The Biochemical and Structural Basis of Get3d's Role in Photosynthesis

Thesis by
Alexandra N. Barlow

In Partial Fulfillment of the Requirements for the
Degree of
Doctorate of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2025
Defended May 21, 2025

Alexandra N. Barlow
ORCID: 0000-0001-6515-2396

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to the many individuals who have supported me throughout my Ph.D. journey and made this dissertation possible.

First and foremost, I extend my sincere thanks to my advisor, Dr. Bil Clemons, for his mentorship, for teaching me how to approach scientific problems thoughtfully and creatively, and for fostering a supportive and inclusive environment in which I could grow as both a scientist and a person. I am also grateful to Dr. Doug Rees for his steady support, thoughtful feedback, and example as a scientist and mentor.

I am fortunate to have been advised by a thoughtful and generous thesis committee—Drs. Rebecca Voorhees, Shu-ou Shan, and Dianne Newman—whose insights, feedback, and support have been invaluable throughout this process.

I would like to acknowledge all members of the Clemons and Rees labs, past and present, who have contributed to this work and helped shape the scientist I am today. I thank Allen Lee, Ann Mao, and Courtney Oaida for their tireless efforts in keeping the lab running smoothly and supporting our research behind the scenes.

I extend my sincere appreciation to our collaborators. Drs. Suresh Ramasamy and M. S. Manu provided essential support in launching this project. I am especially grateful to Dr. Paul T. Tarr for his time, guidance, and insights into the plant-focused aspects of this work. I also thank Drs. Art Laganowsky and Yun Zhu for their technical expertise in native mass spectrometry.

A special thank you to the GET team—Victor Garcia, Shivansh Mahajan, Juliet Lee, Conner Wells, and Storm Robertson—for their valuable feedback, collaborative spirit, and camaraderie throughout this project.

I am especially grateful to Dr. Karen Orta and Victor Garcia for walking this path with me and offering their unwavering support and friendship. I also thank Drs. Michelle Fry, Shyam Saladi, and Nadia Riera for their mentorship and encouragement. I am thankful to Yusrah Kaudeer for bringing joy and warmth to our lab bay, and to Dr. Ailiena Maggiolo for her expertise and friendship and for being a wonderful travel companion in Erice.

I would also like to acknowledge the undergraduate students I have had the honor to get to know including my WAVE mentee, Elizabeth Nelson, as well as Lily

DeBell, Vanessa Mechem, and Helen Brackney, for their enthusiasm, curiosity, and contributions.

This work was supported by funding from the National Institutes of Health, the National Science Foundation, and the Caltech Center for Evolutionary Science. I am also grateful to the University of Massachusetts Chemistry and Biochemistry Departments, and to my undergraduate research mentors, Dr. Mike Knapp and Dr. Vanessa Chaplin Momaney, for helping to establish my foundation in research. A special thanks to Dr. Margaret Stratton for introducing me to my advisor, Dr. Clemons.

To my friends—including those in the recovery community, my Dungeons & Dragons group, and my non-scientist friends—thank you for your grounding presence and continued encouragement. To my family, both biological and chosen: thank you for your love, patience, and unwavering belief in me. A heartfelt thank you to Clay, Julia, and our feline companions for their constant love and support.

Finally, I honor the memory of Dr. Jim Howard, whose excitement for science was contagious, and Dr. Jeff Cameron, who generously shared his time, knowledge, and passion for cyanobacteria. Their legacies continue to inspire me.

I could not have done this without all of you.

