Genome-wide screening reveals a novel class of carbonic anhydrase-like inorganic carbon transporters in chemoautotrophic bacteria

John J. Desmarais[1], Avi I. Flamholz[1], Cecilia Blikstad[1], Eli J. Dugan[1], Thomas G. Laughlin[1], Luke M. Oltrogge[1], Kelly Wetmore[2], Joy Y. Wang[3], David F. Savage[1]

[1]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720
[2]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[3]Department of Chemistry, University of California, Berkeley, CA 94720

Correspondence should be addressed to: savage@berkeley.edu.

Abstract (150 words)

Many bacterial autotrophs rely on CO_2 concentrating mechanisms (CCMs) to assimilate carbon. Although many CCM proteins have been identified, including a 200+ MDa protein organelle called the carboxysome, 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*. Our screen revealed an operon encoding a domain of unknown function (PFAM:PF10070) and putative cation transporter subunit (PFAM:PF00361) is critical for CCM function. These two proteins, which we name DabA and DabB for “DABs accumulate bicarbonate,” function as a heterodimeric, energy-coupled inorganic carbon pump in *E. coli*. Furthermore, DabA has distant homology to a β-carbonic anhydrase and binds a zinc ion necessary for activity. Based on these results, we propose that DABs function as vectorial CAs coupled to cation gradients and serve as inorganic carbon pumps throughout prokaryotic phyla.

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 inorganic carbon into the biosphere. CBB activity is thus critical to agriculture and a major flux removing anthropogenic CO_2 from the atmosphere. Despite its centrality and abundance, Rubisco is not a very fast enzyme (Bar-Even et al., 2011; Bathellier et al., 2018; Flamholz et al., 2018). Nor is Rubisco very specific - all known Rubiscos can use molecular oxygen (O_2) as a substrate in place of CO_2 (Tcherkez, 2016). The resulting oxygenation reaction is often described as "wasteful" as it fails to incorporate inorganic carbon and produces a product, 2-phosphoglycolate, that is not part of the CBB cycle and must be recycled through metabolically-expensive photorespiratory pathways (Bauwe et al., 2010; Buchanan et al., 2015). Many studies support the hypothesis that improvements to Rubisco could improve crop yields, but Rubisco has proven quite recalcitrant to improvement by engineering. Indeed, it remains unclear whether or how Rubisco can be improved (Flamholz et al., 2018; Savir et al., 2010; Tcherkez et al., 2006).

Organisms that depend on Rubisco for growth often employ supplemental physiological mechanisms to improve its rate and specificity. These mechanisms are collectively termed CO_2 concentrating mechanisms (CCMs) because they serve to concentrate CO_2 at the site of Rubisco, ensuring Rubisco is saturated with CO_2, so that carboxylation proceeds at its maximum rate and oxygenation is competitively inhibited (Buchanan et al., 2015; Mangan et al., 2016; Raven et al., 2017). All cyanobacteria and many chemotrophic proteobacteria have a CCM (Badger and Price, 2003; Cannon et al., 2001). The bacterial CCM has garnered particular interest among bioengineers because it is well-understood, composed of only ~20 genes and operates inside single cells (Long et al., 2016). Detailed modeling suggests that transplantation of the bacterial CCM into crops might improve yields (McGrath and Long, 2014; Price et al., 2011) and efforts towards transplantation are already underway (Lin et al., 2014; Long et al., 2018; Occhialini et al., 2016).

Based on diverse experimental studies, a general model of the bacterial CCM function has emerged. This model requires two major components: active transport of inorganic carbon (C_i) leading to the accumulation of HCO_3^- in the cytosol and organization of RuBisCO with carbonic anhydrase (CA) in the lumen of a 200+ MDa protein organelle known as the carboxysome (Mangan et al., 2016). Energy-coupled inorganic carbon pumps ensure that the cytosolic HCO_3^- concentration is high (> 10 mM) and, crucially, out-of-equilibrium with CO_2. Inside the carboxysome, the lumenal CA converts the high cytosolic HCO_3^- concentration into a high carboxysomal CO_2 concentration, which promotes faster carboxylation by Rubisco and inhibits oxygenation (Mangan et al., 2016). Genetic insults to either component - uptake systems or carboxysomes - disrupt the CCM and mutants require elevated CO_2 for growth (Cai et al., 2009; Dou et al., 2008; Ogawa et al., 1987) This high-CO_2 requiring (HCR) mutant phenotype is commonly used to identify CCM components (Mackinder et al., 2016; Marcus et al., 1986; Ogawa et al., 1987; Price and Badger, 1989a).

Despite these early screens, a comprehensive list of bacterial CCM components remains unknown leaving open the possibility that additional activities are required for an *in vivo* reconstitution of the CCM. Although well-assembled carboxysome structures can be be heterologously expressed in bacteria and plants (Bonacci et al., 2012; Fang et al., 2018; Long et al., 2018), functionality of these carboxysomes in a heterologously-expressed CCM has not yet been demonstrated. Moreover, genetic and bioinformatic studies show that several additional genes are associated with carboxysome function (Axen et al., 2014; Jorda et al., 2013). For example, it was recently been demonstrated that carboxysome-associated genes may function as previously unrecognized Rubisco chaperones and assembly factors (Aigner et al., 2017; Wheatley et al., 2014). Moreover, many experimental (e.g. (Price and Badger, 1989a; Shibata et al., 2002)) and modeling studies (e.g. (Hopkinson et al., 2014; Mangan et al., 2016; Reinhold et al., 1991)) make it clear that energy-coupled inorganic carbon uptake systems are required for the CCM to function. Several different C_i pump families, including both transporters and facilitated uptakes systems are now known (Long et al., 2016; Price, 2011). However, as these systems are inherently difficult to purify due to their integral membrane protein nature and, relatedly, since model carbon-fixing bacteria often express multiple C_i uptake systems, a

mechanistic biochemical understanding of their function is limited (Artier et al., 2018; Battchikova et al., 2011; Price, 2011).

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 ϒ-proteobacterial chemoautotroph and a model system for studying α-carboxysomes (Heinhorst et al., 2006; Robertson and Kuenen, 2006). In addition to producing the catalog of essential genes for a bacterial chemotroph, we leverage our pooled mutant library to comprehensively screen for gene knockouts that produce an HCR phenotype. This screen identified all known CCM components and confirmed that a two-gene operon containing both a large, conserved, poorly-characterized protein (PFAM:PF10070, hereafter DabA) and a homolog of the H^+-pumping subunits of respiratory complex I (PFAM:PF00361, hereafter DabB) is required for CCM function. Recent proteomic analyses and physiological experiments have shown that this operon is involved in C_i transport in diverse proteobacteria (Mangiapia et al., 2017; Scott et al., 2018). For reasons outlined below, we term this operon and its gene products DAB for "**D**ABs **A**ccumulate **B**icarbonate."

The genes of the DAB operon form a protein complex that is capable of energetically-coupled C_i uptake when heterologously expressed in *E. coli*. Both proteins are necessary for activity and DabA appears to contain a domain homologous to a type II β-carbonic anhydrase. Like all β-CAs, DabA uses at least one cysteine residue to coordinate a zinc that is required for activity. These results suggest that the inorganic carbon uptake systems of proteobacterial chemotrophs rely on a vectorial CA mechanism that is coupled to a cation gradient (e.g., H^+ or Na^+). Similar mechanisms have been proposed for cyanobacterial C_i uptake systems (Han et al., 2017; Price, 2011; Shibata et al., 2002). These models typically claim that CO_2 hydration is coupled to favorable electron flow through the electron transport chain. However, DAB proteins do not seem to associate with *E. coli* complex I proteins in our experiments. We therefore propose a novel model of vectorial CA activity that is not coupled to electron flow. Rather, based on the experiments presented here, we advance a model of DAB function wherein dissipation of a cation gradient (e.g. of H^+ or Na^+) is coupled to unidirectional hydration of CO_2 to HCO_3^- by the DAB complex.

Results

Transposon mutagenesis and gene essentiality

We generated a randomly-barcoded genome-wide pooled knockout library of *Hnea* by conjugation (Wetmore et al., 2015). The donor strain (*E. coli* APA 766) contains a vector with a barcoded Tn5-derived transposon encoding a kanamycin resistance marker. Conjugation was performed under 5% CO_2 so that CCM genes could be knocked out and the resulting *Hnea* conjugants were selected for growth in the presence of kanamycin at 5% CO_2 to ensure transposon insertion.

The presence of a unique barcode in each transposon simplifies the use of the library for pooled screens (Wetmore et al., 2015). However, transposon insertion sites and associated barcodes must be mapped to the *Hnea* genome in order to perform these screens. We mapped transposon insertions using standard TnSeq methods (Wetmore et al., 2015) and found that our library contains ~10^5 transposon insertions, amounting to one insertion for every ~25 base pairs in the *Hnea* genome. Since the average gene contains ~35 insertions, genes with no insertions are almost certainly essential for growth (Rubin et al., 2015). Following this logic, we used a simple statistical model to identify 551 essential genes and 1787 nonessential genes out of 2408 genes in the *Hnea* genome (Methods, Figure 1A-B, Table S2). The remaining 70 genes were classified as "ambiguous" due either to their short length or because replicate mapping experiments were discordant (Methods). Genes associated with known essential functions including central carbon metabolism, ribosome production, and DNA replication were categorized as essential (Figure 1C). As the library was generated under 5% CO_2 it contains multiple knockouts of known CCM genes, including carboxysome components (Figure 2C).

Comprehensive screen for Hnea CCM components

Knockouts of CCM genes are expected to require high CO_2 for growth (Mackinder et al., 2016; Marcus et al., 1986; Price and Badger, 1989a). Therefore, CCM gene knockouts should have low fitness in ambient CO_2 concentrations. As our pooled library contains ~70,000 barcodes that map to exactly one position in the *Hnea* genome, we were able to use the barseq method to quantify the fitness defects associated with single gene knockouts for all nonessential *Hnea* genes (Figure 2B). In barseq, a preculture of the library is grown in permissive conditions (5% CO_2) and then back-diluted into two conditions: a reference condition (5% CO_2 again) and a condition of interest (e.g. ambient CO_2). Genomic DNA is extracted from the preculture (called t0) and both culture outgrowths and barcodes are PCR-amplified and sequenced. In this pooled competition assay the proportional change in barcode abundance is taken to reflect the fitness effect of gene knockouts (Wetmore et al., 2015). A CCM gene knockout should have no fitness defect in 5% CO_2 but a large defect in ambient CO_2. Since the library contains >20 knockouts with unique barcodes per gene, these screens contain multiple internal biological replicates testing the effect of single gene knockouts.

