Structure-Function Studies in Prokaryotic Photosynthesis and CO_2-Concentrating Mechanisms

By

Thomas Gerard Laughlin

A dissertation submitted in partial satisfaction of the

Requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Associate Professor David Savage, Chair

Assistant Adjunct Professor Karen Davies

Professor Eva Nogales

Professor Krishna Niyogi

Spring 2020

Abstract

Structure-Function Studies in Prokaryotic Photosynthesis and CO_2-Concentrating Mechanisms

by

Thomas Gerard Laughlin

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Associate Professor David Savage, Chair

Photosynthesis encompasses the set of reactions that convert light into stored chemical energy. Photosynthetic organisms are the foundation of life, as they catalyze the assimilation of inorganic carbon (Ci) into the biosphere and are the ultimate source of nutrition on Earth.

Cyanobacteria are an ancient group of oxygenic photosynthetic prokaryotes and remain important primary producers today. The chloroplasts of all land plants can trace their ancestry to cyanobacteria. In many ways, such as the light reactions of photosynthesis that capture solar energy, the protein machinery of cyanobacteria is a representative, simplified form of that found in chloroplasts. In other ways, the cyanobacterial machinery is more complex as in the dark reactions that ultimately store solar energy as chemical energy by fixing CO_2. To overcome inherent limitations in the principal CO_2-fixing enzyme rubisco, cyanobacteria possess an elaborate integrated physiological strategy to increase the local CO_2 concentration near rubisco and facilitate on-pathway catalysis. This physiological strategy, known as a CO_2-concentrating mechanism (CCM), relies on energy-coupled Ci-transporters and encapsulation of rubisco in proteinaceous organelles known as carboxysomes. While some species of land plants possess CCMs, none resemble nor are nearly as productive as those found in cyanobacteria. Therefore, cyanobacteria are an ideal platform to study the basic workings of photosynthetic machinery and potentially provide insight on engineering strategies that may improve photosynthetic productivity in land plants of societal interest.

Here, we investigated the structure and function of protein complexes from cyanobacteria and other prokaryotes to elucidate their roles in photosynthesis and CCMs. First, we determined the structure of the NAD(P)H dehydrogenase-like complex, a key component in the light reactions, and develop putative models for its catalytic mechanism. Next, we discovered a new type of Ci-transporter we named the DABs accumulate bicarbonate (DAB) complex and characterize the complex through a variety of genetic, biochemical, and biophysical approaches. Finally, we determined the structure of simplified carboxysome shells that provide insight into the architecturally principles and assembly of native shell components.

Preface

In the following chapters, I wrote almost exclusively using a first-person-plural point of view (i.e., we and our). The particular style was chosen out of convenience, as well as personal preference. While this is "my" thesis, it is a collection of works that required many people to complete, and this fact is evident by the authors list of Chapters 2-4 on these previously published works. Understandably, some readers may wonder how these works came to be and what my scientific contributions were to each, thus I have disclosed that information here.

The work that lead to the structure of the cyanobacterial NAD(P)H dehydrogenase-like (NDH) complex (Chapter 2), was originally conceived as a project that I could use as the base of my qualifying exam proposal. Through a series of fortunate events, the work was quite fruitful and evolved into the main project of my graduate career. For this chapter, I identified the ideal cyanobacterial strain, performed all purifications, performed the electron microscopy, built the coordinate models, interpreted the data, and drafted and revised the manuscript for publication. This was my first foray into membrane protein biochemistry and cryo-electron microscopy (cryo-EM), thus the majority of my time in graduate school was dedicated to it.

In the discovery and characterization of the Dabs accumulate bicarbonate (DAB) complex (Chapter 3), I contributed to the "characterization" aspect. The DABs are membrane proteins that share structural and functional characteristics with the cyanobacterial forms of NDH possessing vectorial carbonic anhydrase (vCA) activity. Therefore, I predominantly contributed to development of the biochemistry and, based on my familiarity with the NDH vCA literature, design and interpretation of experiments aimed to assess whether DABs displayed accepted hallmarks of a vCA (e.g., energy-coupled, pH-independence, Zn-containing active-site).

For the structures of synthetic beta-carboxysome shells (Chapter 4), I perform the single-particle analysis, and I contributed largely to the modelling, interpretation, and drafting and revising of the resulting manuscript.

As noted, Chapters 2-4 are previously published works adapted to the format of this thesis. The information in these chapters is presented in the context at their original times of writing. Some recent advances that occurred after the are discussed in the Conclusion (Chapter 5). In addition, some continued efforts, data, and possible future experiments on matters pertaining to Chapters 2 and 3 are listed at the end this thesis as appendices.

“Essentially, all models are wrong, but some are useful.”
— George E. P. Box

“The secret to being original is knowing who to imitate.”
— multiple attributions

Acknowledgements

At the end of graduate school, I am a far cry from my high school plans of working on human physiology and pharmaceutical development. Ten years ago, I would have never predicted that I would find immense satisfaction in learning how microbes use perhaps the strangest means possible to conduct processes foundational to all life. I am grateful to my parents, Tom and Teresa, for giving me the opportunities to make it this far in my education, and I thank them for their unconditional support and understanding in my professional pursuits. I am also thankful to my sister, Claire, for her support and contributions to my personal and educational development.

I am indebted to Dr. Emerich for providing my first research opportunity and for taking a genuine interest in my personal and scientific development even after I left his lab. I am thankful to Mario and Min for conversations that helped me determine whether graduate school was right for me. In addition, I thank the Stefan and members of the Sarafianos Lab for providing a breadth of experiences that helped me refine my scientific interests. In particular, I thank Yee for having the time and patience to train me in various techniques and experimental approaches, Dandan for helping build my confidence in lab, and Andrew for helping shape me into a more independent scientist. I also thank my other friends from Mizzou, specifically Shawna and Alessandro whose conversations are a cherished aspect of my undergraduate education.

I am also thankful to my many friends that I have made since starting graduate school. I thank Jay for being a great scientific colleague, fellow Illinoisan, and roommate. I thank Liz for having the patience to teach me about the history, theory, and practice of electron microscopy. In addition, I thank Daniel for teaching me how to keep microscopy and science fun, even when it was difficult. Moreover, I am grateful to Dave and all of the members of the Savage Lab for their support and camaraderie over the years. I am particularly grateful to my baymates and neighbors for our conversations and antics that have made graduate school a memorable experience. A special thanks to Luke for actively and passively helping me to further my education in more topics than I ever anticipated to know, Rob for always being willing to lend a hand in scientific and artistic pursuits, Jack for impromptu history lessons on ancient empires, and Eli for always knowing how to boost my spirits.

I thank my committee, Dave, Karen, Eva, and Krishna, and their respective lab members for their help and guidance in my professional development and completion of this thesis at Berkeley.

To Vickie for her love and support that has helped me get through the most stressful challenges and being there to celebrate the victories.

And, to many other who I have failed to mention specifically who have shaped my scientific career, “Thank you”.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 NDH - the complex I of oxygenic photosynthesis

1.1a Overview of the light reactions of oxygenic photosynthesis

The process of oxygenic photosynthesis converts light energy from the sun to the chemical energy that sustains the majority of life on Earth [1]. This process is divided into the 'light' reactions, which generate ATP and NADPH, and the 'dark' reactions, which consume ATP and NADPH to fix CO_2. The light reactions can be further subdivided into linear and cyclic electron flow. Linear electron flow (LEF) strips electrons from water to yield molecular oxygen and shuttle the electrons through a series of redox-active complexes that ultimately deposit the electrons on NADP+ to form NADPH and also establish a proton gradient. The proton gradient is coupled to ATP synthase to form ATP from ADP and inorganic phosphate (Figure 1.1a,b). The net result of LEF is formation of ATP and NADPH at an approximate ratio of 1.3 [2,3] However, cellular processes (e.g., dark reactions, translation, etc.) typically demand more ATP than NADPH. Cyclic electron flow (CEF) satisfies this need by providing a means of generating ATP independently of NADPH [4,5].

CEF was discovered over 50 years ago and is 'cyclic' in that electrons start and end at photosystem I (PSI) [6]. Electrons excited at PSI enter the plastoquinone pool (PQ) and transit the proton-pumping cytochrome b_6f complex before returning to PSI, thus yielding a proton-motive force (pmf) to drive ATP synthesis via ATP synthase (Figure 1.1c,d). A point of contention in the molecular understanding of CEF has been the redox enzyme and substrate responsible for mediating PQ reduction. While multiple CEF pathways have been proposed, with PGR5/PGRL- and NDH-mediated being the most prevalent, NDH-mediated CEF is the most widely-conserved and is found in cyanobacteria and land plants [7–9].

1.1b Discovery of NDH

Work over decades has revealed multiple possible CEF pathways, which mainly vary at the steps by which electrons from PSI are transferred to a soluble carrier and returned to the PQ pool in the membrane [7,9]. NDH was first discovered when genes encoding complex I homologous core subunits were observed in the sequenced chloroplast genomes of *Nicotiana tabacum* and *Marchantia polymorpha* [10–12]. These genes were later shown to be conserved in cyanobacteria [13]. This discovery was provocative, as it was not immediately clear why a protein complex analogous to the respiratory complex I of mitochondria would be present in the chloroplast. Regulons from diverse photosynthetic organisms were shown to encode homologs for 11 of the 14 essential core complex I subunits, but specifically lacked the three subunits responsible for the NAD(P)H dehydrogenase activity of the respiratory enzyme, raising the question of how electron transfer could be mediated (Figure 1.2) [14]. Furthermore, biochemical analysis of partially purified NDH from plants and cyanobacteria showed several unique, conserved oxygenic photosynthesis specific (OPS) subunits, NdhL-Q, S and V, which are not found in respiratory homologs [15–18].

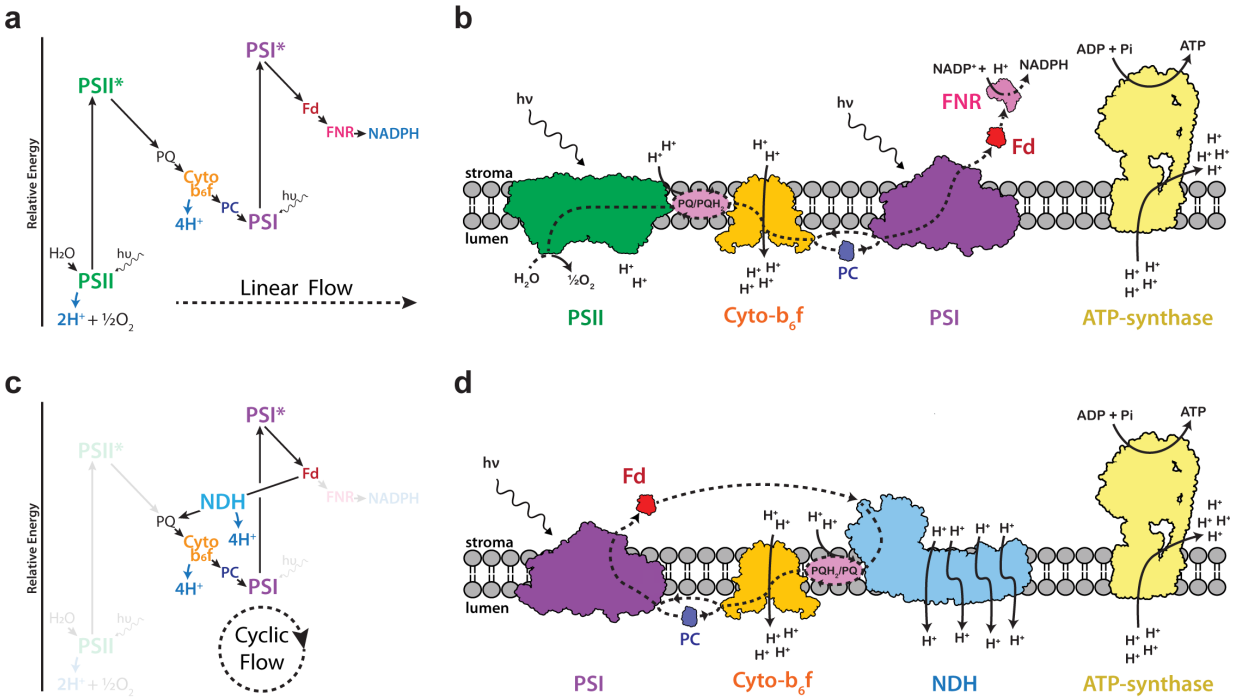


Figure 1.1 Overview of the light reactions of oxygenic photosynthesis. (**a**) Schematic ("Z-scheme") representation of linear electron flow denoting the electron transport chain complexes involved and points of light-energy capture and conversion to proton-translocation and reductant formation. (**b**) Cartoon depiction of the linear electron flow electron transport chain complexes as defined in **a** showing electron transfer, proton-translocation and ATP synthesis via ATP synthase. Dashed lines represent the path of excited electrons starting from water through various carriers to end up on $NADP^+$. (**c**) Schematic representation of cyclic electron flow denoting the electron transport chain complexes involved and points of light-energy capture and conversion to proton-translocation. (**d**) Cartoon depiction of the electron transport chain complexes as defined in **c** showing electron transfer, proton-translocation and ATP synthesis via ATP synthase. Dashed lines represent the path of excited electrons starting from photosystem I through various carriers to end up back on photosystem I.

1.1c Enzymatic function of NDH and its role in cyclic electron flow

Within the last decade, there has been an explosion in our collective understanding of the structure and function of NDH. With regards to function, a seminal results came from the Shikanai group, who demonstrated that plastid-type ferredoxin (Fd, a small iron-sulfur protein), and not NAD(P)H, is the redox donor to chloroplast NDH [19]. Moreover, mutational studies *in planta,* as well as *in vitro* PQ-reduction assays, implicated the OPS subunit NdhS as a critical player in CEF activity of NDH [19,20] . Fortuitously, the structure of isolated NdhS from a mesophilic cyanobacterium (*Synechocystis* sp. PCC6803) had been determined as part of the NIH Protein Structure Initiative. This structure revealed that NdhS adopted an SH3-like fold (PDB ID: 3C4S), a common domain in photosynthetic and Fd-binding proteins [21]. Guided by the available structure and conservation analysis, Yamomato *et al*. performed site-directed mutagenesis to further identify a positive-face on NdhS as important for function. Integrating these results, the authors proposed that NdhS formed part of a docking-site on the peripheral arm of NDH that complimented the negatively-charged Fd [20]. He *et al*. later showed that the cyanobacterial NdhS from *Thermosynechococcus elongatus* BP-1 (*T. elongatus*) also interacted with Fd suggesting that the binding-site is likely conserved between cyanobacteria to chloroplast forms of the complex [22].

More recently, Strand *et al*. used the combination of an *in vitro* luciferase-based assay and *in vivo* spectroscopy to elucidate proton-pumping aspects of the chloroplast NDH [9]. The authors treated osmotically ruptured chloroplasts as 'natural' proteoliposomes and measured ATP production in response to addition of reduced Fd via luciferase activity. The observation of Fd-dependent ATP production in their assay demonstrated for the first time that NDH was in fact a proton-pumping Fd-dependent oxidoreductase. In addition, through spectroscopic measurements of light-induced thylakoid proton flux *in vivo*, they estimated that NDH pumped two protons per electron transferred to PQ. This ~$2H^+$:e^- pumping ratio for NDH is consistent with that of the respiratory complex I. This work provided strong experimental support for the efficiency of ATP production via the NDH-mediated CEF pathway. Furthermore, it demonstrated the high-degree of functional conservation between the proton-pumping membrane arms for NDH and complex I.

These works from the Shikanai and Kramer groups highlighted key similarities and differences between the photosynthetic NDH and respiratory complex I. NDH is an energy-coupled proton-pump similar to complex I [23]. In contrast, the photosynthetic complex is not a NAD(P)H dehydrogenase type-1 like the respiratory complex I but actually a Fd:PQ oxidoreductase [9,19]. Hence, "NAD(P)H dehydrogenase complex" is a misnomer and is generally now referred to as NAD(P)H dehydrogenase"-like" complex or photosynthetic complex I. The homologous eleven core subunits shared between the photosynthetic and respiratory complex explained the similar proton-pumping. However, it was not clear how the distinctive OPS subunits enable NDH to interface with Fd as a redox donor was not completely clear. Furthermore, a critical aspect of the NDH redox mechanism is that Fd shuttles only one electron, but two donations events are required for complete reduction of PQ to PQH_2, which proceeds through a semi-PQ radical intermediate ($PQ^{\cdot}H$). This is in contrast to respiratory complex I that accepts two electrons from a single NADH to complete its redox mechanism (Figure 1.2). Therefore, the OPS subunits may form two distinct Fd-binding sites or enable rapid, successive donation events from Fd at a single site on NDH. Nevertheless, with the exception of NdhS, the structures and functions of the majority of the conserved OPS subunits remained unknown.

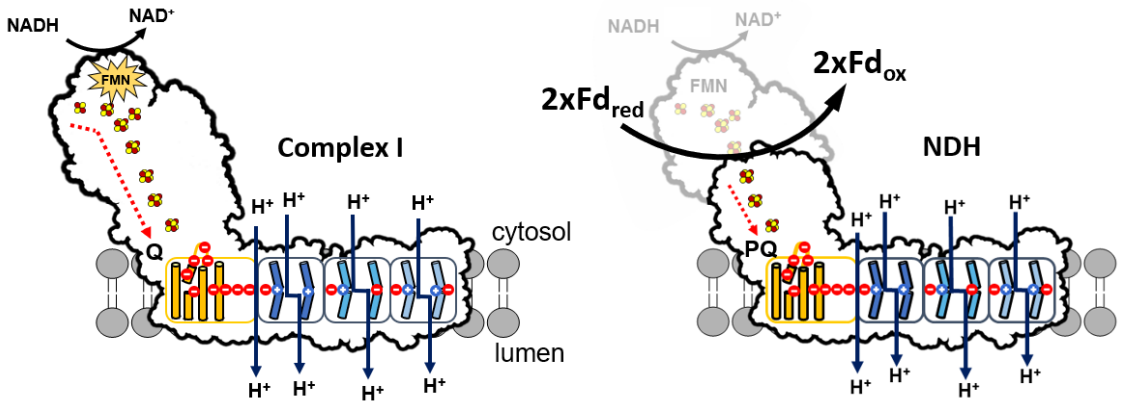


Figure 1.2 Comparison of complex I and NDH redox and proton-pumping activities. (left) Complex I accepts two electrons from NADH as a hydride which is converted by an FMN cofactor in the NADH dehydrogenase domain to single electron transfers through a series of FeS clusters to Q for reduction to QH_2. This redox activity is allosterically coupled to proton-pumping, proposed to occur through a series of conserved charged residues. (right) NDH lacks a NADH dehydrogenase domain and cognate FMN, thus takes single electron transfers directly from two Fds to complete its catalytic cycle to reduce PQ to PQH_2.

1.2 Cyanobacterial NDH and the CO_2-concentrating mechanism

1.2a Overview of the cyanobacterial CO_2-concentrating mechanism

Tangential to the role of NDH in the light reactions of photosynthesis, cyanobacterial forms of the complex has also been implicated in a specific aspect of the light-independent reactions known as the CO_2-concentrating mechanism (CCM) (Figure 1.3) [24–26]. Oxygenic photosynthetic organisms using the Calvin-Benson Bassham (CCB) Cycle are limited by the specificity and kinetics of the principal CO_2-fixing enzyme rubisco, which is both relatively slow and promiscuous [27–29]. A CCM is an integrated physiological strategy to increase the local CO_2 concentration near rubisco and facilitate on-pathway chemistry [1,30]. The cyanobacterial CCM consists of two primary steps: (i) active concentration of CO_2 in the cytosol in its hydrated, anionic form, HCO_3^- and (ii) localization of rubisco with carbonic anhydrase (CA) inside large proteinaceous structures called carboxysomes [31]. HCO_3^- is far less membrane-permeable than CO2, and so bicarbonate accumulation results in high concentrations (>10 mM) of inorganic carbon in the cytosol [32]. Enzymatic dehydration of HCO_3^- to CO_2 occurs only within the carboxysome, yielding a high local concentration of CO_2 near rubisco, thus enabling high-flux by rubisco towards CO_2-fixation and limiting the wasteful photorespiratory side-reaction with O_2 [31].

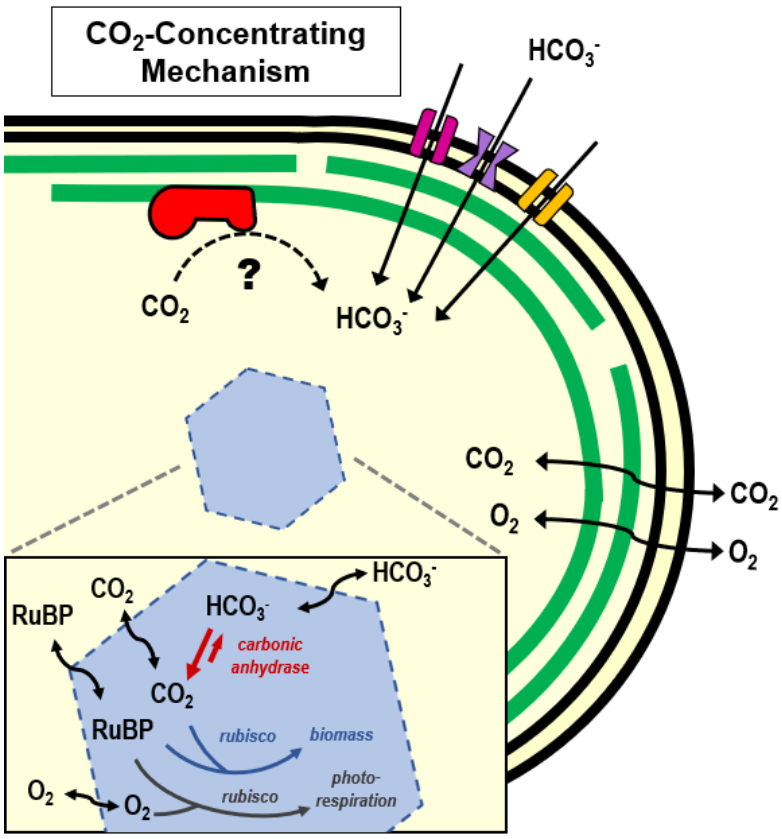


Figure 1.3 Overview of the prokaryotic CO_2-concentrating mechanism. Transporters (top right) at the cell membrane expend various forms of cellular energy (e.g., gradients, ATP, etc.) to import bicarbonate (HCO_3^-) against its concentration gradient into the cell. HCO_3^- accumulates well-above equilibrium concentrations in the cytosol due to its low membrane permeability, unlike CO_2 and O_2 which freely diffuse to reach equilibrium. HCO_3^- is then enzymatically equilibrated with CO_2 by carbonic anhydrase, resulting in a locally elevated concentration of CO_2 proximal to rubisco inside the carboxysome (a large icosahedral, proteinaceous shell). The CCM promotes on-pathway catalysis by rubisco with its substrates CO_2 and RuBP to increase organic carbon in the cell for biomass production and limits the wasteful side-reaction by rubisco with O_2 called photorespiration. NDH-$1_{3/4}$ (top left) are proposed unidirectionally hydrate CO_2 in cytosol by an unknown mechanism of energy-coupling.

1.2b Discovery and characterization of NDH-associated CO_2-uptake proteins

The cyanobacterial NDH is proposed to participate in the active concentration of CO_2 [25,33,34]. Foundational work establishing the cyanobacterial CCM 'parts list' relied on isolating and characterizing mutant strains with high-CO_2 requiring (HCR) phenotypes. HCR phenotypes typically indicate a mutation in a CCM component [35]. Early work by Ogawa established that defects in NDH genes lead to HCR phenotypes, implicating the complex in the cyanobacterial CCM [24,36]. Further work by Ogawa and colleagues using screens, reverse genetics and proteomics indicated that distinct forms of NDH were present in cyanobacteria [25,37–40] The distinct CCM-activate forms possessed paralogous versions of the distal membrane subunits (NdhD and NdhF) along with novel CO_2-uptake proteins (CUPs, a.k.a CO_2-hydration proteins, or Chps) [39,40]. Expression analysis and CO_2-exchange experiments demonstrated that the NdhD4/F4/CupB operon is a low-affinity, constitutively expressed module and the NdhD3/F3/CupA/CupS operon a high-affinity module induced in low-CO_2 conditions [25,37–40]. These forms of the complex were termed NDH-1_3 (containing NdhD3/F3/CupA) and NDH-1_4 (containing NdhD4/F4/CupB) [40]. Targeted disruption the CupA and CupB genes revealed that these novel ~50 kDa proteins, but not the ~16 kDa CupS, were essential to the observed CCM activity of their respective complexes [39,40].

Physical association of the CUP complexes (i.e., NdhD/F/Cup complex) to the NDH 'core' module was initially shown by co-purification and co-migration in BN/SDS-PAGE analysis However, the CUP complexes appeared to readily dissociate from the NDH core module (Figure 1.4a) [17,41]. Moreover, negative-stain TEM on solubilized cyanobacterial thylakoids showed NDH particles with additional density at the distal end of the membrane arm in some particles, giving them a defining U-shape [42,43]. Consistent with the induction of the CupA operon, the population of particles with the additional density was greater at low CO_2. Similar preparations from CupA and CupA/B mutants did not contain any particles with the additional density on the membrane arm [42]. Therefore, this density was assigned to CupA. The assignment would place CupA on top of NdhD3/F3 and CupB on top NdhD4/F4 (Figure 1.4b,c).

1.2c Proposed vectorial carbonic anhydrase model for CO_2-uptake proteins

Unlike other known energy-coupled HCO_3^- transporters in cyanobacteria such as SbtA and BicA, the CUPs have been proposed to be energy-coupled carbonic anhydrases [33,44]. Growth experiments at low pH and membrane-inlet mass-spectrometry of mutants suggest that the CUP complexes act to hydrate CO_2 into HCO_3^- , like carbonic anhydrase (CA) enzymes [25,33,38]. However, a CA in the cytosol of a cyanobacteria decouples the CCM by rapidly equilibrating HCO_3^- with CO_2, which can easily diffuse from the cell [45]. Moreover, neither CupA or CupB obviously resemble any known class of CA, nor possess an identifiable Zn-binding motif necessary for CA activity [46]. To explain the observed phenomena surrounding the CUPs, it has been proposed that the CUPs are energy-coupled, or vectorial, CAs (vCAs) that work exclusively in the direction of CO_2 hydration to HCO_3^-. Moreover, the proposed mode of energy-coupling is to that

of the redox and proton-pumping activity of NDH via a mechanism known as the "alkaline pocket" model (Figure 1.4b,c) [33,44].

The generic CA hydration mechanism involves formation of a Zn-hydroxide (Zn^{2+}-OH^{--}) to react with dissolved CO_2 by abstraction of the proton from an initial Zn-water (Zn^{2+}-H_2O) at the active-site by a residue or residues that act as a proton shuttle to solvent [46]. The alkaline pocket model suggests the vectorial proton-pumping activity of the NDH membrane arm creates a locally elevated pH (i.e., an alkaline environment) and serves to regenerate the proton shuttle and by extension the Zn^{2+}-OH^- in the CUP active-site. The energy-coupled regeneration of the putative Zn^{2+}-OH^- active site biases the CA-activity of the CUP towards the CO_2 hydration reaction rather than dehydration, rendering it vectorial [33,44]. This coupling presumably reduces the number of protons pumped per electron transferred via NDH in order to compensate for the uphill CO_2 hydration reaction. Nevertheless, until recently the molecular basis on how such a coupling could be achieved remained unclear. Unfortunately, the inability to express the CupA and CupB recombinantly in *E. coli* and the fragility of the NDH-$1_{3/4}$ upon detergent solubilization from native thylakoids have impeded rigorous assessment of this proposed mechanism [41–43].

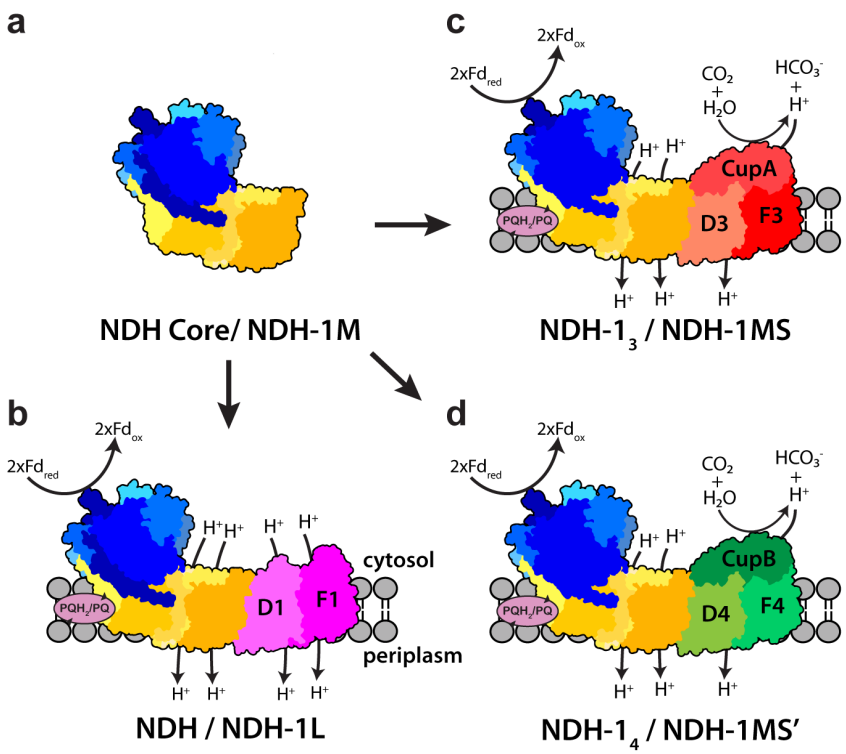


Figure 1.4 Speculative models of NDH-CUP. structures and mechanism. (**a**) Cartoon NDH 'core' module, a.k.a. NDH-1 "Medium" (NDH-1M) complex, containing core peripheral arm subunits (shades of blue) and membrane arm subunits (shades of yellow). (**b**), The addition of NdhD1/F1 module forms the NDH, a.k.a. NDH-1 "Large" (NDH-1L) complex. (**c**), addition of the NdhD3/F3/CupA (CupS not depicted) module, a.k.a. NDH-1 "Small" (NDH-1S) complex, forms the NDH-1_3/NDH-1MS complex. (**d**), addition of the NdhD4/F4/CupB module, a.k.a. NDH-1S' complex, forms the NDH-1_4/NDH-1MS' complex. NDH/NDH-1L accepts electrons from Fd to transfer them to PQ in the membrane and pumps four protons in a complete cycle. The NDH-$1_{3/4}$ are speculated to accept electrons from Fd to transfer to PQ, pump protons, and hydrate CO_2 vectorially. The CO_2 hydration activity is attributed to CupA/B and presumably comes at a reduction of proton pumping stoichiometry.

1.3 Objectives

The initial studies that investigated the structure of the cyanobacterial NDH and its function in the light and dark reactions of photosynthesis were conducted several decades ago. Moreover, as common for a multisubunit membrane protein complex, NDH is difficult to obtain in large quantities and purity-level necessary for high-resolution macromolecular crystallography. Therefore, we sought to take advantage of the relatively recent advances in cryo-electron microscopy that enable high-resolution structure determination with minimal quantities of protein. In addition, we sought to understand the enigmatic vCA activity of the NDH-CUP complexes that are notoriously difficult to study both *in vivo* and *in vitro*. To this end, we used a modern high-throughput to genetic approach to potentially discover more biochemically tractable complexes with similar vCA activity in a chemolithoautotroph. Finally, we sought to understand the architectural principles governing the assembly, stability and permeability of the carboxysome shell by studying a simplified form.

The following chapters describe this research that has led to the first high-resolution structure of NDH, the discovery of the new type of putative vCA known as DABs accumulate bicarbonate (DABs), and determine the structure of a simplified carboxysome shell. The structure of NDH reveals the arrangement and structures of the principal OPS subunits and suggests a mechanism for docking of Fd to the complex. The DABs are a simple two subunit protein complex that appears to possess quasi-convergent vCA activity to the NDH-CUP complexes. However, unlike the NDH-CUP complexes, DABs are biochemically tractable and fully reconstitute activity in recombinant hosts, thus serve as a suitable platform to interrogate the mechanism of a vCA. Lastly, we describe the structures of assembled, synthetic carboxysome shells that provide insight into carboxysomes and related proteinaceous organelles. Collectively, these works further our understanding of photosynthesis and CO_2-concentrating mechanisms.

Chapter 2

Structure of the cyanobacterial NAD(P)H dehydrogenase-like complex

†The work presented in this chapter is adapted from the previously published article:
Laughlin TG, Bayne AN, Trempe J-F, Savage DF, Davies KM. (2019) Structure of the complex I-like particle NDH of oxygenic photosynthesis *Nature* 566 (7744), 411-414.

2.1 Abstract

The fixed stoichiometry of ATP:NADPH generated during the light reactions of photosynthesis by linear electron flow creates an imbalance of energy currencies for assimilation of CO_2 by the Calvin-Benson-Bassham Cycle. Cyclic electron flow (CEF) about photosystem I (PSI) is a mechanism by which photosynthetic organisms can generate ATP, without the net production of reduced species, to thereby balance the photosynthetic budget. The NAD(P)H dehydrogenase-like complex (NDH) is a key component of this pathway in many oxygenic photosynthetic organisms and is the last large photosynthetic membrane protein complex of unknown structure. Related to the respiratory complex I, NDH transfers electrons originating from PSI to the plastoquinone (PQ) pool, while pumping protons across the thylakoid membrane. NDH possesses most of the core complex I subunits as well as several oxygenic photosynthesis specific (OPS) subunits, which are conserved from cyanobacteria to higher plants. However, the three core subunits of complex I involved in accepting electrons from NAD(P)H are notably absent in NDH. Thus, how NDH accepts electrons and transfers them to PQ is presently unclear. We determined two cryo-EM structures of the NDH complex from the thermophilic cyanobacterium *Thermosynechococcus elongatus* BP-1 at 3.1 and 3.5 Å, respectively. Both maps reveal the structure and arrangement of the principle OPS subunits and cofactors, as well as indicate possible modes of interaction with its electron donor, ferredoxin. Taken together, our findings raise interesting questions on how the mechanisms of substrate interaction and electron transfer differ between NDH of photosynthesis and the more extensively studied complex I of respiration.

2.2 Introduction

Cyclic electron flow (CEF) around photosystem I (PSI) is a mechanism by which photosynthetic organisms balance levels of ATP and NADPH necessary for efficient photosynthesis [5,47]. The NAD(P)H dehydrogenase-like complex (NDH) is a key component of this pathway in most oxygenic photosynthetic organisms and the last large photosynthetic membrane protein complex of unknown structure [9,48]. Related to the respiratory NADH dehydrogenase complex (complex I), NDH transfers electrons originating from PSI to the plastoquinone (PQ) pool, while pumping protons across the thylakoid membrane, thereby increasing the amount of ATP produced per $NADP^+$ molecule reduced [9,19]. NDH possesses 11 of the 14 core complex I subunits as well as several oxygenic photosynthesis specific (OPS) subunits, which are conserved from cyanobacteria to higher plants. However, the three complex I core subunits involved in accepting electrons from NAD(P)H are notably absent in NDH [8,48]. Thus, how NDH acquires and transfers electrons to PQ is presently unclear. Nevertheless, the OPS subunits, specifically NdhS, are proposed to enable NDH to accept electrons from ferredoxin (Fd), its electron donor [19,22].

Here we report a 3.1 Å structure of the ~0.42 MDa NDH complex from the thermophilic cyanobacterium *Thermosynechoccous elongatus* BP-1 (*T. elongatus*) obtained by single-particle cryo-electron microscopy (cryo-EM). Our maps reveal the structure and arrangement of the principle OPS subunits in the NDH complex, as well as an unexpected cofactor near the PQ-binding site in the peripheral arm. The location of the OPS subunits supports a role in electron transfer and defines two potential Fd-binding sites at the apex of the peripheral arm. These results

suggest multiple electron transfer routes could be present in NDH, which would serve to maximize PQ reduction and avoid deleterious off-target chemistry of the semi-plastoquinone radical.

2.3 Results

To investigate the role of the OPS subunits in electron transfer through NDH, we performed single-particle cryo-EM on the principle NDH complex of *T. elongatus* (i.e., NDH-1L/NDH-1$_1$) [8,48]. The *T. elongatus* complex represents the fundamental NDH architecture, which is conserved from cyanobacteria to chloroplasts [8,48–51]. NDH complex was detergent-solubilized from thylakoids and purified leveraging the native 'His-loop' distinct to the NdhF(1) subunit of *T. elongatus* NDH-1L/NDH-1$_1$ (see Materials and Methods, Figure 2.1). Mass-spectrometry confirmed the identity of the complex, detecting all 11 core subunits (NdhA-K) including NdhF(1), six of the seven conserved OPS subunits (NdhL-P, S) and cyanobacterial-specific subunit NdhQ were present (Tables 2.1, 2.2). Only conserved OPS subunit NdhV, which is weakly associated with NDH, was absent [52,53]. Through single-particle analysis of two independently purified samples, we obtained a 3.1 and a 3.5 Å resolution map of the overall complex (Figures 2.2-2.4). Both maps are highly consistent and showed clear side-chain density for the majority of residues, thus enabled atomic modeling (Figure 2.3, Table 2.3, specific details in Material and Methods). The distal region of the membrane arm is better resolved in the 3.5 Å map, whereas the remainder is better resolved in the 3.1 Å map (Figure 2.3c). Density was observed for all subunits except NdhV, which was also absent in the mass spectrometry data (Tables 2.1, 2.2).

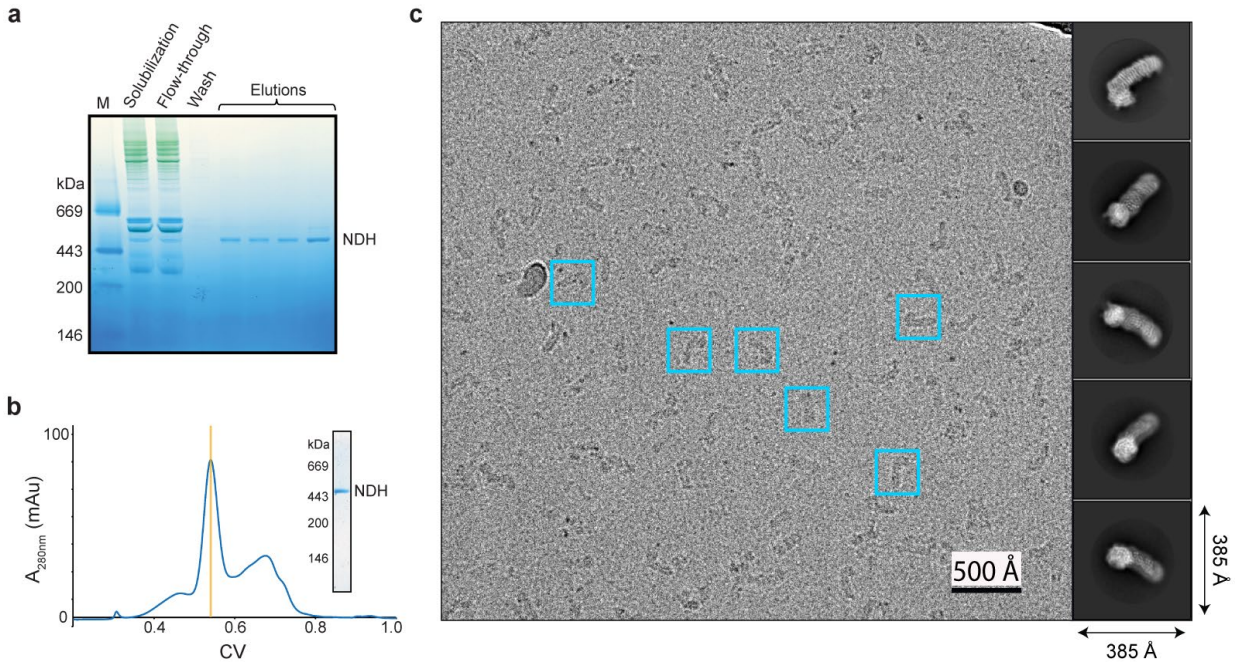


Figure 2.1 Purification and cryo-EM of *T. elongatus* NDH. (**a**) BN-PAGE gel of NDH solubilized by β-DDM and purified by Ni-affinity chromatography with fractions, positions of molecular-weight marker and NDH indicated. (**b**) Size-exclusion chromatography (SEC) trace of concentrated Ni^{2+}-affinity elutions with BN-PAGE gel of SEC fraction, indicated by gold-bar, showing a single band corresponding to NDH. (**c**) Micrograph of frozen-hydrated NDH used in this study with exemplar particles boxed and 2D-class averages generated from dataset-1 showing clear internal features.

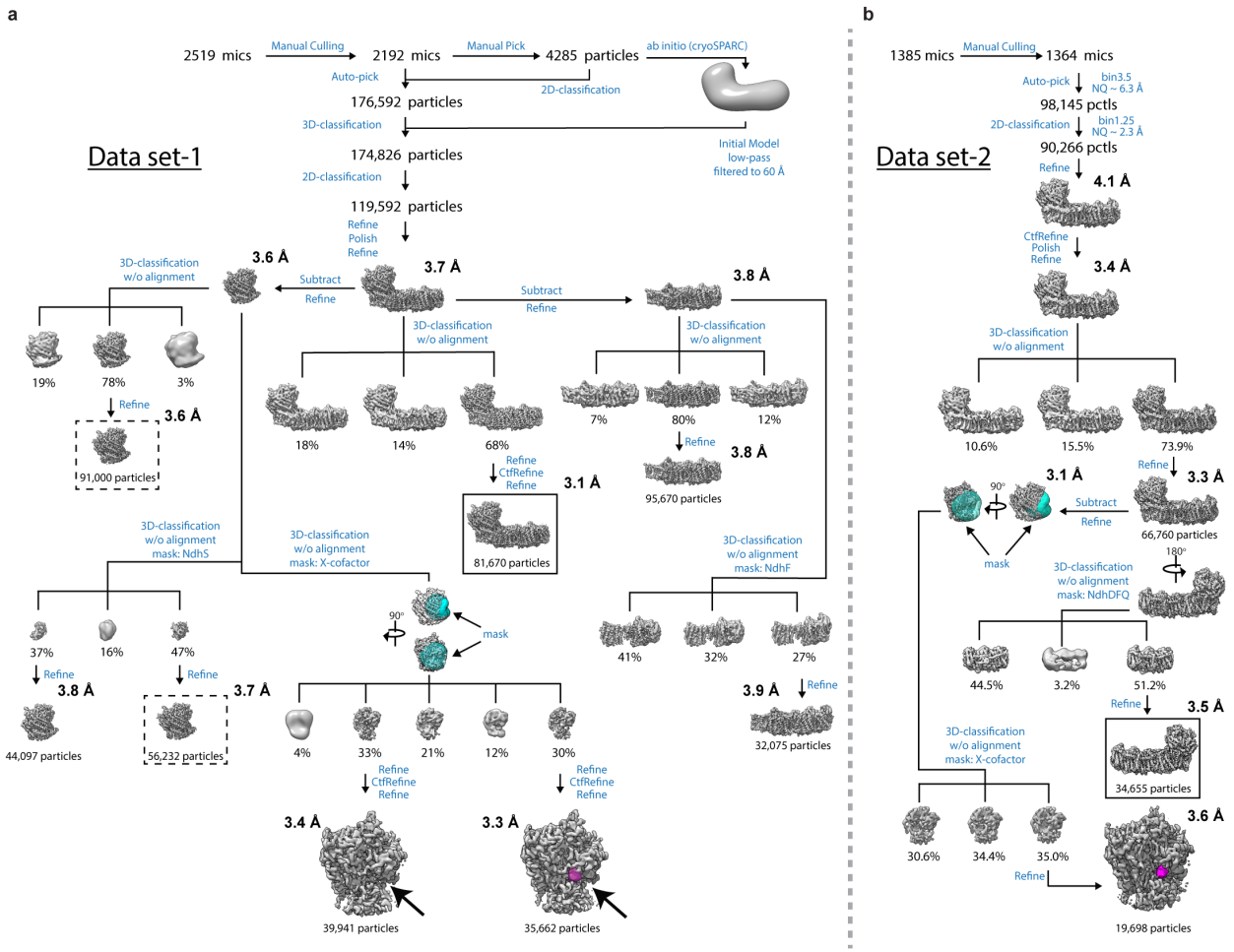


Figure 2.2 NDH cryo-EM data-processing workflow. Schematic of pre-processing, classification, and refinement procedures used to generate the maps obtained in this study (See Methods and Supplementary Methods for details) for dataset-1 (**a**) and dataset-2 (**b**). Boxed maps indicated those used in model-building and solid boxes for coordinate refinement. The masks used for focused-classification of the X-cofactor are indicated in cyan. The apo- and holo- X-cofactor focused-maps are enlarged to emphasize the X-site (colored light-purple).

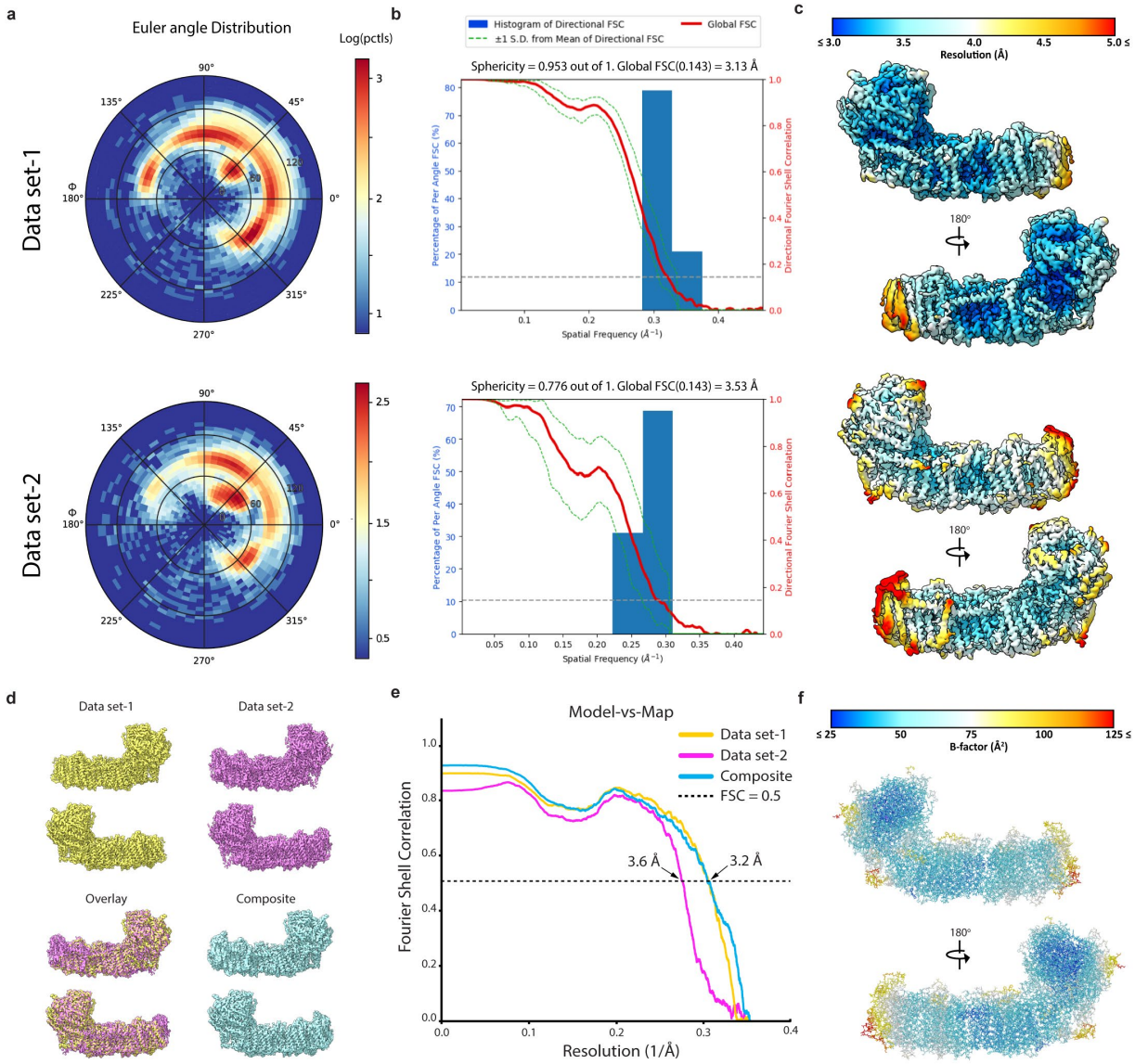


Figure 2.3 Resolution assessment of cryo-EM NDH maps and models. (**a-c**) Euler angle distributions, 3DFSC plots and local resolution (RELION, unsharpened) maps, respectively, for dataset-1 (top) and dataset-2 (bottom) overall cryo-EM maps. (**d**) Model versus map FSC for each dataset and composite map with their respective model calculated with PHENIX. (**e**) Sharpened maps used for coordinate refinement for each dataset, overlay and composite maps. (**f**) Representation of the refined coordinate model colored according to *B*-factor.