ABSTRACT

Tail-anchored (TA) membrane proteins, defined by a single C-terminal transmembrane domain, are inserted into the endoplasmic reticulum (ER) membrane via the guided entry of tail-anchored proteins pathway. The central targeting factor of this pathway is Get3, an ATPase that receives TA clients from upstream chaperones and mediates their delivery to the ER. Here, we identify and characterize a unique Get3 homolog, termed Get3d, distinguished by a C-terminal α-crystallin domain (αCD). We show that Get3d is conserved across plants and photosynthetic bacteria and demonstrate that it localizes to the chloroplast in plants. We present the X-ray crystal structure of Get3d, revealing unique features including the αCD and a client-binding chamber in the closed state. Biochemical analyses confirm that Get3d is an active ATPase capable of binding TA proteins *in vitro*. To investigate its physiological role, we identified the plant-like Get3d homolog in *Synechocystis* sp. PCC 6803 and generated deletion and complementation strains. Loss of Get3d impairs cell growth and pigment production, and proteomic analyses reveal widespread dysregulation, including up-regulation of transcriptional regulators and down-regulation of redox-associated proteins—suggesting a role in redox homeostasis. Complementation studies show that ATPase activity is necessary for restoring the expression of key photosynthesis-related proteins, while the αCD is critical for maintaining Get3d protein stability *in vivo*. Finally, co-immunoprecipitation coupled to mass spectrometry identifies putative Get3d interaction partners enriched in membrane-associated and photosynthetic proteins. Together, these findings establish Get3d as a biochemically distinct and functionally essential member of the Get3 family, with a potential role in redox regulation and photosynthetic homeostasis in diverse photosynthetic organisms.

PUBLISHED CONTENT AND CONTRIBUTIONS

(*1*) Barlow, A. N., Manu, M. S., Saladi, S. M., Tarr, P. T., Yadav, Y., Thinn, A. M. M., Zhu, Y., Laganowsky, A. D., Clemons Jr., W. M., and Ramasamy, S. (2023). Structures of Get3d reveal a distinct architecture associated with the emergence of photosynthesis. *J. Biol. Chem. 299*, DOI: 10.1016/j.jbc.2023.104752,

Contributions: Alexandra N. Barlow wrote the manuscript, prepared figures, performed cloning, expressed and purified proteins for biochemical assays, performed ATPase assays and tail-anchored protein capture assays, and prepared samples for mass spectrometry.

TABLE OF CONTENTS

Acknowledgements . . . iii
Abstract . . . v
Published Content and Contributions . . . vi
Table of Contents . . . vi
List of Illustrations . . . ix
List of Tables . . . xi
Nomenclature . . . xii
Chapter I: Introduction . . . 1
1.1 Photosynthesis . . . 1
1.2 Endosymbiosis . . . 1
1.3 Chloroplast and Cyanobacterial Membranes . . . 1
1.4 Membrane Proteins . . . 2
1.5 Types of Integral Membrane Proteins . . . 2
1.6 Co-Translational Membrane Protein Targeting . . . 3
1.7 Tail-Anchored Membrane Proteins . . . 4
1.8 Post-Translational Membrane Protein Targeting: Guided Entry of Tail-Anchored Proteins Pathway . . . 4
1.9 Tail-Anchored Protein Targeting Factor Get3 . . . 7
1.10 Post-Translational Membrane Protein Targeting: Alternative Pathways 7
1.11 Protein Targeting in Cyanobacteria . . . 8
1.12 Protein Targeting into and in Chloroplasts . . . 8
1.13 Tail-Anchored Proteins and Their Targeting in Plants . . . 9
1.14 Get3 Homolog Conserved in Photosynthetic Organisms – Get3d . . . 10
Chapter II: The Structure of the Conserved Get3d . . . 17
2.1 Abstract . . . 18
2.2 Introduction . . . 19
2.3 Results . . . 21
2.4 Discussion . . . 34
2.5 Experimental Procedures . . . 38
2.6 Acknowledgments . . . 44
2.7 Supporting Information . . . 54
Chapter III: Characterization of the Plant-Like Get3d Homolog in Cyanobacteria 68
3.1 Abstract . . . 69
3.2 Introduction . . . 70
3.3 Results . . . 71
3.4 Discussion . . . 86
3.5 Methods . . . 91
3.6 Supporting Information . . . 100

