCP Microcrystals

LCP hosting the Proteinase K microcrystals was converted by two different strategies. One was phase conversion with seven additives to achieve a less viscous lipid sponge phase, and these additives were MPD, PEG 200, PEG 400, jeffamine M600, 1,4-butanediol, t-butanol, and ethylene glycol. Phase conversion was performed by syringe mixing 15-20 times in Hamilton dual-syringe mixer, in which one syringe contained LCP hosting the microcrystals and the other syringe contained the phase conversion solution. The phase conversion solution was composed of the initial crystallization buffer for

Proteinase K supplemented with each of different additives, and optimized with a gradient series of 10% increments of additive concentration. Finer additive concentration optimization was performed after identifying an initial point capable of converting the cubic phase to a sponge phase. Once the concentration range of each supplemented additive was determined, EM grid blotting experiment was performed to screen the capability of each additive. Among these additives, MPD induced conversion to the least viscous sponge phase, and was later focused for further optimizing crystal survival during phase conversion. Once the cubic phase was converted, the mesophase hosting Proteinase K microcrystals was transferred into a micro-centrifuge tube. Microcrystals were concentrated by low-speed centrifugation and harvested from the bottom by pipetting. Crystal survival after each step was monitored by light microscopy with cross-polarized and UV light. Once crystals survived after each step, MicroED grid preparation would be conducted for analysis in the cryo-TEM.

The other strategy involved treatment of LCP with lipase. The fresh lipase (*Candida rugosa*, Sigma) stock solution was prepared at a concentration of 50 mg/mL in saturated potassium phosphate buffer. The mesophase hosting Proteinase K microcrystals was expelled from the Hamilton syringe into a micro-centrifuge tube. Lipase was introduced at a volume ratio of 1:1:2 (lipase stock:mesophase:crystallization solution) into the same tube without mixing. Incubation for at least 18 h converted the mesophase to a two-phase system of two immiscible liquids: water/glycerol and oleic acid. Proteinase K microcrystals partitioned into the water/glycerol phase. The methods developed for phase conversion by MPD were used.

LCP hosting cholesterol microcrystals was treated with MPD. After removing excess crystallization solution from the Hamilton syringe, the syringe was coupled with another Hamilton syringe containing the conversion solution, and phase conversion was performed by dual-syringe mixing for 3-5 times. The conversion solution was composed of initial crystallization solution supplemented with 7% MPD. MPD was better diffused into the mesophase after incubation for 10 min at 20 °C. The methods developed for Proteinase K after phase conversion by MPD were used.

LCP hosting $A_{2A}AR$ microcrystals was treated with four different additives: MPD, PEG 400, PEG 200 and jeffamine M600. MPD contributed to the highest quality EM grid with clear and thin windows observed in cryo-TEM. The methods developed for cholesterol were used. Besides, single crystal X-ray diffraction data were collected from samples after conversion by MPD at APS microfocus beamline 23-ID-D following the previous protocol [164].

5.2.3 MicroED Grid Preparation

After MPD or lipase treatment, Proteinase K microcrystals were concentrated by centrifugation at 500 x g for 5 min. Microcrystal solution was harvested by pipetting from the bottom of micro-centrifuge tube. For MPD-converted samples, the microcrystal solution was further diluted with equal volume of fresh crystallization solution supplemented with 12.5% MPD. Cholesterol and $A_{2A}AR$ microcrystals were directly applied after phase conversion.

Grids were prepared by standard MicroED sample preparation procedures [99]. To summarize, 2 µL of microcrystal solution was applied on each side of a glow-discharged holey carbon grid (Quantifoil 2/4 grids for Proteinase K and cholesterol, and Quantifoil

multi-A grids for $A_{2A}AR$), and the grid was blotted for 12-16 s in a Vitrobot Mark IV (Thermo Fisher Scientific) followed by plunging into liquid ethane and storing in liquid nitrogen.

5.2.4 MicroED Data Collection

MicroED data collection of Proteinase K was performed by standard methods [99] using a TF20 cryo-TEM (Thermo Fisher Scientific) operated at 200 kV and equipped with a F416 CMOS detector (TVIPS). Diffraction data were recorded in movies at 4 s per frame and electron dosage of approximately 0.01 $e^-/Å^2$ per second as the stage was continuously rotated at 0.09 °/s. Each data set sampled around 45° of reciprocal space and was generated from a total electron dose of up to 5 $e^-/Å^2$.

MicroED data collection of cholesterol and $A_{2A}AR$ was performed by standard methods [99] using a Titan Krios cryo-TEM (Thermo Fisher Scientific) operated at 300 kV and equipped with a CETA CMOS detector (Thermo Fisher Scientific). Each diffraction data set of cholesterol was recorded at 2 s per frame as the stage was continuously rotated at 0.91 °/s, which covered around 45° wedge. Each diffraction data set of $A_{2A}AR$ was recorded at 8 s per frame as the stage was continuously rotated at 0.09 °/s, which covered around 20° wedge.

5.2.5 MicroED Data Processing and Structure Determination

MicroED data of Proteinase K collected from both MPD and lipase treatment were indexed and integrated in iMOSFLM [150]. Data sets from multiple crystals (four for MPD-treated and two for lipase-treated) were merged and scaled in AIMLESS [166] to reach high completeness for structure determination. Phaser [151] was used for running molecular replacement using a published Proteinase K structure (PDB ID: 2ID8 [167]) as search model,

and the solution was refined in phenix.refine [152] of the PHENIX [153] program suites using electron scattering factors. The model building was manually performed using Coot [154], and MolProbity [140] was used for Ramachandran analysis.

MicroED movie data of cholesterol and $A_{2A}AR$ were converted to images in SMV format using published tools [112]. Images were processed in XDS [168] for indexing, integration and scaling. Two data sets of cholesterol were merged in AIMLESS [166], and phased in Phaser [151] using a published Proteinase K structure (CSD entry: CHOLES20 [169]) as search model.

5.3 Results and Discussions

Because LCP is too viscous to be blotted thin for analysis in a cryo-TEM, the standard MicroED sample preparation methods can hardly be applied to microcrystals embedded or grown in the mesophase. Initially, our focus was to identify and optimize sample preparation conditions that would reduce the viscosity of mesophase and allow effective grid blotting. Thus, we introduced two different strategies to reduce the viscosity: One was achieved by mixing LCP with additives, which converted LCP into lipidic sponge phase, a liquid analogue of LCP [63, 102]; The other was realized by incubating LCP with lipase, which hydrolyzed the host lipid and formed a two-liquid phase system consisting of water/glycerol solution and oleic acid [170, 171].

For the first strategy, we screened seven additives at various concentrations: MPD, PEG 200, PEG 400, ethylene glycol, t-butanol, jeffamine M600, and 1,4-butanediol [162, 172]. Other studies have proposed that introducing these agents to LCP was likely to cause a less viscous lipid mesophase [164, 172], however screening and optimizing use of additives are necessary for reducing the viscosity of lipid mesophase without damaging the crystals.

Initial screenings were performed on blank LCP by mixing the host lipid monoolein with the crystallization solution for Proteinase K. All seven additives were able to reduce the viscosity of lipid mesophase by dual-syringe mixing, in the range of 6%-18% with MPD, 24%-40% with PEG 200, 32%-48% with PEG 400, 15%-33% ethylene glycol, 11%-20% t-butanol, 9%-20% jeffamine M600, and 28%-41% 1,4-butanediol, which are consistent with previous studies on phase transition [172].

We then tested the absorption of the viscosity-reduced lipid mesophase on blotting paper, in order to empirically determine how well these phase-converted samples could be blotted to generate a thin sample for MicroED. In this step, samples were expelled out of the mixing syringe and deposited on the blotting paper, without external blotting force applied. Compared with the blank LCP droplet (Figure 5-1 A), which retained its shape and was hardly blotted on the filter paper, the seven phase-converted samples showed absorption on the filter paper (Figure 5-1 B-H). MPD (Figure 5-1 B) showed the most significant blotting relative to other additives, as well as the lowest final ratio added to the LCP (6-18%). When similar blotting tests were conducted on EM grids, the MPD-converted lipid mesophase samples produced the most consistently thin samples for successful TEM visualization. Therefore, MPD was selected for further experiments with Proteinase K, cholesterol and $A_{2A}AR$ crystals embedded in LCP.

For the second strategy, we investigated an alternative enzyme hydrolysis method by treating LCP with lipase to hydrolyze host lipid molecules and transition the LCP to a two-liquid phase system [171]. Blank LCP was again used to find an optimal hydrolysis ratio of LCP with lipase, as well as the minimum treatment time to completely separate LCP into two liquid phases. LCP sample was expelled from the syringe mixer into a 0.2

mL microcentrifuge tube, and freshly prepared lipase solution at 50 mg/mL was directly added on top of LCP sample without additional pipetting. It was found that after a 14-hour treatment in a 1:1 ratio of LCP and freshly prepared lipase solution incubated at 20 °C, the solid cubic phase (Figure 5-1 I) was completely separated into two liquid layers (Figure 5-1 J). The lipase-hydrolyzed sample also penetrated the blotting paper without any visible lipid mesophase residue on the surface (Figure 5-1 K). As with the MPD treated samples described above, this strategy was then tested with Proteinase K microcrystals to study crystal survival and grid preparation for data collection.
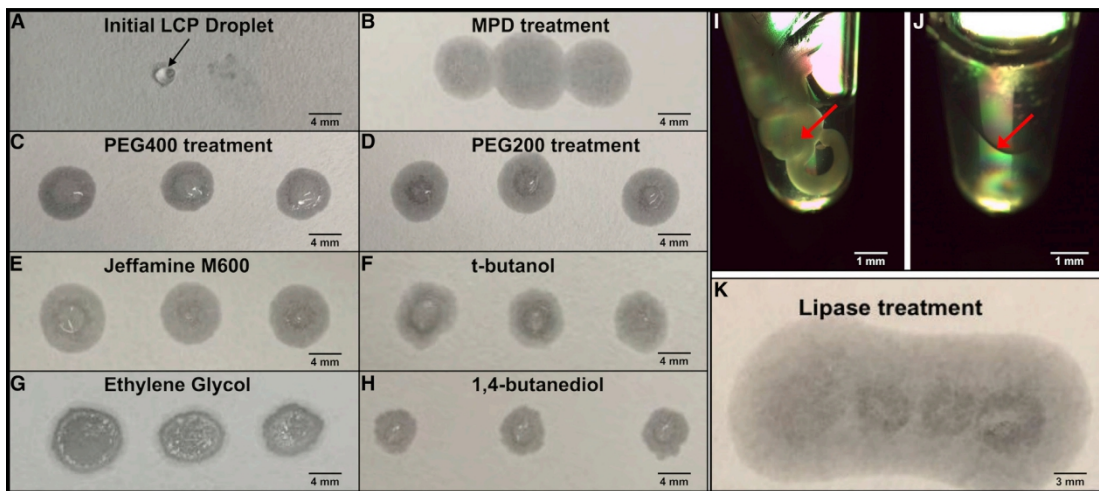
Figure 5-1: Reducing the viscosity of LCP by the introduction of additives or lipase treatment to generate samples suitable for MicroED grid preparation. (A) LCP droplet stayed on the surface of filter paper instead of being blotted away, because of high viscosity. Treating LCP with the phase-converting buffer supplemented with seven additives: (B) MPD, (C) PEG 400, (D) PEG 200, (E) Jeffamine M600, (F) t-butanol, (G) ethylene glycol, and (H) 1,4-butanediol; converted the LCP to a less viscous liquid-like mesophase, which could penetrate the filter paper. Treating LCP with lipase to hydrolyze the lipid to form two liquid phases. (I) The LCP matrix (white solid stream in the tube highlighted by the red arrow) mixed with freshly prepared lipase solution at a ratio of 1:1 (v/v) prior to incubation. (J) The LCP was separated into two liquid phases after 14 h, where the phase separation surface denoted by the red arrow. (K) The sample at the bottom phase after lipase hydrolysis was blotted away on the filter paper without significant mesophase residue on the surface.

We then followed the batch crystallization method to grow Proteinase K microcrystals in solution (Figure 5-2 A) and reconstituted them into LCP (LCP-Proteinase K) by dual-syringe mixing with the host lipid monoolein in a lipid:crystal solution ratio of 3:2 (v/v). Proteinase K microcrystals survived after reconstituted into LCP (Figure 5-2 B) and were used for phase conversion test with MPD or lipase treatment.

For sponge phase conversion, we tested the range of 6-18% MPD supplemented to the Proteinase K crystallization solution in increments of 0.5% as a phase converting buffer. This conversion buffer was then mixed with the LCP-Proteinase K in the syringe. We observed that the converting buffer supplemented with more than 12.5% MPD resulted in fewer/no Proteinase K microcrystals or microcrystals with dissolved edges (crystal image not shown). Therefore, 12.5% MPD was selected as the maximum concentration capable of reducing the viscosity of the LCP sample (Figure 5-1 B) without dissolving the embedded Proteinase K microcrystals (Figure 5-2 C).

LCP-Proteinase K samples were treated with lipase at a ratio of 1:1 (v/v), along with additional Proteinase K crystallization solution as a supplement to ensure the stability of the microcrystals after release from the lipid mesophase. Hydrolysis time was also re-monitored every 1 hour for the first 14 hours, and every 20 minutes afterward. After 18-hours treatment, sample consisting of a 1:1:2 (v/v/v) ratio of lipase:LCP-Proteinase K:crystallization solution was completely digested and separated the LCP into two-liquid layers, and the Proteinase K microcrystals were released from the lipid mesophase into the water/glycerol-rich phase. Similar size and density of Proteinase K microcrystals in

the hydrolyzed liquid phase were observed compared to the LCP-Proteinase K sample before treatment (Figure 5-2 D).

With these two sample preparation methods, the low-viscosity LCP-microcrystal solution was applied to the glow-discharged EM grid coated with holey carbon film, blotted with filter paper vitrified in liquid ethane and stored in liquid nitrogen, which is similar to the standard MicroED sample preparation methods [99]. The MPD-induced less viscous mesophase sample was further diluted by equal volume of the converting buffer, which generated a thin layer on EM grid where microcrystals could still be identified by UV microscopy (Figure 5-2 E). LCP treatment with lipase also produced grids with a similar level of microcrystals visible by UV microscopy (Figure 5-2 F).
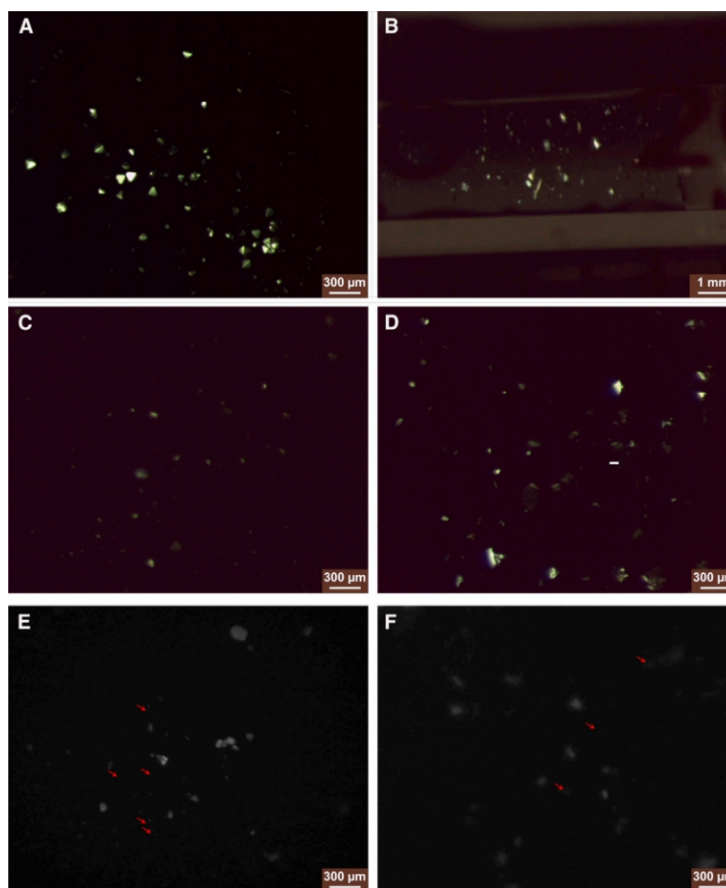
Figure 5-2: Monitoring Proteinase K microcrystals throughout LCP conversion. Proteinase K microcrystals were grown using (A) batch method and (B) reconstituted into the LCP by dual-syringe mixing, viewed with cross-polarizing light. Proteinase K microcrystals reconstituted in LCP survived after converting LCP using the crystallization solution supplemented with (C) 12.5% MPD or (D) lipase to hydrolyze lipids, viewed with cross-polarizing light. Proteinase K microcrystals embedded in LCP treated with (E) MPD and (F) lipase were deposited on the glow-discharged EM grids and successfully blotted thin for MicroED data collection, viewed with UV. Microcrystals could still be detected on the blotted EM grids. Microcrystals were monitored in a Rock Imager (Formulatrix).