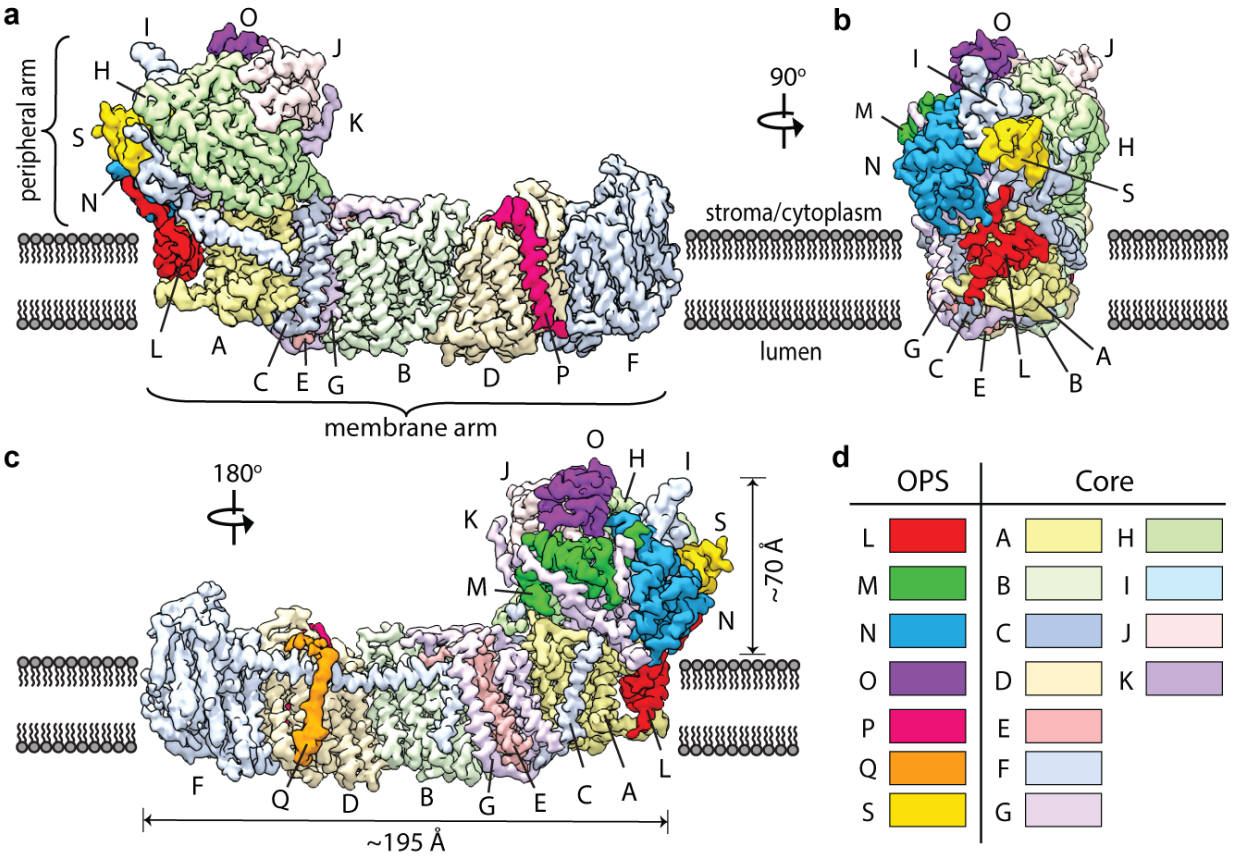
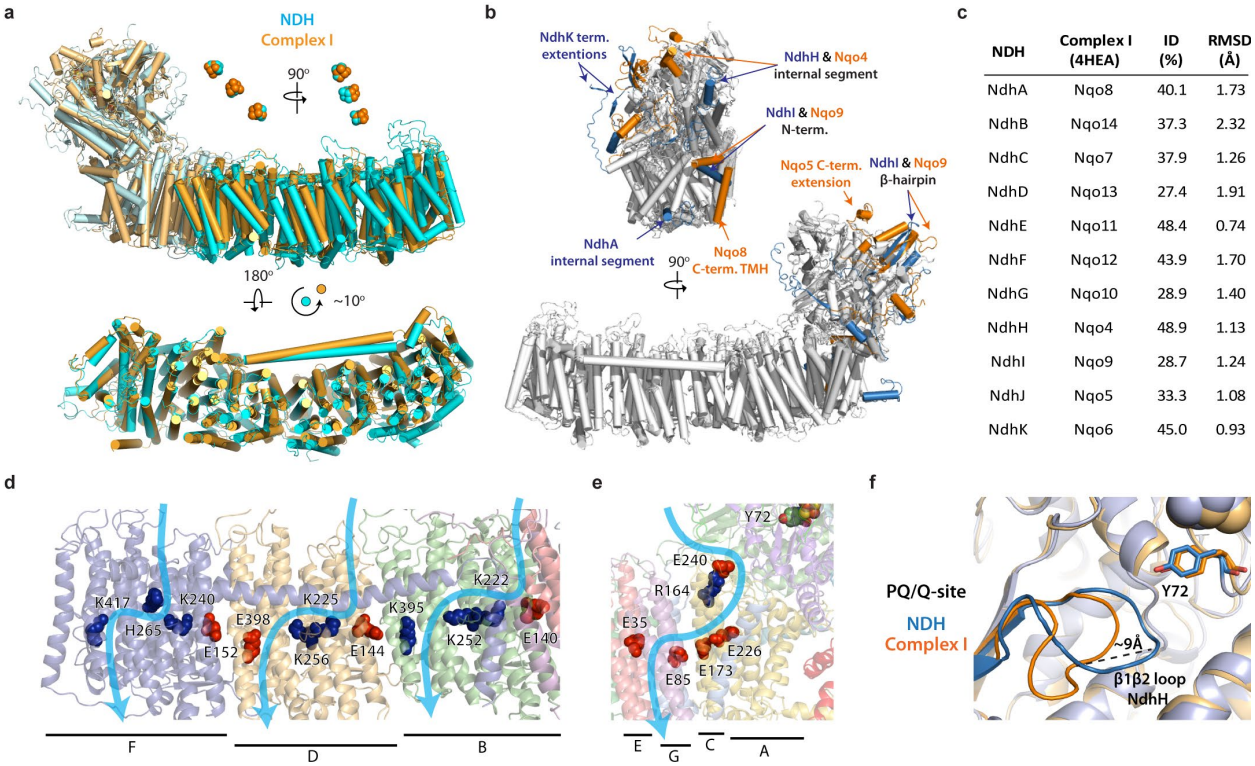


Figure 2.4 Cryo-EM density of the *T. elongatus* NAD(P)H dehydrogenase-like complex. (**a-c**) Different views of the overall cryo-EM reconstruction of the *T. elongatus* NDH complex segmented by subunit (threshold 0.025), coloured as indicated by key in **d**, and labeled by one-letter code. Relative orientations indicated and membrane is depicted schematically. The map was segmented 3 Å about the model and filtered to the $FSC_{0.143}$ resolution and left unsharpened ($B = 0$). Subunits NdhC/F/Q are from the dataset-2 map, all others from dataset-1. (**d**) Color-key defining the oxygenic photosynthesis specific (OPS) and homologous core (Core) subunits.

NDH adopts an L-shaped structure characteristic of respiratory NADH dehydrogenase complexes [49–51] (Figure 2.4). The complex possesses a ~195 Å long membrane arm and a peripheral arm that extends just ~70 Å above the membrane, rather than the ~135 Å observed in respiratory enzymes [9,49–51]. The shorter peripheral arm is consistent with the absence of the NADH dehydrogenase domain subunits [54,55]. The 11 core subunits of NDH are positioned similarly to their counterparts in the respiratory complex (Figure 2.5a-c). NdhB/D/F, homologous to the Mrp Na+/H+ antiporter-like subunits of complex I, are positioned at the distal end of the membrane arm. NdhA/C/E/G form the proximal membrane arm region, which link the antiporter-like subunits to the base of the peripheral arm. As with the respiratory enzyme, these subunits form an additional proton-channel and couple PQ reduction to the ion-pumping activities of the complex [9,49–51] (Figure 2.5d,e). The remaining four homologous subunits NdhH-K form the core of the peripheral arm. The structures of NdhH-K are similar to their respiratory homologs, with the notable exception of NdhK, which possesses N- and C-terminal extensions that interface with OPS subunits (Figure 2.5b, 2.6a).



NDH	Complex I (4HEA)	ID (%)	RMSD (Å)
NdhA	Nqo8	40.1	1.73
NdhB	Nqo14	37.3	2.32
NdhC	Nqo7	37.9	1.26
NdhD	Nqo13	27.4	1.91
NdhE	Nqo11	48.4	0.74
NdhF	Nqo12	43.9	1.70
NdhG	Nqo10	28.9	1.40
NdhH	Nqo4	48.9	1.13
NdhI	Nqo9	28.7	1.24
NdhJ	Nqo5	33.3	1.08
NdhK	Nqo6	45.0	0.93

Figure 2.5 Comparison of NDH and complex I homologous core. (**a**) Core subunits from *T. elongatus* (cyan) and *T.* thermophilus (orange, PDB: 4HEA) superimposed on NdhA/ND1 (the heel subunit) viewed from side (top) and top-down of the membrane arm (bottom). The antiporter domains are related by a ~10º rotation about their central axis. The FeS clusters are similarly positioned. (**b**) Core subunits and transverse helix plus aligned individually with observed differences highlighted in colored and labeled (white, conserved structure; blue, distinct to NDH; orange; distinct to complex I). **c**, Table of core subunit sequence homology and structural similarity (RMSD). (**d,e**) Depiction of putative proton-translocation pathways (blue arrows) based on conserved charged residues for the distal and PQ/Q-site adjacent sites, respectively. (**f**) Close-up view of the PQ/Q-site based on location of coordinating Tyr72 of NdhH (Tyr87 of Nqo4) reveals a difference in the of β1β2 loop of NdhH which is displaced ~9 Å relative to the complex I homolog Nqo4.

The six observed OPS subunits are distributed throughout NDH (Figure 2.4). NdhL, NdhP and NdhQ are located in the membrane arm. NdhL consists of a TMH followed by two short helices, rich in Phe and Leu residues, connected by a sharp Pro turn within the membrane and an extended C-terminal coil reaching from NdhA to NdhI (Figure 2.6b). NdhP and NdhQ are single TMHs adjacent to NdhD on opposite sides of the membrane arm. NdhP traps a lipid molecule at the NdhD/F interface (Figure 2.6c), and NdhQ appears to stabilize the horizontal helix of NdhF (Figure 2.6d). These subunits likely stabilize adjacent elements of the complex, consistent with previous work [55–58].

The remaining four OPS subunits (NdhM/N/O/S) are located on the peripheral arm (Fig. 1). NdhM and NdhN are α/β proteins located on the side of the peripheral arm and make a series of interactions with the terminal extensions of NdhK, such as inter-subunit antiparallel β-sheets (Figure 2.6a). NdhS and NdhO are positioned at the apex of the peripheral arm and on opposing sides (related by a ~170º rotation) of the entry-point FeS cluster in the core subunit NdhI (Figure 2.7a). Their position and distance to the cluster (~25 Å) suggest a possible role in facilitating redox chemistry.

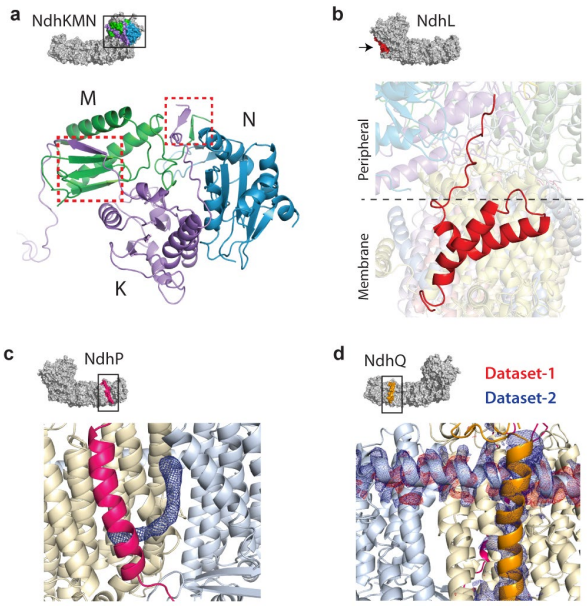


Figure 2.6 OPS subunits NdhL, M, N, P and Q. (**a-d**) Cartoon depictions of NdhL/M/N/P/Q with surface representations indicating location in complex. (**a**) NdhK (purple), M (green), N (blue) with complementary β-strands formed between NdhK and NdhM (dashed red boxes). (**b**) NdhL (red) connects the peripheral and membrane arms (**c**) NdhP (magenta) trapping a lipid (dark blue mesh) between NdhD (tan) and NdhF (light blue). (**d**) NdhQ (orange) with density for dataset-1 (red mesh) and dataset-2 (blue mesh) for NdhQ and the NdhF transverse helix, which is higher quality in the presence of NdhQ.

Fd is the demonstrated electron donor for PQ reduction by NDH [9,19,20,22]. Fd is a 2Fe-2S cluster protein of size ~10 x 25 Å and soluble electron-carrier that leverages protein-protein interactions for targeted redox chemistry. Many Fd-interacting proteins contain SH3-like domains (e.g., PsaE of PSI) [21,59], and it has been shown that NdhS both adopts an SH3-like fold (PDB ID: 3C4S, unpublished) and interacts with Fd [19,22,59]. Surprisingly, our data shows that NdhO also adopts an SH3-like fold, which is highly similar to NdhS (RMSD: 0.92 Å, Figure 2.7b), raising the question as to where, exactly, Fd binds.

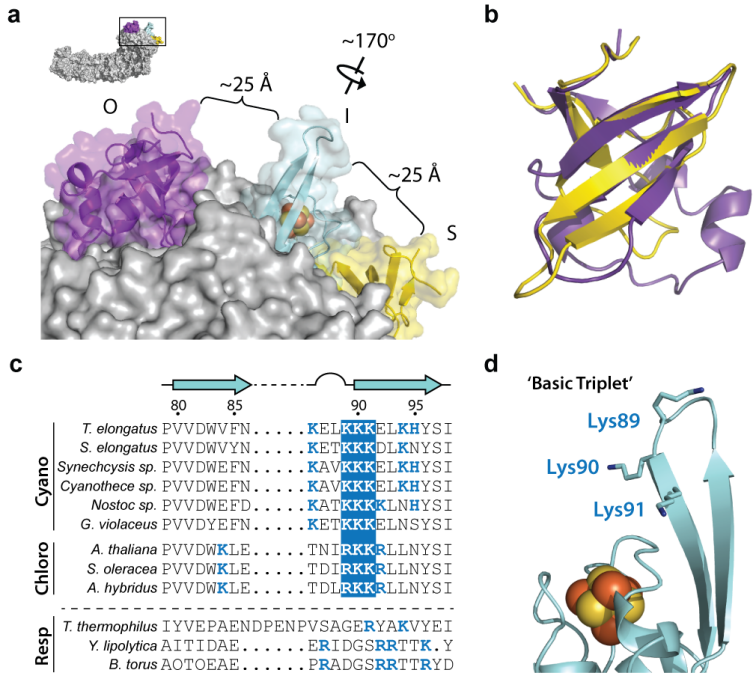


Figure 2.7 OPS subunits NdhO and NdhS and core subunit NdhI β-hairpin. (**a**) Surface model of boxed region in inset. Transparent surfaces of NdhI (blue), NdhO (purple), and NdhS (yellow) are shown along with cartoon atomic models. Distances shown are approximate from nearest edge of NdhO and NdhS to the NdhI β-hairpin. (**b**) Superposition of NdhO and NdhS with corresponding RMSD. Hash pattern represents surface facing respective sites. (**c**) Alignment of NdhI homologues from cyanobacterial (Cyano), chloroplastic (Chloro), and respiratory (Resp) complexes with basic residues in bold blue text and conserved 'basic triplet' (Lys89-91) as white text with blue background. *T. elongatus* secondary structure and numbering indicated above alignment. (**d**) NdhI β-hairpin cartoon model with basic triplet shown as sticks.

The surface of Fd is highly acidic (pI < 4) and its recruitment is often driven by basic patches on the binding partner [60–62]. An electrostatic surface calculated for the peripheral arm (Figure 2.8a) reveals two positively-charged putative binding sites encompassing analogous β-strand elements of either NdhS (S-site) or NdhO (O-site), respectively. Previous work in *Arabidopsis* implicated Arg69 (*T. elongatus* numbering) of NdhS as forming part of a Fd-binding site [19,20]. However, the S-site is a small, narrow cleft (~10 Å wide) and Arg69 is largely inaccessible from the surface (Figure 2.8b) Thus, if NdhS does bind Fd, an alternative surface or conformation would likely be involved.

The NdhO-adjacent site (O-site) is substantially wider (~30 Å) and appears more suitable for Fd-binding (Figure 2.8c). A notable feature of the O-site is the β-hairpin of NdhI that extends ~15 Å from the surface of the peripheral arm above the entry-point FeS cluster. This structural element contains a 'basic triplet' (Lys89-91), which is conserved across photosynthetic homologues but is absent in respiratory forms (Figure 2.7c,d). The basic triplet along with Lys4-5 of NdhO impart the O-site with significant positive charge that is characteristic of a Fd-binding site. Thus, it is possible that NDH has two Fd-binding sites.

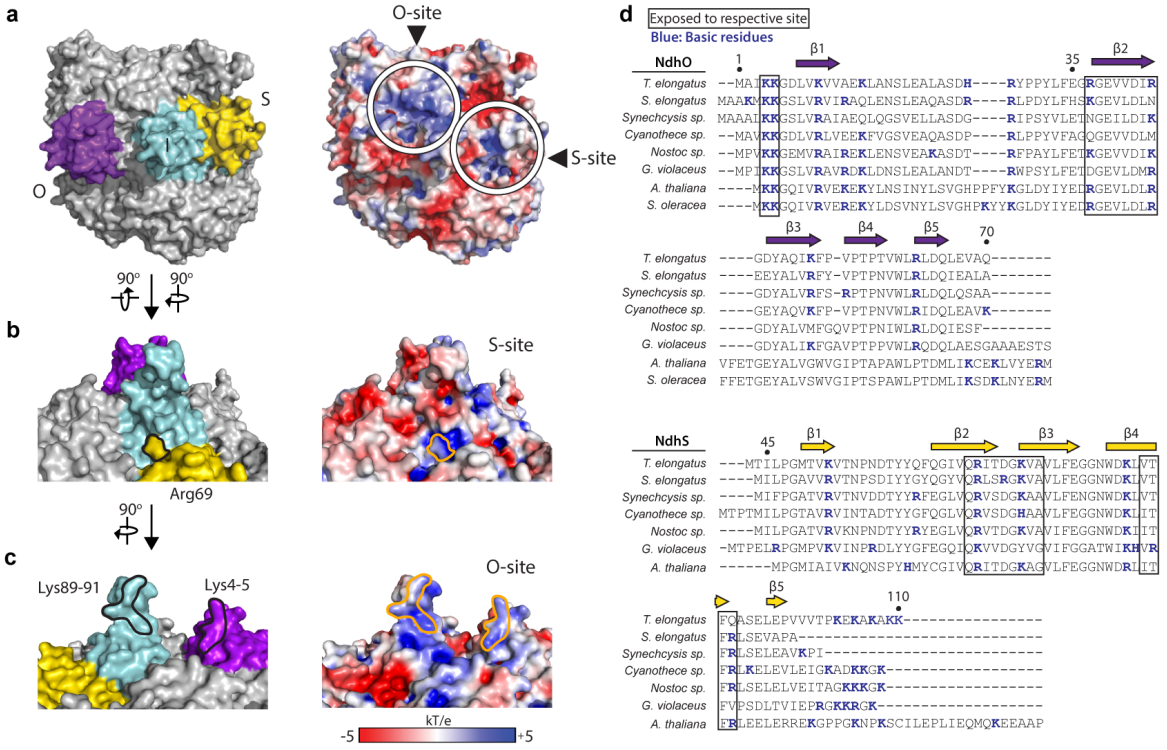


Figure 2.8 Electrostatics of the peripheral arm. (**a-c**) Left, surface models views of peripheral arm with NdhI β-hairpin, NdhO and NdhS colored in pale-cyan, purple and yellow, respectively. Right, corresponding electrostatic potential surfaces calculated at pH 8 with key at bottom. (**a**) Peripheral arm viewed down β-hairpin of NdhI. The electrostatic potential of this view shows two positive patches circled in white and labelled as NdhO-adjacent (O-site) and NdhS-adjacent (S-site). (**b**) S-site viewed as indicated by rotational relationship to **a** with Arg69 of NdhS outlined in orange. (**c**) O-site viewed as indicated by rotational relationship to **b** with the basic triplet (Lys89-91) of NdhI and NdhO Lys4-5 outlined in orange. (d) Alignments of NdhO (top) and NdhS (bottom) homologues with β-strand regions indicated and numbered according to *T. elongatus* sequence. Boxed regions indicated exposed regions facing respective sites and basic residues are colored blue. Predicted chloroplast localization sequences are removed when applicable. NdhS sequence for *S. oleracea* is unavailable. Residues 1-42 for *T. elongatus* NdhS are not conserved and not observed in our structure and have been removed for the purposes of alignment.

Electron transfer through respiratory complex I involves a series of metal-clusters in the peripheral arm [49–51,63]. Solid representation of the NDH maps reveals four strong densities in the peripheral arm (Figure 2.9a,b). Three of these densities (two in NdhI and one in NdhK) correlate with the final three 4Fe-4S clusters of the electron transfer pathway of the respiratory enzyme (Figure 2.5a) [49–51,63], which extends from the β-hairpin loop of NdhI to the predicted PQ-coordinating Tyr72 in NdhH (Figure 2.9c). Furthermore, each of these three densities are coordinated by four cysteine residues characteristic of 4Fe-4S clusters typical of traditional redox pathways in these enzymes.

The fourth density, to our knowledge, has never been observed in any homologous structure. This density, which we term the X-cofactor due to its unknown composition, is at the interface of core subunits NdhA/H/K and OPS subunit NdhM, which we term the 'X-site', and appears solvent-exposed (Figure 2.3, 2.9d). Although the X-cofactor is >25 Å from the PQ-proximal cluster of NdhK (Figure 2.9c), nearby aromatic residues Tyr46 of NdhH and Tyr48 of NdhK could connect the X-cofactor and terminal FeS cluster by acting as an electron transfer conduit, as observed in other redox complexes [64,65]. Focused-classification on the X-site in either dataset revealed only ~30% of particles contained the X-cofactor, but despite a labile nature, the peak is significant (Figure 2.3, 2.9). The local environment around the X-cofactor is composed mainly of hydrophilic residues, including a number of positively charged groups (Figure 2.9d). Despite the uncharacteristic local environment, the strong intensity for the X-cofactor is consistent with a metallocofactor and the location of the two Tyr residues suggest it could be involved in the transfer of electrons to PQ.

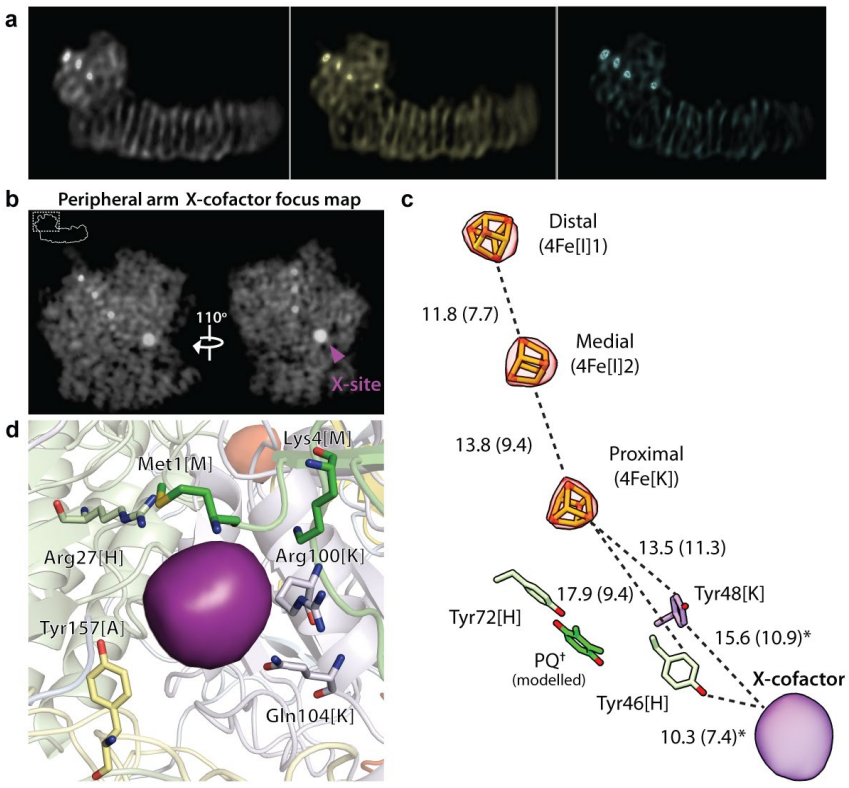


Figure 2.9 Arrangement of redox chain and cofactors within the peripheral arm. (**a**) Solid representation unmasked 3D-classification of the entire complex. (**b**) Solid representation of the X-cofactor focused peripheral arm map from dataset-1 showing strong peaks for the three homologous core FeS clusters and the 'X-cofactor' at the 'X-site' (purple arrow) Similar results for dataset-2 (Figure 2.2b). (**c**) Putative redox chain with transparent isosurface of X-cofactor focus peripheral arm map contoured to show only the density for the FeS clusters (red) and X-cofactor (purple). Center-to-center and edge-to-edge (in parenthesis) distances in Å between the clusters, possible bridging Tyr residues and X-cofactor. *Distances to X-cofactor based on pseudo-atom placed at the approximate center (and edge) of the isosurface as shown. †PQ headgroup positioned based on coordinating Tyr72 of NdhH. (**d**) Environment of the X-site with the X-cofactor density (purple surface, viewed from arrow in **b**). Residues within 8 Å of the X-cofactor center displayed as sticks.

2.4 Discussion

A critical aspect of the NDH redox mechanism is that Fd shuttles only one electron, but two donations events are required for complete reduction of PQ to PQH_2, which proceeds through a semi-PQ radical intermediate ($PQ^{\cdot}H$) (Figure 2.10). This is in contrast to complex I that accepts two electrons from a single NADH to complete its redox mechanism [63]. The PQ-site in our NDH structure is similar to the homologous Q-site of complex I, which has been proposed capable of stabilizing the semi-PQ radical intermediate [50,63]. Therefore, while the mechanism for radical intermediate stabilization may be conserved, our structure reveals how features of the OPS subunits distinct to NDH might further mitigate deleterious effects of a $PQ^{\cdot}H$ radical. These include: 1) NdhS and NdhO may define two binding sites on NDH allowing more rapid, successive donation events than a single-site mechanism, thus maximize complete reduction of PQ to PQH_2. 2) The X-cofactor may participate in redox activity and act as an electron buffer to abstract an electron from a long-lived $PQ^{\cdot}H$ and return the electron upon reformation of $PQ^{\cdot}H$ by a second Fd donation event, thus completing reduction to PQH_2, or 3) the X-cofactor may provide an alternative entry-point for electrons, thus allowing parallel reduction of PQ from two distinct electron donation pathways. The atomic model presented here therefore provides a framework for future functional studies to discern amongst these putative mechanisms.

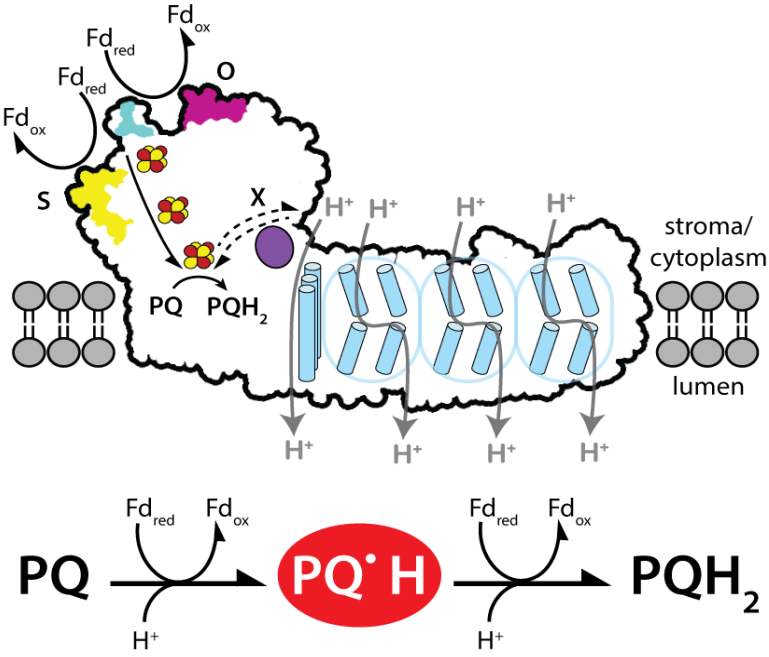


Figure 2.10 Summary of proposed models for NDH electron transfer mechanism. Top, schematic NDH complex with NdhO, NdhS, basic-triplet of NdhI, X-cofactor, and antiporter-like subunits coloured dark-purple, yellow, pale-cyan, light-purple and sky-blue, respectively. O-, S-, X-sites labeled by single letters. Proton-pumping in the membrane arm is coupled to reduction of the PQ to PQH_2 at the base of the peripheral arm. Single-barbed arrows indicate potential electron transfer routes (Solid for ferredoxin and dashed for X-cofactor routes). Bottom, schematic of PQ reduction catalyzed by NDH using two individual ferredoxin donation events, highlighting the radical intermediate ($PQ^{\cdot}H$).

2.5 Tables

Table 2.1 Summary of NDH subunits.

	Subunit name		NDH model summary				Chloroplast homolog	Complex I homologues			
	Common	Other(s)	*Te**	Length (aa)	Range built	Notes	*At*[†]	*Tt*[‡]	*Yl*[§]	*Bt*[‖]
Core	NdhA		tlr0667	372	28-362	PQ-site, X-site	AtCg01100	Nqo8	NU1M	ND1
	NdhB		tll0045	515	11-493	antiporter-like	AtCg00890[¶]	Nqo14	NU2M	ND2
	NdhC		tlr1429	132	18-43; 65-131		AtCg00440	Nqo7	NU3M	ND3
	NdhD	NdhD1	tll0719	529	2-505	antiporter-like	AtCg01050	Nqo13	NU4M	ND4
	NdhE		tlr0670	101	1-101		AtCg01070	Nqo11	NULM	ND4L
	NdhF	NdhF1	tll0720	656	1-655	antiporter-like, traverse helix, His-loop	AtCg01010	Nqo12	NU5M	ND5
	NdhG		tlr0669	200	5-148		AtCg01080	Nqo10	NU6M	ND6
	NdhH		tlr1288	394	3-385	PQ-site, X-site, O-site	AtCg01110	Nqo4	NUCM	49 kDa
	NdhI		tlr0668	196	6-189	2x 4Fe-4S, O-site, S-site	AtCg01090	Nqo9	NUIM	TYKY
	NdhJ		tlr1430	168	8-168		AtCg00420	Nqo5	NUGM	30 kDa
	NdhK		tlr0705	237	7-218	4Fe-4S, PQ-site, X-site	AtCg00430	Nqo6	NUKM	PSST
								Nqo1	NUBM	51 kDa
								Nqo2	NUHM	24 kDa
								Nqo3	NUAM	75 kDa
OPS	NdhL	CRR23	tsr0706	76	11-76	TMH and partial TMH	At1g70760			
	NdhM		tll0447	111	1-111	X-site	At4g37925			
	NdhN		tlr1130	150	2-149	S-site	At5g58260			
	NdhO		tsl0017	70	2-70	O-site, SH3-like	At1g74880			
	NdhP	PnsB4/NDF6	N/A	44	2-43	TMH, trapped lipid	At1g18730			
	NdhQ		N/A	45	2-44	TMH				
	NdhS	CRR31	tlr0636	110	43-99	S-site, SH3-like	At4g23890			
	NdhV		tlr0472	150	0	Not observed	At2g04039			

[†]*Thermosynechococcus elongatus*
[†]*Arabidopsis thaliana*
[‡]*Thermus thermophilus*
[§]*Yarrowia lipolytica*
[‖]*Bos torus*
[¶]Identical to AtCg01250

Table 2.2 Mass spectrometry data from the purified NDH.

				In-solution					In-gel		
				Peptides							
Subunit	Uniprot ID	Sequence length	MW [kDa]	CHYMO	TRYP	TOTAL	Sequence coverage [%]	MaxQuant Score	Peptides	Sequence coverage [%]	MaxQuant Score
NdhA	Q8DL32	372	40.55	14 (12)	8 (8)	22 (20)	44.1 (40.6)	260.88 (176.29)	36 (28)	63.7 (49.7)	323.31 (297.92)
NdhB	Q8DMR6	515	55.14	34 (24)	4 (3)	38 (27)	42.5 (34.2)	228.22 (114.06)	42 (32)	45.4 (38.6)	323.31 (323.31)
NdhC	Q8DJ02	132	15.00	4 (2)	0 (0)	4 (2)	28 (12.9)	41.331 (19.486)	6 (6)	28.8 (28.8)	56.185 (52.667)
NdhD/NdhD1	Q8DKY0	514	56.08	30 (22)	6 (6)	36 (28)	49 (40.1)	323.31 (277.07)	34 (29)	44.9 (37.6)	323.31 (323.31)
NdhE	Q8DL29	101	11.13	2 (2)	2 (2)	4 (4)	21.8 (21.8)	7.367 (5.2195)	9 (8)	66.3 (59.4)	282.59 (104.12)
NdhF/NdhF1	Q8DKX9	656	71.97	37 (21)	5 (5)	42 (26)	46.5 (32)	240.12 (114.94)	49 (34)	57.8 (45.6)	323.31 (323.31)
NdhG	Q8DL30	200	21.57	12 (8)	3 (3)	15 (11)	49.5 (40.5)	107.72 (76.66)	12 (10)	48 (33)	69.291 (85.808)
NdhH	Q8DJD9	394	45.22	38 (31)	23 (23)	61 (54)	90.6 (88.8)	323.31 (323.31)	72 (63)	95.4 (91.4)	323.31 (323.31)
NdhI	Q8DL31	196	22.42	18 (16)	13 (12)	31 (28)	92.3 (90.8)	323.31 (323.31)	26 (26)	81.1 (80.6)	323.31 (323.31)
NdhJ	Q8DJ01	168	19.34	19 (16)	17 (17)	36 (33)	99.4 (99.4)	323.31 (323.31)	21 (21)	81.5 (79.2)	323.31 (323.31)
NdhK	Q8DKZ4	237	25.74	20 (17)	11 (11)	31 (28)	84.4 (81.4)	323.31 (323.31)	38 (33)	94.1 (87.8)	323.31 (323.31)
NdhL	Q8DKZ3	76	8.57	0 (0)	0 (0)	0 (0)	0 (0)	N/A	7 (5)	48.7 (36.8)	184.86 (68.945)
NdhM	Q8DLN5	111	12.57	14 (12)	7 (7)	21 (19)	97.3 (95.5)	323.31 (323.31)	11 (11)	73.9 (73.9)	323.31 (219.89)
NdhN	Q8DJU2	150	16.64	14 (10)	9 (8)	23 (18)	88 (85.3)	323.31 (323.31)	23 (17)	87.3 (80.7)	323.31 (323.31)
NdhO	Q8DMU4	70	7.87	7 (7)	9 (8)	16 (15)	98.6 (98.6)	323.31 (323.31)	7 (7)	85.7 (85.7)	226.92 (66.33)
NdhP	N/A	44	4.87	0	1 (1)	1 (1)	36.4 (36.4)	23.472 (22.536)	4 (4)	75 (75)	92.05 (41.566)
NdhQ	N/A	45	4.71	1 (0)	0 (0)	1 (0)	20.5 (0)	5.554 (N/A)	5 (3)	75 (59.1)	143.55 (34.486)
NdhS	Q8DL61	110	12.45	8 (5)	4 (3)	12 (8)	69.1 (53.6)	323.31 (323.31)	6 (6)	60.9 (55.5)	82.066 (44.438)
NdhV	Q8DLL4	150	16.00	0 (0)	0 (0)	0 (0)	0 (0)	N/A	0 (0)	0	N/A

Values from focused (whole proteome) search

Table 2.3 NDH cryo-EM data collection, refinement and validation statistics.

	Dataset-1 (EMD-0415) (PDB 6NBQ)	Dataset-2 (EMD-0425) (PDB 6NBX)	Composite (PDB 6NBY)
Data collection			
Microscope		FEI Titan Krios	
Voltage (kV)		300	
Detector Model		Gatan K2 Summit	
EFTEM/GIF slit-width (eV)	N/A	25	-
Pixel size (Å)	1.068	0.903	-
Defocus range (μm)	-1.4 to -3.1	-0.8 to -2.7	-
Electron exposure (e–/Å^2)	~50	~60	-
Micrographs Collected	2519	1385	-
Reconstruction			
Software packages	RELION-2.1.0/3.0.b2	RELION-3.0.b2	-
Initial particle images (no.)	176,592	98,145	-
Final particle images (no.)	81,670	34,655	-
Symmetry imposed	C1	C1	-
Accuracy			
rotations (°)	0.900	0.821	-
translations (pixels)	0.402	0.397	-
Map resolution ($FSC_{0.143}$, Å)	3.13	3.53	-
Map resolution range (Å)	2.8-5.9	3.2-7.3	-
Sphericity	0.953	0.776	-
Calculated map *B* factor (Å^2)	-45	-45	-
Coordinate Refinement			
Initial model used (PDB code)[a]		4HEA, 3C4S	
Program package		PHENIX-1.13	
Map sharpening *B* factor (Å^2)	-30	-20	-
Model & Validation			
Model resolution ($FSC_{0.5}$, Å)	3.23	3.56	3.21
Map correlation (CC_{mask})	0.82	0.81	0.82
Model composition			
Non-hydrogen atoms	28295	29400	29400
Protein residues	3754	3853	3853
Ligands (Fe_4S_4)	3	3	3
B factors (Å^2)			
Protein	66	46	57
Ligand (Fe_4S_4)	42	48	41
R.m.s. deviations			
Bond lengths (Å)	0.006	0.007	0.005
Bond angles (°)	0.928	0.985	0.913
MolProbity score	1.81	1.79	1.78
Clashscore	6.08	5.33	5.85
Poor rotamers (%)	0.29	0.30	0.34
C_β deviations	0	0	0
EMRinger score	2.66	2.60	2.66
Ramachandran plot			
Favored (%)	92.22	91.45	92.63
Allowed (%)	7.75	8.52	7.34
Outliers (%)	0.03	0.03	0.03

[a]I-TASSER homology models for core subunits (4HEA) and NdhS (3C4S)

2.6 Materials and methods

Purification of NDH from *T. elongatus*

NDH (NDH-1L/NDH-1$_1$) was purified from *Thermosynechococcus elongatus* BP-1 (*T. elongatus*) by Ni^{2+}-affinity and size exclusion chromatography using the native histidine-rich loop within the most distal antiporter-like subunit NdhF1 of the NDH-1L/NDH-1$_1$ complex [66]. *T. elongatus* membranes were prepared as previously described [66,67] and frozen at -80ºC. All following steps were performed on ice or 4ºC unless otherwise specified. Before extraction, membranes were thawed and washed twice with buffer S (20 mM Bis-Tris, pH 6.0, 100 mM NaCl, 20% v/v glycerol, 1 mM PMSF, one cOmplete™, EDTA-free Protease Inhibitor Cocktail per 50 mL) by resuspension and centrifugation at 25,000g for 30 min. Washed membranes were then resuspended ~4 mg protein per mL in buffer S, mixed 1:1 with buffer S containing 2% (w/v) β-Dodecyl-maltoside (β-DDM, Anatrace), and gently agitated for 30 min. Insoluble material was removed by centrifugation at 165,000g for 45 min and the supernatant passed through a 0.45 um filter. Filtered, solubilized material was bound in batch to Ni-Sepharose FF resin (GE Healthcare) for 1 hr. The resin was then washed with buffer N (20 mM Bis-Tris, pH 6.0, 100 mM NaCl, 10% v/v glycerol, 0.03% β-DDM) in a gravity-flow column until no protein was detected by the Bradford method [68]. The bound protein was subsequently eluted with buffer N + 100 mM imidazole. Fractions were pooled and concentrated in a 100 kDa-cutoff spin-concentrator (Vivaspin) to ~10-12 mg/mL. Concentrated sample was then injected onto a Superose 6 10/300 GL column equilibrated in and ran isocratically with buffer G (20 mM Bis-Tris, pH 6.0, 100 mM NaCl, 0.03% β-DDM). NDH containing fractions were determined by Blue-Native PAGE [69] and negative-stain transmission electron microscopy.

Mass-spectrometry of purified NDH complex

NDH proteins (20 µg) from above were extracted using the methanol/chloroform precipitation method [70]. The pellet was resuspended in 10 µL of 8 M urea. pH adjusted by the addition of 5 µL triethyl ammonium carbonate (TEAB) 50 mM buffer at pH 8.5. Tris(2-carboxyethyl)phosphine (TCEP) was added to 10 mM and the mix incubated 10 min at 50°C to reduce disulfide bonds. Iodoacetamide was added to a final concentration of 50 mM and incubated at room temperature for 30 min in the dark. The sample was diluted to 80 µL with 50 mM TEAB pH 8.5 (final concentration 1 M urea). The sample was split in two and digested overnight at room temperature with 0.5 µg of either trypsin or chymotrypsin (Sigma, proteomics grade). The peptides were then purified using C18 spin columns (Thermo) using the manufacturer's protocol. The peptides were eluted in 2x40 µL 70% acetonitrile, dried in a speed-vac and resuspended in 24 µL of 4% acetonitrile/0.1% trifluoroacetic acid (TFA).

For the in-gel digested sample, BN-PAGE bands corresponding to the NDH complex were isolated and processed using a standard in-gel digestion procedure, as previously described [71]. Briefly, the gel pieces were excised, reduced with 10 mM DTT, alkylated with 55 mM iodoacetamide, destained, and digested overnight with chymotrypsin at 37°C. Chymotryptic peptides were extracted with acetonitrile, dried in a speed-vac, and resuspended in 20 µL of 4% acetonitrile/0.1% trifluoroacetic acid (TFA).

Purified peptides (2.5 µg) were captured on a C18 µ-precolumn and eluted onto an Acclaim PepMap100 C18 column (75 µm × 15 cm) with a 1 h 5-40% gradient of acetonitrile in 0.1% formic acid at 300 nL/min. The eluted peptides were analyzed with an Impact II Q-TOF spectrometer equipped with a CaptiveSpray electrospray source with an acetonitrile-enriched NanoBooster gas (Bruker). Data was acquired using data-dependent auto-MS/MS with a range 150-2200 m/z range, a fixed cycle time of 3 sec, a dynamic exclusion of 1 min, m/z-dependent isolation window (1.5-5 Th) and collision energy in the range 25-75 eV [72].

The raw data was processed using Andromeda, integrated into MaxQuant [73] (version 1.6.1.0). Specific searches with either trypsin (before K,R) or chymotrypsin (before F,Y,W,L,I,M) were performed against the *T. elongatus* (strain BP-1) proteome (2451 proteins; ID: UP000000440; NdhP and NdhQ appended). Focused searches were also performed against only Ndh subunits. In both cases, common contaminants provided by MaxQuant were also included. Tryptic and chymotryptic peptides were permitted a maximum of 2 missed cleavages. All searches were performed with cysteine carbamidomethylation as a fixed modification, as well as N-acetylation and methionine oxidation as variable modifications. Default MaxQuant instrument parameters for Bruker Q-TOF data were utilized, including a first search peptide mass tolerance of 0.07 Da, main search peptide tolerance of 0.006 Da, and isotope match tolerance of 0.005 Da. The false discovery rate threshold was set to 1%.

Cryo-electron microscopy grid preparation and image acquisition

For dataset-1 cryo-EM grid preparation, 3 µl of purified NDH (~35 nM) in buffer G was applied to R1/2 Cu 300 grids (Quantifoil) coated with a thin-layer of homemade continuous carbon film and glow-discharged for 15 s immediately before use. The sample was applied to the grid and incubated for 30 s at 100% humidity and 22ºC prior to blotting for 2-3 s with blot force 18 using Whatman #1 filter paper and plunged into liquid ethane cooled to liquid nitrogen temperatures using a FEI Vitrobot Mark IV. Dataset-2 grids were prepared as above, but at an estimated NDH concentration of ~80 nM and using R2/2 Cu 300 grids (Quantifoil).

Dataset-1 was acquired on a FEI Titan Krios G2 TEM operated at 300 kV and equipped with a Gatan K2 Summit direct electron detector camera at the Bay Area Cryo-EM facility at the University of California-Berkeley. Movies were recorded in super-resolution counting mode at an effective pixel size of 0.534 Å with a cumulative exposure of ~50 e/A^2 distributed uniformly over a 7 s exposure fractionated into 35 frames. Data acquisition was performed using SerialEM [74] with one exposure per hole and focusing for each exposure to a target defocus range of -1.3 to -3.0 µm.

Dataset-2 was acquired as above with the following modifications. The microscope was operated in EFTEM mode with a Gatan Quantum LS Imaging Filter set to a slit width of 25 eV. Three movies per hole were recorded with a super-resolution pixel size of 0.4516 Å, cumulative exposure of ~60 e/A^2 distributed uniformly over a 6 s exposure fractionated into 30 frames and target defocus range of -0.8 to -2.5 µm.

Image processing

For dataset-1, a total of 2519 movie stacks were collected as described above. Preprocessing was performed during data collection within Focus [75]. Super resolution movie stacks were drift-corrected in 5 x 5 patches, dose-weighted, and Fourier-cropped to 1.068 Å pix^{-1} using MotionCor2-v1.0.0 [76]. CTF parameters were estimated from the non-dose weighted, aligned micrographs using GCTF-v1.06 [77]. The data was manually inspected and micrographs of empty carbon, excessive ice-contamination, or poor power spectrum were removed resulting in 2192 micrographs that were used for further processing. All classification and reconstruction steps were performed using RELION-v2.1.0 [78], unless otherwise specified. A total of 4285 particles were picked manually from 100 micrographs and extracted in a 360^2 pixel box. These particles were subjected to reference-free 2D-classification to generate references for auto-picking of all micrographs using Gautomatch-v0.56. A total of 176,394 were extracted using a 360^2 pixel box and subjected to 3D-classification (K = 5) using a 60 Å lowpass-filtered *ab initio* reference generated in cryoSPARC [79]. Selected 3D classes were then subjected to reference-free 2D-classification with selected classes resulting in 119,592 particles. An additional round of 3D-classification (K = 3 or K = 5) did not yield classes with clearly distinguishable features. Particles were re-extracted, re-centered, and CTF parameters refined locally. Unmasked refinement using all particles yielded a reconstruction at 4.1 Å. Local movie-refinement and particle-polishing improved the resolution to 4.0 Å and inclusion of a soft mask during refinement further improved the resolution to 3.7 Å. All further classifications and focused refinements utilise soft masks around the regions of interest without image-alignment and using the consensus refinement as a starting point.

To further improve the overall map, 3D-classification of the entire complex was performed (K = 3, T = 12) and the most abundant class refined using a soft mask and local-searches to obtain a 3.7 Å overall map from 81,670 particles with improved features compared to the consensus refinement. The peripheral and membrane arms were each subjected to refinement with signal-subtraction [80] of the opposite domain. The outputs from refinement were then further classified (K = 3, T = 12, E-step limit = 8 Å) and the dominant class in each case refined. This resulted in a noticeably improved map for the peripheral arm at 3.6 Å resolution and a slight improvement in the connectivity in areas of the membrane arm at an overall resolution of 3.8 Å. A map with marginally improved connectivity for NdhF was obtained after masked classification (K = 3, T = 12, E-step limit = 8 Å) and refinement of the entire membrane arm at 3.9 Å resolution. The map quality of regions suspected of partial-occupancy corresponding to NdhS and the X-site were improved by further classification (K = 3 for NdhS and K = 5 for X-cofactor, T = 20, E-step limit = 8 Å) and subsequent refinement of the dominant bound and apo-states for the entire peripheral arm. More information on the steps used to obtain the X-cofactor map is available in the Supplemental Information of this manuscript.

During preparation of this manuscript, RELION-v3.0.b2 [81] was released with a new CtfRefine function that enabled estimation and correction of beam-tilt. CtfRefine was used for the overall and X-cofactor bound/apo maps which lead to final resolutions of 3.1 Å, 3.4 Å and 3.3 Å.

For dataset-2, a total of 1385 movies were collected as described above and preprocessed through Focus as described for dataset-1 using the appropriate pixel size and exposure. After

manual inspection, 1364 micrographs were chosen for further processing. The 2D-references from dataset-1 were used for auto-picking using Gautomatch-v0.56 with appropriate scaling. All classification and reconstruction steps were performed using RELION-v3.0.b2. A total of 98,145 particles were extracted using a 448^2 pixel box with a binning factor of 3.5 (NQ ~ 6.3 Å) and subjected to 2 rounds of reference-free 2D classification. Classes showing clear secondary structure features were selected resulting in 90,266 particles. These particles were re-extracted using a 424^2 pixel box with a binning factor of 1.25 (NQ ~ 2.3 Å). Refinement using a 35 Å lowpass-filtered reference from dataset-1 resulted in a 4.1 Å with density for an additional TMH adjacent NdhD/F compared to dataset-1, later identified as NdhQ. All further refinement steps are performed using a soft-mask and local-searches about the consensus refinement. Subsequent refinement of both beam-tilt and per-particle defocus with CtfRefine and Bayesian polishing [81,82] resulted in improvements of 0.5 and 0.2 Å, respectively. Further 3D-classification without image-alignment (K= 3, T = 12) and refinement of the dominant class resulted in a 3.3 Å reconstruction from 66,760 particles. To improve the distal region of the membrane arm, focused classification was performed utilizing a soft mask around the distal region of the membrane arm (K = 3, T = 16, E-step limit = 8). The top class displayed clear density for NdhQ and was selected for subsequent refinement, resulting in a 3.5 Å reconstruction of the overall complex from 34,655 particles. The X-cofactor map for this data set was obtained as for dataset 1, but using 3 classes (K = 3) for the classification step.

Specifically, for the X-cofactor maps, The X-factor was initially observed during unmasked 3D-classification of the overall complex in dataset-1 at similar intensity as the 4Fe-4S (Figure 2.9). After model-building the surrounding protein density, however, characteristic residues for FeS clusters (e.g., Cys and His) were found not to be present at this site. To further improve the density for this region, focused 3D-classification on the X-site using a ~45 Å diameter spherical density centered about the X-site and softened with cosine-edge of 8 voxels (~8.5 Å) as depicted in Figure 2.3. The 3D classification was performed on membrane arm signal-subtracted particles without image alignment and using the peripheral arm map lowpass-filtered to 35 Å as the initial reference, K = 5, T = 20 and limiting the expectation step to 8 Å information. The same parameters were used for dataset-2 but with K = 3 due to fewer particles. The X-cofactor classes in both cases were then selected for 3D auto-refinement in RELION using the peripheral arm map low-passed filtered to 35 Å as a reference with a mask over the entire peripheral arm softened with cosine-edge of 8 voxels (~8.5 Å) and local angular searches.

All reported global map resolutions correspond to ‘gold-standard’ Fourier shell correlation (FSC) = 0.143 between independently refined half-maps with phase-randomization to correct for masking artifacts and *B*-factors calculated as implemented in the post-processing function of RELION [83,84]. Local resolution maps were generated with RELION using default parameters. Directional FSC curves and map anisotropy were assessed using the 3DFSC [85]. For visualization, maps were corrected for the modulation transfer function of the detector and filtered to the global resolution.

Model building and refinement

Briefly, the maps from the two datasets were aligned and resampled onto the same grid using UCSF Chimera-v1.11.2 [86]. For the 11 core subunits (NdhA-K), homology models were

generated based on the *T. thermophilus* complex I crystal structure (PDB ID-4HEA) using the I-TASSER web-server [87]. Secondary structure predictions for all subunits were generated using Phyre2, PsiPred and ESPript-3.0 web-servers [88–90]. All manual adjustment and building was performed using COOT-v0.8.8 [91,92]. The homology models were adjusted and rebuilt into the maps. The remaining density was assigned to the novel OPS subunits. The initial models for the two datasets were then refined in PHENIX-v1.13 using *phenix.real_space_refine* with auto-detection of secondary structure assignments [93,94]. The models and maps were then inspected to manually adjust secondary structure assignments. This was followed by iterative rounds of real-space refinement in PHENIX and manual rebuilding in COOT. The 'Composite' map and model was then made from NdhC/F/Q density from dataset-2 and the remainder from dataset-1 using 'vop maximum' in Chimera after normalizing maps to similar grey-scale using 'vop rescale'. The final models were assessed by MolProbity-v4.4 and map-vs-model FSC [95].

Specifically, initial modelling predominantly relied on the 3.7 Å overall complex and 3.6 Å peripheral arm maps from dataset-1 obtained using RELION-v2.1.0. Focused classifications and refinements were performed in conjunction with model building which yielded the improved NdhS region in the 3.8 Å NdhS-focus map of the peripheral arm and marginally improved distal membrane arm region with the 3.9 Å membrane arm map. After placement and rebuilding of homology models for the core subunits, the remaining density was assigned to OPS subunits. The density corresponding to NdhS was initially identified by its all β-strand secondary structure and a homology model of a crystal structure (PDB ID-3C4S) of NdhS from a mesophilic cyanobacterium. NdhL was identified first based on its predicted two TMHs. The density corresponding to a single TMH between NdhD and NdhF was suspected as either NdhP or NdhQ and was ultimately identified as NdhP based on fit of landmark aromatic residues. NdhO was identified based on its predominantly β-strand structure. NdhM and NdhN were identified by testing primary sequence registers after tracing the main-chains for these subunits. No density was observed for NdhQ in dataset-1 nor residues 517-583 of NdhF, which were left unmodelled. After initial model building and refinement against the 3.7 Å overall map, the cryo-EM data was reprocessed with estimated beam-tilt correction in the recently released RELION-v3.0.b2 to obtain the 3.1 Å overall map from dataset-1. The 3.7 Å refined model was consistent with the 3.1 Å map, but with improved side-chain density.