Chapter IV: Conclusions and Perspectives on Get3d Function in Photosynthetic Organisms 116
4.1 Conclusions 117
4.2 Limitations 118
4.3 Future Directions 119

LIST OF ILLUSTRATIONS

Number		*Page*
1.1	Integral membrane proteins.	3
1.2	Membrane protein targeting pathways.	5
1.3	Catalytic cycle of the guided entry of tail-anchored proteins pathway.	6
1.4	Guided entry of tail-anchored proteins pathway in plants.	10
2.1	Identification and features of Get3 homologs.	22
2.2	Cellular localization of *At*Get3d.	25
2.3	Structures of *At*Get3d and *Nos*Get3d.	26
2.4	Structural analysis of the α-crystallin domain of Get3d.	28
2.5	ATPase activity of Get3d.	30
2.6	Hydrophobic chamber of Get3d.	33
S2.1	Comprehensive phylogeny of Get3 homologs.	55
S2.2	Detailed phylogenetic analysis of Get3 homologs.	56
S2.3	Detailed cladogram of Get3 homologs in plants.	58
S2.4	Features of the *At*Get3d structure.	59
S2.5	View of the missing A-loop and CXXC motif.	60
S2.6	Possible lipids in the hydrophobic chamber of Get3d.	61
S2.7	New refinement of *Nos*Get3d.	62
S2.8	Sequence alignment of selected α-crystallin domains.	63
S2.9	Interactions at the α-crystallin domain interface.	64
S2.10	Additional features of the hydrophobic chamber of Get3d.	65
3.1	Features of Get3d homologs.	73
3.2	Structural comparison of Get3d homologs.	74
3.3	Volcano plots of differential protein abundance in whole-cell proteomes.	79
3.4	Spot growth assay of *slr1794* knockout and complementation strains.	80
3.5	Pigment content of *slr1794* knockout and complementation strains.	82
3.6	Identification of putative interaction partners of Slr1794.	82
S3.1	Validation of *Synechocystis* sp. PCC 6803 transformation.	104
S3.2	Predicted local distance difference test for structural predictions of Get3d.	104
S3.3	Sequence features of cyanobacterial Get3d paralogs.	105
S3.4	Structural predictions of Slr1794 with putative interaction partners.	109

S3.5 Predicted local distance difference test for structural predictions of Slr1794 with putative interaction partners. 110
S3.6 Principal component analysis of whole-cell proteomes. 115

LIST OF TABLES

Number *Page*

S2.1 Comparison of data collection and refinement statistics for *At*Get3d and *Nos*Get3d crystal structures. 57
S2.2 Constructs and primers used in this study. 66
3.1 Expression of Slr1794 across strains. 76
3.2 Photosynthesis-related proteins down-regulated in the ATPase-deficient complementation strain. 79
3.3 Proteins significantly enriched in the Slr1794-3xFLAG sample. . . . 83
3.4 Membrane proteins significantly enriched in the Slr1794-3xFLAG sample. 87
3.5 Strains used in this study. 92
S3.1 Tail-anchored proteins in *Synechocystis* sp. PCC 6803. 101
S3.2 Absorbance measurements of pigment extraction. 106
S3.3 Tukey's honestly significant difference post hoc test for chlorophyll *a* concentrations. 107
S3.4 Tukey's honestly significant difference post hoc test for carotenoid concentrations. 108
S3.5 Primer sequences used in this study. 108
S3.6 Whole-cell mass spectrometry protein abundance and metadata. . . . 108
S3.7 Differential protein abundance comparisons in whole-cell proteomic analysis. 111
S3.8 Functional enrichment analysis of annotation clusters derived from proteins significantly up- or down-regulated across strains. 111
S3.9 Protein abundance, enrichment, and metadata for the FLAG co-immunoprecipitation coupled to mass spectrometry. 111
S3.10 Functional enrichment analysis of annotation clusters derived from proteins significantly enriched in the Slr1794-3xFLAG sample. . . . 112

NOMENCLATURE

αCD. α-crystallin domain.