As expected, knockouts to nearly all carboxysome-associated genes produced large fitness defects in ambient CO_2 (Figure 2B&C). These genes include *cbbLS* - the large and small subunits of the α-carboxysomal Rubisco; *csos2* - an intrinsically disordered protein required for α-carboxysome assembly; *csosCA* - the carboxysomal carbonic anhydrase; *csos4AB* - the pentameric proteins thought to form vertices of the α-carboxysome; and *csos1CAB* - the hexamers that form the faces of the α-carboxysome shell (Cannon et al., 2001; Heinhorst et al., 2006). *csos1D*, a shell hexamer with a large central pore, does not appear to be an essential component of the *Hnea* CCM (Bonacci et al., 2012; Roberts et al., 2012). The *Hnea* genome also contains a secondary, non-carboxysomal Form II Rubisco that is likely not involved in CCM activity as its disruption confers no fitness defect in ambient CO_2. A number of genes that are not structurally associated with the carboxysome 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 (Aigner et al., 2017; Wheatley et al., 2014), and two paralogous loci encoding DAB genes (hereafter DAB1 and DAB2) (Figure 2B-F).

dabA2 and dabB2 are necessary and sufficient for energy-coupled C_i accumulation in E. coli

DAB1 is a cluster of 3 genes found in an operon directly downstream of the carboxysome operon (Figure 3A). Though DAB1 is part of a larger 11-gene operon containing several genes associated with Rubisco proteostasis, including acRAF (Aigner et al., 2017; Wheatley et al., 2014) and a cbbOQ-type Rubisco activase (Mueller-Cajar, 2017), 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 (Figure 2B) and possess similar genes. Both contain a large, conserved, poorly-characterized protein (PFAM:PF10070, DabA) and a homolog of the H^+-pumping subunits of respiratory complex I (PFAM:PF00361, DabB). DAB1 also contains a second poorly-described integral membrane protein between DabA and DabB that has weak homology to DabB but is truncated to half the length. Operons of this type were recently demonstrated to be in involved C_i transport in proteobacterial chemotrophs (Mangiapia et al., 2017; Scott et al., 2018).

Since DAB2 disruption is associated with a larger fitness defect than DAB1 (Figure 2B), we used an *E. coli*-based system to test DAB2 for C_i uptake activity. Knocking out carbonic anhydrases produces an HCR phenotype in *E. coli (Merlin and Masters, 2003)* that is complemented by expression of cyanobacterial bicarbonate transporters (Du et al., 2014). We generated an *E. coli* strain, CAfree, that contains no CA genes (Methods) and found that DAB2 expression enables growth of CAfree in ambient CO_2 (Figure 3A). CAfree complementation requires both DabA2 and DabB2 (Figure 3B) and leads to uptake of radiolabeled C_i that is substantially above background (grey bars in Figure 3C). Moreover, DAB2-associated C_i uptake is strongly inhibited by the ionophore CCCP (white bars in Figure 3C), indicating that DAB2 is energetically-coupled to a cation gradient (e.g. H^+ or Na^+).

DabA2 and DabB2 interact to form a complex

In order to determine if the genetic interaction between *dabA2* and *dabB2* is due to a physical interaction, we attempted to purify the two proteins as a complex. DabA2 was genetically fused to a C-terminal Strep-tag, DabB2 fused to a C-terminal GFP and 6xHis-tag and both assayed for heterologous expression in *E. coli.* Tandem-affinity purification revealed that DabA2 and DabB2 interact physically to form a complex in *E. coli* (Figure 4C). 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 4D). Notably, we did not observe co-purification of *E. coli* complex I subunits or any other proteins with the DabA-DabB complex (Figure 4C), suggesting that the DAB2 operates as an independent complex within the membrane.

DabA is distantly homologous to a β-CA and predicted CA active site residues are required for zinc binding and activity

Structural homology modeling software predicted that the middle of DabA2 has sequence elements related to a bacterial type II β-CA (Figure 3A). Specifically, Phyre2 predictions identified C539 and H524 as part of a potential Zn^{2+} binding site distantly homologous to a β-CA (10% of DabA, 90.8% confidence). I-TASSER predicted a similar site comprising the same two residues with an additional cysteine (C351), and an aspartic acid (D353). These residues could make up the active site of a type II β-CA (Cronk et al., 2006, 2001; Supuran, 2016) (Figure 4A). We generated individual alanine mutants for each of these predicted active site residues and tested their capacity to rescue CAfree. All mutants failed to produce growth of CAfree in ambient CO_2 (Figure 4B). We then purified two of the predicted active site mutants (C351A and D353A) in order to determine if this ablation of function was due to failure to bind zinc (Figure 4C). Zinc binding was measured

using X-ray fluorescence spectroscopy. The wild-type protein and D353A mutant both show strong zinc fluorescence while C351A did not (Figure 4E). Notably, this agrees with studies of type II β-CAs where all four residues are required for activity, but the aspartic acid is not essential for zinc binding (Cronk et al., 2006, 2001; Supuran, 2016).

Purified DAB2 does not have conspicuous CA activity.

Finally, we tested whether purified DabAB2 displays CA activity (Figure 4F). In our assay conditions no CA activity was observed. We tested for activity in CO_2 concentrations that are typically saturating for CAs and provided a substantial amount of purified DabAB2 (> 650-fold more protein than the positive control) but did not detect any activity (Figure 4F). We estimate that activities 20x lower than the positive control would have been detected. The absence of activity *in vitro* suggests that DabAB2 must be situated in a cell membrane with a cation gradient in order to function as an activated carbonic anhydrase.

Discussion

Since oxygenic photosynthesis is responsible for our contemporary O_2-rich and relatively CO_2-poor atmosphere, it is likely that Rubisco evolved in CO_2-rich environment where its modest rate and limited specificity posed no problem (Shih et al., 2016; Tabita et al., 2008). However, over the subsequent 2.5 billion years of Earth history, atmospheric O_2 increased and CO_2 declined to the point where, today, autotrophic bacteria that grow in atmosphere appear to uniformly have CCMs (Raven et al., 2017). Bacterial CCMs come in two convergently-evolved forms - α-carboxysomes are found in proteobacteria and marine cyanobacteria while β-carboxysomes are found in freshwater cyanobacteria (Rae et al., 2013). Because the bacterial CCM is well-studied and known to function in single cells it is an attractive target for synthetic biology and efforts to transplant it into crops are already underway (Lin et al., 2014; Long et al., 2018; Occhialini et al., 2016).

In principle, the bacterial CCM requires two major components: i. energy-coupled uptake of inorganic carbon to concentrate HCO_3^- in the cytosol and ii. carboxysome structures that co-localize Rubisco with CA enzymes that convert concentrated HCO_3^- into a high concentration of the Rubisco substrate CO_2 (Mangan et al., 2016). While the carboxysome components are well-documented for both α- and β-CCMs, C_i uptake systems of the proteobacterial α-CCM only identified very recently (Mangiapia et al., 2017; Scott et al., 2018). Although numerous laboratories have spent decades studying the bacterial CCM, it remains unclear whether our current “parts list” for α- and β-CCMs is complete.

Here we undertook an effort to complete the genetic “parts list” of the α-CCM of the proteobacterial chemotroph *H. neapolitanus*. We generated a genome-wide knockout library containing ~35 individual knockouts for every gene in the *Hnea* genome and compiled the first list of essential genes for a chemotroph (Figure 1). Because we generated the library at elevated CO_2 (5%) we were able to knockout all known CCM components, including all genes known to form the α-carboxysome. We subsequently used this library to screen for genes associated with CCM activity by screening for knockouts with fitness defects specific to ambient CO_2 conditions (Figure 2A). As expected, this screen identified most carboxysome components and also highlighted several genes whose relationship to the CCM is not fully understood. These genes include several transcriptional regulators, a putative Rubisco chaperone and two small operons (DAB1 and DAB2) that are involved in CCM-associated C_i uptake in chemotrophic proteobacteria (Mangiapia et al., 2017; Scott et al., 2018).

Freshwater cyanobacteria express several well-studied C_i transporters (Price, 2011) that uptake HCO_3^- and are coupled to energy in the form of ATP or a sodium gradient. The substrate, energy coupling and chemical mechanism are unclear for the recently identified proteobacterial transporters. We note however, that the preferred substrate for C_i uptake will depend on the extracellular pH because pH determines the relative abundance of CO_2, H_2CO_3, HCO_3^- and CO_3^- (Mangan et al., 2016). Since *Hnea* and many other sulfur-oxidizing proteobacteria are acidophilic and CO_2 is more abundant than HCO_3^- at more acidic pH, it stands to reason that they might have evolved a mechanism to take up CO_2 instead of HCO_3^- (Figure 5S2).

We showed that the DAB2 operon encodes a two-component protein complex that has C_i uptake activity in *E. coli* (Figure 3B-C). This complex may be a heterodimer, as suggested by size-exclusion chromatography (Figure 4D). As this activity is strongly inhibited by the ionophore CCCP (Figure 3C), we suspect that DAB2-mediated C_i uptake is energetically-coupled to a cation gradient (Figure 5). Moreover, the DabA unit of this complex has limited homology to a β-carbonic anhydrase and binds a zinc that is required for activity (Figures 3-4). For all these reasons, we propose a model of DAB activity wherein CO_2 is passively taken into the cell and then vectorially (unidirectionally) hydrated to HCO_3^- by DabA. Model carbonic anhydrases are not directly coupled to any energy source (e.g., ATP) and so they only accelerate the equilibration of CO_2 and HCO_3^-. Coupling unidirectional conversion of CO_2 into HCO_3^- via an existing cation gradient, though, would enable the DAB system to actively accumulate C_i and power the CCM. We draw this activity as being coupled

to the H+ gradient in Figure 5 for simplicity, but our results are equally consistent with any other cation gradient, e.g. Na^+. This mechanism requires tight coupling of cation flow to CO_2 hydration by the CA-like DabA protein, which is consistent with our observation that purified DabAB2 displays no measurable CA activity. Notably, the activity of type II β-CAs may be regulated by Zn^{2+} binding and unbinding by the active site aspartic acid (D353 in DabA2), which is postulated to be a mechanism for allosteric control (Cronk et al., 2006). A similar mechanism could be used to allosterically couple ion movement through DabB to the CA-like active site of DabA.

Cyanobacteria possess two distinct uptake systems (CupA/B) that perform vectorial conversion of CO_2 to bicarbonate (Maeda et al., 2002; Price, 2011; Rae et al., 2013; Shibata et al., 2002, 2001). Unlike in *Hnea*, however, these are typically secondary transporters that are not required for growth in standard lab media. For this reason, CupA/B have proven challenging to study. Because CupA/B appear to associate with the NDH complex (cyanobacterial complex I homolog) in transmission electron micrographs (Battchikova et al., 2011; Birungi et al., 2010), it has been proposed that their CO_2 hydration activity is coupled to energetically-favorable electron flow (Figure 5 Supplement 1). As shown in Figures 3-4, however, the DAB complex functions in *E. coli* but does not appear to engage the *E. coli* complex I. Rather, the two subunits of the complex co-purify alone, suggesting that they function as a single unit in the *E. coli* membrane. We therefore propose that DAB activity is coupled to a cation gradient and not electron flow, as shown in Figure 5.