Dataset-2 was collected after model-building for dataset-1 using sample from a different round of purification and processed entirely through RELION-v3.0.b2. Dataset-2 yielded a lower global resolution for the overall complex, but with an improved membrane arm (Figure 2.3c) which revealed density for NdhQ and NdhF residues 517-583.

In addition to the improved distal region of the membrane arm, dataset-2 shows differences in the N-termini of NdhA and NdhC. For dataset-1, weak cylindrical density is present adjacent the N-terminus of NdhA. This density extends at a slant through the detergent-belt from the lumenal to the stromal side of the complex. The density is consistent with an approximately 24 residue helix that is tentatively assigned as NdhA residues 4-27 in the dataset-1 model (PDB ID-6NBQ), but left as UNK due to the absence of clear side-chain density. A similar density is not observed in dataset-2 map, therefore it was not included in the composite model (PDB ID-6NBY) and may represent a detergent-belt artifact. Adjacent to this density, NdhC N-terminal TMH (residues 20-43) also differs between the two datasets. The helix in dataset-2 (PDB ID-6NBX)

most closely resembles the position of its homologue Nqo7 in complex I (PDB ID-4HEA) and appears near upright. The dataset-1 helix is tilted at a ~55^{o} angle towards the distal region of the membrane arm with respect to the pivot point of residue 20. Connecting density from 43-65 is not observed in either map and is predicted to be a disordered region. It is unclear if this difference in position is related to the NdhA N-terminal density. The position from dataset-2 was used for the Composite model due to its similarity to the homologue in complex I.

Visualization and analysis

Figures of maps and models were prepared with Chimera and PyMOL-v1.8.2 [96]. Superposition of structure was performed in PyMOL using the 'super' command. The electrostatic surface representation was generated with the APBS-v2.0 [97] plugin in PyMOL using PDB2PQR's PARSE forcefield and pH = 8, to approximate the daytime stromal/cytosolic pH, in order to prepare the model for calculations. All alignments were performed using the Clustal Omega [98] web-server with sequences obtained from Uniprot [99].

2.7 Accession numbers

Raw micrographs and particles images have been deposited with the Electron Microscopy Public Image Archive Repository under accession number EMPIAR-10257. Cryo-EM maps for the dataset 1 overall, peripheral arm, with NdhS, without NdhS, with X-cofactor and without X-cofactor and dataset 2 overall have been deposited with the Electron Microscopy Data Bank under accession numbers EMD-0415, EMD-0416, EMD-0417, EMD-0418, EMD-0419, EMD-0420 and EMD-0425, respectively. Atomic coordinates for the overall complex for dataset 1, dataset 2 and composite have been deposited with the PDB under accession codes 6NBQ, 6NBX and 6NBY, respectively. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD012206.

Chapter 3

Discovery and characterization of the Dabs accumulate bicarbonate complex

†The work presented in this chapter is adapted from the previously published article:
Desmarais JJ, Flamholz AI, Blikstad C, Dugan EJ, Laughlin TG, Oltrogge LM, Chen AW, Wetmore K, Diamond S, Wang JY, Savage DF. (2019) DABs are inorganic carbon pumps found throughout prokaryotic phyla *Nature Microbiology* 4 (12) 2204-2215.

3.1 Abstract

Bacterial autotrophs often rely on CO_2 concentrating mechanisms (CCMs) to assimilate carbon. Although many CCM proteins have been identified, a systematic screen of CCM components has not been carried out. Here, we performed a genome-wide barcoded transposon screen to identify essential and CCM-related genes in the ɣ-proteobacterium *H. neapolitanus*. Screening revealed that the CCM comprises at least 17 and likely no more than 25 genes most of which are encoded in 3 operons. Two of these operons contain a two-gene locus encoding a domain of unknown function (PFAM:PF10070) and a putative cation transporter (PFAM:PF00361). Physiological and biochemical assays demonstrate that these proteins, which we name DabA and DabB for "DABs accumulate bicarbonate," assemble into a heterodimeric complex, contain a putative β-carbonic anhydrase-like active site, and function as an energy-coupled inorganic carbon (C_i) pump. Surprisingly, DabAB operons are found in diverse bacteria and archaea. We demonstrate that functional DABs are present in the human pathogens *B. anthracis* and *V. cholerae*. Based on these results, we propose that DABs constitute a class of energized C_i pump and play a critical role in C_i metabolism throughout prokaryotic phyla.

3.2 Introduction

Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase (rubisco) is the primary carboxylase of the Calvin-Benson-Bassham (CBB) cycle and the major entry point of C_i into the biosphere. rubisco activity is critical to agriculture and a major flux removing anthropogenic CO_2 from the atmosphere. Despite its centrality and abundance, rubisco is not a fast enzyme [100,101]. Nor is rubisco specific - all known rubiscos can use molecular oxygen (O_2) as a substrate in place of CO_2 [102]. Oxygenation does not fix carbon and produces a product that must be recycled through metabolically-expensive photorespiratory pathways[103]. Many studies support the hypothesis that improvements to rubisco could improve crop yields, but rubisco has proven recalcitrant to protein engineering. It remains unclear whether or how rubisco can be improved [27,101,104].

Organisms that depend on rubisco for growth often employ CO_2 concentrating mechanisms (CCMs) that concentrate CO_2 near rubisco so that carboxylation proceeds at high rate and specificity [105,106]. All cyanobacteria and many chemolithoautotrophic proteobacteria have a CCM [107]. The bacterial CCM has garnered particular interest among bioengineers because it is well-understood, thought to consist of relatively few genes and operates inside single cells [108]. Detailed modeling suggests that transplantation of the bacterial CCM into crops might improve yields [109,110] and efforts towards transplantation are already underway [111].

Diverse experimental studies make it clear that the bacterial CCM requires two components to function: active C_i transport driving accumulation of HCO_3^- in the cytosol and organization of rubisco with carbonic anhydrase (CA) in the lumen of a protein organelle called the carboxysome [105,112,113]. Energy-coupled C_i pumps keep the cytosolic HCO_3^- concentration high (> 10 mM) and, crucially, out-of-equilibrium with CO_2 [112,114,115]. CA activity interconverts $HCO_3^- + H^+$ with $CO_2 + H_2O$, and thus, the carboxysomal CA converts a high cytosolic HCO_3^- concentration into a high carboxysomal CO_2 concentration, promoting faster carboxylation by rubisco and competitively inhibiting oxygenation [105]. Genetic lesions to either component - C_i uptake systems or carboxysomes - disrupt the CCM and mutants have growth defects unless CO_2 is

supplemented [116,117]. This high-CO_2 requiring (HCR) mutant phenotype is commonly-used to identify CCM components in screens [116,117].

Despite early screens, a comprehensive list of bacterial CCM components remains unknown, leaving the possibility that additional activities are required for CCM function. Although well-assembled carboxysome structures can be produced in bacteria and plants [111,118], the functionality of these carboxysomes in a heterologous CCM has not been demonstrated. Bioinformatic studies show that several non-carboxysomal genes are associated with carboxysome operons [119,120]. Further, experimental [112,121] and modeling studies [105,113] make it clear that energy-coupled C_i uptake systems are required for CCM function. Several different C_i pump families, including transporters and facilitated uptake systems are known for cyanobacterial lineages, but mechanistic understanding of C_i uptake is limited [122].

Here we use a genome-wide barcoded transposon mutagenesis screen (RB-TnSeq) to interrogate the CCM of *Halothiobacillus neapolitanus* (henceforth *Hnea*). *Hnea* is a sulfur oxidizing ɣ-proteobacterial chemolithoautotroph and a model system for studying α-carboxysomes [123–125]. Older physiological measurements suggest that *Hnea* possesses an energized C_i uptake system, but the molecular identity of this activity is unknown [115]. In addition to producing the essential gene set for a bacterial chemolithoautotroph, we leverage our pooled mutant library to comprehensively screen for knockouts that produce an HCR phenotype. This screen identified all known CCM components and confirmed that a two-gene operon containing a large, conserved, poorly-characterized protein (PFAM:PF10070, hereafter DabA) and a cation transporter (PFAM:PF00361, hereafter DabB) is required for CCM function. Scott and colleagues have recently identified and validated homologs of these genes as a C_i import system in hydrothermal vent chemolithoautotrophs [126–128]. Based on this work and results further described below, we propose naming this locus the DAB operon for "**D**ABs **A**ccumulate **B**icarbonate."

Here we show that the products of the DAB operon form a protein complex that is capable of energetically-coupled C_i uptake. Both proteins are necessary for activity and treatment with an ionophore abrogates DAB-mediated C_i uptake. Structural homology modeling suggests that DabA contains a domain distantly homologous to a type II β-CA. Indeed, DabA binds zinc, likely in a manner similar to β-CAs. These results are consistent with a model of activity dependant on unidirectional hydration of CO_2 to HCO_3^- in the cytosol via a CA-like mechanism and energized by coupling to a cation gradient. Phylogenomic analysis demonstrates that DAB operons are widespread throughout prokaryotes including carbon-fixing bacteria and archaea. Surprisingly, DAB operons are also found in many heterotrophic bacteria. We demonstrate that functional operons are present in the notable pathogens *V. cholera* and *B. anthracis*. We therefore propose that DABs constitute a class of C_i uptake pump whose biochemical tractability facilitates mechanistic analyses and whose widespread occurrence merits further investigation.

3.3 Results

We constructed a randomly-barcoded genome-wide pooled knockout library of *Hnea* by conjugation (Figure 3.1a). The conjugated vector contained a barcoded Tn5-derived transposon encoding a kanamycin resistance marker. The library was produced in 5% CO_2 enabling isolation of CCM gene knockouts.

Transposon barcodes simplify the use of the library for pooled screens using the 'barseq' approach (Materials and methods) [129]. Transposon insertion sites and cognate barcodes were mapped using standard TnSeq methods (Material and Methods) [129]. The library was found to contain ~10^5 insertions, or one insertion for every ≈25 base pairs in the *Hnea* genome. Since the average gene contains ≈35 insertions, genes with no insertions are very likely essential for growth. A simple statistical model identified 551 essential genes and 1787 nonessential genes out of 2408 genes in the *Hnea* genome (Materials and methods, Figure 3.1a,b). The remaining 70 genes were classified as "ambiguous" due either to their short length or because replicate mapping experiments were discordant (Materials and methods). Genes associated with known crucial functions including central carbon metabolism, ribosome production, and DNA replication were found to be essential (Figure 3.1c, Appendix Figure B.1). Importantly, known CCM genes, including carboxysome components, were not essential for growth at 5% CO_2 (Figure 3.2).

Based on the current model of the bacterial CCM (Figure 3.2b), knockouts of CCM genes are expected to have reduced fitness in atmospheric CO_2 conditions [116,117]. As our pooled library contains ~70,000 barcodes that map to exactly one position in the *Hnea* genome, we were able to use barseq to quantify the fitness effects of single gene knockouts for all nonessential *Hnea* genes in a pooled competition experiment (Materials and methods, Figure 3.2b) [129]. Since the library contains roughly 20 uniquely-mapped knockouts per gene, this screen contains multiple internal biological replicates testing the effect of gene knockouts. Mutants in a particular gene were designated as HCR if the average effect of a knockout in that gene was a twofold (or greater) growth defect in ambient CO_2 as compared to 5% in two replicate experiments. As expected, knockouts of carboxysome genes consistently produced large and specific fitness defects in ambient CO_2 (Figures 3.2b,c and Table 3.1) [125]. These genes include *cbbLS* - the large and small subunits of the α-carboxysomal rubisco [107]; *csoS2* - an intrinsically disordered protein required for α-carboxysome assembly [130]; *csoSCA* - the carboxysomal carbonic anhydrase [107]; *csoS4AB* - the pentameric proteins thought to form vertices of the α-carboxysome [131]; and *csoS1CAB* - the hexamers that form the faces of the α-carboxysome shell [107,123]. Knockouts of *csoS1D*, a shell hexamer with a large central pore [118,132], had too weak a phenotype to be considered HCR (Figures 3.2b,c). The *Hnea* genome also contains a non-carboxysomal Form II rubisco that is likely not involved in CCM activity as its disruption confers no fitness defect. A number of genes that are not associated with the carboxysome structure also exhibited HCR phenotypes. These include two LysR transcriptional regulators, a Crp/Fnr type transcriptional regulator, a protein called acRAF that is involved in rubisco assembly [133,134], and two paralogous loci encoding DAB genes (hereafter DAB1 and DAB2, Figures 3.2b-f).

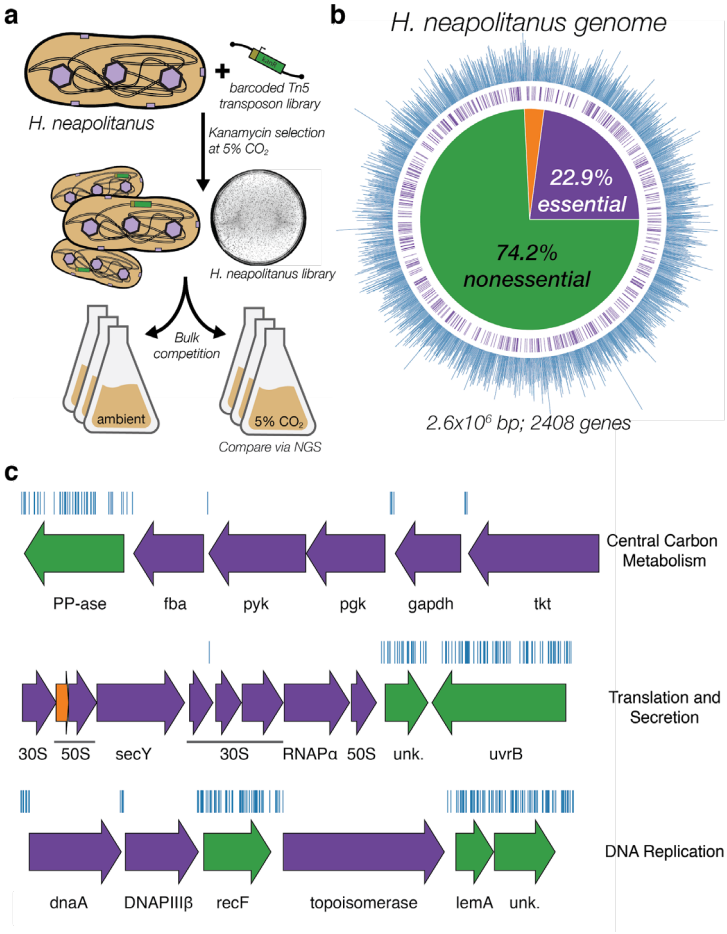


Figure 3.1 Transposon mutagenesis reveals essential gene set of a chemolithoautotrophic organism. (**a**) Schematic depicting the generation and screening of the RB-TnSeq library. Transposons were inserted into the *Hnea* genome by conjugation with an *E. coli* donor strain. The transposon contains a random 20 base pair barcode (yellow) and a kanamycin selection marker (green). Selection for colonies containing insertions was performed in the presence of kanamycin at 5% CO_2 and insertions were mapped by sequencing.. Subsequent screens were carried out as bulk competition assays and quantified by BarSeq. (**b**) Insertions and essential genes are well-distributed throughout the *Hnea* genome. The outer track (blue) is a histogram of the number of barcodes that were mapped to a 1 kb window. The inner track annotates essential genes in purple. The pie chart shows the percentages of the genome called essential (purple), ambiguous (orange), and nonessential (green). (**c**) Representative essential genes and nonessential genes in the *Hnea* genome. The blue track indicates the presence of an insertion. Genes in purple were called essential and genes in green are nonessential. Genes labeled "unk." are hypothetical proteins. The first genomic locus contains 5 essential genes involved in glycolysis or the CBB cycle including pyruvate kinase (pyk) and transketolase (tkt). The 8 essential genes in the second locus encoding 30S and 50S subunits of the ribosome, the secY secretory channel, and an RNA polymerase subunit. Essential genes in the third example locus include topoisomerase and DNA polymerase III β.

DAB1 is a cluster of 2 genes found in an operon directly downstream of the carboxysome operon (Figure 3.2c). Though DAB1 is part of an 11-gene operon containing several genes associated with rubisco proteostasis, including acRAF [133,134] and a cbbOQ-type rubisco activase [135], we refer to DAB1 as an "operon" for simplicity. DAB2 is a true operon and is not proximal to the carboxysome operon in the *Hnea* genome. These "operons" are unified in that they both display HCR phenotypes and possess similar genes (Figures 3.2b-d).

Both operons contain a conserved protein of unknown function (PFAM:PF10070) that we term DabA. DabAs have no predicted transmembrane helices or signal peptides and appear to be large (DabA1: 118.5 kDa, DabA2: 91.7 kDa), soluble, cytoplasmic proteins (Materials and methods, Figure 3.3a). Both DAB operons also contain a member of the cation transporter family (PFAM:PF00361) that includes H^+-pumping subunits of respiratory complex I and Mrp Na^+:H^+ antiporters [136]. This protein, which we call DabB (DabB1: 62.2 kDa, DabB2: 59.3 kDa), is predicted to have 12-13 transmembrane helices (Figure 3.3a). The complex I subunits in PF00361 are H^+-pumping proteins and do not contain redox active groups, e.g. iron-sulfur clusters or quinone binding sites. Phylogenetic analysis suggests DabB proteins form a clade among PF00361 members (Figure 3.4a) distinct from complex I subunits. Therefore, homology between DabB and complex I subunits (e.g., NuoL) suggests cation transport but does not imply redox activity.

Importantly, operons of this type were recently demonstrated to be capable of C_i uptake in the hydrothermal vent chemolithoautotroph *Hydrogenovibrio crunogenus* [126–128].

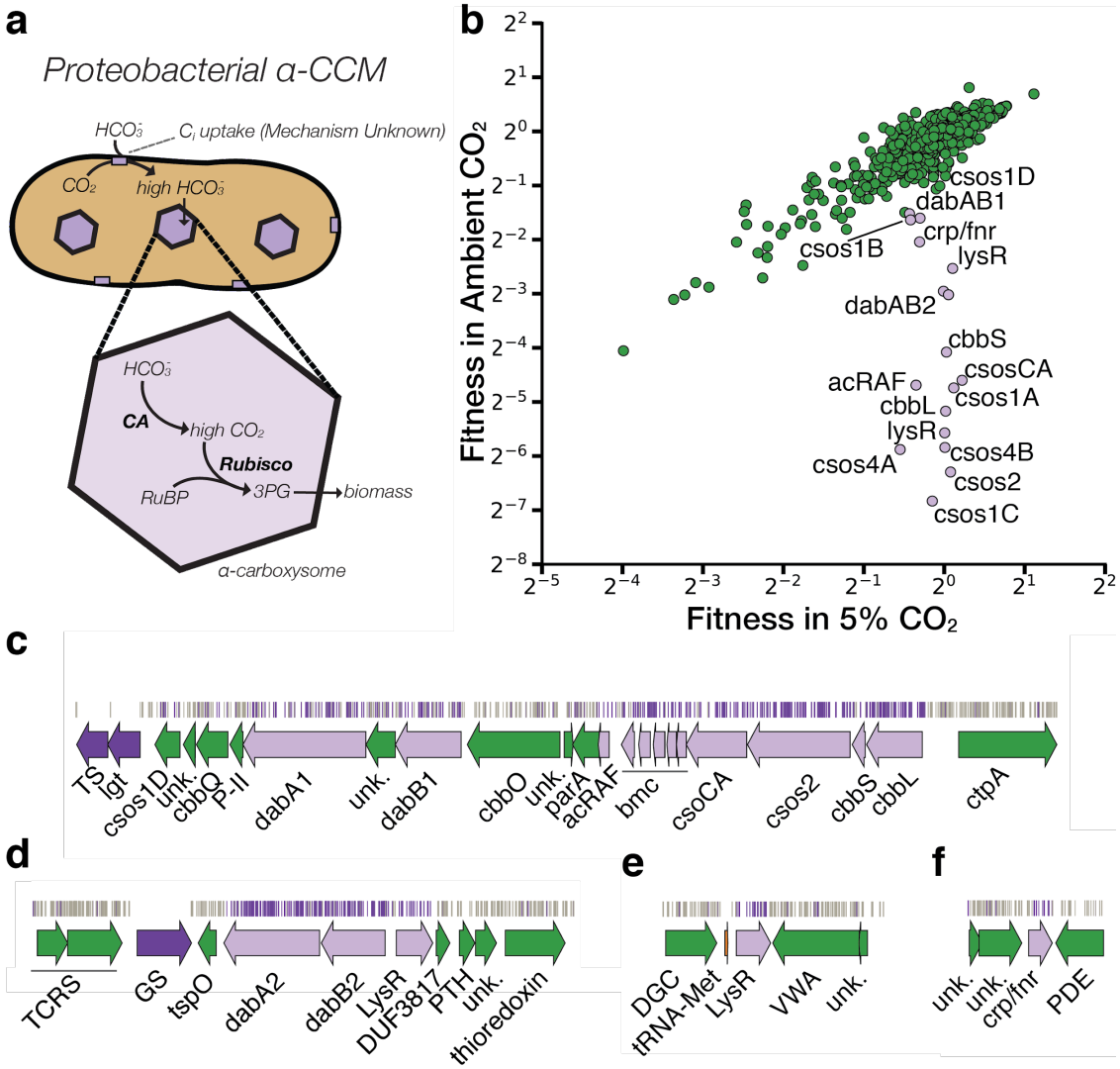


Figure 3.2 A systematic screen for HCR mutants identifies genes putatively associated with the CCM. (**a**) Simplified model of the α-CCM of chemolithoautotrophic proteobacteria. Inorganic carbon is concentrated via an unknown mechanism, producing a high cytosolic HCO_3^- concentration. High cytosolic HCO_3^- is converted into high carboxysomal CO_2 by CA, which is localized only to the carboxysome. (**b**) Fitness effects of gene knockouts in 5% CO_2 as compared to ambient CO_2. Data is from one of two replicates of BarSeq. The effects of single transposon insertions into a gene are averaged to produce the gene-level fitness value plotted. We define HCR mutants as those displaying a twofold fitness defect in ambient CO_2 relative to 5% CO_2 in both replicates. HCR genes are colored light purple. Data from both replicates and the associated standard errors are shown in Supplemental Figure 2 and in supplemental file 3. Panels (**c-f**) show regions of the *Hnea* genome containing genes annotated as HCR in panel A. Essential genes are in dark purple, HCR genes are in light purple, and other genes are in green. The top tracks show the presence of an insertion in that location. Insertions are colored grey unless they display a twofold or greater fitness defect in ambient CO_2, in which case they are colored light purple. (**c**) The gene cluster containing the carboxysome operon and a second CCM-associated operon. This second operon contains acRAF, a Form IC associated cbbOQ-type Rubisco activase and *dabAB1*. (**d**) The DAB2 operon and surrounding genomic context. (**e**) The genomic context of a lysR-type transcriptional regulator that shows an HCR phenotype. (**f**) Genomic context of a crp/fnr-type transcriptional regulator that displays an HCR phenotype. Genes labeled "unk." are hypothetical proteins. Full gene names are given in Appendix Figure B.3 and accession numbers and gi numbers for selected genes can be found in Table 3.1.

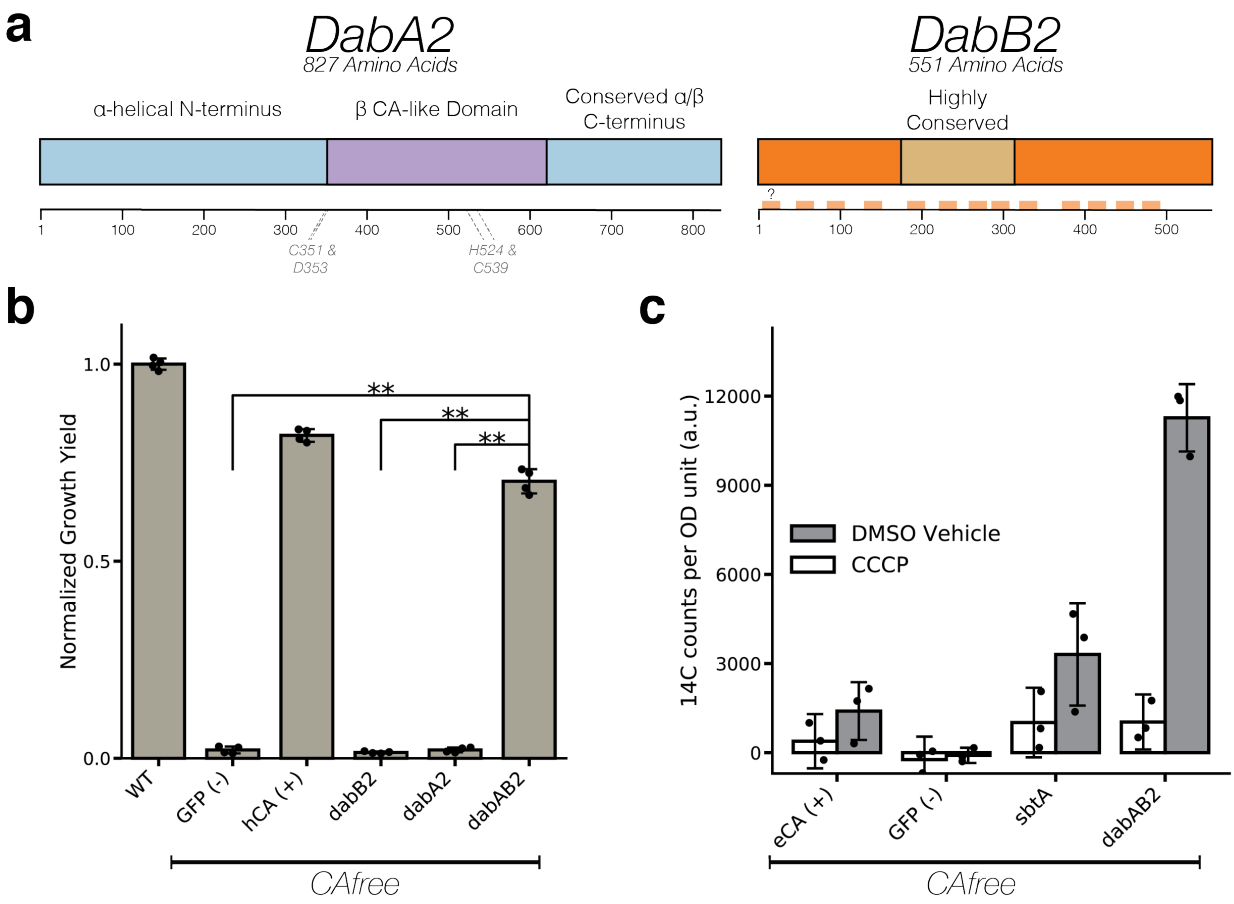


Figure 3.3 The DABs catalyze active transport of C_i and are energized by a cation gradient. (**a**) Diagrammatic representation of DabA2 and DabB2 based on bioinformatic annotation. The four predicted active site residues (C351, D353, H524, C539) are marked on the primary amino acid sequence. Amino acid numbers are marked below each gene and predicted transmembrane helices are marked in light orange. (**b**) DAB2 was tested for ability to rescue growth of CAfree *E. coli* in ambient CO_2 conditions. Expression of the full operon (DabAB2) rescues growth, as does the positive control, and human carbonic anhydrase II (hCA). dabAB2 has a larger rescue than GFP ($t=42.6$, corrected $p=3.4*10^{-8}$), dabA2 ($t=43.4$, corrected $p=3*10^{-8}$), and dabB2 ($t=44.5$, corrected $p=2.6*10^{-8}$). "**" denotes that means are significantly different with Bonferroni corrected $p < 5X10^{-4}$ according to a two-tailed t-test. Bar heights represent means and error bars represent standard deviations of 4 biological replicates. Consistent results were seen after 2 independent experiments. **c).** CAfree *E. coli* were tested for C_i uptake using the silicone-oil centrifugation method. Expression of DabAB2 produced a large increase in ^{14}C uptake as compared to all controls. Moreover, treatment with the ionophore CCCP greatly reduces DabAB2-mediated ^{14}C uptake, suggesting that DabAB2 is coupled to a cation gradient. *E. coli CA* (eCA) was used as a control for a non-vectorial CA. *Synechococcus elongatus* PCC 7942 *sbtA* was used as a known C_i transporter. GFP was used as a vector control. Bar heights represent the mean and error bars represent standard deviations of 3 technical replicates. Consistent results were seen with 3 independent experiments.

In order to facilitate testing for C_i transport activity, we generated an *E. coli* strain, CAfree, that has knockouts of both CA genes (Materials and methods). It was previously shown that deletion of the constitutive CA, *can*, gene produces an HCR phenotype in *E. coli [137]* that is complemented by expression of cyanobacterial bicarbonate transporters [138]. Deleting both CA genes replicates this phenotype and greatly reduces the likelihood of escape mutants. Since DAB2 disruption is associated with a larger fitness defect than DAB1 (Figure 3.2b), we used CAfree to test DAB2 for C_i uptake activity. DAB2 expression enables growth of CAfree in ambient CO_2 while expression of either gene alone is not sufficient (Figure 3.3b, Appendix B.5). $^{14}C_i$ uptake assays demonstrate that DAB2 facilitates import of extracellular C_i to levels above that of the

appropriate control (Figure 3.3c). Moreover, DAB2-associated C_i uptake is strongly inhibited by the ionophore CCCP (white bars in Figure 3.3c), indicating that DAB2 is energetically-coupled, either directly or indirectly, to a cation gradient (e.g., H^+ or Na^+). This is consistent with previous observations that C_i uptake in *Hnea* is powered by a membrane gradient [115].

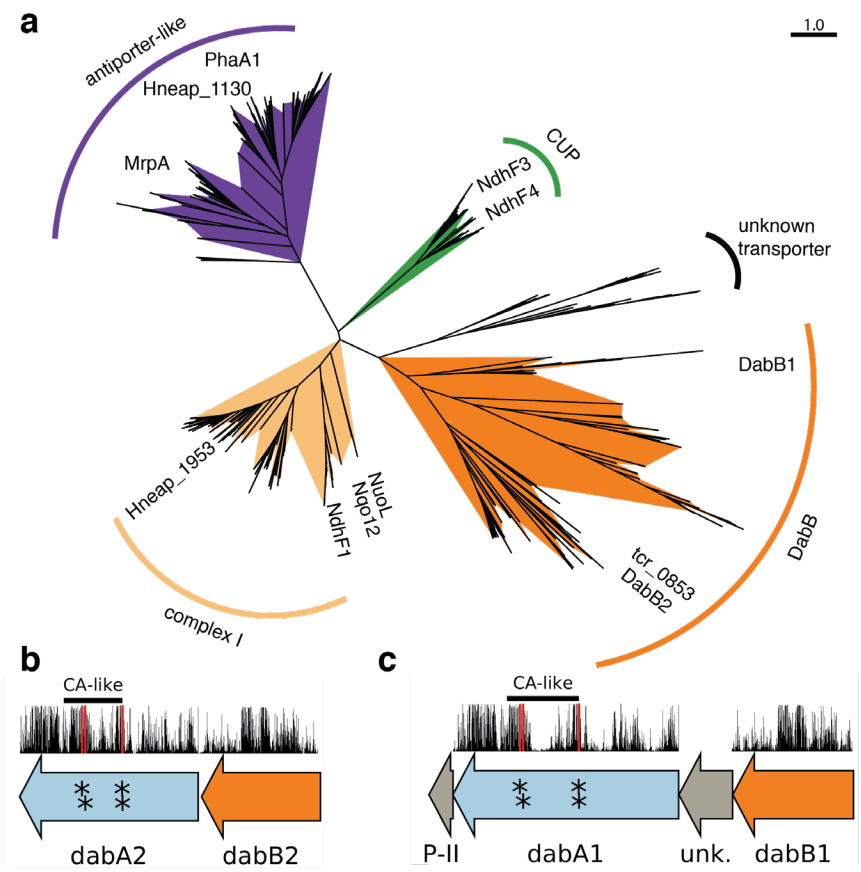


Figure 3.4 PF0361 contains multiple subfamilies with some regions of high conservation. (**a**) PF0361 is a large and diverse protein family containing multiple subfamilies with different reported activities. These subfamilies include Mrp-family cation antiporters, proton translocating subunits of complex I, membrane subunits of CUP (CO_2 uptake protein) complexes, and DabB proteins. These subfamilies are highly diverged and perform a variety of activities. This means that it is not possible to draw conclusions about the mechanism of DAB complexes just from their homology to PF0361. This panel contains an approximate maximum likelihood tree of PF0361 genes. Clades were colored according to the presence of genes with known functions. The purple clade contains *B.subtilis* and *S. aureus* MrpA cation antiporter subunits and *S. meliloti* antiporter PhaA1. The light orange clade contains the known cation translocating subunits of complex I: nuoL from *E. coli*, Nqo12 from *T. thermophilus*, and NdhF1 from both *S.elongatus* PCC7942 and *T. elongatus* BP-1. The green clade contains CUP-associated membrane subunits ndhF3 from both *S. elongatus* PCC7942 and *T. elongatus* BP-1 and ndhF4 from from the same two species. The dark orange clade includes DabB1-2 and tcr_0853 from *T. crunogena*. We note that the clade containing DabB1-2 is distinct from that containing known complex I subunits or to mrp-family antiporters. This tree is consistent with our model, where DabB is not bound to a redox-coupled complex but rather couples redox-independent cation transport to CA activity (as shown in Figure 5). No conclusions should be drawn from the number of sequences in each clade as an exhaustive search for homologs was not performed to ensure that all members of each clade are represented. Scale bar indicates one substitution per site. The tree contains 566 sequences. DAB1 is a segment of an 11-gene operon directly downstream of the carboxysome operon that contains CCM-associated genes. Both DAB1 (**b**) and DAB2 (**c**) "operons" contain two distinct genes that we label DabB and DabA. DabA is annotated as Domain of Unknown Function 2309 (DUF2309, PFAM:PF10070) and appears to be a soluble protein. Approximately one third of dabA is distantly homologous to a type II β-CA. CA-like regions are marked with a line, and the four residues expected to be involved in binding the catalytic zinc ion are marked by asterisks. The height of the asterisks has been varied to make them distinguishable despite proximity in sequence space. DabB is homologous to a cation transporter in the same family as the H+ pumping subunits of respiratory complex I (PFAM:PF00361). The DAB1 operon also contains a protein of unknown function between DabA1 and DabB1. This protein has distant homology to DabA1 but is truncated to half the length. Vertical bars above the genes indicate percent conservation of that particular amino acid position in a multiple sequence alignment. Active site residues are in red. All active site residues are highly conserved with percent identities of greater than 99%. One active site cysteine and the active site aspartate residue are the two most conserved residues in DabA with 99.9% identity each.

In order to determine if the genetic interaction between *dabA2* and *dabB2* reflects a physical interaction, we attempted to co-purify the two proteins. DabA2 was fused to a C-terminal Strep-tag, DabB2 was fused to a C-terminal GFP with 6xHis-tag, and the genes were co-expressed in *E. coli* (Material and methods). Tandem-affinity purification following detergent solubilization revealed that DabA2 and DabB2 form a complex in *E. coli* (Figure 3.5a). The complex runs as a single major peak on size exclusion chromatography and has a retention volume consistent with a heterodimer of DabA2 and DabB2 (Figure 3.5b). We did not observe co-purification of any *E. coli* proteins suggesting that DAB2 operates as an independent complex within the membrane (Figure 3.5a). Moreover, *DAB2* expression rescues CAfree growth even when complex I is knocked out (Δ(*nuoA-nuoN*)) (Figure 3.6), providing further evidence that DAB function is independent of complex I.

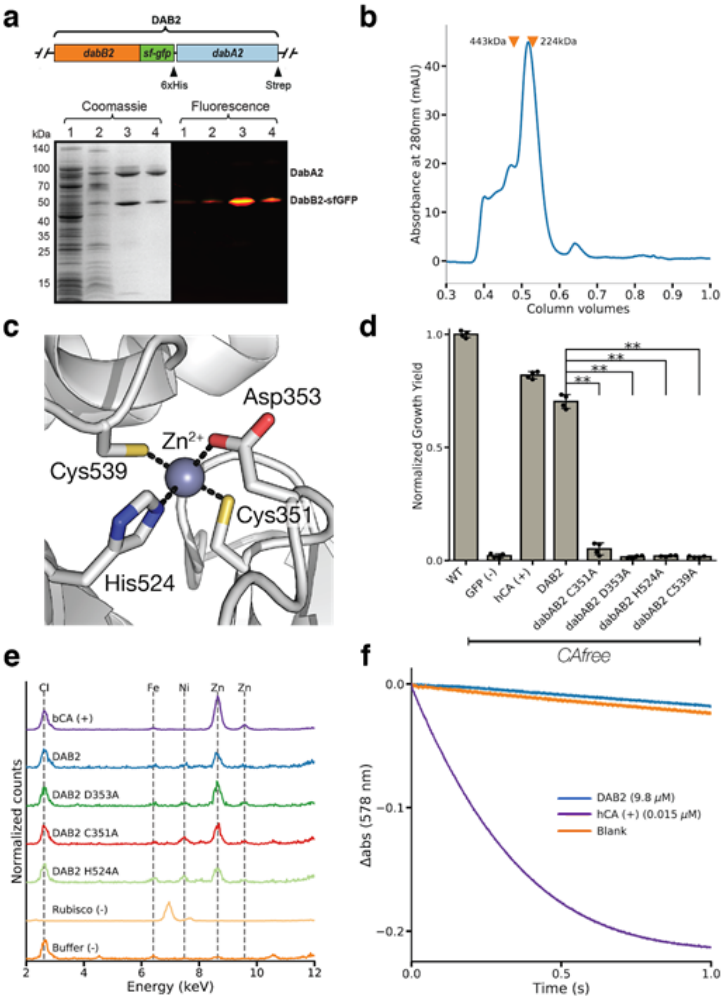


Figure 3.5 DabA contains a β-CA-like active site but is not active outside of the membrane. (**a**) Purification of DabAB2 complex from *E. coli*. DabA2 was C-terminally tagged with a Strep-tag and DabB2 was C-terminally tagged with sf-GFP and a 6xHis-tag. Purification was monitored using SDS-PAGE imaged with fluorescence (right view) before coomassie staining (left view). Lane 1: clarified lysate; 2: solubilized membranes; 3: Ni-NTA resin eluent; 4: strep-tactin resin eluent. DabA2 and DabB2 co-purify as a single complex without any obvious interactors. Similar results were observed after 3 independent purifications. (**b**) Size-exclusion chromatogram of His/Strep purified DabAB2 with retention volumes (orange arrows) and molecular weights (kDa) indicated for standard samples (apoferritin, 443 kDa; β-amylase, 224 kDa). DabAB2 runs with an effective mass of ~270 kDa, which likely reflects an oligomer of DabA and DabB. Similar results were observed after 3 independent purifications. (**c**) Structural model of the DabA2 active site based on a β-CA from *E. coli* (PDB ID 1I6P). Typical β-CAs rely on two cysteines and one histidine to bind Zn^{2+}. The aspartic acid coordinates Zn^{2+} but is likely displaced during catalysis. (**d**) Alanine mutants of the putative DabA2 active site residues abrogate rescue of CAfree *E. coli* compared to wild-type dabAB2 (C351A, t=54.3, $p=1.1*10^{-8}$; D353A, t=144, $p=3.1*10^{-11}$; H524A, t=44, $p=3.7*10^{-8}$; C539A, t=44.3, $p=3.5*10^{-8}$; all p values listed here are Bonferroni corrected). "**" denotes that means are significantly different with Bonferroni corrected $p < 5X10^{-4}$ according to a two-tailed t-test. Bar heights indicate means and error bars give standard deviations of four biological replicate cultures. (**e**) X-ray fluorescence data indicate that DabAB2 binds zinc like all known β-CAs. Single mutations to the active site do not abrogate zinc binding. Curves are from representative samples. Technical replicate traces were concordant (WT: 9, D353A:5, H524A:4, C351A:3, Rubisco: 2, Blank: 4, BCA: 3). Replicate traces for DAB2 H524A include samples from two independent purifications. (**f**) Purified DabAB2 does not display any obvious CA activity despite being present in 650-fold excess over the positive control (Human carbonic anhydrase II, hCA) in our assays. Curves display averages of 7 experimental traces +/- standard error of the mean. Similar results were observed in two independent purifications.

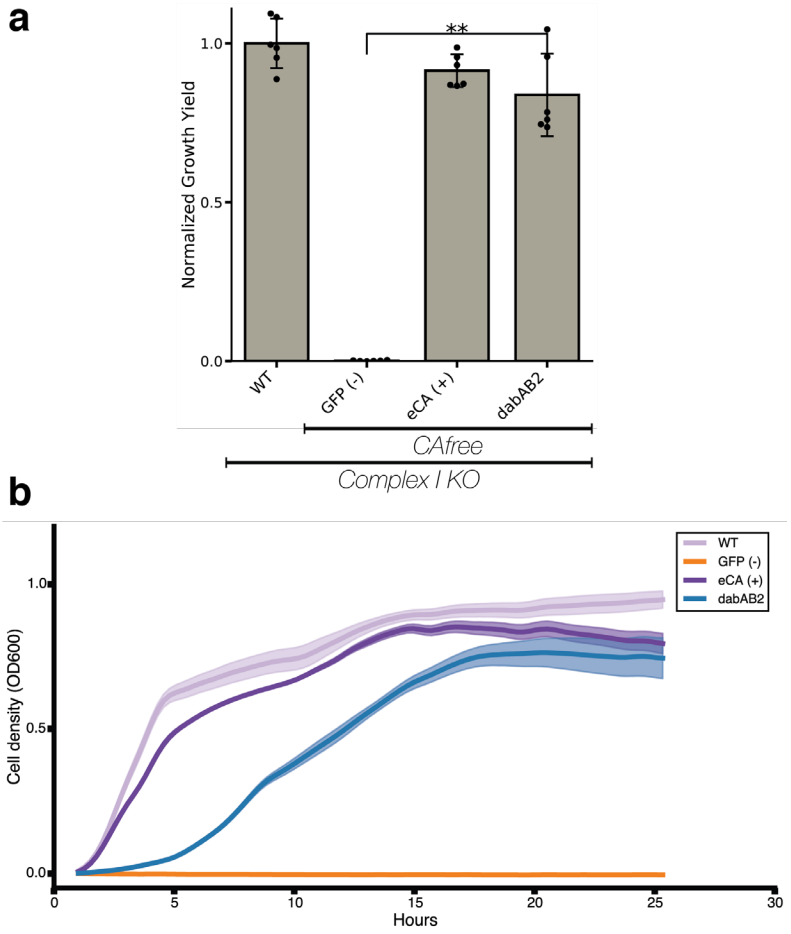


Figure 3.6 DAB2 function is not dependant on complex I. (**a**) DAB2 is still able to rescue growth of CAfree cells in the absence of complex I (Δ(*nuoA-nuoN*)). dabAB2 rescues better than GFP (t=15.7, $p=2.37*10^{-8}$). "**" denotes that means are significantly different with Bonferroni corrected $p < 5X10^{-4}$ according to a two-tailed t-test. Bar heights represent means and error bars represent standard deviation of six biologically replicate cultures. Consistent results were observed in an independent growth experiment. (**b**) These growth curves were used to generate the growth yield values in a. Mean OD600 is plotted +/- standard error of six biologically replicate cultures. Consistent results were observed in an independent growth experiment. All strains are complex I knockout strains. DAB2 is still able to rescue growth of CAfree cells in the absence of complex I.

Aqueous CO_2 spontaneously interconverts with the gas phase as well as hydrated C_i species (H_2CO_3, HCO_3^-, CO_3^{-2}). The equilibrium of $CO_2^{(aq)}$ and $CO_2^{(gas)}$ is not affected by pH, but the conversion from CO_2 to hydrated C_i is pH dependent. Thus, the equilibrium concentration of HCO_3^- increases 100 fold between pH 5 and 7 without an accompanying change in CO_2 concentration (Figure 3.7a) [105]. Expression of SbtA, a known HCO_3^- transporter, rescues CAfree growth at pH 7 but not at pH 5, while DabAB2 rescues growth at both pHs (Figure 3.8). Since DabAB2 rescue is pH-independent in this range, its substrate is likely CO_2 and not H_2CO_3, HCO_3^-, or CO_3^{-2}. This is consistent with previous observations that CO_2 is the likely substrate of *Hnea* C_i uptake [115].

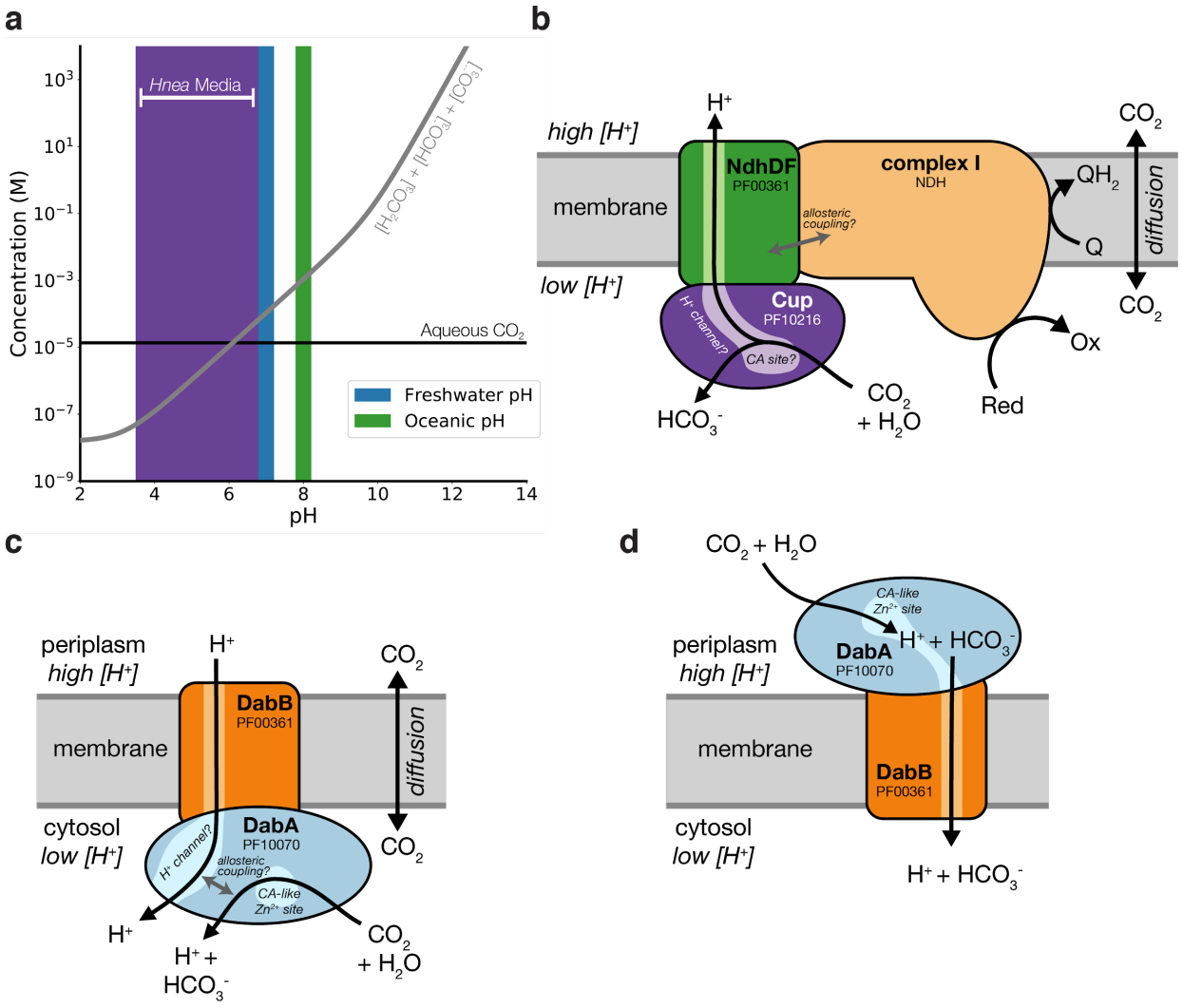


Figure 3.7 Comparison of models of vectorial CA activity for DABs and the Cyanobacterial CUPs. (**a**) Equilibrium concentrations of dissolved inorganic carbon as a function of pH. In this plot we assume the growth medium is in Henry's law equilibrium with present-day atmosphere (400 PPM CO_2) at 25 °C giving a soluble CO_2 concentration of roughly 15 µM. The equilibrium concentrations of hydrated C_i species (H_2CO_3, HCO_3^-, CO_3^{2-}) is determined by the pH. As such, the organisms will "see" a C_i species in very different ratios depending on the environmental pH. In an oceanic pH near 8, HCO_3^- dominates the C_i pool. HCO_3^- is also the dominant constituent of the Ci pool in freshwater, but less so (by a factor of ~10 since freshwater and oceanic environments differ by about 1 pH unit). In acid conditions (pH < 6.1) CO_2 will be the dominant constituent of the C_i pool. The pH of our Hnea culture media ranges from 6.8 (when freshly made) to ~3.5 when cells reach stationary phase (*Hnea* make H_2SO_4 as a product of their sulfur oxidizing metabolism). As such we expect that *Hnea* regularly experiences environments wherein it is advantageous to pump CO_2 and not HCO_3^-. (**b**) CupA/B proteins are CA-like subunits of a class of cyanobacterial Ci uptake systems. Cup-type systems are believed to couple electron transfer to vectorial CA activity and, potentially, outward-directed proton pumping. This model is based on the observation that Cup systems displace the two distal H^+-pumping subunits of the cyanobacterial complex I and replace them with related subunits that bind CupA/B (illustrated in green as NdhD/F). (**c**) As our data are consistent with DAB2 functioning as a standalone complex (i.e., DabAB do not appear to bind nor require the *E. coli* complex I), we propose a different model for DAB function where energy for unidirectional hydration of CO_2 is drawn from the movement of cations along their electrochemical gradient (right panel above). (**d**) An alternative model for DAB activity is that DabA is localized to the periplasm and DabB is functioning as a H^+ : HCO_3^- symporter. In this model DabA CA activity is made vectorial by removal of products. Energy is provided in the form of the *pmf* driving H^+ (and therefore HCO_3^-) uptake. This model is not preferred because no secretion signals were observed in the DabA sequence. Moreover, the *Acidimicrobium ferrooxidans* genome contains an apparent DabA:DabB fusion protein. The predicted architecture the fusion would place DabA in the cytoplasm.