Samples prepared by the methods described above were analyzed in the cryo-TEM to verify the sample thickness and the presence of microcrystals. Regions of grid containing Proteinase K (Figures 5-3 A and E) could be visually identified and confirmed. In the case of Proteinase K, while both treatment strategies produced suitable samples, lipase-

treated samples generally contributed to a thinner layer on the EM grid relative to the MPD-converted samples. Standard MicroED diffraction screening, data collection, and data processing protocols [99, 173] were used to collect high-resolution datasets from each type of sample preparation. For the MPD-converted samples, diffraction data from four Proteinase K microcrystals were merged together to produce a final dataset with a refined structure at 2.0 Å. In the case of the lipase-treated samples, diffraction data from two Proteinase K microcrystals were used to resolve the structure at 2.0 Å. Both of these methods of LCP sample preparation ultimately produced high quality MicroED data, density maps, and models (Appendix Table B; Figure 5-3). When compared with other Proteinase K structures determined by MicroED [93, 96, 174], both of the LCP-Proteinase K structures in this work showed similar levels of data quality (e.g., resolution, R factors), indicating that the LCP phase conversion methods had little adverse impact on the quality of vitrified microcrystals whether induced by MPD or lipase. In addition, the overall root-mean-square deviation (RMSD) is 0.59 Å between both of our models and another MicroED Proteinase K structure (PDB ID: 5I9S [93]), indicating minimal differences between them. When we compared the MPD- and lipase-treated structures, the all-atom RMSD between these two new structures is 0.47 Å.

Coordinates and structure factors of Proteinase K treated with lipase and MPD were deposited in the Protein Data Bank (PDB) under the accession code 6PQ0 and 6PQ4.
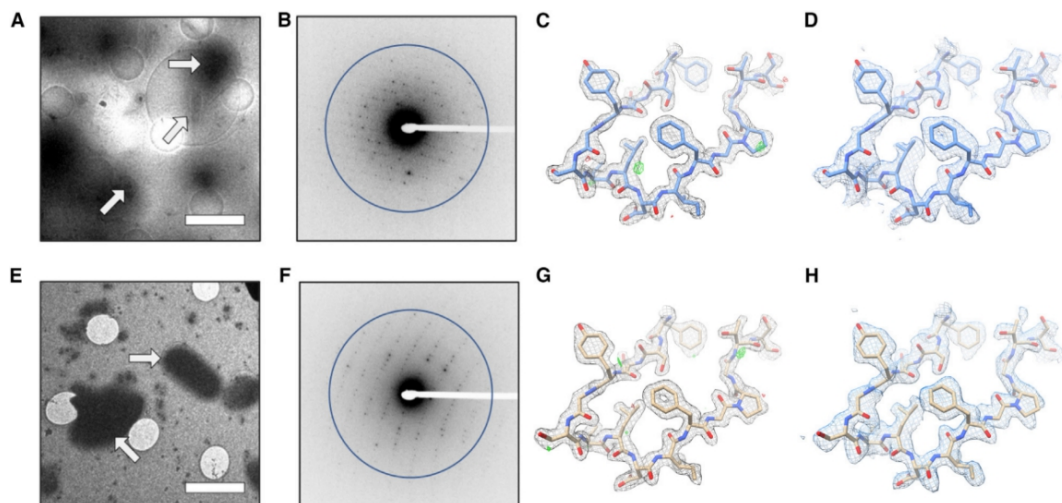
Figure 5-3: Proteinase K structure determined using LCP-MicroED. LCP samples treated with both (A) MPD and (E) lipase produced grids where microcrystals could be identified on the grid surface when loading the grids in the cryo-TEM. Scale bars in represent 4 mm. Representative diffraction patterns collected from microcrystals treated with (B) MPD or (F) lipase with resolution ring of 2.0 Å shown in the images. The $2F_o$-F_c and F_o-F_c maps of Proteinase K after treatment with (C) MPD or (G) lipase show clear density surrounding the models, contoured at 1.5σ and 3.0σ, respectively. The composite omit maps of Proteinase K after treatment with (D) MPD or (H) lipase show clear density surrounding the models, contoured at 1.0σ.

To further test the viability of the sample treatment strategy for microcrystals grown in LCP, cholesterol was crystallized in LCP and then examined using the MPD treatment method. LCP reconstituted with either cholesterol or $A_{2A}AR$ was crystallized in the initial crystallization solution containing 28% PEG 400, which transformed the cubic phase into a less viscous intermediate phase. Once microcrystals formed as needles to the size of the 2-3 μm in the longest dimension from cholesterol (Figure 5-4 A), excess crystallization solution was removed, and the conversion buffer (initial crystallization solution supplemented with 7% MPD) was introduced to convert the LCP by dual-syringe mixing in a gentle manner. Further dilution by adding the conversion solution was applied before depositing microcrystals on the EM grid.

Cholesterol microcrystals grown in LCP were converted to a less viscous phase using the same protocol involving MPD. The vitrified sample on the EM grid (Figure 5-4 A) was diffracted using cryo-TEM, and subsequent diffraction patterns showed well resolved spots to high resolution (slightly beyond 1 Å) (Figure 5-4 B). We were able to index the cholesterol datasets with a resulting P1 space group with unit cell dimensions of 12.257 Å, 12.343 Å, and 34.262 Å with angles of 89.551°, 83.497°, and 78.907° (Table 5-1). These values are consistent with previously published results on cholesterol structure [169, 175]. Also, the X-ray structure of this cholesterol form [169] was used to calculate model amplitudes, and they compared very well with the corresponding amplitudes obtained by MicroED (CC = 84.2%). Unfortunately, because of the low symmetry of the cholesterol crystal space group (P1) and the preferred orientation on the grid, the completeness of the merged dataset was very low (31.1%), which prevented the use of direct methods for phasing. Due to low data completeness, a reliable structure of cholesterol could not be determined, however a projection map could be calculated. By using the X-ray model of cholesterol and molecular replacement, a projection map was generated that shows good maps when viewed along the direction containing high data completeness for the cholesterol and water molecules in the crystals (Figure 5-4 C). This is an important step because it demonstrated that the method previously used for the Proteinase K samples could also be extended to the collection of MicroED data from crystals that had been grown in LCP, even for the small organic molecule samples.

Table 5-1: Cholesterol data collection statistics.

Data collection and processing	
Excitation voltage	300 kV
Electron source	Field emission gun
Wavelength	0.019687 Å
Detector	CETA D
Rotation rate	2 s per frame
Total dose per crystal	~2.0 e$^-$/Å^2
Number of crystals	2
Space group	P1
Cell dimensions	
a, b, c (Å)	12.257, 12.343, 34.262
α, β, ɣ (°)	89.551, 83.497, 78.907
Highest resolution (Å)	1.00
R_{merge} (%)	20.6 (83.5)
$CC_{1/2}$	0.978 (0.393)
Completeness (%)	31.1 (30.9)
Multiplicity	7.7 (6.3)
Total reflections	8,582
Total unique reflections	3,292

Values in parenthesis () for the highest resolution shell
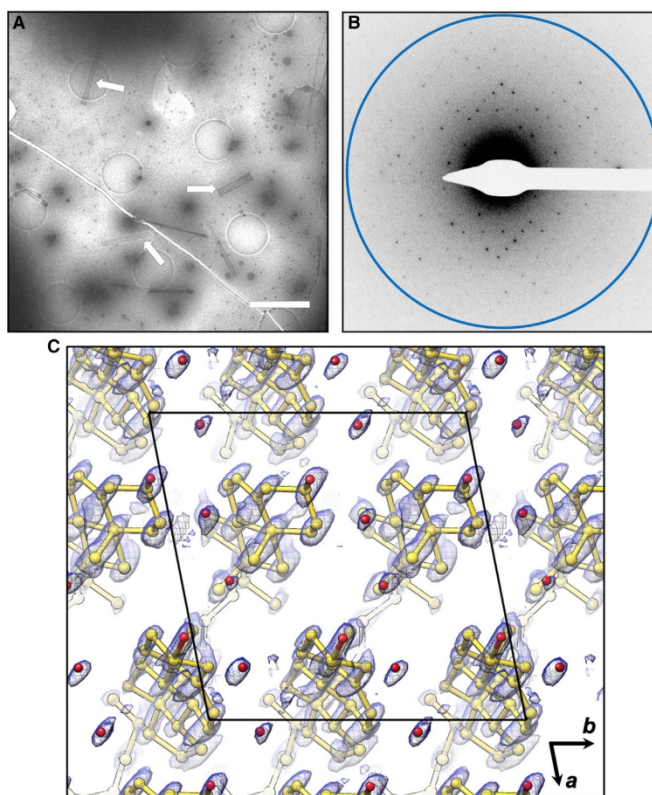
Figure 5-4: Application of LCP-MicroED to cholesterol microcrystals. (A) Cholesterol microcrystals were identified on the EM grid surface after converting LCP using the crystallization solution supplemented with 7% MPD. Scale bar represents 3 μm. (B) A still electron diffraction pattern was collected with a resolution ring of 0.9 Å shown in the image prior to the continuous rotation MicroED data collection. (C) The model and density map of cholesterol viewed along the direction with high completeness using LCP-MicroED data.

The same sample preparation procedures using MPD to treat the cholesterol microcrystals was applied to $A_{2A}AR$ microcrystals once they were grown to the size of 5 x 5 x 2 μm³ (Figure 5-5 A and C). $A_{2A}AR$ microcrystals on the grids were shown to have survived the addition of 7% MPD and the deposition process via UV microscopy (Figure 5-5 B). The diffraction quality of the $A_{2A}AR$ microcrystals following LCP conversion was also examined by X-ray diffraction using a microfocus beamline. These crystals retained their diffracting power to 2.4 Å resolution (Figure 5-5 E), indicating the

treatment with MPD does not significantly reduce the crystal quality. We next applied the same MicroED data collection process to study $A_{2A}AR$ microcrystals and observed the diffraction to be about 4.5 Å (Figure 5-5 D). Although fewer spots were found in the diffraction patterns within a narrow tilt angle series, we were able to collect a small tilt series that showed the unit cell parameters are consistent with previously published $A_{2A}AR$ models, including one from LCP-SFX experiments [163].
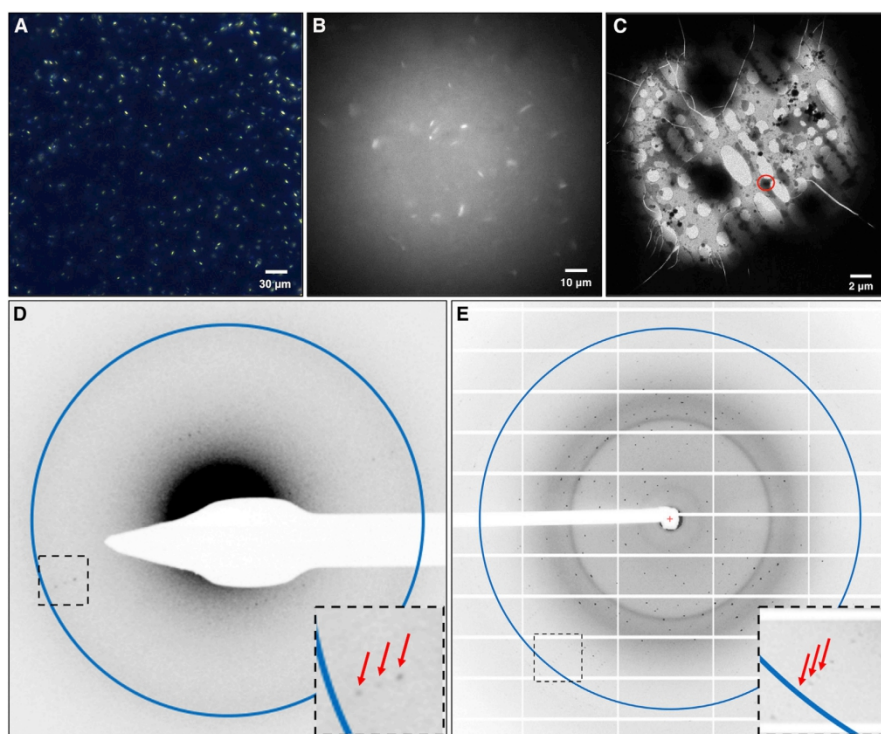
Figure 5-5: Application of LCP-MicroED to human $A_{2A}AR$ microcrystals. (A) $A_{2A}AR$ microcrystals reached average size of 5 x 5 x 2 μm^3 in LCP, viewed with cross-polarizing light. (B) Microcrystals survived on the EM grids after converting LCP using the crystallization solution supplemented with 7% MPD, viewed with UV. (C) Microcrystals were identified on the EM grid surface, highlighted in red circle. (D) A still electron diffraction pattern was recorded with resolution ring of 4.5 Å shown in the image prior to the continuous rotation MicroED data collection. (E) $A_{2A}AR$ microcrystals retained their diffraction power to 2.4 Å resolution (shown as the resolution ring in the image) at a microfocus X-ray beamline, using the same LCP conversion method for MicroED

98

analysis. Red arrows in the black boxes denote the diffraction spots to the highest resolution in a closer view.

5.4 Conclusions

As the first demonstration of LCP-MicroED, these results prove that microcrystals embedded in LCP can yield structure of model protein Proteinase K and data of cholesterol and $A_{2A}AR$ to high resolution using MicroED. This proof-of-concept study and methodology opens the gate for future LCP-MicroED applications for challenging membrane proteins which only form tiny crystals. To expand the application of LCP-MicroED to these tough targets, further development and optimization based on these initial methods need to be explored. Since every microcrystal sample is unique, the extension of the method reported here into a suite of techniques will be important for its broad applicability.

Regarding conversion of lipid mesophase using additives, a broader spectrum of chemicals should be investigated for LCP-MicroED. Certain polar solvents or other additives such as propylene glycol and pentaerythritol propoxylate, are commonly used in membrane protein crystallization solutions and can also be used as phase conversion additives. The identification of a suite of additives compatible with LCP-MicroED would allow users to select chemicals already present in the crystallization solution or with similar chemical properties. Two critical parameters should be kept in mind in order to screen new additives: the viscosity of the additive and the overall percentage of additive required to convert the lipid mesophase. In this study, we found that MPD behaved much better than PEGs in sample blotting on EM grids, which was attributed to its lower viscosity and lower concentration (6%–18%) required for conversion to a less viscous

phase. The intrinsic viscosity is not a parameter that can be tuned easily; however, these additives typically have a wide range of concentrations that can drive conversion to a less viscous lipid mesophase [172]. When screening additives, some of them may affect microcrystal quality at the concentrations required to trigger the desired phase conversion. Thus, their effects on both phase conversion and crystal quality should be carefully examined for novel additives.

After converting the mesophase using additives, the diffraction quality of converted microcrystal samples could be evaluated in a microfocus X-ray diffraction. Once the crystal diffraction quality is confirmed, vitrified EM grids can be loaded into a scanning electron microscope (SEM) to assess the distribution of blotted crystals with the excess LCP residue prior to data collection in cryo-TEM. Under a certain circumstance when the conversion by additives might disrupt the crystal diffraction quality, a lower concentration of additives in a safe range to preserve the crystal integrity could be used to partially convert the phase to a relatively lower viscosity, although inadequate to produce a pure clean MicroED grid. Following Proteinase K, we prepared EM grids of vitrified cholesterol microcrystals grown in LCP for further MicroED data collection. Cholesterol microcrystals appeared in LCP at a ratio of 3:7 (w/w) cholesterol:monoolein, higher than the ratio of 1:9 (w/w) cholesterol:monoolein present in the $A_{2A}AR$ crystallization conditions with LCP. It clears any concern that cholesterol may have crystallized and contaminated the final $A_{2A}AR$ microcrystals, even with MPD-induced phase conversion. Furthermore, high-resolution diffraction could be obtained from cholesterol following LCP conversion, which is likely to be extended to small molecules and pharmaceuticals crystallized *in meso*.

We present LCP-MicroED diffraction images and processed data using human $A_{2A}AR$ microcrystals. Human $A_{2A}AR$ is a model GPCR that has been used extensively in diffraction method development and validation, particularly in cases of using LCP as a carrier medium, such as LCP-SFX [163] and LCP-SMX [179]. Human $A_{2A}AR$ was crystallized in LCP at high concentration of PEG 400 (28%), which resulted in average crystal sizes of 5 x 5 x 2 μm^3 as previously described for LCP-SFX data collection [163]. Due to the high concentration of PEG 400, the matrix was intermediate to that of cubic and sponge phase, therefore we empirically determined that 7% MPD was sufficient to reduce the viscosity of the mesophase for blotting and vitrification onto EM grids. We were able to visually confirm the presence of the $A_{2A}AR$ crystals on the EM grids that retained the appearance and dimensions as observed prior to the MPD-induced phase conversion. Further, MicroED experiments recorded diffraction spots that were visible to approximately 4.5 Å (Figure 5-5 D). The indexed unit cell dimensions were similar to previously published results, suggesting that these were $A_{2A}AR$ diffraction patterns because no other protein crystals were blotted onto the grids at the time of experiment (with the exception of Proteinase K), and no other components of the LCP matrix could have possibly diffracted with such unit cell parameters.

The microfocus X-ray diffraction image (Figure 5-5 E) shows that the $A_{2A}AR$ crystals diffract well even after undergoing the treatment described here, whereas when analyzed by MicroED, the diffraction only extended to 4.5 Å. This suggests that the weaker diffraction seen by MicroED is not a result of the sample processing procedures, but rather due to other factors, most likely due to the thickness and shape of the $A_{2A}AR$ crystals. MicroED requires the microcrystals to be thin (on the order of a few hundred

nanometers) for the electron beam to penetrate and produce reliable diffraction data. In addition, while not a problem for Proteinase K, excess mesophase surrounding the $A_{2A}AR$ microcrystals could also increase the thickness and absorb the electron beam. In the case of $A_{2A}AR$, the average crystal dimension was around 2-5 μm, which was too thick. For microcrystals on the grid that are thin enough for MicroED, the length and width are also greatly reduced to support data collection at high tilt angle. Traditionally, pipetting, vortexing, sonication and micro-seeding are applied to break large protein crystals into smaller fragments or microcrystals. Nevertheless, they barely work for highly viscous sample including LCP. Therefore, the future work of this study would rely on the method optimization for creating small and thin microcrystals on the EM grids.