***A. thaliana*.** *Arabidopsis thaliana*.

***Nostoc* sp.** *Nostoc* sp. PCC 7120.

***Synechocystis* sp.** *Synechocystis* sp. PCC 6803.

ANOVA. Analysis of variance.

CBD. Client binding domain.

Co-IP. Co-immunoprecipitation.

cpSRP. Chloroplast signal recognition particle.

EMC. ER-membrane protein complex.

ER. Endoplasmic reticulum.

FDR. False discovery rate.

GET. Guided entry of tail-anchored proteins.

GFP. Green fluorescent protein.

HCD. Higher-energy collisional dissociation.

HSD. Honestly significant difference.

IMP. Integral membrane protein.

MAFFT. Multiple alignment using fast Fourier transform.

MBP. Maltose binding protein.

MS. Mass spectrometry.

NBD. Nucleotide binding domain.

NTPase. Nucleoside triphosphatase.

PCA. Principal component analysis.

PDB. Protein Data Bank.

PNP. Purine nucleoside phosphorylase.

PSI. Photosystem I.

PSII. Photosystem II.

RMSD. Root mean square deviation.

SA protein. Signal-anchored protein.

Sec. General secretory.

SecE1. Sec translocon subunit E of the thylakoid membrane.

SecE2. Sec translocon subunit E of the inner envelope membrane.

sHSP. Small heat shock protein.

SND. SRP-independent targeting.

SRP. Signal recognition particle.

TA protein. Tail-anchored protein.

TAT. Twin-arginine translocase.

TIC. Translocon of the inner chloroplast envelope.

TMD. Transmembrane domain.

TOC. Translocon of the outer chloroplast envelope.

TP. Transit peptide.

Chapter 1

INTRODUCTION

1.1 Photosynthesis

Photosynthesis, the process by which plants and other organisms convert light energy into chemical energy, is arguably the most important biological process on the planet (reviewed in (*1*)). The chemical energy produced during photosynthesis serves as the basis for the entire food chain, and the oxygen produced is the primary source of atmospheric oxygen.

Oxygenic photosynthesis, performed by a wide range of organisms including cyanobacteria, algae, and plants, is a complex series of chemical reactions that produce molecular oxygen and energy (in the form of sugars) from carbon dioxide, water, and light energy (reviewed in (*2*)). Anoxygenic photosynthesis, performed by some bacteria and archaea, is a specialized form of photosynthesis that differs in the reductant used and byproducts produced (reviewed in (*3*)). Understanding the molecular mechanisms involved in photosynthesis and its regulation is vital to comprehending this essential biological process.

1.2 Endosymbiosis

Approximately one billion years ago, a cyanobacteria-like cell was internalized by an early single-celled eukaryote in a process called primary endosymbiosis (reviewed in (*2*)). This internalized cell was ultimately established as the chloroplast, the photosynthetic organelle in plants. Due to their close evolutionary connection, the chloroplast closely resembles cyanobacteria in both structure and function.

During the evolution of the chloroplast, most genes essential for photosynthesis were transferred from the organelle to the host's nuclear genome via horizontal gene transfer (*4*). Chloroplasts retained a small, circular genome encoding ~120 genes, primarily involved in photosynthesis, transcription, and translation.

1.3 Chloroplast and Cyanobacterial Membranes

Cyanobacteria are Gram-negative bacteria characterized by a double-membrane cell envelope consisting of the plasma (or cell) membrane and the outer membrane. The space between these membranes, known as the periplasm, contains a dense and

highly crosslinked peptidoglycan layer that provides structural integrity to the cell (reviewed in (*5*)). Inside the cell lies the thylakoid membrane, the site of photosynthesis, which separates the cytosol from the enclosed thylakoid lumen. Thylakoid membrane architecture is often intricate and displays considerable variation among cyanobacterial species (*6*).