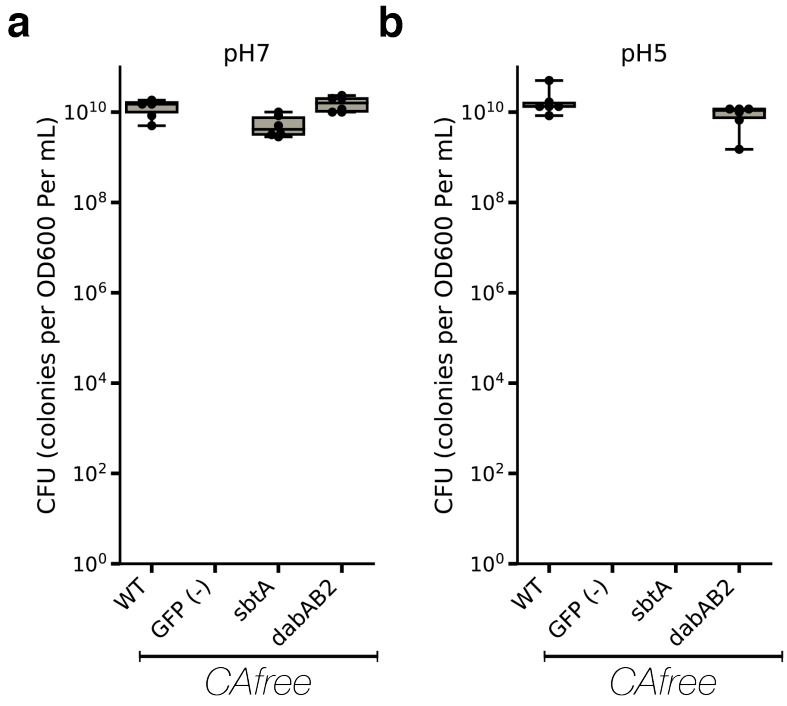


Figure 3.8 pH independence of *dabAB2* rescue of CAfree. Colony forming units per OD600 per ml were measured on LB plates with induction in air at both pH 7 (**a**) and 5 (**b**) dabAB2 rescued growth at both pH7 and pH 5, sbtA only rescued growth at pH 7. Whiskers represent the range of the data, the box represents the interquartile range, and the middle line represents the median. Data is from 6 technical replicate platings of all conditions.

Structural homology modeling software predicts that the middle of DabA2 has sequence elements related to a β-CA (Figure 3.3a). Phyre2 predictions identify C539 and H524 as part of a potential Zn^{2+} binding site distantly homologous to a bacterial type II β-CA (10% coverage of DabA, 90.8% confidence). I-TASSER predicts a Zn^{2+} binding site including the same residues along with an additional cysteine (C351), and aspartic acid (D353). As shown in Figure 3.5c, these residues could make up the active site of a type II β-CA [139]. We generated individual alanine mutants for each of these putative active site residues (C351A, D353A, H524A and C539A) and tested them in CAfree. All mutants failed to rescue CAfree growth in ambient CO_2 (Figure 3.5d). We proceeded to assay zinc binding of purified DabAB2 complex using X-ray fluorescence spectroscopy and found that wild-type DabAB2 and three of the single mutants (C351A, D353A, and H524A) bind zinc (Figure 3.5e). Single mutants retain three of four zinc-coordinating residues [139], which could explain why the mutants bind zinc. Indeed, mutational studies of the human CA II show that single mutations to Zn^{2+}-binding residues reduce but do not abrogate zinc binding [140].

The assay of detergent solubilized, purified DabAB2 did not show carbonic anhydrase activity over controls (Figure 3.5f). DabAB2 was assayed at high protein concentrations (> 650-fold more protein than the positive control) and under CO_2 concentrations that are typically saturating for CAs, but displayed no activity (Figure 3.5f). Absence of activity *in vitro* argues either that DabAB2 has extremely low CA activity or, perhaps, that DabAB2 must reside in a cell membrane holding a cation gradient to function as an energetically-activated carbonic anhydrase.

A query of the Uniprot database with the DabA PFAM (PF10070) yielded 878 putative DabA sequences. DabAs were found in a wide variety of prokaryotes including bacteria and archaea (Figures 3.8a, Appendix B.5), as is consistent with previous work [126]. Represented

clades include not only Proteobacteria, but also Euryarchaeota, Firmicutes, Planctomycetes, and Bacterioides. However, we were surprised to observe many *dabA* sequences were found in genomes of organisms that cannot fix CO_2 including the heterotrophic human pathogens *V. cholera*, *B. anthracis*, and *L. pneumophila* (Figure 3.9a). Notably, 843 (96%) of the identified *dabA* sequences were either within three genes of, or fused to, a *dabB*.

Finally, we assayed whether the DAB homologs from heterotrophic pathogens are functional C_i pumps. Operons from *V. cholera* E7946 El Tor Ogawa and *B. anthracis* Sterne were cloned and expressed in CAfree. Both DAB operons rescued CAfree growth in ambient CO_2 (Figure 3.9b and Appendix Figure B.6). Thus, DAB operons from heterotrophic, human pathogens are functional.

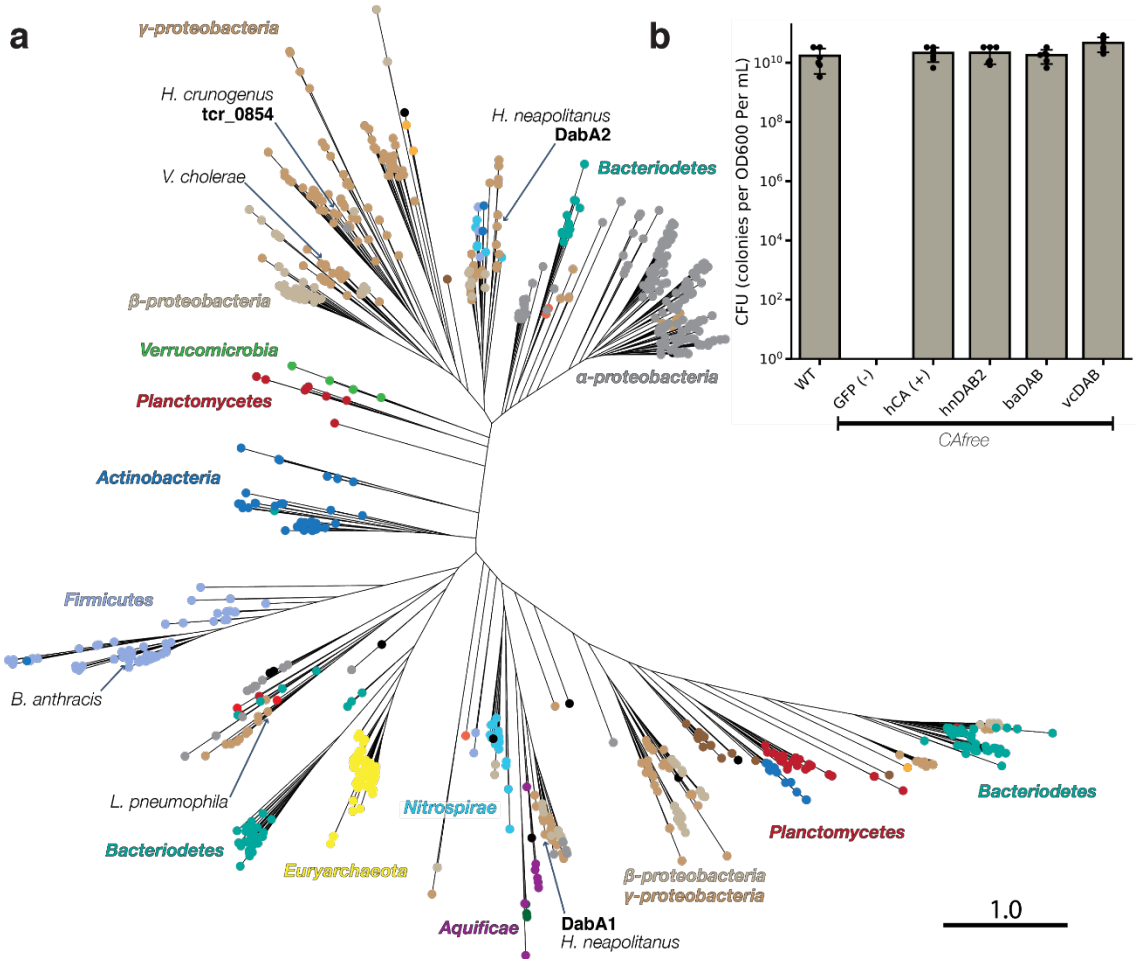


Figure 3.9 DAB operons are widespread among prokaryotes**.** (**a**) Approximate maximum likelihood phylogenetic tree of 878 DabA homologs associated with PF10070.9. DabA homologs are found in > 15 prokaryotic clades, including archaea. *Hnea* DabA1 and DabA2 represent two different groupings that are commonly found in proteobacteria. Inspecting the tree reveals several likely incidents of horizontal transfer, e.g. between Proteobacteria and Firmicutes, Nitrospirae and Actinobacteria. Moreover, the genomes of several known pathogens contain a high-confidence DabA homolog, including *B. anthracis*, *V. cholerae*, and *L. pneumophila*. Detailed annotations are given in Supplemental Figure 9. Scale bar indicates one substitution per site. (**b**) Functional DABs are found in human pathogens. Colony forming units per OD600 per ml were measured on LB plates with induction in air. DAB operons from *B. anthracis* (baDAB) and *V. cholerae* (vcDAB) rescued growth of CAfree cells. The *Hnea* operon DAB2 is abbreviated as hnDAB2. Bars represent means. Error bars represent the standard deviation of 6 technical replicate platings. Consistent results were achieved in biologically independent platings of baDAB and vcDAB.

3.4 Discussion

Here, we generated a knockout library containing ≈35 individual knockouts for every gene in the genome of the proteobacterial chemolithoautotroph *H. neapolitanus.* Using these data, we compiled the essential gene set of a chemolithoautotroph (Figure 3.1) and were able to confidently identify 551 essential genes and 1787 nonessential genes. Mapping essential genes will provide insight into the metabolism and growth physiology of sulfur-oxidizing chemolithoautotrophs.

In addition to mapping essential genes, this library would be used to measure conditional phenotypes for nonessential genes. These mutants were isolated in high CO_2 and so we were able to disrupt all known components of the bacterial CCM (Figure 3.2). The resulting genome-wide knockout library was used to perform a comprehensive screen for bacterial CCM genes. This screen highlighted a small number of genes (17) as having the HCR phenotype associated with the CCM (Figure 3.2b-f), nearly all of these genes are known to be associated with the α-carboxysome. Though it is possible that genetic redundancy, conditional phenotypes, or impairment only at sub-ambient CO_2 permit some genes to escape notice, these data suggest that the proteobacterial CCM is composed of < 30 functionally distinct components. Moreover, none of the genes identified have unexpected functions, suggesting that current models of bacterial CCMs incorporate all necessary functions.

Our screen identified 3 transcriptional regulators as well as 3 distinct CCM operons (Figures 3.2b-f). Identification of transcriptional regulators with HCR phenotypes (Figures 3.2d-f) may inform the study of CCM regulation. The first operon contains nearly all known components of the α-carboxysome, all of which confer HCR phenotypes upon disruption (Figure 3.2c). The second operon is adjacent to the carboxysome operon and contains 11 genes. Only 3 of these genes - the rubisco chaperone *acRAF* and *dabAB1* - displayed HCR phenotypes (Figure 3.2c). The remaining 8 genes had no associated phenotype but might nonetheless have roles in the CCM. These genes include *cbbOQ*, *csos1D*, *p-II*, and a *parA* homolog (Figure 3.2c). The third operon contains two genes, *dabAB2,* both with HCR phenotypes (Figure 3.2d).

A previous physiological study suggested that *Hnea* C_i uptake is coupled to the membrane electrochemical potential and uses CO_2 as a substrate, but the protein(s) responsible for this activity were unknown [115]. DAB1 and DAB2 are homologous to C_i pumps from hydrothermal vent chemolithoautotrophs recently discovered by Scott and colleagues [126,128] and our screen suggests that DAB1 and DAB2 are likely the C_i pumps in *Hnea*. These observations raise many mechanistic questions as to how DABs function, and we therefore sought to establish a biochemical system to investigate DAB structure-function.

We showed that the DAB2 operon encodes a two-component protein complex that has C_i uptake activity when heterologously expressed in *E. coli* (Figures 3.3b,c and 3.5a). This complex is likely a heterodimer as suggested by size-exclusion chromatography (Figure 3.5b). As C_i uptake is inhibited by the ionophore CCCP (Figure 3.3c), we suspect that DAB2 activity is energetically-coupled to a cation gradient. Moreover, since DabAB2 shows pH-independent rescue of CAfree *E. coli*, CO_2 is likely the transported substrate (Figure 3.5c). This is further supported by the fact that DabA has distant homology to a type II β-CA and binds a zinc (Figures 3.3 and 3.5), suggesting that a CA active site hydrates transported CO_2. Finally, mutations to the putative zinc-

binding residues (C351A, D353A, H524A, and C539A) ablate function *in vivo* (Figure 3.5d). We therefore propose a speculative model of DAB activity wherein CO_2 is passively taken into the cell and then unidirectionally hydrated to HCO_3^- by energy-coupled CA activity of DabA (Figure 3.10).

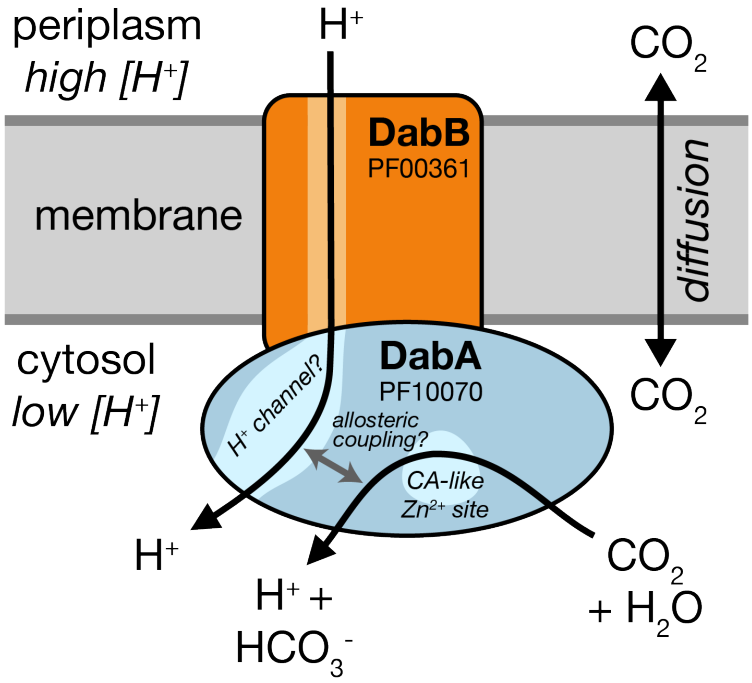


Figure 3.10 A speculative model of unidirectional energy-coupled CA activity of DAB complexes. We propose that DabAB complexes couple CA activity of DabA to a cation gradient across the cell membrane, producing unidirectional hydration of CO_2 to HCO_3^-. The cation gradient could be H+ or Na^+. Energy-coupled CA activity is required for the DABs role as a C_i uptake system in the proteobacterial CCM, as discussed in the text. Because it appears that DabAB2 is not active as a purified complex outside of the membrane, it is assumed protein tightly couples the inflow of cations with CO_2 hydration so that there is no "slippage." Indeed, slippage - i.e., uncoupled CA activity - would be counterproductive for CCM function[105,112]. Notably, Zn^{2+} binding by the active site aspartic acid of type II β-CAs (D353 in DabA2, Figure 3.5c) is thought to allosterically regulate activity[141]. This Asp-mediated activity switch could, therefore, provide a means for allosteric coupling of a β-CA active site to distal ion transport..

Model carbonic anhydrases are not coupled to any energy source (e.g., ATP, cation gradient). Rather, they equilibrate CO_2 and HCO_3^- [140]. However, energy-coupled CA activity could favor CO_2 hydration, allowing the DAB system to actively accumulate HCO_3^- in the cytosol and power the CCM (Figure 3.2a). Given the similarity of DabB to H^+-pumping proteins, we propose that DABs use the H^+ gradient, though our results are equally consistent with other cation gradients, e.g. Na^+. This mechanism would require tight coupling of cation flow to CA activity by DabA, consistent with our observation that purified DabAB2 displays no measurable CA activity. Interestingly, type II β-CAs are the only CAs known to display allosteric regulation [141]. Allosteric control is thought to be mediated by Zn^{2+} capping and uncapping by the active site aspartic acid (D353 in DabA2) [141]. A similar mechanism might couple cation flow through DabB to the active site of DabA.

Cyanobacteria possess vectorial CAs called CUPs, which may provide clues to the DAB mechanism [39,122,142]. Indeed, both DAB and CUP systems contain subunits in the Mrp protein family (DabB and NdhD/F are in PF00361) that also contains the H^+-pumping subunits of complex I. This commonality might suggest similar mechanisms. CO_2 hydration by CUPs is thought to be coupled to energetically-favorable electron flow because CUPs associate with complex I [14] (Figure 3.9b). However, the Mrp protein family (PF00361) is very diverse and contains many cation transporters that do not associate with complex I or any other redox-coupled complex [136].

Moreover, DabB sequences are only distantly related to complex I and CUP subunits (Figure 3.4a), DAB2 subunits do not co-purify with *E. coli* complex I (Figure 3.5A) and DAB2 rescues CAfree growth in a complex I knockout (Figure 3.6). We therefore propose that DAB activity is not coupled to electron flow through complex I but, rather, to a cation gradient across the membrane (Figure 3.10).

DabAB2 functions robustly, as demonstrated by complementation of CAfree (Figure 3.3B) and ^{14}C uptake measurements (Figure 3.3c). Indeed, we observed that DabAB2 functions substantially better in *E. coli* than SbtA, a C_i transporter from cyanobacteria [107,138] (Figure 3.3C). As *E. coli* and *Hnea* are proteobacteria, this observation could result from greater "compatibility" of proteobacterial proteins with *E. coli* expression. It may also be the case, though, that the α-CCM of proteobacteria is more "portable" than the β-CCM of freshwater cyanobacteria. Indeed, α-CCM genes are typically found in a single gene cluster in chemolithoautotrophs throughout α- β- and ɣ-proteobacteria and the α-CCM was clearly horizontally transferred from proteobacteria to marine cyanobacteria [107]. DabA homologs are widespread in prokaryotes and were likely horizontally transferred multiple times (Figure 3.9a). Since DAB complexes are prevalent among prokaryotes and have superlative activity, DAB-family transporters are an attractive target for protein engineering and heterologous expression in plants and industrial microbes, where elevated intracellular C_i could be useful [143].

Finally, DABs are present in a wide variety of bacteria and archaea [126]. High-confidence DabA homologs are found not only in large numbers of autotrophs but also in heterotrophs (Figures 3.9). Moreover, homologs are present in the notable heterotrophic pathogens *V. cholerae*, *B. anthracis*, and *L. pneumophila* (Figure 3.9a). We showed that DABs from *V. cholerae* and *B. anthracis* are active in *E. coli* (Figure 3.9b). This leads us to wonder: what do heterotrophic pathogens use C_i uptake systems for? Carbonic anhydrase activity is essential for growth of the heterotrophs *E. coli* and *S. cerevisiae* in ambient CO_2 [137,144]. In the heterotrophic context, CA activity is thought to supply bicarbonate for biotin-dependent carboxylases in central metabolism, for which HCO_3^- is the substrate [137,144]. Additionally, bicarbonate levels have been linked to virulence in both *V. cholera* and *B. anthracis* [145,146]. Perhaps DAB-family C_i uptake systems play roles in the growth or virulence of these important pathogens? We hope that future research will delineate the role of energetically-activated C_i uptake in heterotrophic and pathogenic organisms.

3.5 Tables

Table 3.1 *Hnea* genes from HCR operons.

Locus Id	NCBI Accession number	NCBI gi number	Gene description	Has HCR phenotype
HNEAP_RS01030	WP_012823110.1	502585319	DabA2	TRUE
HNEAP_RS01035	WP_012823111.1	502585320	DabB2	TRUE
HNEAP_RS01040	WP_012823112.1	502585321	LysR	TRUE
HNEAP_RS04565	WP_012823782.1	502586009	Csos1D	FALSE
HNEAP_RS04570	WP_012823783.1	502586011	unk.	FALSE
HNEAP_RS04575	WP_012823784.1	502586012	CbbQ	FALSE
HNEAP_RS04580	WP_012823785.1	502586013	p-II	FALSE
HNEAP_RS04585	WP_012823786.1	502586014	DabA1	TRUE
HNEAP_RS04590	WP_012823787.1	502586015	unk.	FALSE
HNEAP_RS04595	WP_012823788.1	502586016	DabB1	TRUE
HNEAP_RS04600	WP_012823789.1	502586017	CbbO	FALSE
HNEAP_RS04605	WP_041600361.1	753844744	unk.	FALSE
HNEAP_RS04610	WP_049772467.1	908628434	ParA	FALSE
HNEAP_RS04615	WP_012823792.1	502586020	acRAF	TRUE
HNEAP_RS04620	WP_012823793.1	502586021	Csos1B	TRUE
HNEAP_RS04625	WP_012823794.1	502586022	Csos1A	TRUE
HNEAP_RS04630	WP_012823795.1	502586023	Csos1C	TRUE
HNEAP_RS04635	WP_012823796.1	502586024	Csos4B	TRUE
HNEAP_RS04640	WP_012823797.1	502586025	Csos4A	TRUE
HNEAP_RS04645	WP_012823798.1	502586026	CsosCA	TRUE
HNEAP_RS04650	WP_081441107.1	1174219926	Csos2	TRUE
HNEAP_RS04655	WP_012823800.1	502586028	CbbS	TRUE
HNEAP_RS04660	WP_012823801.1	502586029	CbbL	TRUE
HNEAP_RS05490	WP_012823963.1	502586200	LysR	TRUE
HNEAP_RS07320	WP_081441122.1	1174219941	Crp/Fnr	TRUE

3.6 Materials and Methods

Bacterial strains and growth conditions

E. coli strain APA766 was used as the conjugation donor to transfer the Tn5 transposon to *Halothiobacillus neapolitanus* C2 (*Hnea*) via conjugation [129]. The *E. coli* double CA deletion strain "CAfree" (BW25113 *Δcan ΔcynT*) was generated by curing the KEIO collection *cynT* knockout (BW25113 *ΔcynT*, KEIO strain JW0330) of kanamycin resistance via pCP20-mediated FLP recombination and subsequent P1 transduction (and curing) of kanamycin resistance from the *can* knockout strain EDCM636 (MG1655 *Δcan*, Yale Coli Genomic Stock Center, [137,147]). Complex I knockout strains (Δ(*nuoA-nuoN*)) were generated in the BW25113 and CAfree backgrounds. These strains were generated by lambda red mediated recombination of a Kan^R resistance cassette flanked by FRT sites into the nuo locus such that the entire operon was removed. The pSIM5 plasmid carrying the lambda red recombinase was heat cured at 42 °C. Lysogeny broth (LB) and LB agar were used as *E. coli* growth media unless otherwise specified. *E. coli* strains were grown at 37 °C in the presence of 0.1 mg/ml carbenicillin, 0.06 mg/ml kanamycin, or 0.025 mg/ml chloramphenicol as appropriate. *Hnea* was grown in DSMZ-68 media at 30 °C and in the presence of 0.03 mg/ml kanamycin when appropriate.

Transposon mutagenesis and RB-TnSeq library production

A barcoded library of *Hnea* transposon mutants was generated by adapting the methods of Wetmore *et al*. [129]. Conjugations were performed as follows. *Hnea* and APA766 were cultured and harvested by centrifugation. Both cultures were washed once in 10 mL antibiotic-free growth media per conjugation reaction and resuspended in 100 µl. 5 OD600 units of *Hnea* were mixed with 20 OD600 units of APA766 on a 0.45 µM Millipore MCE membrane filter and cultured overnight at 30 °C in 5% CO_2 on an antibiotic-free LB agar plate containing 0.06 mg/ml diaminopimelic acid. Cells were scraped from the filter into 2 mL DSMZ-68 and collected in a 2 mL microcentrifuge tube. Recovered cells were pelleted by centrifugation at 16000 x g for 1 minute, washed in 2 mL DSMZ-68, pelleted again at 9000 x g for 1 minute, and resuspended in 2 ml DSMZ-68 before 200 µl was plated onto 10 separate DSMZ-68 kanamycin plates (per conjugation). Plates were incubated at 30 °C under 5% CO_2 until colonies formed (~ 7 days). Colonies were counted and scraped into 55 mL DSMZ-68. Two 1.4 OD600 unit samples were taken and used to prepare genomic DNA (Qiagen DNeasy blood and tissue kit). Transposon insertions were amplified from gDNA and transposons were mapped after Illumina sequencing using protocols and software from Wetmore *et al*. [129] 1.6 OD600 unit aliquots were then flash frozen in 50% glycerol for subsequent BarSeq experiments.

Essential gene assignment

Following the logic of Wetmore *et al.* and Rubin *et al*. [129,148], we categorized genes as essential if we observed significantly fewer transposon insertions than would be expected by chance. If insertion occurred uniformly at random, the number of insertions per gene would be expected to follow a binomial distribution. The probability of observing at most *k* insertions into a gene of length *n* is therefore expressed as:

$$P(k;n,p) = \sum_{i=0}^{k} \frac{n!}{k!\,(n-k)!} p^i (1-p)^{n-i}$$

Here, p is the average rate of transposon insertion per base pair genome-wide. Genes were determined to be essential if they received a lower-than-expected number of insertions in both replicates of the library mapping, i.e. if the probability of observing k or fewer insertions was beneath 0.05 after Bonferroni correction. Genes were called "ambiguously essential" in two cases: (i) replicates were discordant or (ii) zero insertions were observed but the gene was short enough that the formula could not yield a Bonferroni-corrected p-value below a 0.05 threshold even in the case of zero insertions.

Gene fitness experiments

Fitness experiments were performed according to a modification of the protocol in Wetmore *et al*. [129]. This method allows pooled library fitness experiments to be performed comparing different growth conditions by comparing barcode abundance changes in order to track changes in the abundance of the transposon mutants. In short, a library aliquot was thawed and used to inoculate three 33 mL cultures. Cultures were grown to OD600 ~0.08 in 5% CO_2. At this point, 20 mL were removed and harvested by centrifugation as two t_0 (input) samples. Cultures were back-diluted 1:64 into 128 mL and incubated for 6.5-7.5 doublings under 5% CO_2 or ambient conditions. 50 mL of culture was harvested by centrifugation. gDNA was prepared and barcodes were amplified for fitness determination via Illumina sequencing as described previously [129]. Fitness values were calculated using existing software [129]. Genes were assigned an HCR phenotype if they had a fitness defect of two-fold or greater in ambient CO_2 compared to 5% CO_2 in two replicate experiments.

CAfree rescue experiments

Electrocompetent CAfree cells were prepared using standard protocols [149] and transformed with pFE plasmids expressing genes of interest by electroporation. CAfree pre-cultures were grown overnight in 10% CO_2 and diluted into 96 well plates (3 µl cells in 250 µl media). Growth curves were measured by culturing cells in a Tecan M1000 microplate reader under ambient conditions with continuous shaking, and measuring OD600 every 15 minutes. When samples are marked "induced," 200 nM anhydrotetracycline (aTc) was added to the media. Growth yields are calculated as the maximum OD600 achieved after 24 hours of growth and normalized to the yield of a wild type control. CFU experiments were performed by back diluting cultures to OD600 0.2 before performing 10X serial dilutions. 3 µl of the OD600 0.2 sample and each of the serial dilutions were then spotted on plates with 200 nM aTc and grown overnight in ambient conditions (400 ppm CO_2). The spot with the highest dilution that yielded more than one colony was counted and a minimum of six replicates were averaged for each strain.

Silicone oil centrifugation measurement of C_i uptake

The silicone oil filtration method was modified from Dobrinski *et al*. [150] and used to measure uptake of radiolabeled inorganic carbon. Assay tubes were generated using 0.6 ml microcentrifuge tubes containing 20 µl of dense kill solution (66.7% v/v 1 M glycine pH 10, 33.3%

v/v triton X-100) covered by 260 µl of silicone oil (4 parts AR20:3.5 parts AR200). Electrocompetent CAfree cells were prepared using standard protocols and transformed with pFA-based plasmids containing genes of interest by electroporation. CAfree cultures were grown overnight in 10% CO_2, back diluted to an OD600 of 0.1 and allowed to grow to mid-log phase in 10% CO_2 in the presence of 200 nM aTc for induction. Cells were then harvested by centrifugation, washed once in PBS (pH 7.55) and resuspended to OD600 0.6 in PBS + 0.4% glucose. ^{14}C-labeled sodium bicarbonate (PerkinElmer) was added to a final concentration of 4.1 nM and an activity of 0.23 µC_i. Cells were incubated with ^{14}C for 4 minutes before centrifugation at 17,000 x g for 4 minutes to separate cells from buffer. Pellets were clipped into scintillation vials containing 5 ml Ultima Gold scintillation fluid and 300 µl 3M NaOH using microcentrifuge tube clippers or medium dog toenail clippers. Counts were measured on a PerkinElmer scintillation counter. ^{14}C counts are normalized to 1 OD600 unit of cells added. During inhibition assays, cells were incubated in PBS pH 7.55 with 0.4% glucose + 0.4% DMSO and the inhibitor (100 µM CCCP) for 10 minutes before assay.

Generation of DabA phylogenetic tree

We searched the Uniprot reference proteome database using the Pfam Hidden Markov Model PF10070.9 with a cutoff e-value of 10^{-4}. Our search recovered 941 candidate DabA proteins. These sequences were aligned using MAFFT and manually pruned to remove fragments and poorly aligning sequences. The remaining 878 candidate DabA sequences were re-aligned with MAFFT and an approximate maximum likelihood phylogenetic tree was constructed using FastTree. Taxonomy was assigned to nodes in the tree based on the NCBI taxonomy information for the genomes harboring each sequence. Genomic neighborhoods for each gene in the tree were determined using the EFIGNT online server [151] and genomes with a *dabB* gene within 3 genes of *dabA* and oriented in the same direction were considered to have full DAB operons. *dabAB* fusions were found by visual inspection of genomic neighborhoods from those genomes that did not have separate *dabB* genes located close to *dabA*.

Generation of DabB phylogenetic tree

DabB homologs were collected manually by searching MicrobesOnline for close homologs of four PF00361 members in the Hnea genome (*dabB1*, *dabB2*, *Hneap_1953*, *Hneap_1130*) and other characterized PF00361 members including Synechococcus *elongatus ndhF1*, Synechococcus *elongatus ndhF3*, and Synechococcus *elongatus ndhF4*. Genes were clustered to 95% similarity and genes with divergent operon structure were removed manually using MicrobesOnline treeview [152]. *nuoL* from *Escherichia coli*, *nqo12* from *Thermus thermophilus*, and *ndhF1/3/4* from *Thermosynechococcus elongatus* BP-1 were added as markers. ClustalW was used to construct a multiple sequence alignment and an approximate maximum likelihood phylogenetic tree was constructed using FastTree [153,154]. The tree was visualized using the Interactive Tree of Life [155].

Protein annotation and structural homology modeling

Secondary structural annotations for DabA and DabB were generated using XtalPred [156]. Structural Homology modeling of DabA was performed using Phyre2 and I-TASSER web

servers with default parameters [88,157]. A list of close DabB homologs was assembled by searching MicrobesOnline for PF00361 members with similar operon structure. A ClustalW alignment was used to calculate residue-level conservation of DabB proteins while the MAFFT alignment generated during the creation of the DabA tree was used to calculate residue level conservation of DabA proteins (Figure 3.4b).

Purification of DAB2

Chemically competent BL21-AI *E. coli* were transformed with a pET14b-based vector containing the *dabAB* genes. 1 liter of 2xYT media was inoculated with 20 ml of an overnight culture of BL21-AI E. coli in LB+CARB and allowed to grow to mid log at 37 °C. When midlog was reached, cells were induced with 20 ml of 50 mg/ml arabinose and transitioned to 20 °C for overnight growth. Cultures were pelleted and resuspended in 10 ml TBS (50 mM Tris, 150 mM NaCl, pH 7.5) supplemented with 1.2 mM phenylmethylsulfonyl fluoride, 0.075 mg/ml lysozyme and 0.8 ug/ml DNAse I per liter of starting culture and then incubated at room temperature on a rocker for 20 minutes. Cells were lysed with four passes through a homogenizer (Avestin). Lysate was clarified at 15,000 x g for 30 minutes. Membranes were pelleted at 140,000 x g for 90 minutes. Membrane pellets were resuspended overnight in 25 ml TBS supplemented with 1 mM phenylmethylsulfonyl fluoride and 1% β-dodecyl-maltoside (DDM, Anatrace) per liter of culture following [158]. Membranes were then re-pelleted at 140,000 - 200,000 x g for 60 minutes and the supernatant was incubated with Ni-NTA beads (Thermo Fisher) for 90 min at 4 °C. The resin was washed with "Ni buffer" (20 mM Tris + 300 mM NaCl + 0.03% DDM, pH 7.5) supplemented with 30 mM imidazole and eluted with Ni buffer supplemented with 300 mM imidazole. Eluent was then incubated with Strep-Tactin (Millipore) resin for 90 min at 4 °C. The resin was washed with "strep buffer" (TBS + 0.03% DDM) and eluted with strep buffer supplemented with 2.5 mM desthiobiotin. Eluent was concentrated using Vivaspin 6 100 kDa spin concentrators and buffer exchanged into strep buffer by either spin concentration or using Econo-Pac 10DG (Biorad) desalting columns. For analytical purposes, 300 µg of strep-purified protein was injected onto a Superdex 200 Increase 3.2/300 size-exclusion column pre-equilibrated in strep buffer and eluted isocratically in the same buffer.

Carbonic anhydrase assays

CA-catalyzed CO_2 hydration of purified DAB2 complex and human carbonic anhydrase (hCA) was measured using the buffer/indicator assay of Khalifah [159] on a KinTek AutoSF-120 stopped-flow spectrophotometer at 25 °C. The buffer/indicator pair used was TAPS/*m*-cresol purple measured at a wavelength of 578 nm using a pathlength of 0.5 cm. Final buffer concentration after mixing was 50 mM TAPS, pH 8.0 with the ionic strength adjusted to 50 mM with Na_2SO_4, and 50 µM of pH-indicator. Final protein concentration used was: 9.8 µM DAB2 (His-elution) and 0.015 µM hCA (positive control; Sigma Aldrich C6624). Saturated solution of CO_2 (32.9 mM) was prepared by bubbling CO_2 gas into milli-Q water at 25 °C. The saturated solution was injected into the stopped-flow using a gas-tight Hamilton syringe, and measurements were performed in a final CO_2 concentration of 16.5 mM. Progression curves were measured in 7 replicates.

X-ray fluorescence spectroscopy for metal analysis

50-100 µg of protein in 20-200 µl of TBS + 0.03% DDM was precipitated by addition of 4 volumes of acetone and incubation at -20 °C for 1 hour. Samples were centrifuged at 21,130 x g for 15 minutes in a benchtop centrifuge and the supernatant was removed. Pellets were stored at 4 °C until analysis. Fluorescence analysis was performed by breaking up the pellet into 5 µl of TBS + 0.03% DDM with a pipette tip. Small pieces of the pellet were looped with a nylon loop and flash frozen at the beamline under a nitrogen stream. The sample was excited with a 14 keV X-ray beam and a fluorescence spectrum was collected. Sample emission spectra were then used to identify metals. Metal analysis was performed on wild-type DAB2, Zn-binding mutants C351A, D353A, and H524A, bovine CA (positive control; Sigma Aldrich C7025), and a buffer blank was used as a negative control. A rubisco crystal containing cobalt salts was also used as a zinc-free control. Experiments were performed at the Lawrence Berkeley National Laboratory Advanced Light Source Beamline 8.3.1.

3.7 Accession numbers

All illumina sequencing data is accessible as an NCBI SRA BioProject accession number PRJNA546024.

Chapter 4

Structures of synthetic beta-carboxysome shells

†The work presented in this chapter is adapted from the previously published article:
Sutter M[#], Laughlin TG[#], Cai F, Sloan N, Serwas D, Davies KM, Kerfeld CA. (2019) Structure of a synthetic beta-carboxysome shell *Plant Phys*. 181 (3) 1050-1058 [#]denotes equal contribution.

4.1 Abstract

Carboxysomes are capsid-like, CO_2-fixing organelles found in all cyanobacteria and some chemoautotrophs and significantly contribute to global primary production. They are composed of a selectively permeable protein shell that encapsulates the principal CO_2-fixing enzyme rubisco, and carbonic anhydrase. As the centerpiece of the CO_2 concentrating mechanism, the co-localization of enzymes enables generation of a locally elevated concentration of substrate CO_2 around RuBisCO, while the shell acts as a barrier to the escape of CO_2, collectively enhancing catalysis. A functional carboxysome consisting of an intact shell and cargo are essential for growth of Cyanobacteria under ambient CO_2 concentrations. Using cryo-electron microscopy we have determined the structure of a recombinantly produced simplified beta-carboxysome shell. The structure reveals the sidedness and the specific interactions between the carboxysome shell proteins. The model provides insights into the structural basis of selective permeability of the carboxysome shell and provides a foundation to design modifications that probe modes of cargo encapsulation and physiochemical properties such as permeability for this important cellular organelle. Notably, the permeability properties are of great interest for modeling and evaluation of this carbon concentrating mechanism in metabolic engineering. Moreover, we find a striking similarity to the structurally characterized evolutionarily distant metabolosome shell, implying universal architectural principles for bacterial microcompartment shells

4.2 Introduction

Carboxysomes are bacterial microcompartments (BMCs) that encapsulate rubisco and carbonic anhydrase (CA) in a selectively permeable protein shell. They are found in some chemoautotrophs and all cyanobacteria and are an important part of global carbon fixation [160]. Bacteria that contain carboxysomes actively accumulate cytosolic HCO_3^- which diffuses across the carboxysome shell, where the CA converts it to CO_2, the substrate of rubisco to enhance carbon dioxide fixation (for an overview see [161,162]). The carboxysome was the first BMC identified, evident as ~200 nm polyhedral particles based on electron microscopy [163] and subsequent isolation and biochemical characterization [124,164]. More recently the availability of genomic sequence data and bioinformatic analysis has shown that compartmentalization of metabolism in BMCs is widespread across the Bacterial Kingdom [120]. While the encapsulated enzymes differ widely, the basic building blocks of the shell are conserved. Two different folds compose the BMC shell architecture, the mixed α-helix/β-sheet PFAM00936 domain that forms regular hexagons consisting of either six BMC-H chains [165] or three BMC-T chains (each BMC-T chain contains two PFAM00936 domains) [166–169] and the all β-sheet PFAM03319 domain that forms pentagonal shaped truncated pyramids consisting of five BMC-P chains [170]. Structural characterization of shells made of proteins of an uncharacterized catabolic BMC from the Myxobacterium *Haliangium ochraceum* (HO shell) shows that proteins from these two folds can form icosahedral structures with the BMC-H and BMC-T on the facets and the BMC-P located at the vertices [171].

There are two types of carboxysomes (alpha- and beta-) that perform the same biological function but differ in composition, assembly and evolution [172]. Our model shell is of the beta type and the prefix of the protein names “Ccm” originates from “CO_2 concentrating mechanism”, starting at CcmK because the extent of the operon was unknown at the time [173]. The conserved

carboxysomal genomic locus generally consist of the shell proteins CcmK1 and CcmK2 (BMC-H), CcmL (BMC-P), CcmO (a BMC-T of unknown structure) as well as the lumenal proteins CcmM and CcmN. There are additional proteins in satellite chromosomal locations such as the BMC-T type CcmP and the BMC-H proteins CcmK3 and CcmK4 [174]. A recent study shows in detail how rubisco interacts with parts of the CA CcmM to form networks [175] around which the beta-carboxysome shell is thought to assemble [176]. The alpha-carboxysome shell has been shown to provide a barrier to CO_2 leakage [131], contributing to the carbon concentrating properties of carboxysomes.

We have previously showed formation of shells from the beta-carboxysome proteins CcmK1, CcmK2, CcmL and CcmO from *Halothece* sp. PCC 7418 and demonstrated their potential for engineering purposes [177]. Here we optimized purification of this particle and determined three- dimensional structures composed of CcmK1/2 and CcmL with cryo-electron microscopy (cryo-EM) single particle analysis. The shell proteins self-assemble predominantly into two families of shells with pseudo-T=3 and pseudo-T=4 icosahedral symmetry with diameters of 210 and 245 Å, respectively. The resolution is sufficient to identify the specific amino acid interactions among the various shell proteins which are likely the same as in native carboxysomes. The complete structural description of this synthetic beta-carboxysome shell provides the foundational system for design of cargo loading strategies and mutational studies that will enable direct measurement of carboxysome shell permeability to gases and metabolites.

4.3 Results

A synthetic beta-carboxysome shell operon encoding BMC-H proteins CcmK1 and CcmK2, BMC-T type CcmO and C-terminally Strep-tagged BMC-P/CcmL was recombinantly expressed in *E. coli* and self-assembled shells were purified by affinity and anion-exchange chromatography. The composition of the purified sample was assessed by SDS-PAGE and negative-stain transmission electron microscopy prior to vitrification for cryo-EM imaging (Figure 4.1).

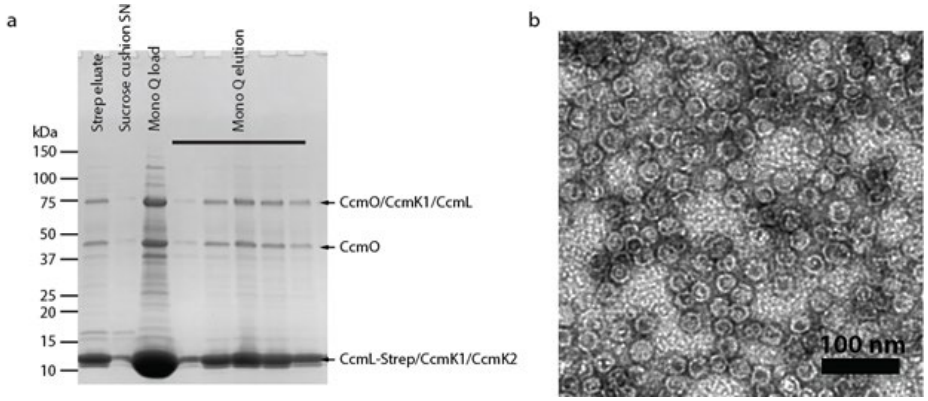


Figure 4.1 Purification of synthetic carboxysome shells. (**a**) SDS-PAGE of the sample at different steps of the purification. The top band below 75 kDa corresponds to a mixture of CcmO/K1/L as identified in (Cai, Bernstein et al. 2016). (**b**) Negative stain TEM of the final purified sample. Scale bar = 100 nm.

Cryo-EM imaging and subsequent single particle analysis revealed that multiple classes of shell assemblies were present in the sample (Figures 4.2-4.4). We obtained three-dimensional reconstructions of three classes (Figures 4.2b-e, 4.3). The most prominent and highest resolution class (75% picked particles, 2.6 Å) was a pseudo-T=4 icosahedral shell (designated pseudo because not all subunits are identical), with a diameter of approximately 245 Å (Figure 4.2b,c). A

smaller, pseudo-T=3 icosahedral shell with a diameter of 210 Å is also found in the sample at a lower occurrence of 3.4 % but we were still able to obtain good resolution (3.0 Å) due to the high inherent particle symmetry (Figure 4.2b,d). An elongated shell (Figure 4.2b,e), similar to prolate capsids of some bacteriophages [178], is also found at low occurrence (4.5%). It has dimensions of 245 Å x 310 Å and we were only able to obtain a low resolution structure ($FSC_{0.143}$ = 4.1 Å), likely due to the lower symmetry and possibly structural variability. There were also some larger shells (>300 Å diameter) that frequently appeared filled (Figure 4.3), indicating that cargo encapsulation (albeit non-specific, because the shell was not produced in the native organism) might favor larger shells that are closer to the full carboxysome size. However, density can also be seen in pseudo T=3 and T=4 shells occasionally (Figures 4.2a, 4.3a,b), indicating that *E.coli* cytosolic cargo is commonly packaged.

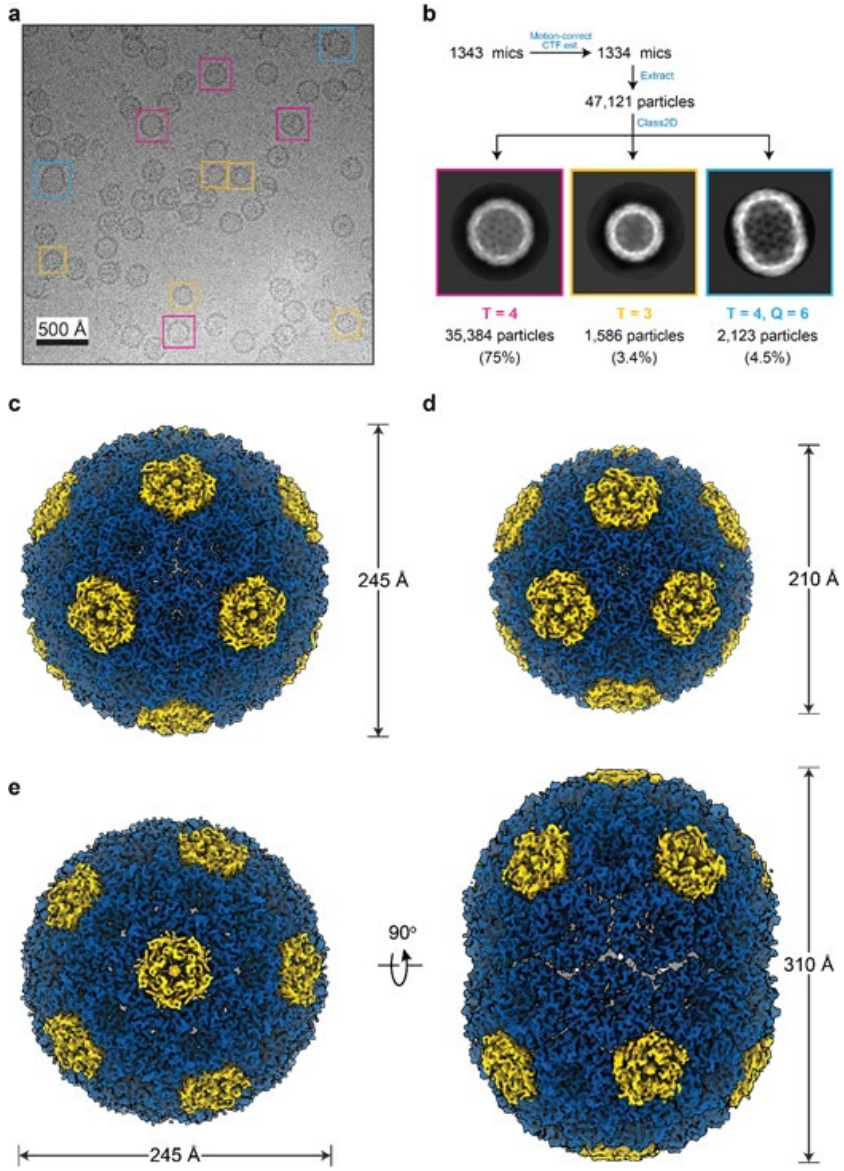


Figure 4.2 Cryo-EM reconstruction of synthetic shells. (**a**) Example cryo-EM micrograph used in reconstruction with example particles of different shell types are boxed (T=4 in magenta, T=3 in yellow, and T=4, Q=6 "prolate" in cyan). All particles were used for reconstruction but only selected ones are highlighted for clarity. (**b**) Single particle analysis pre-processing workflow with example 2D class-averages for the different shell types. Particle counts and percent of input particles for each class are indicated. (**c-e**) 3D reconstructions for the T=4 (**c**), T=3 (**d**) and prolate (**e**) shell types with dimensions indicated**.**

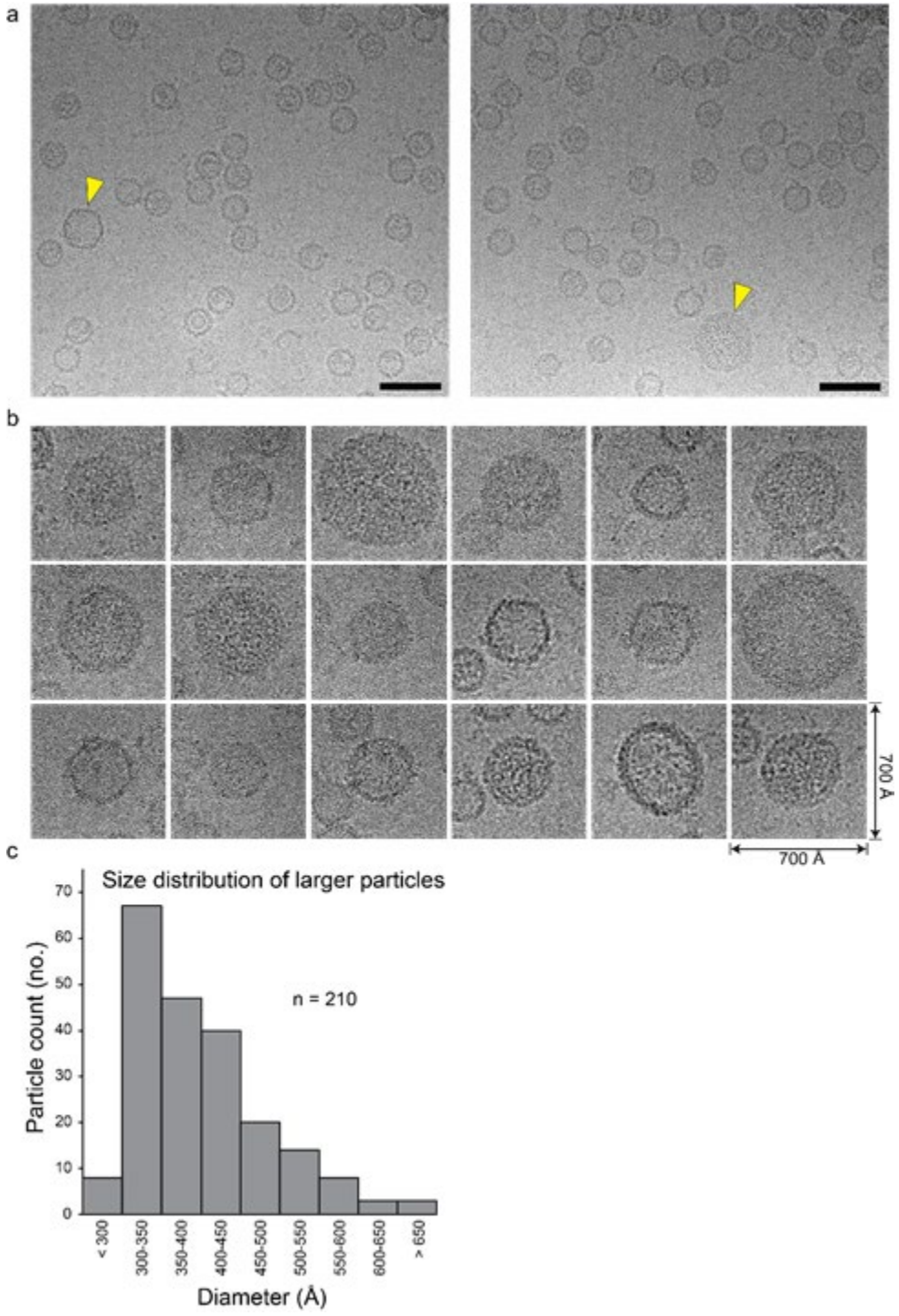


Figure 4.4 Larger shell particles observed in the cryo-EM data. (**a**) Two examples of micrographs containing putative larger shells or aggregates indicated by yellow arrows. Scale bars correspond to 500 Å. (**b**) Subset of larger particles (total n=210) extracted in a 700 Å box shows the heterogeneity among these particles. (**c**) Histogram of particle counts for the larger particles, grouped by particle diameter in Å in bins of 50 Å.

