Title

Elucidation of genes of unknown function in alpha carboxysome operons: acRAF, BMVs and carbon regulatory PII proteins.

Author

Wheatley, Nicole Marie

Publication Date

2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,

Los Angeles

ELUCIDATION OF GENES OF UNKNOWN FUNCTION IN ALPHA CARBOXYSOME OPERONS: ACRAF, BMVs, AND CARBON PII REGULATORY PROTEINS.

A dissertation submitted according to the requirements for the degree Doctor of Philosophy in Molecular Biology

By

Nicole Marie Wheatley

2014

ABTRACT OF THE DISSERTATION

Elucidation of genes of unknown function in alpha carboxysome operons: acRAF, BMVs and carbon regulatory PII proteins.

By

Nicole Marie Wheatley

Doctor of Philosophy in Molecular Biology

University of California, Los Angeles 2014

Professor Todd Yeates, Chair

The chapters of this thesis describe the structure and function of three proteins involved in assembly and function of bacterial microcompartments. Bacterial microcompartments (MCPs) are often thought of as bacterial organelles; they are proteinaceous structures that spatially compartmentalize metabolic reactions for increased catalytic efficiency. There are three types of characterized MCPs: those that catalyze either the catabolic utilization of propanediol (Pdu MCP), ethanolamine (Eut MCP), or carbon dioxide (called *carboxysomes*).

Chapter 1 describes studies of bacterial microcompartment vertex proteins (BMVs) from various microcompartments. BMVs are thought to occupy the 12 vertices of the MCP icosahedron capsule. However, past structural studies of the BMV from *E. coli,* EutN*,* reveal pseudo-hexagonal shapes, whose geometry is unfit for incorporation into pentagonal vertex spaces of microcompartments. We developed a novel method called OCAC (Oligomeric Characterization by Addition of Charge) to probe BMV

oligomeric states in solution. We also endeavored to determine the oligomeric states of the two conserved BMV proteins from alpha-carboxysome operons. The reason for the existence of paralogous BMV proteins in alpha-carboxysome is not understood. The contents and figures of Part-I of Chapter-I were published in *Protein Science* in 2013 (1). Yuxi Lui performed the experiment reported in Figure 2. Soheil Gidaniyan and Nicole Marie Wheatley performed protein expression, purification and crystallization of GrpN and other required proteins. Duilio Cascio provided assistance in solving the GrpN crystal structure. Nicole Marie Wheatley and Todd Yeates coordinated and designed experiments and prepared the manuscript. The contents of Part-II have not been published. Joanna Ngo expressed, purified and performed gels shown in that chapter sub-section. Nicole Marie Wheatley designed and coordinated efforts. Chapter 2 describes the discovery of a novel Rubisco (Ribulose-1,5-bisphosphate carboxylase/oxygenase) chaperone encoded within the operons of alpha-carboxysomes. We named this distant pterin-4a-carbinolamine dehydratase homolog acRAF, for alpha carboxysome Rubisco Assembly Factor. Heterologous co-overexpression of acRAF, GroELS and CbbLS yields greatly increased amounts of soluble, folded Rubisco compared to co-overexpression with GroELS and CbbLS alone. The contents and figures of Chapter-II were published in the Journal of Biological Chemistry in 2014 (2). Christopher Sundberg created the DNA constructs used in Figure 5 of that publication, or Figure 9 of this thesis. Soheil Gidanyian and Nicole Marie Wheatley performed protein expression, purification and crystallization experiments on acRAF. Nicole Marie Wheatley performed the experiment shown in Figure 10A and AB of this thesis. Soheil Gidanyian performed experiment shown in Figure 10C. Duilio

Cascio assisted in solving the crystal structure of acRAF. Todd Yeates made Figure 7A. Nicole Marie Wheatley and Todd Yeates prepared the manuscript. Nicole Marie Wheatley designed and coordinated he efforts of the work. Chapter 3 describes a nitrogen regulatory PII-like protein encoded within several proteobacterial alpha-carboxysome. Crystal structures of this protein, here named ndhPII, display several unique characteristics compared to canonical nitrogen regulatory PII proteins. We structurally analyze both unbound and ADP-bound ndhPII crystal structures, and speculate about its regulation of inorganic carbon assimilation in these organisms. The work contained in this chapter is not yet published. Joanna Ngo and Nicole Marie Wheatley performed protein expression, purification and crystallization of ndhPII. Duilio Cascio and Michael Sawaya solved and refined both crystal structures. Nicole Marie Wheatley designed and coordinated the efforts of this work.

The dissertation of Nicole Marie Wheatley is approved.

David Eisenberg

Beth Lazazzera

Todd Yeates, Committee Chair

University of California, Los Angeles

2014

Dedicated to CDS

Table of Contents

	Dissertation Abstract	ii
	Thesis Committee Page	v
	Dedication Page	vi
	Table of Contents (this page)	vii
	List of Acronyms	viii
	List of Figures, Tables and Symbols	ix
	Acknowledgements	xi
	Curriculum Vitae	xii
Chapter I	Bacterial Microcompartment Vertex Proteins, BMVs	1
Part I.	Bacterial Microcompartment Shells of Diverse Functional Types Possess Pentameric Vertex Proteins	1
Part II	A short report regarding the ambiguous oligomeric state alpha-carboxysomes BMVs	15
Chapter II	Structure and identification of a pterin dehydratase-like protein as a Rubisco assembly factor in the alpha-carboxysome	20
Chapter III	Putative Carbon Regulatory PII proteins of alpha carboxysomes	46
	Thesis Conclusion	68
	Appendix	73
	BIBILOGRAPHY	77

List of Acronyms

BMC – bacterial microcompartment (hexameric) shell protein

BMV – bacterial microcompartment vertex protein (pentameric)

Grp – glycyl radical-based propanediol utilizing microcompartment

MCP – bacterial microcompartment

OCAC – oligomeric characterization by addition of charge

OCAM - oligomeric characterization by addition of mass

acRAF - alpha-carboxysome Rubisco Assembly Factor

PCD - pterin-4a-carbinolamine dehydratase

DHPR - dihydropterin reductase

BMC - bacterial microcompartment shell protein

Rubisco- Ribulose-1,5-bisphosphate carboxylase/oxygenase

ndhPII – nitrogen regulatory PII-like protein encoded within Ndh 4-member gene cluster

2-OG – 2-oxoglutarate, or alpha-ketoglutarate

nrPII - nitrogen regulatory PII protein

crPII - putative carbon regulatory PII protein

Eut – ethanolamine utilization (microcompartment)

Pdu – propanediol utilization (microcompartment)

SDS-PAGE – sodium dodecyl sulfate polyacrylamide gel electrophoresis

RMSD - Root-Mean-Square Deviation

PFOR – pyruvate-ferredoxin oxidoreductase, or pyruvate synthease

Ci – inorganic carbon: carbon dioxide and/or bicarbonate

List of Figures, Tables and Symbols

Figure 1	Comparison of the structure of GrpN with other bacterial microcompartment vertex (BMV) proteins.	Page 05
Figure 2	Application of the OCAC (oligomeric characterization by the addition of charge) method to the EutN shell protein.	Page 09
Figure 3	BVM-1-MBP and BMV-2 characterization by SDS-PAGE	Page 17
Figure 4	BMV2-MBP Native Mass Spectrometry reveals monomer, dimer, trimer, tetramer and hexameric assemblies	Page 18
Figure 5	AcRAF is a defining member of a-carboxysome operons	Page 24
Figure 6	Crystal structure and conserved surface regions in the acRAF dimer	Page 35
Figure 7	Structural comparisons of acRAF and PCD	Page 37
Figure 8	Ferredoxin-like protein folds associated with carboxysomes	Page 38
Figure 9	acRAFs do not compliment PCD activity *in vivo*	Page 41
Figure 10	acRAF increases the quantity of natively assembled Rubisco when heterologously co-expressed with GroELS in *E. coli*	Page 43
Figure 11	Mechanism of inhibition of ammonium uptake by PII under nitrogen sufficient conditions	Page 51
Figure 12	NdhPII genetic linkage to proteobacterial alpha-carboxysome and other operons	Page 58

Figure 13 Structural comparison of nrPIIs and putative carbon regulatory PII proteins, crPII proteins Page 60
Figure 14 Apo-ndhPII aligned to ADP-ndhPII Page 61
Figure 15 Cartoon guide for interpretation of the gels shown in Figure 16 Page 74
Figure 16 Native PAGE of Eut MCP shell protein mixtures Page 75

List of Tables

Table 1 X-ray Data Collection and Model Refinement Statistics of GrpN Page 07
Table 2 X-ray data collection and model refinement statistics of acRAF Page 33
Table 3 Genetic Linkages of PII proteins Page 54
Table 4 X-ray Data Collection and Model Refinement Statistics of ADP-bound Apo-ndhPII models Page 62

List of Symbols

˚C degree Celsius

Acknowledgments

We thank Dr. Andrew Hanson and Mike Ziemak for providing PCD constructs, Dr. Sabeeha Merchant and Dr. Rikard Fristedt for valuable discussions and *Arabidopsis* constructs, Drs. Alice Barkan, postdoc, David Stern, Michael Andrew and José Ramón Botella for discussions on acRAF proteins in plants. We thank Dr. Thomas Bobik for help discussions about microcompartments and microbiology. We thank Dr. Robert Gusalus for use of facilities and helpful discussion. We thank M. Capel, K. Rajashankar, N. Sukumar, J. Schuermann, I. Kourinov and F. Murphy at NECAT for synchrotron X-ray assistance. Use of the APS is supported by the DOE under contract DE-AC02-06CH11357. We thank Mike Collazo, Michael Sawaya and Duilio Cascio at the UCLA-DOE X-ray Crystallization and Crystallography Core Facilities, which is supported by DOE Grant DE-FC02-02ER63421. This work was supported by a NIH grant R01AI08114 to TOY and by the BER program of the DOE Office of Science. NMW was supported by the Ruth L. Kirschstein National Research Service Award GM007185 and the Philip Whitcome Pre-Doctoral Training Fellowship. Christopher D Sundberg was supported by the NSF Graduate Research Fellowship Program. Rikard Fristedt

Curriculum VItae

September 2014

EDUCATION

Microbiology and Immunology; University of California, Irvine. GPA 4.0

PROFESSIONAL EXPERIENCE

2009-*Present*: Graduate Student. MBIDP Ph.D. Program, University of California, Los Angeles. Todd Yeates, Ph.D. PI.

2008-2009: Undergraduate Researcher. Celia Goulding Ph.D., Supervisor. University of California, Irvine. *Structural Studies in M. tuberculosis.*

REASEARCH FELLOWSHIP AWARDS

2013-2014 Philip Whitcome Pre-doctoral Training Fellowship. University of California, Los Angeles

2010-2013 Cellular and Molecular Biology Training Grant. University of California, Los Angeles

PUBLICATIONS and CURRENT MANUSCRIPTS

Thompson MC, **Wheatley NM**, Jorda J, Sawaya MR, Gidaniyan SD, Ahmed H, Yang Z, McCarty KN, Whitelegge JP, Yeates TO. *Identification of a unique fe-s cluster binding site in a glycyl-radical type microcompartment shell protein*. J Mol Biol. **2014** Sep 23;426(19):3287-304. doi: 10.1016/j.jmb.2014.07.018. Epub 2014 Aug 4. PubMed PMID: 25102080

Khachatoorian R, Ganapathy E, Ahmadieh Y, **Wheatley N,** Sundberg C, Jung CL, Arumugaswami V, Raychaudhuri S, Dasgupta A, French SW. *The NS5A-binding heat*

shock proteins HSC70 and HSP70 play distinct roles in the hepatitis C viral life cycle. Virology. **2014** Apr;454-455:118-27. doi: 10.1016/j.virol.2014.02.016. PMID: 24725938

Wheatley NM, Sundberg CD, Gidaniyan SD, Cascio D, Yeates TO. *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. **2014** Mar 14;289(11):7973-81. doi: 10.1074/jbc.M113.531236. PubMed PMID: 24459150

Wheatley NM, Gidaniyan SD, Liu Y, Cascio D, Yeates TO. *Bacterial microcompartment shells of diverse functional types possess pentameric vertex proteins.* Protein Sci. **2013** May;22(5):660-5. doi: 10.1002/pro.2246. PMID: 23456886

Jorda J, Lopez D, **Wheatley NM**, Yeates TO. *Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria.* Protein Sci. **2013** Feb;22(2):179-95. doi: 10.1002/pro.2196. PubMed PMID: 23188745

Torres R, Swift RV, Chim N, **Wheatley N**, Lan B, Atwood BR, Pujol C, Sankaran B, Bliska JB, Amaro RE, Goulding CW. *Biochemical, structural and molecular dynamics analyses of the potential virulence factor RipA from Yersinia pestis*. PLoS One. **2**

TEACHING ACTIVITIES

Teaching Assistant - University of California, Los Angeles. Introduction to Biochemistry: CHEM 153A. Spring 2011.

Teaching Assistant - University of California, Los Angeles. Introduction to Biochemistry: CHEM 153A. Fall 2011

CHAPTER I : BACTERIAL MICROCOMPARTMENT VERTEX PROTEINS, BMVs

PART I : Bacterial Microcompartment Shells of Diverse Functional Types Possess Pentameric Vertex Proteins

ABSTRACT

Bacterial microcompartments (MCPs) are large proteinaceous structures comprised of a roughly icosahedral shell and a series of encapsulated enzymes. Experimental studies have characterized MCPs carrying out three different metabolic functions, and bioinformatics studies have implicated other types, including one believed to perform glycyl radical-based metabolism of 1,2-propanediol (Grp). Here we report the crystal structure of a protein (GrpN), which is presumed to be part of the shell of a Grp-type MCP in *Rhodospirillum rubrum* F11. GrpN is homologous to a family of proteins (EutN/PduN/CcmL/CsoS4) whose members have been implicated in forming the vertices of MCP shells. Consistent with that notion, the crystal structure of GrpN revealed a pentameric assembly. That observation revived an outstanding question about the oligomeric state of this protein family: pentameric forms (for CcmL and CsoS4A) and a hexameric form (for EutN) had both been observed in previous crystal structures. To clarify these confounding observations, we revisited the case of EutN. We developed a molecular biology-based method for accurately determining the number of subunits in homo-oligomeric proteins, and found unequivocally that EutN is a

pentamer in solution. Based on these convergent findings, we propose the name bacterial microcompartment vertex (BMV) for this special family of MCP shell proteins.

INTRODUCTION

Microcompartments (MCPs) are polyhedrally shaped supramolecular protein assemblies that physically encapsulate select metabolic pathways in prokaryotes (reviewed in refs. (3–7)). This encapsulation serves to concentrate and separate specific enzymes and their metabolic intermediates from the cytoplasm, thereby increasing reaction efficiency, retaining volatile intermediates, and/or protecting the cellular milieu from toxic intermediates.(7) The carboxysome MCP increases carbon-fixation efficiency by encapsulating carbonic anhydrase and Rubisco together, so the CO_2 produced by the first enzyme can be delivered at high concentration to the second enzyme.(6, 8) Cofactor B_{12}-dependent MCPs for ethanolamine utilization (Eut) and 1,2-propandiol utilization (Pdu) function by retaining volatile or toxic aldehyde compounds – propionaldehyde and acetaldehyde – that occur as intermediates in those pathways. (7, 9–13) Bioinformatic studies have suggested the existence of an MCP for glycyl radical-based 1,2-propanediol metabolism, referred to hereafter as Grp.(14) Sequence analysis suggests that the Grp MCP, like the Pdu MCP, houses enzymes for metabolizing 1,2-propanediol, but that it uses a glycyl radical enzyme for the key dehydration reaction rather than B_{12}-dependent enzymes.(14) The Grp MCP has not yet been studied biochemically.

Structural analyses have answered a number of questions regarding the geometry and mechanism of MCP assembly and function (reviewed in ref. (4)). Current models describe MCPs as being comprised of sheets of hexameric shell proteins (belonging to the BMC family of proteins) forming polyhedral facets (15–17), along with a much smaller number of special pentameric proteins (belonging to the EutN/PduN/CcmL/CsoS4 protein family) placed at the vertices of the polyhedral shell.(18) While crystallographic studies of (hexameric) BMC shell proteins have consistently supported this model of MCP organization, experiments on the putative vertex proteins have provided a less coherent picture. Two crystal structures of carboxysome vertex proteins, CcmL and CsoS4A, are indeed pentameric(18) (PDB accession code 2QW7 and 2RCF). However, a third structure of a presumptive vertex protein, EutN from *E. coli* (PDB code 2Z9H), revealed a hexameric quaternary assembly with nearly hexagonal shape (19) Likewise, multiple genetic studies have shown that deleting the presumptive vertex proteins compromises the formation of closed polyhedral shells (20–22), yet closed shells can apparently be formed with lower efficiency in some cases (23). These equivocal results have complicated the interpretation of the architectural role of this protein family.

Electron microscopy studies indicate that carboxysomes tend to be more geometrically regular and more nearly icosahedral in shape compared to other types of MCPs that have been visualized(24, 25); this has raised the possibility that pentameric units required for symmetric icosahedral architecture might be present only in carboxysome MCPs. Here we provide evidence that this family of proteins serves as the pentameric

vertex element across multiple divergent types of MCPs. Specifically, we report the pentameric crystal structure of GrpN from the Grp MCP, and show also that EutN is a pentamer in solution, contrary to previous crystallographic findings.

RESULTS AND DISCUSSION

An operon encoding enzymes and shell proteins for a presumptive Grp MCP has been described in *R. rubrum* (14). One of the encoded proteins, referred to hereafter as GrpN (Gene ID: 12642037), was identified as a representative of the EutN/PduN/CcmL/CsoS4 family of shell proteins (Pfam domain: PF03319), hereafter described as bacterial microcompartment vertex (or BMV) proteins. GrpN was overexpressed in *E. coli* and purified by metal affinity chromatography on the basis of a polyhistidine tail added to the protein. Single crystals were obtained in space group I23, and an X-ray structure was determined by molecular replacement and refined at a resolution of 3.2 Å. The final atomic model is 96% complete; residues 65-70 are in a disordered loop and could not be modeled accurately.

As expected, the tertiary fold follows closely that reported earlier for other members of this protein family (Fig 1). The core structure of GrpN is comprised of five antiparallel beta strands that curve to create a small beta barrel. The C-termini and loop regions between beta strands extend away from the beta-barrels, making contacts with adjoining subunits. The atomic coordinates align well to CcmL, CsoS4A and EutN, with RMSD values of 0.73 Å, 0.81 Å, and 0.80 Å, respectively, over C-alpha positions.

Stabilization of the quaternary structure is largely mediated by contacts between loop regions from adjacent beta-barrels; minor contributions are made by barrel-to-barrel contacts. Structural comparisons between BMV structures show that the N-termini and the small beta barrels are highly conserved, while the C-termini and regions adjoining the core beta sheets differ slightly in length and secondary structure content. For example, GrpN and CsoS4A lack C-terminal beta hairpins present in both CcmL and EutN.

Figure 1

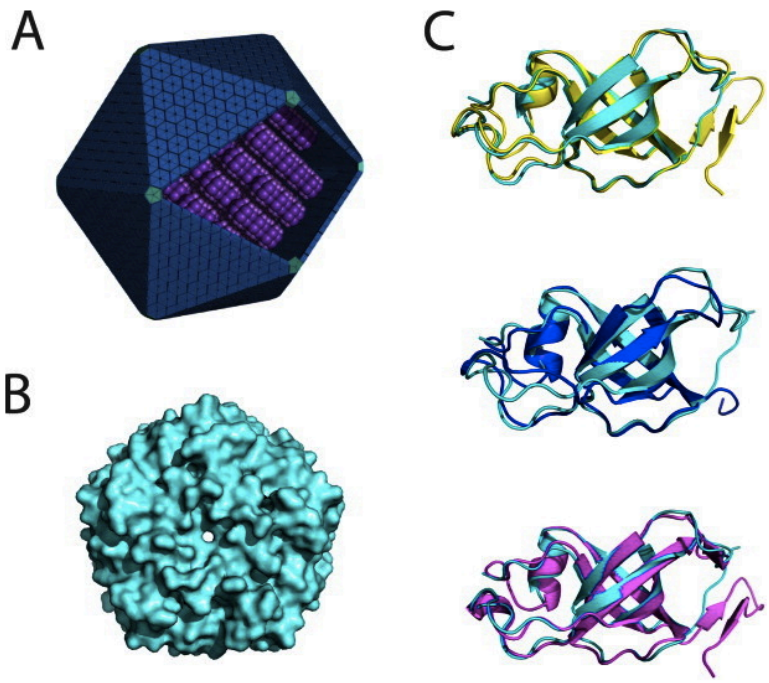


Figure 1. Comparison of the structure of GrpN with other bacterial microcompartment vertex (BMV) proteins. (A) An idealized model of an MCP showing pentameric units at the vertices of the polyhedral shell. (B) Space-filling model of the pentameric structure of GrpN. (C) Superposition of a GrpN monomer with CcmL (top, yellow), CsoS4A (middle, blue), and EutN (bottom, magenta).