We observed that DabAB2 functions substantially better in CAfree *E. coli* than SbtA (Figure 3C&S3), the primary inorganic carbon transporter of model freshwater cyanobacteria (Du et al., 2014; Rae et al., 2013). This is likely due to greater “compatibility” of *Hnea* proteins with *E. coli* expression as compared to proteins derived from cyanobacteria. Indeed, *Hnea* is a proteobacterium like *E. coli*. It may also be the case 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 chemoautotrophs throughout α- β- and ɣ-proteobacteria and the α-CCM was clearly horizontally transferred at least once from proteobacteria to marine cyanobacteria (Rae et al., 2013). Since DAB2 appears to be so much more active in *E. coli* than SbtA and the α-CCM appears to have undergone widespread horizontal transfer, we believe that the DAB-family transporters are an attractive target for protein engineering and heterologous expression of CCM components in plants and industrial microbes, where a elevated intracellular C_i levels could be technologically useful (Antonovsky et al., 2016).

Materials and Methods

Bacterial strains and growth conditions

E. coli strain APA 766 was used as conjugation donor to transfer the Tn5 transposon to *Halothiobacillus neapolitanus* C2 (*Hnea*) via conjugation (Wetmore et al., 2015). The *E. coli* double CA deletion strain "CAfree" (BW25113 ΔcanA 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 *canA* knockout strain EDCM636 (MG1655 *ΔcanA*, Yale Coli Genomic Stock Center, (Baba et al., 2006; Merlin and Masters, 2003)). Lysogeny broth (LB) and LB agar were used as *E. coli* growth media unless otherwise specified. *E. coli* strains were grown at 37 ℃ 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 ℃ 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., 2015). Conjugations were performed as follows. *Hnea* and APA 766 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 APA 766 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 following protocols in (Wetmore et al., 2015). Transposons were mapped after Illumina sequencing using software developed in (Wetmore et al., 2015) 1.6 OD600 unit aliquots were then flash frozen in 50% glycerol for subsequent Bar-seq experiments.

Essential gene assignment

Following the logic of (Rubin et al., 2015; Wetmore et al., 2015), we categorized genes as essential if we observed significantly fewer transposon insertions than would be expected by chance. If insertion occurred uniformly at random, we expect the number of insertions per gene to follow a binomial distribution. The probability of observing at most *k* insertions into a gene of length *n* is therefore expressed as:

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

Here *p* is the average rate of transposon insertion per base pair genome-wide. Genes were determined to be essential if they received a low-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) the replicates were discordant or (ii) zero insertions were observed but the gene was short enough that the formula could not yield a Bonferroni-corrected probability below 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., 2015). 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 in (Wetmore et al., 2015).

CAfree rescue experiments

Electrocompetent CAfree cells were prepared using standard protocols and transformed with constructs 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 with shaking in a Tecan M1000 microplate reader under ambient conditions, 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.

Silicone oil centrifugation measurement of inorganic carbon uptake

The silicone oil filtration method was modified from (Dobrinski et al., 2005) and used to measure uptake of labeled 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). 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 7.0) 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 with 0.4% glucose + 0.4% DMSO and the inhibitor (100 µM CCCP) for 10 minutes before assay.

Structural Homology Modeling

Homology modeling was performed using Phyre2 and I-TASSER web servers with default parameters (Kelley et al., 2015; Roy et al., 2010).

Purification of DAB2

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 ℃. When midlog was reached, cells were induced with 20 ml of 50 mg/ml arabinose and transitioned to 20 ℃ 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 (Newby et al., 2009). Membranes were then repelleted 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. Eluant was then incubated with Strep-Tactin (Millipore) resin for 90 min at 4 ℃. 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.

CA Assays

CA catalyzed CO_2 hydration of purified DAB2 complex and human carbonic anhydrase (hCA) was measured using the buffer/indicator assay of Khalifah (Khalifah, 1971) on a KinTek AutoSF-120 stopped-flow spectrophotometer at 25 °C. The buffer/indicator pair used were TAPS/*m*-cresol purple measured at a wavelength of 578 nm using a pathlength of 0.5 cm. Final buffer concentration after mixing were 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 were: 9.8 µM DAB2 (His-elution) and 0.015 µM hCA (positive control; Sigma Aldrich C6624). Saturated solution of CO_2 (32.9 mM) were 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 dissolved in 20-200 µl of TBS + 0.03% DDM was precipitated by mixing with 4 volumes of acetone and incubated 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 ℃ until analysis. Fluorescence analysis was performed by breaking up the pellet into 5 ul of TBS + 0.03% DDM with a pipette tip. Small pieces of the pellet were looped with a nylon loop and flash frozen in place on a goniometer 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 and D353A, bovine CA (positive control; Sigma Aldrich C7025) and a buffer blank was used as a negative control. Experiments were performed at the Lawrence Berkeley National Laboratory Advanced Light Source Beamline 8.3.1.

Acknowledgements

We thank Adam Deutschbauer and Morgan Price for assistance with RB-TnSeq experiments and analysis, respectively. We also thank Andreas Martin and Jared Bard for assistance with stopped flow experiments. X-ray-based experiments were performed at the Lawrence Berkeley National Laboratory Advanced Light Source Beamline 8.3.1. J.J.D. was supported by National Institute of General Medical Sciences grant-T32GM066698. A.F. and T.G.L. were supported by a National Science Foundation Graduate Research Fellowship. C.B. was supported by an International Postdoctoral grant from the Swedish Research Council. D.F.S. was supported by the US Department of Energy Grant DE-SC00016240.

Competing Interests

UC Regents have filed a patent related to this work on which J.J.D., A.F.. and D.F.S. are inventors. D.F.S. is a co-founder of Scribe Therapeutics and a scientific advisory board member of Scribe Therapeutics and Mammoth Biosciences. All other authors declare no competing interests.

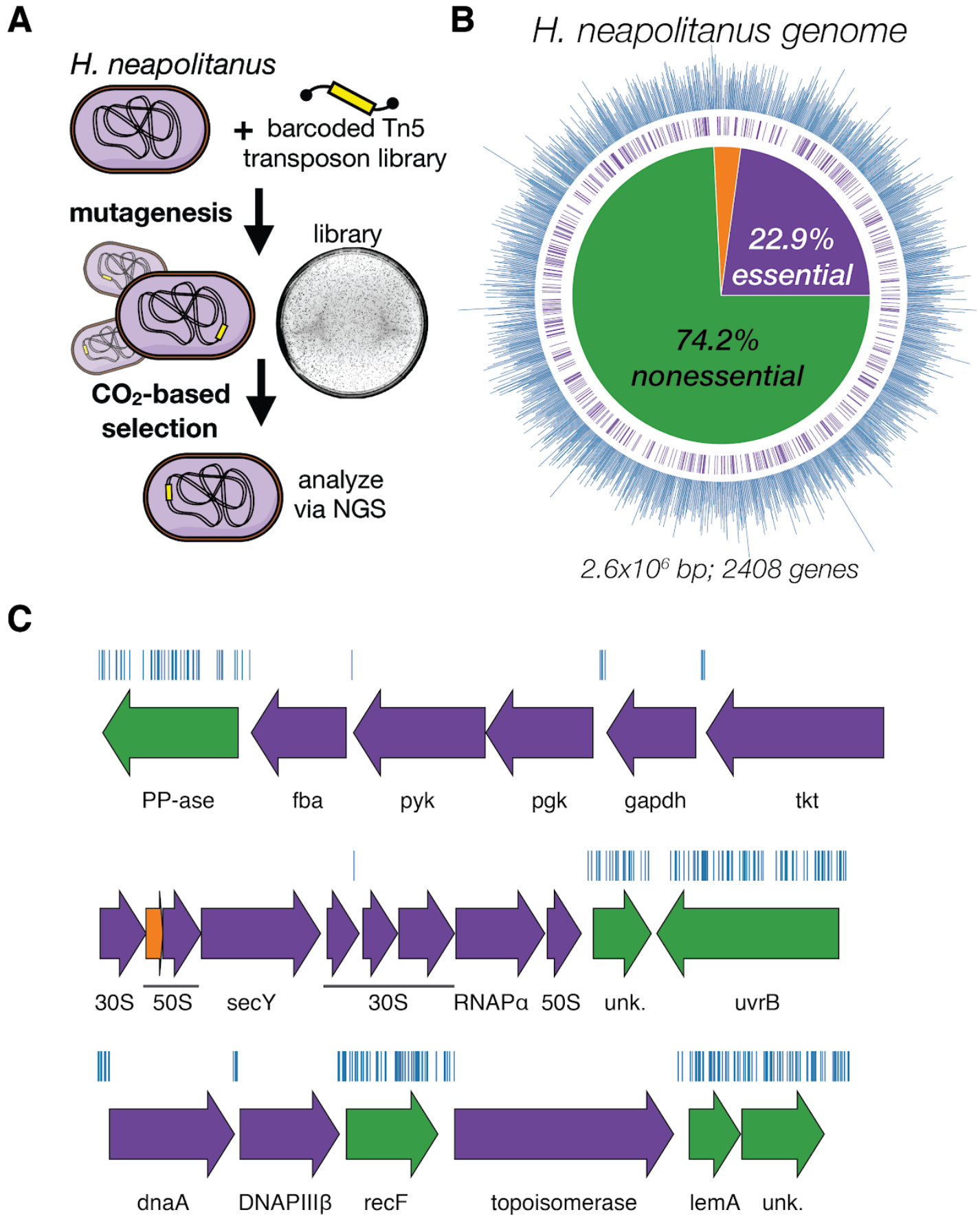

Figure 1. Transposon mutagenesis reveals the essential gene set of a chemoautotrophic organism. A. Schematic depicting the generation of the RB-TnSeq library and detection of HCR genes using a CO_2 based selection. **B.** Insertions

and essential genes are well-distributed throughout the genome. The outer track (blue) is a histogram of the number of barcodes that were mapped to a 1 kb window in the *Hnea* genome. The inner track annotates essential genes in purple. The pie chart shows the percentages of the genome called as 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 top operon contains 5 genes involved in either glycolysis or the CBB cycle. The second operon contains genes encoding 30S and 50S subunits of the ribosome, the secY secretory channel, and an RNA polymerase subunit. The third operon contains genes involved in DNA replication. Acronyms: 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).