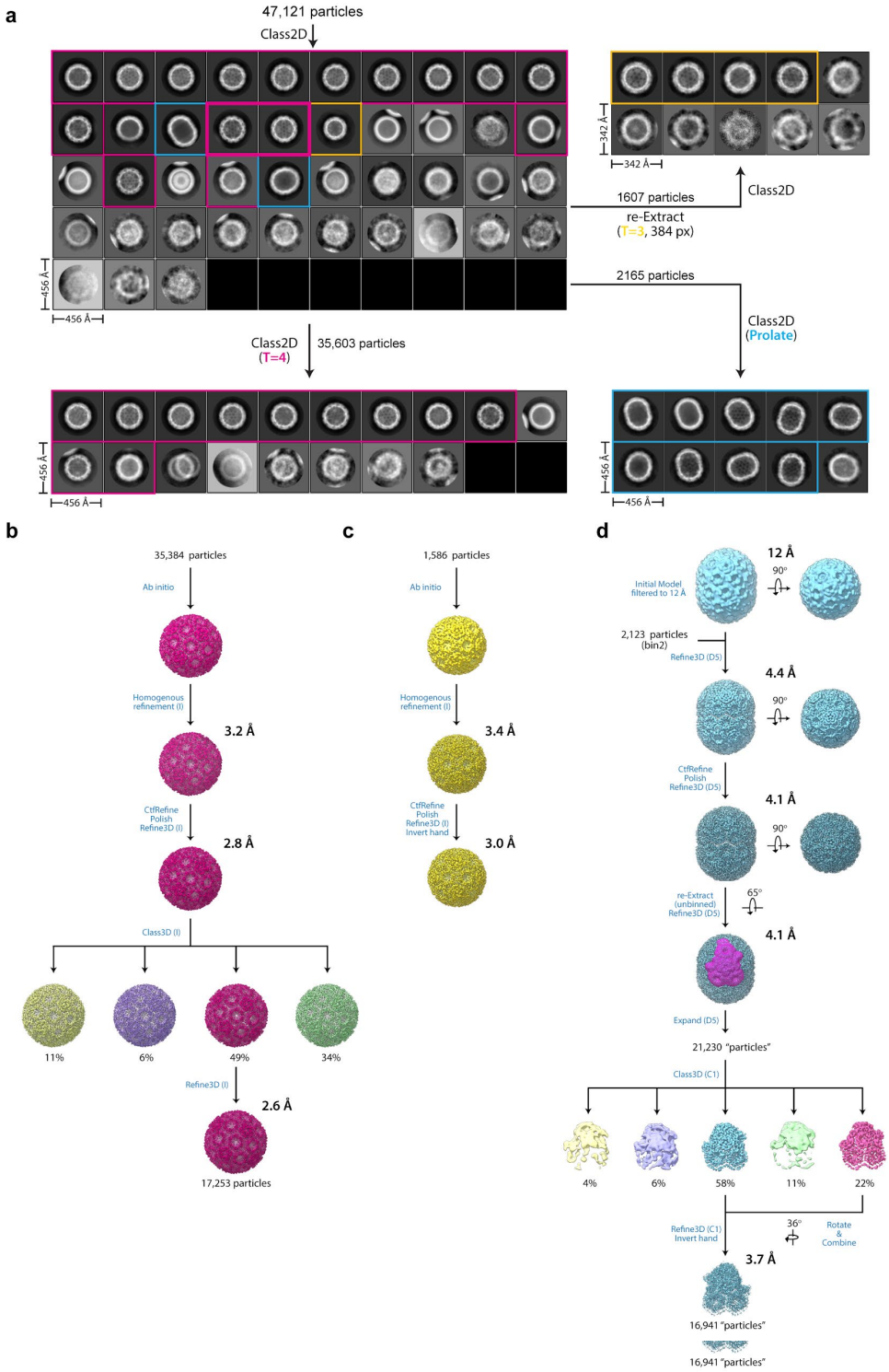


Figure 4.4 Cryo-EM reconstruction workflow of synthetic carboxysome shells. (**a**) Hierarchical 2D-classification scheme used to identify and segregate the difference major shell types prior to 3D-reconstruction. The different major shell types (T=3,4 and prolate) were discerned after the initial round of 2D-classification. At this point, the shell types were segregated and each subject to an additional round of 2D-classification to check for broken particles and misassignment of shell type prior to 3D-reconstruction. Structural 3D-overview of the T=4 (**b**), T=3 (**c**) and prolate shells (**d**). The indicated resolutions are for mask-corrected $FSC_{0.143}$ between two independently refined half-maps The mask used for the C1 reconstruction of the symmetry expanded prolate shell is shown in magenta in **d**.

Atomic modelling into the density was facilitated by available high-resolution crystal structures of homologous carboxysome shell protomers [165,170]. The T=3 and T=4 icosahedral shells reconstructions were of exceptional quality and atomic models could be readily built and refined into the density (Table 4.1, Figure 4.5). For the prolate shell, shell proteins could be reliably docked but not refined because of the poor quality of the density. In all cases, the shell reconstructions are composed entirely of the BMC-H protomers CcmK1/CcmK2 and the BMC-P protomers CcmL. The BMC-H paralogs CcmK1 and CcmK2 are 92% identical over 101 aligned residues (CcmK1 has a 12 residue C-terminal extension which is not visible in the density) and could not be distinguished in the T=4 reconstruction at 2.6 Å resolution. However, ambiguous density at the side chain positions where the two proteins differ possibly indicates that the two proteins are used interchangeably in those shells (Figure 4.6). For simplicity, all BMC-H protomers are modelled as CcmK1 (hereafter, CcmK refers to either unless specified). Density corresponding to the BMC-T protomer CcmO was not observed in any of the 3D reconstructions, indicating that it is not part of the shell structures we observed. It is possible that CcmO is binding to particles on the inside or outside or it could be a structural part of particles with low occurrence (e.g., some of the larger particles shown in Figure 4.3).

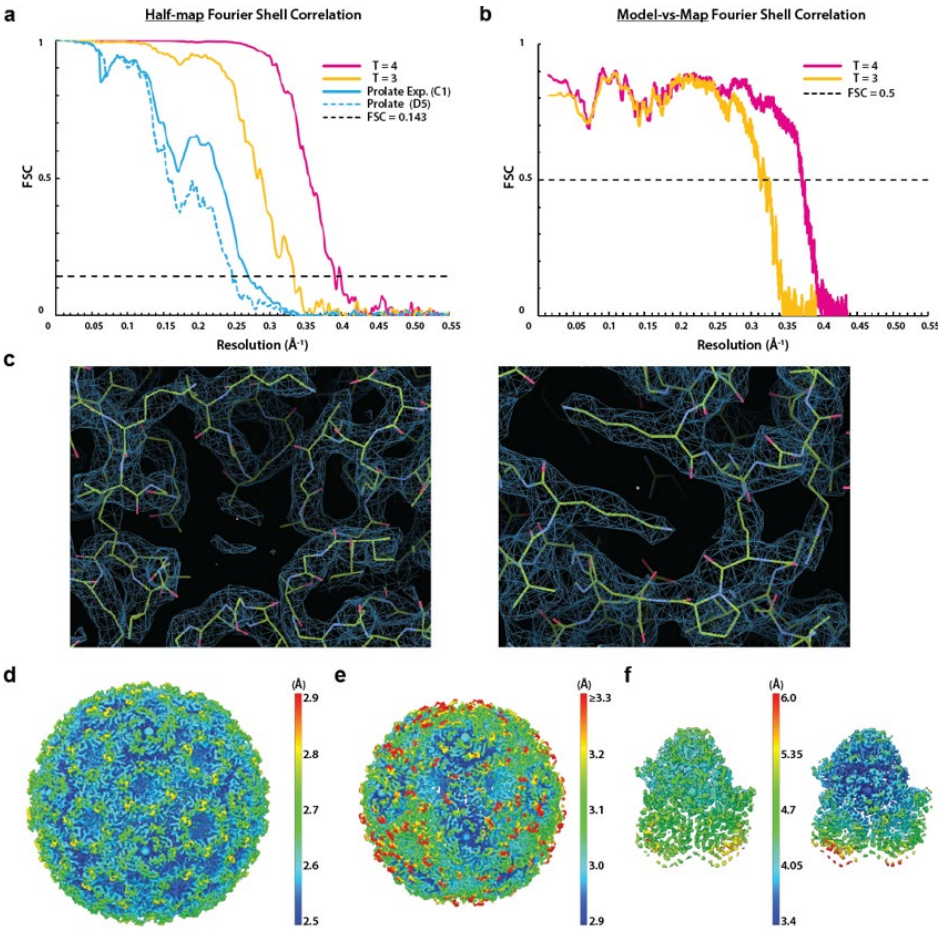


Figure 4.5 Synthetic shell cryo-EM maps and models quality. (**a**) Half-map Fourier shell correlation (FSC) curves for reconstructions, computed with RELION. (**b**) Model-vs.-map FSC curves computed with PHENIX. (**c**) Example fits-to-density of for the T=4 map with the model shown as sticks and EM density shown as mesh at 2rmsd (left: pore region of CcmK, right: antiparallel lysines at the CcmK-CcmK interface). (**d-f**) Local resolution maps computed with RELION for the T=4 (**d**), T=3 (**e**) and prolate (**f**) for the D5-symmetrized (left) and expanded C1 (right) reconstructions.

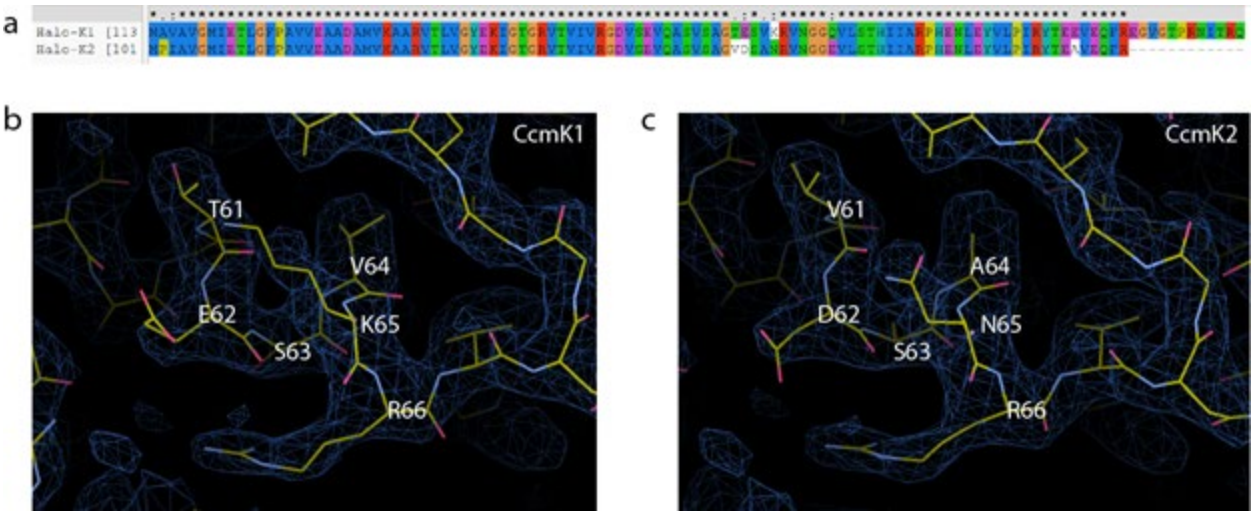


Figure 4.6 CcmK1 vs CcmK2 comparison. (**a**) Sequence alignment of CcmK1 and CcmK2, coloring by residue properties. (**b**) Density of the T=4 shell showing residues 61-66 with the CcmK1 sequence. E62 and K65 do not have full side chain density but V64 fits well. (**c**) same as (b) but for the CcmK2 sequence; D62 and N65 fit well but there is more density than can be explained with the alanine side chain of A64.

The T=3 shell icosahedral asymmetric unit consists of one CcmL protomer and two CcmK protomers with the complete shell composed of 20 CcmK hexamers and 12 CcmL pentamers at the vertices (Figure 4.7a). The T=4 shell icosahedral asymmetric unit contains an additional CcmK protomer relative to the T=3 shell for a total of 30 CcmK hexamers and 12 CcmL pentamers (Figure 4.7b). In the prolate capsid the asymmetric unit consists of 6 CcmL protomers and 24 CcmK protomers with the complete shell composed of 40 CcmK hexamers and 12 CcmL pentamers.

In all of the shells, the orientation of the subunits was consistent. For the CcmL pentamers, which are shaped like truncated pyramids, the broader side (pyramid base) and C-termini face outward. The CcmK proteins have a distinct concave and convex surfaces. In all shells the concave side, which contains both termini, faces outward (Figure 4.7). The sidedness of the carboxysome shell proteins is therefore consistent with the previously characterized shell of a metabolosome [171,179].

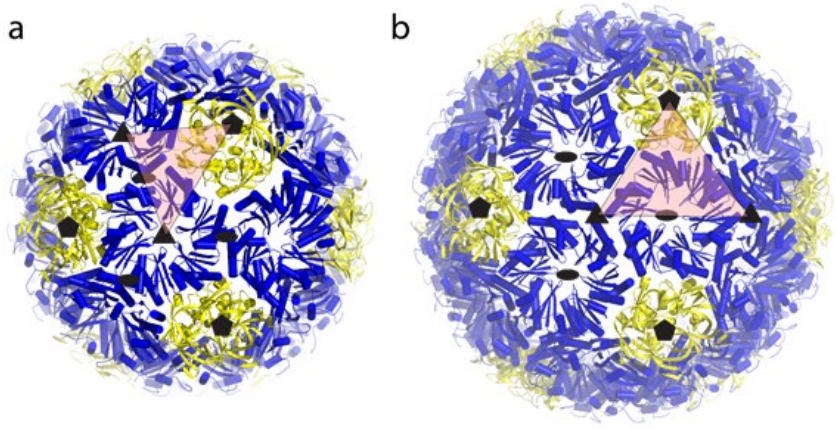


Figure 4.7 Geometrical description of the regular icosahedral shell types. (**a**,**b**) Close-up of cartoon model for each shell type with the asymmetric unit indicated by a translucent red triangle. Subunits corresponding to the T=3 and T=4 lattices are indicated by solid lines for the icosahedral shells in **a** and **b**, respectively.

The interfaces between the individual proteins for the T=3 and the T=4 shell are almost perfectly identical; we show only the T=4 interfaces for clarity. The residues at the interfaces between the BMC-H:BMC-H and BMC-P:BMC-H are highly conserved across different microcompartment types [171]. For the CcmK-CcmK interface we can recognize the absolutely conserved K25 making an antiparallel interaction with its counterpart (Figure 4.8a). Another highly conserved interaction is R80 of the (A/P)-R-P-H motif inserting into a pocket formed by H82 and backbone oxygens of the neighboring CcmK (Figure 4.8a). This CcmK-CcmK interface is at an angle of 30° and, as observed previously for the HO shell, the same interface can probably also adopt a planar configuration where the R80 is extending all the way into the pocket. An interaction that seems to be specific to beta-carboxysomes is the hydrogen bonds formed between the R28 sidechain to the sidechain of E52, its own chain, as well as to the backbone oxygen of A79 and A2 across the interface (Figure 4.8a). Looking at the CcmK-CcmK interface from the interior of the shell we notice additional intra-chain interactions between residues E62 and K65 as well as R66 and E18 (Figure 4.8b); the E62 and R66 residues are positioned such that they could possibly interact across the interface in a planar configuration. Interestingly, this is a position where *Halothece* CcmK1 and CcmK2 differ (E62/K65 in CcmK1, D62/N65 in CcmK2) which could cause a preference of one shell subunit over another for angled/planar interfaces. This residue pattern is not consistent though for all beta-carboxysomal CcmK1/2; there is no distinguishing feature that separates CcmK1 from CcmK2 other than the C-terminal extension [174].

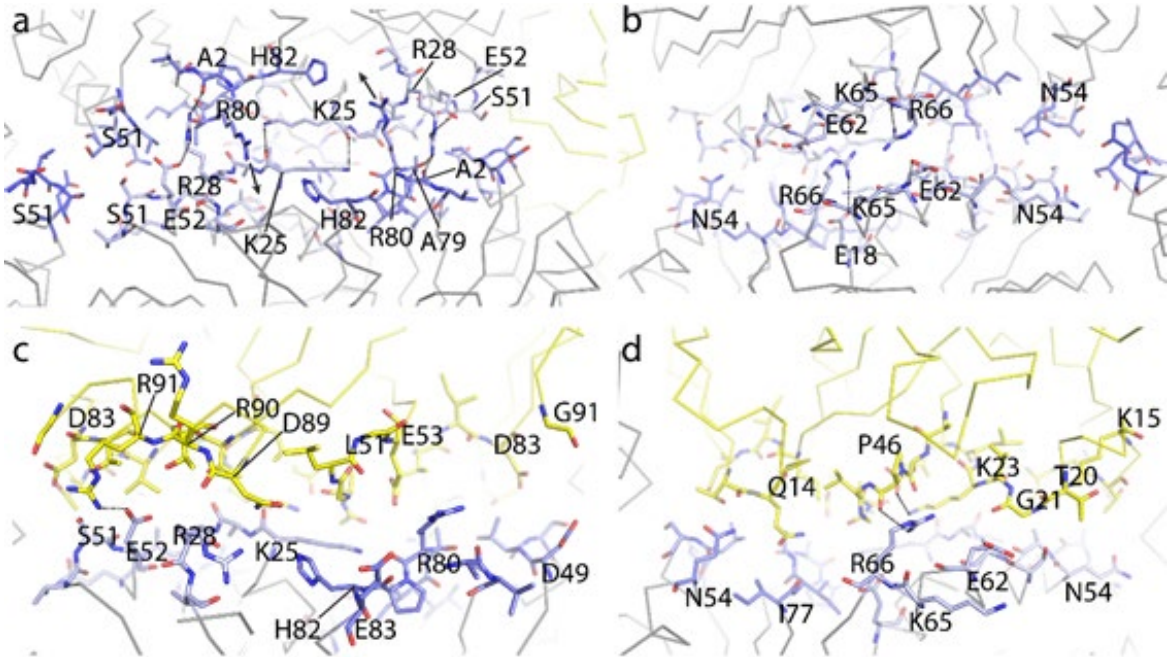


Figure 4.8 Shell subunit interfaces. (**a**) View onto the CcmK-CcmK interface from the outside and (**b**) from the inside. (**c**) View of the CcmK-CcmL interface from the inside and (**d**) outside. Interface residues are shown as sticks, CcmK in shades of blue and CcmL in yellow. Important residues are labeled and dashed lines indicate specific interactions.

The composition of our T=4 shell enables visualization for the first time, in the context of a complete BMC shell, the interface where three BMC-H proteins meet; in the HO shell a BMC-T is always one of the three oligomers at this interface. The small side chain of residue S51 seems

to prevent steric clashes (Figure 4.8a) and explains why the residues at this position are generally glycine, alanine or serine.

The interface between CcmL and CcmK seems to rely mainly on shape complementarity (Figure 4.8c,d). There is a notable interaction between K23 of CcmL and the backbone oxygen of P46 (Figure 4.8d) which, similar to the antiparallel lysines of CcmK, creates a flat surface. There are a few charged interactions, including CcmK's E52 side chain and CcmL's R91 as well as CcmK's R66 side chain to the backbone oxygen of P46.

The CcmK proteins are the majority building blocks of the facets of the beta-carboxysome shell. Protein-protein interactions in the facets completely seal the synthetic shell from bulk solvent, and the pores formed by the cyclic axes of symmetry provide the only conduits into and out of the shell. The EM density for the residues around the CcmK pore of the T=4 shell is of good quality and allows placement of the backbone and sidechains for all residues (Figure 4.5c). CcmK1 and CcmK2 are identical in their pore-forming residues, and therefore even if the observed CcmK proteins are heterohexamers (i.e., mixtures or CcmK1 and CcmK2), the diameter of the pore is unlikely to be affected. The pores in the CcmK hexamers are about 7 Å in diameter, consistent with the sizes estimated from the hexamers in isolation [165]. Electrostatic surface analyses of the shells reveal a generally more negative charged exterior and positively charged interior (Figure 4.9a,b). The pores of both CcmK and CcmL are positively charged, especially pronounced on the exterior surface, possibly favorable to interaction with the negatively charged metabolites that must cross the shell.

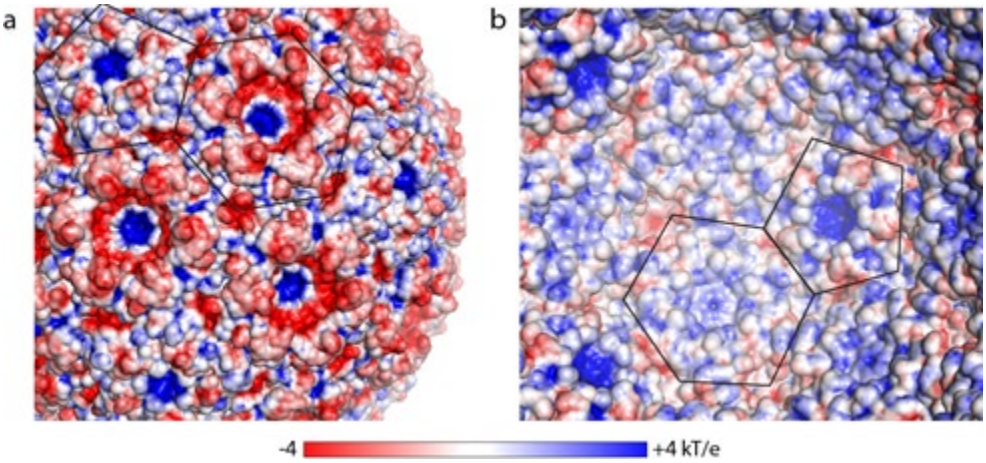


Figure 4.9 Electrostatics of the T=4 shell. (**a**) Surface as seen from the outside colored by electrostatic potential (red: negative, blue: positive) with an outline of CcmK and CcmL indicated (**b**) Same as **a** but for the inside view. Scale bar below indicates range from -4 to +4 kT/e.

There is some residual density at the center of the CcmK pore (Figure 4.5c) but it is weaker than the surrounding protein density, unlike the very strong density observed at center of CcmL (Figure 4.10a). At the center of the pentamer pore we observe a large spherical density on the concave, outside facing side of the shell. The density is approximately 8 Å in diameter and appears relatively smooth and featureless. Asymmetric reconstructions still result in featureless density occupying the CcmL pore (Figure 4.10b), therefore the density likely corresponds to an average of a molecule in multiple orientations. The density is surrounded by positively charged residues K39 and R59 and it is possible that a negatively charged molecule from the *E. coli* cytosol is bound and retained throughout the purification process.

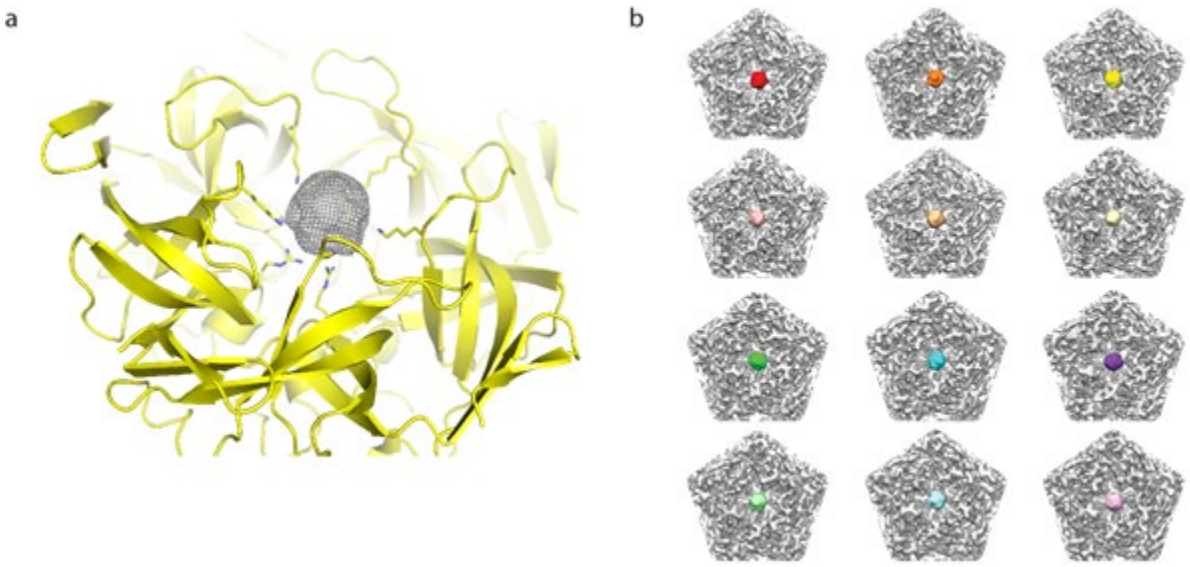


Figure 4.10 Density occupying the CcmL pore. (**a**) View of the pore from the T=4 shell with the cryo-EM density shown as a mesh contoured at 2 sigma and a cartoon model shown with residues lining the pore shown as sticks (K39 and R59). (**b**) Each of the twelve vertices of CcmL from a C1 reconstruction of the T=4 shell using the 35,384 T=4 particle set. The reconstruction was performed by low pass-filtering the initial reference to 12 Å, global angular searches and using a soft, circular mask.

There are a number of crystal structures available for both CcmK1/2 and CcmL. When we align the structure of one chain of the CcmK hexamer extracted from the shell to other available beta-carboxysomal CcmK1/2 structures (Figure 4.11a, left), the rmsd values are very low (0.3-0.5 Å over 80-90 aligned C-α atoms). However, when looking at the opposite side of the locally aligned hexamer we note that the hexamer extracted from the shell is more compact than its isolated counterparts (Figure 4.11a, right). This indicates that the BMC-H monomer fold is quite rigid but that there is some flexibility in the interactions between individual chains that allow the hexamer to adopt slightly different overall shapes. The difference could be due to the fact that the hexamer in the shell is more constrained by surrounding proteins relative to a single type of hexamer in the context of only crystal packing interactions. However, the flexibility could also have a role in the native particles where CcmK presumably interacts not only with itself and CcmL but also CcmO and CcmP. The CcmL monomer also aligns very well to structures of available beta-carboxysomal counterparts (Figure 4.11b, left; 0.4-0.55 Å over 75-85 aligned C-α atoms) but in contrast to the CcmK hexamer the whole pentamer also aligns very well (Figure 4.11b, right), indicating that the shape of the pentamer is quite rigid.

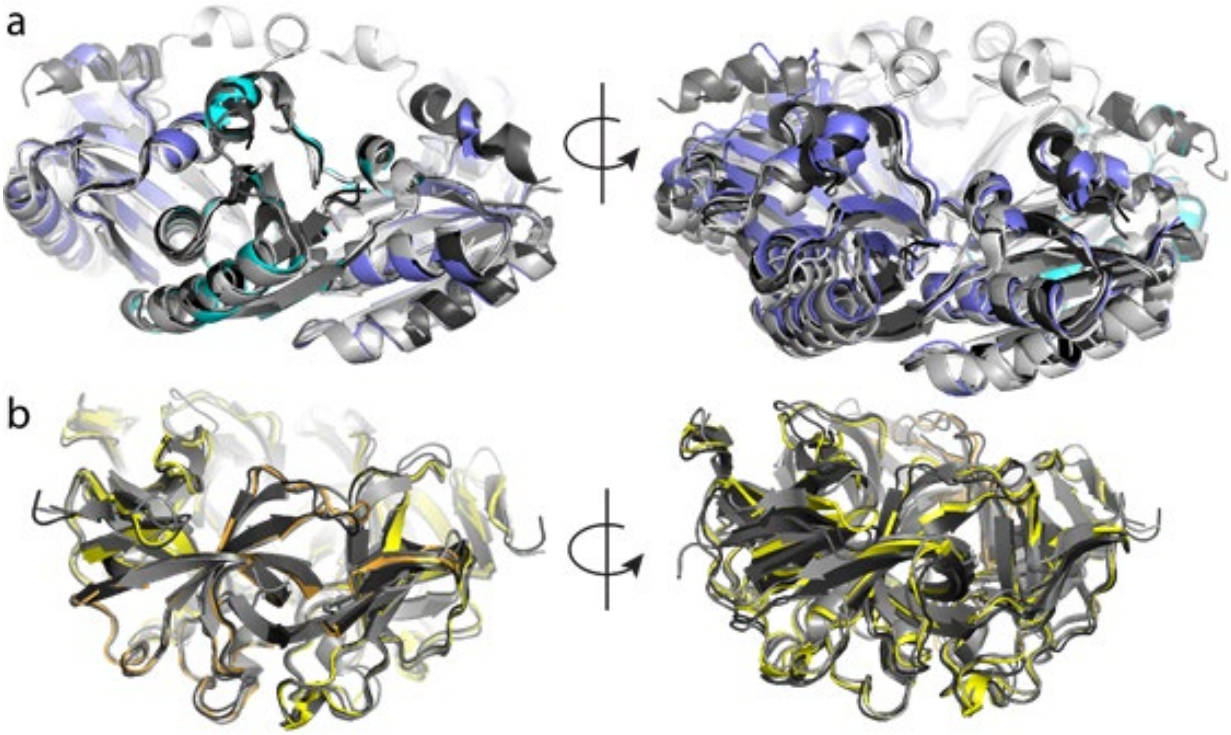


Figure 4.11 Structural alignment of *in situ* BMC-H and BMC-P protomer with isolated structures. (**a**) Alignment of *Halothece* CcmK1/2 (blue) with available crystal structures of CcmK1/2 (PDB IDs: 2A1B, 3BN4, 3CIM, 3SSQ, 3SSS, 4OX7) in different shades of grey. View of the aligned chain in cyan (left) and opposite side (right). (**b**) Alignment of *Halothece* CcmL (yellow) with homologous structures (PDB IDs: 2QW7, 4JVZ, 4JW0, 4N8X) in different shades of grey.

In regards to the peculiar architecture of the prolate shell, it consists of two hemispheres similar to the T=4 shell connected by an additional 'belt' of hexamer subunits, described by an elongation number of six (Q=6) [178]. Due to the additional hexamer belt, the prolate shell lacks the three-fold symmetry that would otherwise be present if the shell were a regular icosahedron, thus the structure is D5-symmetric. The low abundance of this class necessitated enforcement of this D5-symmetry during 3D-classification and refinement to yield sufficient resolution reconstructions to discern the identity of the constitutive shell protomers. While symmetric refinement of the entire complex resulted in density maps of sufficient quality to resolve secondary structure elements, the density on the circumference of the hemisphere regions appeared as a partial mix of subunit density and central belt was very weak relative to the hemisphere regions.

To address the structural heterogeneity and improve the resolution of the reconstruction, we performed symmetry expansion on the prolate particle images and treated the artificially increased dataset without enforcing symmetry. Classification on the expanded 'particles' with a soft-mask about the vertices and without angular searches revealed a mix of pentamer- and hexamer-centered particles (Figure 4.3c). This result indicates a mixture of prolate shells in *levo* and *dextro* configurations in which the T=4 hemispheres are rotated 36º relative to one another (Figure 4.12). The angular assignments for hexamer-centered particles were appropriately rotated, combined with the pentamer-centered class and refined to yield a 3.7 Å reconstruction. However, the central belt hexamers are still much lower quality than the remainder of the reconstruction. Nevertheless, despite the low quality, we could identify the shell protomers as CcmK and CcmL.

The low quality of parts of the structure and low abundance indicates that this is likely an incorrectly assembled particle rather than a relevant structural form.

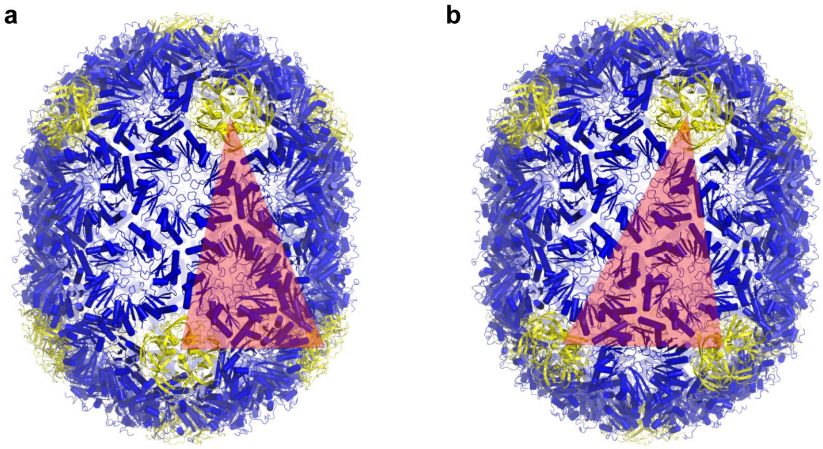


Figure 4.12 Geometries of different prolate shell geometries. (a,b) Close-up of cartoon models for the *laevo* (**a**) and *dextro* (**b**) prolate shell configurations, highlighted by transparent red triangles connecting vertices in the top and bottom hemispheres.

4.4 Discussion

The structures we have determined for the synthetic beta-carboxysome shells are formed by the major carboxysome shell proteins CcmK1, CcmK2 and the pentamer protein, CcmL [165,170,180]. Despite co-expression of CcmO and its presence throughout the purification, we could not unambiguously identify this protein in the final structure. It is possible that it is bound to the shell randomly or present in particles that are not ordered well enough to enable classification. Purified CcmO has so far eluded structural characterization (Kerfeld group, unpublished) and it is possible that it might only play a role and even be properly folded in native (enzymes + shell) carboxysomes. We observe two main types of structures, corresponding to pseudo- (because not all subunits are identical) T=3 and T=4 icosahedral particles as well as an elongated prolate type structure that is likely a result of misassembly. Native beta-carboxysomes are much larger (>150 nm diameter) [180], but it has been shown that the assembly of functional carboxysomes starts with an aggregation of the interior components [176] which might influence their size. Larger particles presumably have extended sheets of planar CcmK1/2 but the pentamer caps and the surrounding hexamers are expected to be the same as in our model because the angles are conserved among all sizes of icosahedral particles.

The permeability of carboxysome shells has not been measured directly; it is an essential parameter in estimating the CO_2 fixation efficiency of natural and designed carboxysomes, including the goal of enhancing CO_2 fixation in plants by heterologous expression of carboxysomes. The surface of the synthetic beta-carboxysome shell is tightly sealed, perforated only by 7 Å openings at the cyclic symmetry axis of each CcmK hexamer. These openings are large enough to readily allow the passage of bicarbonate/HCO_3^- (Figure 4.13a,b), and potentially PGA (3-Phosphoglyceric acid) (Figure 4.13c,d) but are unlikely to conduct RuBP (Ribulose 1,5-bisphosphate). Moreover, by revealing the sidedness of the shell proteins, our model provides new structural information about the orientation of the pores and the electrostatic potential on the interior and exterior surface of the shell (Figure 4.9); these provide constraints that now can be

used to inform theoretical models of the influence of shell permeability (i.e., CO_2 barrier, substrate and product permeability) on CO_2 fixation efficiency [30,110,181,182]. Notably the exterior surface (Figure 4.9a) is predominantly negative, with relatively strong positive charges focused at the pores. The metabolites (HCO_3^-, PGA and RuBP) that cross the carboxysome shell all are negatively charged. The overall global charge distribution is likely important for electrostatic steering of HCO_3^- to the openings in the shell. The synthetic shell system and its atomic resolution model provide an ideal system to probe the shell permeability.

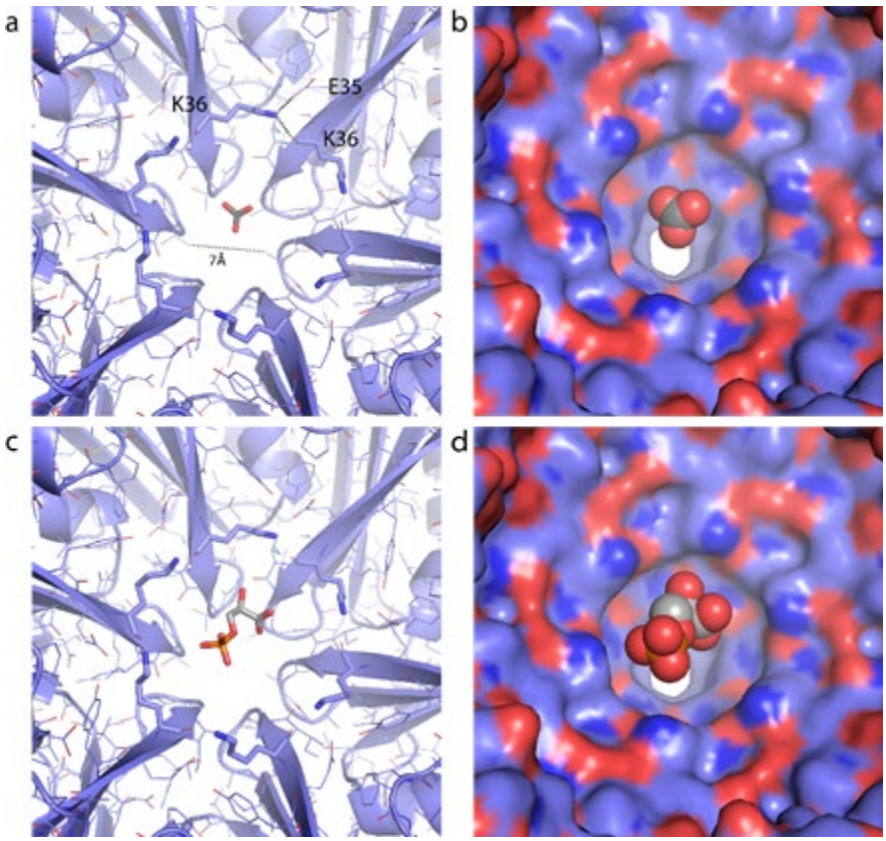


Figure 4.13 Analysis of BMC-H subunit CcmK pore (**a**) Close-up of the CcmK pore of the *Halothece* shell with a bicarbonate modeled in the center. K36 lines the concave side of the pore and hydrogen bonds to the E35 side chain and the backbone oxygen of K36 of the adjacent chain. (**b**) Surface representation of **a**. (**c**,**d**) same as **a**,**b** with 3-phosphoglyceric acid (PGA).

In addition, knowing the structure will facilitate the design of modifications, such as the SpyTag/Catcher system that has been used to "program" synthetic metabolosome shells [183] to reliably encapsulate cargo. Since the CcmK structures are highly conserved, a substitution of some subunits with homologues is possible, for example to test the effect of capping of CcmK4 subunits with CcmK3-CcmK4 heterohexamers as shown recently [184]. Our detailed structure of a synthetic carboxysome shell, together with our previous biochemical characterization [177] makes this an attractive model system to study aspects like encapsulation and shell permeability.

We recently reported the structure of a synthetic catabolic BMC shell derived from the phylogenetically remote BMC from *Haliangium ochraceum*. This HO shell is constructed from a BMC-P protein, a single type of hexamer (there are no BMC-H paralogs in the genome) and three BMC-T proteins. The carboxysome and HO BMC are functionally distinct and are found in organisms remote from one another on a phylogenetic tree (Deltaproteobacteria vs Cyanobacteria)

[177,185]. The amino acid sequences of HO-BMC-P and CcmL are 49% identical and a structural alignment yields an rmsd of 0.7 Å over 72 aligned C-α atoms. Similar values are observed when comparing the HO-BMC-H and CcmK1/2; the sequences are 53% identical and the structures align with an rmsd of 0.65 Å over 78 aligned C-α atoms. We also compared the interfaces between two hexamers and between the hexamer and pentamer for the beta-carboxysomal shell and the HO shell. Despite a number of amino acid substitutions, the backbones of the subunits involved in both the hexamer-hexamer and hexamer-pentamer interface align almost perfectly (Figure 4.14a,b).

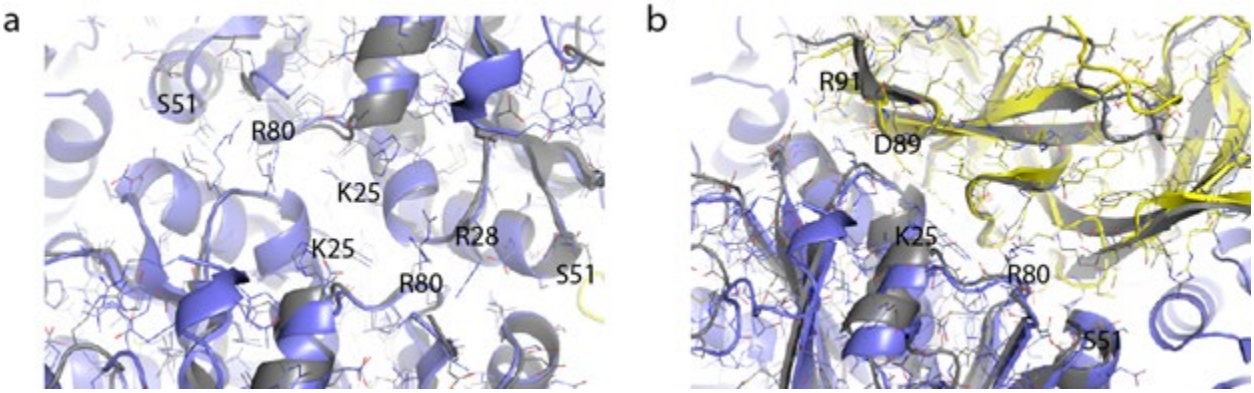


Figure 4.14 Alignment of the interfaces with the corresponding HO shell proteins. (**a**) Alignment of the angled BMC-H-BMC-H interface of the Halo shell (blue) with the HO shell (grey). (**b**) Alignment of the BMC-H-BMC-P interface, Halo shell BMC-H in blue, BMC-P in yellow; HO shell in grey. Selected residues are labeled for orientation.

In a phylogenectic tree of BMC-P proteins (BMC-H trees are much more difficult to generate and interpret due to permutation and duplication events) the *Halothece* CcmL and *Haliangium* BMC-P are in very distinct clades and separated almost all the way to the base of the tree (Figure 4.15). In this context, the structural conservation between these functionally distinct BMC types is remarkable and suggests that this is a very stable architectural solution for capping interaction between the pentamer and its surrounding hexamers, one that has persisted through evolution and functional diversification. Indeed, BMC-H and BMC-P seem to form an optimal structure that both allows the formation of the vertices, where a BMC-P pentamer is surrounded by five BMC-H hexamers while at the same time allowing the BMC-H to form specific interactions between each other. The CcmK-CcmK interaction around the pentamer is at a 30° angle while a planar interaction (as observed in the HO shell) is necessary to form sheets for larger facets. These findings provide detailed structural information on key inter-subunit interfaces (such as BMC-P:BMC-H and BMC-H:BMC-H and subunit sidedness, providing a foundation for the direct study of permeability of carboxsyome shells.

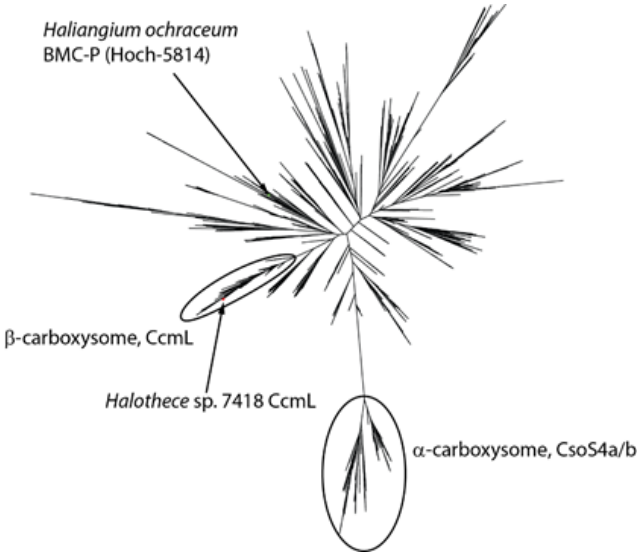


Figure 4.15. Phylogenetic tree of BMC-P sequences. BMC-P amino acid sequences were obtained by searching the RP75 database using the PFAM03319 HMM using hmmsearch (www.ebi.ac.uk). The sequences were made non-redundant to 99%, trimmed with trimAl [186], aligned with ClustalW [187] and the sequence alignment was used to build the tree in PhyML [188] which was then visualized using Archaeopteryx [189]. The alpha/beta carboxysome branches were identified by inspecting the genomic neighborhood of several clade members.

4.5 Tables

Table 4.1 Synthetic carboxysome shell cryo-EM data collection, refinement and validation statistics.

	T = 4 (EMD-20207) (PDB 6OWG)	T = 3 (EMD-20208) (PDB 6OWF)	Prolate (T = 4, Q = 6) (EMD-20209)	Prolate Expanded (T = 4, Q = 6) (EMD-20210)
Data collection and processing				
Magnification		56,000		
Voltage (kV)		300		
Electron exposure (e–/Å^2)		50		
Defocus range (μm)		-0.5 to -2		
Pixel size (Å)		0.8908		
Symmetry imposed	I	I	D5	C1
Initial particle images (no.)		47,121		
Final particle images (no.)	17,253	1,586	2,123	16,942
Map resolution (Å)	2.6	3.0	4.1	3.7
FSC threshold	0.143	0.143	0.143	0.143
Map resolution range (Å)	2.5-2.9	2.9-3.4	3.8-6.0	3.4-6.0
Refinement				
Initial model used (PDB code)		2A1B,2QW7		
CC_mask	0.87	0.87		
Model resolution (Å)	2.7	3.2		
FSC threshold	0.5	0.5		
Map sharpening *B* factor (Å^2)		phenix.auto_sharpen		
Model composition				
Non-hydrogen atoms	178,080	133,500		
Protein residues	23,640	17,700		
Ligands				
B factors (Å^2)				
Protein	12	71		
Ligand				
R.m.s. deviations				
Bond lengths (Å)	0.006	0.006		
Bond angles (°)	0.79	0.85		
Validation				
MolProbity score	1.2	1.5		
Clashscore	3.7	4.0		
Poor rotamers (%)	0.1	0.38		
Ramachandran plot				
Favored (%)	97.7	95.8		
Allowed (%)	2.3	4.2		
Disallowed (%)	0	0		

4.6 Materials and methods

Expression and purification of synthetic carboxysome shell

The plasmid Halo-1 as described in [177] was modified using Gibson cloning to include a Strep-II tag (WSHPQFEK) at the C-terminus of CcmL connected by a GS linker. *E. coli* (DE3) cells were transformed with this plasmid and grown in Studier's autoinducing medium at 25 °C O/N. The cell pellet was resuspended in 100 mM Tris pH 8, 150 mM NaCl, 1 mM EDTA (buffer A), Benzonase and Lysozyme added (EMD Millipore, Novagen. manufacturer's instructions) and lysed using French Press. The lysate was centrifuged in a SS-34 rotor at 20,000 rpm for 20 min. The soluble lysate was then applied to a 5 mL GE StrepTrap equilibrated in buffer A, the column was then washed with 6 column volumes (CV) buffer A and the protein eluted with 5 CV 20 mM Tris pH 8.0, 50 mM NaCl, 1 mM EDTA, 2.5 mM d-Desthiobiotin. The eluate was then applied on a sucrose cushion (30 % sucrose in 20 mM Tris pH 8.0, 50 mM NaCl) and centrifuged in a Ti-70 rotor for 16 h at 42,000 rpm. The glassy pellet was then resuspended in 20 mL 20 mM Tris pH 8.0, 50 mM NaCl and applied to a MonoQ 10/100 anion exchange column. Shells eluted at around 35% of a 50-1000 mM NaCl gradient in 20 mM Tris pH 8.0.

Cryo-electron microscopy grid preparation and data collection

In the chamber of a Vitrobot Mark IV at 4ºC and 100% humidity, 3 µL of sample at ~0.6 mg/mL was applied to a freshly glow-discharged R1/2 Cu 300 grid (Quantifoil) and immediately blotted with Whatman no. 1 filter paper for 6 s prior to plunging into liquid ethane cooled to liquid nitrogen temperatures. Data was collected on a Titan Krios G2 transmission electron microscope operated in EFTEM mode at 300 kV and equipped with a K2 summit direct electron detector and Quantum LS Imaging Filter (slit width of 25 eV). Movies were recorded in super-resolution mode at an effective pixel size of 0.4454 Å with a cumulative exposure of 50 e^{-} $Å^{-2}$ distributed uniformly over a 6 s exposure into 25 frames. Data acquisition was performed using SerialEM [74] with one exposure per hole and focusing for each exposure to a target defocus range of -0.5 to -2 µm. A total of 1343 movies were collected.

Single-particle analysis image processing

Preprocessing was performed concurrently with data-collection through Focus [75]. Super-resolution movies were drift-corrected, dose-weighted, and Fourier cropped to 0.8908 Å per pixel using MotionCor2-v1.1.0 [76] and CTF parameters estimated using Gctf-v1.06 [77]. Aligned micrographs were manually curated and 1334 selected for further processing. Non-templated particle-picking was performed using Gautomatch-v0.56 (https://www.mrc-lmb.cam.ac.uk/kzhang/Gautomatch) and results curated to remove empty picks resulting in 47,121 particles. Particles were extracted in a 512 pixel box and binned by a factor of two using RELION-v3.0.b2 [81], subjected to reference-free 2D-classification and the different assembly types selected individually for further processing. All reported global map resolutions correspond to the 'gold-standard' Fourier shell correlation (FSC) at 0.143 between independently refined half-maps, with phase-randomization to correct for masking artefacts [83]. Local resolution maps were generated using RELION.

For the T=4 assembly, 35,383 particles were re-extracted in a 424 pixel box and imported into cryoSPARC-v2.2 [79] for ab initio reference generation and subsequent homogenous refinement with icosahedral symmetry enforced resulting in a 3.1 Å map. The cryoSPARC refinement metadata was convert to RELION STAR format for further processing using Daniel Asarnow's pyem scripts (unpublished, https://github.com/asarnow/pyem). Contrast transfer function (CTF) parameter refinement followed by 3D auto-refine improved the global resolution to 2.8 Å. Bayesian polishing [82] further improved the map to 2.6 Å. Further 3D-classification without alignment into four classes resulted in a class of 15,354 particles that yielded a 2.6 Å map with marginal improvements.

The T=3 assembly was processed similarly as the T=4 for 1,586 particles extracted in a 384 pixel box and achieving final global resolution of 3.0 Å. Further classification of the particles did not yield an improved map.

For the prolate (T=4, Q=6) assembly, 2,123 particles were extracted in a 484 pixel box and binned by a factor of two (Nyquist-limit ~ 3.6 Å). An ab initio reference could not successfully be generated using either cryoSPARC or RELION. Nevertheless, based on the known dimensions of the shell protomers and 2D class averages, a coordinate model was built and converted to a pseudo-map using Chimera-v1.12 [86]. The pseudo-map was low-pass filtered to 12 Å and used as an initial reference for auto-refinement in RELION using D5 symmetry, which resulted in a 4.4 Å. Refinement of CTF parameters and Bayesian polishing improved the map to 4.1 Å. To further improve the resolution, the particle images were expanded with D5 symmetry, resulting in 21,230 expanded-particles. Focused classification on the expanded-particles was performed into five classes, using C1 symmetry and applying a soft-mask encompassing an asymmetric unit of the complex. This resulted in two dominant classes, one a pentamer-centered face (58%) and the other hexamer-centered (22%). A tenth of a rotation was applied to the hexamer-centered particle metadata and particles merged with the pentamer-centered class resulting in 16,941 that yielded a 3.7 Å map after auto-refinement. The necessity to invert the handedness of the map suggests low initial reference bias and FSC curves dropping to around zero indicate absence of duplicate particles.