GrpN crystallized as a pentamer in the asymmetric unit of the crystal (Table 1). The pentameric unit of GrpN matches those reported earlier for CcmL from the beta-type carboxysome and CsoS4A from the alpha-type carboxysome. One surprising aspect of the GrpN crystal structure is the arrangement of pentamers in the unit cell. Twelve crystallographically-related pentamers are arranged in a nearly icosahedral fashion. This 60-subunit icosahedral arrangement is unlikely to reflect the natural assembly of the much larger MCP; in native MCPs, pentameric units are presumably surrounded by BMC-type hexamers. In addition, the interactions between pentamers in the crystal involve relatively small sites of contact. Nonetheless, this spontaneous assembly could be an interesting starting point for engineering a novel icosahedral cage. The design of self-assembling protein cages by various methods has been discussed in a series of recent studies (26–30) The observation that GrpN is a pentamer establishes that other MCPs besides carboxysomes possess special pentameric shell proteins to explain the closure of an otherwise flat layer of hexameric units provided by the BMC-type proteins.

Table 1

X-ray Data Collection and Model Refinement Statistics

Statistics	**Value**
Wavelength (Å)	0.9791
Resolution range (Å)	19.8–3.2 (3.314–3.2)
Space group	I23
Unit cell (Å)	$a = b = c = 150.7$
Total reflections recorded	49,986 (3856)
Unique reflections	9,248 (700)
Multiplicity	5.4 (5.5)
Completeness (%)	96.6 (99.6)
Mean I/sigma(I)	21.0 (3.0)
Wilson B-factor (Å^2)	84.4
R-meas[25]	7.3% (62%)
Model R-work	0.267 (0.340)
Model R-free	0.288 (0.381)
Number of non-hydrogen atoms	2,930
Macromolecules	2,926
Chloride	3
Water	1

Statistics	**Value**
Protein residues	423
Geometric deviations (rms)	
Bonds (Å)	0.004
Angles (°)	0.69
Ramachandran favored (%)	95
Ramachandran outliers (%)	0
Clashscore[26]	3813.35
Average B-factor (Å^2)	87.6
Protein atoms	87.6
Solvent (water)	82.2

Statistics for the highest-resolution shell are shown in parentheses.

Our finding that GrpN forms a pentameric assembly left the earlier reported hexameric structure for *E. coli* EutN as an outlying observation. The hexameric state reported for EutN could either indicate an unusual structural role for this homolog, or it could represents a spurious or minor oligomeric form selected during crystallization. To distinguish between these two possibilities, we devised a novel method to determine the oligomeric state of EutN in solution (Fig 2).

Figure 2

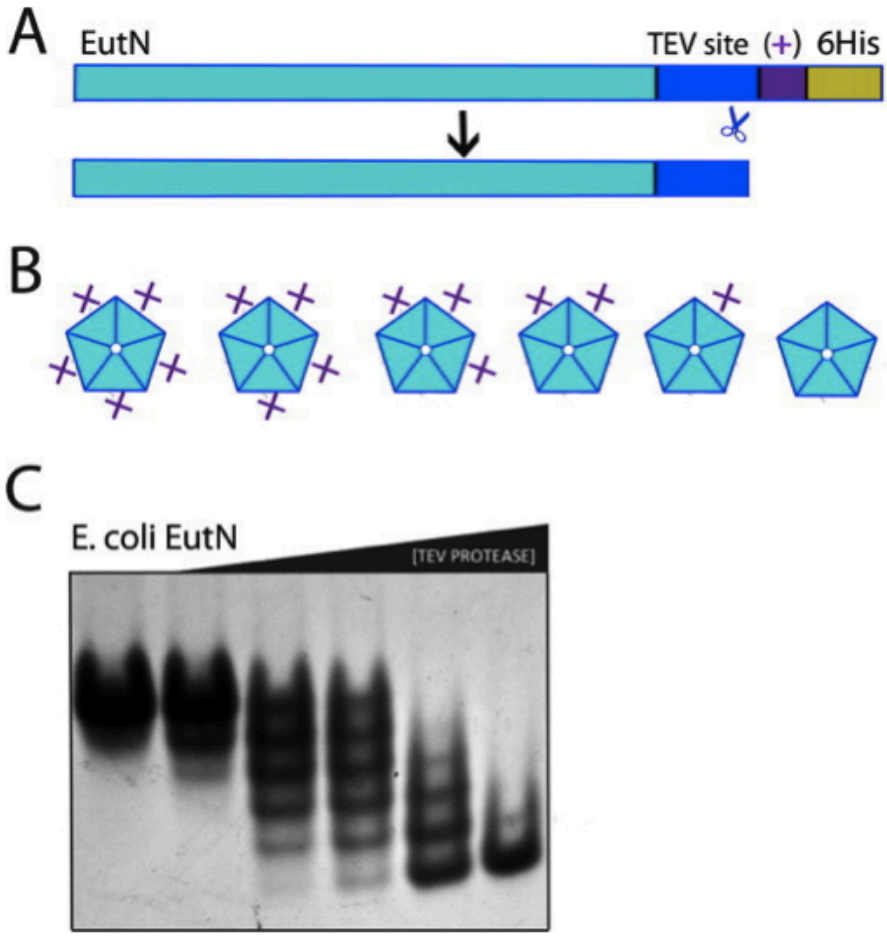


Figure 2. Application of the OCAC (oligomeric characterization by the addition of charge) method to the EutN shell protein. (A) Diagram of primary structure of EutN before and after cleavage with TEV protease. The plus symbol " + " denotes one additional positive charge. (B) For an *n*-oligomer, *n + 1* possible charge states exist. Shown are the six possible charge-states of a pentamer. (C) EutN OCAC native gel. The engineered EutN(+) oligomer was incubated with TEV protease for 15 min, quenched with iodoacetamide, and then run on a native gel. From left to right, TEV protease concentrations are 0.0, 0.002, 0.01, 0.02, 0.1, and 0.5 mg/mL. The presence of six distinct bands shows that EutN is a pentamer.

The method we developed for determining oligomeric states is based on the change in charge-to-mass ratio that occurs after proteolytic cleavage of a small, charged N- or C-terminal sequence that is genetically appended at the end of the native protein subunit

(Fig 2A). Following different degrees of exposure to protease, a homo-oligomer composed of *n* subunits can be partially proteolyzed to generate a total of *n+1* distinct charge forms. The distinct charge forms can be enumerated by native gel electrophoresis. To be consistent with the nomenclature of established methods, we refer to our approach as OCAC, an acronym for oligomeric characterization by addition of charge (Fig 2B). A related method described earlier, OCAM (oligomeric characterization by addition of mass), was developed to determine the oligomeric state of membrane proteins (31). OCAM determines oligomeric state based on changes in the mobility of protein complexes on blue native gels, which arise from differential removal of relatively massive domains fused to the native protein subunits. In contrast, OCAC requires addition of relatively small terminal extensions to the native protein: essentially a protease cleavage site and as few as 1 or 2 directly adjacent charged amino acids. As with OCAM, the success of OCAC depends on a slow rate of exchange between subunits in different oligomeric complexes. Under conditions where we applied the method, native gels were capable of distinguishing between oligomers with just one unit charge difference. Another method based on counting distinct oligomeric forms was described several years ago, wherein the chemical modification of a genetically engineered cysteine residue by a charged reagent resulted in mobility shifts on SDS-PAGE (32). Unlike this earlier method, OCAC uses limited proteolysis to create alternative charge forms and does not require the complex to be stable in the presence of sodium dodecyl sulfate. The OCAC method also avoids the potential challenges of using chemical approaches when multiple cysteine residues are present in a protein.

We applied the OCAC method to examine the oligomeric state of EutN, which we had earlier reported to be a hexamer based on its crystal structure. EutN from *E. coli* was cloned to include a C-terminal TEV site followed by LEKK-6His. The resulting net charge difference between uncleaved and cleaved EutN is one charge unit (at pH 8.6 of the native gel running buffer). After recombinant expression and purification, this construct was subjected to a series of increasing concentrations of TEV protease, and then run on a native gel.

Surprisingly, the results clearly indicate that EutN is a pentamer in solution (Fig 2C). Six sharp bands are obtained by separation on a native gel. These arise from oligomers having 0, 1, 2, 3, 4, or 5 tails cleaved. This pentameric behavior contradicts two independent crystal structures in the PDB showing EutN in a hexameric arrangement (19)(PDB ID 2Z9H and 2HD3). A comparison of conditions used to obtain crystals of EutN did not suggest a clear explanation for the observed difference between crystalline and solution oligomer states. Although we see no evidence in the OCAC method for hexameric assembly, it is possible that a minor hexameric species of EutN exists in solution, which happens to crystallize preferentially. If the protein can equilibrate between pentameric and (minor) hexameric forms in solution, a low energy crystal form could drive the protein to a hexameric configuration.

Our combined results demonstrating the pentameric states of both GrpN and EutN resolve a looming doubt regarding a particular aspect of microcompartment formation. While the role of BMC type hexamers in forming the flat facets of MCPs has been

apparent since the first structural studies (4, 15), the universality of pentameric bacterial microcompartment vertex (BMV) proteins in MCPs has remained uncertain (19, 23). Here we confirm the first pentameric assemblies of BMV proteins from MCPs besides the carboxysome: GrpN from *Rhodospirillum rubrum* by X-ray crystallography, and EutN from *E. coli* in solution by the OCAC method. These results help reunify ideas for how different types of MCPs are constructed.

MATERIAL AND METHODS

Cloning

A gene sequence was designed, using codons optimized for expression in *E. coli*, to encode GrpN with a C-terminal hexahistidine tag, using the online program DNAWorks.(33) The resulting nucleotide sequence was synthesized by Biomatik, and then transferred to pET22b(+) vector using the NdeI restriction site via Isothermal Assembly (a.k.a. Gibson Assembly) (34) We followed the isothermal protocol utilizing 20 nucleotide base pair complementary overhangs.

Full length EutN was amplified from *E. coli* genomic DNA. Synthetic DNA oligomers, purchased from IDT Inc., were used to add a TEV cleavage site and the positively charged tail onto the C-terminus of EutN with PCR. We refer to the resulting protein as EutN(+). Using Isothermal Assembly, EutN(+) was transferred into a pET22b(+) vector between NdeI and XhoI restriction sites in order to append a C-terminal 6xHis tag. The sequences of both GrpN and EutN(+) were verified by Laragen, Inc.

Expression and Purification

The expression of GrpN-6His and EutN(+) were induced with 1mM IPTG in BL21 cells, shaking at 250 rpm, for 3 to 5 hours at 37˚C. Cells were spun down for 5 min at 6,000 rpm and stored at -20˚C. Cells were suspended in 50mM Tris-HCl pH 7.6, 300mM NaCl, 20 mM imidazole, with Protease Inhibitor Cocktail (Sigma, Cat # P8849) and lysed by sonication. Cells were spun down in rotor SS-34 at 16,500 rpm for 30 min, filtered through a 0.2um filter, and applied to a Hi-trap Nickel column by syringe at room temperature. Protein was eluted in one step with of 50mM Tris–HCl pH 7.6, 300 mM NaCl, 400mM Imidazole. GrpN was dialyzed into 2L of 10mM Tris-HCl pH 7.6, 20 mM NaCl for 1 hour at 4˚C, and then again in 2L of fresh buffer overnight. GrpN-6His tended to precipitate even at 4˚C, so the protein solution was subjected to centrifugation at 20,800 rcf for 1-2 minutes throughout the protein preparation process to pellet protein precipitate.

Crystal Structure Determination

Initial crystallization screens were performed in 96-well, hanging drop trays, set up with the nanoliter liquid handling Mosquito from TTP LabTech. Upon optimization of condition G8 from Hampton Research screen HR2-110, cube-shaped crystals were obtained within the following condition ranges: 1.4 M - 1.6 M ammonium sulfate, 0.1 M NaCl, 0.1 M HEPES pH 7.0 – 7.6, at protein concentrations between 20 mg/mL and 40 mg/mL. Hanging drops were 1:1 well: protein with a total drop size between 2 and 4 uL. Crystals generally took between 1 and 10 days to grow.

Diffraction data extending to 3.2 Å resolution were collected at the Argonne National Laboratory, Advanced Photon Source (APS), beamline 24-ID-C. The structure of GrpN was phased by molecular replacement using the program PHASER (35). Coordinates for the CcmL pentamer (PDB accession code 2QW7) were used as the search model. The structure was built using the program COOT (36) and refined using PHENIX (37) and BUSTER (38) with a final R_{work} and R_{free} of 0.2671 and 0.2885 respectively. 95% of the backbone dihedral angles are within the favored regions of a Ramachandran diagram. Coordinates and structure factors have been deposited with the PDB with ID code 4I7A.

OCAC Assay

Aliquots of purified EutN(+) (in 50mM Tris-HCl pH 7.6, 300mM NaCl and 300mM imidazole) were incubated on ice with fresh 1mM DTT for 30 minutes. Due to weak UV absorbance at 280 nm, concentrations were adjusted based on band intensity visualized by SDS-PAGE. TEV protease was added to EutN(+) aliquots in a series of dilutions, resulting in the following final concentrations: 0.0, 0.002, 0.01, 0.02, 0.1, 0.5 mg/mL TEV protease. These TEV: EutN(+) samples were incubated at room temperature for 15 minutes. Protease reactions were stopped with addition of 5X native loading dye containing 10mM iodoacetamide. Reactions were then run on a native gel (BioRad CAT# 456-1096) at 100V volts for 2 hours at room temperature.

PART II: A short report regarding the ambiguous oligomeric state alpha-carboxysomes BMVs

Following the publication of the our manuscript “Bacterial Microcompartment Shells of Diverse Functional Types Possess Pentameric Vertex Proteins”, I stumbled upon a startling fact that there exist two putative BMV shell proteins encoded in alpha-carboxysome operons. To my knowledge, alpha-carboxysome operons are the only type of microcompartment encoding more than one putative BMV shell protein. Even though it is very common for MCP operons to encode several paralogous BMC shell proteins, it is not common for an operon to have paralogous BMV genes. I attempted to perform OCAC on both BMV shell proteins from *T. intermedia* K12, in order to determine their oligomeric states in solution. Unfortunately, the proteins proved to be highly unstable when heterologously overexpressed in *E. coli* BL21(DE3). Co-overexpression of both BMVs did not relieve the problem. C-terminal MBP fusions were then made of each BMV shell protein, and this seemed to stabilize the protein well enough for purification by Nickel column chromatography (via the poly-histidine tag on the C-terminus of MBP).

BMV1-MBP forms large aggregates that are not dissolved by 15 minute incubations at 95˚C in 1% SDS. This construct was not amenable to OCAC or OCAM. Figure 3a shows SDS-PAGE samples of an attempted BMV1-MBP OCAM (31) experiment. Figure 3b shows an SDS-PAGE containing samples of whole cell lysates and soluble fractions of cells overexpressing ether BMV1-MBP-6His and BMV2-6His (20˚C, 24 hours,1 mM IPTG in LB media). No BMV2-6His is detected; BMV1-MBP appears only whole cell lysates and does not fully enter the gel.

C-terminal MBP fusion to BMV2 proved to be much more stable. BVM2-MBP enters SDS-PAGE gels, and appears at the predicted monomeric molecular weight (not shown). Purified BMV2-MBP protein (in 100 mM ammonium acetate pH 7) was sent to Dr. Laganowski in Carol Robinson's laboratory for analysis by native mass spectrometry, in an attempt to determine its native molecular weight, and hence infer oligomeric state. BMV2-MBP samples produced polydisperse MS spectra, detecting masses corresponding to monomeric, dimeric, trimeric, tetrameric and hexameric BMV2-MBP assemblies. This is a compelling result. Since it is common for native assemblies to fall apart during vaporization in the MS experiment (this is, in fact, the only major limitation of this valuable technique), the polydispersity need not be interpreted literally. The largest mass detected is a hexamer. Not only is the pentamer not the largest mass detected, it isn't represented among the other putatively 'disassembled' sub-complexes. OCAM was also performed on BMV2-MBP protein, but due to BMV2 instability after proteolysis of the MBP linker region, the gel (not shown) did not reveal its oligomeric state. It is worth noting, however, that native BMV2-MBP protein samples (not digested with TEV) migrated on a blue native gel as something much larger than a hexamer, and not inconsistent with a mass that would correspond to a dimer of pentamers or hexamers. The resolution of Invitrogen Blue Native Ladder, however, does not allow for distinguishing between the two possibilities.

Figure 3

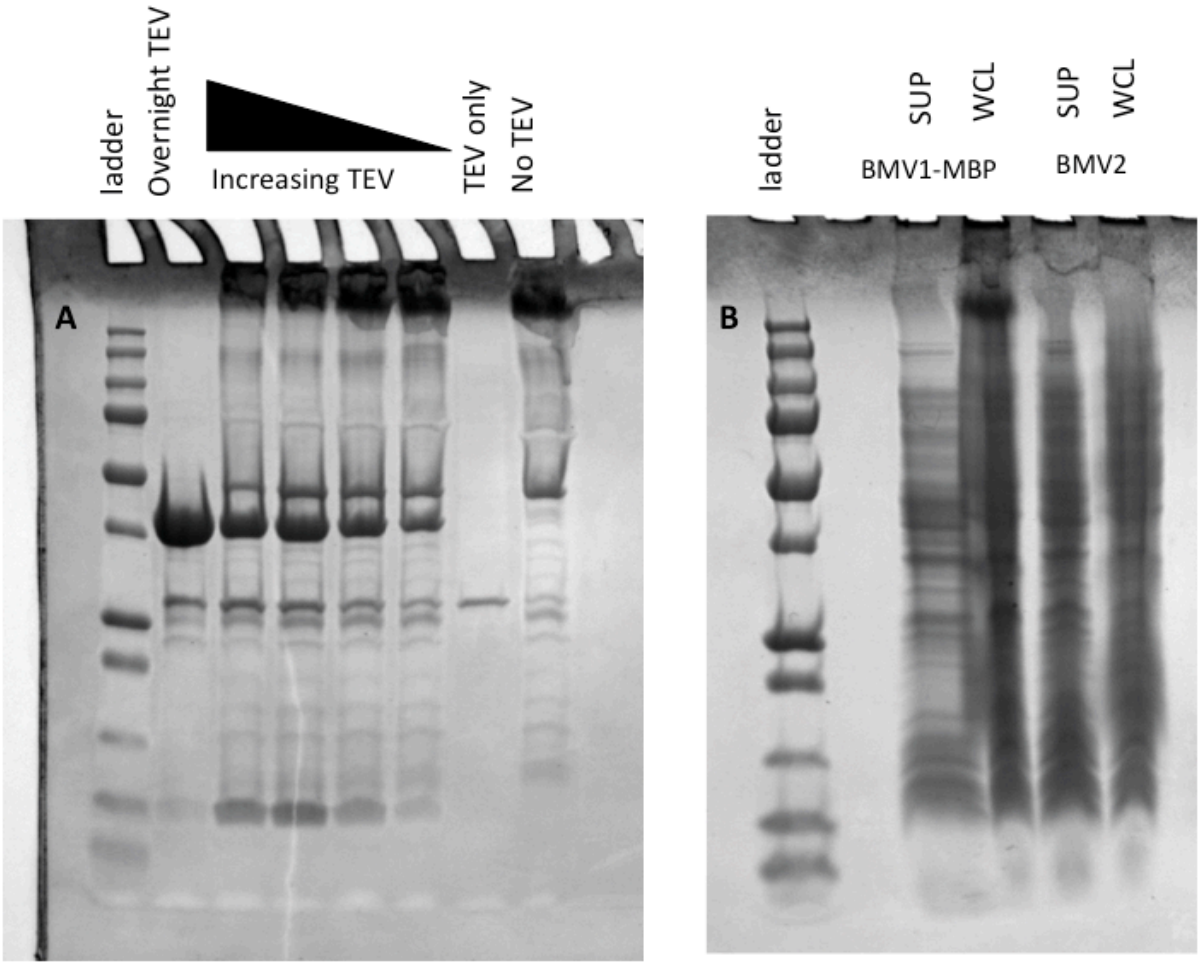
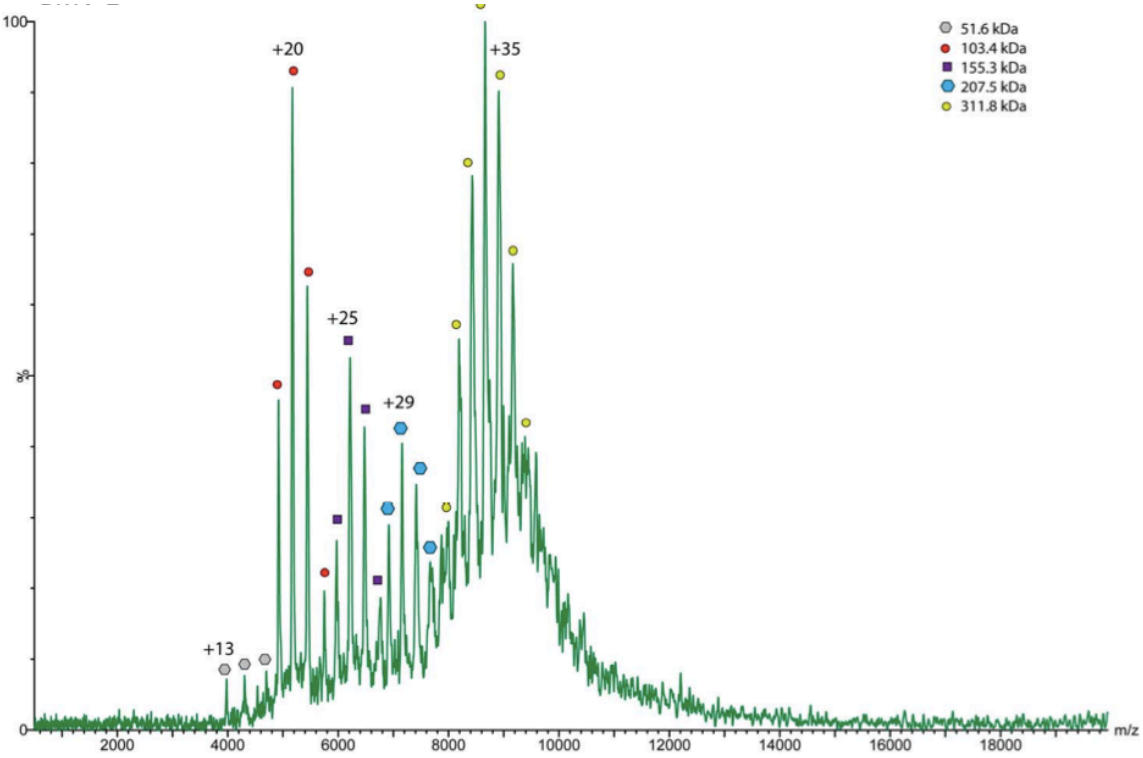


Figure 3. BVM-1 and BMV-2 characterization by SDS-PAGE
3a. OCAM samples were prepared similarly to OCAC samples described in Chapter 1. "Overnight TEV" refers to a sample incubated with the highest amount of TEV protease for over 24 hours, compared to a 1-hour time limit for other samples. In this sample, monomeric BMV2 is degraded. **3b**. SUP refers to supernatant, or soluble fractions of whole cell lysates (WCL) of heterologous over-expression studies.