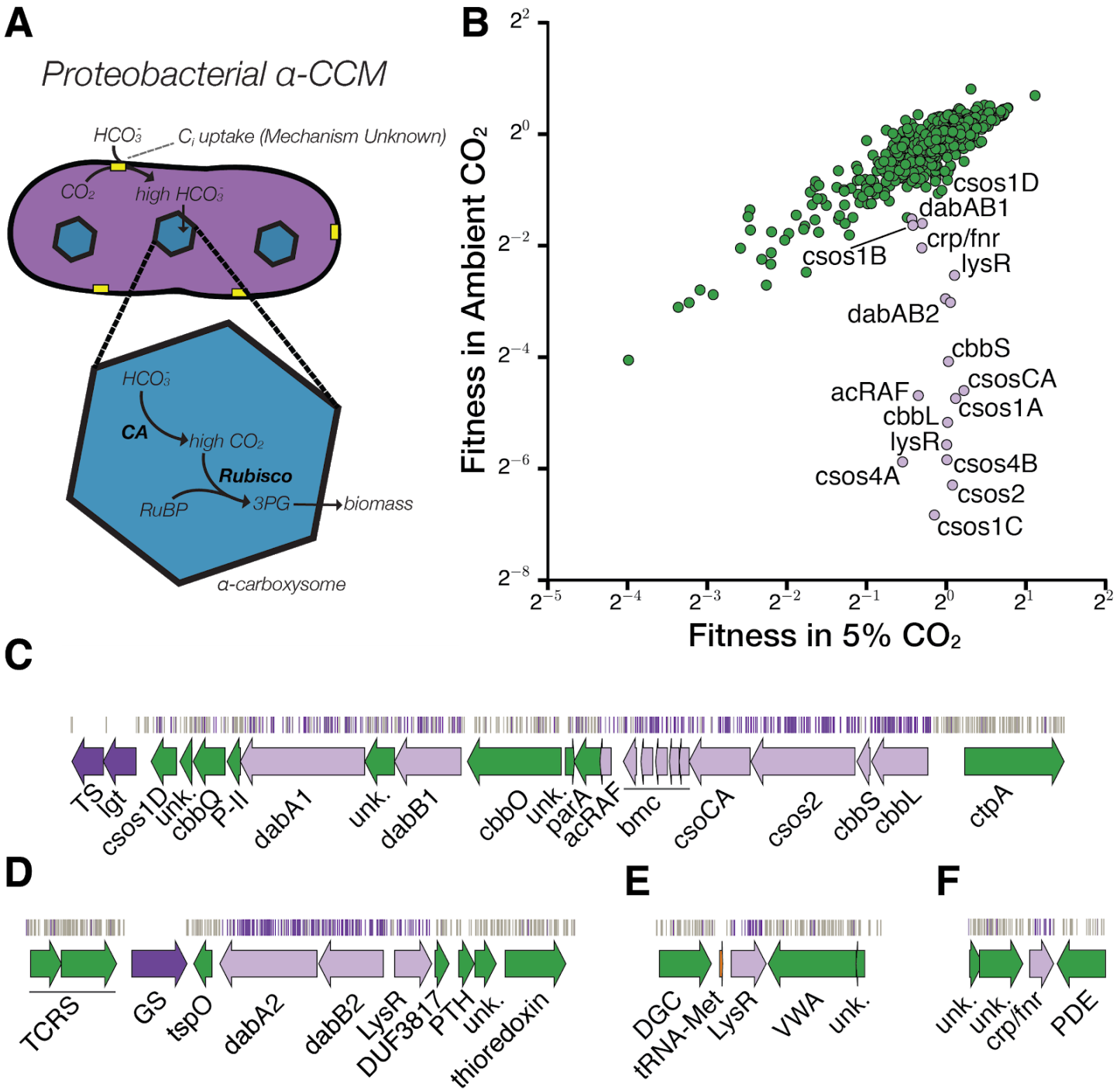

Figure 2. A systematic screen for high CO_2-requiring mutants identifies genes putatively associated with the CCM. A. Simplified model of the α-CCM of chemotrophic proteobacteria. Inorganic carbon is concentrated via an unknown mechanism, producing a high cytosolic HCO3- concentration. High cytosolic HCO3- is converted into high carboxysomal CO2 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 a one of two replicates of the BarSeq - the second replicate gives consistent results. When the effect of single transposon insertions into a gene are mutually consistent, those effects are averaged to produce the gene-level fitness value plotted (Wetmore et al., 2015). We define HCR mutants as those displaying a twofold fitness defect in ambient CO_2 relative to 5% CO_2 (i.e. a fitness difference of 1 on the $\log_2$ scale plotted). HCR genes are colored light purple. 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 purple if they display a twofold fitness defect in ambient CO_2 relative to 5% CO_2, otherwise they are colored grey. **C.** The gene cluster containing the carboxysome operon and a second CCM-associated operon annotated as in Figure 1C. This second operon contains acRAF, a FormIC associated

cbbOQ-type Rubisco activase and DAB1. **D.** The DAB2 operon and surrounding genomic context. **E.** The genomic context of a lysR-type transcriptional regulator that shows an HCR phenotype. **F** The genomic context of a crp/fnr-type transcriptional regulator that displays an HCR phenotype. Genes labeled “unk.” are hypothetical proteins. Acronyms: 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).

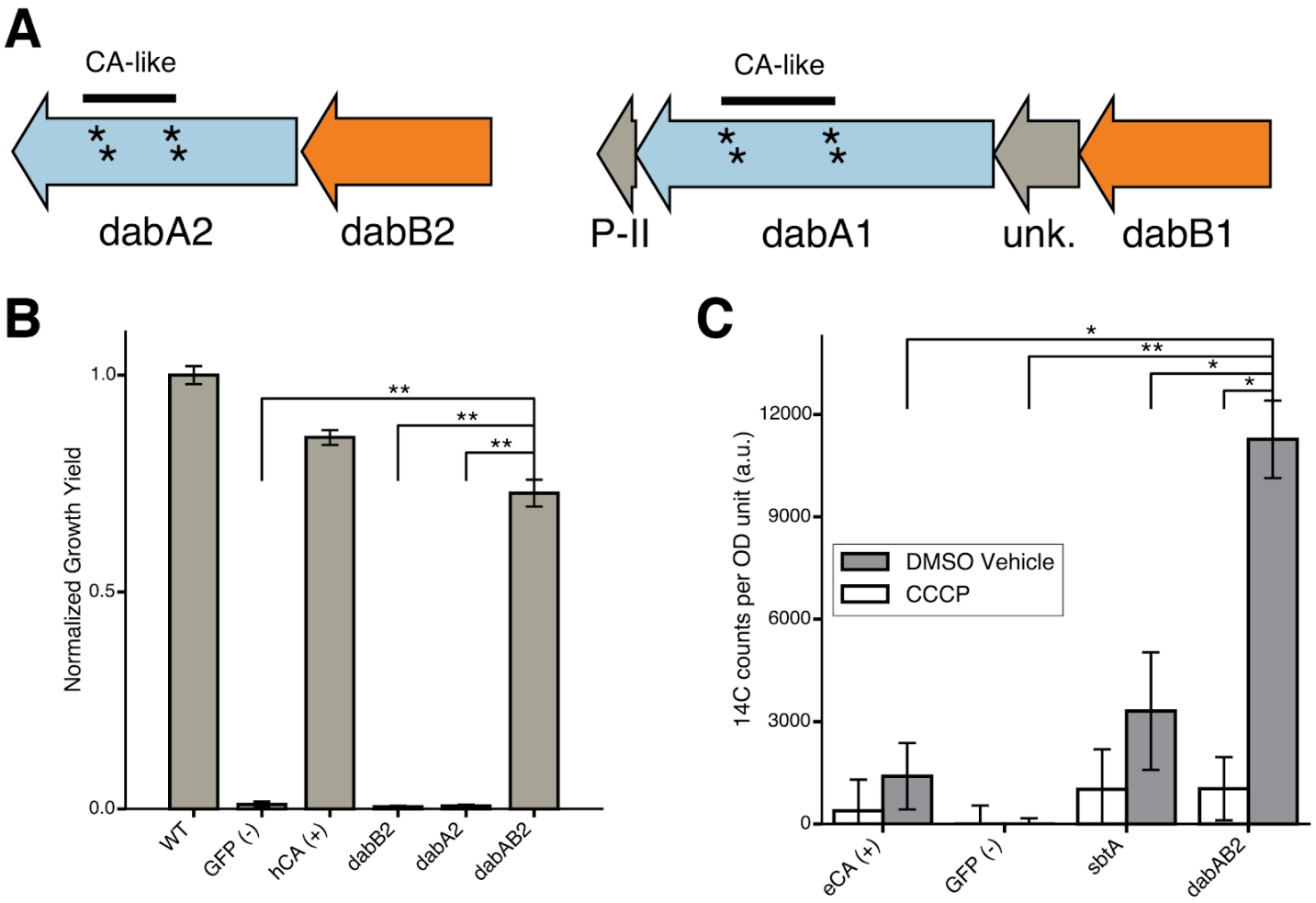

Figure 3. The DABs catalyze active transport of C_i energized by a cation gradient. A. Operonic structure of the DAB1 and DAB2 operons. As noted in the text and shown in Figure 2B, DAB1 is actually a piece of a larger 11-gene operon directly downstream of the carboxysome operon and containing CCM-associated genes. Both DAB1 and DAB2 "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 Zn ion are marked by asterisks. The height of the asterisks has been varied because they are placed close in sequence space and are otherwise hard to distinguish. DabB is homologous to a cation transporter in the same family as the H+ pumping subunits of respiratory complex 1 (PFAM:PF00361). The DAB1 operon also contains a protein of unknown function between DabA1 and DabB1. **B.** DAB2 was tested for ability to rescue growth of CAfree *E. coli* in ambient CO_2 conditions. The full operon (DabAB2) rescues growth as well as heterologous expression of the human carbonic anhydrase II (hCA), but rescue is contingent of the expression of both genes. Error bars represent standard deviations of 4 replicate cultures. **C.** CAfree *E. coli* were tested for C_i uptake using the silicone-oil centrifugation method. Expression of DabAB2 produced a large and statistically significant increase in 14C uptake as compared to all controls. Moreover, treatment with the ionophore CCCP greatly reduces DabAB2-mediated 14C uptake, suggesting that DabAB2 is coupled to a cation gradient. canA (eCA) was used as a control for a non-vectorial CA. sbtA was used as a known C_i importer. GFP was used as a vector control. Error bars represent standard deviations of 3 technical replicates. In (B) and (C) "*" denotes that the means are significantly different with $P < 0.05$ according to a two-tailed T-test. "**" denotes $P < 5X10^{-4}$.