Regarding hydrolysis of host lipids from the mesophase using lipase, it is capable of producing thinner EM grids than those made by additive-induced phase conversion because the resulting water/glycerol liquid phase with microcrystals enriched is much easier to be blotted away on the filter paper. As seen in Figure 5-3, the grid window of Proteinase K microcrystals is more visible from the lipase-treated sample with smaller average crystal size and clearer shape, compared with the MPD-treated sample. It has been previously shown that lipase treatment of bacteriorhodopsin crystals could be performed without degrading the crystals [180]. However, lipid molecules may interact with membrane proteins and mediate crystal contact, which is common in crystallizing GPCRs in LCP [164, 181]. The lipase treatment method may cause deleterious effects to the target membrane protein microcrystals. Thus, the strategy of lipase treatment should be evaluated extensively for different membrane protein targets prior to LCP-MicroED studies.

Another strategy to generate thin EM grids for LCP-MicroED analysis is to apply cryo-focused ion beam (cryo-FIB) milling to the vitrified samples prior to MicroED data collection. The combination methodology of cryo-FIB and MicroED has been applied to several structure determinations by MicroED without damage to the underlying crystal lattice [176-178]. With cryo-FIB milling, the large and thick crystals can be milled leaving thin lamellae (around several hundred nanometer) that still show reasonable areas for diffraction. Cryo-FIB can significantly benefit the application of LCP-MicroED since it bypasses the time spent on growing and optimizing many microcrystals, which are often hard to be identified under a light microscopy. Both large crystals and microcrystals embedded in the less viscous lipid mesophase can be directly applied onto the EM grid, milled in a cryo-FIB-SEM, and screened in a cryo-TEM for MicroED analysis. This combination method was recently applied to bacteriorhodopsin to generate high-resolution while low-completeness diffraction dataset [183]. Later, the structure of human $A_{2A}AR$ was determined at 2.8 Å by MicroED of a single crystal grown in LCP followed by cryo-FIB milling, where a novel cholesterol binding site was identified [184].

MicroED has shown promises in the identifying new structural information by improving resolution of poorly ordered samples [96], resolving structures with minimal radiation damage [174] and modeling charge within structures [95, 182]. LCP-MicroED paves the way for a robust method to extend the application of MicroED to the challenging membrane proteins, such as GPCRs, that are more successfully crystallized in LCP instead of vapor diffusion. Structures of GPCRs are often resolved to high resolution by LCP crystallization, which are necessary for investigating the interaction with ligands or pharmaceuticals. Because GPCRs are generally difficult to be expressed, purified and

crystallized, it is important to make great use of the limited protein resources. As mentioned above, the structure of human $A_{2A}AR$ was determined at high resolution by applying LCP-MicroED methods to a single crystal of $A_{2A}AR$ and new cholesterol binding site was identified, which demonstrates the success and potential of LCP-MicroED.

CHAPTER 6: MICROED STRUCTURE DETERMINATION OF ORGANIC SEMICONDUCTOR THIN FILMS

6.1 Introduction

In chapter 5, microcrystal electron diffraction (MicroED) was applied to collecting and phasing diffraction data of a small organic molecule cholesterol. Despite its original development for biological samples, MicroED has been extensively applied in the field of small organic molecules such as natural products and pharmaceuticals [120-123]. It can be challenging to grow crystals of small organic molecules to the sizes (at least 1×10^3 μm^3) necessary for conventional single crystal X-ray diffraction (SXRD), however MicroED is able to collect diffraction data from crystallites several orders of magnitude smaller than what is required for SXRD [185]. The application of MicroED facilitates rapid structure determination by bypassing the difficult and time-consuming crystallization steps, and even work well with powders from in-house purification or purchased from chemical supplier [120-123].

Organic semiconductor films consist of organic molecules that conduct electricity through charge injection or excitation with light. They are active components in electronic devices such as organic photovoltaics (OPV) [186-188] and organic field effect transistors (OFET) [189-191]. The optoelectronic responses of these devices are sensitive to the packing geometry of the organic semiconductors in the functional films [192-194], so it is necessary to determine their crystal structures in order to understand and, ultimately, predict device properties, involving charge transport mechanism, conductance, and

mobility. However, it is difficult to apply SXRD to these organic semiconductors since they can hardly be crystallized to sufficient dimensions.

Thus, MicroED was applied to high-resolution structure determination of three organic semiconductors for the application of organic optoelectronic devices: two rylene bisimides [195-198] - dipyrrolodine perylene diimide (dPyr PDI) [199, 200] and dicyano naphthalene diimide (dCN NDI) [201] - and a diketopyrrolopyrrole (dDPP) [202-204]. These three structures show the unit cell parameters and crystal packing geometry from unrefined powders of each organic semiconductor. It demonstrates that the application of MicroED to the organic semiconductors is successful and ideal for investigating the structure-activity relationships on compounds whose properties depends on the relative spatial arrangement in the solid state. Besides, grazing-incidence wide-angle X-ray scattering (GIWAXS) was applied to the active films in order to determine whether the films contain the same unit cell as those determined by MicroED and details on crystal packing orientation with respect to the substrate. Together, these techniques offer a model for how organic semiconductors organize in the thin films.

6.2 Materials and Methods

6.2.1 Synthesis of Chemicals

dPyr PDI [200], dCN NDI [201], and dDPP [199] were synthesized using published protocols. In brief, thin-layer chromatography was applied using aluminum sheets pre-coated with a thin layer of silica gel 60 (EMD 40-60 mm, 230-400 mesh with 254 nm dye). Silica gel (BDH 60 Å) was used for flash column chromatography. All solvents were dried prior to use, and all reactions were carried out under Argon atmosphere using standard Schlenk

techniques. Deuterated solvents (Cambridge Isotope Laboratories Inc.) were used as received. Proton NMR spectra were obtained on Bruker AVANCE 300 MHz spectrometer and all spectroscopic data were consistent with the previous reports.

6.2.2 MicroED Sample Preparation

3 μL of 10 mM each organic semiconductor crystals in toluene solutions were applied onto CF400-Cu continuous carbon grids (Electron Microscope Sciences) for 1 min and then dried from the grid edge using filter paper (Whatman). Grids were then loaded into a Titan Krios electron microscope (Thermo Fisher Scientific) operated at 300 kV and cryogenic temperature and equipped with a CETA D detector (Thermo Fisher Scientific).

6.2.3 MicroED Data Collection

MicroED data were collected using standard protocols [91, 99] with slight alterations. Low-dose settings for the Titan Krios electron microscope were applied throughout screening and data collection. With the search mode at at low magnification (660 X), the overall quality of grid and promising microcrystals that were well-separated from other crystals were screened. Once single microcrystal was identified, it was placed at the center of the electron beam and initial diffraction pattern was recorded by switching the microscope into the exposure mode, where the diffraction mode is applied. The selected area aperture and beam stop were inserted prior to switching the microscope into the exposure mode. If a high quality diffraction pattern was obtained, the microscope was switched back into the search mode and the maximal tilt range for data collection (i.e. the range in which the microcrystal remained centered without other crystals or grid bars in the beam) was determined. Upon switching the microscope into the exposure mode, MicroED data were

collected by continuously rotating the stage within the maximal tilt angles as the detector was continuously acquiring diffraction images.

6.2.4 MicroED Data Processing and Structure Determination

Diffraction data were converted from movies to images in SMV format [112]. XDS was used for data indexing, integration, merging and scaling [168]. Structures were phased by direct methods using SHELXT [205], refined using SHELXLE [206], and analyzed in Mercury [207].

6.2.5 GIWAXS Sample Preparation

Thin films of dPyr PDI and dDPP were prepared by drop casting 60 μL of 10 mM toluene solutions onto clean glass slides and drying in air. dCN NDI was thermally evaporated on glass slides using a RADAK I thermal evaporator. Glass slides were sonicated in detergent water for 10 min, acetone for 10 min, 2-propanol for 10 min, and then plasma cleaned in a Harrick plasma cleaner PDC-32 G for 10 min prior to use.

6.2.6 GIWAXS Data Collection and Processing

GIWAXS was performed at the NSLS-II beamline 12-ID SMI. Films on glass slides were illuminated by an X-ray beam at a wavelength of 0.77 Å and with beam size of 200 μm by 30 μm in a grazing incident geometry with an incident angle of 0.1º. Scattering patterns were recorded on a Pilatus300K detector at a distance of 273.9 mm downstream of the substrate. The detector scanned across the q-range required for the work and multiple scattering patterns were stitched together to form the final scattering data. Plots of integrated 1D intensity versus the scattering vector q were acquired through in-house beamline analysis program. Based on the crystal structures from MicroED, the

experimental GIWAXS data were indexed using in-house software to estimate the molecular orientation with respect to the substrate.

6.3 Results and Discussions

The first organic semiconductor we analyzed by MicroED was dPyr PDI whose structure had been resolved previously by SXRD, and hence it was ideal for validating our experimental approach. When the grid was loaded into the TEM and screened in the low-magnification search mode, it was found to contain single nanocrystalline materials (Figure 6-1 A) for examining the diffraction ability. After switching the microscope into the exposure mode and acquiring an initial still diffraction pattern, crystallites that diffracted well with sharp spots and high resolution were used to collect continuous rotation MicroED data at a total electron dose of approximately 5 $e^-/Å^2$ for each data set [91]. It is important to find that the average dimensions of crystals suitable for high-resolution MicroED data collection were estimated to be 0.8 μm x 0.3 μm x 0.1 μm, which is around 4 x 10^4 times smaller in volume than the samples used for SXRD. The diffraction spots extended to approximately 0.60 Å (Figure 6-2 A). Three dPyr PDI diffraction data sets were merged for generating the final data set for phasing using direct methods. The best solution was provided that dPyr PDI crystals were processed in the Cc space group with a unit cell of a=22.05 Å, b=10.76 Å, c=9.34 Å, α=γ=90° and β=101.287°. Most hydrogen atoms were resolved in the MicroED structure. It shows deviations of 0.68%, 1.01%, 0.64%, and 0.55% between the MicroED and previous SXRD data for a, b, c, and β, respectively (Appendix Table C, Figure 6-3 A), which

confirms that the MicroED structure is nearly identical to the published X-ray structure and validates the methods.
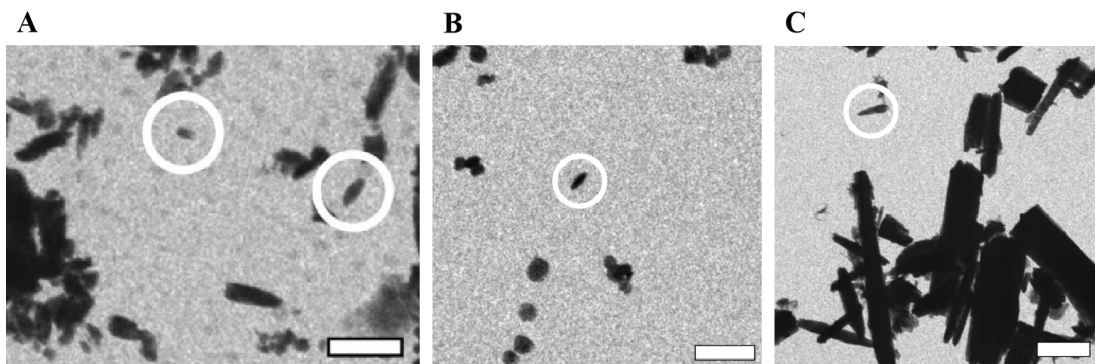
Figure 6-1: Representative images of crystallite samples viewed on the EM grid. (A) dPyr PDI, (B) dCN NDI and (C) dDPP. Samples used for electron diffraction are circled. Scale bars represent 2 μm.

MicroED was subsequently applied to an unknown crystal structure of a naphthalene diimide, dCN NDI, which is increasingly adopted in OFET [198, 209] as an air-stable, electron acceptor with a low lying LUMO [201, 210]. Because of the shape of dCN NDI crystallites, they adopted a preferred orientation on the grid surface (Figure 6-1 B) and eight diffraction data sets were merged into the final data set to reach an overall completeness of 71.5% at a resolution of 0.57 Å (Figure 6-2 B, Appendix Table C). Despite the relatively low completeness of the data, the structure of dCN NDI was determined by direct methods in the space group $P2_1/c$, with a unit cell of a=8.09 Å, b=6.39 Å, c=11.63 Å, $\alpha=\gamma=90°$ and $\beta=104.711°$. Hydrogen atoms were not well resolved and, nevertheless, some of them were calculated based on the chemistry.
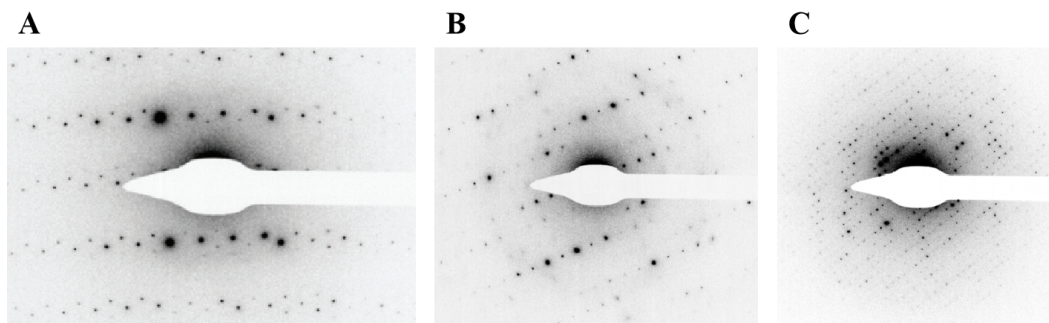
Figure 6-2: Representative diffraction patterns during MicroED data collection. (A) dPyr PDI diffracted to 0.6 Å; (B) dCN NDI diffracted beyond 0.6 Å; and (C) dDPP diffracted to 0.90 Å.

The third organic semiconductor, dDPP, whose crystal structure has not been previously determined, was analyzed due to its potentials to show singlet fission (SF) with high yields and lifetimes in films [199]. SF is highly influenced by packing geometry [211-213], and lack of a crystal structure in our previous studies made it impossible to correlate SF yields and lifetimes to the solid-state packing. A dDPP molecule is composed of a diketopyrrolidine core, two chiral alkyl side chains branching from the core nitrogens, and two diamidopyridine (DAP) moieties designed for hydrogen bonding with adjacent molecules. For dDPP nanocrystals identified for MicroED (Figure 6-1 C), they were diffracted to approximately beyond 0.90 Å (Figure 6-2 C), and data sets from seven crystals were merged (Appendix Table C) together to generate a final data set at 0.90 Å. The structure was determined using direct methods in the space group $P2_1/n$, with a unit cell of a=15.09 Å, b=19.55 Å, c=34.77 Å, $\alpha=\gamma=90°$ and $\beta=94.627°$. Not all hydrogen atoms were well resolved and, nevertheless, some of them were calculated based on the chemistry.
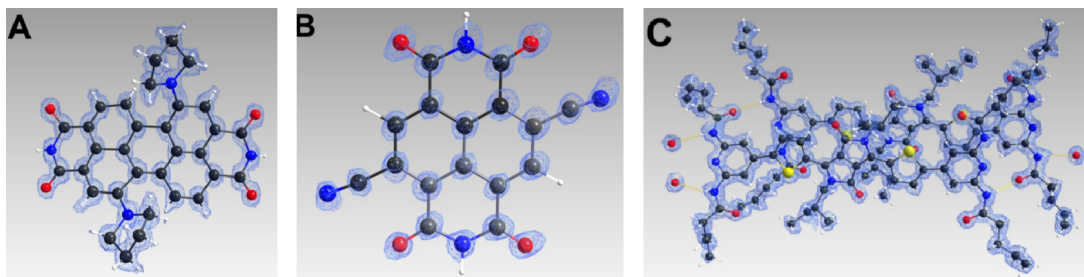
Figure 6-3: F_o density maps with structures displayed. (A) dPyr PDI at 0.60 Å resolution; (B) dCN NDI at 0.57 Å resolution; and (C) dDPP at0.90 Å resolution; all contoured at 1.5σ. Atom color: carbon, black; nitrogen, blue; oxygen, red; hydrogen, white.

CrystalDiffract was used for generating simulated powder diffraction pattern of the MicroED structures of these three samples. GIWAXS data (Figure 6-4 A) of each sample were integrated to plot a spectra of 1D intensity versus the scattering vector. Each integrated 1D GIWAXS spectrum was compared with the corresponding MicroED-derived powder diffraction pattern. Despite some deviations in dDPP, the peaks from the GIWAXS matches well with those from the simulated powder diffraction pattern (Figure 6-4 B-D), which corroborates that crystals on the TEM grids are in good agreement with those prevalent in the films and offers a pathway for determining orientation.
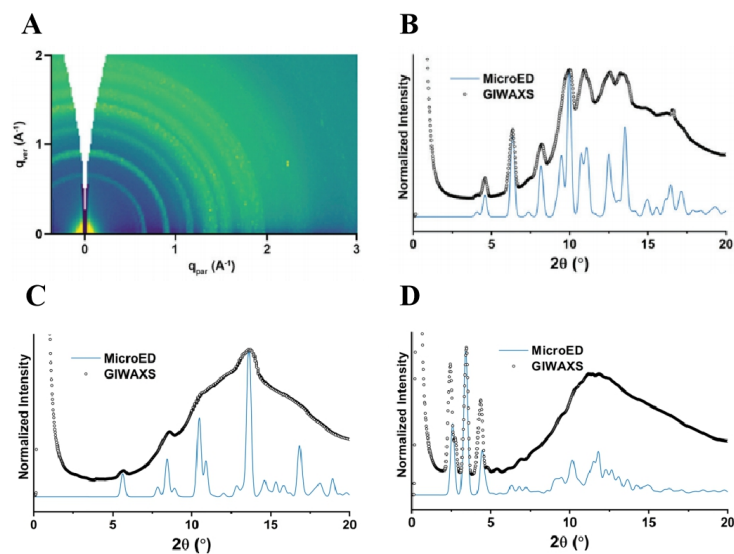
Figure 6-4: Combining MicroED data with GIWAXS data. (A) GIWAXS pattern of dPyr PDI thin film drop-cast on clean glass slides shows predominantly rings, indicating no preferential orientation of crystals with respect to the substrate. Comparison of GIWAXS pattern with simulated powder diffraction pattern derived from MicroED data on (B) dPyr PDI; (C) dCN NDI; and (D) dDPP; all shows good agreement.