Figure 4. BMV2-MBP Native Mass Spectrometry reveals monomer, dimer, trimer, tetramer and hexameric assemblies.



The theoretical monomeric mass of BMV2-MBP is 51.4 kDa

In the aftermath of these investigations, I seems that the current models of BMV structure and mechanism are lacking in some way. Why would a microcompartment need two BMV shell proteins? Could it impart three-dimensional asymmetry to an icosahedron microcompartment? Might one of them be a hexamer within the shell wall? Results suggested that, at least in solution while fused to MBP, BMV2 forms hexameric, or possibly dodecameric structures (Fig 4 and data not shown). Why do BMVs so easily convert to hexamers? Recall that *E. coli* ethanolamine microcompartment BMV shell protein, EutN, crystallized twice as a hexamer, but was later determined to be pentameric in solution by OCAC (see Chapter One of this thesis). Perhaps eutN is a

‘strange’ protein, but it is more likely that these BMV anomalies are indicative of their mechanisms of incorporation into microcompartments. Although not the only hypothetical mechanism, BMVs could incorporate into growing shell walls as hexamers, and only under the select conditions, convert to pentamers. Some icosahedral viruses use similar mechanisms of icosahedral vertex formation (***Virus Structure and Assembly***, ISBN 0-12-039863-X).

CHAPTER II : Discovery of alpha-carboxysome Rubisco Assembly Factor, acRAF

ABSTRACT

Carboxysomes are proteinaceous bacterial microcompartments that increase the efficiency of the rate-limiting step in carbon fixation by sequestering reaction substrates. Typically, a-carboxysomes are genetically encoded as a single operon expressing the microcompartment's structural proteins and its encapsulated enzymes. In addition, depending on phylogeny, as many as thirteen other genes are found to co-occur near or within a-carboxysome operons. One of these genes codes for a protein with distant homology to pterin-4a-carbinolamine dehydratase (PCD) enzymes. It is present in all a-carboxysome containing bacteria and has homologs in algae and higher plants. Canonical PCDs play an important role in amino acid hydroxylation, a reaction not associated with carbon fixation. We determined the crystal structure of an a-carboxysome PCD-like protein from the chemoautotrophic bacterium *Thiomonas intermedia* K12, at a resolution of 1.3 Å. The protein retains a three-dimensional fold similar to canonical PCDs, though the prominent active site cleft present in PCD enzymes is disrupted in the a-carboxysome PCD-like protein. Using a cell-based complementation assay, we tested the PCD-like proteins from *T. intermedia* and two additional bacteria, and found no evidence for PCD enzymatic activity. However, we discovered that heterologous co-expression of the PCD-like protein from *H. neapolitanus* with Rubisco and GroELS in *E. coli* increased the amount of soluble,

assembled Rubisco recovered from cell lysates compared to co-expression of Rubisco with GroELS alone. We conclude that this conserved PCD-like protein, renamed here alpha carboxysome Rubisco Assembly Factor (or acRAF), is a novel Rubisco chaperone integral to α-carboxysome function.

Rubisco (Ribulose Bisphosphate Carboxylase/Oxygenase) catalyzes the rate limiting reaction of carbon fixation: the covalent attachment of carbon dioxide to the sugar ribulose-1,5-bisphosphate . Despite its central role in metabolism, Rubisco is a notoriously inefficient enzyme. Its turnover rate varies between species but is in the range of seconds (39, 40). The enzyme also shows relatively poor discrimination between its substrate, carbon dioxide, and molecular oxygen (41, 42). Higher plants have evolved various mechanisms to overcome these inherent catalytic inefficiencies, including strategies that rely on compartmentalization of the steps of carbon fixation across multiple cell types. Many bacteria, on the other hand, rely on a giant proteinaceous structure known as the carboxysome to sequester the cell's carbon fixation reactions. By encapsulating Rubisco and carbonic anhydrase together, concentrated CO_2 can be provided in the vicinity of Rubisco (15, 18, 25, 43, 44). In this way, the carboxysome serves as an essential carbon fixing organelle in many photosynthetic and some chemoautotrophic bacteria.

The major protein constituents of the carboxysome have been well characterized. Bacterial microcompartment shell proteins (BMCs) assemble into hexagonal lattices that form the walls of the carboxysome shell (15, 18). Pentameric bacterial microcompartment vertex proteins (BMVs) enable full closure of these icosahedral structures by occupying the spaces at each of the twelve vertices (18, 23, 1, 45). A

variety of paralogous BMC proteins are typically encoded in carboxysome operons; they are believed to play distinct roles in the shell (18). Carboxysomes are phylogenetically classified as either alpha- or beta-type (46, 47). The two divergent carboxysome lineages are distinguished by differences in Rubisco phylogeny (Rubisco type IA and type IB are found in a- and b-carboxysomes, respectively), gene composition and operon organization. a-carboxysome genes generally appear within a single operon, while b-carboxysome genes typically appear in multiple dispersed gene clusters. Functional differences resulting from this divergence, if any, are not currently understood.

Numerous proteins are implicated in the complex process of Rubisco folding and activation in diverse organisms. In some instances known Rubisco chaperones are encoded near α-carboxysome operons or β-carboxysome Rubisco genes. RbcX, a small (~15 kD) dimeric protein encoded between Rubisco large and small subunits of β-carboxysomes, has been shown to facilitate hexadecameric assembly of Rubisco in the b-carboxysome system (48–50). RbcX homologs are not present in α-carboxysome containing organisms, suggesting that either the subunit assembly function performed by RbcX is unnecessary for carboxysome type 1A Rubisco or that a distinct, unknown mechanism promotes hexadecameric assembly of α-carboxysome Rubisco. CbbX, a Rubisco activase, modulates Rubisco activity by dislodging inhibitory ribulose-1,5-bisphosphate molecules from the active site (51). Other ATP-dependent chaperones CbbO and CbbQ regulate and enhance Rubisco catalysis by altering Rubisco conformational states in a variety of carbon-fixing prokaryotes (52). CbbO, CbbQ and CbbX homologs are sometimes present in α-carboxysome operons (Fig 1). GroELS

(also referred to as chaperonin) is a ubiquitous ATP-dependent chaperone complex essential for the proper folding of Rubisco large subunits (53, 54). Rubisco assembly factor 1, RAF1, was recently identified as a novel Rubisco chaperone in *Z. maize* (55). RAF1 is conserved in green plant lineages, but is not present in carbon-fixing prokaryotes. To date, plant Rubisco has not been successfully assembled *in vitro* or in a heterologous *E. coli* host, hinting at the probable existence of yet unidentified folding or assembly factors.

Figure 5

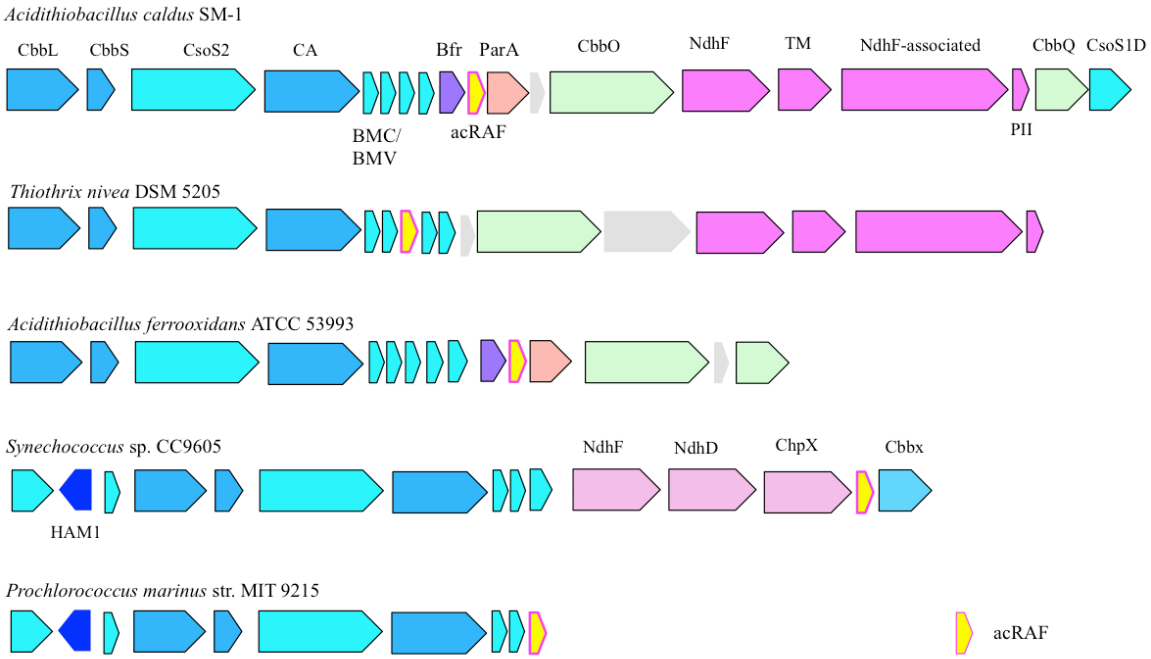


Fig 5. acRAF is a defining member of a-carboxysome operons.
Carboxysome operons of *Acidithiobacillus caldus* SM-1, *Thiothrix nivea* DSM 5205, *Acidithiobacillus ferrooxidans* ATCC 53993 and *Synechococcus sp*. CC9605 and *Prochlorococcus marinus* str. MIT 9215 operons are shown to demonstrate the extent of acRAF association with α-carboxysome operons. acRAFs are present within all α-carboxysome operons examined. CbbL: Rubisco large subunit. CbbS: Rubisco small subunit. CsoS2: shell protein of unknown function. CA: carbonic anhydrase. BMC: bacterial microcompartment shell protein. BMV: bacterial microcompartment vertex shell protein. Bfr: bacterioferritin family. acRAF: alpha-carboxysome Rubisco assembly factor. ParA: Partitioning A family (example YP_003262802). CbbO: CbbQ activase. NdhF (NuoL/Nqo12-like): Complex I NADH oxidoreductase chain F family protein (example *YP_003262799*). TM: transmembrane protein. (example *YP_003262798*). NdhF-associated: ORF commonly encoded adjacent to or near NdhF (example *YP_003262797*). PII: PII nitrogen regulatory family (example *YP_003262796*). CbbQ: ATP-dependent Rubisco activase. CsoS1D: Double domain BMC shell protein. HAM1: Ham1 domain containing. NdhF: Complex I NADH dehydrogenase oxidoreductase subunit F family. NdhD: Complex I NADH dehydrogenase oxidoreductase M family. ChpX: CO2 hydration protein. CbbX: Rubisco activase.

Recent bioinformatics studies have identified several genes of unknown function that consistently occur near a-carboxysome BMC genes (14, 56). This conserved genetic

association strongly suggests that the encoded proteins are involved in the catalytic function, regulation or structure of the a-carboxysome. One of these proteins is remotely related in sequence (often with BLAST E values ≥0.05) to pterin-4a-carbinolamine dehydratase (PCD) enzymes, and is represented in all a-carboxysome operons we examined. It appears to be a defining member of this carboxysome subtype. Figure 1 illustrates the organization of a number of diverse a-carboxysome operons and highlights the presence of this remote homolog of PCD.

PCD enzymes are ubiquitous throughout the three kingdoms of life, where they act in concert with aromatic amino acid hydroxylases (AAH) in amino acid metabolism pathways (57, 58). AAHs use tetrahydrobiopterin (also known as BH4 and THB) cofactors and molecular oxygen to catalyze the addition of a hydroxyl group onto aromatic amino acids. A specific example of an AAH is bacterial phenylalanine hydroxylase, which adds a hydroxyl group to phenylalanine to produce tyrosine (59). In this reaction, the tetrahydrobiopterin cofactor is simultaneously hydroxylated on the 4a carbon, generating petrin-4a-carbinolamine. Two enzymes are required to regenerate pterin-4a-carbinolamine to tetrahydrobiopterin: DHPR (dihydropterin reductase) and PCD (58). Functional PCDs exist as homodimers in prokaryotes. In animals, DCoH (dimerization cofactor of HNF-1 or hepatocyte nuclear factor-1) is a tetrameric bifunctional enzyme that serves both as a PCD and as a transcriptional activator of HNF-1 (60). Despite its dual role in animals, DCoHs are similar in sequence and tertiary structure to bacterial PCDs (61). In prokaryotes lacking AAHs, 'orphan' PCDs are sometimes present and are suspected to participate in other unknown reactions or cellular processes (58). Catalytically inactive PCD-like proteins of unknown function are

present both in chloroplasts (58) and, as shown here, in many bacteria, where they occur in close association with a-carboxysome operons. No structures have been reported to date for these PCD-like proteins. Figure 6*A* summarizes current information on PCDs and PCD-like proteins.

Here we report our work investigating the structure and function of the conserved PCD-like family of proteins found in a-carboxysome containing bacteria. We show that the protein is an enzymatically inactive relic of PCD, and that its true cellular function relates to the folding or assembly of hexadecameric Rubisco. Based on our results, we rename this protein the a-carboxysome Rubisco Assembly Factor, or acRAF.

EXPERIMENTAL PROCEDURES

Cloning, expression, and purification of T. intermedia acRAF

The construct used for crystallization of codon-optimized acRAF (with amino acids "MHHHHHH" replacing the first 5 amino acids of the protein coding sequence) was ordered from Biomatik Corp. gene synthesis company. Codon optimization was performed by DNAworks (www.DNAworks.com). The resulting gene sequence was cloned into pET22b(+) vector between the 5' NdeI and 3' EcoRI sites. acRAF was over-expressed in *E. coli* BL21(DE3). One liter flasks were inoculated with 15 mL overnight culture, and grown under ampicillin selection in LB media at 37˚C, shaking at 240 rpm, until cells reached an optical density absorbance of 0.8 at 600 nm (OD_{600} = 0.8). Isopropyl b-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1.0 mM. After 3-5 hours, cells were spun down and frozen at -20˚C until further use. Frozen cell pellets collected from two liters of culture were resuspended in 50 mL of buffer (50

mM Tris-HCl pH 7.6, 300 mM NaCl, Roche complete EDTA-free protease inhibitor cocktail) and lysed by sonication. Cell debris was pelleted by centrifugation at 16,500 rpm in SS-34 rotor for 30 minutes at 4°C. The resulting supernatant was filtered through a 0.22 μm Whatman filter. Filtered supernatant was then loaded onto a HisTrap 5 mL column by syringe, washed with 30 mL wash-buffer (50 mM Tris-HCl pH 7.6, 300 mM NaCl, 50 mM imidazole), and eluted in one step with elution buffer (50 mM Tris-HCl pH 7.6, 300 mM NaCl, 400 mM imidazole). Purified acRAF was dialyzed into 10 mM Tris-HCl pH 7.6, 50 mM NaCl, 5% glycerol. acRAF was concentrated to 11 mg/mL, as calculated from absorbance at 280 nm using the theoretical extinction coefficient.

acRAF crystallization

Hanging-drop crystallization screens were performed using the nanoliter liquid handling Mosquito from TTP LabTech at the High Throughput Macromolecular Crystallization Facility at UCLA. Optimization of needle-like crystals obtained in JCSF+ Suit crystallization screen condition A8 were pursued in 24-well (500 μL well) trays. Crystals were collected from 2 μL hanging drops of 1:1 ratio protein:well-solution after 16 days of slow but continuous growth at room temperature. The optimized well solution contained 17.5% PEG 3500 and 800 mM ammonium formate. A cryoprotectant was prepared containing 17.5% PEG 3500, 800 mM ammonium formate, and 25% ethylene glycol, with or without 200 mM NH_4I_2 for subsequent phasing by anomalous scattering. Crystals were soaked in cyroprotectant for 1-2 minutes before freezing and storage in liquid nitrogen.

Phasing and Refinement

Diffraction data extending to 1.3 Å resolution were collected at the Argonne National Laboratory, Advanced Photon Source (APS), beamline 24-ID-C (Table 1). Phases were obtained from single wavelength data using isomorphous replacement with anomalous scattering from iodide. Five iodide-bound sites were identified by anomalous scattering, and all five sites were used for phasing. Iodide sites were located and phases calculated using the program HKL2MAP (62). The program Buccaneer (63) was used to automatically trace the main chain. Subsequent rounds of model building were performed with COOT (36) and the structure was refined using BUSTER (64) to a final Rwork and Rfree of 0.203 and 0.204, respectively. 98% of the backbone dihedral angles are within the favored regions of a Ramachandran diagram. Coordinates and structure factors have been deposited with the PDB ID code 4LOW.

PCD Complementation Assay

All PCD constructs were cloned between the 5' NdeI and 3' KpnI restriction enzyme sites in a pBSKII(-) vector. The PCD construct from *P. aeruginosa* (NP_249562) in the vector pBSKII(-) and *P. aeruginosa* phenylalanine hydroxylase (NP_249563) in the vector pACYC177 were gifts from Dr. Andrew Hanson. A cDNA template of the *A. thaliana* chloroplast acRAF gene (NP_199924) was provided by Dr. Sabeeha Merchant. Codon optimized *T. intermedia* K12 acRAF (YP_003641868) DNA was purchased through Biomatik Corp. The codon-optimized *T. intermedia* K12 PCD (YP_003642661) gene was synthesized by assembly PCR from oligonucleotides designed by DNAworks and purchased from IDT. Codon-optimized acRAF genes from *H. neapolitanus* (YP_003262803) and *P. marinus* (NP_874951) were purchased as IDT gBlocks. All PCDs and acRAFs were transferred into clonal descendants of Hanson's original *P.*

aeruginosa PCD pBSKII(-) construct using Gibson assembly cloning methods, thereby replacing the original insert. Codon optimized *T. intermedia* K12 phenylalanine hydroxylase (YP_003642727) (purchased from Biomatik) was transferred into that PCR-linearized pACYC177 using Gibson assembly (34). The starting ATG of pACYC177 ampicillin resistance open reading frame became the beginning ATG of *T. intermedia* K12 phenylalanine hydroxylase, and largely replaced the ampicillin resistance gene. A pBSKII(-) empty vector control and pACYC177 empty vector controls were selected from failed cloning experiments. All constructs were sequenced by Laragen, Co. Sequences are available upon request.

Cloning of Rubisco and acRAF

Several experiments were performed to determine the effects of acRAF and/or GroELS co-expression on the solubility and assembly of Rubisco subunits. Ten different *E. coli* BL21(DE3) strains were created by triple or double electroporation into Lucigen Crop.'s "*E. cloni* EXPRESS BL21 (DE3) Electrocompetent cells". CbbLS, CbbL and CbbS were PCR-amplified from *H. neapolitanus* genomic DNA and cloned into pRSF-duet vector between 5' NcoI and 3' XhoI restriction enzyme sites using Gibson assembly. acRAF, also PCR-amplified from *H. neapolitanus* genomic DNA, was cloned into pET22b(+) between 5' NdeI and 3' EcoRI restriction enzyme sites. Plasmid pBB541, a GroELS expressing vector with a p15a origin of replication, was purchased from Addgene. Empty pRSF and pET22b(+) plasmids were used as negative controls where appropriate. Exact sequences are available upon request.