Given that plant chloroplasts originated through the endosymbiosis of a cyanobacteria-like ancestor, it is unsurprising that chloroplast membranes share structural and functional similarities with those of cyanobacteria (*7*). Like mitochondria—which also arose via the endosymbiosis of a bacterial progenitor (*8*)—chloroplasts are double-membrane organelles. They possess an inner envelope membrane and an outer envelope membrane separated by an intermembrane space, analogous to the cyanobacterial cell wall. Chloroplasts also contain thylakoid membranes that organize into stacked disc-shaped structures known as grana (*7*). The thylakoid membrane separates the stroma, the equivalent of the cyanobacterial cytosol, and the enclosed thylakoid lumen.

1.4 Membrane Proteins

The correct targeting of proteins to and across membranes is essential for maintaining cellular homeostasis. The mislocalization of proteins can lead to severe cellular defects and diseases, including neurological disorders and cancers (reviewed in (*9*)). Therefore, a comprehensive understanding of protein targeting is critical for elucidating cellular functions. The targeting of integral membrane proteins (IMPs) is of special interest, as they constitute approximately 30% of the human proteome and serve as targets for nearly 60% of clinical drugs (reviewed in (*10*)). IMPs perform diverse cellular functions, including signal transduction, ion transport, and cell adhesion (reviewed in (*11*)), and are present in every membrane in the cell, including the plasma membrane and the membranes of eukaryotic organelles.

1.5 Types of Integral Membrane Proteins

There are three primary classes of IMPs: monotopic, single-pass (bitopic), and multi-pass (polytopic) (Fig. 1.1) (reviewed in (*12*)). Monotopic IMPs, which make up a small portion of IMPs, associate with a single face of the membrane. Single-pass IMPs have a single membrane-spanning α-helix. Polytopic IMPs are either the highly abundant α-helical bundle proteins or the less abundant β-barrel proteins primarily found in bacterial outer membranes. Most IMPs fall into the category of single-pass and α-helical bundle proteins (together called α-helical IMPs).

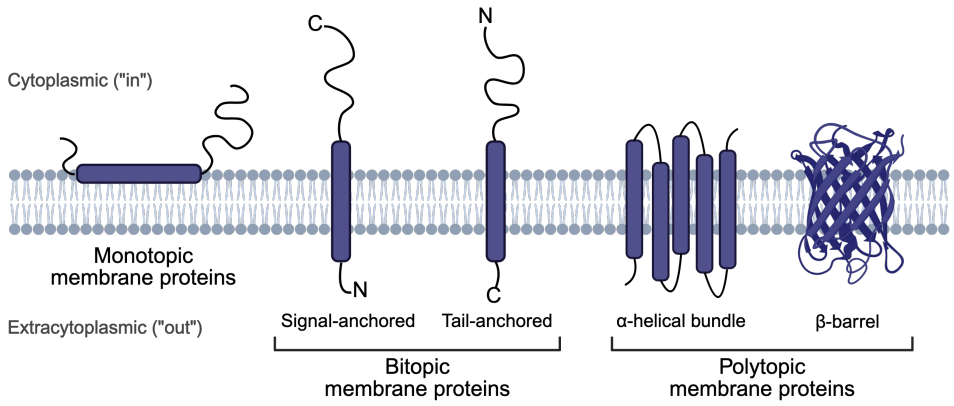


Figure 1.1: Integral membrane proteins. Monotopic, bitopic (signal-anchored and tail-anchored), and polytopic (α-helical bundle and β-barrel) membrane proteins are shown as cartoons. α-helices are depicted as cylinders and β-strands are depicted as arrows. The lipid bilayer is shown, and the cytoplasmic ("in") and extracytoplasmic ("out") sides of the membrane are labeled. The N- and C-termini of the bitopic membrane proteins are shown. Key as follows: N, N-terminus; C, C-terminus. Figure created in BioRender.com.

α-helical IMPs contain short (~18 amino acids) transmembrane α-helices termed transmembrane domains (TMDs). These highly hydrophobic TMDs are not stable in the hydrophilic environment of the cytosol where they are synthesized, presenting a challenge for the cell. Thus, special pathways have been developed to stabilize these proteins until they are targeted to and integrated into their destination membranes.