In the MicroED structure of dPyr PDI (Figure 6-5), the molecular packing is cofacial and slip-stacked instead of the common herringbone pattern in the molecular packing of rylenes [198, 208]. This may be a result of steric crowding imposed by the pyrrolidine groups, which causes bowing and prevents hydrogen bonding between the imide groups which is commonly observed in herringbone packing of rylenes. The GIWAXS pattern shows predominantly well-defined, radially uniform rings, indicating no preferential orientation of crystallites in the films with respect to the glass slides.
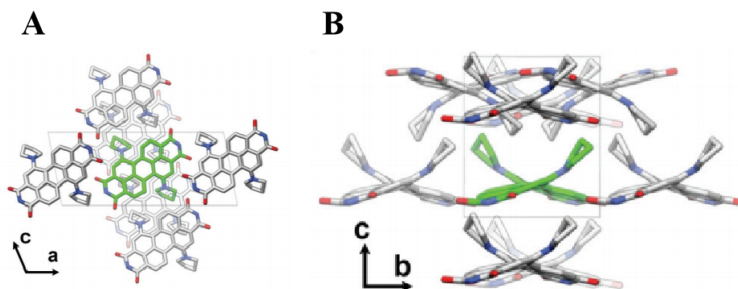
Figure 6-5: Unit cell of MicroED-determined crystal structure of dPyr PDI, with asymmetric Cc space group packing viewed along (A) b-axis and (B) a-axis with a single molecule of dPyr PDI shown in green color, carbon in grey, nitrogen in blue and oxygen in red.

Regarding dCN NDI, it was found to organize into herringbone motif (Figure 6-6). Each dCN NDI molecule makes a total of four hydrogen bonds with four other adjacent molecules between the carbonyl and amide groups of adjacent molecules (Figure 6-6 C). The hydrogen bonds are uniformly 1.85 Å and 175.151°. Smooth films of dCN NDI for GIWAXS was carried out on the glass slides by thermal evaporation rather than drying in air, because dCN NDI is not well soluble in common organic solvents. The GIWAXS pattern shows the films have preferential out-of-plane orientation with respect to the glass slide, which is perpendicular to the substrate, along the (113) plane and minor orientation along (100), (112), and (102). π-stacking in dCN NDI occurs either 60° or parallel to the substrate (Figure 6-6 B).
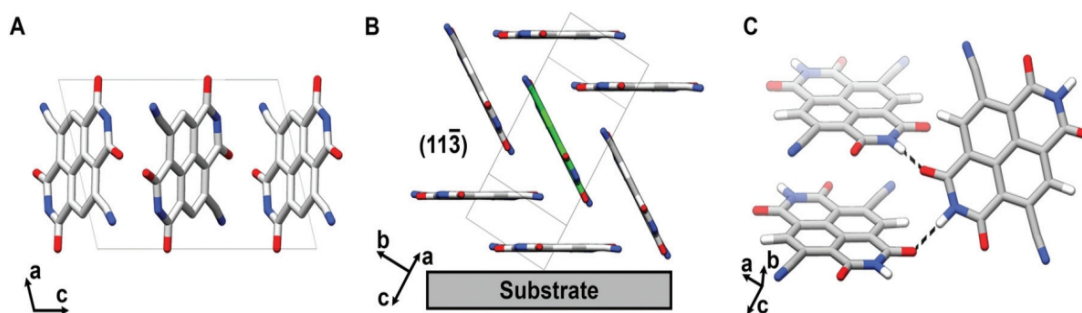
Figure 6-6: Unit cell of MicroED-determined crystal structure of dCN NDI, with asymmetric P2$_1$/c space group packing viewed along (A) b-axis and (B) the preferential out-of-plane orientation with respect to the substrate with a single molecule of dCN NDI shown in green color. (C) Four hydrogen bonds form within a dCN NDI molecule between the carbonyl and amide groups of adjacent molecules. Atom colors: carbon in grey, nitrogen in blue, oxygen in red and hydrogen in white. Hydrogen have been omitted from (A) and (B) for clarity.

In the MicroED-solved crystal structure of dDPP, hydrogen bonds between neighboring dDPPs occurs between the amide groups of the DAP substituents. Only one of the two amide groups on each DAP is tied up in hydrogen bonding with the amide groups of DAP from a neighboring dDPP molecule. Hydrogen bonds are 2.20 Å and 156.741° or 1.98 Å and 169.601°. Molecules adopt π-stacking which are 3.55 Å, 3.66 Å, or 3.36 Å if measured from different views of asymmetric unit (Figure 6-7 B). This supramolecular interaction in dDPP is the similar to the previously reported structure of mDPP, which contains one DAP group instead of two and whose crystal structure has been previously determined by SXRD [214]. The comparable superstructures between dDPP and mDPP are the result of a conserved supramolecular assembly mechanism. GIWAXS pattern shows preferred out-of-plane orientation along the planes of (011) (Figure 6-7 A), (012), and (013), which orient close to each other. This geometry places the a-axis, the direction of π-stacking, in a preferential orientation which is parallel to the substrate.
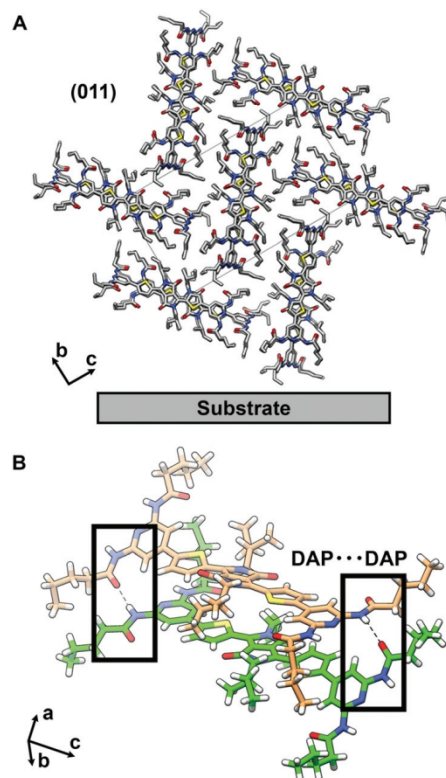
Figure 6-7: Unit cell of MicroED-determined crystal structure of dDPP, with asymmetric P2$_1$/n space group packing viewed along (A) a-axis. (B) Hydrogen bonds form between the DAP groups of adjacent, asymmetric dDPP molecules. A single dDPP molecule is shown in green. Atom colors: carbon in grey, nitrogen in blue, oxygen in red and hydrogen in white.

The first and third major peaks of the simulated powder diffraction pattern are shifted slightly toward larger 2θ, however the second major peak matches well with the GIWAXS pattern (Figure 6-4 D). The first peak in the simulated powder diffraction pattern actually corresponds to two overlapping peaks of (002) and (011). The second and third peaks in the simulated powder diffraction pattern correspond to (012) and (013), respectively. Because the mismatches are all in 0kl, there may be some variations in the b and c axes of the samples used for MicroED and GIWAXS, likely due to interactions

with the glass substrates during crystallization. The a-axis falls along the π-stacking direction and can be constant, whereas b and c axes may be variable due to different packing arrangements along the flexible alkyl chains. Since only the best diffracting crystals were selected for data collection, they are likely to represent only one of multiple polymorphs in the sample. This highlights the importance of combining other methods, such as GIWAXS, with MicroED to understand molecular packing in the films.

6.4 Conclusions

MicroED was used to solve the structures of three organic semiconductors - dPyr PDI, dCN NDI and dDPP - to high resolution from thin films without the need for producing large crystals. Future investigations will involve absolute structure determination through polymorphism, modeling dynamic scattering [215] and multi-component supramolecular crystals [199, 214]. In combination with GIWAXS, MicroED is a valuable tool to investigate 3D molecular packing within the films. However, it must be cautious to correlate crystals and films as shown in the case of dDPP.

CHAPTER 7: BEAM-SENSITIVE METAL-ORGANIC FRAMEWORK STRUCTURE DETERMINATION BY MICROED

7.1 Introduction

Metal-organic frameworks represent a new class of microporous materials. Zeolitic imidazolate frameworks (ZIFs) are a subcategory of highly crystalline metal-organic framework materials bearing zeolitic topologies finding wide variety of applications in adsoption, separation, gas storage, and heterogeneous catalysis. ZIF-8 with sodalite (SOD) topology has been among the most extensively studied ZIF material [216, 217]. Many MOF structures are determined by single crystal X-ray diffraction [218], however this requires the growth of large crystals to withstand the effects of radiation damage and allow the collection of sufficient diffraction data. In some cases, powder diffraction data alone can be used for structure determination; however, this can be a more difficult approach. Because MOFs are often synthesized as micro and nanocrystals, the use of transmission electron microscopy offers a way to study the MOF structure directly from the synthesized material. Collecting data at cryogenic temperatures improves data collection from beam-sensitivity MOFs, and this can be further improved by using low dose conditions (in the range of 4-12 $e^-/\text{Å}^2$ total exposure) and direct electron detection [219-222]. Imaging thin sections of MOF fragments provides an excellent approach for directly visualizing interfacial, surface, and local structure of these beam sensitive materials.

As an alternative to imaging based TEM structure determination methods for overall structure determination, electron diffraction can also be employed for high-resolution structure determination from crystalline materials, including MOFs [223-226]. In recent years,

electron microscopy methods, such as microcrystal electron diffraction (MicroED) [185], have been employed to determine the structures of a variety of samples from various fields including materials science, organic chemistry, and structural biology [120-122, 185, 215, 227, 228]. A key benefit of electron diffraction is that crystals several orders of magnitude smaller than those used for single crystal X-ray methods can be used for high-resolution structure determination, which bypasses the often times difficult and time-consuming steps of growing large single crystals for structural analysis.

As is the case with imaging, very low electron dose must be used when beam-sensitive materials are being analyzed by electron diffraction. MicroED is a technique that uses ultra-low dose to collect electron diffraction data at cryogenic temperature from very small microcrystals as the stage continually rotates [91, 185]. By using sufficiently low dose and continuous rotation of the sample, large volumes of reciprocal space can be sampled before the crystal is overcome by radiation damage [89, 174], which allows for the indexing and integration of the diffraction data set.

In this work, we make use of ZIF-8 as a model beam-sensitive MOF [219] and demonstrate that MicroED data collection procedures are suitable for rapid structure determination of these materials at very high-resolution. The ZIF-8 MicroED data were collected and processed using a cryo-TEM equipped with a high-speed camera for diffraction data collection, and did not require any other specialized hardware. The ZIF-8 data from a single crystal were processed to 0.87 Å and an ab initio structure was determined, which compares well to previously solved structures of ZIF-8 determined by single crystal XRD.

7.2 Materials and methods

7.2.1 Synthesis

ZIF-8 crystals were synthesized by modifying the procedure reported by Tanaka and co-workers [229]. All chemicals were purchased from Sigma-Aldrich (U.S.A) unless otherwise stated. Zinc nitrate hexahydrate [$Zn(NO_3) \cdot 6H2O$] (98%, 0.297 g) was dissolved in 20 ml of deionized water and 3.28 g of 2-methylimidazole (MeIM) was dissolved separately in 20 ml of deionized water. The separate dissolved solutions were stirred for 10 min. The imidazole solution was poured into the zinc solution and stirred for 1h. The solution quickly became cloudy and a suspension was obtained. The final molar composition of the synthesis solution was Zn^+:MeIM:water = 1:40:2,227. After mixing, the resultant white solution was centrifuged at 5,000 g for 10 min and washed with methanol (VWR, 99.8%) three times. The Products were then dried overnight in vacuum oven at room temperature. The morphology and crystallographic properties of ZIF-8 crystals was examined using SEM (Amray 1910) and X-ray powder diffraction (XRD, AXS-D8, Bruker) with a scan step of 0.015° using Cu Kα radiation (λ = 1.543 Å), respectively.

7.2.2 MicroED sample preparation

Dried, lightly ground ZIF-8 crystals (0.004 g) were placed in a glass vial containing 100 ml of methanol to form 0.004% (w/v) colloidal ZIF-8 solution. 2 µL of prepared solution were applied onto the surface of glow-discharged Quantifoil 300-mesh 2/2 grids, and dried by evaporation for a few minutes. Grids were subsequently placed into autoloader cartridges under liquid nitrogen and loaded into a Titan Krios cryo-TEM (Thermo Fisher Scientific) operated at 300 kV and equipped with a CETA D detector (Thermo Fisher Scientific), which was used for diffraction data collection.

7.2.3 MicroED data collection

MicroED data were collected using similar protocols described previously [91, 99]. Low-dose settings for the Titan Krios cryo-TEM were employed throughout screening and data collection steps. Crystals that were well-separated from other crystals on the grid were identified using low magnification search mode. Upon the identification of single microcrystals suitable for data collection, initial diffraction patterns were collected by switching the microscope into diffraction mode using the pre-calibrated and aligned exposure mode within the low dose settings. If a high quality diffraction pattern was obtained, the microscope was put back into search mode and the maximal tilt range of for data collection (i.e. the range where crystal remained centered and no other crystals or grid bars moved into the path of the beam) was determined. The stage was tilted to the maximum angle and the selected area aperture and beam stop were inserted. The selected area aperture size was selected such that the entire crystal was within the aperture while also minimizing the area surrounding the crystal. MicroED data sets were collected by continuously rotating the crystal (0.7 degrees per second) in the beam as the CETA D camera was continuously acquiring diffraction frames at a rate of 2 seconds per frame. The exposure rate was approximately 0.02 e$^-$/Å^2 per frame.

7.2.4 Data processing and structure determination

Diffraction data sets were converted to SMV format [112] and indexed, integrated and scaled using XDS [168]. The structure of ZIF-8 was ultimately determined with data from a single crystal by direct methods using SHELXT [205] followed by refinement using SHELXLE [206].

7.3 Results and discussions

The morphology of the synthesized ZIF-8 crystals were observed by SEM as shown in Figure 7-1 A. ZIF-8 crystals shows an approximate isotropic hexagon envelope with particle size range of 200 to 800 nm. The XRD powder diffraction pattern from prepared ZIF-8, is consistent with that of standard ZIF-8 samples reported in literature [230-232] (Figure 7-1 B), thus revealing the successful synthesis of ZIF-8. The as-synthesized ZIF-8 particles were suspended in methanol, and this solution was directly deposited on a holey carbon coated EM grid. Following the evaporation of the methanol from the surface of the grid, the EM grid was loaded into the Titan Krios cryo-TEM and the ZIF-8 particle distribution was assessed. The distribution of particles is important for MicroED because individual crystals need to be isolated for data collection to ensure that data from multiple crystals is not collected. When viewed at lower magnification, it can be seen that while many of the ZIF-8 nanoparticles preferred to clump together, there were a significant fraction of particles that were well-dispersed and amenable to data collection (Figure 7-2 A).
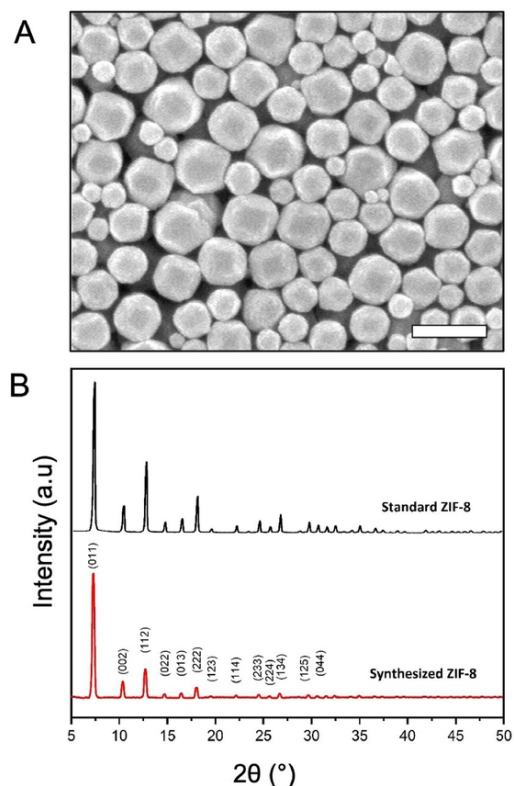
Figure 7-1: Characterization of synthesized ZIF-8 particles. (A) SEM micrograph of synthesized ZIF-8 and (B) XRD powder diffraction data of standard [232] and synthesized ZIF-8 crystals. Scale bar in A represents 1 µm.