Heterologous Rubisco and acRAF co-expression assay

Transformed BL21(DE3) cells were grown overnight in LB broth supplemented with the following antibiotics as appropriate: 15 μM kanamycin, 50 μM ampicillin, 12.5 μM streptomycin. 2 mL overnight cultures were added to 250 mL flasks containing 100 mL LB. Cells were grown at 37˚C until the OD_{600} reached 0.8. Cells were induced with 1 mM ITPG, shaking at 240 rpm, for 3 hours at 37˚C. Cells were spun down and resuspended in the following buffer: 50 mM Tris-HCl pH 7.6, 300 mM NaCl, 10 mM $MgCl_2$, 10 mM $NaHCO_3$, 10 mM DTT. Buffer volume was adjusted so that all cell cultures achieved an OD_{600} of 54. Cells were lysed by sonication. Insoluble cell debris was pelleted in 1.6 mL eppendorf tubes spinning at 20,000 x *g* in a tabletop microcentrifuge. Supernatant was then loaded onto BioRAD MiniProtean TGX gels (Cat# 456-9036) and analyzed by SDS-PAGE (data not shown) and native gels (Fig 4).

Associated gene analysis

Alpha-carboxysome operons were detected by searching the non-redundant protein sequence database for sequences homologous to the a-carboxysome protein CsoS2, using the program Blast. Twenty-one a-carboxysome operons were selected with the aim of maximizing diversity of both genomic backgrounds and a-carboxysome components. Operons from the following organisms were analyzed by hand: *Nitrobacter winogradsky*i NB-255, *Thiomonas intermedia* K12, *Nitrosomonas eutropha* C91, *Thiobacillus denitrificans* ATCC 25259, *Nitrococcus mobilis* NB-231, *Halothiobacillus neapolitanus*, *Thioalkalivibrio sulfidophilus* HL-EbGr7 , *Thiothrix nivea* DSM 5205, *Acidithiobacillus caldus* SM-1, *Acidithiobacillus thiooxidans* ATCC 19377, *Acidithiobacillus ferrooxidans* ATCC 53993, *Synechococcus* sp. CC9311, *Synechococcus* sp. CC9902, *Synechococcus* sp. CC9605, *Cyanobium* sp. PCC 7001,

Prochlorococcus marinus MED4., *Prochlorococcus marinus* str. MIT 9215, *Prochlorococcus marinus* str. MIT 9202, *Prochlorococcus marinus* str. NATL2A, *Prochlorococcus marinus* str. MIT 9211.

In order to develop a better picture of the various uncharacterized proteins that appear to be associated with a-carboxysomes, we have used the comparative genomics program SEED to systematize the known a-carboxysome operons and their associated genes. Those results are publicly accessible (http://pubseed.theseed.org - SEED curator: nwheatley; Subset name: alpha carboxysome).

Data deposition - Atomic coordinates and diffraction data for acRAF have been deposited in the PDB with ID 4LOW.

RESULTS

Structure of T. intermedia acRAF

We determined the structure of *T. intermedia* acRAF (GI: 296134626, YP_003641868) to 1.3 Å resolution by X-ray crystallography (Fig 6B and 6C). An initial acRAF construct with a C-terminal hexa-histidine tag did not crystallize despite many attempts. Secondary structure prediction programs indicated that the first six amino acids of native acRAF were unstructured. We therefore moved the polyhistidine tag to the N-terminus. We chose to replace the presumptively disordered N terminus with the polyhistidine tag rather than to extend the disordered region, and this construct crystallized. The structure was determined by anomalous scattering methods (Table 2). The asymmetric unit of the $P4_1$ crystal contained two protein subunits, or one dimer. A

final refined model was obtained in which a total of 148 of the 164 residues of the combined subunits of the acRAF dimer could be reliably modeled (Fig 6B). The electron density for residues 66-70, corresponding to the region of catalytic activity in PCD enzymes, was not interpretable, indicating either multiple conformations and/or high structural disorder in this region of acRAF. This segment was therefore not included in the final deposited structure, though the loop was included in a plausible conformation for the purposes of structure visualization in some cases. In the crystal structure, the histidine residues that were inserted at the N terminus emerge from an ordered beta strand, implying that the deleted N-terminal residues of acRAF might have contributed to a beta sheet, contrary to our predictions. Further analysis revealed that these ordered histidine residues are involved in a chelation-dependent crystal contact with a divalent cation, thereby creating a fortuitous interaction important for crystal formation.

TABLE 2. acRAF X-ray data collection and model refinement statistics

Statistics	Value
Wavelength (Å)	0.9791
Resolution range (Å)	51.1 - 1.30 (1.34-1.30)
Space group	$P4_1$
Unit cell (Å.°)	a= b=51.1, c=60.9; α=β=γ=90
Total reflections recorded	253115 (16941)
Unique reflections	37451 (2610)
Multiplicity	6.7 (6.5)
Completeness (%)	96.9 (92.6)
Mean I/sigma(I)	17.2 (2.05)
Wilson B-factor ($Å^2$)	23.1
CC(1/2)	99.9% (69.4%)
Model R-work	0.203 (0.211)
Model R-free	0.204 (0.241)
Number of chains	2
Number of modeled residues (chain A/ chain B)	83/78
Number of non-hydrogen atoms	1452
Macromolecules	1305
Formate	18
Water	127
Ni^{+}*	2

Protein residues	336
Geometric deviations (rms)	
bonds (Å)	0.001
angles (°)	1.03
Ramachandran statistics (%)	
Favored	97.8
Additionally allowed	2.2
Generously allowed	0
Disallowed	0
Average B-factor ($Å^2$)	
protein main chain	29.5
protein side chains	33.6
solvent (water)	39.1

Statistics for the highest resolution shell are shown in parentheses. *The divalent metal cation was defined as Ni^+during the refinement process, based on the possibility of contamination from nickel-bound chromatography columns.

Figure 6

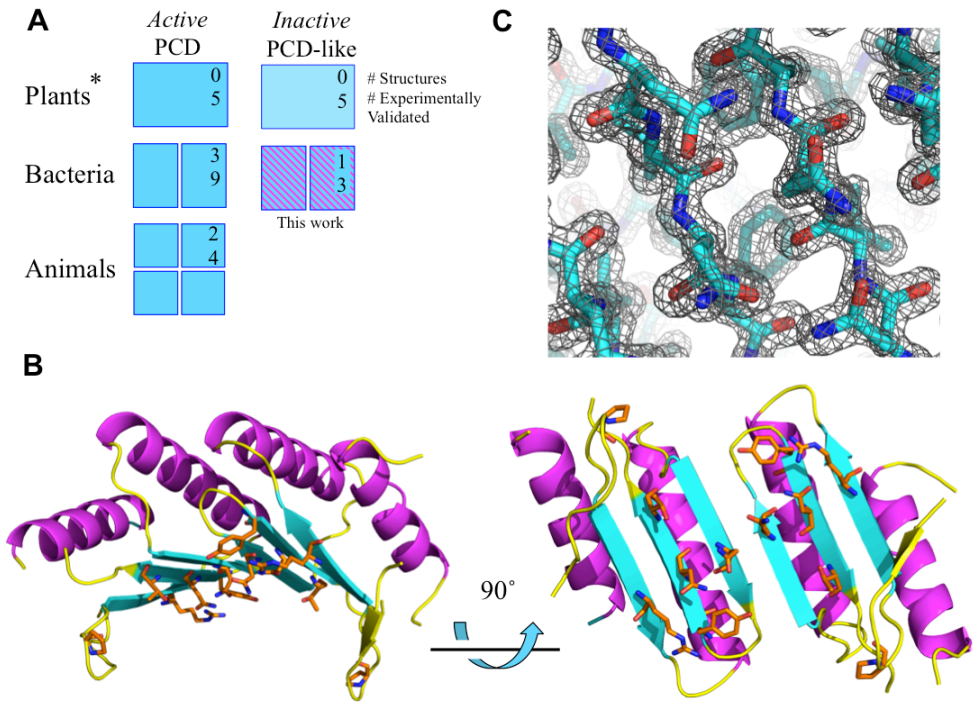


Figure 6. Crystal structure and conserved surface regions in the acRAF dimer
A) Diagram summarizing current knowledge of PCDs and inactive PCD-like proteins. PCDs are capable of pterin-4 a-carbinolamine dehydratase activity. PCD-like proteins, while structurally similar to PCDs, do not show evidence of pterin-4a-carbinolamine dehydratase activity. Oligomeric states are represented by number of subunits shown, except for plant PCD and PCD-like proteins (denoted by *), whose oligomeric states are unknown. Numbers located in upper right hand corners represent the number of representative proteins with crystal structures deposited in the PDB. Numbers below indicate the number of representative proteins examined experimentally, to our knowledge. Information on the bacterial carboxysomal acRAFs (hatched) refers to the present study. **B)** Solvent-exposed, conserved residues of carboxysome acRAF are shown as orange sticks. **C)** σ_A-weighted 2mFo-DFc electron density map of the dimer interface viewed in the same orientation as the bottom right model in panel B.

Comparisons between acRAF and true PCD enzymes show significant structural similarities despite low sequence identity. Generally, about 75% of a typical acRAF sequence can be aligned to PCD. Within that region of similarity, 25-35% of residues are identically conserved. Structural superposition of acRAF onto *Toxoplasma gondii* PCD (PDB 2V6T) resulted in a backbone RMS deviation of only 1.5 Å (Fig 7). The main

difference between the PCD and acRAF structures is the N-terminal helix. The PCD enzyme from *T. gondii* contains an 11-residue N-terminal helix that is lacking in acRAF. The tertiary structures of PCD and acRAF are similar to the ferredoxin superfamily, but with a minor circular permutation near the chain termini. We note here the intriguing observation that a number of other carboxysome-related proteins also show a similar fold. This includes the main BMC shell proteins, the N-terminal domain of the Rubisco large subunit, and a PII nitrogen regulatory-like protein encoded in several proteobacteria a-carboxysome operons (65) (Fig 5). The oligomeric state of acRAF in the crystal structure is a dimer, which like true PCDs is composed of two subunits arranged as a symmetric dimer. The dimeric interface is formed by a beta strand (residues 50-55) and an alpha helix (residues 28-43). The strand pairs with the cognate strand from the other subunit to form an anti-parallel beta sheet.

A presumptive catalytic motif for PCD enzymatic activity can be inferred by analyzing the arrangement and sequence conservation in representative PCDs known to be active compared to those proven earlier to be inactive. The PCD catalytic motif, [EDKH]-x(3)-H-[HN]-[PCS]-x(5,6)-[YWF]-x(9)-[HW]-x(8,15)-D, (58) is not conserved in acRAF; only one histidine residue (residue 48) remains in a potentially catalytic position in the structure determined here. Consistent with this divergence, the structure of acRAF showed that its active site region (vis-à-vis PCD) is dramatically altered (Fig 7A and 7C). The narrow cleft that binds the pterin cofactor in PCD is substantially opened up in acRAF.

Figure 7

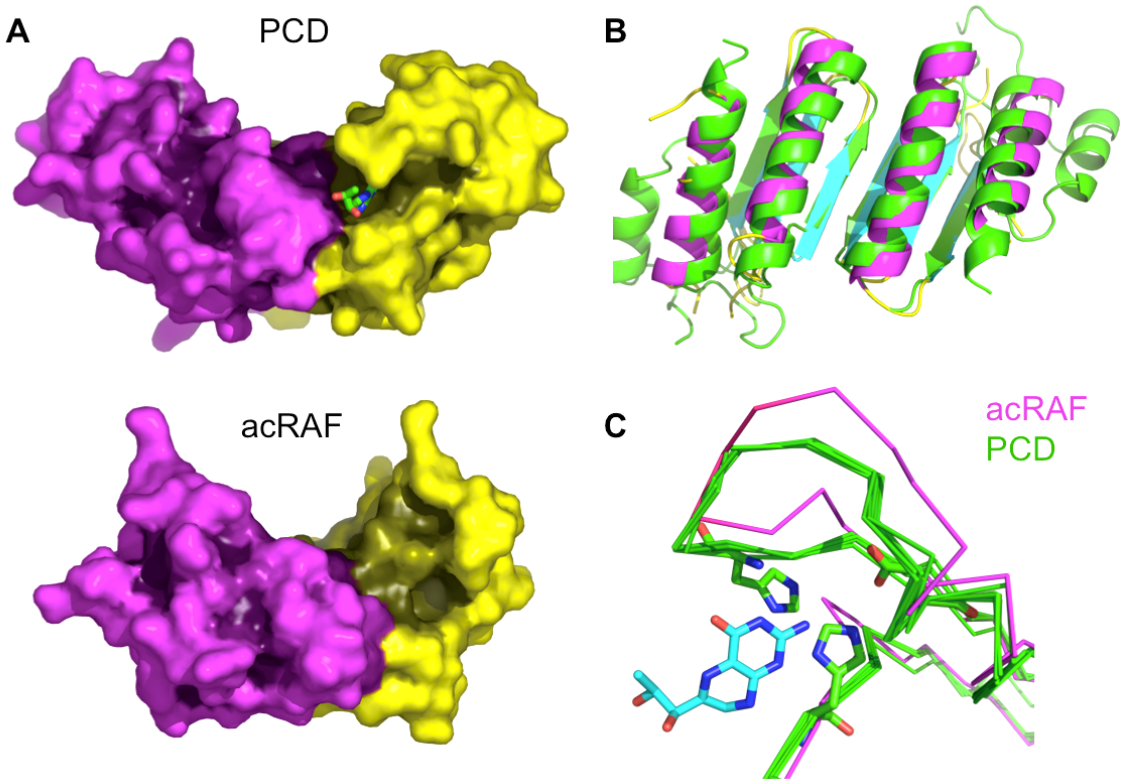


Figure 7. Structural comparison of acRAF and PCD
A) Comparison of the surface shapes and binding clefts of a true pterin-4a-carbinolamine dehydratase (PCD) with acRAF from the a-carboxysome operon. A bound dihydrobiopterin is shown in the active site cleft of PCD (PDB ID 2V6T). The two proteins are shown in the same orientation. The active site region of PCD is dramatically altered and much less pronounced in acRAF. Surfaces are shaded according to depth based on a calculation of diffusion accessibility (66) (http://nihserver.mbi.ucla.edu/diff_acc/). Three partially disordered residues of a loop were included in a plausible configuration in order to convey a more realistic view of the shape of acRAF. **B)** Structural alignment of acRAF to PCD from *Toxoplasma gondii* (PDB 2V6T). *T. gondii* PCD is colored green. **C)** Structural alignments of active sites among five PCDs shown in green (PDB IDs: 1DC0, 2EBB, 2V6T, 3JST, 4C45). acRAF is shown in magenta. The pterin cofactor (from model 2V6T) is shown in blue. Residues required for PCD catalysis that are present in all PCDs, but not in acRAF, are shown as sticks. The structural alignment is tilted roughly 45˚ upwards into the plane of the page compared to the orientation of PCD and acRAF in panel A.

Figure 8

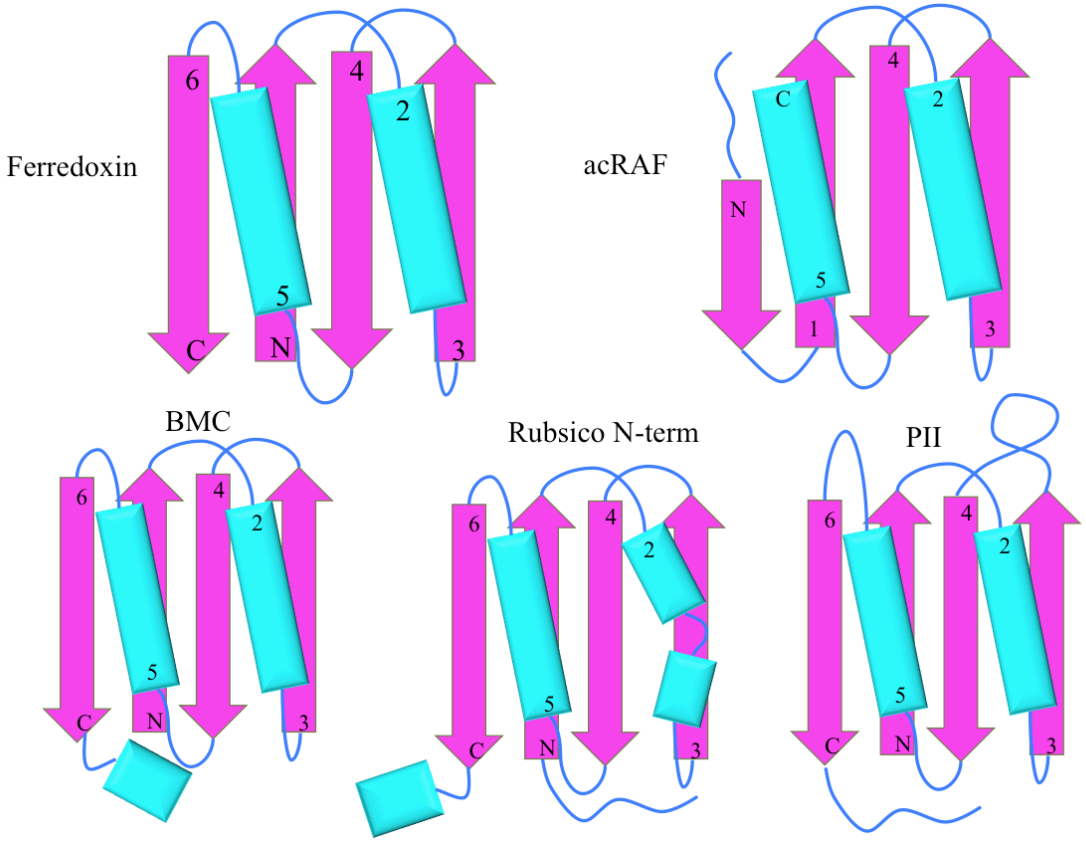


Figure 8. Ferredoxin-like protein folds associated with carboxysomes
Simplified models of secondary and tertiary structural motifs of select carboxysome-associated proteins. Alpha helices are colored cyan and beta sheets are colored magenta. Top left: Ferredoxin. Top Right: PCD/acRAF. Bottom Left: BMC (bacterial microcompartments shell proteins). Bottom Center: N-terminal domain of Rubisco large subunit. Bottom Right: nitrogen regulatory PII family protein (encoded near several chemoautotrophic a-carboxysome operons).

Despite structural divergence in the active site, we deemed it necessary to experimentally test acRAFs for PCD activity for two reasons. First, the catalytic mechanism of PCD is not well understood, making it difficult to rule out enzymatic activity of acRAF from the structure alone. Second, previous studies demonstrating a

lack of catalytic activity in PCD-like proteins from plant chloroplasts (also referred to as type 2 PCDs) left open questions about the reason for this inactivity. The plant PCD-like proteins previously studied show a lack of conservation in the active site region, but they also bear a 40-60 amino acid extension at their N-termini (not including the predicted chloroplast targeting sequence), which could, in principle, block or regulate the active site. No such extension is present in the acRAF proteins from a-carboxysome operons.

Sequence alignments among seventy-three carboxysome associated acRAFs were performed using the program PRALINE (67) to investigate potentially conserved motifs specific to this family. Pro47 and Phe52 are the only fully conserved residues; both likely fulfill structural roles required for overall acRAF stability. Several other solvent accessible residues, however, also show substantial conservation (Fig 6B). Interestingly, these conserved residues localize to a contiguous surface region of acRAF, distant from the vestigial active site, implicating this surface region as a putative functional interface.

Absence of PCD enzyme activity

To test acRAF for PCD enzyme activity, we employed a cell-based complementation assay that requires PCD activity for cell growth (18). The complementation assay was performed in *E. coli* strain JP2255, a tyrosine and phenylalanine auxotroph that can replicate on media lacking tyrosine if supplemented with both a functional phenylalanine hydroxylase and a functional pterin-4α-carbinolamine dehydratase (PCD). *E. coli* species naturally lack AAHs and PCDs but do have genes encoding DHPRs, which are required in combination with PCD dehydration to regenerate the pterin cofactor. We

used *E. coli* strain JP2255 to confirm the catalytic inactivity of three different acRAFs: those from *T. intermedia* (the same organism as our crystal structure), *Halothiobacillus neapolitanus* C2, and *Prochlorococcus marinus* CCMP1375. As a negative control, we tested the type 2 PCD from *A. thaliana* and its N-terminal truncation mutant. Two positive controls were used: a well-characterized PCD from *Pseudomonas aeruginosa* POA1 and a predicted canonical PCD from *T. intermedia*, a paralog to the acRAF protein structurally characterized here. *T. intermedia* also contains a presumptive phenylalanine hydroxylase. We used this phenylalanine hydroxylase in our complementation assay along with another well-characterized phenylalanine hydroxylase from *Pseudomonas aeruginosa* PAO1.

Figure 9 shows representative results of the PCD complementation experiments. In combination with either of the phenylalanine hydroxylases, none of the acRAFs we tested restored cell growth on media lacking tyrosine. In contrast, both the PCD from *P. aeruginosa* and the non-carboxysomal PCD from *T. intermedia* supported cell growth when co-transformed with plasmids encoding either of the phenylalanine hydroxylases. These results are consistent with the suggestion from structural and sequence analysis that acRAFs are not pterin dehydratases. These results also show that *T. intermedia* encodes enzymes capable of phenylalanine hydroxylation independent of acRAF, further supporting the hypothesis that acRAF fulfills a novel, non-enzymatic role in the function and/or structure of α-carboxysomes.