4.7 Accession numbers

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers AFZ45643.1 (ccmK1), AFZ45642.1 (ccmK2), and AFZ45644.1 (ccmL) for *Halothece* sp. PCC 7418 shell proteins. Density maps were deposited with the Electron Microscopy Data Bank with accession codes EMD-20207 (T = 4 shell), EMD-20208 (T = 3 shell), EMD-20209 (prolate shell), and EMD-20210 (symmetry-expanded prolate shell). Raw frames were deposited with the Electron Microscopy Public Image Archive with accession code EMPIAR-10275. Coordinate models were deposited with the Protein Data Bank with accession codes 6OWF (T = 3 shell) and 6OWG (T = 4 shell).

Chapter 5

Conclusion

5.1 Summary

Our work has revealed the structure and function of different components of the protein machinery - both enzymes and structural scaffolds - that give rise to cyanobacterial photosynthesis and CO_2-concentrating mechanism (CCM).

Chapter 1 begins by providing the background on the protein machinery that facilitates the remarkable processes of photosynthesis and the CCM in cyanobacteria. In particular, it highlights the importance the NAD(P)H dehydrogenase-like (NDH) complex in balancing energy inputs for CO_2-fixation and other aspects of downstream metabolism. Furthermore, it describes the enigmatic vectorial carbonic anhydrase (vCA) activity of the NDH-associated CO_2-uptake proteins (CUPs).

In Chapter 2, we show the cryo-electron microscopy (cryo-EM) structure of the cyanobacterial NDH complex. We reveal the structures and arrangements of the conserved oxygenic photosynthesis specific subunits distinct to the complex compared to respiratory homologs. Finally, from the structure we are able to derive a putative model for the interaction between NDH and its redox substrate ferredoxin (Fd).

Chapter 3 starts with a systematic genetic screen to identify all the components that give rise to the CCM of a non-photosynthetic, CO_2-fixing bacterium. This screen resulted in the discovery of the DABs accumulate bicarbonate complex (DAB). Our physiological and biochemical characterization of the DAB reveal it to possess similar vCA activity to the NDH-CUP complexes. However, our work indicates the DABs vCA activity is achieved by different biochemical mechanism compared to the CUPs. Furthermore, unlike the NDH-CUP complexes, DABs are ubiquitous throughout prokaryotic phyla including heterotrophic and pathogenic organisms.

Finally, in Chapter 4, we determine the structures of several different assembled forms of simplified carboxysome shell using cryo-EM. These structures reveal general architectural principles for carboxysome shell components, as well as related proteinaceous organelles.

5.2 Discussion and outlook

Our work and others have been enlightening in regards to the understanding of NDH structure and function. In spite of this progress, many questions remain. The structure of the cyanobacterial NDH resolved most of the conserved OPS subunits and provided initial insight as to how Fd may interface with the complex for electron donation. Furthermore, one of the more provocative results of our cryo-EM structure was the observation of a strong, unexplained density near the base of the peripheral arm that we termed the ‘X-cofactor’. The X-cofactor appeared in a third of the particles in each dataset of our datasets and was found in a solvent exposed cleft formed by the N-terminal region of the OPS subunit, NdhM. The X-cofactor may be a purification artifact or a small molecule which represents an unknown function of the enzyme complex. The X-cofactor density, however, was not observed in the cryo-EM structure from Schuller *et al*. (or other structures described below), which used different purification and classification strategies [190-192]. Thus, additional studies are required to confirm the presence, identity and significance of the

X-cofactor in regards to NDH function. The nature and identity of the X-cofactor is an ongoing aspect of research in our group.

During the writing of this thesis, Pan *et al.* determined high-resolution single-particle cryo-EM structures of the *T. elongatus* NDH bound to Fd (NDH-Fd) and plastoquinone (NDH-PQ) [191]. In the NDH-PQ structure the authors also saw low-occupancy for NdhS (Figure 2.2) and absence of NdhV. This result led the authors to incubate recombinantly prepared NdhS, NdhV and Fd in excess to the complex to obtain the NDH-Fd structure. Surprisingly, neither NdhS nor NdhO is directly involved in Fd binding. NdhV, a small α/β protein, binds the interface of NdhS, NdhS, and core subunit NdhI. Specifically, the C-terminal segment of NdhV binds on the opposite side of the "basic-triplet" (K89-K90-K91) of NdhI. The binding of NdhV appears to buttress the NdhI β-hairpin and enable formation of a component Fd-binding site which consists of the aforementioned basic-triplet and other residues on core subunit NdhH. This Fd-binding site is consistent with what we initially identified as the "O-site", although Fd does not appear to interact with NdhO. Moreover, the NDH-PQ suggest a possible mechanism to overcome the stoichiometry problem between the single electron donor Fd and the two electrons necessary for full PQ reduction. The canonical PQ-site proximal to the terminal FeS cluster is occluded by a loop of NdhH and the observed PQ density is stably positioned too far away from the cluster for electron transfer. The authors suggest this loop may serve to prevent PQ reduction until the terminal cluster is reduced and another Fd is bound, at which point a conformational change in the loop would allow PQ access to the terminal cluster for rapid, successive reduction to PQH_2.

Also, during the writing of this thesis, Schuller *et al.* determined the high-resolution structure of the NDH-1_3/NDH-1MS complex from *T. elongatus* by single-particle cryo-EM and performed simulations to begin to elucidate the mechanism of a vCA [192]. Consistent with prior negative-stain work, CupA and CupS are positioned at the distal end of the membrane arm above NdhF3 giving the complex its defining U-shape. CupA adopts a fold distinct from any known CA. Remarkably, the CA active-site is at the interface of CupA and NdhF3 and contains the necessary Zn^{2+}, which the authors observe in the EM density as well as by spectroscopy. The catalytic Zn^{2+} is coordinated by CupA residues His130 and Arg135, as well as density assigned as a putative water/hydroxide or carbonate molecule. The active-site is completed by NdhF3 residues Arg37, Glu114 and Tyr41. Of particular note is Tyr41, which acts as part of the putative proton shuttle in the proposed CO_2 hydration mechanism.

However, despite the interfacial active-site between CupA and NdhF3, the mechanism of coupling does not appear to be proton-abstraction via NdhF3 antiporter activity to regenerate the catalytic Zn^{2+}. The authors note NdhF3 lacks the charged residues proposed form the proton-tunnel in other Mrp-like antiporter homologs and is incompetent for proton-translocation in their simulations. In its place, the authors propose NdhF3 possess a putative gas-tunnel that leads to the interfacial active-site with CupA and may serve as a conduit for CO_2, which is supported by their simulations. Moreover, while possibly lacking proton-antiporter activity, NdhF3 still possesses the conserved ion-pair at the interface with the adjacent proton-antiporter subunit NdhD3 as seen in other complex I homologs. This connection may enable coordination of the CO_2-hydration activity by NdhF3/CupA to the proton-pumping activity of the other subunits, thus providing support for the "alkaline pocket" model of vCA activity.

However, our recent discovery and characterization of different putative class of vCA complexes provides a comparative model for interrogating the structure and energy-coupling mechanism of the CUP complexes. The proposed DAB mechanism speculates dissipation of an inward monovalent cation gradient (likely H^+ or Na^+) ensures unidirectional hydration of CO_2 in the cytosol and is distinct from the outward pumping of H^+ in the NDH-coupled CUP complex mechanism. Furthermore, functional DABs exist in a wide variety of prokaryotes and archaea including pathogens. Despite the lack of sequence homology, DabA and CupA/B appear to possess quasi-convergent vCA activity. Therefore, owing to the existence and characterization of DAB complexes, it is curious that the CUPs complexes associate and couple to the function of the NDH complex in cyanobacteria.

The structure of NDH-CUP complex by Schuller *et al.* provided many novel insights and hypotheses on residues important for the function of CUPs. However, future work on the CUPs using mutagenesis to assess the CUP mechanism will be necessary and insightful. Unfortunately, an *in vitro* assay for the complex's CO_2-hydration activity has yet to be established. In this regard, the Burnap group has made progress in developing cyanobacterial strains and plasmids to facilitate more thorough structure-function studies of the CUPs *in vivo* [193]. Moreover, in light of the discovery and characterization of the two subunit DAB complexes, alternative mechanisms for vCA activity should be considered and investigated, such as coupling to a cation-gradient. Similar to the NDH-CUP complex, a structure of a DAB complex will be valuable in elucidating the mechanism of its vCA activity. Finally, functional reconstitution of vCAs *in vitro* will be fundamental in understanding this class of enzymes.

Appendix A

Further analysis and considerations on the NDH cofactors

A.1 Further analysis and considerations on the X-cofactor

One of the more provocative results of our previous cryo-EM work on the cyanobacterial NAD(P)H dehydrogenase-like complex was an unexplained density near the base of the peripheral arm that we termed 'the X-cofactor' (Figure 2.9). The density for the X-cofactor was on par with that of the bonafide 4Fe-4S clusters in the complex, thus leading us to believe that the density could be explained by a metallocofactor. Therefore, our experiments to identify the composition of the X-cofactor have been analytical methods canonically used for the detection of metals such as X-ray fluorescence (XRF) and inductively coupled plasma mass spectrometry (ICP-MS).

A thorough explanation of XRF can be found elsewhere [194]. Briefly, XRF is a spectroscopic technique that relies on the fact that the difference in energies between electron shells is characteristic of a particular element. A sample is excited with an x-ray beam at or beyond the energy necessary to dislodge an inner-shell, or "K-shell" electron. For a particular element the minimum energy for dislodging an inner-shell electron is called the "K-edge". For second to third row elements this energy is on the order of 10^3-10^4 eV. After the inner-shell (lower energy) electron has been ejected, an outer-shell (higher energy) electron will move down to fill the vacancy and emit the difference in energy between the two shells as a fluorescence signal that is characteristic of the element. The use of this technique is chiefly used to identify metals or other elements in a protein crystal to be used for phasing by anomalous scattering.

Similarly, a thorough explanation of ICP-MS can be found elsewhere [195]. Briefly, "inductively coupled plasma" refers to the ionization source which is an inert gas (typically Ar) brought to the plasma phase via inductive heating with an electromagnetic coil and render the medium electrically conductive. ICP typically decomposes a sample into constituent elements and transforms those into atomic or small poly-atomic ions for mass spectrometry. ICP-MS is a quantitative and highly-sensitive technique (with high-end models detecting species at parts-per-trillion). It is particularly well-suited for detecting metals due to the method of ionization, although some organic molecules can persist as small poly-atomic ions upon ionization.

We used XRF as a qualitative technique on precipitated, purified NDH (as described in Section 3.6) in an attempt to probe the chemical identity of the X-cofactor and showed signal for Fe, Cu, Zn, S and Cl in the sample (Figure A.1). ICP-MS was also performed on purified NDH provided to our collaborators (A. Bayne and J-F. Trempe), who detected Li, Al, Ti, Mn, Rb, Ba, La, Pr, Nd at the µg/kg level, Cu, Zn, Ce at the 100 µg/kg level and K at the mg/kg level. Fe was also detected by ICP-MS. Unfortunately, an appropriate calibration standard for Fe was not available but it was stated to be higher than buffer background and the other protein sample ran that day. The Fe iron detected by both techniques is expected due to the known iron-sulfur clusters along with the S detected by XRF, while the Cl is can be explained by the buffer. The Cu and Zn detected by both techniques are interesting, as well as the Mn detected by ICP-MS. However, a metallocofactor composed of any of those elements is still hard to rectify with the context of the X-site (Figure 2.9d) which contains mostly positive moieties. The Ba-La-Ce-Nd-Pr signal detected by ICP-MS are certainly peculiar since these elements are rare-earth metals and unexpected in biological specimen. An alternative explanation for the Ba-La-Ce-Nd-Pr signal is a small poly-atomic ion of an organic molecule with a MW = 133-140. Interestingly, this mass range is

consistent with the nitrogenous base adenine (MW = 135.13) when also considering isotopic distribution.

Cryo-EM work from another group did not display the same density for the X-cofactor; however, the purification and reconstruction strategies were quite different [190-192]. Nevertheless, the structure did agree on the open cleft (suggestive of a binding-site) at the base of the peripheral arm where we see the X-cofactor in our reconstructions. Moreover, recent cryo-EM work from the Brohawn group found a similarly strong density for a small, pi-conjugated drug molecule in their reconstruction [196]. Therefore, our initial assumption that the X-cofactor is metallic in nature may be incorrect, and it may in fact be a conjugated organic molecule. Interestingly, the Kramer group have demonstrated the inhibition of the chloroplast NDH activity by adenylates (e.g., ATP, ADP, etc.) [197]. Thus, along with the anomalous 133-140 mass signal from the ICP-MS, it is tempting to speculate that the identity of the X-cofactor could be an adenylate or another metabolite that we co-purified with the complex. However, the mode of inhibition of NDH by adenylates concluded by Fischer *et al.* was competitive, which is inconsistent with distance of the X-cofactor from Fd and PQ binding sites. Therefore, while the anomalous signal ICP-MS may be adenine, it may not be bound at the X-site.

The X-cofactor density is certainly an interesting feature of our cryo-EM reconstructions. However, the chemical nature of the X-cofactor and its biological relevance remain enigmatic. The possibility that the X-cofactor is an artifact of the purification strategy should not be excluded.

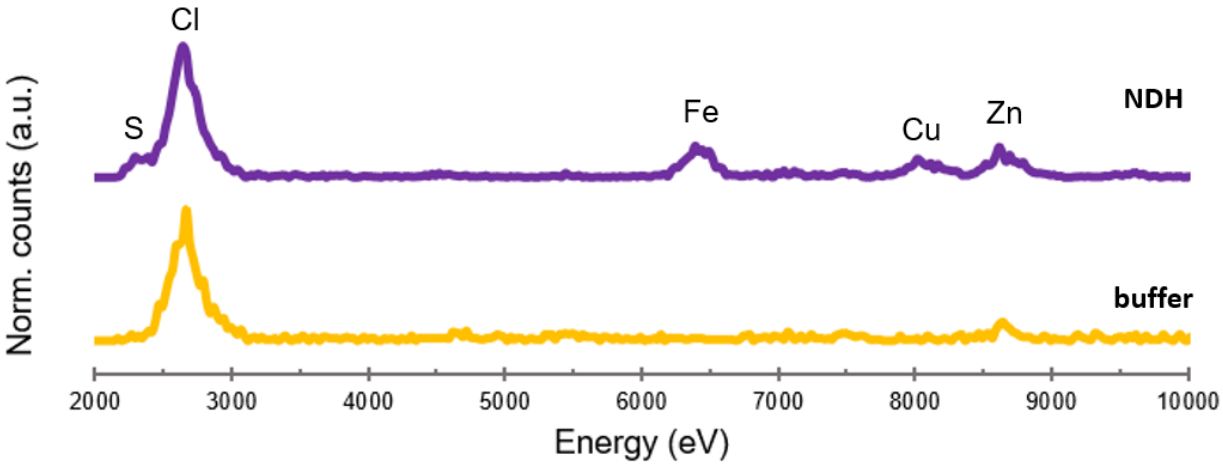


Figure A.1 XRF emission spectrum of purified *T. elongatus* NDH. XRF spectrum covering 2000-10000 eV for NDH (purple) and buffer control (yellow) normalized by respective Cl peaks and offset for display purposes. Discernible peaks (S, Cl, Fe, Cu, Zn) are labelled.

Nevertheless, it is entertaining to speculate on the mechanism(s) of electron transfer through NDH as briefly described in Chapter 2.4 and are elaborated on here that involve the X-cofactor (Figure A.2). To reiterate Chapter 2.4, a significant difference between the mechanisms of complex I and NDH is the stoichiometry of redox substrates. In the case of NDH, ferredoxin (Fd) shuttles only one electron to the complex at a time, but two donations events are required for complete reduction of plastoquinone (PQ) to plastoquinol (PQH_2), which proceeds through a radical intermediate ($PQ^{\cdot}H$). This is in contrast to respiratory complex I that accepts two electrons simultaneously from a single NADH to complete its redox mechanism. Thus, unlike complex I, in the mechanism of NDH the radical intermediate persists until the first oxidized Fd dissociates from

the complex and a second reduced Fd binds the complex. During this time, the radical has the opportunity to form adducts or other deleterious chemistry to the NDH complex. The 'null' hypothesizes are that the PQ-binding site of NDH can stabilize the radical intermediate to mitigate its reactivity or somehow only position the PQ for reduction when two successive electron transfers can be assured. However, our cryo-EM structures presented possibilities for alternative hypothesis for how NDH could overcome the Fd-PQ stoichiometry mismatch. The following models consider the S- and O-sites as two possible Fd-binding sites, and that the X-cofactor maybe a redox-active cofactor (metal or otherwise).

The models are namely: 1) 'Double-barrel'- The SH3-like subunits NdhS and NdhO may define two binding sites on NDH allowing more rapid, successive donation events than a single-site mechanism, thus maximize complete reduction of PQ to PQH_2 (Figure A.2a). This model in some sense needs to invoke coordination between the sites since simultaneous donation from both sites is not possible due to the nature of the iron-sulfur clusters. Either each beta-'barrel' adjacent site must be loaded with a reduced Fd before the privileged site can 'fire' first, or only one site is competent for electron donation to NDH and the second site serves to hold a reduced Fd in wait for reloading the competent site. 2) 'Flux-capacitor'-The X-cofactor may participate in redox activity and act as an electron buffer to abstract a primary electron from a long-lived PQ·H and return the electron upon reformation of PQ·H by a secondary Fd donation event, thus completing reduction to PQH_2 (Figure A.2b). This model requires that the X-cofactor be near-isopotential with the PQ·H to enable both abstraction and return of an electron to the same species. Note the product species from the abstraction and return processes are different, i.e., PQ vs PQH_2. 3) 'Side-entry'- The X-cofactor may provide an alternative entry-point for electrons, thus allowing parallel reduction of PQ from two distinct electron donation pathways (Figure A.2c). 4) 'Side-exit'-The converse of Side-entry, the X-cofactor serves as an exit-point for a PQ·H to quench by releasing the high energy electron to some donor molecule or solvent (Figure A.2d).

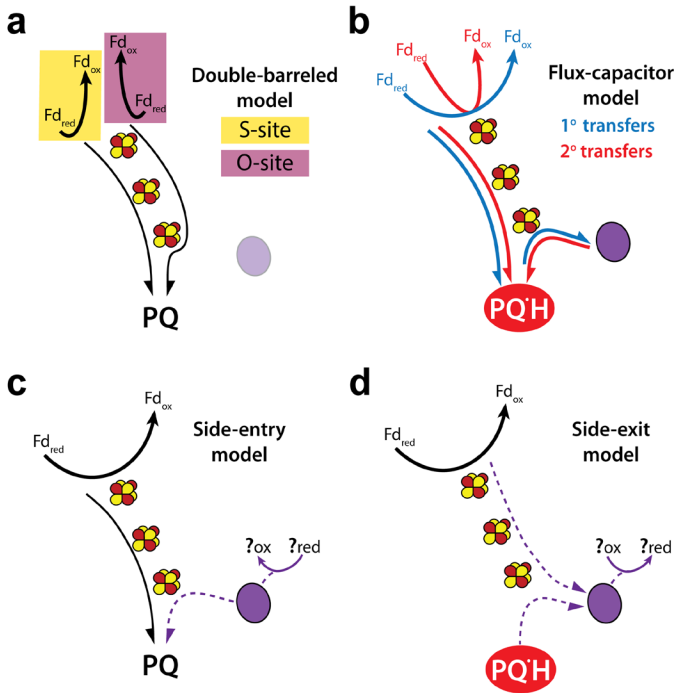


Figure A.2 Select speculative models on NDH electron transfer mechanism. As described in the text (**a**) Double-barreled model, (**b**) Flux-capacitor model, (**c**) Side-entry model and (**d**) Side-exit model.

However, as stated in Chapter 5.2, during the writing of this thesis Pan *et al.* were able to obtain cryo-EM structures of both NDH bound to Fd (NDH-Fd) and NDH bound to endogenous PQ (NDH-PQ) [191]. The NDH-Fd structure revealed that only a single Fd-binding site is present on NDH. This site was consistent with what we proposed as the 'O-site' in Chapter 2. The positioning of Fd at this site is such that it sterically occludes binding at what we proposed as the 'S-site', therefore discrediting the 'Double-barrel' model. In the NDH-PQ structure, however, Pan *et al*. observe putative density for co-purified PQ in a pocket different from the canonical PQ-binding site adjacent the terminal iron-sulfur cluster. The PQ molecule is blocked from the canonical site by an extended loop from the core subunit NdhH. Seeing only one Fd-binding site exists and noting the Fd-PQ stoichiometry problem, Pan *et al.* propose a model in which the PQ is bound by NDH but impeded from reaching the terminal iron-sulfur cluster until two successive electron transfers can be assured. In their model, the first Fd donates an electron that is transferred and held on the terminal iron-sulfur cluster, changing its charge state. Then, a second Fd electron donation event triggers a conformational change in the NdhH loop that allows PQ to access the terminal iron-sulfur cluster for reduction by two successive electron transfers, thus limiting the time spent by the PQ as the radical intermediate. Furthermore, as with author published NDH structures, Pan *et al.* do not observe density for the X-cofactor, thus supporting that the X-cofactor may be an artifact of some kind.

A.2 Possible future experiments on the X-cofactor

Given that others have not observed density for the X-cofactor suggests it may be an artifact. Our previous attempts to ascertain its chemical identity have been inconclusive. Furthermore, it is difficult to conceive an experiment that would unambiguously determine the chemical nature of the X-cofactor. Nevertheless, experiments can be conceived to assess if the X-cofactor is an entity of biological relevance.

One possible *in vivo* strategy to assess the significance of the X-cofactor is to engineer point mutants in the X-cofactor binding site (i.e., the 'X-site', see Figure 2.9d). NDH is essential in cyanobacteria, with the exception of the glucose-tolerant strain of *Synechocystis sp.* PCC6803 which can be cultured mixotrophically. Therefore, one could compare the growth and cyclic electron flow rates of PCC6803 X-site mutants to determine effects of X-site mutations.

Alternatively, a possible *in vitro* strategy to determine the significance of the X-cofactor is to perform electron paramagnetic resonance (EPR) spectroscopy. Like other spectroscopic technique, EPR measures the absorption of light. In particular, EPR measures the absorption of frequencies in the microwave range by a specimen in an applied magnetic field. Specimen yield a signal under these conditions when they possess one or more unpaired electrons (i.e., paramagnetic centers). Therefore, EPR is ideal for assessing the redox activity of protein cofactors since gain or loss of an electron will change the number of unpaired electrons, thus the EPR signal. To assess whether the X-cofactor is redox active, one could measure the EPR signal of fully oxidized and fully reduced NDH. Oxidized iron-sulfur clusters are paramagnetic, thus should yield distinct EPR signals in the fully oxidized condition. The iron-sulfur cluster signals should then disappear in the reduced specimen. If an anomalous signal is not observed under either condition for the X-cofactor, it is not a redox active cofactor and thus not likely involved in the NDH electron transfer mechanism.

A.3 Further analysis and considerations on lipids

Lipids and lipid-bilayers are essential components of life that are often overlooked (the reader is encouraged to examine the lack of detail this component in most diagrams of various cells, organelles and electron transport chains). Lipid-bilayers, or membranes, along with cognate membrane proteins provide the selective permeability barriers that enable the fundamental processes such as photosynthesis and respiration. Lipid membranes inhibit diffusion of ions and thus are a prerequisite in generation of the electrochemical gradients used for ATP synthesis and active-transport. Moreover, in recent years, appreciation has grown for the role of lipids as critical for the structural and functional integrity of membrane proteins [198–200].

Multiple studies attest to the importance of specific lipid requirements for the structural integrity and proper function of membrane proteins [201–204]. Lipids may be necessary for stability in multisubunit membrane complexes by increasing shape complementarity between adjacent subunits [205,206], have a chaperonin-like function in folding or directly participate in the mechanism of the protein (such as in mechanosensitive channels) [204,207]. Many examples can be found in respiratory complexes, such as the essentially of cardiolipin for the enzymatic function of the respiratory complex I and the so-called "interhelical" lipid in the cytochrome bc_1 complex (i.e., complex III) [201,208]. In addition, lipids have been shown to mediate many of the contact sites among various super-complexes and may explain the fragility of these complexes upon detergent-solubilization and 'delipidation'[209,210].

Similar to the respiratory complex I, co-purified lipids are also observed in or cryo-EM maps of NDH (Figures A.3). All of the observed lipids are observed at subunit interfaces, primarily at the distal end (i.e., about NdhD1/F1/P/Q) (Figure A.4). Therefore, these lipids likely contribute to the structural integrity of the complex. Indeed, more lipids are observed in dataset-2 which has improved resolution in the NdhD1 and NdhF1 subunits where the majority of the lipids are present. With the exception of time from solubilization to grid-preparation, purifications for dataset-1 and dataset-2 were performed with the same protocol and lot number of β-DDM. Dataset-2 was prepared faster due to streamlining of the workflow and user experience. Moreover, with the exception of a β-carotene at the NdhD1/F1/P interface (Figure A.4b), all lipid densities appear to be diacyl (Figure A.4). The diacyl densities were modelled tentatively as monogalactosyl diacylglycerol (MGDG), the most prevalent and simplest head-group lipid in cyanobacterial thylakoids [211]. However, due to the poor resolution and quality of these diacyl densities, they could represent other diacyl lipids found in cyanobacteria.

Interestingly, the majority of the observed lipids occur in regions specific to the NDH-1L form of the complex and it is tempting to speculate that lipids may play a role in dictating the assembly of the different forms of NDH (e.g., NDH-1L, NDH-1MS and NDH-1MS') in cyanobacteria. Of course, further work is necessary to ascertain the importance of lipids in the structure and function of NDH. To this end, lipidomics and work in native membrane or reconstituted saposin A nanodiscs (MSP discs likely have an incompatible geometry for NDH, which is greatly elongated and irregular in shape compared to most membrane proteins) will greatly facilitate or understanding on the significance of lipids for NDH function.

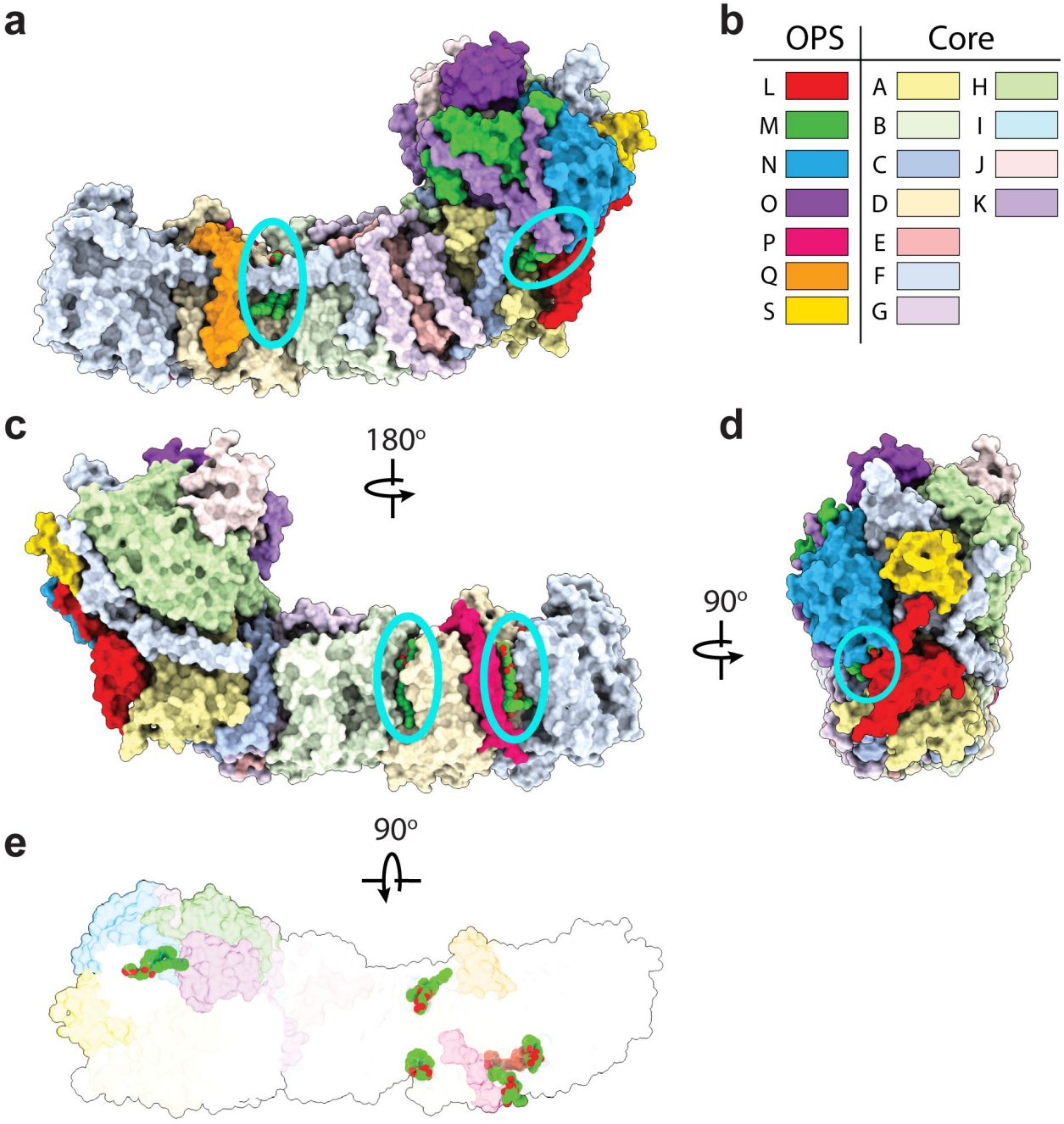


Figure A.3 Overview of co-purified lipids within *T. elongatus*. (**a**, **c-e**) Different orientations of *T. elongatus* structural model (unpublished) with NDH subunits shown as surface representation (transparent in **e**) and colored according to key in **b**. Lipids are shown as spheres and circled in cyan

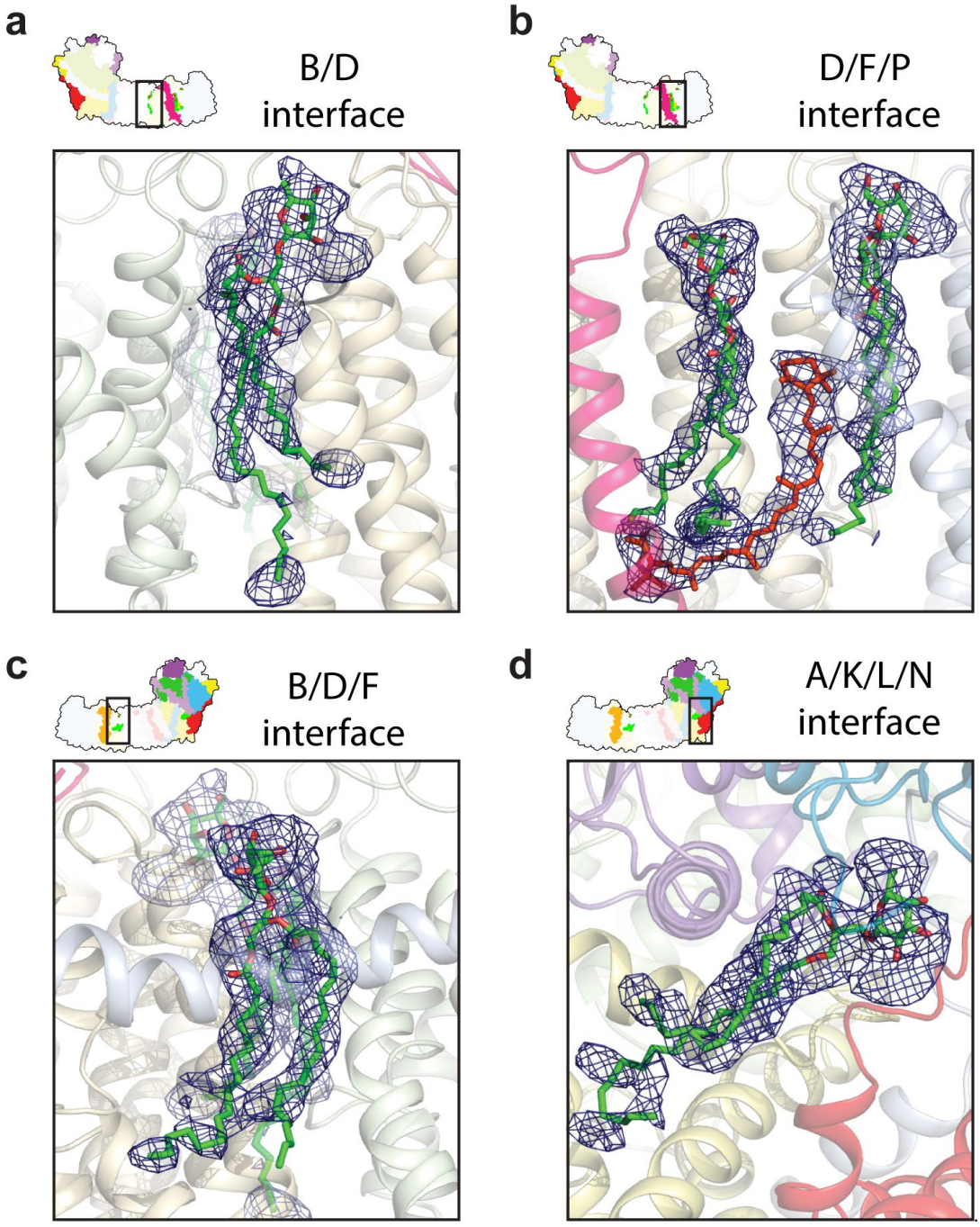


Figure A.4 NDH subunit interfaces containing co-purified lipids. (**a-d**) Unpublished structural model of NDH and co-purified lipids shown as stick representation. Interfaces are the zoom-in of boxed area on NDH silhouette cartoon with each panel and the labelled interface. Density for each lipid is shown as blue mesh from the dataset-2 map (EMD-0425) and segemented 5 Å about the model. Subunits colored as in Figure A.3. Lipids colored by standard heteroatom scheme with carbons for MGDG in green and β-carotene in orange.

Appendix B

Dabs accumulate bicarbonate complex additional figures

†The work presented in this chapter is adapted from the previously published article:
Desmarais JJ, Flamholz AI, Blikstad C, Dugan EJ, Laughlin TG, Oltrogge LM, Chen AW, Wetmore K, Diamond S, Wang JY, Savage DF. (2019) DABs are inorganic carbon pumps found throughout prokaryotic phyla *Nature Microbiology* 4 (12) 2204-2215.

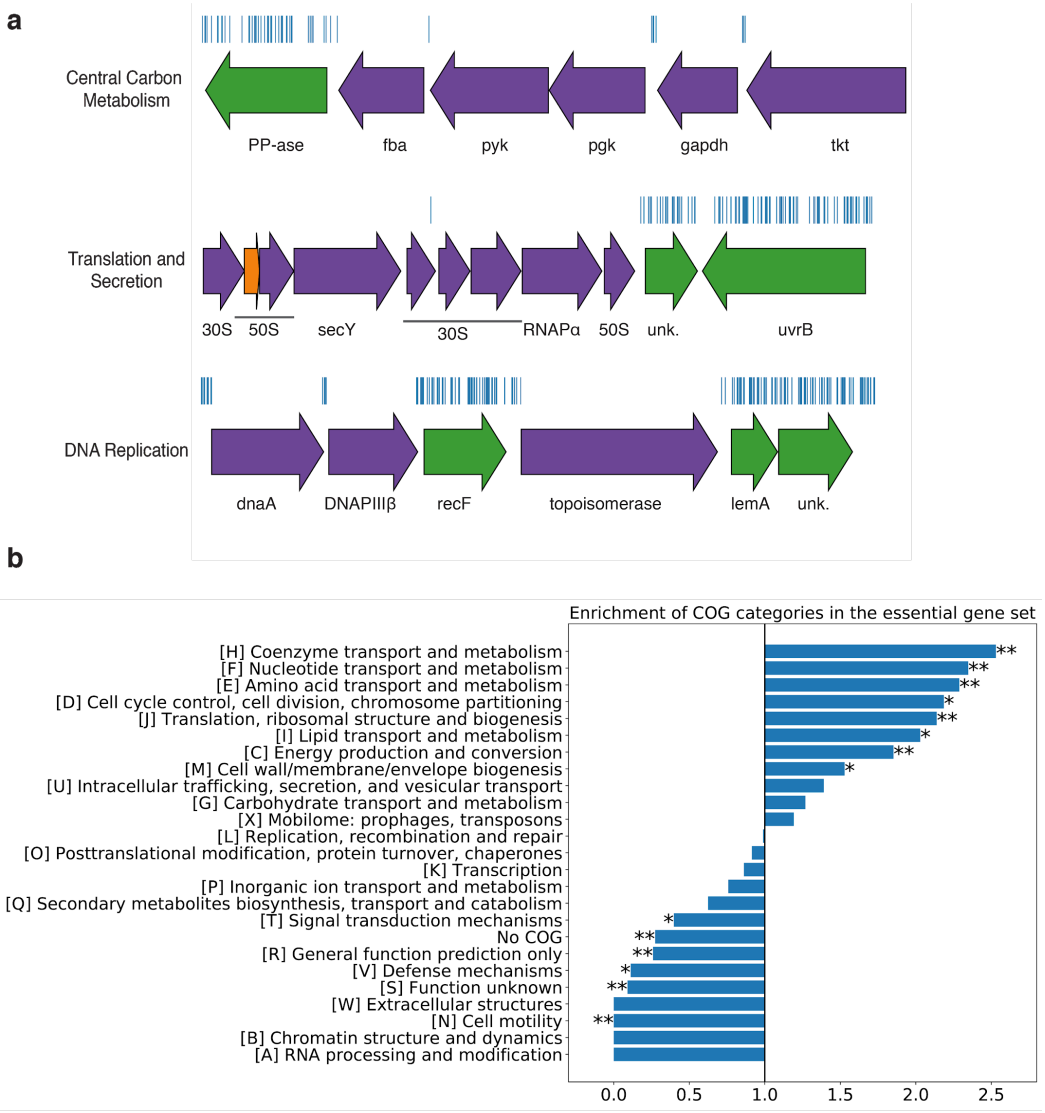


Figure B.1 The essential gene set is enriched for COGs associated with essential cellular processes. (**a**) Representative essential genes and nonessential genes in the *Hnea* genome. The blue track indicates the presence of an insertion. Genes in purple were called essential and genes in green are nonessential. Genes labeled "unk." are hypothetical proteins. The first genomic locus contains 5 essential genes involved in glycolysis or the CBB cycle including pyruvate kinase (pyk) and transketolase (tkt). The 8 essential genes in the second locus encodi 30S and 50S subunits of the ribosome, the secY secretory channel, and an RNA polymerase subunit. Essential genes in the third example locus include topoisomerase and DNA polymerase III β. The following abbreviations are used: exopolyphosphatase (PP-ase), fructose-bisphosphate aldolase class II (fba), pyruvate kinase (pyk), phosphoglycerate kinase (pgk), type I glyceraldehyde-3-phosphate dehydrogenase (gapdh), transketolase (tkt), 30S ribosomal protein (30S), 50S ribosomal protein (50S), preprotein translocase subunit SecY (SecY), DNA-directed RNA polymerase subunit alpha (RNAPα), hypothetical protein (unk.), excinuclease ABC subunit UvrB (UvrB), chromosomal replication initiator protein dnaA (dnaA), DNA polymerase III subunit beta (DNAPIIIβ), DNA replication and repair protein recF (recF), DNA topoisomerase (ATP-hydrolyzing) subunit B (topoisomerase), lemA family protein (LemA). (**b**) COG enrichments were calculated by dividing the fraction of genes in the essential gene set associated with this COG category by the fraction of genes in the genome associated with this category. "*" denotes that this COG is enriched (or depleted) with Bonferroni corrected $p < 0.05$ by a hypergeometric test, and "**" denotes $p < 5X10^{-4}$.

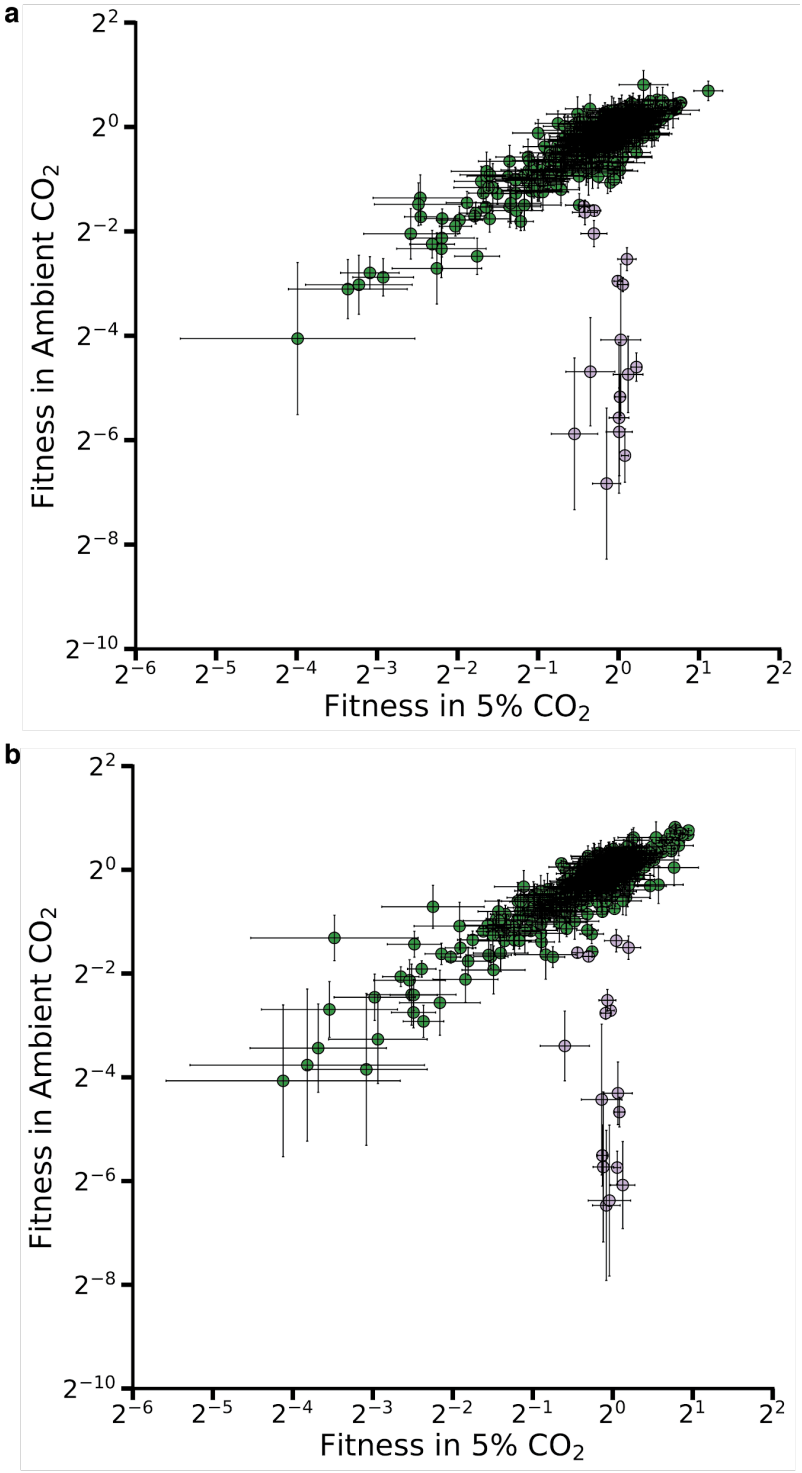


Figure B.2 Gene fitnesses measurements for each replicate. Fitness effects of gene knockouts in 5% CO_2 as compared to ambient CO_2.The effects of single transposon insertions into a gene are averaged to produce the gene-level fitness value plotted. Points represent means of insertions in the middle 80% of the gene. Error bars represent one standard error of the mean. Fitness effects and HCR phenotype by gene. We define HCR mutants as those displaying a twofold fitness defect in ambient CO_2 relative to 5% CO_2. HCR genes are colored light purple. Panel (**a**) contains data from the first replicate experiment and panel (**b**) contains data from the second replicate experiment.

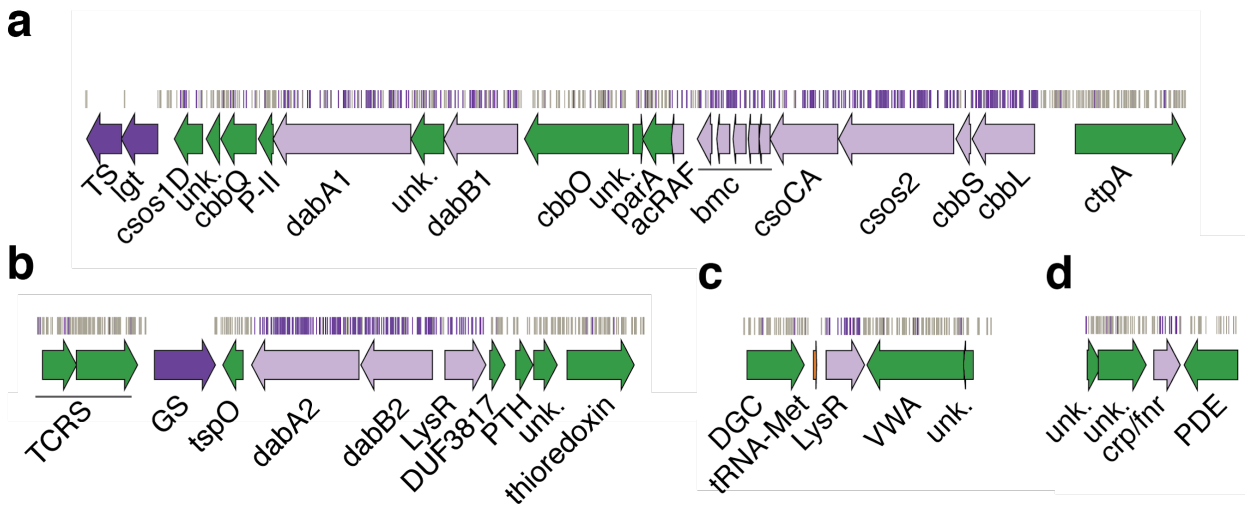


Figure B.3 Genomic context of *Hnea* HCR genes identified in our genome-wide screen. Panels (**a-d**) show regions of the *Hnea* genome containing genes annotated as HCR. Essential genes are in dark purple, HCR genes are in light purple, and other genes are in green. The top tracks show the presence of an insertion in that location. Insertions are colored colored grey unless they display a twofold or greater fitness defect in ambient CO_2, in which case they are colored purple. (**a**) The gene cluster containing the carboxysome operon (HNEAP_RS04660-HNEAP_RS04620) and a second CCM-associated operon. This second operon contains acRAF (HNEAP_RS04615), a FormIC associated cbbOQ-type Rubisco activase (HNEAP_RS04575 and HNEAP_RS04600), parA (HNEAP_RS04610), P-II (HNEAP_RS04580) and dabAB1 (dabA1: HNEAP_RS04585 and dabB1: HNEAP_RS04620). (**b**) The DAB2 operon and surrounding genomic context (lysR: HNEAP_RS01040, dabA2: HNEAP_RS01030, and dabB2: HNEAP_RS01035). (**c**) The genomic context of a lysR-type transcriptional regulator (HNEAP_RS05490) that shows an HCR phenotype. (**d**) Genomic context of a crp/fnr-type transcriptional regulator that displays an HCR phenotype (HNEAP_RS07320). Accession numbers and gi numbers for selected genes can be found in Supplemental Table 1. Abbreviations: thymidylate synthase (TS), prolipoprotein diacylglyceryl transferase (lgt), Rubisco activase Rubisco activase subunits (cbbOQ), nitrogen regulatory protein P-II (P-II), ParA family protein (parA), csos1CAB and csos4AB (bmc), copper-translocating P-type ATPase (ctpA), DNA-binding response regulator and two-component sensor histidine kinase (TCRS), glutamate--ammonia ligase (GS), tryptophan-rich sensory protein (tspO), DUF3817 domain-containing protein (DUF3817), aminoacyl-tRNA hydrolase (PTH), thioredoxin domain-containing protein (thioredoxin), sensor domain-containing diguanylate cyclase (DGC), methionine tRNA (tRNA-Met), VWA domain-containing protein (VWA), diguanylate phosphodiesterase (PDE).

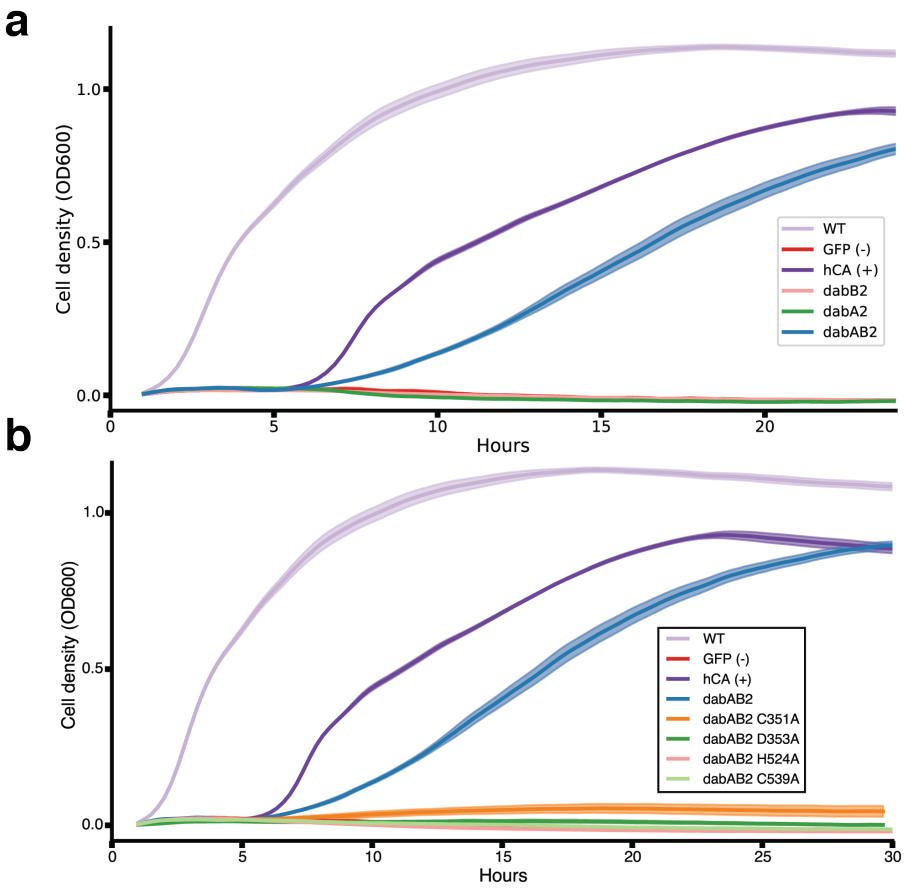


Figure B.4 Expression of DabAB2 rescues growth of CAfree *E. coli* in ambient CO_2. (**a**) These growth curves were used to generate the growth yield values in Figure 3b. Mean OD600 is plotted +/- standard error for four biologically replicate cultures. Wild-type *E. coli* (BW25113) and CAfree strains expressing either dabAB2 or human carbonic anhydrase II (hCA) grow in ambient CO_2 while CAfree expressing GFP, dabB2 alone, or dabA2 alone fail to grow. (**b**) These growth curves were used to generate the growth yield values in Figure 3.4b. Mean OD600 is plotted +/- standard error of four biologically replicate cultures. Wild type cells and CAfree expressing either DabAB2 or human carbonic anhydrase II (hCA) grow robustly. CAfree cells expressing putative active site mutants of DabAB2 (C351, D353, H524, or C539) grow as poorly as the negative control – CAfree expressing sfGFP in the same plasmid backbone.