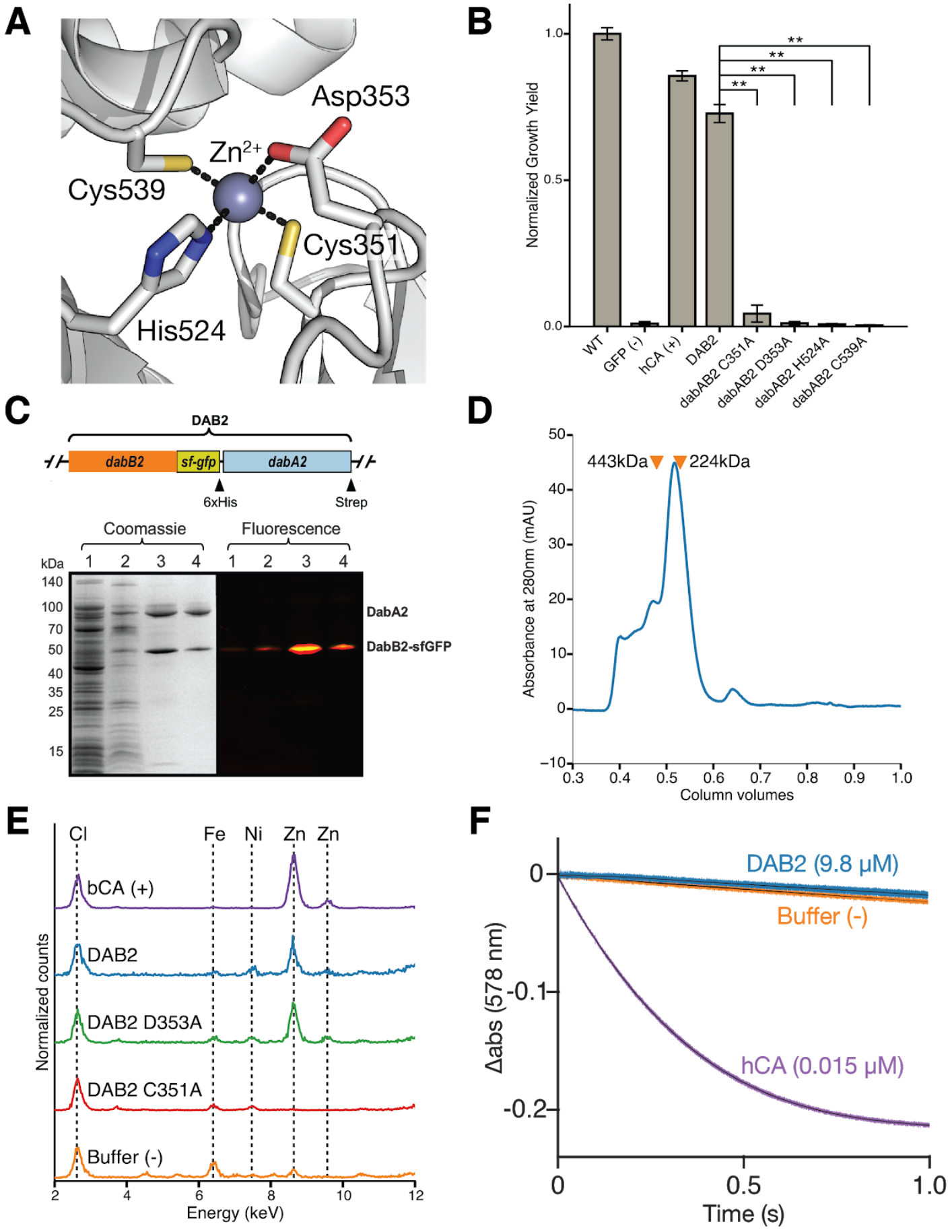

Figure 4. DabA contains a β-CA-like active site but is not constitutively active. A. Structural model of DabA2 active site based on the constitutive β-CA of *E. coli* (PDB 1I6P). Typical β-CAs rely on two cysteine and one histidine residues to

bind Zn^{2+}. A fourth residue - an aspartic acid - coordinates Zn^{2+} in the structure but is thought to be displaced in order to enable catalysis (Cronk et al., 2006). **B.** Alanine mutants of the putative DabA2 active site residues (C351A, D353A, H524A, C539A) abrogate rescue of CAfree *E. coli*. "*" denotes that means differ significantly with $P < 0.05$ according to a two-tailed T-test, and "**" denotes $P < 5X10^{-4}$. Error bars represent standard deviations of four replicate cultures. **C.** We purified the DabAB2 complex from *E. coli* BL21(AI) cells using a purification construct in which DabB2 was C-terminally fused to sf-GFP and a 6xHis-tag and DabA2 was C-terminally fused to a Strep-tag. Lane 1: clarified lysate; 2: solubilized membranes; 3: Ni resin eluent; 4: strep-tactin resin eluent. DabA2 and DabB2 co-purify as a single complex without any obvious interactors. **D.** Size-exclusion chromatography trace 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 at an estimated mass of ~270 kDa, which must be an oligomer of DabA and DabB. Given the additional size contributed by the detergent-belt, a heterodimer is consistent with these data. **E.** Comparing to X-ray fluorescence of Bovine CA (bCA), it is clear that purified DabAB2 and the aspartic acid mutant (D353A) bind zinc, as is expected based on the current model of type II β-CA activity (Cronk et al., 2006). A cysteine mutant (C351A) fails to bind zinc, which is also consistent with panel A. **F.** Purified DabAB2 does not display any obvious CA activity even though it was present in 650-fold excess over the positive control (Human carbonic anhydrase II, hCA) in our assays. In (B) "**" denotes that the means are significantly different with $P < 5X10^{-4}$ according to a two-tailed T-test.

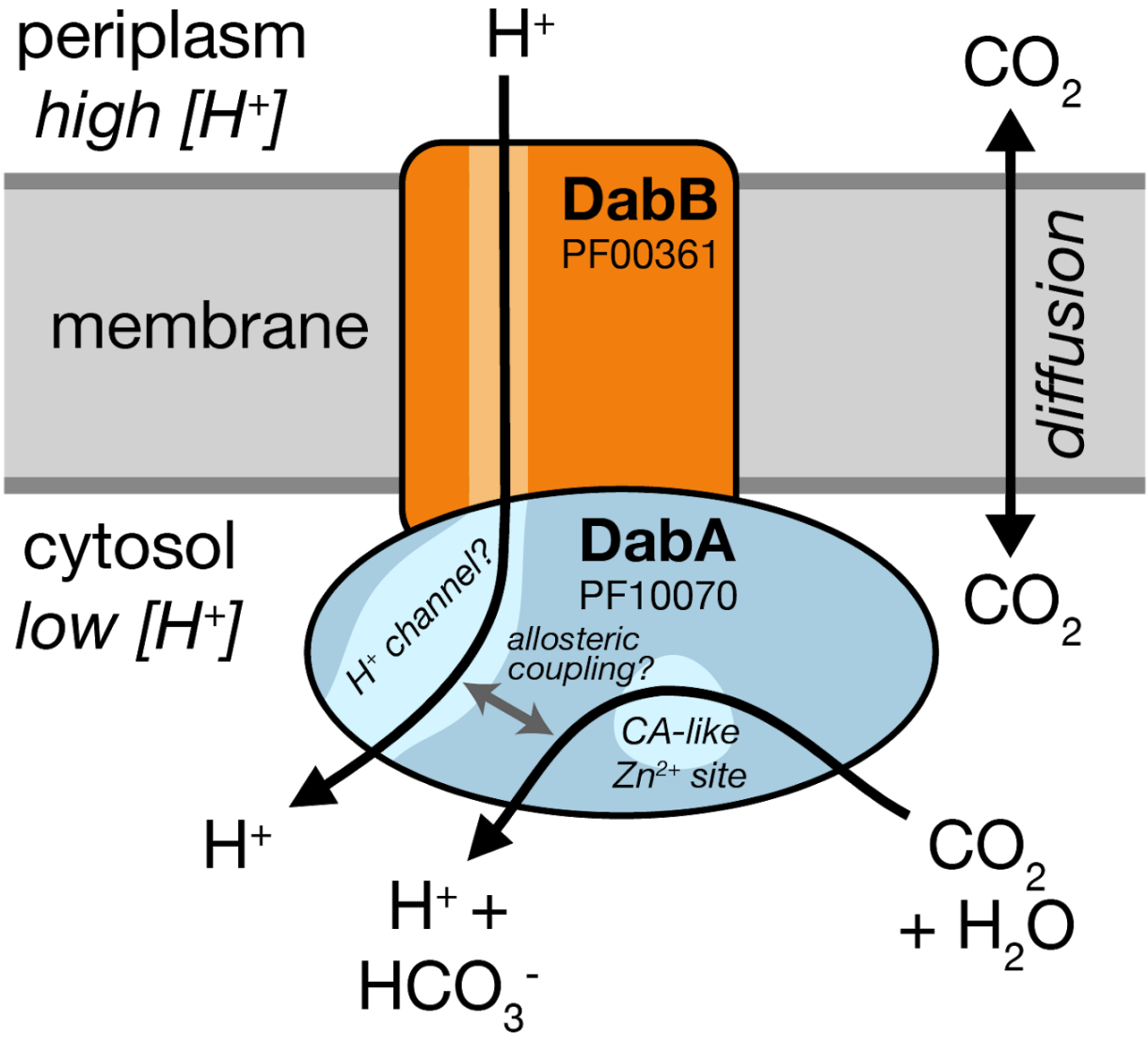

Figure 5. A model of the unidirectional energy-coupled CA activity of DAB complexes. We propose that DabAB complexes couple the β-CA-like active site of DabA to a cation gradient across the cell membrane, thereby producing unidirectional hydration of CO_2 to HCO_3^-. We draw this activity as being coupled to the H+ gradient (more generally, to the proton-motive force) for simplicity but our results are equally consistent with another cation gradient, e.g. Na^+. This model of energy-coupled CA activity is consistent with the DABs role as a C_i uptake system in the proteobacterial CCM since the CCM requires a high and, crucially, out-of-equilibrium HCO_3^- concentration in the cytosol in order for the carboxysomal CA to produce a high CO_2 concentration near Rubisco. Because it appears that DabAB2 is not active as a purified complex, the protein must tightly couple the inflow of cations with CO_2 hydration so that there is no "slippage." Indeed, slippage - i.e., uncoupled CA activity - would be counterproductive from the perspective of the CCM (Mangan et al., 2016; Price and Badger, 1989b). Notably, Zn^{2+} binding by the active site aspartic acid of type II β-CAs (D353 in DabA2, Figure 4A) is thought allosterically regulate activity (Cronk et al., 2006). This Asp-mediated activity switch could, therefore, provide a means for allosteric coupling of a β-CA active site to distal ion transport.

References

Aigner H, Wilson RH, Bracher A, Calisse L, Bhat JY, Hartl FU, Hayer-Hartl M. 2017. Plant RuBisCo assembly in E. coli with five chloroplast chaperones including BSD2. *Science* **358**:1272–1278. doi:10.1126/science.aap9221

Antonovsky N, Gleizer S, Noor E, Zohar Y, Herz E, Barenholz U, Zelcbuch L, Amram S, Wides A, Tepper N, Davidi D, Bar-On Y, Bareia T, Wernick DG, Shani I, Malitsky S, Jona G, Bar-Even A, Milo R. 2016. Sugar Synthesis from CO2 in Escherichia coli. *Cell* **166**:115–125. doi:10.1016/j.cell.2016.05.064

Artier J, Holland SC, Miller NT, Zhang M, Burnap RL. 2018. Synthetic DNA system for structure-function studies of the high affinity CO2 uptake NDH-13 protein complex in cyanobacteria. *Biochim Biophys Acta Bioenerg*. doi:10.1016/j.bbabio.2018.06.015

Axen SD, Erbilgin O, Kerfeld CA. 2014. A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. *PLoS Comput Biol* **10**:e1003898. doi:10.1371/journal.pcbi.1003898

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko K a., Tomita M, Wanner BL, Mori H. 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**:2006.0008. doi:10.1038/msb4100050

Badger MR, Price GD. 2003. CO2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. *J Exp Bot* **54**:609–622.

Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS, Milo R. 2011. The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry*. doi:10.1021/bi2002289

Bathellier C, Tcherkez G, Lorimer GH, Farquhar GD. 2018. Rubisco is not really so bad. *Plant Cell Environ* **41**:705–716. doi:10.1111/pce.13149

Battchikova N, Eisenhut M, Aro EM. 2011. Cyanobacterial NDH-1 complexes: Novel insights and remaining puzzles. *Biochimica et Biophysica Acta - Bioenergetics* **1807**:935–944. doi:10.1016/j.bbabio.2010.10.017

Bauwe H, Hagemann M, Fernie AR. 2010. Photorespiration: players, partners and origin. *Trends Plant Sci* **15**:330–336. doi:10.1016/j.tplants.2010.03.006

Birungi M, Folea M, Battchikova N, Xu M, Mi H, Ogawa T, Aro EM, Boekema EJ. 2010. Possibilities of subunit localization with fluorescent protein tags and electron microscopy examplified by a cyanobacterial NDH-1 study. *Biochimica et Biophysica Acta - Bioenergetics* **1797**:1681–1686. doi:10.1016/j.bbabio.2010.06.004

Bonacci W, Teng PK, Afonso B, Niederholtmeyer H, Grob P, Silver P a., Savage DF. 2012. Modularity of a carbon-fixing protein organelle. *Proc Natl Acad Sci U S A* **109**:478–483. doi:10.1073/pnas.1108557109

Buchanan BB, Gruissem W, Jones RL. 2015. Biochemistry and Molecular Biology of Plants. Wiley.

Cai F, Menon BB, Cannon GC, Curry KJ, Shively JM, Heinhorst S. 2009. The pentameric vertex proteins are necessary for the icosahedral carboxysome shell to function as a CO2 leakage barrier. *PLoS One* **4**:e7521. doi:10.1371/journal.pone.0007521

Cannon GC, Bradburne CE, Aldrich HC, Baker SH, Heinhorst S, Shively JM. 2001. Microcompartments in prokaryotes: carboxysomes and related polyhedra. *Appl Environ Microbiol* **67**:5351–5361. doi:10.1128/AEM.67.12.5351-5361.2001

Cronk JD, Endrizzi J a., Cronk MR, O'neill JW, Zhang KY. 2001. Crystal structure of E. coli beta-carbonic anhydrase, an enzyme with an unusual pH-dependent activity. *Protein Sci* **10**:911–922. doi:10.1110/ps.46301

Cronk JD, Rowlett RS, Zhang KYJ, Tu C, Endrizzi JA, Lee J, Gareiss PC, Preiss JR. 2006. Identification of a novel noncatalytic bicarbonate binding site in eubacterial beta-carbonic anhydrase. *Biochemistry* **45**:4351–4361. doi:10.1021/bi052272q

Dobrinski KP, Longo DL, Scott KM. 2005. The Carbon-Concentrating Mechanism of the Hydrothermal Vent Chemolithoautotroph Thiomicrospira crunogena. *J Bacteriol* **187**:5761–5766. doi:10.1128/JB.187.16.5761-5766.2005

Dou Z, Heinhorst S, Williams EB, Murin CD, Shively JM, Cannon GC. 2008. CO2 fixation kinetics of Halothiobacillus neapolitanus mutant carboxysomes lacking carbonic anhydrase suggest the shell acts as a diffusional barrier for CO2. *J Biol Chem* **283**:10377–10384. doi:10.1074/jbc.M709285200

Du J, Förster B, Rourke L, Howitt SM, Price GD. 2014. Characterisation of Cyanobacterial Bicarbonate Transporters in E. coli Shows that SbtA Homologs Are Functional in This Heterologous Expression

System. *PLoS One* **9**:e115905. doi:10.1371/journal.pone.0115905
Fang Y, Huang F, Faulkner M, Jiang Q, Dykes GF, Yang M, Liu L-N. 2018. Engineering and Modulating Functional Cyanobacterial CO2-Fixing Organelles. *Front Plant Sci* **9**:739. doi:10.3389/fpls.2018.00739
Flamholz A, Prywes N, Moran U, Davidi D, Bar-On Y, Oltrogge L, Savage D, Milo R. 2018. Revisiting tradeoffs in Rubisco kinetic parameters. *bioRxiv*. doi:10.1101/470021
Han X, Sun N, Xu M, Mi H. 2017. Co-ordination of NDH and Cup proteins in CO2 uptake in cyanobacterium Synechocystis sp. PCC 6803. *J Exp Bot* **68**:3869–3877. doi:10.1093/jxb/erx129
Heinhorst S, Cannon GC, Shively JM. 2006. Carboxysomes and Carboxysome-like Inclusions In: Shively JM, editor. Complex Intracellular Structures in Prokaryotes. Berlin, Heidelberg: Springer Berlin Heidelberg. pp. 141–165. doi:10.1007/7171_023
Hopkinson BM, Young JN, Tansik a. L, Binder BJ. 2014. The Minimal CO2-Concentrating Mechanism of Prochlorococcus spp. MED4 Is Effective and Efficient. *Plant Physiol* **166**:2205–2217. doi:10.1104/pp.114.247049
Jorda J, Lopez D, Wheatley NM, Yeates TO. 2013. Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. *Protein Sci* **22**:179–195. doi:10.1002/pro.2196
Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**:845–858. doi:10.1038/nprot.2015.053
Khalifah RG. 1971. The Carbon Dioxide Hydration Activity of Carbonic Anhydrase. *J Biol Chem* **246**:2561–2573.
Lin MT, Occhialini A, Andralojc PJ, Parry MAJ, Hanson MR. 2014. A faster Rubisco with potential to increase photosynthesis in crops. *Nature* **513**:547–550. doi:10.1038/nature13776
Long BM, Hee WY, Sharwood RE, Rae BD, Kaines S, Lim Y-L, Nguyen ND, Massey B, Bala S, von Caemmerer S, Badger MR, Price GD. 2018. Carboxysome encapsulation of the CO2-fixing enzyme Rubisco in tobacco chloroplasts. *Nat Commun* **9**:3570. doi:10.1038/s41467-018-06044-0
Long BM, Rae BD, Rolland V, Förster B, Price GD. 2016. Cyanobacterial CO2-concentrating mechanism components: function and prospects for plant metabolic engineering. *Curr Opin Plant Biol* **31**:1–8. doi:10.1016/j.pbi.2016.03.002
Mackinder LCM, Meyer MT, Mettler-Altmann T, Chen VK, Mitchell MC, Caspari O, Freeman Rosenzweig ES, Pallesen L, Reeves G, Itakura A, Roth R, Sommer F, Geimer S, Mühlhaus T, Schroda M, Goodenough U, Stitt M, Griffiths H, Jonikas MC. 2016. A repeat protein links Rubisco to form the eukaryotic carbon-concentrating organelle. *Proc Natl Acad Sci U S A* **113**:5958–5963. doi:10.1073/pnas.1522866113
Maeda S-I, Badger MR, Price GD. 2002. 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**:425–435.
Mangan NM, Flamholz A, Hood RD, Milo R, Savage DF. 2016. pH determines the energetic efficiency of the cyanobacterial CO2 concentrating mechanism. *Proc Natl Acad Sci U S A* **113**:E5354–62. doi:10.1073/pnas.1525145113
Mangiapia M, USF MCB4404L, Brown T-RW, Chaput D, Haller E, Harmer TL, Hashemy Z, Keeley R, Leonard J, Mancera P, Nicholson D, Stevens S, Wanjugi P, Zabinski T, Pan C, Scott KM. 2017. Proteomic and mutant analysis of the CO2 concentrating mechanism of hydrothermal vent chemolithoautotroph Thiomicrospira crunogena. *J Bacteriol*. doi:10.1128/JB.00871-16
Marcus Y, Schwarz R, Friedberg D, Kaplan A. 1986. High CO(2) Requiring Mutant of Anacystis nidulans R(2). *Plant Physiol* **82**:610–612.
McGrath JM, Long SP. 2014. Can the cyanobacterial carbon-concentrating mechanism increase photosynthesis in crop species? A theoretical analysis. *Plant Physiol* **164**:2247–2261. doi:10.1104/pp.113.232611
Merlin C, Masters M. 2003. Why is carbonic anhydrase essential to Escherichia coli? *J Bacteriol* **185**. doi:10.1128/JB.185.21.6415
Mueller-Cajar O. 2017. The Diverse AAA+ Machines that Repair Inhibited Rubisco Active Sites. *Front Mol Biosci* **4**:31. doi:10.3389/fmolb.2017.00031
Newby ZER, O'Connell JD 3rd, Gruswitz F, Hays FA, Harries WEC, Harwood IM, Ho JD, Lee JK, Savage DF, Miercke LJW, Stroud RM. 2009. A general protocol for the crystallization of membrane proteins for X-ray structural investigation. *Nat Protoc* **4**:619–637. doi:10.1038/nprot.2009.27

Occhialini A, Lin MT, Andralojc PJ, Hanson MR, Parry MAJ. 2016. Transgenic tobacco plants with improved cyanobacterial Rubisco expression but no extra assembly factors grow at near wild-type rates if provided with elevated CO2. *Plant J* **85**:148–160. doi:10.1111/tpj.13098

Ogawa T, Kaneda T, Omata T. 1987. A Mutant of Synechococcus PCC7942 Incapable of Adapting to Low CO2 Concentration. *Plant Physiol* **84**:711–715. doi:10.1104/pp.84.3.711

Price GD. 2011. Inorganic carbon transporters of the cyanobacterial CO2 concentrating mechanism. *Photosynth Res* **109**:47–57. doi:10.1007/s11120-010-9608-y

Price GD, Badger MR. 1989a. 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**:514–525.

Price GD, Badger MR. 1989b. 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**:505–513.