Figure 9

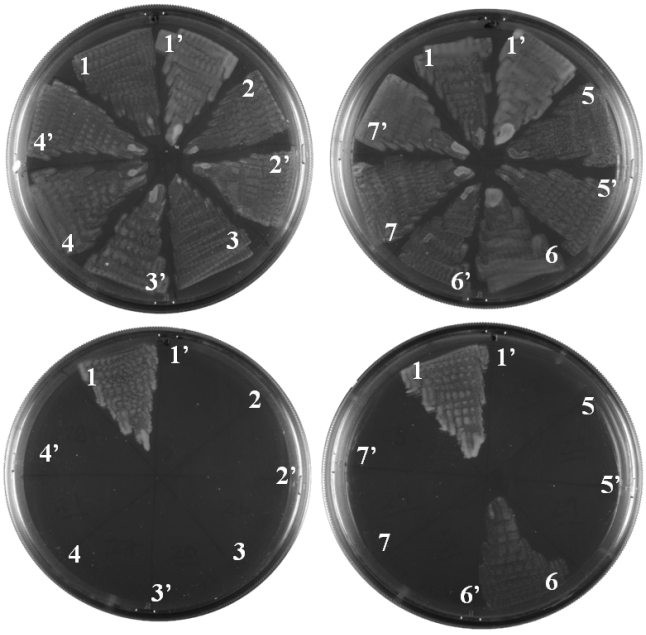


Figure 9. acRAFs do not compliment PCD activity *in vivo*

An *E. coli* tyrosine auxotroph, strain JP2255, was doubly transformed with two compatible plasmids, one plasmid containing the phenylalanine hydroxylase from *T. intermedia* and the second plasmid containing a putative or experimentally verified PCD. Cells were streaked in wedges on minimal media agar plates supplemented with or without tyrosine (top and bottom rows, respectively). Putative PCDs are numbered as follows: 1, *P. aeruginosa* PCD (NP_249562). 2, type-2 PCD from *A. thaliana* missing an N-terminal 46 residue chloroplast-targeting sequence (NP_199924). 3, type-2 PCD from *A. thaliana* missing 101 residues of N terminus. 4, *T. intermedia* acRAF with a 5 residue N-terminal truncation (YP_003641868). 5, *H. neapolitanus* acRAF (YP_003262803). 6, *T. intermedia* PCD (YP_003642661). 7, *P. marinus* acRAF (NP_874951). The prime symbol, as in 1', indicates empty plasmid transformed in place of a putative PCD. None of the PCDs or acRAFs were able to support tyrosine-independent growth without co-transformation with a phenylalanine hydroxylase (data not shown).

The absence of enzyme activity prompted us to search for an alternative function for this apparent enzyme relic. Clues regarding acRAF function arose in comparing gene content between a- and b-carboxysomes. The two types of carboxysomes are functionally similar in their abilities to concentrate and fix CO_2. They are morphologically similar, and their shells are assembled mainly from homologous proteins that follow similar architectural principles. Yet those similarities belie key differences in genomic

organization and protein composition. The presence of mutually exclusive genes in the two carboxysome systems suggests that some functions might be satisfied by functionally analogous but evolutionarily distinct proteins. The use of highly divergent, or even unrelated, carbonic anhydrases in the two systems is a case in point (68). We wondered if acRAF might play a role in Rubisco folding and assembly similar to that played by RbcX in the b-carboxysome.

Heterologous Rubisco Assembly Assay

RbcX is a Rubisco chaperone understood to function in b-carboxysomes by stabilizing Rubisco L_2 dimers before they assemble into L_8S_8 hexadecamers (48, 50). Noting that a-carboxysome-containing organisms do not contain RbcX proteins but always encode acRAF, we tested the effects of co-expressing *H. neapolitanus* acRAF with *H. neapolitanus* Rubisco. In a first set of experiments, we found that co-expressing acRAF with Rubisco in *E. coli* did not improve the yield of soluble native Rubisco. However, Rubisco folding and assembly is a complex process: GroELS is important in subunit folding, but in b-carboxysome-containing cyanobacteria RbcX is required for assembly of the native (hexadecameric) complex. Indeed, in a second set of experiments we found that when acRAF was expressed together with GroELS, the yield of natively assembled Rubisco, as judged by native PAGE, increased dramatically compared to when GroELS was expressed with Rubisco without acRAF (Fig 10). This result identifies the PCD-like family of proteins as a new Rubisco chaperone.

Figure 10

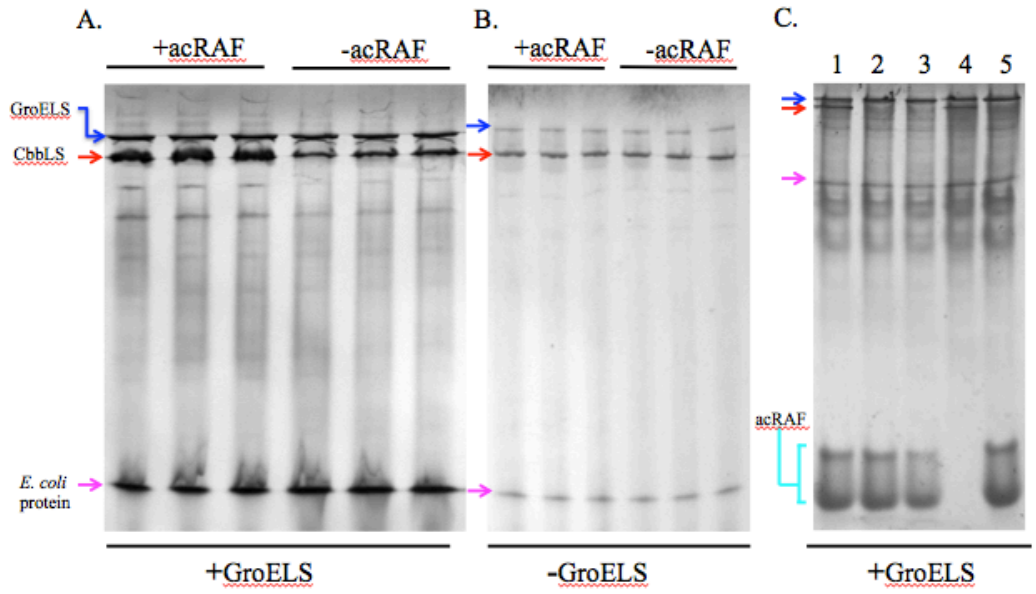


Fig 10. acRAF increases the quantity of natively assembled Rubisco when heterologously co-expressed with GroELS in *E. coli*
(A) CbbLS was co-over-expressed with GroELS in *E. coli* BL21(DE3) either with (+) or without (-) acRAF. Soluble, whole cell lysates were analyzed by native non-denaturing PAGE. Independent co-expression experiments were performed in triplicate. The amount of natively assembled Rubisco is substantially increased by the presence of acRAF (when GroELS is also overexpressed). The band corresponding to assembled CbbLS was deduced by comparisons to control gels run on lysates from cells expressing or lacking specific proteins (panel C). **(B)** CbbLS was over-expressed in *E. coli* BL21(DE3) either with (+) or without (-) acRAF, but without GroELS overexpression. Soluble, whole cell lysates were analyzed by native non-denaturing PAGE. Independent co-expression experiments were performed in triplicate. No effect of acRAF on Rubisco production is seen. **(C)** CbbLS and acRAF bands were identified from soluble fractions of lysates of *E. coli* BL21 (DE3) cells, which were transformed with different combinations of compatible plasmids. Strain 1: CbbLS, acRAF and GroELS. Strain 2: CbbL, acRAF, GroELS. Strain 3: CbbS, acRAF, GroELS. Strain 4: CbbLS, empty pET22b, GroELS. Strain 5: empty pRSF, acRAF, GroELS. Rubisco (hexadecameric L_8S_8) and acRAF band identities were deduced by comparisons of strain-specific band patterns. The acRAF from *H. neapolitanus* migrated as two species, in contrast to acRAF from *T. intermedia*, which migrates as one species (not shown). Note that only strains 1 and 4 are expressing both large and small Rubisco subunits, consistent with the assignment of native Rubisco to the band present exclusively in those lanes. Blue, red and magenta arrows indicated bands that correspond to GroELS, CbbLS and an unidentified *E. coli* protein respectively.

DISCUSSION

Previous studies have noted the occurrence of genes for PCD-like proteins – proteins related to the pterin dehydratase enzyme family – in a-carboxysome operons from both cyanobacteria (56) and chemoautotrophs (14). A systematic examination suggests that these proteins (now renamed acRAFs) are universally conserved in a-carboxysome operons, but absent from genomes that encode b-carboxysomes, making acRAF a defining gene of a-carboxysome operons. By determining the three-dimensional structure of acRAF from *Thiomonas intermedia* we showed that, despite overall structural similarity to bacterial PCDs, the active site cleft of acRAF is drastically altered. Conserved residues believed to be catalytically important in PCDs are not replaced by potentially catalytic residues in the acRAF structure. An *in vivo* enzyme complementation assay in *E. coli* confirmed that carboxysome-related acRAFs lack PCD enzymatic activity. After confirming its catalytic inactivity, we were able to establish a role for this protein in Rubisco assembly.

When heterologously co-expressed in combination with GroELS, acRAF enhances the production of assembled Rubisco. The mechanism by which acRAF interacts with and supports Rubisco expression, stability and/or assembly is not understood yet. All that can be asserted based on these first experiments is that acRAF likely acts at the subunit assembly stage (in a fashion analogous to RbcX), since its effect is only manifested in the presence of GroELS, which acts generally to promote subunit folding. Further mechanistic details await additional studies.

The evolution of enzymes for alternate functions is well known, as are examples of pseudoenzymes or “dead enzymes” wherein the loss of catalytic activity evolves in

concert with the acquisition of regulatory or structural functions. Lens crystallins are classic examples of pseudoenzymes (69). Genomic and biochemical studies have focused attention on the adaptation of enzymes for other purposes as a widespread evolutionary phenomenon ((70, 71), reviewed in refs. (72) and (73)).

Our findings here regarding a cellular function for a-carboxysomal acRAFs align with previous studies of Type 2 PCDs from plant chloroplasts. Proteins from that group have been shown to lack catalytic activity (58), and recent experiments (Alice Barkan and David Stern, personal communication) show that PCD-like proteins affect Rubisco assembly in plants as well. Now that a function has been identified, detailed mechanistic studies on acRAF and its homologs in plants can be undertaken.

CHAPTER III : PUTATIVE CARBON REGULATORY PII PROTEINS

ABSTRACT

We present here two X-ray crystallographic structures of a nitrogen regulatory PII-like protein genetically linked to alpha-carboxysome operons in proteobacteria. Alpha and beta-carboxysomes are proteinaceous microcompartments that play large roles in bacterial carbon concentrating mechanisms. Carboxysomes inundate two encapsulated enzymes, carbonic anhydrase and ribulose-1,5-bisphosphate (Rubisco), with their respective substrates, bicarbonate and carbon dioxide, resulting in higher efficiencies of carbon fixation. Although carboxysomes are mainly thought of as microcompartments of cyanobacteria, many species of proteobacteria rely on carboxysomes for autotrophic carbon metabolism, especially under Ci-limiting conditions. Interestingly, our laboratory identified several instances of nitrogen regulatory PII-like proteins encoded adjacent to these alpha-carboxysome operons. Canonical nitrogen regulatory PII proteins are ubiquitous signal-transducers that function to regulate nitrogen metabolism within diverse bacteria, archea and plant species. Generally, these small trimeric proteins undergo conformational changes upon binding of 2-oxogluarate, a central nitrogen metabolite, enabling 2-oxoglutarate-bound PII proteins to release direct inhibition of a variety of enzymes, transcription factors and membrane transporters involved in nitrogen metabolism. Very often, PII proteins are genetically encoded adjacent to their target effectors. It is intriguing, therefore, to find nitrogen regulatory PII protein homologs genetically encoded adjacent to carbon metabolizing genes, rather than

nitrogen metabolizing genes. We describe here several unique structural attributes belonging to this new family of carboxysome-linked nitrogen regulatory PII-like proteins, named here ndhPII, from *Thiomonas intermedia*'s carboxysome operon.

INTRODUCTION

Carboxysomes

Carboxysomes are proteinaceous, carbon fixation micro-organelles found within cyanobacteria and several clades of proteobacteria (43, 47). Resembling intracellular viral capsids under a microscope, carboxysomes are in fact icosahedral assemblies of host-encoded shell proteins encapsulating enzymes Ribulose-1,5-bisphosphate Oxygenase/carboxylase (Rubisco) and carbonic anhydrase. The carboxysome walls consist of sheets of selectively permeable bacterial microcompartment shell proteins, BMCs (44), while its twelve vertices are occupied and by bacterial microcompartment vertex proteins, or BMVs (23, 1). Within the carboxysome, Rubisco catalyzes the first rate-limiting step of carbon fixation: the attachment of gaseous carbon dioxide onto the sugar ribulose-1,5-bisphosphate to produce two molecules of 3-phosphoglycerate. The evolutionary advantage of encapsulating these reactions is thought to arise from the carboxysome's ability to concentrate carbon dioxide and exclude oxygen from its interior, allowing Rubisco to carboxylate ribulose-1,5-bisophoshate with enhanced efficiency. There are two known classes of carboxysomes, designated either alpha or beta-type, which are distinguished by Rubisco phylogeny and gene content (47). Alpha-carboxysomes encode type 1A Rubisco, while beta-carboxysomes encode type 1B Rubisco. There are several genes that occur exclusively in alpha-carboxysomes:

CsoS2, acRAF, carboxysome peptides A and B (paralogous BMV proteins) are prominent examples. Conversely, CcmM, CcmN and RbcX are some proteins found only in beta-carboxysomes. Despite these differences, both carboxysome subtypes appear morphologically similar and likely perform identical metabolic functions within their host species.

Nitrogen-regulatory PII proteins

Nitrogen-regulatory PII proteins, referred to here nrPIIs, are a ubiquitous class of signal transduction molecules that regulate various other proteins implicated in primary nitrogen metabolic pathways (74). The first nrPII protein was isolated from *E. coli* in an attempt to unravel mechanisms of glutamine synthetase regulation (75). Since then, PIIs have been shown to directly inhibit glutamine synthetase, as well as several other nitrogen metabolism enzymes such as NAD+ synthase, ammonium transporter AmtB, and nitrogenase transcription factors (76, 77). Table I provides a summary of common PII genetic linkages, including those of nrPIIs and non-canonical type PII proteins, including the ndhPII family described in this work.

Structure and Regulatory Mechanisms of Nitrogen Regulatory PII Proteins

PII proteins undergo several layers of structural modification in response to cellular nitrogen status. These layers of structural regulation include post-translational modifications (such as uridylation, adenylation or phosphorylation), ligand binding (ATP/ADP, Mg2+, 2-oxoglutarate), and ATP-hydrolysis activity. Figure 11 describes the regulation of the ammonium transporter, AmtB, by the nrPII protein GlnK.

Not shown in this figure is the reversible uridylation of GlnK. In *E. coli,* GlnK is uridylated on residue Tyr51 under low ammonium conditions, sterically precluding its binding to

the *amtB* ammonium channel, but is de-uridylated with increasing concentrations of ammonium (78). In cyanobacteria, a conserved serine residue in the T-loop is differentially phosphorylated in response to cellular nitrogen status. In yet other species, PII's are subject to adenylation or to no post-translational modifications.

PIIs have long been known to bind both ATP and ADP, conferring 'active' and 'inactive' states. The historical inclination to imagine nrPII sensing cellular energy status, in addition to nitrogen status, by competitive binding to ATP and ADP has been repeatedly muddied by *in-vitro* experimentation. In short, PII's binding affinities towards both ATP and ADP have proven to be far below physiologically relevant concentrations. Only recently, however, Radchenko *et al* discovered nrPIIs intrinsic ability to hydrolyze ATP to ADP, thereby proposing a non-competitive mechanism for switching between "active" and "inactive" states (79). This mechanism is reminiscent of well-known families of GTP-ases, whose rates of GTP hydrolysis are modulated by other protein-protein or protein-ligand interactions (80). Upon ATP hydrolysis, ADP-bound glnK binds to amtB and effectively blocks ammonium transport (79). The only level of PII regulation directly related to intracellular nitrogen status is the binding of its ligand 2-oxoglutarate, a metabolite whose intracellular concentration is inversely related to ammonium availability (76, 81, 82). Binding of 2-oxogluatarate to ATP-PII inhibits ATP-hydrolysis activity, promoting the retention of its "non-binding" conformation. Numerous crystallographic structures neatly convey how 2-OG binding induces slight structural changes hindering ATP hydrolysis: 2-OG binds a magnesium cation coordinating the phosphates of the ATP bound within the nucleotide-binding pocket of PII (82, 83).

Residues responsible for ATP hydrolysis are not definitively known, but there are several obvious candidate motifs. Aside from the central ferredoxin-like domain shared among all homotrimeric nrPII proteins, there are three loop regions thought to impart specific functional attributes to the PIIs. The T-loop, named after the glnK's conserved tyrosine residue within that loop, affords PII a flexible region for post-translational modification and conformational flexibility to enable nrPII's direct binding to target effector proteins. The B-loop contains a highly conserved Walker A motif "TGxxGDGKI", thought to be critical for ATP binding and ATP hydrolysis. The C-loop, so named due to its location at the C-terminus, and is also implicated in both ATP/ADP binding and promoting conformational changes within the T-loop. The nrPII C-loop contains highly conserved Arg residues known to interact with the phosphates of ATP and ADP. Figure 11 shows an ideological scheme of AmtB regulation by GlnK, PDB ID 2NUU (84).

Figure 11. Ammonium transporter, AmtB, inhibition by GlnK under low 2-OG concentrations

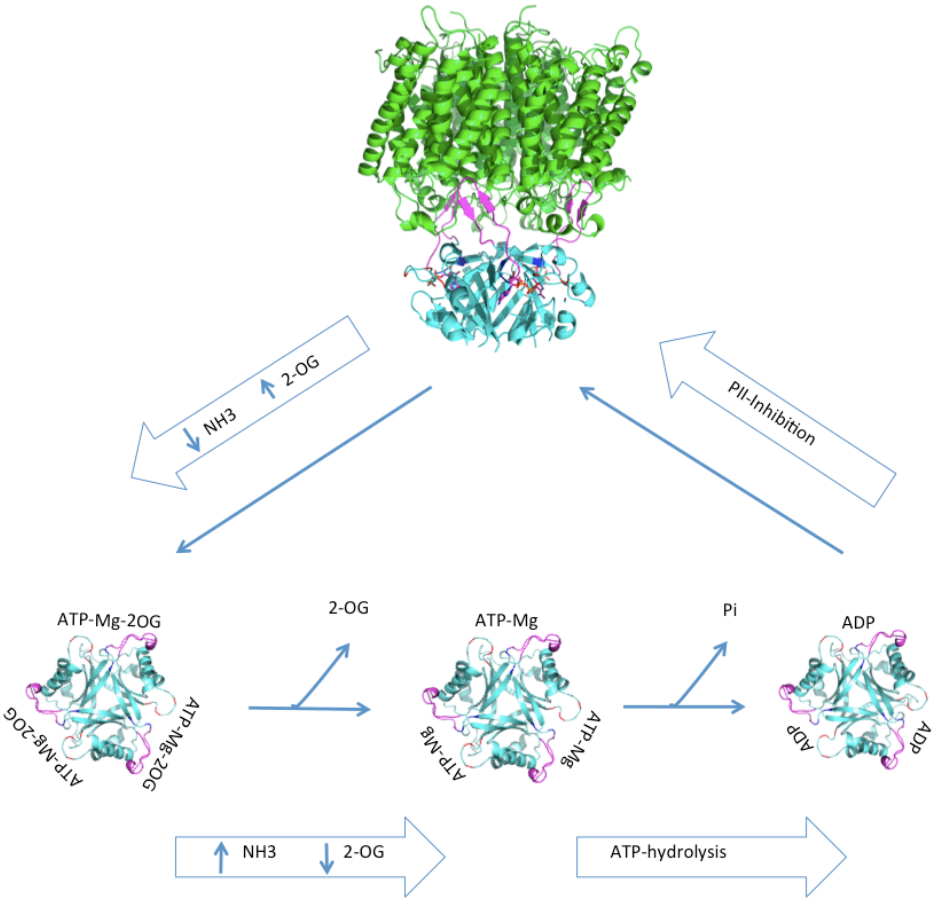


Figure 11. Mechanism of inhibition of ammonium uptake by PII under nitrogen sufficient conditions
Beginning at the lower left-hand: High levels of 2-OG, indicative of ammonium limitation, enable 2-OG binding to ATP-bound GlnK, preventing ATP-hydrolysis. However, upon ammonium increases, 2-OG levels decrease. 2-OG is released from ATP-bound PII proteins, leading to ATP hydrolysis. GlnK-ADP adopts a conformation enabling binding, and thus blocking, of the AmtB channel. When nitrogen levels again decrease, 2-OG antagonizes ADP binding to GlnK, accelerating nucleotide exchange for ATP, promoting its return to the cytoplasm. Other layers of regulation, including glnK uridylation, are not shown.