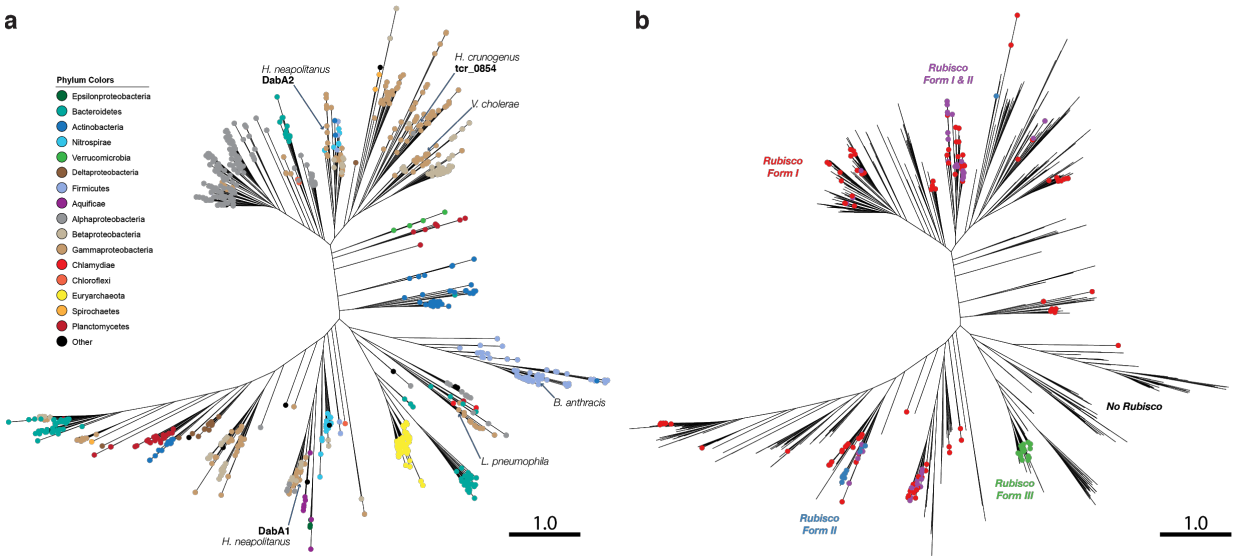


Figure B5. Fully annotated approximate maximum likelihood phylogenetic trees of DabA. (**a**) A phylogenetic tree emphasizing the clades containing high-confidence DabA homologs. DabA homologs are found in > 15 prokaryotic clades, including some archaea. *Hnea* DabA1 and DabA2 represent two different groupings that are commonly found in proteobacteria. The tcr_0854 gene of *H. crunogenus* is more closely related to DabA2 than DabA1. Inspecting the tree reveals several likely incidents of horizontal transfer, e.g. between proteobacteria and Firmicutes, Nitrospirae and Actinobacteria. Moreover, the genomes of several known pathogens contain a high-confidence DabA homolog, including *B. anthracis*, *L. pneumophila*, *V. cholerae*. (**b**) Association of various Rubisco isoforms with DabA homologs. Many organisms that have DabA also have a Rubisco. However, there are numerous examples of DabA homologs that are found in genomes with no Rubisco (denoted by leaves with no colored marking), suggesting that this uptake system might play a role in heterotrophic metabolism. DabA is most-frequently associated with Form I Rubiscos (red and purple leaves in panel B), which is sensible because all known bacterial CCMs involve a Form I Rubisco exclusively. Some DabA-bearing genomes have only a Form II Rubisco (blue) and the Euryarchaeota genomes have that DabA have a Form III Rubisco (green) or none at all. For both panels, scale bars indicate one substitution per site.

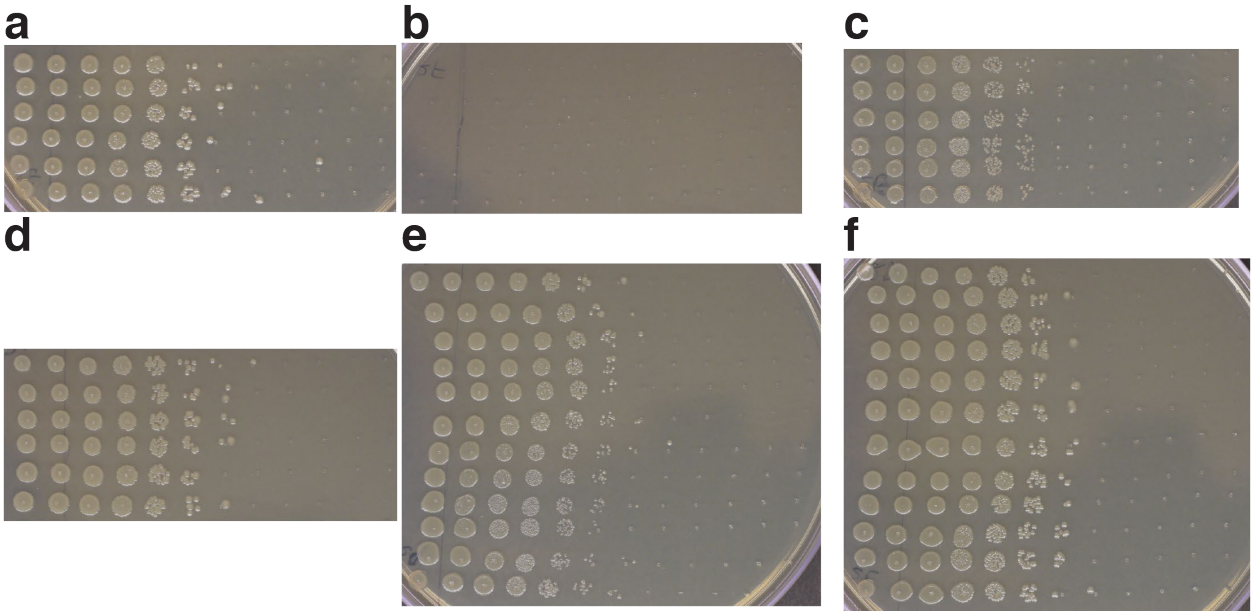


Figure B6. Plates used for determining CFU counts for Figure 3.9b. (**a**) Wt positive control. (**b**) CAfree sfGFP negative control does not rescue. (**c**) CAfree hCA positive control rescues growth. (**d**) CAfree DAB2 rescues growth. (**e**) baDAB from *Bacillus anthracis* rescues growth of CAfree. (**f**) vcDAB from *Vibrio cholera* rescues growth of CAfree. Panels **a-d** represent 6 technical replicates of the plating. Panels **e** and **f** represent 6 technical replicates each of 2 biological replicates. In all panels, the first spot represents 3ul of an OD 0.2 culture grown at 10% CO_2 each subsequent spot is 3 ul of a 1:10 dilution of the previous spot.

Appendix C

Towards a high-resolution structure of the *Hnea* DAB2 complex

C.1 Introduction

Chapter 3 describes the discovery and characterization of the remarkable DABs accumulate bicarbonate complex (DAB) from *Halothiobacillus neapolitanus* c2 (*Hnea*). The work described therein was largely driven by my fellow graduate student John James "Jack" Desmarais, and I certainly feel fortunate to have contributed to that exciting and fun work. I hope Jack can appreciate the subtitles to these sections.

The DABs are one of only two examples of a vectorial carbonic anhydrase (vCA) with the other being the NDH-1$_{3/4}$ complexes from cyanobacteria. While being a vCA is remarkable by its own merits, DABs are widespread throughout prokaryotic phyla and potentially participate in a variety of physiological processes (i.e., not just the CO_2-concentrating mechanism, the context in which they were discovered). Indeed, DABs are present in heterotrophic and even pathogenic bacteria. Moreover, the DABs are much simpler than the NDH-1$_{3/4}$ complexes. A DAB consists of just two proteins (DabA and DabB), in stark contrast to the 18-19 of the NDH-1$_{3/4}$ complexes. Furthermore, DAB function can be fully reconstituted in *E. coli*, unlike the NDH-1$_{3/4}$ complexes. Thus, DABs represent a highly-tractable system to probe the structure-function relationships of a vCA protein and elucidate its biochemical mechanism.

To this end, part of my thesis was dedicated towards determining the structure of the DAB2 complex, which would serve as a foundation for mutational studies. Unfortunately, I was unable to achieve this goal during my time as a graduate student in the Savage Lab. Nevertheless, I am certain it is possible given time, iteration and learning from my failures. Here, I described my attempts towards high-resolution structure determination of the DAB2 complex and rationale (as best as I can remember). In addition, I describe ideas that I think are worth trying but was unable to implement.

C.2 Structure determination of *Hnea* DAB2 complex: The Road So Far

A functional DAB complex consists of a ~90 kDa "soluble" subunit DabA (DUF2309, PF10070) and a ~60 kDa membrane subunit DabB (PF00361). DabA lacks any known structural homologs, with the notable exception of a Zn-binding motif with distant β-CA homology. DabA is "soluble" in the sense that it does not possess predicted transmembrane helices (TMHs), but its expression appears to be contingent on expression of DabB. An attempt to express a *Hnea* MBP-DabA resulted in soluble expression of the fusion, but appeared to bring the entire soluble *E. coli* proteome with it upon subsequent purification steps. Therefore, isolated DabA may be "soluble" but "sticky" and likely exposing hydrophobic patches in a potentially misfolded state. I am personally against pursuing an isolated DabA structure. The integrity of DabA seems contingent on DabB and perhaps the most important information for understanding the enzyme will be at the interface between the two subunits. DabB is related to a ubiquitous family of Mrp-like cation-antiporters with several homologs of known structure. Based on homology DabB is likely ~50 Å in its largest dimension and consists of 12-14 TMHs (13 predicted).

Given my expertise, my efforts towards *Hnea* DAB2 complex structure were focused on using single-particle cryo-electron microscopy (cryo-EM). While the DAB complex is relatively small at ~150 kDa, its size is similar to a number of GPCRs that have been successively resolved to better than 4 Å by cryo-EM and a recent ~3.7 Å structure of a 57 kDa membrane protein by the Brohawn lab bodes well for a high-resolution cryo-EM structure of a DAB [212]. However, crystallography equipment, reagents and expertise have since become available to the Savage Lab and should be considered (structural biology is a lot like good improvisational comedy: "yes, and..."). The solution to the problem is limited by the creativity, time and perseverance of the reader.

On this front, I only made attempts on β-DDM-solubilized *Hnea* DAB2-sfGFP (pJJD060, Figure 3.5a). Purifications from solubilization to grid usually occurred over the course of two days, and consisted of a tandem-affinity purification (Ni-NTA, StrepTactin) followed by size-exclusion chromatography (Superdex-200 3.2/300). The sfGFP-6xHis on the C-terminus of DabB2 greatly facilitates the purification but is presently uncleavable. Furthermore, large losses at the StrepTactin step seem to suggest that the Strep-II tag on the C-terminus of DabA2 may have limited accessibility intrinsically or in the β-DDM micelle. Cryo-specimen were prepared understandard conditions using equipment in Donner Hall (Pelco SC, Vitrobot MarkV) using open-hole Quantifoil Cu grids using a concentration of ~5-10 μM. Images were acquired on the JEOL-3100/K2 in Donner (Figure C.1) or Arctica/K3 in Stanley (Figure C.2). Data from both scopes showed particles ~50-70 Å in diameter, which could be explained by DabB2 and micelle alone. This may suggest that the *Hnea* DAB2-sfGFP particles are denaturing upon freezing. Images from Arctica sessions also showed denatured films in micrographs, suggesting the particle may be sensitive to the air-water interface. Finally, no dataset yielded well-resolved 2D-class averages nor a 3D-reconstruction that I regarded with confidence.

Towards improving the sample, I tested different solubilization strategies for the *Hnea* DAB2-sfGFP sample to reduce detergent-background. At ≥1% (w/v) both β-DDM and LMNG seem to readily solubilize most of the available DAB2-sfGFP from membranes within 3 hours 4ºC with agitation (not shown). β-DDM appears slightly better at solubilization, but LMNG has the advantage of a slow off-rate from the micelle thus enabling techniques like GraDeR that remove free detergent from the sample [213]. Moreover, *Hnea* DAB2-sfGFP readily exchanges into A8-35 amphipols which eliminates all detergent from the sample.

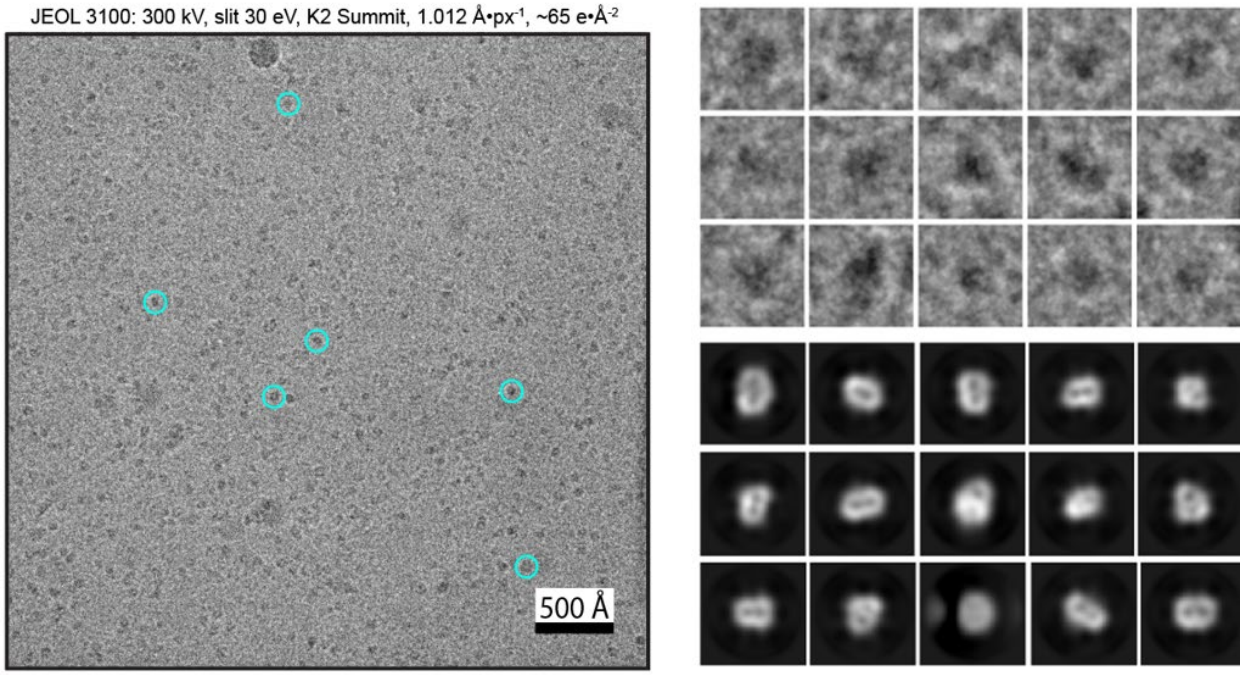


Figure C.1 Cryo-EM of *Hnea* DAB2-sfGFP on the JEOL-3100. Left, example cryo-EM micrograph of DAB2-sfGFP from JEOL-3100 recorded on a K2 Summit camera in counting-mode at conditions listed above micrograph at 1.3 e/Å^2/Fm. Example particles circled in cyan and scale indicate on image. Top right, particle images extracted, normalized and deconvolved to the first peak of their CTF. Bottom right, example unmasked 2D-class averages in a ~140 Å box. A total of 315 movies were recorded in counting-mode at conditions described using a target defocus-range of -1 to -3 µm. Movies were aligned, anisotropy-corrected and exposure-weighted with MotionCor2 and CTF parameters estimated with Gctf. Particles were picked using the general BoxNet2 model in Warp. The particles (127,979) were extracted and binned to ~4 Å/px for 2D-classification in RELION-3.

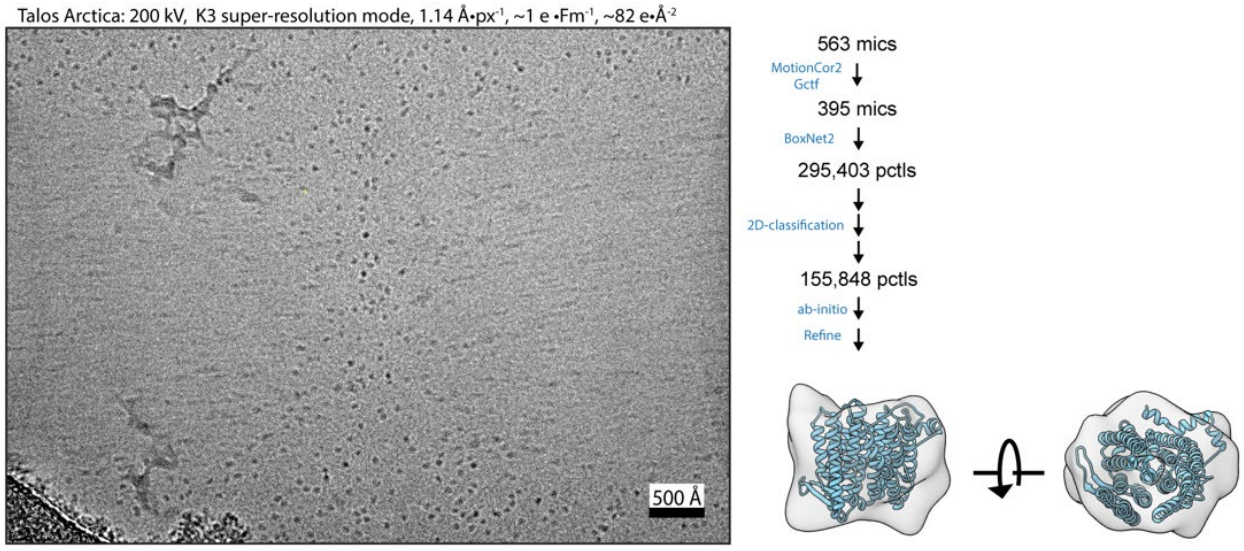


Figure C.2 Cryo-EM of *Hnea* DAB2-sfGFP on the Arctica. Left, example cryo-EM micrograph of DAB2-sfGFP from Talos Arctica recorded on a K3 camera in super-resolution mode at conditions listed above micrograph. Micrographs were collected and pre-processed as described in in Figure C.2. The micrograph shows denatured protein films. Right, processing workflow for the dataset that lead to a questionable *ab initio* and refined reference from cryoSPARC-v2 that can be explained by DabB2 alone (homology model for DabB2 from Phyre2 in blue docked).

C.3 Amphipol-exchange procedure

DAB2-sfGFP (pJJD060) was expressed in BL21-AI by growth in 2xYT+Carb at 37ºC and 250 rpm to OD_{600} ~ 0.4-0.6 at which point the temperature was dropped to 18ºC, 1 g/L culture of arabinose was added to cultures and continued overnight. Cells were harvested and washed once with 20 mM HEPES, pH 7.5, 100 mM NaCl before freezing at -20ºC until use.

Membranes were prepared by thawing cell pellets at RT and resuspension in 20 mM HEPES, pH 7.5, 100 mM NaCl, 20% glycerol, 0.1 mg/mL DNaseI, 1 mM PMSF and dash of lysozyme up to 40 mL. The cell suspension was incubated at RT for 30 min with gentle agitation. Cells were then lysed by passage 3x through a homogenizer at ~17,000 psi. Large cell debris was pelleted by 2 x 15 min centrifugation steps at 15,000g and 4ºC taking the supernatant each time. Membranes were collected by ultracentrifugation at ~100,000g for 1.5 hrs at 4ºC (28,500 rpm in SW32 Ti rotor). Supernatant was discarded and the membrane pellet resuspended in ~1 mL/L culture in 20 mM HEPES, pH 7.5, 100 mM NaCl, 20% glycerol using a magnetic stir bar. Bradford was used to estimate a relative protein concentration for the suspension (i.e., this is not an accurate measure of total protein, but seems internally consistent). The suspended membranes were snap-frozen in liquid-nitrogen and stored at -80ºC until use.

Resuspended membranes from ~1.5 L of culture were thawed in a RT water bath and combined 1:2 with solubilization buffer (premixed 2xBuffer, H_2O, β-DDM) to result in 20 mM HEPES, pH 7.5, 100 mM NaCl, 20% glycerol, 1% β-DDM). The solubilization solution was incubated with gentle agitation for 2 hrs at 4ºC. Unsolubilizated material was pellet in a tabletop ultracentrifuge at 150,000g for 1 hr at 4ºC. Solubilized material was flown over 3 mL of Ni-NTA resin pre-equilibrated in binding buffer (20 mM HEPES, pH 7.5, 300 mM NaCl, 10% glycerol, 0.03% β-DDM) at a rate of ~ 1 mL/min. The column was then washed with 10 CVs of binding buffer + 40 mM imidazole before eluting in 3 x 2 CVs fractions of binding buffer + 300 mM imidazole. Samples were assessed by fluorescence and unboiled SDS-PAGE (Figure C.3). Desired fractions were then pooled, desalted into 20 mM HEPES, pH 7.5, 100 mM NaCl, 10% glycerol, 0.03% β-DDM, and concentrated with a 100 kDa-cutoff spin-column.

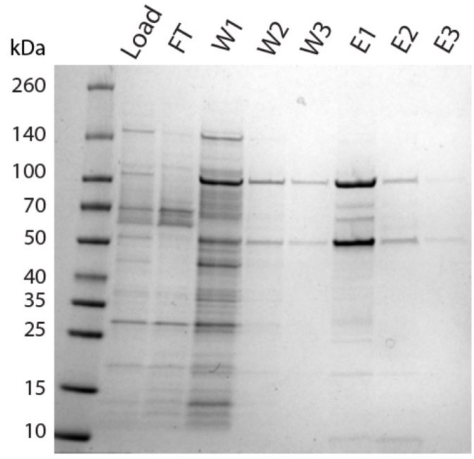


Figure C.3 Example *Hnea* DAB2-sfGFP purification. Unboiled SDS-PAGE (4-20%) FT-flow through, W-wash, E-elution. DabA2 corresponds to band at ~100 kDa in wash and elution lanes. DabB2-sfGFP corresponds to band at ~50 kDa in the wash and elution lanes. Samples left unboiled to avoid irreversible aggregation of DabB2, as well as maintain sfGFP fluorescence in the unstained gel.

Purified DAB2-sfGFP was quantified based on the sfGFP absorbance (ε[sfGFP] ~ 5.5×10^4, MW(DAB2-sfGFP) ~ 1.8×10^5). Approximately 250 µg aliquots were prepared for each condition: negative-control (no amphipol, no detergent removal), positive-control (no amphipol, detergent removal), 3:1, 5:1 and 10:1 amphipol:protein mass ratios. A 10% (w/v) stock of A8-35 amphipol was mixed with H_2O such that consistent volumes were added to all samples. A8-35 solutions were mixed and incubated with each sample at RT for 30 min before adding samples to tubes containing 3 volumes detergent-free buffer and ~20 mg of damp BioBeads (prepared according to manufacturer's protocol). The negative-control was diluted with detergent containing buffer and without BioBeads. Detergent-free buffer was added to the remaining samples to drop the β-DDM below the CMC (~0.0087%). The particular mass of BioBeads yielded a 62.5 mass ratio with the detergent and was chosen for the convenience of massing this amount (not a specific scientific purpose). The samples were incubated with magnetic stir-rods overnight at 4ºC. The exchange process is likely complete when the positive-control is obviously viscous, indicating denaturation of the target protein. This process may occur faster than overnight, which was chosen here for convenience. For SDS-PAGE, 10 µL was taken from each sample (T-total). Remaining supernatant was pelleted by ultracentrifugation at 150,000g for 1 hr at 4ºC and 10 µL of the resulting supernatant taken for SDS-PAGE (S-soluble) (Figure C.4). Successfully exchanged DAB2-sfGFP samples were pooled, concentrated and ran on a Superdex 200 Increase 3.2/300 in detergent-free buffer (Figure C.4). The major peak was pooled, concentrated and ran again on the same column in detergent-free buffer to assess if residual detergent was maintaining protein solubility (Figure C.4). The exchange into A8-35 appears well-tolerated by DAB2-sfGFP and occurs at as low as a 3:1 ratio.

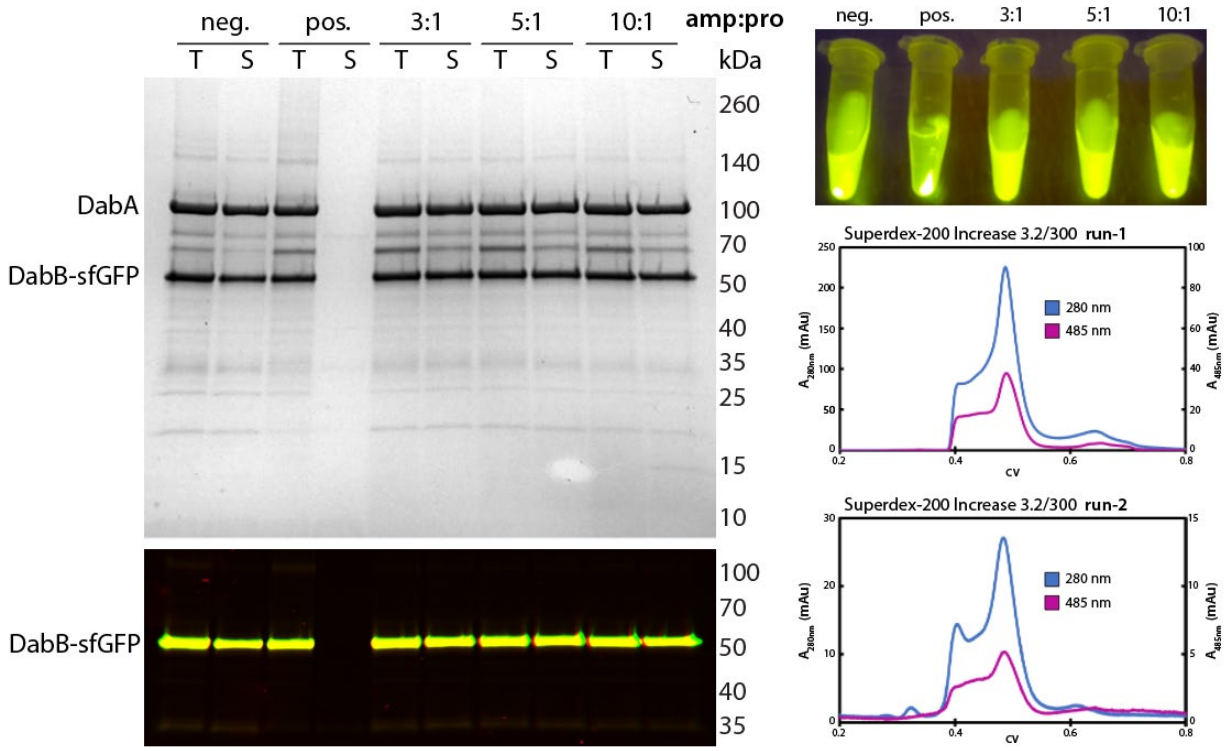


Figure C.4 Amphipol exchange of *Hnea* DAB2-sfGFP. Top left, Coomassie-stained SDS-PAGE of DAB2-sfGFP in A8-35 amphipol. DabA2 and DabB2-sfGFP bands indicated. Negative (neg.) control, positive (pos) control and amphipol:protein (amp:pro) ratios explained in text. T-total, S-soluble. Bottom left, fluorescence image of gel above cropped to the region of DabB2-sfGFP. Top right, ultracentrifuged samples after detergent removal. Middle right, size-exclusion chromatography of DAB2-sfGFP in A8-35 without detergent showing traces for A280nm (general protein) in blue and A485nm (sfGFP) in magenta. Bottom right, major peak from middle right panel ran again as described.

C.4 Structure determination of *Hnea* DAB2 complex: The Road Ahead

The following section is more of a list of approaches that might resolve a DAB complex by cryo-EM, along with perceived advantages and disadvantages. The list is in no particular order. As always, many of these are easier said than done.

Collect more data

The datasets I collected of the DAB were all less than 600 micrographs. It is not unheard of for people to collect over 10,000 micrographs for a particular protein due to intrinsic flexibility, preferred-orientation and/or a low percentage of assembled particles per image. I think the DAB falls into the last case. Therefore, improving the stability of the particle through biochemistry is likely a better option than collecting large amounts of bad data. Collecting more data also means more time and resources during collection and processing.

However, increasingly larger image-shift collection strategies are being used successfully due to better active beam tilt compensation and correction for residual aberrations in post [214–217]. Some people are reporting close to ~7,000 micrographs a day and being limited by write-speed of the data. The Berkeley facility can now readily achieve 3,000 in an overnight session. Whenever the DAB is collected on again, the dataset will likely be much larger than what I previously collected. Just make sure the sample is worth collecting that much data.

Collect at a higher magnification

The DAB is relatively small by EM standards and collecting at higher-magnification should always be better (better signal-to-noise), *ceteris paribus*. The smaller field of view at higher magnification is offset by collecting more images (image-shift collection makes this quite practical). Again, however, I do not think the DAB is limited by the magnification. Michelle Soto Reid of the Brohawn lab successfully reconstructed a 57 kDa membrane protein to ~3.7 Å at 1.137 Å/px. The DAB (~150 kDa) is likely limited by biochemistry.

Crosslinking stabilization

Gentle crosslinking is a common strategy in EM for preserving transient complexes and/or improving particle stability upon freezing (many of the Nogales Lab's nucleic acid binding protein structures are crosslinked). Crosslinking does not appear to limit the achievable resolution of a reconstruction in practice. Popular crosslinkers are glutaraldehyde and BS3, which crosslink primary amines. BS3 appears to be less aggressive than glutaraldehyde and can be added directly to the sample at ~ 1 mM concentration for a few minutes to achieve the desired effect. Glutaraldehyde seems best applied through a GRaFix approach in which the sample exposed to a gradient of crosslinker [218]. GRaFix can be combined with GraDeR for an LMNG-solubilized sample to crosslink and remove excess detergent simultaneously [213,218]. The BS3 crosslinking approach is certainly simpler and probably worth trying first, unless one has experience with GRaFix.

Different solubilization method

The β-DDM-solubilized *Hnea* DAB2 complex is clearly not ideal for cryo-EM. Changing the solubilization method can certainly lead to different behavior on the grid and in the reconstruction either by changing the orientation of the particle, reducing background or improving particle rigidity. An amphipol strategy is described above for A8-35 (other amphipols are available with different properties). A different detergent could also achieve the desired effect. Glyco-diosgenin (GDN) is an increasingly popular synthetic analog of digitonin (still expensive, just less toxic), that appears to preserve native-lipid interactions in a protein-detergent complex better than other popular detergents [192,220]. In addition, a note on detergents, solubilization in my experience works within 1-3 hrs using 0.5-2% (w/v) for most membrane proteins (mitochondrial and cyanobacterial at least). The 20 x CMC rule-of-thumb does not seem accurate for the majority of detergents. Also know that the detergent used for solubilization need not be the detergent used in downstream steps. Finally, lipid nanodiscs in there many different forms are also an alternative strategy. Nanodiscs are routinely employed by the Brohawn lab [196,212], so certainly worth getting protocols and advice from them.

Alter the grid surface

Altering the grid surface is a popular method to preserve particles upon freezing. Continuous-carbon is by far the most popular and easiest to implement to spare a particle from the air-water interface. The main disadvantage of continuous-carbon comes at the expense of increased imaging thickness. Homemade carbon is difficult to work with when it is much less than 5 nm. One method successfully used for Δ4 holocomplex (~177 kDa, $FSC_{0.143}$ ~ 6 Å) was to float ~5 nm carbon onto a batch of grids and glow-discharge the batch. Check one by TEM for carbon integrity. Re-perform glow-discharge on just that one, examine again, glow-discharge all of batch if previous one still looks good. Keep checking the one since it is two glow-discharges ahead of the rest until the carbon blows out (procedure courtesy the Cheng lab). This leaves the majority of batch two glow-discharges away from blow out (leaving one more glow discharge for just before use, which is the thinnest carbon you are going to be able to use).

Alternatively one can use graphene (single-atom thick carbon) or graphene oxide (GO) [220,221]. Preparing these types of grids successfully seems to be more of an art than an exact science. Adam Yokom of the Hurley lab seems to have had success in making relatively consistent GO grids. Graphene has an issue of getting contaminated readily according to Bong-Gyoon “BG” Han and Robert “Bob” Glaeser. Premade graphene grids are also pretty expensive (even by cryo-EM standards).

Affinity grids are another attractive option as they protect particle from the air-water interface, as well as any other non-protein surface (some proteins also denature when adhered to carbon or graphene). Bob Glaeser and BG Han have established a protocol for streptavidin monolayer grids (SA-grids) that immobilize biotinylated particles [222], and David Agard’s group have a protocol for functionalized PEG grids [223]. The SA-grids appear easier to prepare and BG (the expert) is right on campus. The grids are prepared using carbon-film on Au grids (Cu grids will oxidize) on which a monolayer of lipids with biotinylated head-groups is deposited followed by a crystalline layer of streptavidin. A thin layer of carbon is evaporated on the backside of the

grid to preserve the lipid monolayer if the grids are not used immediately or if detergent is used. BG recommended removing detergent from the sample (i.e., use amphipols or nanodiscs) to avoid having to deposit the additional carbon. Biotinylated amphipols are also available if this strategy is used. Of note if using this technique, the streptavidin lattice needs to be subtracted from the experimental images with a script that is not currently implemented Berkeley on-the-fly monitoring set-up (Paul Tobias, system analyst, can likely implement it if asked). The subtraction is quite robust, but the lattice can make it difficult to distinguish small particles from the streptavidin when screening grids.

Alter the construct

Perhaps the most classic of the approaches in the structural biology toolbox is to change the protein expression construct (e.g., homologs, fusions, truncations, etc.). There exists a wide variety of DAB complexes from a variety of organisms with presumably the same enzymatic activity and structure as the *Hnea* DAB2. Some are likely more stable than our favorite DAB complex, thus better candidates for structure determination by cryo-EM or crystallography. Homologs as sfGFP fusions can be conveniently screened in CAfree *E.coli* strain (described in Chapter 3) to ensure the fusions do not perturb protein function. Functional sfGFP fusions can then be conveniently screened for expression and purification in BL21-derived *E.coli* strains by tracking fluorescence. Furthermore, some DABs occur as nature fusions of DabA and DabB proteins, and the *Hnea* DAB2 tolerates direct concatenation without an additional linker. A DabAB fusion is advantageous as it removes possible concerns about DabA:DabB stoichiometry of purified complexes. Unfortunately, the *Hnea* fusion does not purify well at the moment, likely due to accessibility issues for the affinity tag. Finally, making any sfGFP tag used for a DAB construct should be made cleavable. Tag cleavage is a simple, effective purification step. Moreover, the sfGFP (or any large, flexible tag) will behave like noise in single-particle analysis since the mass is likely not coordinated with the complex of interest.

Find a binder

Another classic structural biology approach is to develop a “binder” for the target protein or complex. Popular approaches are to pan for Fabs or nanobodies via phage- or yeast-display. Binders can add rigid-mass to the complex to aid in image alignment, as well as improve general complex stability. The Savage Lab has a nanobody yeast-display library from McMahon *et al.* [224] in the -80^{o}C and a protein-engineering expertise for those interested in this route.

Bibliography

[1] J.A. Raven, Contributions of anoxygenic and oxygenic phototrophy and chemolithotrophy to carbon and oxygen fluxes in aquatic environments, Aquatic Biology. (2009). https://doi.org/10.3354/ab01315.

[2] C.A. Sacksteder, A. Kanazawa, M.E. Jacoby, D.M. Kramer, The proton to electron stoichiometry of steady-state photosynthesis in living plants: A proton-pumping Q cycle is continuously engaged, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 14283–14288. https://doi.org/10.1073/pnas.97.26.14283.

[3] H. Seelert, A. Poetsch, N.A. Dencher, A. Engel, H. Stahlberg, D.J. Müller, Proton-powered turbine of a plant motor, Nature. 405 (2000) 418–419. https://doi.org/10.1038/35013148.

[4] D.M. Kramer, T.J. Avenson, G.E. Edwards, Dynamic flexibility in the light reactions of photosynthesis governed by both electron and proton transfer reactions, Trends in Plant Science. 9 (2004) 349–357. https://doi.org/10.1016/j.tplants.2004.05.001.

[5] D.M. Kramer, J.R. Evans, The Importance of Energy Balance in Improving Photosynthetic Productivity, Plant Physiology. 155 (2011) 70–78. https://doi.org/10.1104/pp.110.166652.

[6] D.I. Arnon, Conversion of Light into Chemical Energy in Photosynthesis, Nature. 184 (1959) 10–21. https://doi.org/10.1038/184010a0.

[7] D.D. Strand, D.M. Kramer, Control of Non-Photochemical Exciton Quenching by the Proton Circuit of Photosynthesis, Advances in Photosynthesis and Respiration. (2014) 387–408. https://doi.org/10.1007/978-94-017-9032-1_18.

[8] G. Peltier, E.-M. Aro, T. Shikanai, NDH-1 and NDH-2 Plastoquinone Reductases in Oxygenic Photosynthesis, Annu. Rev. Plant Biol. 67 (2016) 55–80. https://doi.org/10.1146/annurev-arplant-043014-114752.

[9] D.D. Strand, N. Fisher, D.M. Kramer, The higher plant plastid NAD (P) H dehydrogenase-like complex (NDH) is a high efficiency proton pump that increases ATP production by cyclic electron flow, J. Biol. Chem. (2017). http://www.jbc.org/content/early/2017/05/30/jbc.M116.770792.short.

[10] K. Shinozaki, M. Ohme, M. Tanaka, T. Wakasugi, N. Hayashida, T. Matsubayashi, N. Zaita, J. Chunwongse, J. Obokata, K. Yamaguchi-Shinozaki, C. Ohto, K. Torazawa, B.Y. Meng, M. Sugita, H. Deno, T. Kamogashira, K. Yamada, J. Kusuda, F. Takaiwa, A. Kato, N. Tohdoh, H. Shimada, M. Sugiura, The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression, The EMBO Journal. 5 (1986) 2043–2049. https://doi.org/10.1002/j.1460-2075.1986.tb04464.x.

[11] T. Matsubayashi, T. Wakasugi, K. Shinozaki, K. Yamaguchi-Shinozaki, N. Zaita, T. Hidaka, B.Y. Meng, C. Ohto, M. Tanaka, A. Kato, T. Maruyama, M. Sugiura, Six chloroplast genes (ndhA-F) homologous to human mitochondrial genes encoding components of the respiratory chain NADH dehydrogenase are actively expressed: Determination of the splice sites in ndhA and ndhB pre-mRNAs, Molecular and General Genetics MGG. 210 (1987) 385–393. https://doi.org/10.1007/bf00327187.

[12] K. Ohyama, H. Fukuzawa, T. Kohchi, H. Shirai, T. Sano, S. Sano, K. Umesono, Y. Shiki, M. Takeuchi, Z. Chang, S.-I. Aota, H. Inokuchi, H. Ozeki, Chloroplast gene organization deduced from complete sequence of liverwort Marchantia polymorpha chloroplast DNA, Nature. 322 (1986) 572–574. https://doi.org/10.1038/322572a0.

[13] S. Berger, U. Ellersiek, K. Steinmüller, Cyanobacteria contain a mitochrondrial complex I-homologous NADH-dehydrogenase, FEBS Letters. 286 (1991) 129–132.

https://doi.org/10.1016/0014-5793(91)80957-5.
[14] N. Battchikova, M. Eisenhut, E.-M. Aro, Cyanobacterial NDH-1 complexes: novel insights and remaining puzzles, Biochim. Biophys. Acta. 1807 (2011) 935–944. https://doi.org/10.1016/j.bbabio.2010.10.017.
[15] P. Prommeenate, A.M. Lennon, C. Markert, M. Hippler, P.J. Nixon, Subunit composition of NDH-1 complexes of Synechocystis sp. PCC 6803: identification of two new ndh gene products with nuclear-encoded homologues in the chloroplast Ndh complex, J. Biol. Chem. 279 (2004) 28165–28173. https://doi.org/10.1074/jbc.M401107200.
[16] D. Rumeau, N. Bécuwe-Linka, A. Beyly, M. Louwagie, J. Garin, G. Peltier, New subunits NDH-M, -N, and -O, encoded by nuclear genes, are essential for plastid Ndh complex functioning in higher plants, Plant Cell. 17 (2005) 219–232. https://doi.org/10.1105/tpc.104.028282.
[17] N. Battchikova, P. Zhang, S. Rudd, T. Ogawa, E.-M. Aro, Identification of NdhL and Ssl1690 (NdhO) in NDH-1L and NDH-1M Complexes ofSynechocystissp. PCC 6803, Journal of Biological Chemistry. 280 (2005) 2587–2595. https://doi.org/10.1074/jbc.m410914200.
[18] N. Battchikova, L. Wei, L. Du, L. Bersanini, E.-M. Aro, W. Ma, Identification of novel Ssl0352 protein (NdhS), essential for efficient operation of cyclic electron transport around photosystem I, in NADPH:plastoquinone oxidoreductase (NDH-1) complexes of Synechocystis sp. PCC 6803, J. Biol. Chem. 286 (2011) 36992–37001. https://doi.org/10.1074/jbc.M111.263780.
[19] H. Yamamoto, L. Peng, Y. Fukao, T. Shikanai, An Src homology 3 domain-like fold protein forms a ferredoxin binding site for the chloroplast NADH dehydrogenase-like complex in Arabidopsis, Plant Cell. 23 (2011) 1480–1493. https://doi.org/10.1105/tpc.110.080291.
[20] H. Yamamoto, T. Shikanai, In planta mutagenesis of Src homology 3 domain-like fold of NdhS, a ferredoxin-binding subunit of the chloroplast NADH dehydrogenase-like complex in Arabidopsis: a conserved Arg-193 plays a critical role in ferredoxin binding, J. Biol. Chem. 288 (2013) 36328–36337. https://doi.org/10.1074/jbc.M113.511584.
[21] S. Veit, A. Nagadoi, M. Rögner, S. Rexroth, R. Stoll, T. Ikegami, The cyanobacterial cytochrome b 6 f subunit PetP adopts an SH3 fold in solution, Biochimica et Biophysica Acta (BBA) - Bioenergetics. 1857 (2016) 705–714. https://doi.org/10.1016/j.bbabio.2016.03.023.
[22] Z. He, F. Zheng, Y. Wu, Q. Li, J. Lv, P. Fu, H. Mi, NDH-1L interacts with ferredoxin via the subunit NdhS in Thermosynechococcus elongatus, Photosynth. Res. 126 (2015) 341–349. https://doi.org/10.1007/s11120-015-0090-4.
[23] L.A. Sazanov, A giant molecular proton pump: structure and mechanism of respiratory complex I, Nat. Rev. Mol. Cell Biol. 16 (2015) 375–388. https://doi.org/10.1038/nrm3997.
[24] T. Ogawa, A gene homologous to the subunit-2 gene of NADH dehydrogenase is essential to inorganic carbon transport of Synechocystis PCC6803, Proceedings of the National Academy of Sciences. 88 (1991) 4275–4279. https://doi.org/10.1073/pnas.88.10.4275.
[25] B. Klughammer, D. Sultemeyer, M.R. Badger, G. Dean Price, The involvement of NAD(P)H dehydrogenase subunits, NdhD3 and NdhF3, in high-affinity CO2 uptake in Synechococcus sp. PCC7002 gives evidence for multiple NDH-1 complexes with specific roles in cyanobacteria, Molecular Microbiology. 32 (1999) 1305–1315. https://doi.org/10.1046/j.1365-2958.1999.01457.x.
[26] H. Ohkawa, H.B. Pakrasi, T. Ogawa, Two Types of Functionally Distinct NAD(P)H

Dehydrogenases inSynechocystissp. Strain PCC6803, Journal of Biological Chemistry. 275 (2000) 31630–31634. https://doi.org/10.1074/jbc.m003706200.
[27] G.G.B. Tcherkez, G.D. Farquhar, T.J. Andrews, Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 7246–7251. https://doi.org/10.1073/pnas.0600605103.
[28] Y. Savir, E. Noor, R. Milo, T. Tlusty, Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape, Proceedings of the National Academy of Sciences. 107 (2010) 3475–3480. https://doi.org/10.1073/pnas.0911663107.
[29] A.I. Flamholz, N. Prywes, U. Moran, D. Davidi, Y.M. Bar-On, L.M. Oltrogge, R. Alves, D. Savage, R. Milo, Revisiting Trade-offs between Rubisco Kinetic Parameters, Biochemistry. 58 (2019) 3365–3376. https://doi.org/10.1021/acs.biochem.9b00237.
[30] N.M. Mangan, A. Flamholz, R.D. Hood, R. Milo, D.F. Savage, pH determines the energetic efficiency of the cyanobacterial CO2 concentrating mechanism, Proceedings of the National Academy of Sciences. 113 (2016) E5354–E5362. https://doi.org/10.1073/pnas.1525145113.
[31] B.D. Rae, B.M. Long, M.R. Badger, G.D. Price, Functions, Compositions, and Evolution of the Two Types of Carboxysomes: Polyhedral Microcompartments That Facilitate CO2 Fixation in Cyanobacteria and Some Proteobacteria, Microbiology and Molecular Biology Reviews. 77 (2013) 357–379. https://doi.org/10.1128/mmbr.00061-12.
[32] L. Whitehead, B.M. Long, G. Dean Price, M.R. Badger, Comparing the in Vivo Function of α-Carboxysomes and β-Carboxysomes in Two Model Cyanobacteria, Plant Physiology. 165 (2014) 398–411. https://doi.org/10.1104/pp.114.237941.
[33] G.D. Price, G. Dean Price, S.-I. Maeda, T. Omata, M.R. Badger, Modes of active inorganic carbon uptake in the cyanobacterium, Synechococcus sp. PCC7942, Functional Plant Biology. 29 (2002) 131. https://doi.org/10.1071/pp01229.
[34] M.R. Badger, G.D. Price, CO2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution, J. Exp. Bot. 54 (2003) 609–622. https://doi.org/10.1093/jxb/erg076.
[35] H. Ohkawa, M. Sonoda, H. Katoh, T. Ogawa, The use of mutants in the analysis of the CO2-concentrating mechanism in cyanobacteria, Canadian Journal of Botany. 76 (1998) 1035–1042. https://doi.org/10.1139/b98-076.
[36] T. Ogawa, Identification and Characterization of the ictA/ndhL Gene Product Essential to Inorganic Carbon Transport of Synechocystis PCC6803, Plant Physiol. 99 (1992) 1604–1608. https://doi.org/10.1104/pp.99.4.1604.
[37] H. Ohkawa, H.B. Pakrasi, T. Ogawa, Two types of functionally distinct NAD(P)H dehydrogenases in Synechocystis sp. strain PCC6803, J. Biol. Chem. 275 (2000) 31630–31634. https://doi.org/10.1074/jbc.M003706200.
[38] H. Ohkawa, G. Dean Price, M.R. Badger, T. Ogawa, Mutation of ndh Genes Leads to Inhibition of CO2 Uptake Rather than HCO3 −Uptake in Synechocystis sp. Strain PCC 6803, Journal of Bacteriology. 182 (2000) 2591–2596. https://doi.org/10.1128/jb.182.9.2591-2596.2000.
[39] M. Shibata, H. Ohkawa, T. Kaneko, H. Fukuzawa, S. Tabata, A. Kaplan, T. Ogawa, Distinct constitutive and low-CO2-induced CO2 uptake systems in cyanobacteria: genes involved and their phylogenetic relationship with homologous genes in other organisms, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 11789–11794. https://doi.org/10.1073/pnas.191258298.
[40] S.-I. Maeda, M.R. Badger, G. Dean Price, Novel gene products associated with NdhD3/D4-

containing NDH-1 complexes are involved in photosynthetic CO2 hydration in the cyanobacterium, Synechococcus sp. PCC7942, Molecular Microbiology. 43 (2002) 425–435. https://doi.org/10.1046/j.1365-2958.2002.02753.x.
[41] P. Zhang, N. Battchikova, T. Jansen, J. Appel, T. Ogawa, E.-M. Aro, 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 (2004) 3326–3340. https://doi.org/10.1105/tpc.104.026526.
[42] I.M. Folea, P. Zhang, M.M. Nowaczyk, T. Ogawa, E.-M. Aro, E.J. Boekema, Single particle analysis of thylakoid proteins from Thermosynechococcus elongatus and Synechocystis 6803: localization of the CupA subunit of NDH-1, FEBS Lett. 582 (2008) 249–254. https://doi.org/10.1016/j.febslet.2007.12.012.
[43] M. Birungi, M. Folea, N. Battchikova, M. Xu, H. Mi, T. Ogawa, E.-M. Aro, E.J. Boekema, Possibilities of subunit localization with fluorescent protein tags and electron microscopy examplified by a cyanobacterial NDH-1 study, Biochim. Biophys. Acta. 1797 (2010) 1681–1686. https://doi.org/10.1016/j.bbabio.2010.06.004.
[44] A. Kaplan, L. Reinhold, CO2CONCENTRATING MECHANISMS IN PHOTOSYNTHETIC MICROORGANISMS, Annual Review of Plant Physiology and Plant Molecular Biology. 50 (1999) 539–570. https://doi.org/10.1146/annurev.arplant.50.1.539.
[45] G.D. Price, M.R. Badger, Isolation and Characterization of High CO(2)-Requiring-Mutants of the Cyanobacterium Synechococcus PCC7942 : Two Phenotypes that Accumulate Inorganic Carbon but Are Apparently Unable to Generate CO(2) within the Carboxysome, Plant Physiol. 91 (1989) 514–525. https://doi.org/10.1104/pp.91.2.514.
[46] C.T. Supuran, Structure and function of carbonic anhydrases, Biochemical Journal. 473 (2016) 2023–2032. https://doi.org/10.1042/bcj20160115.
[47] D.I. Arnon, The light reactions of photosynthesis, Proc. Natl. Acad. Sci. U. S. A. 68 (1971) 2883–2892. https://www.ncbi.nlm.nih.gov/pubmed/4400251.
[48] T. Shikanai, Chloroplast NDH: A different enzyme with a structure similar to that of respiratory NADH dehydrogenase, Biochim. Biophys. Acta. 1857 (2016) 1015–1022. https://doi.org/10.1016/j.bbabio.2015.10.013.
[49] R. Baradaran, J.M. Berrisford, G.S. Minhas, L.A. Sazanov, Crystal structure of the entire respiratory complex I, Nature. 494 (2013) 443–448. https://doi.org/10.1038/nature11871.
[50] V. Zickermann, C. Wirth, H. Nasiri, K. Siegmund, H. Schwalbe, C. Hunte, U. Brandt, Structural biology. Mechanistic insight from the crystal structure of mitochondrial complex I, Science. 347 (2015) 44–49. https://doi.org/10.1126/science.1259859.
[51] J. Zhu, K.R. Vinothkumar, J. Hirst, Structure of mammalian respiratory complex I, Nature. 536 (2016) 354–358. https://doi.org/10.1038/nature19095.
[52] X. Fan, J. Zhang, W. Li, L. Peng, The NdhV subunit is required to stabilize the chloroplast NADH dehydrogenase-like complex in Arabidopsis, Plant J. 82 (2015) 221–231. https://doi.org/10.1111/tpj.12807.
[53] X. Chen, Z. He, M. Xu, L. Peng, H. Mi, NdhV subunit regulates the activity of type-1 NAD(P)H dehydrogenase under high light conditions in cyanobacterium Synechocystis sp. PCC 6803, Sci. Rep. 6 (2016) 28361. https://doi.org/10.1038/srep28361.
[54] A.A. Arteni, P. Zhang, N. Battchikova, T. Ogawa, E.-M. Aro, E.J. Boekema, Structural characterization of NDH-1 complexes of Thermosynechococcus elongatus by single particle electron microscopy, Biochim. Biophys. Acta. 1757 (2006) 1469–1475.