Price GD, Badger MR, von Caemmerer S. 2011. The prospect of using cyanobacterial bicarbonate transporters to improve leaf photosynthesis in C3 crop plants. *Plant Physiol* **155**:20–26. doi:10.1104/pp.110.164681

Rae BD, Long BM, Badger MR, Price GD. 2013. 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**:357–379. doi:10.1128/MMBR.00061-12

Raven JA, Beardall J, Sánchez-Baracaldo P. 2017. The possible evolution, and future, of CO2-concentrating mechanisms. *J Exp Bot*. doi:10.1093/jxb/erx110

Reinhold L, Kosloff R, Kaplan A. 1991. A model for inorganic carbon fluxes and photosynthesis in cyanobacterial carboxysomes. *Can J Bot* **69**:984–988. doi:10.1139/b91-126

Roberts EW, Cai F, Kerfeld CA, Cannon GC, Heinhorst S. 2012. Isolation and characterization of the Prochlorococcus carboxysome reveal the presence of the novel shell protein CsoS1D. *J Bacteriol* **194**:787–795. doi:10.1128/JB.06444-11

Robertson LA, Kuenen JG. 2006. The Genus Thiobacillus In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E, editors. The Prokaryotes: Volume 5: Proteobacteria: Alpha and Beta Subclasses. New York, NY: Springer New York. pp. 812–827. doi:10.1007/0-387-30745-1_37

Roy A, Kucukural A, Zhang Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* **5**:725–738. doi:10.1038/nprot.2010.5

Rubin BE, Wetmore KM, Price MN, Diamond S, Shultzaberger RK, Lowe LC, Curtin G, Arkin AP, Deutschbauer A, Golden SS. 2015. The essential gene set of a photosynthetic organism. *Proceedings of the National Academy of Sciences* 201519220. doi:10.1073/pnas.1519220112

Savir Y, Noor E, Milo R, Tlusty T. 2010. Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape. *Proc Natl Acad Sci U S A* **107**:3475–3480. doi:10.1073/pnas.0911663107

Scott KM, Williams J, Porter CMB, Russel S, Harmer TL, Paul JH, Antonen KM, Bridges MK, Camper GJ, Campla CK, Casella LG, Chase E, Conrad JW, Cruz MC, Dunlap DS, Duran L, Fahsbender EM, Goldsmith DB, Keeley RF, Kondoff MR, Kussy BI, Lane MK, Lawler S, Leigh BA, Lewis C, Lostal LM, Marking D, Mancera PA, McClenthan EC, McIntyre EA, Mine JA, Modi S, Moore BD, Morgan WA, Nelson KM, Nguyen KN, Ogburn N, Parrino DG, Pedapudi AD, Pelham RP, Preece AM, Rampersad EA, Richardson JC, Rodgers CM, Schaffer BL, Sheridan NE, Solone MR, Staley ZR, Tabuchi M, Waide RJ, Wanjugi PW, Young S, Clum A, Daum C, Huntemann M, Ivanova N, Kyrpides N, Mikhailova N, Palaniappan K, Pillay M, Reddy TBK, Shapiro N, Stamatis D, Varghese N, Woyke T, Boden R, Freyermuth SK, Kerfeld CA. 2018. 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**:2686–2708. doi:10.1111/1462-2920.14090

Shibata M, Ohkawa H, Kaneko T, Fukuzawa H, Tabata S, Kaplan A, Ogawa T. 2001. 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**:11789–11794. doi:10.1073/pnas.191258298

Shibata M, Ohkawa H, Katoh H, Shimoyama M, Ogawa T. 2002. Two CO2 uptake systems in cyanobacteria: four systems for inorganic carbon acquisition in Synechocystis sp. strain PCC6803. *Funct Plant Biol*

29:123–129. doi:10.1071/pp01188
Shih PM, Occhialini A, Cameron JC, Andralojc PJ, Parry MAJ, Kerfeld CA. 2016. Biochemical characterization of predicted Precambrian RuBisCO. *Nat Commun* **7**:10382. doi:10.1038/ncomms10382
Supuran CT. 2016. Structure and function of carbonic anhydrases. *Biochem J* **473**:2023–2032. doi:10.1042/BCJ20160115
Tabita FR, Hanson TE, Satagopan S, Witte BH, Kreel NE. 2008. Phylogenetic and evolutionary relationships of RubisCO and the RubisCO-like proteins and the functional lessons provided by diverse molecular forms. *Philos Trans R Soc Lond B Biol Sci* **363**:2629–2640. doi:10.1098/rstb.2008.0023
Tcherkez G. 2016. The mechanism of Rubisco-catalysed oxygenation. *Plant Cell Environ* **39**:983–997.
Tcherkez GGB, Farquhar GD, Andrews TJ. 2006. Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proc Natl Acad Sci U S A*. doi:10.1073/pnas.0600605103
Wetmore KM, Price MN, Waters RJ, Lamson JS, He J, Hoover CA, Blow MJ, Bristow J, Butland G, Arkin AP, Deutschbauer A. 2015. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *MBio* **6**:e00306–15. doi:10.1128/mBio.00306-15
Wheatley NM, Sundberg CD, Gidaniyan SD, Cascio D, Yeates TO. 2014. 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**:7973–7981. doi:10.1074/jbc.M113.531236

Figure Supplements

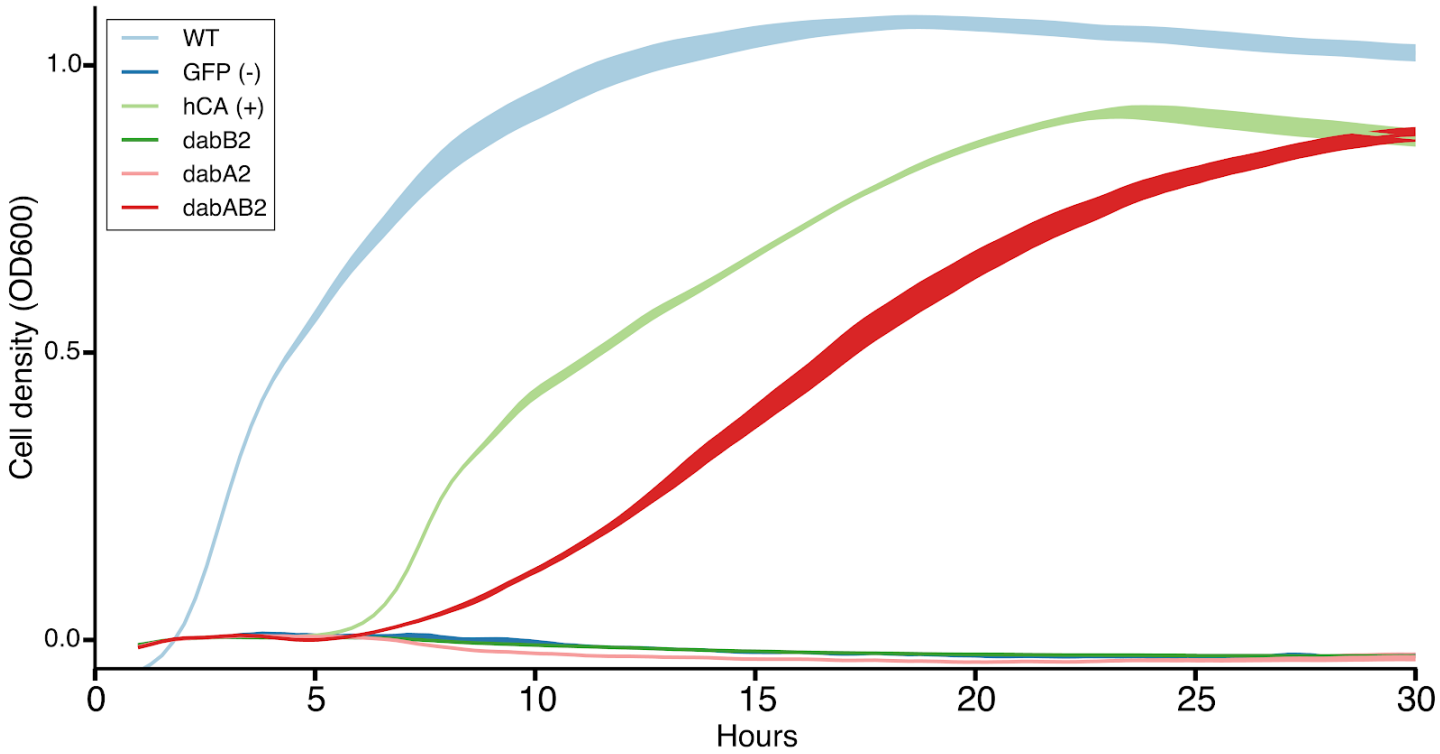

Figure 3 S1. Growth curves of CAfree *E. coli* growth rescue in ambient CO_2. These growth curves were used to generate the growth yield graph in figure 3B. Mean OD600 is graphed +/- standard error for four replicate cultures. Wild-type *E. coli* (BW25113) and CAfree strains expressing either dabAB2 or human carbonic anhydrase II (hCA) grow in ambient CO_2 while CAfree expressing GFP, dabB2 alone, or dabA2 alone fail to grow.

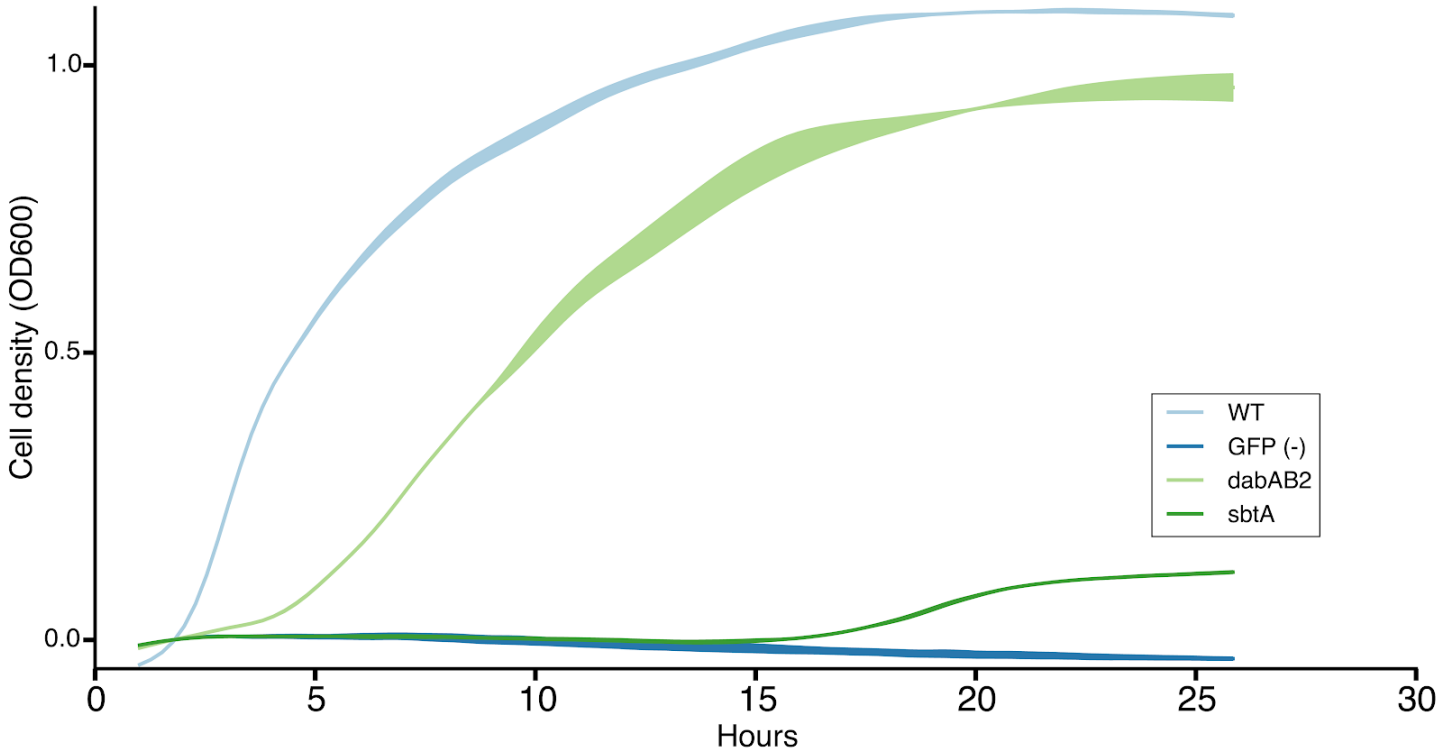

Figure 3 S3. Growth curves of CAfree *E. coli* rescued with dabAB2 or the cyanobacterial HCO_3^- transporter, sbtA. Mean OD600 is graphed +/- standard error for four replicate cultures. Wild-type *E. coli* (BW25113) and CAfree strains expressing dabAB2 grow in ambient CO_2 conditions, while a GFP-expressing negative control fails to grow. Expression of the cyanobacterial HCO3- transporter, sbtA, is noticeably less effective at rescuing CAfree *E. coli* than dabAB2 expression.