Connections between autotrophic carbon metabolism and nitrogen metabolism

Most nrPII protein targets function strictly within nitrogen metabolism. However, in 2010, researchers identified an unprecedented nrPII target enzyme in plant chloroplasts: acetyl-CoA carboxylase, or ACCase (85). ACCase is a carbon biosynthesis enzyme, catalyzing the carboxylation of acetyl-CoA to produce the lipid precursor metabolite malonyl-CoA. An ACCase subunit is directly inhibited by PII when concentrations of 2-oxoglutarate are low, making it the first known instance of a nitrogen regulatory PII protein inhibiting carbon metabolic enzymes. The inter-dependence of carbon and nitrogen metabolic regulation has long been acknowledged in many autotrophic organisms, but direct evidence of nitrogen/carbon co-dependent regulation is limited. Studies in cyanobacteria have shown that nitrate/nitrite transport is regulated by PII proteins and is dependent on carbon availability (86, 87). Another putative example of carbon regulation in response to nitrogen availability is the discovery that CcmR, a ubiquitous transcriptional regulator of bicarbonate transporter and carboxysome genes, is allosterically regulated by 2-oxoglutarate, the same metabolic intermediate recognized by nrPIIs (88), suggests that Ci-uptake (Ci meaning inorganic carbon), and consequentially carbon fixation, is dependent on nitrogen availability in cyanobacteria.

Inorganic Carbon Concentrating Mechanisms in Cyanobacteria

Cyanobacteria have a diverse collection of bicarbonate transporters, many species containing 3 or more transport systems per genome (89). The systematic redundancy of Ci-uptake affords the capacity to regulate mechanisms of Ci-uptake in accordance to intracellular metabolism and extracellular carbon concentrations. For instance, low- and high-affinity Ci-uptake systems are utilized under conditions of high and low extracellular Ci concentration, respectively. Some Ci-uptake complexes recognize and

transport bicarbonate. Others concentrate carbon dioxide via intracellular hydration to HCO_3^-. Moreover, the energy source of the various Ci transporters varies; some siphon potential energy from ionic gradients, while others utilize ATP hydrolysis. A comprehensive review of the known bicarbonate transporters in cyanobacteria is available (90). Complex I NDH-homologs (Complex I NADH:ubiquinone oxidoreductase, also known as NADH Dehydrogenase) are members of cyanobacterial CCMs (carbon concentrating mechanisms). NDH subunit homologs incorporate into complex I assemblies, converting them into carbon concentrating complexes coupled to the electron transport chain (91, 92). These CO_2-uptake operons are comprised of genes encoding for two Complex I subunit homologs NdhF and NdhD, along with either ChpX or ChpY, whose incorporation into modified Complex I affords either low- and high-affinity carbon uptake, respectively. SbtA, a gene induced under inorganic carbon limitation, encodes for a Na+/HCO3- symporter and has been studied in beta-cyanobacteria. SbtB, encoded adjacent to SbtA, is a small gene with detectable sequence similarity to nitrogen regulatory PII proteins. Although putative SbtAB operons are found in most alpha-cyanobacteria, they are have not yet been evaluated in these organisms. As of yet, the function of SbtB has not been identified; its genomic deletion does not impair SbtA-mediated bicarbonate concentration in beta-cyanobacteria *in vivo* (90).

Table 3. **Genetic Linkages of PII proteins***

Gene Name	Target Protein(s)/Genetic linkages	B-loop motif	Sample Protein ID
glnB	Glutamine Synthetase (glnA) and NAD Synthetase (nadE)	yes	WP_012563053
glnK	Ammonium transporter (amtB)	yes	WP_001700364
$nifH_1$ / $nifH_2$	Nitrogenase (nifD, nifK, nifH)	yes	WP_017212870/ WP_026889541
PII-NX	Nudix hydrolase family	yes	WP_006633346
PII-like proteins			
PII-NG	Metal toxicity genes	no	YP_550304.1
ndhPII	NDH gene cluster	no	YP_003641857
SbtB	SbtA, Na^+/HCO3- transporter	no	WP_010872646

* PII is, in many cases, genetically linked to more than one gene at a time. Here we display only the most obvious genetic linkage per category.

METHODS

Genetic Linkage Detection

PII genetic linkage analysis was performed, analyzed and verified using the following online programs: SEED (93), STRING (94), Clustal Omega (95), NCBI protein domain search and NCBI protein BLAST tool.

Cloning and protein expression of ndhPII

A gBlock (purchased from Integrated DNA Technologies, Inc.) encoding *T. intermedia* ndhPII sequence with a C-terminal 6-His tag was Gibson assembled (34) into pET28a between the NcoI and HindIII restriction enzyme sites. A sequence-verified clone was transformed into BL21(DE3) cells. Cells were grown overnight in 100 mL LB broth, 15 uM kanamycin. 15 mL of the overnight culture was used to inoculated 1L LB baffled flasks. These inoculates were grown, shaking, at 37˚C until the OD_{600} reached 0.8. At this point, protein expression was induced with 1 mM ITPG. Cells were let shake at 23 degrees Celsius for 24 hours. Afterwards, culture pellets were kept frozen at -20˚C.

Purification of NdhPII

Cell pellets were resuspended in the following buffer: 50 mM Tris-HCl pH 7.6, 300 mM NaCl, Roche complete EDTA-free protease inhibitor cocktail, and lysed by sonication. Cell debris was pelleted by centrifugation at 16,500 rpm in SS-34 rotor for 30 minutes at 4°C. The resulting supernatant was filtered through a 0.22 µm Whatman filter. Filtered supernatant was then loaded onto a HisTrap 5 mL column by syringe, washed with 30 mL wash-buffer (50 mM Tris-HCl pH 7.6, 300 mM NaCl, 50 mM imidazole), and eluted in one step with elution buffer (50 mM Tris-HCl pH 7.6, 300 mM NaCl, 500 mM imidazole). Purified ndhPII was dialyzed into 10mM Tris-HCl pH 7.6, 50mM NaCl. NdhPII was concentrated to target concentrations by measurement of absorbance at 280 nm, using its theoretical extinction coefficient.

nhdPII crystallization

Hanging-drop crystallization screens were performed using the nanoliter liquid handling Mosquito from TTP LabTech at the High Throughput Macromolecular Crystallization Facility at UCLA. The crystallization condition for the apo-ndhPII crystal structure was

38 mg/mL protein, combined in a 1:1 ratio with 20% PEG 3500, 0.15 M LD Malic Acid pH 7.0. The crystallization condition for the ADP-ndhPII crystal structure was 25 mg/mL protein, combined in a 1:1 ratio with 2M Ammonium Sulfate, 0.1 M BisTris pH 5.5. Crystals grow within 1-4 days at room temperature.

Phasing and Refinement

Diffraction data was collected with the UCLA X-ray diffractometer. Phases of the ndhPII structure were determined by molecular replacement with a putative SbtB structure from *T. thermophilus* H8, PDB ID 2CZ4, as a search model. ADP-ndhPII phases were provided by molecular replacement using the previously solved apo-ndhPII monomer as a search model. Molecular replacement, model development and refinement was performed with PHASER(35), COOT (36) and BUSTER (64), respectively. Final R-work and R-free statistics were 0.17 and 0.20 for the ndhPII structure and 0.22 and 0.24 for the ADP-ndhPII structure, respectively. Table 4 reports detailed crystallographic information of both these structures. Coordinates have not yet been deposited into the PDB.

RESULTS AND DISCUSSION

Identification of a novel NDH-PII gene cluster

Our laboratory was intrigued to discover a putative PII protein genetically linked to alpha-carboxysome operons in several species of proteobacteria. Further bioinformatic analysis confirmed the significance of PII's genetic linkage to alpha-carboxysomes and, surprisingly, led to the discovery that carboxysome associated PII-like proteins are members of a larger, novel class of NdhF-associated (Complex I NADH dehydrogenase

subunit L homolog) PII-like proteins. For the purposes of this paper, we call this new subcategory of PII-like proteins “ndhPII”. NdhPII is member of a highly conserver 4-gene operon. Only two of these genes bear sequence similarity to known proteins: respiratory complex I subunit homolog, NdhF, and ndhPII. The third is a putative 8-pass transmembrane protein, whose primary structure is loosely conserved. The fourth member of the NDH-PII operon is a large, predicted cytoplasmic, protein of unknown function. Figure 12 displays examples of the genetic linkages of this PII-NDH operon, encoded adjacent to a variety CO2/HCO3- utilizing metabolic pathway genes, including pyruvate:ferredoxin oxidoreductase subunits and alpha-carboxysome operons. There exists an unacknowledged assumption that Complex I modification as a mechanism of carbon concentration is specific to cyanobacteria. However, genetic evidence provided in this work challenges that assumption. NDH-PII clusters (clearly distinguishable from cyanobacterial NdhF/D/ChpX(Y) gene clusters) occur in a variety of different types of prokaryotic species – even those without carboxysomes - suggesting that modification of Complex I towards the purposes of regulated, intracellular carbon concentration might be a wide-spread phenomena.

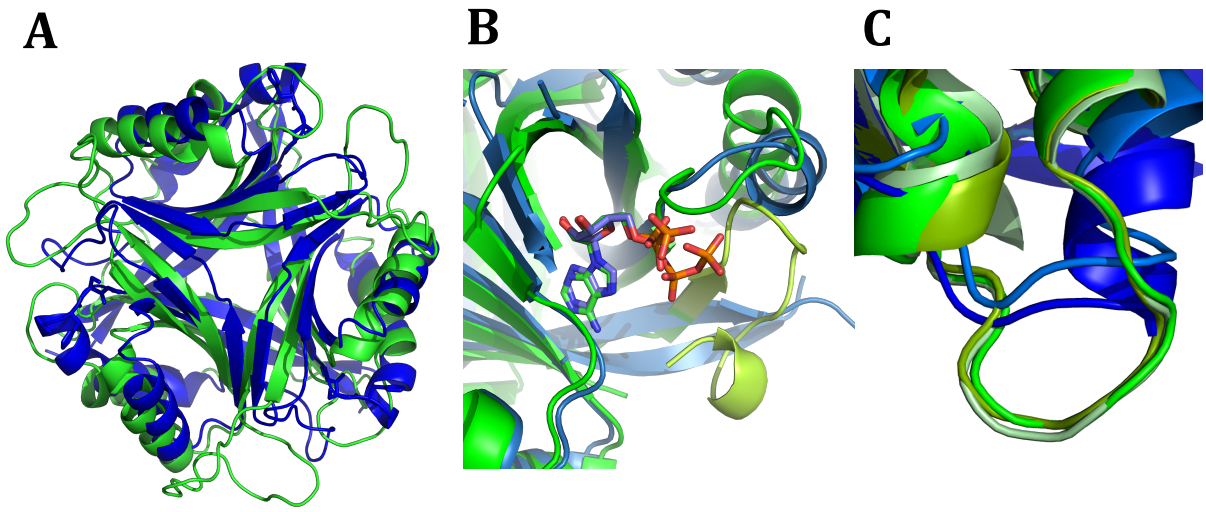
Figure 12. ndhPII genetic linkage to proteobacterial alpha-carboxysome and other operons

Figure 12
Diagram of genetic linkages between NDH-PII gene clusters and the following: alpha-carboxysome, pyruvate:ferredoxin oxidoreductase (PFOR) and genes encoding putative bicarbonate transporter system SbtAB.

The PFOR enzyme complex, also known as pyruvate synthase, catalyzes the reversible fixation of CO_2 onto acetyl-CoA to produce pyruvate and CoA (co-enzyme A). Ferredoxin is used by PFOR as an electron source and sink, depending on the reaction's directionality. The PFOR genetic association with this putative Ci-

concentrating NDH-PII cluster might be a consequence of the reversibility of PFOR catalysis; increasing the concentration of dissolved bicarbonate could slant the PFOR reaction towards the production of pyruvate. Also shown are genetic linkages of NDH-PII genes to putative SbtAB operons. Although the NDH-PII – SbtAB association is not common enough to warrant its own classification or any hurrah of that sort, it is included due to the interesting fact that in both instances of NDH-PII gene linkage to SbtAB, the NDH-PII operon is missing the ndhPII gene. It is tempting to speculate that SbtB might fulfill the role of ndhPII in these organisms.

Figure 13. Structural comparison of nrPIIs and putative carbon regulatory PII proteins, crPII proteins



D

Variable T-loop B-loop C-loop

```
glnK    ..SSIGIQGLTVTEV-KGFGRQ...VNFLPKVKIDVAIADDQLDEVIDIVSKAAYTGKIGDGKIFVAELQRVIRIRTGEADEAAL
glnB    ..AEVGITGMTVTEV-KGFGRQ...VDFLPKVKIEIVVPDDIVDTCVDTIIRTAQTGKIGDGKIFVFDVARVIRIRTGEEDDAAI
nifI1   ..LASGCLALTKMDV-IGRGKQ...YDELPKTMLMLVVEDDTAENVIELITKTAYTGSFGYGKIFVSPVDEAYTVRTRSCGL---
nifI2   ..EGLGFPAMTANRV-LGRGKQ...MRYIPKTMLTLIVPDEDASLVVEAIMKVNKSGQYGDGKIFVCPIEDIITVRTSERGEAAI
ndhPII  ..DRAGVKGYTIVGNLSGKGSH...DD--ALIMIIAAVPEELVGPLLEGFQ-PFFE--AHSGVVFVHDIQVGRPIKFRN------
sbtB    ..DESGAKGYTVMNT-GGKGSR...EA--NIKFEILTETREMAEEIADRVAVKYFN--DYAGIIYICSAEVLYGHTFCGPEGC—
```

Figure 13. Structural comparisons between nitrogen regulatory PII proteins and putative carbon regulatory PII proteins (crPIIs)
A. GlnB from *E. coli* aligned to ndhPII from *T. intermedia*: RMSD of 5.4 Angstroms. **B.** Structural overlay of GlnB-ADP (green) and ndhPII-ADP (blue). GlnB C-loop shown in lighter green. No similar C-loop is found in ndhPII. 2 Angstrom RMSD **C.** Putative crPIIs posses a relatively short B-loop. Both crPIIs, ndhPII from *T. intermedia* and putative SbtB from *T. thermophilus* H8, are colored blue. NrPII proteins are colored green; GlnB from *E. coli* PDB ID 2PII, GlnK2 from *A. brazilenses* PDB ID 3NCP, GlnK from *Synechococcus* sp. 7942 PDB ID 1QY7 **D.** Sequence alignment of select members of the PII superfamily: GlnK from *E. coli* WP_001700364, GlnB from *E. coli* NP_417048, NifI1 from *M. maripaludis* P0CW44 or GLNB1_METMI, NifI2 from *M. maripaludis* WP_011170799, NdhPII from *T. intermedia K12* YP_003641857, SbtB from *Synechocystis* PCC 6803 WP_010872646.

Putative crPII B-loop sequences are shorter, and have no detectable Walker A-like

“TGxxGxGK motif, thought to be involved in ATP-hydrolysis. NdhPII, however, does

contain a conserved His-Ser-Gly motif with its respective B-loop. CrPIIs do not have conserved arginine/lysine residues involved ATP-phosphate interactions.

Figure 14. Apo-ndhPII aligned to ADP-ndhPII

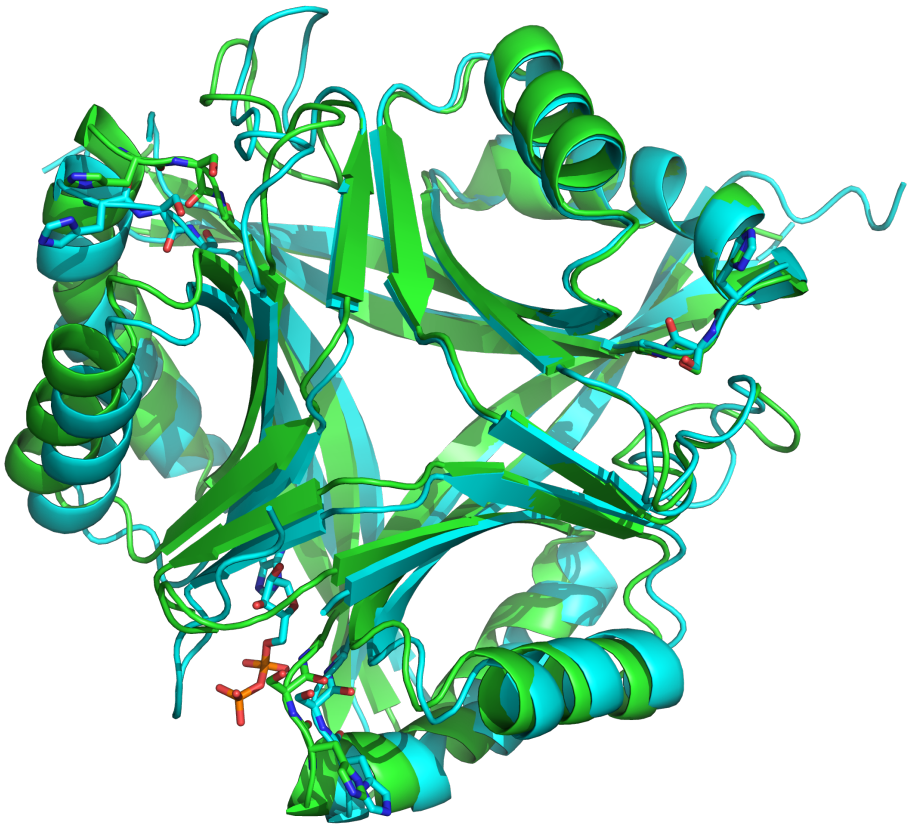


Figure 14
ADP-bound ndhPII (cyan) aligned to apo-ndhPII (green) 1.2 Angstrom RMSD. ADP shown in sticks. Conserved residues HSG, or His-Ser-Gly, along ndhPII B-loop shown in sticks. A single ADP-binding event induces significant structural changes in only one of its two adjacent nucleotide binding sites.

Figure 14 shows the structural overlay of ligand-free and ADP-bound ndhPII. Strangely, ADP occupies only one of the three binding sites, even though co-crystallization trials were performed with 2 mM ligand concentrations. It is interesting, however, to notice that the single ADP induces asymmetrically structural distortions within the other two adjacent nucleotide-binding pockets. This asymmetric distortion caused by a single

binding event could possibly be evidence that ndhPII is perhaps capable of a cooperative nucleotide binding.

Table 4. X-ray Data Collection and Model Refinement Statistics of ADP-bound an Apo-ndhPII models

Description	Apo ndhPII	ADP ndhPII
Unit cell (Å)	*a* = 82.69 *b* = 82.69 *c* = 31.5	*a* = 55.250 *b* = 57.670 *c* = 120.850
Unit cell angles	90, 90, 120	90, 90, 90
Total reflections recorded	49,986 (3856)	130336 (7457)
Resolution	2.301	1.9
Space Group	P 3	P 21 21 21
Unique reflections	10491 (747)	30220 (1852)
Multiplicity	11.54 (10.26)	4.312 (4.02)
Completeness (%)	97.70 (97.7)	95.42 (95.42)
Mean I/sigma(I)	17.0 (3.2)	15.27 (3.02)
Wilson B-factor (Å^2)	34.25	35.79
R-meas[25]	12.8% (69.8%)	4.47% (44.2%)
Model R-work	0.177 (0.203)	0.2187 (0.4638)
Model R-free	0.204 (0.229)	0.244 (0.4519)
Number of non-hydrogen atoms	1515	2379
Macromolecules in asymmetric unit	2	3
Sulfate	1	0
Water	54	121
Protein residues	406	618
Geometric deviations (rms)		
Bonds (Å)	0.01	0.010
Angles (°)	1.18	1.20
Ramachandran favored (%)	99	99

Ramachandran outliers (%)	0	0
Average B-factor (Å^2)		
Protein Atoms Chain A	35.14	61.2
Protein Atoms Chain B	42.7	53.2
Protein Atoms Chain C	NA	43.6
Adenosine Diphosphate	NA	81.5
Solvent (Water)	38.66	48.6
Solvent (Sulfate)	92.02	NA

Statistics for the highest-resolution shell are shown in parentheses.

NrPII – Nudix hydrolase genetic linkage

The PII superfamily is a diverse set of signal transducers, honed throughout evolution to adjust an organism's intracellular metabolism to the reality of its extracellular environment. GlnB and GlnK, and to lesser extent NifI1 and NifI2, are well-studied nitrogen regulatory molecules. Many comprehensive reviews are available for detailed information regarding canonical nrPIIs (87, 96, 97). PII-NX proteins, acknowledged in this work for the first time, are seemingly indistinguishable from glnK/glnB family sequences, but are genetically encoded adjacent to Nudix hydrolase-type proteins. Nudix is an acronym for Nucleotide Diphosphate linked to X, "X" being any nucleotide according to experimentally determined specificities. Nudix hydrolase representatives recognize a diverse set of di-phoshpo-nucleotides, catalyzing the hydrolysis of their pyrophosphate moieties to produce a nucleotide monophosphate and, in many cases, ortho-phosphate. The significance of this genetic linkage is unknown. Further investigation into to the nature of PII-Nudix hydrolase associations may result in the discovery of a novel PII function, or reveal yet another layer of fine-tuned regulation.