ZIF-8 crystals ranging in diameter from approximately 200 nm to 800 nm could be seen on the grids. While screening crystal quality by taking initial diffraction patterns from each crystal, it was found that the diffraction data became qualitatively worse (i.e. lower resolution and more background noise from the increased diffuse scattering) as the size of the ZIF-8 nanoparticles approached the larger end of the size range (approximately 400 nm to 800 nm in diameter). The crystals which were on the smaller end of the size range (approximately 200 nm to 400 nm in diameter), generally produced better diffraction, and crystals that produced the data with the highest resolution and sharpest diffraction spots were used for data collection (Figure 7-2 B). Ultimately, the crystal that produced the

data used for structure determination was approximately 200 nm across and 200 nm thick. This is significantly smaller than the size of crystals used for single crystal X-ray methods which have dimensions on the order of tens to hundreds of micrometers.
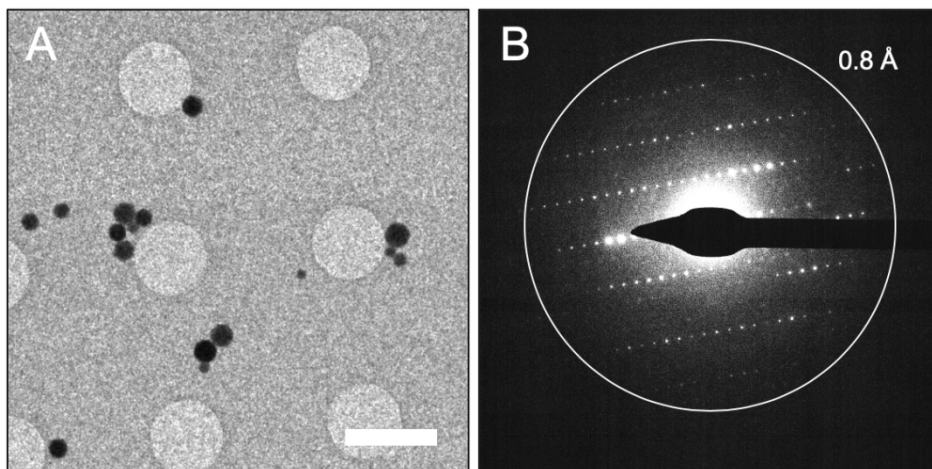
Figure 7-2: High-resolution diffraction patterns obtained from ZIF-8 particles. (A) A low magnification search of the EM grid shows well-dispersed ZIF-8 crystals on the surface of the holey-carbon. (B) Electron diffraction pattern collected from a high-quality ZIF-8 nanocrystal shows sharp and well-separated spots extending to high-resolution. This diffraction pattern represents what was typically seen for crystals which were used for MicroED data collection. Scale bar in A represents 3 μm.

In a previous study, it was demonstrated that ZIF-8 samples rapidly lose their crystallinity in the electron beam [219]. Therefore, the total exposure for each crystal during the course of MicroED data collection was kept to ~1 e⁻/Å², which is much lower than the what was used the recent high-resolution imaging studies of ZIF-8 particles [219-221]. The time to collect a MicroED data set is relatively short (in the range of 1 to 2 min for the ZIF-8 samples), and because many of the ZIF-8 nanocrystals in the 200 to 400 nm size range diffracted to beyond 1 Å, we could specifically focus on crystals of this size. This allowed the collection of many high-resolution data sets in a relatively short time (about

1-2 min per crystal). With sample loading, microscope setup, crystal screening for only the best crystals, and data collection are all taken into account, we collected twelve high-quality data sets in approximately 2 h. The ability to rapidly collect a large number of data sets is a great advantage for the use of MicroED in the study of MOFs and other organic, biological, or beam-sensitive crystalline samples.

While data from many crystals were collected, because of the high symmetry of the ZIF-8 nanocrystals, data from a single crystal could ultimately be used for structure determination to 0.87 Å (Appendix Table D). When processed, data from one single crystal resulted in a data set that was 99.8% complete. It should be noted that many other crystals in the data sets collected could also be used for structure determination as well. The direct methods solution found was in space group I-43m and five atoms were placed within the asymmetric unit, which is what would be expected for ZIF-8. Following structure refinement ($R_1 = 0.1779$, $wR_2 = 0.3411$, and GooF = 1.011) and expanding the asymmetric unit the structure of ZIF-8 can be seen, with each zinc being coordinated by four MeIM ligands and each ligand bridges two zinc atoms (Fig 25). This produces the typical SOD zeolite type structure with hex-zinc-atomic open ring channels of 3.4 Å in diameter. Both the unit cell parameters and atom positions of the MicroED structure compare well with previous ZIF-8 structures determined by single crystal X-ray diffraction [230, 233, 234], with differences in bond lengths between the MicroED structures and the X-ray structures being in the range of 0.03 to 0.07 Å.
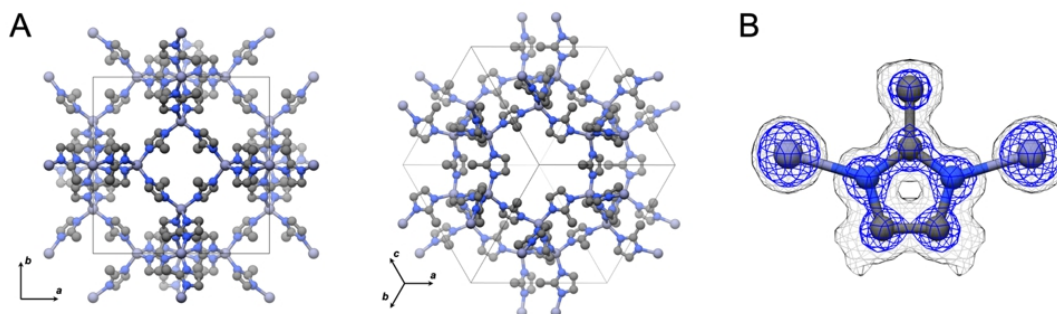
Figure 7-3: MicroED structure and unit cell of ZIF-8, (A) viewed in different orientations and (B) with the surrounding density. The maps in panel B are contoured at 0.8σ (grey) and 2.4σ (blue). Hydrogens are omitted from the figures for clarity.

7.4 Conclusions

The rapid and accurate determination of ZIF-8 structure to high resolution paves the way for understanding the properties of MOF materials using cryo-EM method MicroED. A benefit of the MicroED method is that it can be performed on many of the same cryo-TEM instruments used for high-resolution imaging. Future combinations of MicroED structure determination with low-dose high-resolution imaging could allow multi-leveled analysis of beam-sensitive materials from a single sample, where diffraction is used for high-resolution structure determination and the images can analyze surface and interfacial structure. This would further cement low-dose cryo-TEM as an invaluable tool for understanding MOF structure and function.

CHAPTER 8: CONCLUSIONS

Cyanobacterial bicarbonate transporters SbtA and BicA are active components of CCM, which accumulate Ci, primarily bicarbonate, into the cells by approximately 1,000-fold than the extracellular level, and the pooled Ci can be utilized by the primary carbon fixing enzyme Rubisco to improve photosynthesis. The determination of their structures is important to investigate the bicarbonate transport mechanisms, and facilitate engineering these transporters to enhance the photosynthetic products in cyanobacteria or in plants. *E. coli* was selected as the expression host for SbtA-6803 and BicA-6803, as well as the SbtA-6803 regulator protein SbtB-6803, and optimization of expression and protein purification conditions was performed. Single particle cryo-EM or protein crystallography was carried out for each target.

As described in chapter 2, SbtB-6803 was highly expressed in *E. coli*, and the standard IMAC and SEC purification were capable of acquiring approximately 16.0 ± 3.0 mg purified samples per liter of BW25113 cell culture. Thus no optimization of protein expression and purification was performed. Tetragonal or trigonal crystal types were grown in the sitting drop vapor diffusion plates without or with the addition of reducing agent TCEP, respectively. They were resolved at 1.8 Å and 2.01 Å by synchrotron single crystal X-ray diffration, showing a dimer and trimer in the ASU, respectively. The dimer conformation, not a canonical form of P_{II}-like signaling proteins, contains disulfide bonds at the C-terminal hairpin loop between C105 and C110, along with C94 and another C94 from adjacent ASU. Therefore, a dimer of dimer spacial arrangement occurs in this crystal form. The trigonal crystal form is the canonical form of P_{II}-like proteins, whose

structure is very similar to the published structure by other groups [129], except that the C-terminal loop is too flexible to be seen in the map. When comparing the monomeric structure of SbtB-6803 from both crystal forms, it is found that in the tetragonal crystal form, the C-terminus starting from C94 is less rigid and likely to promote the formation of intermolecular disulfide bonds and the dimer of dimer spacial arrangement.

In chapter 3, protein expression and purification conditions have been screened for SbtA-6803. After optimization, the yield of SbtA-6803 after membrane solubilization and purification using LMNG reached 0.1 ± 0.04 mg per liter of BL21 (DE3) pLysS culture. SEC elution profiles, meanwhile, confirm that LMNG helps to keep SbtA-6803 in a stable form. Besides, the SEC elution time of SbtA-6803 along with results from negative-stain EM indicates that the developed protocol is capable of yielding stable, homotrimeric complexes. After screening multiple conditions involving different adenosyl nucleotides and TCEP, addition of ADP without TCEP has been selected for generating the SbtA-SbtB complex for structural studies. Single particle cryo-EM was used for both SbtA-6803 and SbtA-SbtB complex. Although SbtA-6803 is too small to be properly aligned for data processing, SbtA-SbtB is well aligned during 2D classification and the features of SbtA-6803 and potential trimeric SbtB-6803 shown with the secondary structure information are clear on the class averages.

In chapter 4, optimization of the expression of BicA-6803 in *E. coli* and purification methods were performed, and approximately 6.5 ± 0.1 mg of purified BicA-6803 using DDM detergent can be obtained from one liter of C43 (DE3) Rosetta2 cell culture. SEC elution profiles confirm that DDM helps to keep BicA-6803 in a stable form. Besides, the SEC elution time of BicA-6803 along with results from negative-stain EM and

comparisons with other, full-length SulP/SLC26 family transporters indicate that the developed protocol is capable of yielding stable, homodimeric complexes. Crystals in the shape of plate have been obtained by vapor diffusion methods, however they diffracted to approximately 8 Å, which is too low but indicative of protein crystals.

For crystallization experiments of SbtA-6803 and BicA-6803, LCP was used, although often times the crystals grown are generally too tiny to withstand radiation damage from the X-ray beam during an X-ray diffraction experiment. An alternative approach for diffraction experiments, MicroED, has been developed for these tiny microcrystals grown in lipid mesophase. The new technique is termed as LCP-MicroED, however, prior to applying LCP-MicroED to the bicarbonate transporters, methods needed to be developed for LCP-MicroED.

Therefore in chapter 5, the first-time application of LCP-MicroED has been demonstrated. Various treatment methods, involving additives or lipase, have been performed to determine the best strategies for reducing the viscosity of LCP. MPD or lipase has been selected as the best for LCP conversion. The model protein Proteinase K was used and its structure was determined to 2.0 Å by MicroED, using both MPD and lipase to treat LCP. Additionally we showed that, using MPD to treat LCP, we could collect electron diffraction data from cholesterol and human A_{2A} adenosine receptor crystals at 1.0 Å and 4.5 Å, respectively.

In chapter 6, MicroED was extended to be applied to small organic molecules. Crystal structures of three organic semiconductors - dPry PDI, dCN NDI, and dDPP - were determined to high resolution from thin films by MicroED. It skips the time spent on producing large crystals and enables to collect electron diffraction data and determine

their structures in the same day. In addition, GIWAXS was used for determining the relative orientation of crystals in the thin films with respect to the substrate, as well as serving as reference for the accuracy of MicroED data. It has been concluded that the combination of MicroED with GIWAXS is important for understanding crystal packing in the films.

In chapter 7, MicroED was applied to a beam-sensitive model MOF sample, ZIF-8. ZIF-8 structure has been resolved to 0.87 Å using MicroED data from a single crystal. The MicroED-determined ZIF-8 structure matches very well with the previously determined ZIF-8 structure using SXRD, with RMSD bond length deviations less than 0.07 Å. The rapid and accurate determination of MOF structure is important to understand the properties of these beam-sensitive samples.

In the future work, screening expression level of SbtA-6803 can be performed in Rosetta2 (DE3) pLysS because Rosetta2 (DE3) strain showed improvement in the yield of SbtA-6803 at 37 $^{\circ}$C and the combination of pLysS and Rosetta2 is likely to be beneficial. With higher yield of SbtA-6803, crystallization can be set up using vapor diffusion methods. More single particle cryo-EM data of SbtA-SbtB complex need to be collected using A8-35 stabilized SbtA-6803 and ADP. It is generally assumed that the number of good particles with different views needs to be more than 100,000 for producing a high-resolution map. Optimization of crystallization conditions needs to be performed in the case of BicA-6803. New detergent, such as LDAO, will be tested for crystallizing BicA-6803. LCP-MicroED can be applied to the microcrystals of SbtA-6803 and BicA-6803 grown in LCP, along with other tough membrane protein targets, especially the GPCRs. Cryo-FIB milling can be incorporated into the developed LCP-

MicroED technique to generate thin crystal samples for electron diffraction. In addition, MicroED will be applied to other small organic molecules, including the engineered organic semiconductor thin films, engineered MOF sample, purified natural products, pharmaceuticals, and so on. MicroED can be used along with the low-dose high-resolution imaging settings from the same cryo-TEM for multi-leveled analysis of these materials.

REFERENCES

1.   Solomon, S., Plattner, G. K., Knutti, R., & Friedlingstein, P. (2009). Irreversible climate change due to carbon dioxide emissions. *Proceedings of the national academy of sciences*, 106(6), 1704-1709.

2.   Amouroux, J., Siffert, P., Massué, J. P., Cavadias, S., Trujillo, B., Hashimoto, K., ... & Wang, X. (2014). Carbon dioxide: A new material for energy storage. Progress in Natural Science: Materials International, 24(4), 295-304.

3.   Ding, M., Flaig, R. W., Jiang, H. L., & Yaghi, O. M. (2019). Carbon capture and conversion using metal-organic frameworks and MOF-based materials. Chemical Society Reviews, 48(10), 2783-2828.

4.   Albo, J., Vallejo, D., Beobide, G., Castillo, O., Castaño, P., & Irabien, A. (2017). Copper-based metal–Organic porous materials for CO2 electrocatalytic reduction to alcohols. ChemSusChem, 10(6), 1100-1109.

5.   Blankenship, R. E. (2014). Molecular mechanisms of photosynthesis. John Wiley & Sons.

6.   Geider, R. J., Delucia, E. H., Falkowski, P. G., Finzi, A. C., Grime, J. P., Grace, J., ... & Platt, T. (2001). Primary productivity of planet earth: biological determinants and physical constraints in terrestrial and aquatic habitats. Global Change Biology, 7(8), 849-882.

7.   Bryant, D. A., & Frigaard, N. U. (2006). Prokaryotic photosynthesis and phototrophy illuminated. Trends in microbiology, 14(11), 488-496.

8.   Rasmussen, B., Fletcher, I. R., Brocks, J. J., & Kilburn, M. R. (2008). Reassessing the first appearance of eukaryotes and cyanobacteria. Nature, 455(7216), 1101-1104.

9.   Paumann, M., Regelsberger, G., Obinger, C., & Peschek, G. A. (2005). The bioenergetic role of dioxygen and the terminal oxidase (s) in cyanobacteria. Biochimica et Biophysica Acta (BBA)-Bioenergetics, 1707(2-3), 231-253.

10.   Welsh, E. A., Liberton, M., Stoeckel, J., Loh, T., Elvitigala, T., Wang, C., ... & Aurora, R. (2008). The genome of Cyanothece 51142, a unicellular diazotrophic cyanobacterium important in the marine nitrogen cycle. Proceedings of the National Academy of Sciences, 105(39), 15094-15099.

11. Grossman, A. R., Schaefer, M. R., Chiang, G. G., & Collier, J. L. (1993). The phycobilisome, a light-harvesting complex responsive to environmental conditions. Microbiology and Molecular Biology Reviews, 57(3), 725-749.

12. Badger, M. R., & Price, G. D. (2003). CO_2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. Journal of experimental botany, 54(383), 609-622.

13. Karagouni, A. D., & Slater, J. H. (1979). Enzymes of the Calvin cycle and intermediary metabolism in the cyanobacterium Anacystis nidulans grown in chemostat culture. Microbiology, 115(2), 369-376.

14. Price, G. D., Badger, M. R., Woodger, F. J., & Long, B. M. (2008). Advances in understanding the cyanobacterial CO_2-concentrating-mechanism (CCM): functional components, Ci transporters, diversity, genetic regulation and prospects for engineering into plants. Journal of experimental botany, 59(7), 1441-1461.

15. Kupriyanova, E. V., Sinetova, M. A., Cho, S. M., Park, Y. I., Los, D. A., & Pronina, N. A. (2013). CO_2-concentrating mechanism in cyanobacterial photosynthesis: organization, physiological role, and evolutionary origin. Photosynthesis research, 117(1-3), 133-146.

16. Price, G. D. (2011). Inorganic carbon transporters of the cyanobacterial CO_2 concentrating mechanism. Photosynthesis Research, 109(1-3), 47-57.

17. Burnap, R. L., Hagemann, M., & Kaplan, A. (2015). Regulation of CO_2 concentrating mechanism in cyanobacteria. Life, 5(1), 348-371.

18. Higgins, C. F. (2001). ABC transporters: physiology, structure and mechanism–an overview. Research in microbiology, 152(3-4), 205-210.

19. Omata, T., Price, G. D., Badger, M. R., Okamura, M., Gohta, S., & Ogawa, T. (1999). Identification of an ATP-binding cassette transporter involved in bicarbonate uptake in the cyanobacterium Synechococcus sp. strain PCC 7942. Proceedings of the National Academy of Sciences, 96(23), 13571-13576.

20. Omata, T., Takahashi, Y., Yamaguchi, O., & Nishimura, T. (2002). Structure, function and regulation of the cyanobacterial high-affinity bicarbonate transporter, BCT1. Functional plant biology, 29(3), 151-159.

21. Shibata, M., Katoh, H., Sonoda, M., Ohkawa, H., Shimoyama, M., Fukuzawa, H., ... & Ogawa, T. (2002). Genes essential to sodium-dependent bicarbonate transport in cyanobacteria function and phylogenetic analysis. Journal of Biological Chemistry, 277(21), 18658-18664.