In addition to being targeted to the correct membrane, the correct topology (i.e., orientation) of an IMP is also essential for its proper function (reviewed in (*12*)). Topology is typically defined by the positioning of the protein's N- and C-termini, where "in" refers to the cytoplasmic side of the membrane and "out" denotes the extracytoplasmic side. This orientation is primarily determined during membrane integration.

1.6 Co-Translational Membrane Protein Targeting

For most α-helical IMPs, the first TMD functions as a targeting signal, directing the protein to its specific destination membrane (reviewed in (*13*, *14*)). IMPs with a single α-helix near the N-terminus (known as signal-anchored [SA] proteins) as well as multi-pass α-helical bundle proteins are primarily targeted via the universal signal recognition particle (SRP) pathway (Fig. 1.2) (reviewed in (*15*)). In eukaryotes, the SRP directs proteins to the endoplasmic reticulum (ER) membrane, while in prokaryotes, it targets them to the plasma membrane.

The SRP binds to both the N-terminal TMD of the nascent polypeptide and the

ribosome, forming the ribosome-nascent chain complex. This complex is delivered to the SRP receptor at the ER membrane. Upon arrival, the nascent polypeptide is transferred to the general secretory (Sec) translocon, enabling continued translation and integration of the membrane protein into the membrane.

1.7 Tail-Anchored Membrane Proteins

Tail-anchored (TA) membrane proteins, or TA proteins, are a distinct class of membrane proteins characterized by a single TMD located within ~30 residues of the C-terminus (*16*). They adopt an "N-in" topology, with the N-terminus positioned in the cytosol or its equivalent (e.g., the chloroplast stroma or mitochondrial matrix).

TA proteins comprise ~3–5% of membrane proteins and play critical roles in various cellular processes, including protein translocation, vesicle transport, protein quality control, and apoptosis (reviewed in (*13*, *17*, *18*)). Notably, many proteins involved in photosynthesis, including the PSI-associated linker protein CpcL (*19*) and Photosystem II reaction center protein H (*20*) are TA proteins. Thus, understanding the mechanisms governing TA protein targeting is essential. Unlike most IMPs, TA proteins are not candidates for co-translational targeting via the SRP pathway. Because their TMD remains sequestered in the ribosome exit tunnel until translation is complete, they must rely on alternative post-translational targeting pathways.

1.8 Post-Translational Membrane Protein Targeting: Guided Entry of Tail-Anchored Proteins Pathway

ER-destined TA proteins are delivered post-translationally primarily via the guided entry of tail-anchored proteins (GET) pathway (Fig. 1.2) (reviewed in (*21*)). This pathway has been studied in detail, particularly in opisthokonts (i.e., animals and fungi). The yeast GET system consists of six proteins: the chaperone Sgt2 and Get1–5.

Upon release from the ribosome, the TA protein is captured by Sgt2, with assistance from other chaperones, which shields the hydrophobic TMD from the cytosol. Sgt2 then interacts with Get3/4/5, forming a pre-targeting complex. The TA protein is subsequently transferred to the homodimeric targeting factor Get3. Following ATP hydrolysis and the dissociation of Get4/5, Get3 delivers the TA protein to the ER membrane complex Get1/2, which functions as an insertase, embedding the TMD into the lipid bilayer (Fig. 1.3).