Unclear phylogenetic heritage between PII-like subfamilies

Given the observed interdependence of carbon and nitrogen metabolism in autotrophic bacteria and chloroplasts, one can imagine that carbon and nitrogen regulatory PII proteins might share a common ancestry. However, the significance of the sequence relationships between these groups are difficult ascertain for several reasons. PII subunits are often fewer than 100 amino acids in length; resulting in reduced statistical likelihoods that observed sequence similarities are reflections of shared protein ancestry. Moreover, the PII-fold belongs to the ubiquitous and highly diverse ferredoxin-fold superfamily, leaving ample space-time for independent PII-like proteins to have branched off from functionally unrelated ancestors. A exemplary case for consideration is the “New Group” PII-like subfamily (PII-NGs), described in (96). PII-NGs are a highly sequence-diverse collection of PII-like proteins which are found encoded adjacent to putative heavy metal transporters; indicating their role in either sensing heavy metals or, plausibly, controlling cellular stress responses in accordance to intracellular metabolic status. Although no PII-NG studies have been reported to date, it is helpful to consider the possibility that these diverse PII-NGs might have arisen from a set of selection pressures entirely distinct from the selection pressure of nitrogen limitation, and hence from nrPII proteins. CutA1 proteins, a well-studied group of heavy metal sensors (98), bear no detectable sequence similarity to PII-NGs, in spite of the surprising fact that their three-dimensional structures overlay surprisingly well (Often >2 Angstrom RMSD compared to ndhPII, example cutA1 PDB ID 4E98). The disparity between their high structure similarity and undetectable sequence similarity is easily understood when one notices these align with opposite N to C-terminal orientations. This surprising

observation enforces the interpretation that these proteins evolved the same fold and putative function without sharing any common ancestor. If these inherently promiscuous signal-sensing folds are capable of such a type of convergent evolution – does it not reflect the insignificance of inferring shared ancestry between mildly-sequence-similar PII-like subfamilies whose structures overlay in shared N- to C-terminal orientations? Even though ancestral relationships may certainly exist between the various PII-like subfamilies, due to observations like the above convergent evolution, and the ubiquitous nature of the ferredoxin-fold superfamily to which the PII-fold belongs, we limit our attempts to infer shared ancestry between PII-sequence-similar subfamilies, including the relationship between carbon regulatory sbtBs and ndhPIIs considered in this work.

SbtB and ndhPII as regulatory signal transducers to putative inorganic carbon transporters

SbtB and ndhPII proteins are genetically encoded adjacent to SbtA and a novel NDH gene cluster, respectively. Experiments have shown that SbtA encodes for a transmembrane Na^+/HCO_3^- symporter. NDH gene clusters, encoded adjacent to alpha-carboxysome operons in proteobacteria, are hypothesized to function analogously to the known (but phylogenetically distinct) NDH gene clusters in cyanobacteria and chloroplasts, wherein paralogous NDH subunits modify core respiratory Complex-I assemblies, imparting on them the ability to concentrate carbon dioxide. This work evaluates for the first time two novel categories of crPIIs, or carbon-regulatory PII proteins. Both identified crPII families are genetically linked to distinct, non-homologous putative Ci concentrating proteins. It is interesting to note that, not only are the two Ci-uptake proteins non-homologous, their methods of concentration differ fundamentally.

SbtA is likely a sodium-bicarbonate symporter, transcriptionally induced under conditions of low inorganic carbon; it is thought that high affinity Ci-uptake could be achieved through the use of a sustainable sodium gradient across the outer to inner membranes (90). It is possible that the function of SbtB might be to protect and maintain the sodium concentration gradient, upon which several other transport-processes depend. Might SbtB block unnecessary transport under conditions of adequate intracellular bicarbonate concentrations?

The NDH-PII protein assemblies, if they function analogously to cyanobacterial NhdD/NdhF/ChpX(Y) systems (91, 92), would be coupled to an electron transport chain. In cyanobacteria, Complex I mediated Ci-uptake is coupled to photosynthetic electron flow (99). However, NHD-PII gene clusters appear in a diverse set of largely chemoautotrophic bacteria, or non-photosynthetic organisms, suggesting its functional integration into a variety of distinct electron transport pathways. Many questions remain unanswered. For instance, does the NDH-PII assembly concentrate CO2 only when its electron transport chain is flowing? Furthermore, does the NDH-PII assembly siphon energy from that electron transport chain, thereby inciting a necessity for redox regulation? Several metabolites reflect the health of an electron transport chain: those involved in the electron transport chain itself, and those functioning within the tricyclic acid cycle. ATP, NAD+, NADH, 2-oxoglutarate and succinate are examples of likely key metabolites whose concentrations, or relative concentrations, reflect the health of an electron transport chain. NADP+:NADPH ratios, on the other hand, are reflective of an organism’s autotrophic capacity. Primary and tertiary structural analysis of ndhPII affords no meaningful predictions as to its ability to bind either 2-oxoglutarate or

succinate. However, structural analysis of crPIIs strongly suggests both an impaired ATP-hydrolysis motif and perhaps an impaired ability to distinguish between ATP and ADP, since both nitrogen-regulatory B- and C-loop motifs are absent in crPIIs (Fig 13). We analyze the ADP-ndhPII crystal structure to evaluate ndhPII's potential to bind other types of redox-relevant nucleotides.

NhdPII binds ADP and ATP in vitro

NdhPII successfully co-crystallized bound to 1 molecule of ADP (Fig 13 and Fig 14), proving its adenosine nucleotide binding potential. It is a glaring inconsistency, however, that only one of three identical ADP-binding sites of ndhPII is occupied. This could arguably be a crystallographic artifact; perhaps ADP binding to either of the two other binding pockets would obliterate crystallization due to the T-loops' participation in crystal contact formation. Alternative explanations could be that ndhPII truly only binds one ADP, or that it binds ATP-ADP in combination (not attempted), or that the ADP binding itself is a crystallographic artifact attributable to non-physiological concentrations of ADP, which was 2 mM in these co-crystallization trials. Crystals diffracting to low-resolutions were also grown in the presence of 2 mM ATP, again revealing just one binding event (ATP-bound structures not shown). Taking into consideration the major structural differences between the crPII and nrPII ATP-interaction loops (B-loops and C-loops), we wonder if the crPII ATP binding pocket might be better suited for other nucleotides. All nicotinamide-type nucleotides, NAD(P)(H), are structurally similar to ADP. In fact, NAD+ and NADH molecules have a complete ADP moiety in common. What differentiates between them is what is attached to the terminal ADP phosphate: either oxidized or reduce versions of nicotine-riboside. NADP+ and NADPH are

structurally identical to NAD+ and NADPH, save for a distinguishing phosphate replacing the 2' hydroxyl group on the adenosine ribose. One can imagine that NAD(P)(H) might bind the ndhPII nucleotide pocket similar to how ADP is shown binding it in the ndhPII-ADP crystal structure.

Interestingly, nrPII from cyanobacteria *Synechocystis* PCC 6803 is modulated by intracellular redox state (86). That same cyanobacterium utilizes an NDH-type CO_2 carbon concentration mechanism. Specifically, phosphorylation of nrPII in this organism is affected by the cellular redox state. Although it is unknown whether ndhPII is subject to posttranslational modifications, and it is unlikely that *Synechocystis* PII binds alternative redox nucleotides *in vivo*, it is entertaining to speculate that they both might regulate intercellular metabolism in response to their redox-dependent bicarbonate uptake mechanisms, albeit in their own unique ways.

THESIS CONCLUSION

Many questions remain about bacterial microcompartments. Proteins, whose structure and function are surely of utmost importance to bacterial microcompartment function, remain classified as 'proteins of unknown function'. Many of these hypothetical proteins– strongly associated with bacterial microcompartment sub-types – offer some bioinformatic clues as to their functions. For example, partitioning enzyme (ParA-like) and bacterioferretin (Bfr-like) proteins are often encoded within alpha-carboxysome operons (2). ParA may plausibly participate in carboxysome distribution among daughter bacterial cells. Beta-carboxysomes, for example, employ distinct parA enzymes that partition carboxysomes between dividing cells. Iron chelation by

bacterioferritin, which may either be physically located within the carboxysome or simply expressed during carboxysome formation, could protect encapsulated enzymes from Fe-induced reactive oxygen species. Prevention of oxidative damage to Rubisco and carbonic anhydrase could be preferable to repairing them, considering their enclosure within microcompartments. There are still others however, whose functions are absolutely mysterious. HAM-1 domain containing protein, strictly conserved and encoded opposite and adjacent to CsoS1D in cyanobacteria alpha-carboxysome containing bacteria, is an example (56). I had hoped to discover its function, but none of the various methods of heterologous expression have yet produced a soluble form carboxysome associated HAM-1 protein. The structure and function of CsoS2, a long acknowledged member of alpha-carboxysome operon, is still undetermined. In other microcompartments, the Eut and Pdu MCPs, still more types of proteins are awaiting elucidation. GrpU (PDB ID: 4OLP and 4OLO) (100), PduT (PDB ID: 3PAC and 3N79) (101) and PduK are examples of shell proteins that contain iron sulfur clusters occupying their central pore regions. Do they exist to transfer electrons in and out of the microcompartment? Do they sense intracellular redox status? The fact that Pdu microcompartment shell faces are redox-capable is apparent. However, the purpose and mechanism of the redox-capability remains elusive. Additionally, more often than not, several paralogous BMC shell proteins are encoded within single MCP operons. To which are functional niche does each belong? The EutK BMC shell protein paralog contains a domain capable of binding nucleic acids (19). The consequences of the nucleic acid binding capability remain unclear.

Although much remains to be understood, I hope this work has filled in at least some of the void representing all unknowns of MCP formation and function. The dual pentameric - hexameric assembly states of BMV shell proteins was inconsistent with our understanding of the geometry of microcompartments (19). Icosahedrons, whose faces are comprised of numerous hexagonal BMC shell proteins, must by geometric definition have five-sided vertices. This work, combined with on-going experiments, attempts to reconcile the apparent BMV inconsistency. Our hypothesis now is that all BMV shell proteins are capable of forming either pentameric or hexameric assemblies. Their binding partners and environment, however, specify which oligomeric state is thermodynamically favored. Studies of virus assembly has shown this to be viable mechanism of pentagonal vertex formation; BMVs could stably incorporate into growing layers of BMC shell proteins as hexamers, and under correct conditions, adopt pentagonal geometry by ejection of a single subunit to fulfill the function of a microcompartment vertex protein.

Two years ago, a new question regarding MCPs emerged: why is a PCD-like protein encoded in the alpha-carboxysome microcompartment operon? (14, 56) And why wasn't it noticed before? Even though the PCD-like protein is present in 100% of alpha-carboxysome operons sequenced so far, the historical significance of the linkage was under-appreciated for several reasons (2). Firstly, canonical PCD enzymes have no understandable place in carboxysome function. Secondly, in many operons, the PCD-like protein is simply annotated as 'hypothetical protein', greatly reducing the likelihood of observers realizing the significance of the PCD-like genetic linkage. Thirdly, and importantly, the PCD-like protein's position within alpha-carboxysome is highly variable

(unlike CbbLS, CsoS2, CA, BMVs and BMCs) and is often found next to genes that were *also* previously unknown to be alpha-carboxysome associated. To many observers, it could easily seem that the putative PCD or hypothetical protein is simply near an alpha-carboxysome operon, but not apart of it. In fewer than two years, our lab showed it to function as a Rubisco folding chaperone, and we named it acRAF for alpha-carboxysome Rubisco Assembly Factor.

Currently we are trying to understand the reason for the genetic linkage between alpha carboxysomes and nitrogen regulatory PII-like proteins. Online genetic analysis tools, including STRING and SEED, proved to be very helpful in our endeavor: the nitrogen regulatory PII like protein was shown to be but one gene of a four-member gene cluster. As independent linkages to alpha-carboxysomes

, these genes of unknown function gave no satisfactory hints as to their functions: one distantly related Complex I NADH oxidoreductase subunit homolog, one very large soluble protein of unknown function, a very-low-complexity transmembrane protein of unknown function, and one PII-like protein. Considered as a group, however, testable hypotheses are easily imagined. The four-member cluster bears just enough resemblance to well-known clusters in cyanobacteria, encoding proteins that modify Complex I towards the means of carbon dioxide concentration. When compared to this analogous system, it could be hypothesized that the large soluble protein of unknown function is yet another novel carbonic anhydrase, that the NdhL subunit integrates into Complex I with structural support from the low complexity transmembrane protein, and that PII regulates the overall carbon concentration faculty in response to transient changes cellular metabolic status.

Careful genomic analysis has been crucial for all of the work presented in this thesis. Using appropriate bioinformatic tools is of paramount importance – it is impossible to determine genetic linkages by eye with our annotation system in its current state. The biggest obstacle to noticing highly penetrant genetic linkages is reading an annotated protein as ‘hypothetical protein’ or ‘protein of unknown function’; those terms preclude humans from noticing even the most obvious genomic patterns, time and again. DUF or PFAM numbers, and unrestricted combinations thereof, would be a much better way to annotate proteins of unknown function in public gene databases.

Appendix

Ethanolamine Microcompartment Shell Protein-Protein Interactions

INTRODUCTION

Mixtures of two distinct shell proteins (0.5 mg/mL each) from all the BMC/BMV shell proteins of the Eut MCP form E. coli were co-incubated for 5 days at 4˚C. No other time scales were tested. Samples of proteins incubated either alone or in together were analyzed by native gel.

METHOD

Methods of cloning, expression and purification are described in “Structure and mechanisms of a protein-based organelle in *Escherichia coli*” published in Science in 2010, PMID: 20044574. After separate purifications, each Eut shell protein was dialyzed against 50 mM Tris pH 7.6, 300 mM NaCl, 12% glycerol. Protein solutions were diluted to 0.5 mg/mL, as determined by BCA assay using bovine serum albumen as a standard. Shell proteins were mixed with themselves and with other shell proteins and stored at 4˚C for 5 days. On the fifth day, a fraction of each sample was mixed with 50% PEG 8000 in a 1:1 ratio, and let set at 4˚C for 45 min. Incubations were then run on a native gel (BioRad CAT# 456-1096, 4-20%) for 20 min at 90V and then 1 hour at 150V and stained with Coomasie Blue.

Pictures of resulting native gels were taken under identical camera settings (including zoom, and camera’s distance from gel) for all gels. Afterwards, gels photos were imported into Adobe Photoshop, where each lane was made into its own smaller file, but whose dimensions did not vary from the original source photo. Gel lanes were then re-aligned to afford simple side-by-side comparisons of mixed and unmixed proteins.

Figure 15 shows a cartoon gel with the following lane identities from right to left: shell protein A alone, shell protein A with 25% PEG 8000, shell proteins A and B mixture, shell proteins A and B mixture with 25% PEG, shell protein B alone, shell protein B with 25 % PEG. Figure 15 describes, by cartoon, how to interpret the images in Figure 16.

Figure 15

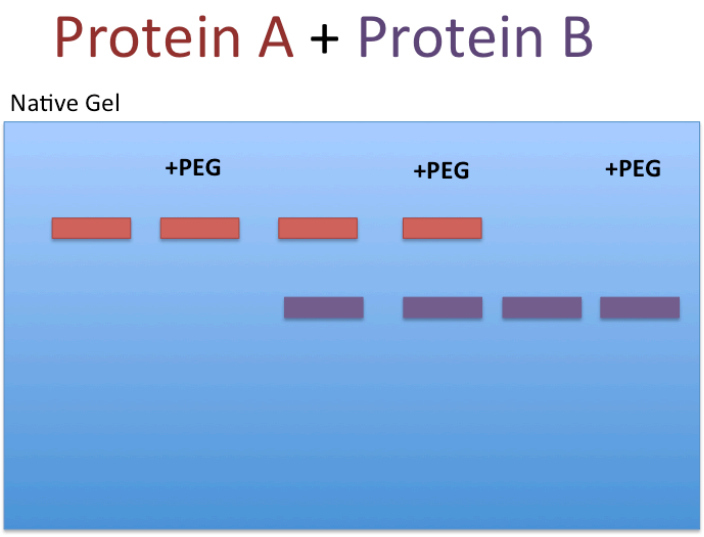


Figure 15. Cartoon guide for interpretation of the gels shown in Figure 16. This particular gel does not show any detectable shift in band migration, indicative of a failure to detect protein: protein interaction.

Figure 16

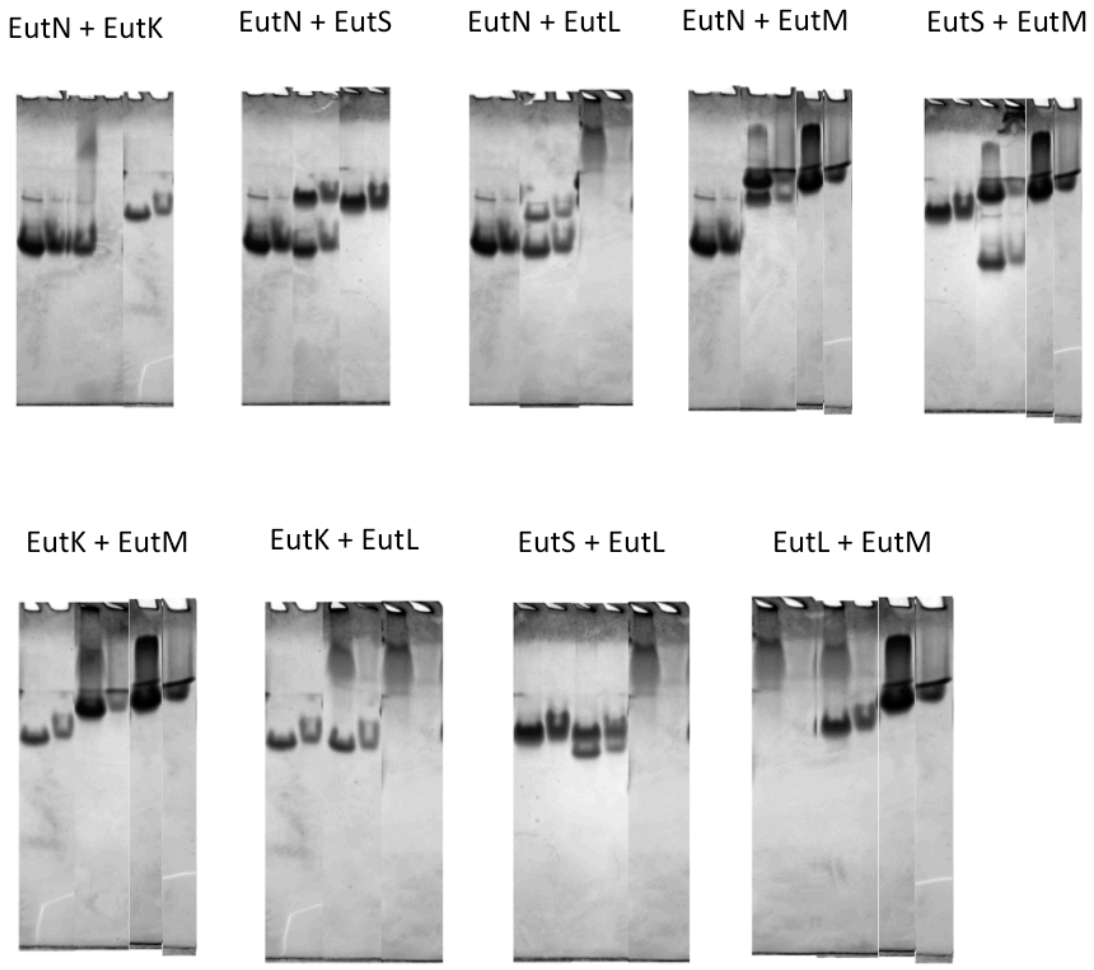


Figure 16. Native PAGE of Eut MCP shell protein mixtures**.** See Figure 15 for proper interpretation of Figure 16.

RESULTS AND DISCUSION

The information one might obtain from the above gels is limited, due in part to the fact that within the mixture lanes, it is impossible to known which band corresponds to which

protein. These images provide some insight into the common ideal that assumes that certain shell proteins interact only with a select set of other shell proteins. Cautions interpretation of data is left to the reader.