22. Reizer, J., Reizer, A., & Saier Jr, M. H. (1994). A functional superfamily of sodium/solute symporters. Biochimica et Biophysica Acta (BBA)-Reviews on Biomembranes, 1197(2), 133-166.

23. Price, G. D., Shelden, M. C., & Howitt, S. M. (2011). Membrane topology of the cyanobacterial bicarbonate transporter, SbtA, and identification of potential regulatory loops. Molecular membrane biology, 28(5), 265-275.

24. Krishnamurthy, H., Piscitelli, C. L., & Gouaux, E. (2009). Unlocking the molecular secrets of sodium-coupled transporters. Nature, 459(7245), 347-355.

25. Yamashita, A., Singh, S. K., Kawate, T., Jin, Y., & Gouaux, E. (2005). Crystal structure of a bacterial homologue of Na+/Cl--dependent neurotransmitter transporters. Nature, 437(7056), 215-223.

26. Faham, S., Watanabe, A., Besserer, G. M., Cascio, D., Specht, A., Hirayama, B. A., ... & Abramson, J. (2008). The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na+/sugar symport. Science, 321(5890), 810-814.

27. Du, J., Förster, B., Rourke, L., Howitt, S. M., & Price, G. D. (2014). Characterisation of cyanobacterial bicarbonate transporters in E. coli shows that SbtA homologs are functional in this heterologous expression system. PloS one, 9(12).

28. Zhang, P., Battchikova, N., Jansen, T., Appel, J., Ogawa, T., & Aro, E. M. (2004). Expression and functional roles of the two distinct NDH-1 complexes and the carbon acquisition complex NdhD3/NdhF3/CupA/Sll1735 in Synechocystis sp PCC 6803. The Plant Cell, 16(12), 3326-3340.

29. Conroy, M. J., Durand, A., Lupo, D., Li, X. D., Bullough, P. A., Winkler, F. K., & Merrick, M. (2007). The crystal structure of the Escherichia coli AmtB–GlnK complex reveals how GlnK regulates the ammonia channel. Proceedings of the National Academy of Sciences, 104(4), 1213-1218.

30. Villarino, M., Etxebeste, O., Mendizabal, G., Garzia, A., Ugalde, U., & Espeso, E. A. (2017). Boron tolerance in Aspergillus nidulans is sustained by the SltA pathway through the SLC-family transporters SbtA and SbtB. Genes, 8(7), 188.

31. Price, G. D., Woodger, F. J., Badger, M. R., Howitt, S. M., & Tucker, L. (2004). Identification of a SulP-type bicarbonate transporter in marine cyanobacteria. Proceedings of the National Academy of Sciences, 101(52), 18228-18233.

32. Saier, M. H. (2000). A functional-phylogenetic classification system for transmembrane solute transporters. Microbiol. Mol. Biol. Rev., 64(2), 354-411.

33. Mount, D. B., & Romero, M. F. (2004). The SLC26 gene family of multifunctional anion exchangers. Pflügers Archiv, 447(5), 710-721.

34. Ohana, E., Yang, D., Shcheynikov, N., & Muallem, S. (2009). Diverse transport modes by the solute carrier 26 family of anion transporters. The Journal of physiology, 587(10), 2179-2185.

35. Price, G. D., & Howitt, S. M. (2011). The cyanobacterial bicarbonate transporter BicA: its physiological role and the implications of structural similarities with human SLC26 transporters. Biochemistry and Cell Biology, 89(2), 178-188.

36. Alper, S. L., & Sharma, A. K. (2013). The SLC26 gene family of anion transporters and channels. Molecular aspects of medicine, 34(2-3), 494-515.

37. Cordat, E., & Reithmeier, R. A. (2014). Structure, function, and trafficking of SLC4 and SLC26 anion transporters. In Current topics in membranes (Vol. 73, pp. 1-67). Academic Press.

38. Wang, C., Sun, B., Zhang, X., Huang, X., Zhang, M., Guo, H., ... & Yu, F. (2019). Structural mechanism of the active bicarbonate transporter from cyanobacteria. Nature plants, 5(11), 1184-1193.

39. Walter, J. D., Sawicka, M., & Dutzler, R. (2019). Cryo-EM structures and functional characterization of murine Slc26a9 reveal mechanism of uncoupled chloride transport. Elife, 8.

40. Chi, X., Jin, X., Chen, Y., Lu, X., Tu, X., Li, X., ... & Pan, X. (2020). Structural insights into the gating mechanism of human SLC26A9 mediated by its C-terminal sequence. Cell discovery, 6(1), 1-10.

41. Pandey, A., Shin, K., Patterson, R. E., Liu, X. Q., & Rainey, J. K. (2016). Current strategies for protein production and purification enabling membrane protein structural biology. Biochemistry and Cell Biology, 94(6), 507-527.

42. Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. Journal of molecular biology, 305(3), 567-580.

43. Overington, J. P., Al-Lazikani, B., & Hopkins, A. L. (2006). How many drug targets are there?. Nature reviews Drug discovery, 5(12), 993-996.

44.   Stillwell, W. (2016). An introduction to biological membranes: composition, structure and function (pp. 89-110). Elsevier.

45.   Marsh, D., Horváth, L. I., Swamy, M. J., Mantripragada, S., & Kleinschmidt, J. H. (2002). Interaction of membrane-spanning proteins with peripheral and lipid-anchored membrane proteins: perspectives from protein-lipid interactions. Molecular membrane biology, 19(4), 247-255.

46.   Garber, E. A., & Margoliash, E. (1990). Interaction of cytochrome c with cytochrome c oxidase: an understanding of the high-to low-affinity transition. Biochimica et Biophysica Acta (BBA)-Bioenergetics, 1015(2), 279-287.

47.   Johnson, J. E., & Cornell, R. B. (1999). Amphitropic proteins: regulation by reversible membrane interactions. Molecular membrane biology, 16(3), 217-235.

48.   Newton, A. C. (1995). Protein kinase C: structure, function, and regulation. Journal of biological chemistry, 270(48), 28495-28498.

49.   Schenkman, J. B., & Jansson, I. (2003). The many roles of cytochrome b5. Pharmacology & therapeutics, 97(2), 139-152.

50.   Tomita, M., & Marchesi, V. T. (1975). Amino-acid sequence and oligosaccharide attachment sites of human erythrocyte glycophorin. Proceedings of the National Academy of Sciences, 72(8), 2964-2968.

51.   Lanyi, J. K. (2004). Bacteriorhodopsin. Annu. Rev. Physiol., 66, 665-688.

52.   Jackups Jr, R., Cheng, S., & Liang, J. (2006). Sequence motifs and antimotifs in β-barrel membrane proteins from a genome-wide analysis: the Ala-Tyr dichotomy and chaperone binding motifs. Journal of molecular biology, 363(2), 611-623.

53.   Aitken, A., Cohen, P., Santikarn, S., Williams, D. H., Calder, A. G., Smith, A., & Klee, C. B. (1982). Identification of the NH2-terminal blocking group of calcineurin B as myristic acid. FEBS letters, 150(2), 314-318.

54.   H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. (2000) The Protein Data Bank. Nucleic Acids Research, 28: 235-242

55.   Parker, J. L., & Newstead, S. (2016). Membrane protein crystallisation: current trends and future perspectives. In The Next Generation in Membrane Protein Structure Determination (pp. 61-72). Springer, Cham.

56. Newby, Z. E., D O'Connell III, J., Gruswitz, F., Hays, F. A., Harries, W. E., Harwood, I. M., ... & Stroud, R. M. (2009). A general protocol for the crystallization of membrane proteins for X-ray structural investigation. Nature protocols, 4(5), 619.

57. Rhodes, G. (2010). Crystallography made crystal clear: a guide for users of macromolecular models. Elsevier.

58. Deisenhofer, J., Epp, O., Miki, K., Huber, R., & Michel, H. (1985). Structure of the protein subunits in the photosynthetic reaction centre of Rhodopseudomonas viridis at 3Å resolution. Nature, 318(6047), 618-624.

59. Raman, P., Cherezov, V., & Caffrey, M. (2006). The membrane protein data bank. Cellular and Molecular Life Sciences, 63(1), 36.

60. Delmar, J. A., Bolla, J. R., Su, C. C., & Edward, W. Y. (2015). Crystallization of membrane proteins by vapor diffusion. In Methods in enzymology (Vol. 557, pp. 363-392). Academic Press.

61. Landau, E. M., & Rosenbusch, J. P. (1996). Lipidic cubic phases: a novel concept for the crystallization of membrane proteins. Proceedings of the National Academy of Sciences, 93(25), 14532-14535.

62. Cherezov, V., Yamashita, E., Liu, W., Zhalnina, M., Cramer, W. A., & Caffrey, M. (2006). In meso structure of the cobalamin transporter, BtuB, at 1.95 Å resolution. Journal of molecular biology, 364(4), 716-734.

63. Caffrey, M. (2015). A comprehensive review of the lipid cubic phase or in meso method for crystallizing membrane and soluble proteins and complexes. Acta Crystallographica Section F: Structural Biology Communications, 71(1), 3-18.

64. Bowler, M. W., Guijarro, M., Petitdemange, S., Baker, I., Svensson, O., Burghammer, M., ... & Leonard, G. A. (2010). Diffraction cartography: applying microbeams to macromolecular crystallography sample evaluation and data collection. Acta Crystallographica Section D: Biological Crystallography, 66(8), 855-864.

65. Dubochet, J., & McDowall, A. W. (1981). Vitrification of pure water for electron microscopy. Journal of Microscopy, 124(3), 3-4.

66. Thomson, G. P., & Reid, A. (1927). Diffraction of cathode rays by a thin film. Nature, 119(3007), 890-890.

67. Dubochet, J., & Knapek, E. (2018). Ups and downs in early electron cryo-microscopy. PLoS biology, 16(4), e2005550.

68. Cheng, Y. (2018). Single-particle cryo-EM - how did it get here and where will it go. Science, 361(6405), 876-880.

69. Fernandez-Leiro, R., & Scheres, S. H. (2016). Unravelling biological macromolecules with cryo-electron microscopy. Nature, 537(7620), 339-346.

70. Cheng, Y., Grigorieff, N., Penczek, P. A., & Walz, T. (2015). A primer to single-particle cryo-electron microscopy. Cell, 161(3), 438-449.

71. Vinothkumar, K. R. (2015). Membrane protein structures without crystals, by single particle electron cryomicroscopy. Current opinion in structural biology, 33, 103-114.

72. Murata, K., & Wolf, M. (2018). Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. Biochimica et Biophysica Acta (BBA)-General Subjects, 1862(2), 324-334.

73. Lyumkis, D. (2019). Challenges and opportunities in cryo-EM single-particle analysis. Journal of Biological Chemistry, 294(13), 5181-5197.

74. Zhang, L., Song, J., Newhouse, Y., Zhang, S., Weisgraber, K. H., & Ren, G. (2010). An optimized negative-staining protocol of electron microscopy for apoE4• POPC lipoprotein. Journal of lipid research, 51(5), 1228-1236.

75. Rubinstein, J. L. (2007). Structural analysis of membrane protein complexes by single particle electron microscopy. Methods, 41(4), 409-416.

76. Tribet, C., Audebert, R., & Popot, J. L. (1996). Amphipols: polymers that keep membrane proteins soluble in aqueous solutions. Proceedings of the National Academy of Sciences, 93(26), 15047-15050.

77. Le Bon, C., Marconnet, A., Masscheleyn, S., Popot, J. L., & Zoonens, M. (2018). Folding and stabilizing membrane proteins in amphipol A8-35. Methods, 147, 95-105.

78. Tonggu, L., & Wang, L. (2020). Cryo-EM sample preparation method for extremely low concentration liposomes. Ultramicroscopy, 208, 112849.

79. Almgren, M., Edwards, K., & Karlsson, G. (2000). Cryo transmission electron microscopy of liposomes and related structures. Colloids and Surfaces A: Physicochemical and Engineering Aspects, 174(1-2), 3-21.

80. Bayburt, T. H., Grinkova, Y. V., & Sligar, S. G. (2002). Self-assembly of discoidal phospholipid bilayer nanoparticles with membrane scaffold proteins. Nano Letters, 2(8), 853-856.

81. Autzen, H. E., Myasnikov, A. G., Campbell, M. G., Asarnow, D., Julius, D., & Cheng, Y. (2018). Structure of the human TRPM4 ion channel in a lipid nanodisc. Science, 359(6372), 228-232.

82. Carlson, M. L., Young, J. W., Zhao, Z., Fabre, L., Jun, D., Li, J., ... & Beatty, J. T. (2018). The Peptidisc, a simple method for stabilizing membrane proteins in detergent-free solution. Elife, 7, e34085.

83. Angiulli, G., Dhupar, H. S., Suzuki, H., Wason, I. S., Van Hoa, F. D., & Walz, T. (2020). New approach for membrane protein reconstitution into peptidiscs and basis for their adaptability to different proteins. Elife, 9, e53530.

84. Henderson, R. (1995). The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. Quarterly reviews of biophysics, 28(2), 171-193.

85. Kimura, Y., Vassylyev, D. G., Miyazawa, A., Kidera, A., Matsushima, M., Mitsuoka, K., ... & Fujiyoshi, Y. (1997). Surface of bacteriorhodopsin revealed by high-resolution electron crystallography. Nature, 389(6647), 206-211.

86. Bill, R. M., Henderson, P. J., Iwata, S., Kunji, E. R., Michel, H., Neutze, R., ... & Vogel, H. (2011). Overcoming barriers to membrane protein structure determination. Nature biotechnology, 29(4), 335-340.

87. Nannenga, B. L., Iadanza, M. G., Vollmar, B. S., & Gonen, T. (2013). Overview of electron crystallography of membrane proteins: crystallization and screening strategies using negative stain electron microscopy. Current protocols in protein science, 72(1), 17-15.

88. Martynowycz, M. W., & Gonen, T. (2018). From electron crystallography of 2D crystals to MicroED of 3D crystals. Current opinion in colloid & interface science, 34, 9-16.

89. Shi, D., Nannenga, B. L., Iadanza, M. G., & Gonen, T. (2013). Three-dimensional electron crystallography of protein microcrystals. Elife, 2, e01345.

90. Nannenga, B. L., & Gonen, T. (2014). Protein structure determination by MicroED. Current opinion in structural biology, 27, 24-31.

91. Nannenga, B. L., Shi, D., Leslie, A. G., & Gonen, T. (2014). High-resolution structure determination by continuous-rotation data collection in MicroED. Nature methods, 11(9), 927-930.

92. Nannenga, B. L., Shi, D., Hattne, J., Reyes, F. E., & Gonen, T. (2014). Structure of catalase determined by MicroED. Elife, 3, e03600.

93. Hattne, J., Shi, D., de la Cruz, M. J., Reyes, F. E., & Gonen, T. (2016). Modeling truncated pixel values of faint reflections in MicroED images. Journal of applied crystallography, 49(3), 1029-1034.

94. Rodriguez, J. A., Ivanova, M. I., Sawaya, M. R., Cascio, D., Reyes, F. E., Shi, D., ... & Jiang, L. (2015). Structure of the toxic core of α-synuclein from invisible crystals. Nature, 525(7570), 486-490.

95. Yonekura, K., Kato, K., Ogasawara, M., Tomita, M., & Toyoshima, C. (2015). Electron crystallography of ultrathin 3D protein crystals: atomic model with charges. Proceedings of the National Academy of Sciences, 112(11), 3368-3373.

96. de la Cruz, M. J., Hattne, J., Shi, D., Seidler, P., Rodriguez, J., Reyes, F. E., ... & Hinck, C. S. (2017). Atomic-resolution structures from fragmented protein crystals with the cryoEM method MicroED. Nature methods, 14(4), 399-402.

97. Liu, S., & Gonen, T. (2018). MicroED structure of the NaK ion channel reveals a Na+ partition process into the selectivity filter. Communications biology, 1(1), 1-6.

98. Nannenga, B. L., & Gonen, T. (2016). MicroED opens a new era for biological structure determination. Current opinion in structural biology, 40, 128-135.

99. Shi, D., Nannenga, B. L., de la Cruz, M. J., Liu, J., Sawtelle, S., Calero, G., ... & Gonen, T. (2016). The collection of MicroED data for macromolecular crystallography. Nature Protocols, 11(5), 895-904.

100. Caffrey, M. (2008). On the mechanism of membrane protein crystallization in lipidic mesophases. Crystal Growth and Design, 8(12), 4244-4254.

101. Caffrey, M. (2015). A comprehensive review of the lipid cubic phase or in meso method for crystallizing membrane and soluble proteins and complexes. Acta Crystallographica Section F: Structural Biology Communications, 71(1), 3-18.

102. Qiu, H., & Caffrey, M. (2000). The phase diagram of the monoolein/water system: metastability and equilibrium aspects. Biomaterials, 21(3), 223-234.

103. Cherezov, V., Clogston, J., Papiz, M. Z., & Caffrey, M. (2006). Room to move: crystallizing membrane proteins in swollen lipidic mesophases. Journal of molecular biology, 357(5), 1605-1618.

104. Caffrey, M., & Porter, C. (2010). Crystallizing membrane proteins for structure determination using lipidic mesophases. JoVE (Journal of Visualized Experiments), (45), e1712.

105. Caffrey, M., Li, D., & Dukkipati, A. (2012). Membrane protein structure determination using crystallography and lipidic mesophases: recent advances and successes. Biochemistry, 51(32), 6266-6288.

106. Misquitta, Y., Cherezov, V., Havas, F., Patterson, S., Mohan, J. M., Wells, A. J., ... & Caffrey, M. (2004). Rational design of lipid for membrane protein crystallization. Journal of structural biology, 148(2), 169-175.