https://doi.org/10.1016/j.bbabio.2006.05.042.
[55] H. Wulfhorst, L.E. Franken, T. Wessinghage, E.J. Boekema, M.M. Nowaczyk, The 5 kDa protein NdhP is essential for stable NDH-1L assembly in Thermosynechococcus elongatus, PLoS One. 9 (2014) e103584. https://doi.org/10.1371/journal.pone.0103584.
[56] H. Shimizu, L. Peng, F. Myouga, R. Motohashi, K. Shinozaki, T. Shikanai, CRR23/NdhL is a subunit of the chloroplast NAD(P)H dehydrogenase complex in Arabidopsis, Plant Cell Physiol. 49 (2008) 835–842. https://doi.org/10.1093/pcp/pcn058.
[57] M.M. Nowaczyk, H. Wulfhorst, C.M. Ryan, P. Souda, H. Zhang, W.A. Cramer, J.P. Whitelegge, NdhP and NdhQ: two novel small subunits of the cyanobacterial NDH-1 complex, Biochemistry. 50 (2011) 1121–1124. https://doi.org/10.1021/bi102044b.
[58] J. Zhang, F. Gao, J. Zhao, T. Ogawa, Q. Wang, W. Ma, NdhP is an exclusive subunit of large complex of NADPH dehydrogenase essential to stabilize the complex in Synechocystis sp. strain PCC 6803, J. Biol. Chem. 289 (2014) 18770–18781. https://doi.org/10.1074/jbc.M114.553404.
[59] H. Kubota-Kawai, R. Mutoh, K. Shinmura, P. Sétif, M.M. Nowaczyk, M. Rögner, T. Ikegami, H. Tanaka, G. Kurisu, X-ray structure of an asymmetrical trimeric ferredoxin-photosystem I complex, Nat Plants. 4 (2018) 218–224. https://doi.org/10.1038/s41477-018-0130-0.
[60] G. Kurisu, M. Kusunoki, E. Katoh, T. Yamazaki, K. Teshima, Y. Onda, Y. Kimata-Ariga, T. Hase, Structure of the electron transfer complex between ferredoxin and ferredoxin-NADP+ reductase, Nat. Struct. Biol. 8 (2001) 117. https://doi.org/10.1038/84097.
[61] S. Dai, R. Friemann, D.A. Glauser, F. Bourquin, W. Manieri, P. Schürmann, H. Eklund, Structural snapshots along the reaction pathway of ferredoxin–thioredoxin reductase, Nature. 448 (2007) 92. https://doi.org/10.1038/nature05937.
[62] A.P. Srivastava, E.P. Hardy, J.P. Allen, B.J. Vaccaro, M.K. Johnson, D.B. Knaff, Identification of the Ferredoxin-Binding Site of a Ferredoxin-Dependent Cyanobacterial Nitrate Reductase, Biochemistry. 56 (2017) 5582–5592. https://doi.org/10.1021/acs.biochem.7b00025.
[63] C. Wirth, U. Brandt, C. Hunte, V. Zickermann, Structure and function of mitochondrial complex I, Biochim. Biophys. Acta. 1857 (2016) 902–914. https://doi.org/10.1016/j.bbabio.2016.02.013.
[64] M. Ekberg, M. Sahlin, M. Eriksson, B.-M. Sjöberg, Two Conserved Tyrosine Residues in Protein R1 Participate in an Intermolecular Electron Transfer in Ribonucleotide Reductase, Journal of Biological Chemistry. 271 (1996) 20655–20659. https://doi.org/10.1074/jbc.271.34.20655.
[65] H. Ishikita, E.-W. Knapp, Function of redox-active tyrosine in photosystem II, Biophys. J. 90 (2006) 3886–3896. https://doi.org/10.1529/biophysj.105.076984.
[66] P. Zhang, N. Battchikova, V. Paakkarinen, H. Katoh, M. Iwai, M. Ikeuchi, H.B. Pakrasi, T. Ogawa, E.-M. Aro, Isolation, subunit composition and interaction of the NDH-1 complexes from Thermosynechococcus elongatus BP-1, Biochem. J. 390 (2005) 513–520. https://doi.org/10.1042/BJ20050390.
[67] J. Kern, B. Loll, C. Lüneberg, D. DiFiore, J. Biesiadka, K.-D. Irrgang, A. Zouni, Purification, characterisation and crystallisation of photosystem II from Thermosynechococcus elongatus cultivated in a new type of photobioreactor, Biochim. Biophys. Acta. 1706 (2005) 147–157. https://doi.org/10.1016/j.bbabio.2004.10.007.
[68] M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities

of protein utilizing the principle of protein-dye binding, Anal. Biochem. 72 (1976) 248–254. https://www.ncbi.nlm.nih.gov/pubmed/942051.
[69] I. Wittig, H.-P. Braun, H. Schägger, Blue native PAGE, Nat. Protoc. 1 (2006) 418–428. https://doi.org/10.1038/nprot.2006.62.
[70] D. Wessel, U.I. Flügge, A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids, Anal. Biochem. 138 (1984) 141–143. https://www.ncbi.nlm.nih.gov/pubmed/6731838.
[71] K. Gevaert, J. Vandekerckhove, In-Gel Digestion of Protein Spots for Mass Spectrometry, Cell Biology. (2006) 379–382. https://doi.org/10.1016/b978-012164730-8/50233-1.
[72] S. Beck, A. Michalski, O. Raether, M. Lubeck, The impact II, a very high resolution quadrupole time-of-flight instrument for deep shotgun proteomics, Molecular & Cellular. (2015). http://www.mcponline.org/content/early/2015/05/19/mcp.M114.047407.abstract.
[73] J. Cox, N. Neuhauser, A. Michalski, R.A. Scheltema, J.V. Olsen, M. Mann, Andromeda: a peptide search engine integrated into the MaxQuant environment, J. Proteome Res. 10 (2011) 1794–1805. https://doi.org/10.1021/pr101065j.
[74] D.N. Mastronarde, Automated electron microscope tomography using robust prediction of specimen movements, J. Struct. Biol. 152 (2005) 36–51. https://doi.org/10.1016/j.jsb.2005.07.007.
[75] N. Biyani, R.D. Righetto, R. McLeod, D. Caujolle-Bert, D. Castano-Diez, K.N. Goldie, H. Stahlberg, Focus: The interface between data collection and data processing in cryo-EM, J. Struct. Biol. 198 (2017) 124–133. https://doi.org/10.1016/j.jsb.2017.03.007.
[76] S.Q. Zheng, E. Palovcak, J.-P. Armache, K.A. Verba, Y. Cheng, D.A. Agard, MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy, Nat. Methods. 14 (2017) 331–332. https://doi.org/10.1038/nmeth.4193.
[77] K. Zhang, Gctf: Real-time CTF determination and correction, J. Struct. Biol. 193 (2016) 1–12. https://doi.org/10.1016/j.jsb.2015.11.003.
[78] D. Kimanius, B.O. Forsberg, S.H.W. Scheres, E. Lindahl, Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2, Elife. 5 (2016) e18722. https://doi.org/10.7554/eLife.18722.
[79] A. Punjani, J.L. Rubinstein, D.J. Fleet, M.A. Brubaker, cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination, Nat. Methods. 14 (2017) 290–296. https://doi.org/10.1038/nmeth.4169.
[80] X.C. Bai, E. Rajendra, G.H. Yang, Y.G. Shi, S.H.W. Scheres, Cryo-EM structure of gamma secretase in class 1 of the apo- state ensemble, (2015). https://doi.org/10.2210/pdb5fn3/pdb.
[81] J. Zivanov, T. Nakane, B.O. Forsberg, D. Kimanius, W.J. Hagen, E. Lindahl, S.H. Scheres, New tools for automated high-resolution cryo-EM structure determination in RELION-3, Elife. 7 (2018). https://doi.org/10.7554/eLife.42166.
[82] J. Zivanov, T. Nakane, S.H.W. Scheres, A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis, IUCrJ. 6 (2019) 5–17. https://doi.org/10.1107/S205225251801463X.
[83] S. Chen, G. McMullan, A.R. Faruqi, G.N. Murshudov, J.M. Short, S.H.W. Scheres, R. Henderson, High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy, Ultramicroscopy. 135 (2013) 24–35. https://doi.org/10.1016/j.ultramic.2013.06.004.
[84] P.B. Rosenthal, R. Henderson, Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy, J. Mol. Biol. 333 (2003) 721–

745. https://www.ncbi.nlm.nih.gov/pubmed/14568533.
[85] Y.Z. Tan, P.R. Baldwin, J.H. Davis, J.R. Williamson, C.S. Potter, B. Carragher, D. Lyumkis, Addressing preferred specimen orientation in single-particle cryo-EM through tilting, Nat. Methods. 14 (2017) 793–796. https://doi.org/10.1038/nmeth.4347.
[86] T.D. Goddard, C.C. Huang, T.E. Ferrin, Visualizing density maps with UCSF Chimera, J. Struct. Biol. 157 (2007) 281–287. https://doi.org/10.1016/j.jsb.2006.06.010.
[87] Y. Zhang, I-TASSER server for protein 3D structure prediction, BMC Bioinformatics. 9 (2008) 40. https://doi.org/10.1186/1471-2105-9-40.
[88] L.A. Kelley, S. Mezulis, C.M. Yates, M.N. Wass, M.J.E. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis, Nat. Protoc. 10 (2015) 845–858. https://doi.org/10.1038/nprot.2015.053.
[89] L.J. McGuffin, K. Bryson, D.T. Jones, The PSIPRED protein structure prediction server, Bioinformatics. 16 (2000) 404–405. https://www.ncbi.nlm.nih.gov/pubmed/10869041.
[90] P. Gouet, E. Courcelle, D.I. Stuart, F. Métoz, ESPript: analysis of multiple sequence alignments in PostScript, Bioinformatics. 15 (1999) 305–308. https://doi.org/10.1093/bioinformatics/15.4.305.
[91] P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics, Acta Crystallogr. D Biol. Crystallogr. 60 (2004) 2126–2132. https://doi.org/10.1107/S0907444904019158.
[92] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of Coot, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 486–501. https://doi.org/10.1107/S0907444910007493.
[93] P.D. Adams, P.V. Afonine, G. Bunkóczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.-W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger, P.H. Zwart, PHENIX: a comprehensive Python-based system for macromolecular structure solution, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 213–221. https://doi.org/10.1107/S0907444909052925.
[94] P.V. Afonine, B.K. Poon, R.J. Read, O.V. Sobolev, T.C. Terwilliger, A. Urzhumtsev, P.D. Adams, Real-space refinement in PHENIX for cryo-EM and crystallography, Acta Crystallogr D Struct Biol. 74 (2018) 531–544. https://doi.org/10.1107/S2059798318006551.
[95] V.B. Chen, W.B. Arendall 3rd, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson, D.C. Richardson, MolProbity: all-atom structure validation for macromolecular crystallography, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 12–21. https://doi.org/10.1107/S0907444909042073.
[96] DELANO, W. L, The PyMOL Molecular Graphics System, Http://www.pymol.org. (2002). https://ci.nii.ac.jp/naid/10020095229/ (accessed August 24, 2018).
[97] N.A. Baker, D. Sept, S. Joseph, M.J. Holst, J.A. McCammon, Electrostatics of nanosystems: application to microtubules and the ribosome, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 10037–10041. https://doi.org/10.1073/pnas.181342398.
[98] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J.D. Thompson, D.G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, Mol. Syst. Biol. 7 (2011) 539. https://doi.org/10.1038/msb.2011.75.
[99] The UniProt Consortium, UniProt: the universal protein knowledgebase, Nucleic Acids Res. 45 (2017) D158–D169. https://doi.org/10.1093/nar/gkw1099.
[100] C. Bathellier, G. Tcherkez, G.H. Lorimer, G.D. Farquhar, Rubisco is not really so bad,

Plant Cell Environ. 41 (2018) 705–716. https://doi.org/10.1111/pce.13149.
[101] A. Flamholz, N. Prywes, U. Moran, D. Davidi, Y. Bar-On, L. Oltrogge, D. Savage, R. Milo, Revisiting tradeoffs in Rubisco kinetic parameters, bioRxiv. (2018) 470021. https://doi.org/10.1101/470021.
[102] G. Tcherkez, The mechanism of Rubisco-catalysed oxygenation, Plant Cell Environ. 39 (2016) 983–997. https://onlinelibrary.wiley.com/doi/abs/10.1111/pce.12629.
[103] H. Bauwe, M. Hagemann, A.R. Fernie, Photorespiration: players, partners and origin, Trends Plant Sci. 15 (2010) 330–336. https://doi.org/10.1016/j.tplants.2010.03.006.
[104] Y. Savir, E. Noor, R. Milo, T. Tlusty, Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 3475–3480. https://doi.org/10.1073/pnas.0911663107.
[105] N.M. Mangan, A. Flamholz, R.D. Hood, R. Milo, D.F. Savage, pH determines the energetic efficiency of the cyanobacterial CO2 concentrating mechanism, Proc. Natl. Acad. Sci. U. S. A. 113 (2016) E5354–62. https://doi.org/10.1073/pnas.1525145113.
[106] J.A. Raven, J. Beardall, P. Sánchez-Baracaldo, The possible evolution and future of CO2-concentrating mechanisms, J. Exp. Bot. 68 (2017) 3701–3716. https://doi.org/10.1093/jxb/erx110.
[107] B.D. Rae, B.M. Long, M.R. Badger, G.D. Price, Functions, compositions, and evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate CO2 fixation in cyanobacteria and some proteobacteria, Microbiol. Mol. Biol. Rev. 77 (2013) 357–379. https://doi.org/10.1128/MMBR.00061-12.
[108] B.M. Long, B.D. Rae, V. Rolland, B. Förster, G.D. Price, Cyanobacterial CO2-concentrating mechanism components: function and prospects for plant metabolic engineering, Curr. Opin. Plant Biol. 31 (2016) 1–8. https://doi.org/10.1016/j.pbi.2016.03.002.
[109] G.D. Price, M.R. Badger, S. von Caemmerer, The prospect of using cyanobacterial bicarbonate transporters to improve leaf photosynthesis in C3 crop plants, Plant Physiol. 155 (2011) 20–26. https://doi.org/10.1104/pp.110.164681.
[110] J.M. McGrath, S.P. Long, Can the cyanobacterial carbon-concentrating mechanism increase photosynthesis in crop species? A theoretical analysis, Plant Physiol. 164 (2014) 2247–2261. https://doi.org/10.1104/pp.113.232611.
[111] B.M. Long, W.Y. Hee, R.E. Sharwood, B.D. Rae, S. Kaines, Y.-L. Lim, N.D. Nguyen, B. Massey, S. Bala, S. von Caemmerer, M.R. Badger, G.D. Price, Carboxysome encapsulation of the CO2-fixing enzyme Rubisco in tobacco chloroplasts, Nat. Commun. 9 (2018) 3570. https://doi.org/10.1038/s41467-018-06044-0.
[112] G.D. Price, M.R. Badger, Expression of Human Carbonic Anhydrase in the Cyanobacterium Synechococcus PCC7942 Creates a High CO2-Requiring Phenotype Evidence for a Central Role for Carboxysomes in the CO2 Concentrating Mechanism, Plant Physiol. 91 (1989) 505–513. http://www.plantphysiol.org/content/91/2/505.short.
[113] B.M. Hopkinson, J.N. Young, a. L. Tansik, B.J. Binder, The Minimal CO2-Concentrating Mechanism of Prochlorococcus spp. MED4 Is Effective and Efficient, Plant Physiol. 166 (2014) 2205–2217. https://doi.org/10.1104/pp.114.247049.
[114] L. Whitehead, B.M. Long, G.D. Price, M.R. Badger, Comparing the in Vivo Function of α-Carboxysomes and β-Carboxysomes in Two Model Cyanobacteria, Plant Physiol. 165 (2014) 398–411. https://doi.org/10.1104/pp.114.237941.
[115] Y.A. Holthuijzen, F.F.M. van Dissel-Emiliani, J.G. Kuenen, W.N. Konings, Energetic

aspects of CO2 uptake in Thiobacillus neapolitanus, Arch. Microbiol. 147 (1987) 285–290. https://doi.org/10.1007/BF00463489.

[116] G.D. Price, M.R. Badger, Isolation and characterization of high CO2-requiring-mutants of the cyanobacterium Synechococcus PCC7942: two phenotypes that accumulate inorganic carbon but are apparently unable to generate CO2 within the carboxysome, Plant Physiol. 91 (1989) 514–525. http://www.plantphysiol.org/content/91/2/514.short.

[117] Y. Marcus, R. Schwarz, D. Friedberg, A. Kaplan, High CO(2) Requiring Mutant of Anacystis nidulans R(2), Plant Physiol. 82 (1986) 610–612. https://www.ncbi.nlm.nih.gov/pubmed/16665079.

[118] W. Bonacci, P.K. Teng, B. Afonso, H. Niederholtmeyer, P. Grob, P. a. Silver, D.F. Savage, Modularity of a carbon-fixing protein organelle, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 478–483. https://doi.org/10.1073/pnas.1108557109.

[119] J. Jorda, D. Lopez, N.M. Wheatley, T.O. Yeates, Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria, Protein Sci. 22 (2013) 179–195. https://doi.org/10.1002/pro.2196.

[120] S.D. Axen, O. Erbilgin, C.A. Kerfeld, A taxonomy of bacterial microcompartment loci constructed by a novel scoring method, PLoS Comput. Biol. 10 (2014) e1003898. https://doi.org/10.1371/journal.pcbi.1003898.

[121] M. Shibata, H. Ohkawa, H. Katoh, M. Shimoyama, T. Ogawa, Two CO2 uptake systems in cyanobacteria: four systems for inorganic carbon acquisition in Synechocystis sp. strain PCC6803, Funct. Plant Biol. 29 (2002) 123–129. https://doi.org/10.1071/pp01188.

[122] G.D. Price, Inorganic carbon transporters of the cyanobacterial CO2 concentrating mechanism, Photosynth. Res. 109 (2011) 47–57. https://doi.org/10.1007/s11120-010-9608-y.

[123] S. Heinhorst, G.C. Cannon, J.M. Shively, Carboxysomes and Carboxysome-like Inclusions, in: J.M. Shively (Ed.), Complex Intracellular Structures in Prokaryotes, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006: pp. 141–165. https://doi.org/10.1007/7171_023.

[124] J.M. Shively, F. Ball, D.H. Brown, R.E. Saunders, Functional organelles in prokaryotes: polyhedral inclusions (carboxysomes) of Thiobacillus neapolitanus, Science. 182 (1973) 584–586. https://doi.org/10.1126/science.182.4112.584.

[125] G.C. Cannon, C.E. Bradburne, H.C. Aldrich, S.H. Baker, S. Heinhorst, J.M. Shively, Microcompartments in prokaryotes: carboxysomes and related polyhedra, Appl. Environ. Microbiol. 67 (2001) 5351–5361. https://doi.org/10.1128/AEM.67.12.5351-5361.2001.

[126] M. Mangiapia, USF MCB4404L, T.-R.W. Brown, D. Chaput, E. Haller, T.L. Harmer, Z. Hashemy, R. Keeley, J. Leonard, P. Mancera, D. Nicholson, S. Stevens, P. Wanjugi, T. Zabinski, C. Pan, K.M. Scott, Proteomic and mutant analysis of the CO2 concentrating mechanism of hydrothermal vent chemolithoautotroph Thiomicrospira crunogena, J. Bacteriol. (2017). https://doi.org/10.1128/JB.00871-16.

[127] K.M. Scott, J. Williams, C.M.B. Porter, S. Russel, T.L. Harmer, J.H. Paul, K.M. Antonen, M.K. Bridges, G.J. Camper, C.K. Campla, L.G. Casella, E. Chase, J.W. Conrad, M.C. Cruz, D.S. Dunlap, L. Duran, E.M. Fahsbender, D.B. Goldsmith, R.F. Keeley, M.R. Kondoff, B.I. Kussy, M.K. Lane, S. Lawler, B.A. Leigh, C. Lewis, L.M. Lostal, D. Marking, P.A. Mancera, E.C. McClenthan, E.A. McIntyre, J.A. Mine, S. Modi, B.D. Moore, W.A. Morgan, K.M. Nelson, K.N. Nguyen, N. Ogburn, D.G. Parrino, A.D. Pedapudi, R.P. Pelham, A.M. Preece, E.A. Rampersad, J.C. Richardson, C.M. Rodgers, B.L. Schaffer, N.E.

Sheridan, M.R. Solone, Z.R. Staley, M. Tabuchi, R.J. Waide, P.W. Wanjugi, S. Young, A. Clum, C. Daum, M. Huntemann, N. Ivanova, N. Kyrpides, N. Mikhailova, K. Palaniappan, M. Pillay, T.B.K. Reddy, N. Shapiro, D. Stamatis, N. Varghese, T. Woyke, R. Boden, S.K. Freyermuth, C.A. Kerfeld, Genomes of ubiquitous marine and hypersaline Hydrogenovibrio, Thiomicrorhabdus and Thiomicrospira spp. encode a diversity of mechanisms to sustain chemolithoautotrophy in heterogeneous environments, Environ. Microbiol. 20 (2018) 2686–2708. https://doi.org/10.1111/1462-2920.14090.

[128] K.M. Scott, J.M. Leonard, R. Boden, D. Chaput, C. Dennison, E. Haller, T.L. Harmer, A. Anderson, T. Arnold, S. Budenstein, R. Brown, J. Brand, J. Byers, J. Calarco, T. Campbell, E. Carter, M. Chase, M. Cole, D. Dwyer, J. Grasham, C. Hanni, A. Hazle, C. Johnson, R. Johnson, B. Kirby, K. Lewis, B. Neumann, T. Nguyen, J. Nino Charari, O. Morakinyo, B. Olsson, S. Roundtree, E. Skjerve, A. Ubaldini, R. Whittaker, Diversity in CO2-Concentrating Mechanisms among Chemolithoautotrophs from the Genera Hydrogenovibrio, Thiomicrorhabdus, and Thiomicrospira, Ubiquitous in Sulfidic Habitats Worldwide, Appl. Environ. Microbiol. 85 (2019). https://doi.org/10.1128/AEM.02096-18.

[129] K.M. Wetmore, M.N. Price, R.J. Waters, J.S. Lamson, J. He, C.A. Hoover, M.J. Blow, J. Bristow, G. Butland, A.P. Arkin, A. Deutschbauer, Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons, MBio. 6 (2015) e00306–15. https://doi.org/10.1128/mBio.00306-15.

[130] T. Chaijarasphong, R.J. Nichols, K.E. Kortright, C.F. Nixon, P.K. Teng, L.M. Oltrogge, D.F. Savage, Programmed Ribosomal Frameshifting Mediates Expression of the α-Carboxysome, J. Mol. Biol. 428 (2016) 153–164. https://doi.org/10.1016/j.jmb.2015.11.017.

[131] F. Cai, B.B. Menon, G.C. Cannon, K.J. Curry, J.M. Shively, S. Heinhorst, The pentameric vertex proteins are necessary for the icosahedral carboxysome shell to function as a CO2 leakage barrier, PLoS One. 4 (2009) e7521. https://doi.org/10.1371/journal.pone.0007521.

[132] E.W. Roberts, F. Cai, C.A. Kerfeld, G.C. Cannon, S. Heinhorst, Isolation and characterization of the Prochlorococcus carboxysome reveal the presence of the novel shell protein CsoS1D, J. Bacteriol. 194 (2012) 787–795. https://doi.org/10.1128/JB.06444-11.

[133] N.M. Wheatley, C.D. Sundberg, S.D. Gidaniyan, D. Cascio, T.O. Yeates, Structure and Identification of a Pterin Dehydratase-like Protein as a Ribulose-bisphosphate Carboxylase/Oxygenase (RuBisCO) Assembly Factor in the α-Carboxysome, J. Biol. Chem. 289 (2014) 7973–7981. https://doi.org/10.1074/jbc.M113.531236.

[134] H. Aigner, R.H. Wilson, A. Bracher, L. Calisse, J.Y. Bhat, F.U. Hartl, M. Hayer-Hartl, Plant RuBisCo assembly in E. coli with five chloroplast chaperones including BSD2, Science. 358 (2017) 1272–1278. https://doi.org/10.1126/science.aap9221.

[135] O. Mueller-Cajar, The Diverse AAA+ Machines that Repair Inhibited Rubisco Active Sites, Front Mol Biosci. 4 (2017) 31. https://doi.org/10.3389/fmolb.2017.00031.

[136] T.A. Krulwich, D.B. Hicks, M. Ito, Cation/proton antiporter complements of bacteria: why so large and diverse?, Mol. Microbiol. 74 (2009) 257–260. https://doi.org/10.1111/j.1365-2958.2009.06842.x.

[137] C. Merlin, M. Masters, Why is carbonic anhydrase essential to Escherichia coli?, J. Bacteriol. 185 (2003). https://doi.org/10.1128/JB.185.21.6415.

[138] J. Du, B. Förster, L. Rourke, S.M. Howitt, G.D. Price, Characterisation of Cyanobacterial Bicarbonate Transporters in E. coli Shows that SbtA Homologs Are Functional in This

Heterologous Expression System, PLoS One. 9 (2014) e115905. https://doi.org/10.1371/journal.pone.0115905.
[139] J.D. Cronk, J. a. Endrizzi, M.R. Cronk, J.W. O'neill, K.Y. Zhang, Crystal structure of E. coli beta-carbonic anhydrase, an enzyme with an unusual pH-dependent activity, Protein Sci. 10 (2001) 911–922. https://doi.org/10.1110/ps.46301.
[140] V.M. Krishnamurthy, G.K. Kaufman, A.R. Urbach, I. Gitlin, K.L. Gudiksen, D.B. Weibel, G.M. Whitesides, Carbonic anhydrase as a model for biophysical and physical-organic studies of proteins and protein- ligand binding, Chem. Rev. 108 (2008) 946–1051. https://pubs.acs.org/doi/abs/10.1021/cr050262p.
[141] J.D. Cronk, R.S. Rowlett, K.Y.J. Zhang, C. Tu, J.A. Endrizzi, J. Lee, P.C. Gareiss, J.R. Preiss, Identification of a novel noncatalytic bicarbonate binding site in eubacterial beta-carbonic anhydrase, Biochemistry. 45 (2006) 4351–4361. https://doi.org/10.1021/bi052272q.
[142] S.-I. Maeda, M.R. Badger, G.D. Price, Novel gene products associated with NdhD3/D4-containing NDH-1 complexes are involved in photosynthetic CO2 hydration in the cyanobacterium, Synechococcus sp. PCC7942, Mol. Microbiol. 43 (2002) 425–435. https://febs.onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2958.2002.02753.x.
[143] N. Antonovsky, S. Gleizer, E. Noor, Y. Zohar, E. Herz, U. Barenholz, L. Zelcbuch, S. Amram, A. Wides, N. Tepper, D. Davidi, Y. Bar-On, T. Bareia, D.G. Wernick, I. Shani, S. Malitsky, G. Jona, A. Bar-Even, R. Milo, Sugar Synthesis from CO2 in Escherichia coli, Cell. 166 (2016) 115–125. https://doi.org/10.1016/j.cell.2016.05.064.
[144] J. Aguilera, J.P. Van Dijken, J.H. De Winde, J.T. Pronk, Carbonic anhydrase (Nce103p): an essential biosynthetic enzyme for growth of Saccharomyces cerevisiae at atmospheric carbon dioxide pressure, Biochem. J. 391 (2005) 311–316. https://doi.org/10.1042/BJ20050556.
[145] J.C. Sirard, M. Mock, A. Fouet, The three Bacillus anthracis toxin genes are coordinately regulated by bicarbonate and temperature, J. Bacteriol. 176 (1994) 5188–5192. https://www.ncbi.nlm.nih.gov/pubmed/8051039.
[146] B.H. Abuaita, J.H. Withey, Bicarbonate Induces Vibrio cholerae virulence gene expression by enhancing ToxT activity, Infect. Immun. 77 (2009) 4111–4120. https://doi.org/10.1128/IAI.00409-09.
[147] T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. a. Datsenko, M. Tomita, B.L. Wanner, H. Mori, Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection, Mol. Syst. Biol. 2 (2006) 2006.0008. https://doi.org/10.1038/msb4100050.
[148] B.E. Rubin, K.M. Wetmore, M.N. Price, S. Diamond, R.K. Shultzaberger, L.C. Lowe, G. Curtin, A.P. Arkin, A. Deutschbauer, S.S. Golden, The essential gene set of a photosynthetic organism, Proceedings of the National Academy of Sciences. (2015) 201519220. https://doi.org/10.1073/pnas.1519220112.
[149] B.L. Oakes, D.C. Nadler, D.F. Savage, Chapter Twenty-Three - Protein Engineering of Cas9 for Enhanced Function, in: J.A.D.A.E.J. Sontheimer (Ed.), Methods in Enzymology, Academic Press, 2014: pp. 491–511. https://www.sciencedirect.com/science/article/pii/B9780128011850000246 (accessed February 6, 2017).
[150] K.P. Dobrinski, D.L. Longo, K.M. Scott, The Carbon-Concentrating Mechanism of the Hydrothermal Vent Chemolithoautotroph Thiomicrospira crunogena, J. Bacteriol. 187

(2005) 5761–5766. https://doi.org/10.1128/JB.187.16.5761-5766.2005.
[151] R. Zallot, N.O. Oberg, J.A. Gerlt, “Democratized” genomic enzymology web tools for functional assignment, Curr. Opin. Chem. Biol. 47 (2018) 77–85. https://doi.org/10.1016/j.cbpa.2018.09.009.
[152] P.S. Dehal, M.P. Joachimiak, M.N. Price, J.T. Bates, J.K. Baumohl, D. Chivian, G.D. Friedland, K.H. Huang, K. Keller, P.S. Novichkov, I.L. Dubchak, E.J. Alm, A.P. Arkin, MicrobesOnline: an integrated portal for comparative and functional genomics, Nucleic Acids Res. 38 (2010) D396–400. https://doi.org/10.1093/nar/gkp919.
[153] F. Sievers, D.G. Higgins, Clustal Omega for making accurate alignments of many protein sequences, Protein Sci. 27 (2018) 135–145. https://doi.org/10.1002/pro.3290.
[154] M.N. Price, P.S. Dehal, A.P. Arkin, FastTree: computing large minimum evolution trees with profiles instead of a distance matrix, Mol. Biol. Evol. 26 (2009) 1641–1650. https://doi.org/10.1093/molbev/msp077.
[155] I. Letunic, P. Bork, Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees, Nucleic Acids Res. 44 (2016) W242–5. https://doi.org/10.1093/nar/gkw290.
[156] L. Slabinski, L. Jaroszewski, L. Rychlewski, I.A. Wilson, S.A. Lesley, A. Godzik, XtalPred: a web server for prediction of protein crystallizability, Bioinformatics. 23 (2007) 3403–3405. https://doi.org/10.1093/bioinformatics/btm477.
[157] A. Roy, A. Kucukural, Y. Zhang, I-TASSER: a unified platform for automated protein structure and function prediction, Nat. Protoc. 5 (2010) 725–738. https://doi.org/10.1038/nprot.2010.5.
[158] Z.E.R. Newby, J.D. O’Connell 3rd, F. Gruswitz, F.A. Hays, W.E.C. Harries, I.M. Harwood, J.D. Ho, J.K. Lee, D.F. Savage, L.J.W. Miercke, R.M. Stroud, A general protocol for the crystallization of membrane proteins for X-ray structural investigation, Nat. Protoc. 4 (2009) 619–637. https://doi.org/10.1038/nprot.2009.27.
[159] R.G. Khalifah, The Carbon Dioxide Hydration Activity of Carbonic Anhydrase, J. Biol. Chem. 246 (1971) 2561–2573. http://www.jbc.org/content/246/8/2561.
[160] F. Garcia-Pichel, J. Belnap, S. Neuer, F. Schanz, Estimates of global cyanobacterial biomass and its distribution, Algological Studies. 109 (2003) 213–227. https://doi.org/10.1127/1864-1318/2003/0109-0213.
[161] T.O. Yeates, C.A. Kerfeld, S. Heinhorst, G.C. Cannon, J.M. Shively, Protein-based organelles in bacteria: carboxysomes and related microcompartments, Nat. Rev. Microbiol. 6 (2008) 681–691. https://doi.org/10.1038/nrmicro1913.
[162] C.A. Kerfeld, C. Aussignargues, J. Zarzycki, F. Cai, M. Sutter, Bacterial microcompartments, Nat. Rev. Microbiol. 16 (2018) 277–290. https://doi.org/10.1038/nrmicro.2018.10.
[163] G. Drews, W. Niklowitz, [Cytology of Cyanophycea. II. Centroplasm and granular inclusions of Phormidium uncinatum], Arch. Mikrobiol. 24 (1956) 147–162. https://www.ncbi.nlm.nih.gov/pubmed/13327992.
[164] Y.A. Holthuijzen, J. Gijs Kuenen, W.N. Konings, Activity of ribulose-1,5-bisphosphate carboxylase in intact and disrupted carboxysomes ofThiobacillus neapolitanus, FEMS Microbiology Letters. 42 (1987) 121–124. https://doi.org/10.1111/j.1574-6968.1987.tb02057.x.
[165] C.A. Kerfeld, M.R. Sawaya, S. Tanaka, C.V. Nguyen, M. Phillips, M. Beeby, T.O. Yeates, Protein structures forming the shell of primitive bacterial organelles, Science. 309

(2005) 936–938. https://doi.org/10.1126/science.1113397.
[166] D. Heldt, S. Frank, A. Seyedarabi, D. Ladikis, J.B. Parsons, M.J. Warren, R.W. Pickersgill, Structure of a trimeric bacterial microcompartment shell protein, EtuB, associated with ethanol utilization in Clostridium kluyveri, Biochemical Journal. 423 (2009) 199–207. https://doi.org/10.1042/bj20090780.
[167] M.G. Klein, P. Zwart, S.C. Bagby, F. Cai, S.W. Chisholm, S. Heinhorst, G.C. Cannon, C.A. Kerfeld, Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport, J. Mol. Biol. 392 (2009) 319–333. https://doi.org/10.1016/j.jmb.2009.03.056.
[168] M. Sagermann, A. Ohtaki, K. Nikolakakis, Crystal structure of the EutL shell protein of the ethanolamine ammonia lyase microcompartment, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 8883–8887. https://doi.org/10.1073/pnas.0902324106.
[169] C.S. Crowley, D. Cascio, M.R. Sawaya, J.S. Kopstein, T.A. Bobik, T.O. Yeates, Structural insight into the mechanisms of transport across the Salmonella enterica Pdu microcompartment shell, J. Biol. Chem. 285 (2010) 37838–37846. https://doi.org/10.1074/jbc.M110.160580.
[170] S. Tanaka, C.A. Kerfeld, M.R. Sawaya, F. Cai, S. Heinhorst, G.C. Cannon, T.O. Yeates, Atomic-level models of the bacterial carboxysome shell, Science. 319 (2008) 1083–1086. https://doi.org/10.1126/science.1151458.
[171] M. Sutter, B. Greber, C. Aussignargues, C.A. Kerfeld, Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell, Science. 356 (2017) 1293–1297. https://doi.org/10.1126/science.aan3289.
[172] C.A. Kerfeld, M.R. Melnicki, Assembly, function and evolution of cyanobacterial carboxysomes, Curr. Opin. Plant Biol. 31 (2016) 66–75. https://doi.org/10.1016/j.pbi.2016.03.009.
[173] G.D. Price, S.M. Howitt, K. Harrison, M.R. Badger, Analysis of a genomic DNA region from the cyanobacterium Synechococcus sp. strain PCC7942 involved in carboxysome assembly and function, J. Bacteriol. 175 (1993) 2871–2879. https://doi.org/10.1128/jb.175.10.2871-2879.1993.
[174] M. Sommer, F. Cai, M. Melnicki, C.A. Kerfeld, β-Carboxysome bioinformatics: identification and evolution of new bacterial microcompartment protein gene classes and core locus constraints, J. Exp. Bot. 68 (2017) 3841–3855. https://doi.org/10.1093/jxb/erx115.
[175] H. Wang, X. Yan, H. Aigner, A. Bracher, N.D. Nguyen, W.Y. Hee, B.M. Long, G.D. Price, F.U. Hartl, M. Hayer-Hartl, Rubisco condensate formation by CcmM in β-carboxysome biogenesis, Nature. 566 (2019) 131–135. https://doi.org/10.1038/s41586-019-0880-5.
[176] J.C. Cameron, S.C. Wilson, S.L. Bernstein, C.A. Kerfeld, Biogenesis of a Bacterial Organelle: The Carboxysome Assembly Pathway, Cell. 155 (2013) 1131–1140. https://doi.org/10.1016/j.cell.2013.10.044.
[177] F. Cai, S.L. Bernstein, S.C. Wilson, C.A. Kerfeld, Production and Characterization of Synthetic Carboxysome Shells with Incorporated Luminal Proteins, Plant Physiol. 170 (2016) 1868–1877. https://doi.org/10.1104/pp.15.01822.
[178] A. Luque, D. Reguera, The structure of elongated viral capsids, Biophys. J. 98 (2010) 2993–3003. https://doi.org/10.1016/j.bpj.2010.02.051.
[179] B.J. Greber, M. Sutter, C.A. Kerfeld, The Plasticity of Molecular Interactions Governs

Bacterial Microcompartment Shell Assembly, Structure. 27 (2019) 749–763.e4. https://doi.org/10.1016/j.str.2019.01.017.
[180] B.D. Rae, B.M. Long, M.R. Badger, G. Dean Price, Structural Determinants of the Outer Shell of β-Carboxysomes in Synechococcus elongatus PCC 7942: Roles for CcmK2, K3-K4, CcmO, and CcmL, PLoS ONE. 7 (2012) e43871. https://doi.org/10.1371/journal.pone.0043871.
[181] F. Hinzpeter, U. Gerland, F. Tostevin, Optimal Compartmentalization Strategies for Metabolic Microcompartments, Biophys. J. 112 (2017) 767–779. https://doi.org/10.1016/j.bpj.2016.11.3194.
[182] S. Tsitkov, H. Hess, Design Principles for a Compartmentalized Enzyme Cascade Reaction, ACS Catalysis. 9 (2019) 2432–2439. https://doi.org/10.1021/acscatal.8b04419.
[183] A. Hagen, M. Sutter, N. Sloan, C.A. Kerfeld, Programmed loading and rapid purification of engineered bacterial microcompartment shells, Nat. Commun. 9 (2018) 2881. https://doi.org/10.1038/s41467-018-05162-z.
[184] M. Sommer, M. Sutter, S. Gupta, H. Kirst, A. Turmo, S. Lechno-Yossef, R.L. Burton, C. Saechao, N.B. Sloan, X. Cheng, L.-J.G. Chan, C.J. Petzold, M. Fuentes-Cabrera, C.Y. Ralston, C.A. Kerfeld, Heterohexamers Formed by CcmK3 and CcmK4 Increase the Complexity of Beta Carboxysome Shells, Plant Physiol. 179 (2019) 156–167. https://doi.org/10.1104/pp.18.01190.
[185] F. Schulz, E.A. Eloe-Fadrosh, R.M. Bowers, J. Jarett, T. Nielsen, N.N. Ivanova, N.C. Kyrpides, T. Woyke, Towards a balanced view of the bacterial tree of life, Microbiome. 5 (2017) 140. https://doi.org/10.1186/s40168-017-0360-9.
[186] S. Capella-Gutierrez, J.M. Silla-Martinez, T. Gabaldon, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses, Bioinformatics. 25 (2009) 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.
[187] M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins, Clustal W and Clustal X version 2.0, Bioinformatics. 23 (2007) 2947–2948. https://doi.org/10.1093/bioinformatics/btm404.
[188] S. Guindon, J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, O. Gascuel, New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0, Syst. Biol. 59 (2010) 307–321. https://doi.org/10.1093/sysbio/syq010.
[189] M.V. Han, C.M. Zmasek, phyloXML: XML for evolutionary biology and comparative genomics, BMC Bioinformatics. 10 (2009). https://doi.org/10.1186/1471-2105-10-356.
[190] J.M. Schuller, J.A. Birrell, H. Tanaka, T. Konuma, H. Wulfhorst, N. Cox, S.K. Schuller, J. Thiemann, W. Lubitz, P. Sétif, T. Ikegami, B.D. Engel, G. Kurisu, M.M. Nowaczyk, Structural adaptations of photosynthetic complex I enable ferredoxin-dependent electron transfer, Science. 363 (2019) 257–260. https://doi.org/10.1126/science.aau3613.
[191] X. Pan, D. Cao, F. Xie, F. Xu, X. Su, H. Mi, et al., Structural basis for electron transport mechanism of complex I-like photosynthetic NAD(P)H dehydrogenase., Nat. Commun. 11 (2020) 610. doi:10.1038/s41467-020-14456-0.
[192] J.M. Schuller, P. Saura, J. Thiemann, S.K. Schuller, A.P. Gamiz-Hernandez, G. Kurisu, M.M. Nowaczyk, V.R.I. Kaila, Redox-coupled proton pumping drives carbon concentration in the photosynthetic complex I, Nat. Commun. 11 (2020) 494. https://doi.org/10.1038/s41467-020-14347-4.

[193] J. Artier, S.C. Holland, N.T. Miller, M. Zhang, R.L. Burnap, Synthetic DNA system for structure-function studies of the high affinity CO uptake NDH-1 protein complex in cyanobacteria, Biochim. Biophys. Acta Bioenerg. (2018). https://doi.org/10.1016/j.bbabio.2018.06.015.
[194] N.P. Cheremisinoff, ELEMENTAL AND STRUCTURAL CHARACTERIZATION TESTS, Polymer Characterization. (1996) 43–81. https://doi.org/10.1016/b978-081551403-9.50006-6.
[195] D. Beauchemin, Inductively coupled plasma mass spectrometry, Anal. Chem. 78 (2006) 4111–4136. https://doi.org/10.1021/ac060712t.
[196] D.M. Kern, S. Oh, R.K. Hite, S.G. Brohawn, Cryo-EM structures of the DCPIB-inhibited volume-regulated anion channel LRRC8A in lipid nanodiscs, Elife. 8 (2019). https://doi.org/10.7554/eLife.42636.
[197] N. Fisher, T.M. Bricker, D.M. Kramer, Regulation of photosynthetic cyclic electron flow pathways by adenylate status in higher plant chloroplasts, Biochim. Biophys. Acta Bioenerg. 1860 (2019) 148081. https://doi.org/10.1016/j.bbabio.2019.148081.
[198] R. Phillips, T. Ursell, P. Wiggins, P. Sens, Emerging roles for lipids in shaping membrane-protein function, Nature. 459 (2009) 379. https://doi.org/10.1038/nature08147.
[199] C. Hunte, S. Richers, Lipids and membrane protein structures, Curr. Opin. Struct. Biol. 18 (2008) 406–411. https://doi.org/10.1016/j.sbi.2008.03.008.
[200] P.L. Yeagle, Non-covalent binding of membrane lipids to membrane proteins, Biochim. Biophys. Acta. 1838 (2014) 1548–1559. https://doi.org/10.1016/j.bbamem.2013.11.009.
[201] M. Fry, D.E. Green, Cardiolipin requirement for electron transfer in complex I and III of the mitochondrial respiratory chain, J. Biol. Chem. 256 (1981) 1874–1880. https://www.ncbi.nlm.nih.gov/pubmed/6257690.
[202] D. Tang, W.L. Dean, D. Borchman, C.A. Paterson, The influence of membrane lipid structure on plasma membrane Ca2+ -ATPase activity, Cell Calcium. 39 (2006) 209–216. https://doi.org/10.1016/j.ceca.2005.10.010.
[203] Q.-X. Jiang, T. Gonen, The influence of lipids on voltage-gated ion channels, Current Opinion in Structural Biology. 22 (2012) 529–536. https://doi.org/10.1016/j.sbi.2012.03.009.
[204] S.G. Brohawn, Z. Su, R. MacKinnon, Mechanosensitivity is mediated directly by the lipid membrane in TRAAK and TREK1 K+ channels, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) 3614–3619. https://doi.org/10.1073/pnas.1320768111.
[205] K. Fiedorczuk, J.A. Letts, G. Degliesposti, K. Kaszuba, M. Skehel, L.A. Sazanov, Atomic structure of the entire mammalian mitochondrial complex I, Nature. 538 (2016) 406–410. https://doi.org/10.1038/nature19794.
[206] A.-N.A. Agip, J.N. Blaza, H.R. Bridges, C. Viscomi, S. Rawson, S.P. Muench, J. Hirst, Cryo-EM structures of complex I from mouse heart mitochondria in two biochemically defined states, Nat. Struct. Mol. Biol. 25 (2018) 548–556. https://doi.org/10.1038/s41594-018-0073-1.
[207] S.G. Brohawn, E.B. Campbell, R. MacKinnon, Physical mechanism for gating and mechanosensitivity of the human TRAAK K+ channel, Nature. 516 (2014) 126–130. https://doi.org/10.1038/nature14013.
[208] C. Lange, J.H. Nett, B.L. Trumpower, C. Hunte, Specific roles of protein-phospholipid interactions in the yeast cytochrome bc1 complex structure, EMBO J. 20 (2001) 6591–6600. https://doi.org/10.1093/emboj/20.23.6591.

[209] J.A. Letts, G. Degliesposti, K. Fiedorczuk, M. Skehel, L.A. Sazanov, Purification of Ovine Respiratory Complex I Results in a Highly Active and Stable Preparation, J. Biol. Chem. 291 (2016) 24657–24675. https://doi.org/10.1074/jbc.M116.735142.
[210] J.A. Letts, K. Fiedorczuk, G. Degliesposti, M. Skehel, L.A. Sazanov, Structures of Respiratory Supercomplex I III2 Reveal Functional and Conformational Crosstalk, Molecular Cell. 75 (2019) 1131–1146.e6. https://doi.org/10.1016/j.molcel.2019.07.022.
[211] I. Sakurai, J.-R. Shen, J. Leng, S. Ohashi, M. Kobayashi, H. Wada, Lipids in oxygen-evolving photosystem II complexes of cyanobacteria and higher plants, J. Biochem. 140 (2006) 201–209. https://doi.org/10.1093/jb/mvj141.
[212] M.S. Reid, D.M. Kern, S.G. Brohawn, Cryo-EM structure of the potassium-chloride cotransporter KCC4 in lipid nanodiscs, (n.d.). https://doi.org/10.1101/805267.
[213] F. Hauer, C. Gerle, N. Fischer, A. Oshima, K. Shinzawa-Itoh, S. Shimada, K. Yokoyama, Y. Fujiyoshi, H. Stark, GraDeR: Membrane Protein Complex Preparation for Single-Particle Cryo-EM, Structure. 23 (2015) 1769–1775. https://doi.org/10.1016/j.str.2015.06.029.
[214] A. Cheng, E.T. Eng, L. Alink, W.J. Rice, K.D. Jordan, L.Y. Kim, C.S. Potter, B. Carragher, High resolution single particle cryo-electron microscopy using beam-image shift, J. Struct. Biol. 204 (2018) 270–275. https://doi.org/10.1016/j.jsb.2018.07.015.
[215] C. Wu, X. Huang, J. Cheng, D. Zhu, X. Zhang, High-quality, high-throughput cryo-electron microscopy data collection via beam tilt and astigmatism-free beam-image shift, J. Struct. Biol. 208 (2019). https://doi.org/10.1016/j.jsb.2019.09.013.
[216] J.N. Cash, S. Kearns, Y. Li, M.A. Cianfrocco, High-resolution cryo-EM using beam-image shift at 200 keV, (n.d.). https://doi.org/10.1101/2020.01.21.914507.
[217] J. Zivanov, T. Nakane, S.H.W. Scheres, Estimation of High-Order Aberrations and Anisotropic Magnification from Cryo-EM Datasets in RELION-3.1, (n.d.). https://doi.org/10.1101/798066.
[218] B. Kastner, N. Fischer, M.M. Golas, B. Sander, P. Dube, D. Boehringer, K. Hartmuth, J. Deckert, F. Hauer, E. Wolf, H. Uchtenhagen, H. Urlaub, F. Herzog, J.M. Peters, D. Poerschke, R. Lührmann, H. Stark, GraFix: sample preparation for single-particle electron cryomicroscopy, Nat. Methods. 5 (2008) 53–55. https://doi.org/10.1038/nmeth1139.
[219] T. Vasanthakumar, S.A. Bueler, D. Wu, V. Beilsten-Edmands, C.V. Robinson, J.L. Rubinstein, Structural comparison of the vacuolar and Golgi V-ATPases from Saccharomyces cerevisiae, Proceedings of the National Academy of Sciences. 116 (2019) 7272–7277. https://doi.org/10.1073/pnas.1814818116.
[220] E. Palovcak, F. Wang, S.Q. Zheng, Z. Yu, S. Li, M. Betegon, D. Bulkley, D.A. Agard, Y. Cheng, A simple and robust procedure for preparing graphene-oxide cryo-EM grids, J. Struct. Biol. 204 (2018) 80–84. https://doi.org/10.1016/j.jsb.2018.07.007.
[221] C.J. Russo, L.A. Passmore, Controlling protein adsorption on graphene for cryo-EM using low-energy hydrogen plasmas, Nat. Methods. 11 (2014) 649–652. https://doi.org/10.1038/nmeth.2931.
[222] B.-G. Han, Z. Watson, H. Kang, A. Pulk, K.H. Downing, J. Cate, R.M. Glaeser, Long shelf-life streptavidin support-films suitable for electron microscopy of biological macromolecules, J. Struct. Biol. 195 (2016) 238–244. https://doi.org/10.1016/j.jsb.2016.06.009.
[223] F. Wang, Z. Yu, M. Betegon, M.G. Campbell, T. Aksel, J. Zhao, S. Li, S.M. Douglas, Y. Cheng, D.A. Agard, Amino and PEG-amino graphene oxide grids enrich and protect

samples for high-resolution single particle cryo-electron microscopy, J. Struct. Biol. (2019) 107437. https://doi.org/10.1016/j.jsb.2019.107437.

[224] C. McMahon, A.S. Baier, R. Pascolutti, M. Wegrecki, S. Zheng, J.X. Ong, S.C. Erlandson, D. Hilger, S.G.F. Rasmussen, A.M. Ring, A. Manglik, A.C. Kruse, Yeast surface display platform for rapid discovery of conformationally selective nanobodies, Nat. Struct. Mol. Biol. 25 (2018) 289–296. https://doi.org/10.1038/s41594-018-0028-6.