References

1. Wheatley, N. M., Gidaniyan, S. D., Liu, Y., Cascio, D., and Yeates, T. O. (2013) Bacterial microcompartment shells of diverse functional types possess pentameric vertex proteins. *Protein Sci.* **22**, 660–665
2. Wheatley, N. M., Sundberg, C. D., Gidaniyan, S. D., Cascio, D., and Yeates, T. O. (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
3. Kerfeld, C. A., Heinhorst, S., and Cannon, G. C. (2010) Bacterial microcompartments. *Annu. Rev. Microbiol.* **64**, 391–408
4. Yeates, T. O., Thompson, M. C., and Bobik, T. A. (2011) The protein shells of bacterial microcompartment organelles. *Curr. Opin. Struct. Biol.* **21**, 223–231
5. Yeates, T. O., Crowley, C. S., and Tanaka, S. (2010) Bacterial microcompartment organelles: protein shell structure and evolution. *Annu. Rev. Biophys.* **39**, 185–205
6. Cannon, G. C., Bradburne, C. E., Aldrich, H. C., Baker, S. H., Heinhorst, S., and Shively, J. M. (2001) Microcompartments in prokaryotes: carboxysomes and related polyhedra. *Appl. Environ. Microbiol.* **67**, 5351–5361
7. Cheng, S., Liu, Y., Crowley, C. S., Yeates, T. O., and Bobik, T. A. (2008) Bacterial microcompartments: their properties and paradoxes. *BioEssays News Rev. Mol. Cell. Dev. Biol.* **30**, 1084–1095
8. Price, G. D., Coleman, J. R., and Badger, M. R. (1992) Association of carbonic anhydrase activity with carboxysomes isolated from the cyanobacterium *Synechococcus* PCC7942. *Plant Physiol.* **100**, 784–793

9. Sampson, E. M., and Bobik, T. A. (2008) Microcompartments for B12-dependent 1,2-propanediol degradation provide protection from DNA and cellular damage by a reactive metabolic intermediate. *J. Bacteriol.* **190**, 2966–2971
10. Penrod, J. T., and Roth, J. R. (2006) Conserving a Volatile Metabolite: a Role for Carboxysome-Like Organelles in *Salmonella enterica*. *J. Bacteriol.* **188**, 2865–2874
11. Brinsmade, S. R., Paldon, T., and Escalante-Semerena, J. C. (2005) Minimal Functions and Physiological Conditions Required for Growth of *Salmonella enterica* on Ethanolamine in the Absence of the Metabolosome. *J. Bacteriol.* **187**, 8039–8046
12. Stojiljkovic, I., Bäumler, A., and Heffron, F. (1995) Ethanolamine utilization in *Salmonella typhimurium*: nucleotide sequence, protein expression, and mutational analysis of the cchA cchB eutE eutJ eutG eutH gene cluster. *J. Bacteriol.* **177**, 1357–66
13. Bobik, T., Xu, Y., Jeter, R., Otto, K., and Roth, J. (1997) Propanediol utilization genes (pdu) of *Salmonella typhimurium*: three genes for the propanediol dehydratase. *J. Bacteriol.* **21**, 6633–9
14. Jorda, J., Lopez, D., Wheatley, N. M., and Yeates, T. O. (2012) Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. *Protein Sci.* **22**, 179–95
15. Kerfeld, C. A., Sawaya, M. R., Tanaka, S., Nguyen, C. V., Phillips, M., Beeby, M., and Yeates, T. O. (2005) Protein structures forming the shell of primitive bacterial organelles. *Science* **309**, 936–938
16. Samborska, B., and Kimber, M. S. (2012) A dodecameric CcmK2 structure suggests β-carboxysomal shell facets have a double-layered organization. *Structure* **20**, 1353–1362

17. Tsai, Y., Sawaya, M. R., Cannon, G. C., Cai, F., Williams, E. B., Heinhorst, S., Kerfeld, C. A., and Yeates, T. O. (2007) Structural analysis of CsoS1A and the protein shell of the *Halothiobacillus neapolitanus* carboxysome. *PLoS Biol.* **5**, e144

18. Tanaka, S., Kerfeld, C. A., Sawaya, M. R., Cai, F., Heinhorst, S., Cannon, G. C., and Yeates, T. O. (2008) Atomic-level models of the bacterial carboxysome shell. *Science* **319**, 1083–1086

19. Tanaka, S., Sawaya, M. R., and Yeates, T. O. (2010) Structure and mechanisms of a protein-based organelle in *Escherichia coli*. *Science* **327**, 81–84

20. Price, G. D., Howitt, S. M., Harrison, K., and Badger, M. R. (1993) Analysis of a genomic DNA region from the cyanobacterium *Synechococcus sp. strain* PCC7942 involved in carboxysome assembly and function. *J. Bacteriol.* **175**, 2871–2879

21. Cheng, S., Sinha, S., Fan, C., Liu, Y., and Bobik, T. A. (2011) Genetic analysis of the protein shell of the microcompartments involved in coenzyme B12-dependent 1,2-propanediol degradation by *Salmonella*. *J. Bacteriol.* **193**, 1385–1392

22. Parsons, J. B., Dinesh, S. D., Deery, E., Leech, H. K., Brindley, A. A., Heldt, D., Frank, S., Smales, C. M., Lunsdorf, H., Rambach, A., Gass, M. H., Bleloch, A., McClean, K. J., Munro, A. W., Rigby, S. E. J., Warren, M. J., and Prentice, M. B. (2008) Biochemical and Structural Insights into Bacterial Organelle Form and Biogenesis. *J. Biol. Chem.* **283**, 14366–14375

23. Cai, F., Menon, B. B., Cannon, G. C., Curry, K. J., Shively, J. M., and 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

24. Schmid, M. F., Paredes, A. M., Khant, H. A., Soyer, F., Aldrich, H. C., Chiu, W., and Shively, J. M. (2006) Structure of *Halothiobacillus neapolitanus* carboxysomes by cryo-electron tomography. *J. Mol. Biol.* **364**, 526–535

25. Iancu, C. V., Ding, H. J., Morris, D. M., Dias, D. P., Gonzales, A. D., Martino, A., and Jensen, G. J. (2007) The structure of isolated *Synechococcus* strain WH8102 carboxysomes as revealed by electron cryotomography. *J. Mol. Biol.* **372**, 764–773

26. Lai, Y.-T., King, N. P., and Yeates, T. O. (2012) Principles for designing ordered protein assemblies. *Trends Cell Biol.* **22**, 653–661

27. Lai, Y.-T., Cascio, D., and Yeates, T. O. (2012) Structure of a 16-nm cage designed by using protein oligomers. *Science* **336**, 1129

28. Yang, Y., and Burkhard, P. (2012) Encapsulation of gold nanoparticles into self-assembling protein nanoparticles. *J. Nanobiotechnology* **10**, 42

29. King, N. P., Sheffler, W., Sawaya, M. R., Vollmar, B. S., Sumida, J. P., André, I., Gonen, T., Yeates, T. O., and Baker, D. (2012) Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **336**, 1171–1174

30. Wörsdörfer, B., Woycechowsky, K. J., and Hilvert, D. (2011) Directed evolution of a protein container. *Science* **331**, 589–592

31. Gandhi, C. S., Walton, T. A., and Rees, D. C. (2011) OCAM: A new tool for studying the oligomeric diversity of MscL channels. *Protein Sci.* **20**, 313–326

32. Gouaux, J., Braha, O., Hobaugh, M., Song, L., Cheley, S., Shustak, C., and Bayley, H. (1994) Subunit stoichiometry of staphylococcal alpha-hemolysin in crystals and on membranes: a heptameric transmembrane pore. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 12828–31

33. Hoover, D. M. (2002) DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.* **30**, 43e–43

34. Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., 3rd, and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345

35. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674

36. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132

37. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221

38. Headd, J. J., Echols, N., Afonine, P. V., Grosse-Kunstleve, R. W., Chen, V. B., Moriarty, N. W., Richardson, D. C., Richardson, J. S., and Adams, P. D. (2012) Use of knowledge-based restraints in phenix.refine to improve macromolecular refinement at low resolution. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 381–390

39. Sage, R. F. (2002) Variation in the k(cat) of Rubisco in C(3) and C(4) plants and some implications for photosynthetic performance at high and low temperature. *J. Exp. Bot.* **53**, 609–620

40. Marcus, Y., Altman-Gueta, H., Wolff, Y., and Gurevitz, M. (2011) Rubisco mutagenesis provides new insight into limitations on photosynthesis and growth in *Synechocystis* PCC6803. *J. Exp. Bot.* **62**, 4173–4182

41. Andersson, I. (2007) Catalysis and regulation in RuBisCO. *J. Exp. Bot.* **59**, 1555–1568

42. Satagopan, S., Scott, S. S., Smith, T. G., and Tabita, F. R. (2009) A RuBisCO Mutant That Confers Growth under a Normally "Inhibitory" Oxygen Concentration. *Biochemistry (Mosc.)* **48**, 9076–9083

43. Espie, G. S., and Kimber, M. S. (2011) Carboxysomes: cyanobacterial RuBisCO comes in small packages. *Photosynth. Res.* **109**, 7–20

44. Yeates, T. O., Kerfeld, C. A., Heinhorst, S., Cannon, G. C., and Shively, J. M. (2008) Protein-based organelles in bacteria: carboxysomes and related microcompartments. *Nat. Rev. Microbiol.* **6**, 681–691

45. Sutter, M., Wilson, S. C., Deutsch, S., and Kerfeld, C. A. (2013) Two new high-resolution crystal structures of carboxysome pentamer proteins reveal high structural conservation of CcmL orthologs among distantly related cyanobacterial species. *Photosynth. Res.* **118**, 9–16

46. Badger, M. R. (2003) CO2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. *J. Exp. Bot.* **54**, 609–622

47. Rae, B. D., Long, B. M., Badger, M. R., and Price, G. D. (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. MMBR* **77**, 357–379

48. Onizuka, T., Endo, S., Akiyama, H., Kanai, S., Hirano, M., Yokota, A., Tanaka, S., and Miyasaka, H. (2004) The rbcX gene product promotes the production and assembly of ribulose-1,5-bisphosphate carboxylase/oxygenase of *Synechococcus* sp. PCC7002 in *Escherichia coli*. *Plant Cell Physiol.* **45**, 1390–1395

49. Liu, C., Young, A. L., Starling-Windhof, A., Bracher, A., Saschenbrecker, S., Rao, B. V., Rao, K. V., Berninghausen, O., Mielke, T., Hartl, F. U., Beckmann, R., and Hayer-Hartl, M. (2010) Coupled chaperone action in folding and assembly of hexadecameric RuBisCO. *Nature* **463**, 197–202

50. Bracher, A., Starling-Windhof, A., Hartl, F. U., and Hayer-Hartl, M. (2011) Crystal structure of a chaperone-bound assembly intermediate of form I RuBisCO. *Nat. Struct. Mol. Biol.* **18**, 875–880

51. Mueller-Cajar, O., Stotz, M., Wendler, P., Hartl, F. U., Bracher, A., and Hayer-Hartl, M. (2011) Structure and function of the AAA+ protein CbbX, a red-type Rubisco activase. *Nature* **479**, 194–199

52. Hayashi, N. R., Arai, H., Kodama, T., and Igarashi, Y. (1997) The novel genes, cbbQ and cbbO, located downstream from the RubisCO genes of Pseudomonas hydrogenothermophila, affect the conformational states and activity of RubisCO. *Biochem. Biophys. Res. Commun.* **241**, 565–569

53. Chen, D.-H., Madan, D., Weaver, J., Lin, Z., Schröder, G. F., Chiu, W., and Rye, H. S. (2013) Visualizing GroEL/ES in the act of encapsulating a folding protein. *Cell* **153**, 1354–1365

54. Ellis, R. J. (2001) Molecular chaperones: inside and outside the Anfinsen cage. *Curr. Biol. CB* **11**, R1038–1040

55. Feiz, L., Williams-Carrier, R., Wostrikoff, K., Belcher, S., Barkan, A., and Stern, D. B. (2012) Ribulose-1,5-bis-phosphate carboxylase/oxygenase accumulation factor1 is required for holoenzyme assembly in maize. *Plant Cell* **24**, 3435–3446

56. Roberts, E. W., Cai, F., Kerfeld, C. A., Cannon, G. C., and Heinhorst, S. (2011) Isolation and Characterization of the *Prochlorococcus* Carboxysome Reveal the Presence of the Novel Shell Protein CsoS1D. *J. Bacteriol.* **194**, 787–795

57. Thöny, B., Auerbach, G., and Blau, N. (2000) Tetrahydrobiopterin biosynthesis, regeneration and functions. *Biochem. J.* **347 Pt 1**, 1–16

58. Naponelli, V., Noiriel, A., Ziemak, M. J., Beverley, S. M., Lye, L.-F., Plume, A. M., Botella, J. R., Loizeau, K., Ravanel, S., Rébeillé, F., de Crécy-Lagard, V., and Hanson, A. D. (2008) Phylogenomic and functional analysis of pterin-4a-carbinolamine dehydratase family (COG2154) proteins in plants and microorganisms. *Plant Physiol.* **146**, 1515–1527

59. Volner, A., Zoidakis, J., and Abu-Omar, M. M. (2003) Order of substrate binding in bacterial phenylalanine hydroxylase and its mechanistic implication for pterin-dependent oxygenases. *J. Biol. Inorg. Chem.* **8**, 121–128

60. Cronk, J. D., Endrizzi, J. A., and Alber, T. (1996) High-resolution structures of the bifunctional enzyme and transcriptional coactivator DCoH and its complex with a product analogue. *Protein Sci. Publ. Protein Soc.* **5**, 1963–1972

61. Sourdive, D. J., Transy, C., Garbay, S., and Yaniv, M. (1997) The bifunctional DCOH protein binds to HNF1 independently of its 4-alpha-carbinolamine dehydratase activity. *Nucleic Acids Res.* **25**, 1476–1484

62. Pape, T., and Schneider, T. R. (2004) HKL2MAP : a graphical user interface for macromolecular phasing with SHELX programs. *J. Appl. Crystallogr.* **37**, 843–844

63. Cowtan, K. (2006) The *Buccaneer* software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011

64. Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., Smart, O. S., Vonrhein, C., and Womack, T. O. (2011) BUSTER version X.Y.Z. Cambridge, United Kingdom: Global Phasing Ltd.

65. Starkenburg, S. R., Chain, P. S. G., Sayavedra-Soto, L. A., Hauser, L., Land, M. L., Larimer, F. W., Malfatti, S. A., Klotz, M. G., Bottomley, P. J., Arp, D. J., and Hickey, W. J. (2006) Genome sequence of the chemolithoautotrophic nitrite-oxidizing bacterium *Nitrobacter winogradskyi* Nb-255. *Appl. Environ. Microbiol.* **72**, 2050–2063

66. Yeates, T. O. (1995) Algorithms for evaluating the long-range accessibility of protein surfaces. *J. Mol. Biol.* **249**, 804–815

67. Simossis, V. A., and Heringa, J. (2005) PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic Acids Res.* **33**, W289–W294

68. So, A. K.-C., Espie, G. S., Williams, E. B., Shively, J. M., Heinhorst, S., and Cannon, G. C. (2004) A Novel Evolutionary Lineage of Carbonic Anhydrase (Class) Is a Component of the Carboxysome Shell. *J. Bacteriol.* **186**, 623–630

69. Piatigorsky, J. (2003) Crystallin genes: specialization by changes in gene regulation may precede gene duplication. *J. Struct. Funct. Genomics* **3**, 131–137

70. Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002) The protein kinase complement of the human genome. *Science* **298**, 1912–1934

71. Zeqiraj, E., Filippi, B. M., Deak, M., Alessi, D. R., and van Aalten, D. M. F. (2009) Structure of the LKB1-STRAD-MO25 complex reveals an allosteric mechanism of kinase activation. *Science* **326**, 1707–1711

72. Leslie, M. (2013) "Dead" Enzymes Show Signs of Life. *Science* **340**, 25–27

73. Jeffery, C. J. (2004) Molecular mechanisms for multitasking: recent crystal structures of moonlighting proteins. *Curr. Opin. Struct. Biol.* **14**, 663–668

74. Ninfa, A. J., and Atkinson, M. R. (2000) PII signal transduction proteins. *Trends Microbiol.* **8**, 172–179

75. Adler, S. P., Purich, D., and Stadtman, E. R. (1975) Cascade control of Escherichia coli glutamine synthetase. Properties of the PII regulatory protein and the uridylyltransferase-uridylyl-removing enzyme. *J. Biol. Chem.* **250**, 6264–6272

76. Dodsworth, J. A., Cady, N. C., and Leigh, J. A. (2005) 2-Oxoglutarate and the PII homologues NifI1 and NifI2 regulate nitrogenase activity in cell extracts of Methanococcus maripaludis. *Mol. Microbiol.* **56**, 1527–1538

77. Llácer, J. L., Contreras, A., Forchhammer, K., Marco-Marín, C., Gil-Ortiz, F., Maldonado, R., Fita, I., and Rubio, V. (2007) The crystal structure of the complex of PII and acetylglutamate kinase reveals how PII controls the storage of nitrogen as arginine. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 17644–17649

78. Coutts, G. (2002) Membrane sequestration of the signal transduction protein GlnK by the ammonium transporter AmtB. *EMBO J.* **21**, 536–545

79. Radchenko, M. V., Thornton, J., and Merrick, M. (2013) PII signal transduction proteins are ATPases whose activity is regulated by 2-oxoglutarate. *Proc. Natl. Acad. Sci. U. S. A.*

80. Goody, R. S. (2003) The significance of the free energy of hydrolysis of GTP for signal-transducing and regulatory GTPases. *Biophys. Chem.* **100**, 535–544

81. Forchhammer, K. (2010) The network of P(II) signalling protein interactions in unicellular cyanobacteria. *Adv. Exp. Med. Biol.* **675**, 71–90

82. Fokina, O., Chellamuthu, V.-R., Forchhammer, K., and Zeth, K. (2010) Mechanism of 2-oxoglutarate signaling by the Synechococcus elongatus PII signal transduction protein. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 19760–19765

83. Truan, D., Bjelić, S., Li, X.-D., and Winkler, F. K. (2014) Structure and Thermodynamics of Effector Molecule Binding to the Nitrogen Signal Transduction PII Protein GlnZ from Azospirillum brasilense. *J. Mol. Biol.* **426**, 2783–2799

84. Conroy, M. J., Durand, A., Lupo, D., Li, X.-D., Bullough, P. A., Winkler, F. K., and Merrick, M. (2007) The crystal structure of the Escherichia coli AmtB-GlnK complex reveals how GlnK regulates the ammonia channel. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 1213–1218

85. Feria Bourrellier, A. B., Valot, B., Guillot, A., Ambard-Bretteville, F., Vidal, J., and Hodges, M. (2010) Chloroplast acetyl-CoA carboxylase activity is 2-oxoglutarate-regulated by interaction of PII with the biotin carboxyl carrier subunit. *Proc. Natl. Acad. Sci.* **107**, 502–507

86. Hisbergues, M., Jeanjean, R., Joset, F., Tandeau de Marsac, N., and Bédu, S. (1999) Protein PII regulates both inorganic carbon and nitrate uptake and is modified by a redox signal in synechocystis PCC 6803. *FEBS Lett.* **463**, 216–220

87. Chellamuthu, V. R., Alva, V., and Forchhammer, K. (2012) From cyanobacteria to plants: conservation of PII functions during plastid evolution. *Planta* **237**, 451–462

88. Daley, S. M. E., Kappell, A. D., Carrick, M. J., and Burnap, R. L. (2012) Regulation of the Cyanobacterial CO2-Concentrating Mechanism Involves Internal Sensing of NADP+ and α-Ketogutarate Levels by Transcription Factor CcmR. *PLoS ONE* **7**, e41286

89. Wang, H.-L., Postier, B. L., and Burnap, R. L. (2004) Alterations in global patterns of gene expression in Synechocystis sp. PCC 6803 in response to inorganic carbon limitation and the inactivation of ndhR, a LysR family regulator. *J. Biol. Chem.* **279**, 5739–5751

90. Price, G. D. (2011) Inorganic carbon transporters of the cyanobacterial CO2 concentrating mechanism. *Photosynth. Res.* **109**, 47–57

91. Klughammer, B., Sultemeyer, D., Badger, M. R., and Price, G. D. (1999) 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. *Mol. Microbiol.* **32**, 1305–1315

92. Zhang, P. (2004) Expression and Functional Roles of the Two Distinct NDH-1 Complexes and the Carbon Acquisition Complex NdhD3/NdhF3/CupA/Sll1735 in Synechocystis sp PCC 6803. *PLANT CELL ONLINE* **16**, 3326–3340

93. Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H.-Y., Cohoon, M., de Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E. D., Gerdes, S., Glass, E. M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L., Kubal, M., Larsen, N., Linke, B., McHardy, A. C., Meyer, F., Neuweger, H., Olsen, G., Olson, R., Osterman, A., Portnoy, V., Pusch, G. D., Rodionov, D. A., Rückert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye, Y., Zagnitko, O., and Vonstein, V. (2005) The

subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* **33**, 5691–5702

94. Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M., Bork, P., and von Mering, C. (2009) STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* **37**, D412–416
95. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., and Higgins, D. G. (2014) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539–539
96. Sant'Anna, F. H., Trentini, D. B., de Souto Weber, S., Cecagno, R., da Silva, S. C., and Schrank, I. S. (2009) The PII superfamily revised: a novel group and evolutionary insights. *J. Mol. Evol.* **68**, 322–336
97. Huergo, L. F., Chandra, G., and Merrick, M. (2013) P(II) signal transduction proteins: nitrogen regulation and beyond. *FEMS Microbiol. Rev.* **37**, 251–283
98. Tanaka, Y., Tsumoto, K., Nakanishi, T., Yasutake, Y., Sakai, N., Yao, M., Tanaka, I., and Kumagai, I. (2004) Structural implications for heavy metal-induced reversible assembly and aggregation of a protein: the case of Pyrococcus horikoshii CutA. *FEBS Lett.* **556**, 167–174
99. Maeda, S., Badger, M. R., and Price, G. D. (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

100. Thompson, M. C., Wheatley, N. M., Jorda, J., Sawaya, M. R., Gidaniyan, S. D., Ahmed, H., Yang, Z., McCarty, K. N., Whitelegge, J. P., and Yeates, T. O. (2014) Identification of a unique fe-s cluster binding site in a glycyl-radical type microcompartment shell protein. *J. Mol. Biol.* **426**, 3287–3304

101. Crowley, C. S., Cascio, D., Sawaya, M. R., Kopstein, J. S., Bobik, T. A., and Yeates, T. O. (2010) Structural insight into the mechanisms of transport across the Salmonella enterica Pdu microcompartment shell. *J. Biol. Chem.* **285**, 37838–37846