107. Rasmussen, S. G., DeVree, B. T., Zou, Y., Kruse, A. C., Chung, K. Y., Kobilka, T. S., ... & Mathiesen, J. M. (2011). Crystal structure of the β 2 adrenergic receptor–Gs protein complex. Nature, 477(7366), 549-555.

108. Katona, G., Andreasson, U., Landau, E. M., Andreasson, L. E., & Neutze, R. (2003). Lipidic cubic phase crystal structure of the photosynthetic reaction centre from Rhodobacter sphaeroides at 2.35 Å resolution. Journal of molecular biology, 331(3), 681-692.

109. Ostermeier, C., & Michel, H. (1997). Crystallization of membrane proteins. Current opinion in structural biology, 7(5), 697-701.

110. Weierstall, U., James, D., Wang, C., White, T. A., Wang, D., Liu, W., ... & Fromme, R. (2014). Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. Nature communications, 5(1), 1-6.

111. Liu, W., Ishchenko, A., & Cherezov, V. (2014). Preparation of microcrystals in lipidic cubic phase for serial femtosecond crystallography. Nature protocols, 9(9), 2123.

112. Hattne, J., Martynowycz, M. W., Penczek, P. A., & Gonen, T. (2019). MicroED with the Falcon III direct electron detector. IUCrJ, 6(5), 921-926.

113. Martin-Garcia, J. M., Zhu, L., Mendez, D., Lee, M. Y., Chun, E., Li, C., ... & Henning, R. (2019). High-viscosity injector-based pink-beam serial crystallography of microcrystals at a synchrotron radiation source. IUCrJ, 6(3).

114. Scott, A. I. (2013). Interpretation of the Ultraviolet Spectra of Natural Products: International Series of Monographs on Organic Chemistry (Vol. 7). Elsevier.

115. Coates, J. (2006). Interpretation of infrared spectra, a practical approach. Encyclopedia of analytical chemistry: applications, theory and instrumentation.

116. De Hoffmann, E. (2000). Mass spectrometry. Kirk-Othmer Encyclopedia of Chemical Technology.

117. Günther, H. (2013). NMR spectroscopy: basic principles, concepts and applications in chemistry. John Wiley & Sons.

118. David, W. I. F., & Shankland, K. (2008). Structure determination from powder diffraction data. Acta Crystallographica Section A: Foundations of Crystallography, 64(1), 52-64.

119. Dunitz, J. D. (1979). X-ray Analysis and the Structure of Organic Molecules. Cornell University Press.

120. Kunde, T., & Schmidt, B. M. (2019). Microcrystal Electron Diffraction (MicroED) for Small-Molecule Structure Determination. Angewandte Chemie International Edition, 58(3), 666-668.

121. Jones, C. G., Martynowycz, M. W., Hattne, J., Fulton, T. J., Stoltz, B. M., Rodriguez, J. A., ... & Gonen, T. (2018). The CryoEM method MicroED as a powerful tool for small molecule structure determination. ACS central science, 4(11), 1587-1592.

122. Gruene, T., Wennmacher, J. T., Zaubitzer, C., Holstein, J. J., Heidler, J., Fecteau-Lefebvre, A., ... & Li, T. (2018). Rapid structure determination of microcrystalline molecular compounds using electron diffraction. Angewandte Chemie International Edition, 57(50), 16313-16317.

123. Danelius, E., Halaby, S., van der Donk, W. A., & Gonen, T. (2021). MicroED in natural product and small molecule research. Natural Product Reports, 38(3), 423-431.

124. Ting, C. P., Funk, M. A., Halaby, S. L., Zhang, Z., Gonen, T., & Van Der Donk, W. A. (2019). Use of a scaffold peptide in the biosynthesis of amino acid–derived natural products. Science, 365(6450), 280-284.

125. Dick, M., Sarai, N. S., Martynowycz, M. W., Gonen, T., & Arnold, F. H. (2019). Tailoring tryptophan synthase TrpB for selective quaternary carbon bond formation. Journal of the American Chemical Society, 141(50), 19817-19822.

126. Gleason, P. R., Nannenga, B. L., & Mills, J. H. (2020). Rapid structural analysis of a synthetic non-canonical amino acid by microcrystal electron diffraction. Frontiers in molecular biosciences, 7.

127. Kaczmarski, J. A., Hong, N. S., Mukherjee, B., Wey, L. T., Rourke, L., Forster, B., ... & Jackson, C. J. (2019). Structural basis for the allosteric regulation of the SbtA bicarbonate transporter by the PII-like protein, SbtB, from Cyanobium sp. PCC7001. Biochemistry, 58(50), 5030-5039.

128. Förster, B., Mukherjee, B., Rourke, L., Kaczmarski, J. A., Jackson, C. J., & Price, G. D. (2021). Regulatory adenylnucleotide-mediated binding of the PII-like protein SbtB to the cyanobacterial bicarbonate transporter SbtA is controlled by the cellular energy state. bioRxiv.

129. Selim, K. A., Haase, F., Hartmann, M. D., Hagemann, M., & Forchhammer, K. (2018). PII-like signaling protein SbtB links cAMP sensing with cyanobacterial inorganic carbon response. Proceedings of the National Academy of Sciences, 115(21), E4861-E4869.

130. Huergo, L. F., Chandra, G., & Merrick, M. (2013). P_{II} signal transduction proteins: nitrogen regulation and beyond. FEMS microbiology reviews, 37(2), 251-283.

131. Nannenga, B. L., & Baneyx, F. (2011). Reprogramming chaperone pathways to improve membrane protein expression in Escherichia coli. Protein Science, 20(8), 1411-1420.

132. Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature methods, 6(5), 343-345.

133. Datsenko, K. A., & Wanner, B. L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proceedings of the National Academy of Sciences, 97(12), 6640-6645.

134. Otwinowski, Z., & Minor, W. (1997). [20] Processing of X-ray diffraction data collected in oscillation mode. Methods in enzymology, 276, 307-326.

135. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R., & Leslie, A. G. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallographica Section D: Biological Crystallography, 67(4), 271-281.

136. Bunkóczi, G., Echols, N., McCoy, A. J., Oeffner, R. D., Adams, P. D., & Read, R. J. (2013). Phaser. MRage: automated molecular replacement. Acta Crystallographica Section D: Biological Crystallography, 69(11), 2276-2286.

137. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., ... & Adams, P. D. (2012). Towards automated crystallographic structure refinement with phenix. refine. Acta Crystallographica Section D: Biological Crystallography, 68(4), 352-367.

138. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., ... & McCoy, A. J. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D: Biological Crystallography, 66(2), 213-221.

139. Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallographica Section D: Biological Crystallography, 60(12), 2126-2132.

140. Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., ... & Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallographica Section D: Biological Crystallography, 66(1), 12-21.

141. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. Journal of computational chemistry, 25(13), 1605-1612.

142. Conroy, M. J., Durand, A., Lupo, D., Li, X. D., Bullough, P. A., Winkler, F. K., & Merrick, M. (2007). The crystal structure of the Escherichia coli AmtB–GlnK complex reveals how GlnK regulates the ammonia channel. Proceedings of the National Academy of Sciences, 104(4), 1213-1218.

143. Studier, F. W. (2005). Protein production by auto-induction in high-density shaking cultures. Protein expression and purification, 41(1), 207-234.

144. Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nature methods, 9(7), 671-675.

145. Grant, T., Rohou, A., & Grigorieff, N. (2018). cisTEM, user-friendly software for single-particle image processing. Elife, 7, e35383.

146. Mastronarde, D. N. (2018). Advanced data acquisition from electron microscopes with SerialEM. Microscopy and Microanalysis, 24(S1), 864-865.

147. Punjani, A., Rubinstein, J. L., Fleet, D. J., & Brubaker, M. A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nature methods, 14(3), 290-296.

148.   Rath, P., Hilton, J. K., Sisco, N. J., & Van Horn, W. D. (2016). Implications of human transient receptor potential melastatin 8 (TRPM8) channel gating from menthol binding studies of the sensing domain. Biochemistry, 55(1), 114-124.

149.   Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G., & Deber, C. M. (2009). Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. Proceedings of the National Academy of Sciences, 106(6), 1760-1765.

150.   Rath, A., Cunningham, F., & Deber, C. M. (2013). Acrylamide concentration determines the direction and magnitude of helical membrane protein gel shifts. Proceedings of the National Academy of Sciences, 110(39), 15668-15673.

151.   Gupta, J. K., Rai, P., Jain, K. K., & Srivastava, S. (2020). Overexpression of bicarbonate transporters in the marine cyanobacterium Synechococcus sp. PCC 7002 increases growth rate and glycogen accumulation. Biotechnology for Biofuels, 13(1), 1-12.

152.   Kamennaya, N. A., Ahn, S., Park, H., Bartal, R., Sasaki, K. A., Holman, H. Y., & Jansson, C. (2015). Installing extra bicarbonate transporters in the cyanobacterium Synechocystis sp. PCC6803 enhances biomass production. Metabolic engineering, 29, 76-85.

153.   Price, G. D., Badger, M. R., & von Caemmerer, S. (2011). The prospect of using cyanobacterial bicarbonate transporters to improve leaf photosynthesis in C3 crop plants. Plant Physiology, 155(1), 20-26.

154.   Price, G. D., Pengelly, J. J., Forster, B., Du, J., Whitney, S. M., von Caemmerer, S., ... & Evans, J. R. (2013). The cyanobacterial CCM as a source of genes for improving photosynthetic CO2 fixation in crop species. Journal of experimental botany, 64(3), 753-768.

155.   Rolland, V., Badger, M. R., & Price, G. D. (2016). Redirecting the cyanobacterial bicarbonate transporters BicA and SbtA to the chloroplast envelope: soluble and membrane cargos need different chloroplast targeting signals in plants. Frontiers in Plant Science, 7, 185.

156.   Puertas, J. M., Nannenga, B. L., Dornfeld, K. T., Betton, J. M., & Baneyx, F. (2010). Enhancing the secretory yields of leech carboxypeptidase inhibitor in Escherichia coli: influence of trigger factor and signal recognition particle. Protein expression and purification, 74(1), 122-128.

157.   Pebay-Peyroula, E., Rummel, G., Rosenbusch, J. P., & Landau, E. M. (1997). X-ray structure of bacteriorhodopsin at 2.5 angstroms from microcrystals grown in lipidic cubic phases. Science, 277(5332), 1676-1681.

158. Lan, Z., Lee, M. Y., Chun, E., Liu, B., & Liu, W. (2019). Overview of Biochemical Assays in Lipidic Cubic Phase. Trends in biochemical sciences, 44(4), 295-299.

159. Xiang, J., Chun, E., Liu, C., Jing, L., Al-Sahouri, Z., Zhu, L., & Liu, W. (2016). Successful strategies to determine high-resolution structures of GPCRs. Trends in pharmacological sciences, 37(12), 1055-1069.

160. Johansson, L. C., Stauch, B., Ishchenko, A., & Cherezov, V. (2017). A bright future for serial femtosecond crystallography with XFELs. Trends in biochemical sciences, 42(9), 749-762.

161. Martin-Garcia, J. M., Conrad, C. E., Nelson, G., Stander, N., Zatsepin, N. A., Zook, J., ... & Hilgart, M. C. (2017). Serial millisecond crystallography of membrane and soluble protein microcrystals using synchrotron radiation. IUCrJ, 4(4), 439-454.

162. Caffrey, M., & Cherezov, V. (2009). Crystallizing membrane proteins using lipidic mesophases. Nature protocols, 4(5), 706.

163. Batyuk, A., Galli, L., Ishchenko, A., Han, G. W., Gati, C., Popov, P. A., ... & Cherezov, V. (2016). Native phasing of x-ray free-electron laser data for a G protein-coupled receptor. Science advances, 2(9), e1600292.

164. Liu, W., Chun, E., Thompson, A. A., Chubukov, P., Xu, F., Katritch, V., ... & Stevens, R. C. (2012). Structural basis for allosteric regulation of GPCRs by sodium ions. Science, 337(6091), 232-236.

165. Liu, W., Ishchenko, A., & Cherezov, V. (2014). Preparation of microcrystals in lipidic cubic phase for serial femtosecond crystallography. Nature protocols, 9(9), 2123-2134.

166. Evans, P. R., & Murshudov, G. N. (2013). How good are my data and what is the resolution?. Acta Crystallographica Section D: Biological Crystallography, 69(7), 1204-1214.

167. Wang, J., Dauter, M., & Dauter, Z. (2006). What can be done with a good crystal and an accurate beamline?. Acta Crystallographica Section D: Biological Crystallography, 62(12), 1475-1483.

168. Kabsch, W. (2010). Xds. Acta Crystallographica Section D: Biological Crystallography, 66(2), 125-132.

169. Craven, B. M. (1976). Crystal structure of cholesterol monohydrate. Nature, 260(5553), 727-729.

170. Nollert, P., Navarro, J., & Landau, E. M. (2002). Crystallization of membrane proteins in cubo. Methods in enzymology, 343, 183-199.

171. Nollert, P., & Landau, E. M. (1998). Enzymic release of crystals from lipidic cubic phases. Biochemical Society Transactions, 26, 709-714.

172. Wadsten, P., Wöhri, A. B., Snijder, A., Katona, G., Gardiner, A. T., Cogdell, R. J., ... & Engström, S. (2006). Lipidic sponge phase crystallization of membrane proteins. Journal of molecular biology, 364(1), 44-53.

173. Hattne, J., Reyes, F. E., Nannenga, B. L., Shi, D., De La Cruz, M. J., Leslie, A. G., & Gonen, T. (2015). MicroED data collection and processing. Acta Crystallographica Section A: Foundations and Advances, 71(4), 353-360.

174. Hattne, J., Shi, D., Glynn, C., Zee, C. T., Gallagher-Jones, M., Martynowycz, M. W., ... & Gonen, T. (2018). Analysis of global and site-specific radiation damage in cryo-EM. Structure, 26(5), 759-766.

175. Varsano, N., Beghi, F., Elad, N., Pereiro, E., Dadosh, T., Pinkas, I., ... & Addadi, L. (2018). Two polymorphic cholesterol monohydrate crystal structures form in macrophage culture models of atherosclerosis. Proceedings of the National Academy of Sciences, 115(30), 7662-7669.

176. Duyvesteyn, H. M., Kotecha, A., Ginn, H. M., Hecksel, C. W., Beale, E. V., de Haas, F., ... & Stuart, D. I. (2018). Machining protein microcrystals for structure determination by electron diffraction. Proceedings of the National Academy of Sciences, 115(38), 9569-9573.

177. Martynowycz, M. W., Zhao, W., Hattne, J., Jensen, G. J., & Gonen, T. (2019). Collection of continuous rotation MicroED data from ion beam-milled crystals of any size. Structure, 27(3), 545-548.

178. Martynowycz, M. W., Zhao, W., Hattne, J., Jensen, G. J., & Gonen, T. (2019). Qualitative analyses of polishing and precoating FIB milled crystals for MicroED. Structure, 27(10), 1594-1600.

179. Martin-Garcia, J. M., Zhu, L., Mendez, D., Lee, M. Y., Chun, E., Li, C., ... & Henning, R. (2019). High-viscosity injector-based pink-beam serial crystallography of microcrystals at a synchrotron radiation source. IUCrJ, 6(3).

180. Belrhali, H., Nollert, P., Royant, A., Menzel, C., Rosenbusch, J. P., Landau, E. M., & Pebay-Peyroula, E. (1999). Protein, lipid and water organization in

bacteriorhodopsin crystals: a molecular view of the purple membrane at 1.9 Å resolution. Structure, 7(8), 909-917.

181. Hanson, M. A., Cherezov, V., Griffith, M. T., Roth, C. B., Jaakola, V. P., Chien, E. Y., ... & Stevens, R. C. (2008). A specific cholesterol binding site is established by the 2.8 Å structure of the human β2-adrenergic receptor. Structure, 16(6), 897-905.

182. Yonekura, K., & Maki-Yonekura, S. (2016). Refinement of cryo-EM structures using scattering factors of charged atoms. Journal of Applied Crystallography, 49(5), 1517-1523.

183. Polovinkin, V., Khakurel, K., Babiak, M., Angelov, B., Schneider, B., Dohnalek, J., ... & Hajdu, J. (2020). Demonstration of electron diffraction from membrane protein crystals grown in a lipidic mesophase after lamella preparation by focused ion beam milling at cryogenic temperatures. Journal of applied crystallography, 53(6).

184. Martynowycz, M. W., Shiriaeva, A., Ge, X., Hattne, J., Nannenga, B. L., Cherezov, V., & Gonen, T. (2020). MicroED structure of the human adenosine receptor determined from a single nanocrystal in LCP. bioRxiv.

185. Nannenga, B. L., & Gonen, T. (2019). The cryo-EM method microcrystal electron diffraction (MicroED). Nature methods, 16(5), 369-379.

186. Hou, J., Inganäs, O., Friend, R. H., & Gao, F. (2018). Organic solar cells based on non-fullerene acceptors. Nature materials, 17(2), 119-128.

187. Cheng, P., Li, G., Zhan, X., & Yang, Y. (2018). Next-generation organic photovoltaics based on non-fullerene acceptors. Nature Photonics, 12(3), 131-142.

188. Zhang, J., Tan, H. S., Guo, X., Facchetti, A., & Yan, H. (2018). Material insights and challenges for non-fullerene organic solar cells based on small molecular acceptors. Nature Energy, 3(9), 720-731.

189. Paterson, A. F., Singh, S., Fallon, K. J., Hodsden, T., Han, Y., Schroeder, B. C., ... & Anthopoulos, T. D. (2018). Recent progress in high-mobility organic transistors: a reality check. Advanced Materials, 30(36), 1801079.