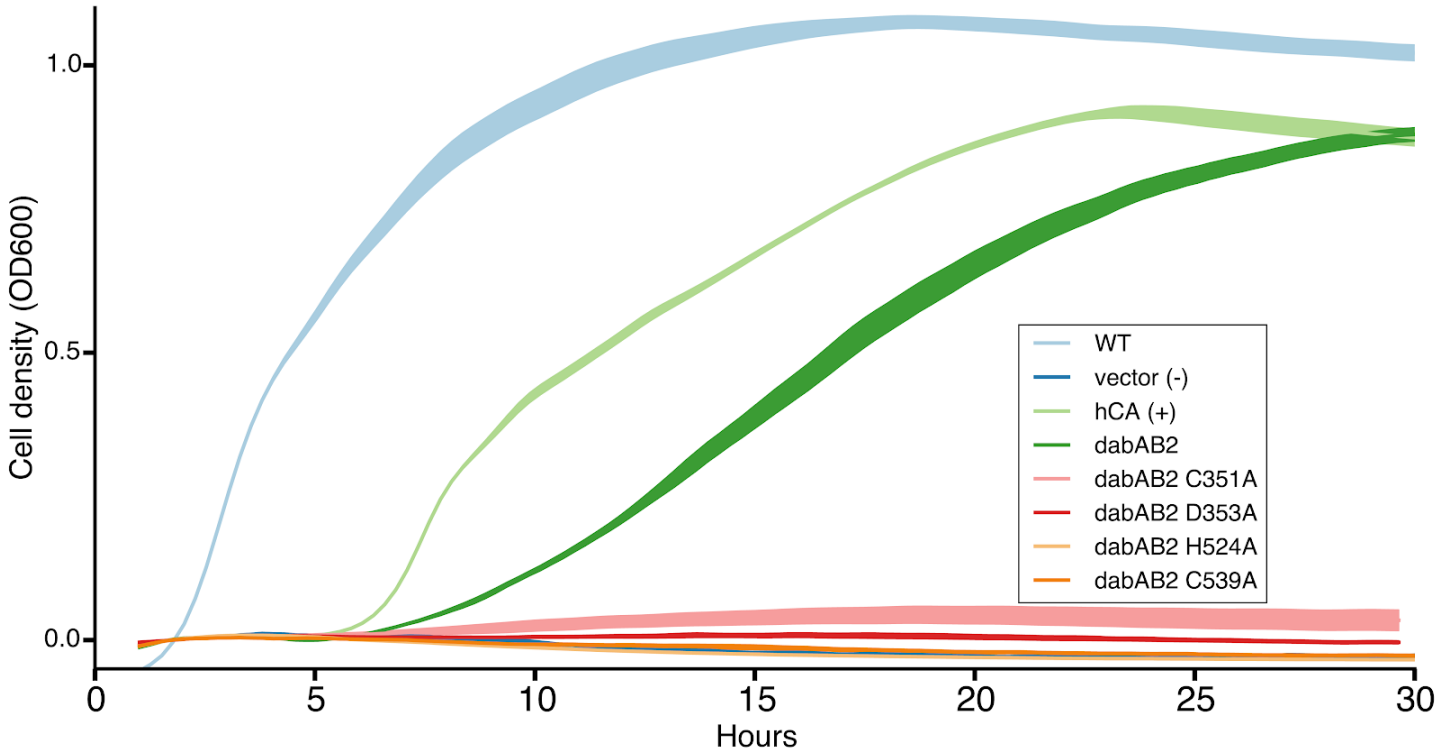

Figure 4 S1. Growth curves of CAfree *E. coli* rescued with dabAB2 active site mutants. These growth curves were used to generate the yield graph in figure 4B. The lines are mean plus and minus standard deviation of four replicate cultures. Wild type cells and those rescued with either dabAB2 or human carbonic anhydrase II (hCA) grow while those rescued with GFP, or active site mutants fail to grow.

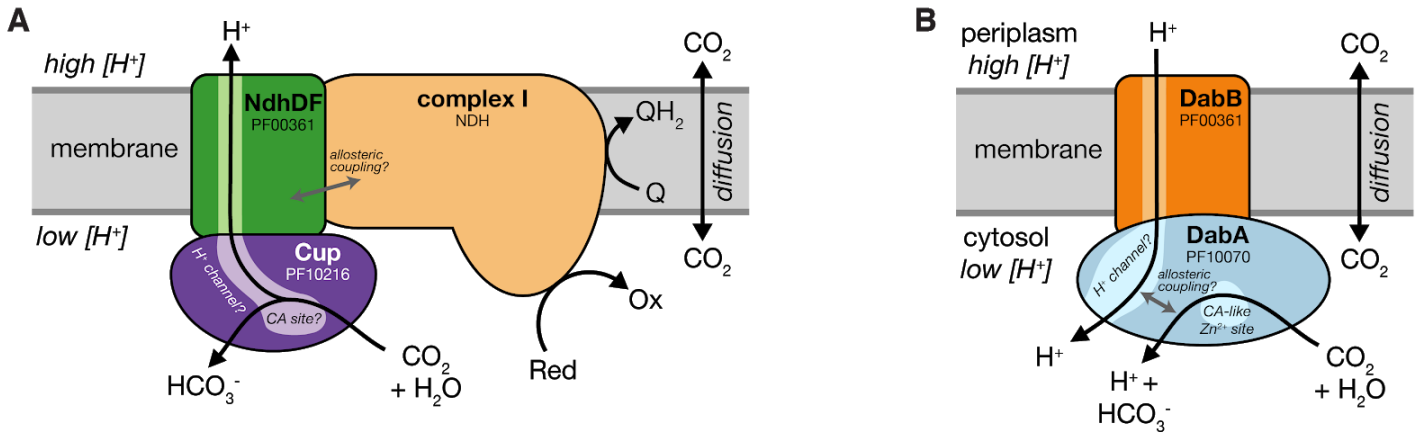

Figure 5 S1. Comparison of models of vectorial CA activity for DABs and the Cyanobacterial Cup systems. A. Cup proteins are CA-like subunits of a class of cyanobacterial Ci uptake systems. Cup-type systems are believed to couple of electron transfer to vectorial CA activity and, potentially, outward-directed 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 to bind CupA/B (illustrated in green as NdhDF). **B.** As our data are consistent with DAB2 functioning as a standalone complex (i.e. DabAB do not appear to bind the *E. coli* complex 1), we propose a different model for DAB function where energy for unidirectional hydration of CO2 is drawn from the movement of cations along their electrochemical gradient (left panel above).

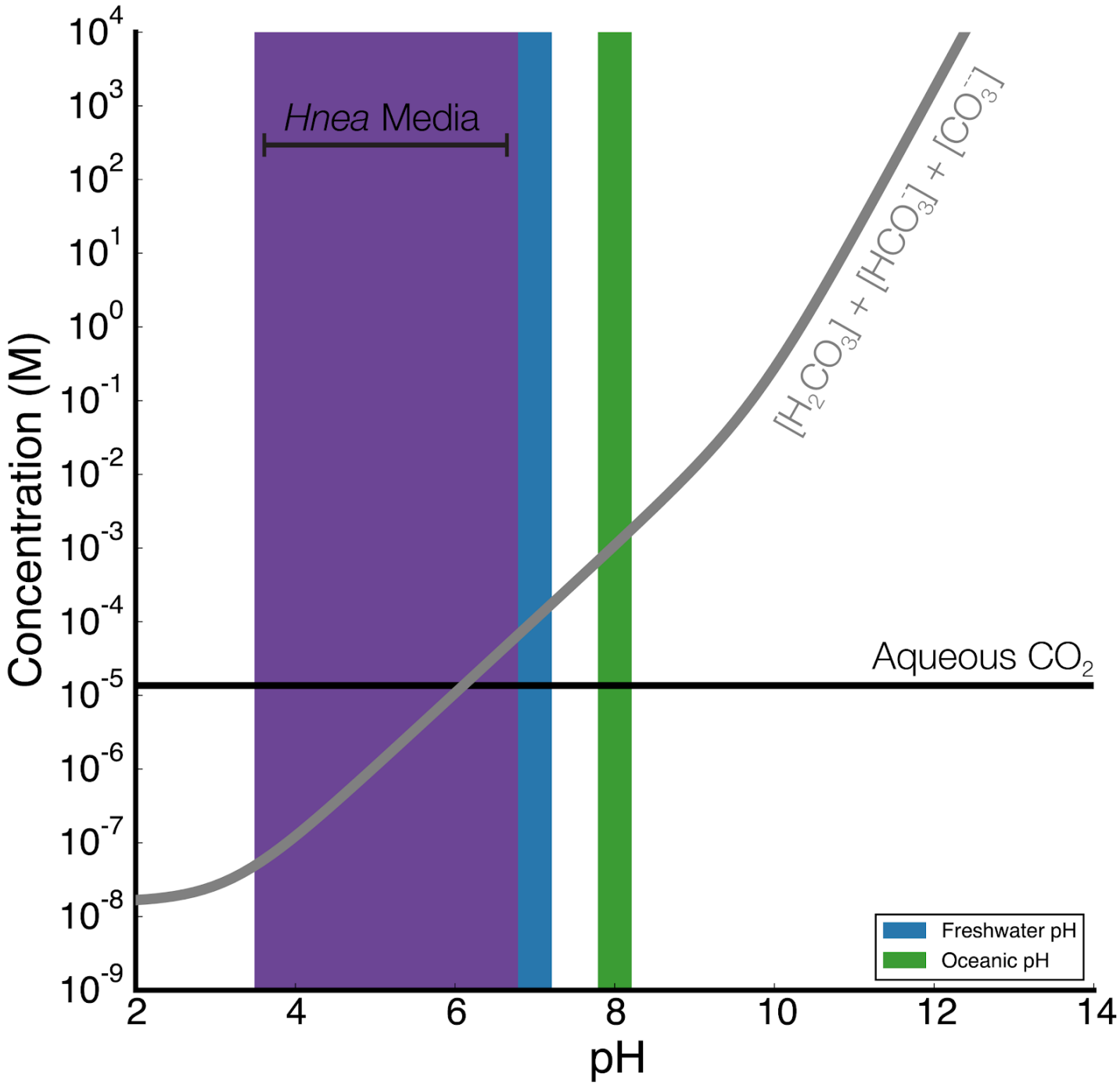

Figure 5 S2. Equilibrium concentrations of dissolved inorganic carbon as a function of pH. In this plot we assume the growth medium is in Henry's law equilibrium with present-day atmosphere (400 PPM CO_2) giving a soluble CO_2 constant of roughly 15 μM. The equilibrium concentrations of hydrated C_i 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 a oceanic pH near 8, HCO_3^- is 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^-.