190. Yang, F., Cheng, S., Zhang, X., Ren, X., Li, R., Dong, H., & Hu, W. (2018). 2D organic materials for optoelectronic applications. Advanced Materials, 30(2), 1702415.

191. Wang, C., Dong, H., Jiang, L., & Hu, W. (2018). Organic semiconductor crystals. Chemical Society Reviews, 47(2), 422-500.

192. Levine, A. M., Biswas, S., & Braunschweig, A. B. (2019). Photoactive organic material discovery with combinatorial supramolecular assembly. Nanoscale Advances, 1(10), 3858-3869.

193. Guo, X., Zhou, N., Lou, S. J., Smith, J., Tice, D. B., Hennek, J. W., ... & Marks, T. J. (2013). Polymer solar cells with enhanced fill factors. Nature Photonics, 7(10), 825-833.

194. Zhao, F., Wang, C., & Zhan, X. (2018). Morphology control in organic solar cells. Advanced Energy Materials, 8(28), 1703147.

195. Feng, J., Jiang, W., & Wang, Z. (2018). Synthesis and Application of Rylene Imide Dyes as Organic Semiconducting Materials. Chemistry–An Asian Journal, 13(1), 20-30.

196. Jiang, W., Li, Y., & Wang, Z. (2014). Tailor-made rylene arrays for high performance n-channel semiconductors. Accounts of chemical research, 47(10), 3135-3147.

197. Eaton, S. W., Shoer, L. E., Karlen, S. D., Dyar, S. M., Margulies, E. A., Veldkamp, B. S., ... & Wasielewski, M. R. (2013). Singlet exciton fission in polycrystalline thin films of a slip-stacked perylenediimide. Journal of the American Chemical Society, 135(39), 14701-14712.

198. Zhan, X., Facchetti, A., Barlow, S., Marks, T. J., Ratner, M. A., Wasielewski, M. R., & Marder, S. R. (2011). Rylene and related diimides for organic electronics. Advanced Materials, 23(2), 268-284.

199. Levine, A. M., Schierl, C., Basel, B. S., Ahmed, M., Camargo, B. A., Guldi, D. M., & Braunschweig, A. B. (2019). Singlet fission in combinatorial diketopyrrolopyrrole-rylene supramolecular films. The Journal of Physical Chemistry C, 123(3), 1587-1595.

200. Smieska, L. M., Li, Z., Ley, D., Braunschweig, A. B., & Marohn, J. A. (2016). Trap-clearing spectroscopy in perylene diimide derivatives. Chemistry of Materials, 28(3), 813-820.

201. Vadehra, G. S., Maloney, R. P., Garcia-Garibay, M. A., & Dunn, B. (2014). Naphthalene diimide based materials with adjustable redox potentials: Evaluation for organic lithium-ion batteries. Chemistry of Materials, 26(24), 7151-7157.

202. Li, W., Hendriks, K. H., Wienk, M. M., & Janssen, R. A. (2016). Diketopyrrolopyrrole polymers for organic solar cells. Accounts of chemical research, 49(1), 78-85.

203.   Nielsen, C. B., Turbiez, M., & McCulloch, I. (2013). Recent advances in the development of semiconducting DPP-containing polymers for transistor applications. Advanced Materials, 25(13), 1859-1880.

204.   Wu, Y., & Zhu, W. (2013). Organic sensitizers from D-π-A to D-A-π-A: effect of the internal electron-withdrawing units on molecular absorption, energy levels and photovoltaic performances. Chemical Society Reviews, 42(5), 2039-2058.

205.   Sheldrick, G. M. (2015). SHELXT - Integrated space-group and crystal-structure determination. Acta Crystallographica Section A: Foundations and Advances, 71(1), 3-8.

206.   Hübschle, C. B., Sheldrick, G. M., & Dittrich, B. (2011). ShelXle: a Qt graphical user interface for SHELXL. Journal of applied crystallography, 44(6), 1281-1284.

207.   Macrae, C. F., Edgington, P. R., McCabe, P., Pidcock, E., Shields, G. P., Taylor, R., ... & Streek, J. V. D. (2006). Mercury: visualization and analysis of crystal structures. Journal of applied crystallography, 39(3), 453-457.

208.   See, K. C., Landis, C., Sarjeant, A., & Katz, H. E. (2008). Easily synthesized naphthalene tetracarboxylic diimide semiconductors with high electron mobility in air. Chemistry of Materials, 20(11), 3609-3616.

209.   Bhosale, S. V., Jani, C. H., & Langford, S. J. (2008). Chemistry of naphthalene diimides. Chemical Society Reviews, 37(2), 331-342.

210.   Chang, J., Ye, Q., Huang, K. W., Zhang, J., Chen, Z. K., Wu, J., & Chi, C. (2012). Stepwise cyanation of naphthalene diimide for n-channel field-effect transistors. Organic letters, 14(12), 2964-2967.

211.   Smith, M. B., & Michl, J. (2013). Recent advances in singlet fission. Annual review of physical chemistry, 64, 361-386.

212.   Busby, E., Xia, J., Wu, Q., Low, J. Z., Song, R., Miller, J. R., ... & Sfeir, M. Y. (2015). A design strategy for intramolecular singlet fission mediated by charge-transfer states in donor-acceptor organic materials. Nature materials, 14(4), 426-433.

213.   Sanders, S. N., Pun, A. B., Parenti, K. R., Kumarasamy, E., Yablon, L. M., Sfeir, M. Y., & Campos, L. M. (2019). Understanding the bound triplet-pair state in singlet fission. Chem, 5(8), 1988-2005.

214.   Guzman, C. X., Calderon, R. M. K., Li, Z., Yamazaki, S., Peurifoy, S. R., Guo, C., ... & Scott, A. M. (2015). Extended charge carrier lifetimes in hierarchical

donor-acceptor supramolecular polymer films. The Journal of Physical Chemistry C, 119(34), 19584-19589.

215. Brázda, P., Palatinus, L., & Babor, M. (2019). Electron diffraction determines molecular absolute configuration in a pharmaceutical nanocrystal. Science, 364(6441), 667-669.

216. Zhang, C., Lively, R. P., Zhang, K., Johnson, J. R., Karvan, O., & Koros, W. J. (2012). Unexpected molecular sieving properties of zeolitic imidazolate framework-8. The journal of physical chemistry letters, 3(16), 2130-2134.

217. Zhao, Z., Li, Z., & Lin, Y. S. (2009). Adsorption and diffusion of carbon dioxide on metal-organic framework (MOF-5). Industrial & Engineering Chemistry Research, 48(22), 10015-10020.

218. Gandara, F., & Bennett, T. D. (2014). Crystallography of metal-organic frameworks. IUCrJ, 1(6), 563-570.

219. Zhu, Y., Ciston, J., Zheng, B., Miao, X., Czarnik, C., Pan, Y., ... & Pinnau, I. (2017). Unravelling surface and interfacial structures of a metal-organic framework by transmission electron microscopy. Nature materials, 16(5), 532-536.

220. Li, Y., Wang, K., Zhou, W., Li, Y., Vila, R., Huang, W., ... & Wang, H. (2019). Cryo-EM structures of atomic surfaces and host-guest chemistry in metal-organic frameworks. Matter, 1(2), 428-438.

221. Zhang, D., Zhu, Y., Liu, L., Ying, X., Hsiung, C. E., Sougrat, R., ... & Han, Y. (2018). Atomic-resolution transmission electron microscopy of electron beam-sensitive crystalline materials. Science, 359(6376), 675-679.

222. Li, X., Wang, J., Liu, X., Liu, L., Cha, D., Zheng, X., ... & Han, Y. (2019). Direct imaging of tunable crystal surface structures of MOF MIL-101 using high-resolution electron microscopy. Journal of the American Chemical Society, 141(30), 12021-12028.

223. Zhu, L., Zhang, D., Xue, M., Li, H., & Qiu, S. (2013). Direct observations of the MOF (UiO-66) structure by transmission electron microscopy. CrystEngComm, 15(45), 9356-9359.

224. Feyand, M., Mugnaioli, E., Vermoortele, F., Bueken, B., Dieterich, J. M., Reimer, T., ... & Stock, N. (2012). Automated Diffraction Tomography for the Structure Elucidation of Twinned, Sub-micrometer Crystals of a Highly Porous, Catalytically Active Bismuth Metal-Organic Framework. Angewandte Chemie, 124(41), 10519-10522.

225. Yuan, S., Qin, J. S., Xu, H. Q., Su, J., Rossi, D., Chen, Y., ... & Zhou, H. C. (2018). [Ti8Zr2O12 (COO) 16] cluster: An ideal inorganic building unit for photoactive metal-organic frameworks. ACS central science, 4(1), 105-111.

226. Denysenko, D., Grzywa, M., Tonigold, M., Streppel, B., Krkljus, I., Hirscher, M., ... & Volkmer, D. (2011). Elucidating Gating Effects for Hydrogen Sorption in MFU-4-Type Triazolate-Based Metal-Organic Frameworks Featuring Different Pore Sizes. Chemistry-A European Journal, 17(6), 1837-1848.

227. Gemmi, M., Mugnaioli, E., Gorelik, T. E., Kolb, U., Palatinus, L., Boullay, P., ... & Abrahams, J. P. (2019). 3D electron diffraction: the nanocrystallography revolution. ACS central science, 5(8), 1315-1329.

228. Palatinus, L., Brázda, P., Boullay, P., Perez, O., Klementová, M., Petit, S., ... & Mintova, S. (2017). Hydrogen positions in single nanocrystals revealed by electron diffraction. Science, 355(6321), 166-169.

229. Kida, K., Okita, M., Fujita, K., Tanaka, S., & Miyake, Y. (2013). Formation of high crystalline ZIF-8 in an aqueous solution. CrystEngComm, 15(9), 1794-1801.

230. Park, K. S., Ni, Z., Côté, A. P., Choi, J. Y., Huang, R., Uribe-Romo, F. J., ... & Yaghi, O. M. (2006). Exceptional chemical and thermal stability of zeolitic imidazolate frameworks. Proceedings of the National Academy of Sciences, 103(27), 10186-10191.

231. Liu, G., Jiang, Z., Cao, K., Nair, S., Cheng, X., Zhao, J., ... & Pan, F. (2017). Pervaporation performance comparison of hybrid membranes filled with two-dimensional ZIF-L nanosheets and zero-dimensional ZIF-8 nanoparticles. Journal of Membrane Science, 523, 185-196.

232. Liu, G., Xu, Y., Han, Y., Wu, J., Xu, J., Meng, H., & Zhang, X. (2017). Immobilization of lysozyme proteins on a hierarchical zeolitic imidazolate framework (ZIF-8). Dalton Transactions, 46(7), 2114-2121.

233. Morris, W., Stevens, C. J., Taylor, R. E., Dybowski, C., Yaghi, O. M., & Garcia-Garibay, M. A. (2012). NMR and X-ray study revealing the rigidity of zeolitic imidazolate frameworks. The Journal of Physical Chemistry C, 116(24), 13307-13312.

234. Kwon, H. T., Jeong, H. K., Lee, A. S., An, H. S., & Lee, J. S. (2015). Heteroepitaxially grown zeolitic imidazolate framework membranes with unprecedented propylene/propane separation performances. Journal of the American Chemical Society, 137(38), 12304-12311.

APPENDIX A

LIST OF PUBLICATIONS

LIST OF PUBLICATIONS DURING GRADUATE STUDY

* Authors contributed equally

1. Levine, A. M., He, G., Bu, G., Ramos, P., Wu, F., Soliman, A., ... & Braunschweig, A. B. (2021). Efficient Free Triplet Generation Follows Singlet Fission in Diketopyrrolopyrrole Polymorphs with Goldilocks Coupling. The Journal of Physical Chemistry C. in press.

2. Bu, G., & Nannenga, B. L. (2021). MicroED Sample Preparation and Data Collection For Protein Crystals. In CryoEM (pp. 287-297). Humana, New York, NY.

3. Bu, G., Parrish, S., Gleason, P. R., Nielsen, D. R., & Nannenga, B. L. (2020). Heterologous expression and purification of the bicarbonate transporter BicA from Synechocystis sp. PCC 6803. Protein Expression and Purification, 175, 105716.

4. Zhu, L., Bu, G., Jing, L., Shi, D., Lee, M. Y., Gonen, T., ... & Nannenga, B. L. (2020). Structure determination from lipidic cubic phase embedded microcrystals by MicroED. Structure, 28(10), 1149-1159.

5. Bu, G., Simmons, C. R., Nielsen, D. R., & Nannenga, B. L. (2020). Tetragonal crystal form of the cyanobacterial bicarbonate-transporter regulator SbtB from Synechocystis sp. PCC 6803. Acta Crystallographica Section F: Structural Biology Communications, 76(9).

6. *Banihashemi, F., *Bu, G., Thaker, A., Williams, D., Lin, J. Y., & Nannenga, B. L. (2020). Beam-sensitive metal-organic framework structure determination by microcrystal electron diffraction. Ultramicroscopy, 216, 113048.

7. Levine, A. M., Bu, G., Biswas, S., Tsai, E. H., Braunschweig, A. B., & Nannenga, B. L. (2020). Crystal structure and orientation of organic semiconductor thin films by microcrystal electron diffraction and grazing-incidence wide-angle X-ray scattering. Chemical Communications, 56(30), 4204-4207.

8. Nannenga, B. L., Bu, G., & Shi, D. (2018). The evolution and the advantages of MicroED. Frontiers in molecular biosciences, 5, 114.

APPENDIX B

MICROED DATA STATISTICS OF LCP-PROTEINASE K

MICROED DATA STATISTICS OF LCP-PROTEINASE K

Table B: Proteinase K data processing and refinement statistics.

	MPD-treated	Lipase-treated
Data collection		
Excitation voltage	200 kV	
Electron source	Field emission gun	
Wavelength	0.025079 Å	
Detector	TemCam-F416	
Rotation rate	4 s per frame	
Total dose per crystal	~4.0 e$^-$/Å^2	
Data processing		
Number of crystals	4	2
Space group	P4$_3$2$_1$2	P4$_3$2$_1$2
Cell dimensions		
a, b, c (Å)	67.4, 67.4, 106.5	67.6, 67.6, 106.8
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	17.4-2.0 (2.05-2.0)	16.6-2.0 (2.05-2.0)
R$_{merge}$ (%)	32.4 (53.2)	40.4 (70.5)
CC$_{1/2}$	0.987 (0.368)	0.900 (0.275)
I/σ$_I$	5.5 (3.4)	4.0 (2.6)

Completeness (%)	84.6 (63.3)	94.6 (94.8)
Multiplicity	7.7 (6.3)	5.2 (5.4)
Total reflections	111,081	85,421
Total unique reflections	14,491	16,351
Data refinement		
R_{work} (%)	21.7	24.4
R_{free} (%)	26.7	28.2
RMSD bonds (Å)	0.003	0.003
RMSD angles (°)	0.571	0.478
Ramachandran (%) (favored, allowed, outlier)	96.8, 2.8, 0.4	97.1, 2.9, 0

Values in parentheisis () for highest resolution shell

APPENDIX C

MICROED DATA STATISTICS OF ORGANIC SEMICONDUCTORS

Table C: Organic semiconductor data collection and refinement statistics.

	dPyr PDI	dCN NDI	dDPP
Data collection			
Excitation Voltage	300 kV	300 kV	300 kV
Wavelength (Å)	0.019687	0.019687	0.019687
Number of crystals	3	8	7
Data Processing			
Space group	Cc	P2$_1$/c	P2$_1$/n
Unit cell dimensions			
a, b, c (Å)	22.05, 10.76, 9.34	8.09, 6.39, 11.63	15.09, 19.55, 34.77
β (°)	101.287	104.711	94.627
Resolution (Å)	0.60	0.57	0.90
Total reflections	23,894	27,418	134,916
Unique reflections	4,661	2,512	14,926
R$_{obs}$ (%)	17.4 (41.2)	19.2 (43.8)	28.0 (129.0)
R$_{meas}$ (%)	19.0 (52.1)	19.9 (53.3)	29.7 (146.7)
I/σ$_I$	4.48 (1.14)	6.81 (0.98)	4.15 (0.66)
CC$_{1/2}$ (%)	99.4 (45.5)	97.7 (79.4)	97.7 (20.5)
Completeness (%)	84.7 (41.8)	71.5 (49.0)	97.5 (77.3)
Refinement			

Stoichiometric Formula	$C_{32}H_{24}N_4O_4$	$C_{16}H_4N_4O_4$	$C_{54}H_{70}N_8O_6S_2$
R_1	0.2355 (0.1991)	0.1690 (0.1376)	0.2908 (0.2350)
wR_2	0.5085	0.3919	0.5555
GooF	1.423	1.141	1.813

Values in parenthesis () stand for high resolution shell.

APPENDIX D

MICROED DATA STATISTICS OF ZIF-8

MICROED DATA STATISTICS OF ZIF-8

Table D: ZIF-8 data collection and refinement statistics.

	ZIF-8
Data collection	
Excitation Voltage	300 kV
Wavelength (Å)	0.019687
Number of crystals	1
Data Processing	
Space group	I-43m
Unit cell length a = b = c (Å)	16.880
Angles $\alpha = \beta = \gamma$ (°)	90.00
Resolution (Å)	8.45 - 0.87
Number of total reflections	7,466
Unique reflections	407
R_{obs} (%)	34.3 (159.0)
R_{meas} (%)	35.8 (163.3)
I/σ_I	4.37 (1.11)
$CC_{1/2}$ (%)	97.2 (57.0)
Completeness (%)	99.8 (100.0)
Structure Refinement	
R1	0.1779

wR2	0.3411
GooF	1.011

Values in parentheses represent numbers in the highest resolution shell.