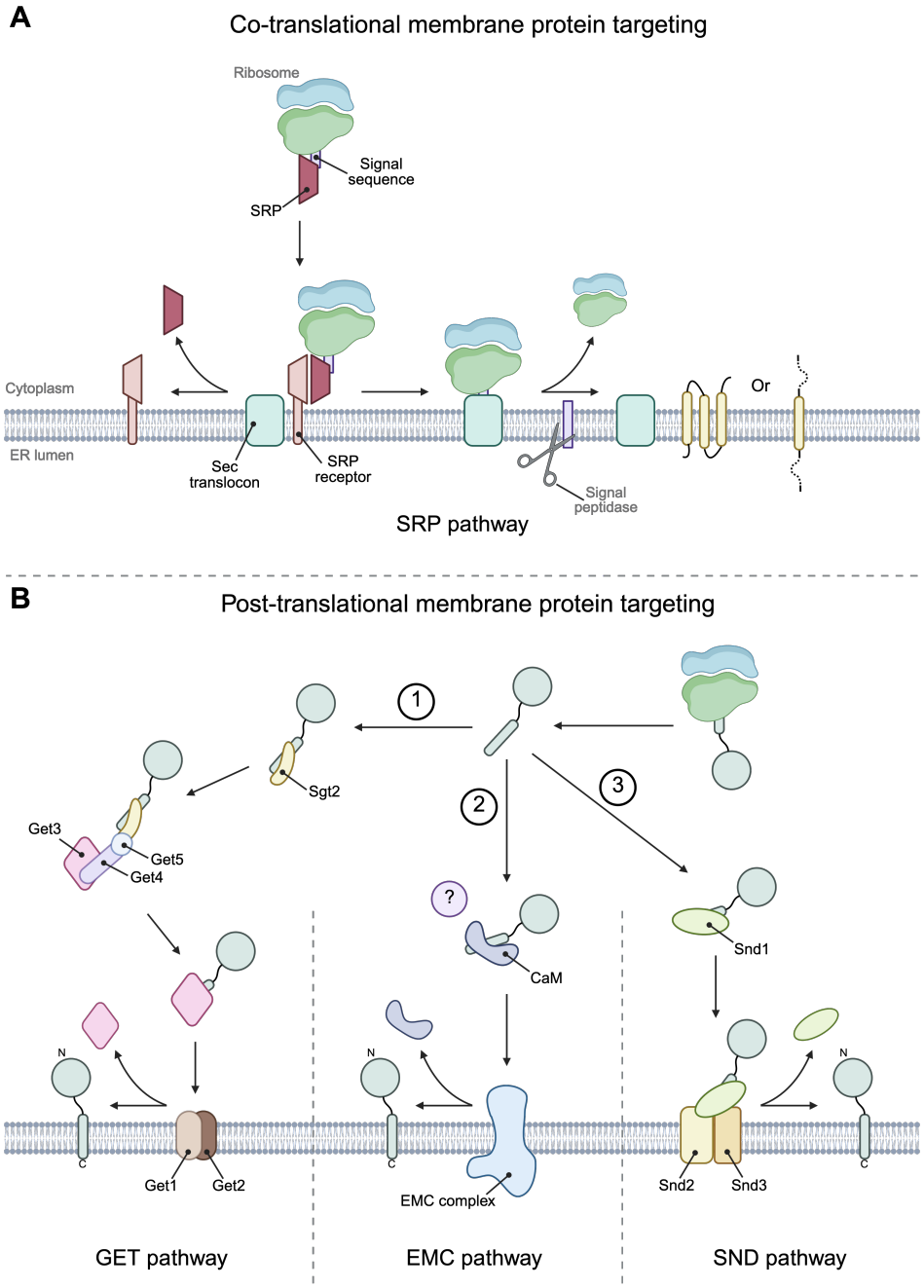


Figure 1.2: Membrane protein targeting pathways. Cartoon representation of (A) co-translational membrane protein targeting (SRP pathway) and (B) post-translational protein targeting ([1] GET, [2] EMC, and [3] SND pathways). The N- and C-termini of selected membrane proteins are shown. The lipid bilayer is shown, and the cytoplasm and ER lumen are labeled. Key as follows: ER, endoplasmic reticulum; SRP, signal recognition particle; Sec, general secretory; GET, guided entry of tail-anchored proteins; EMC, ER membrane protein complex; CaM, calmodulin; ?, unknown protein(s); SND, SRP-independent; N, N-terminus; C, C-terminus. Figure created in BioRender.com.

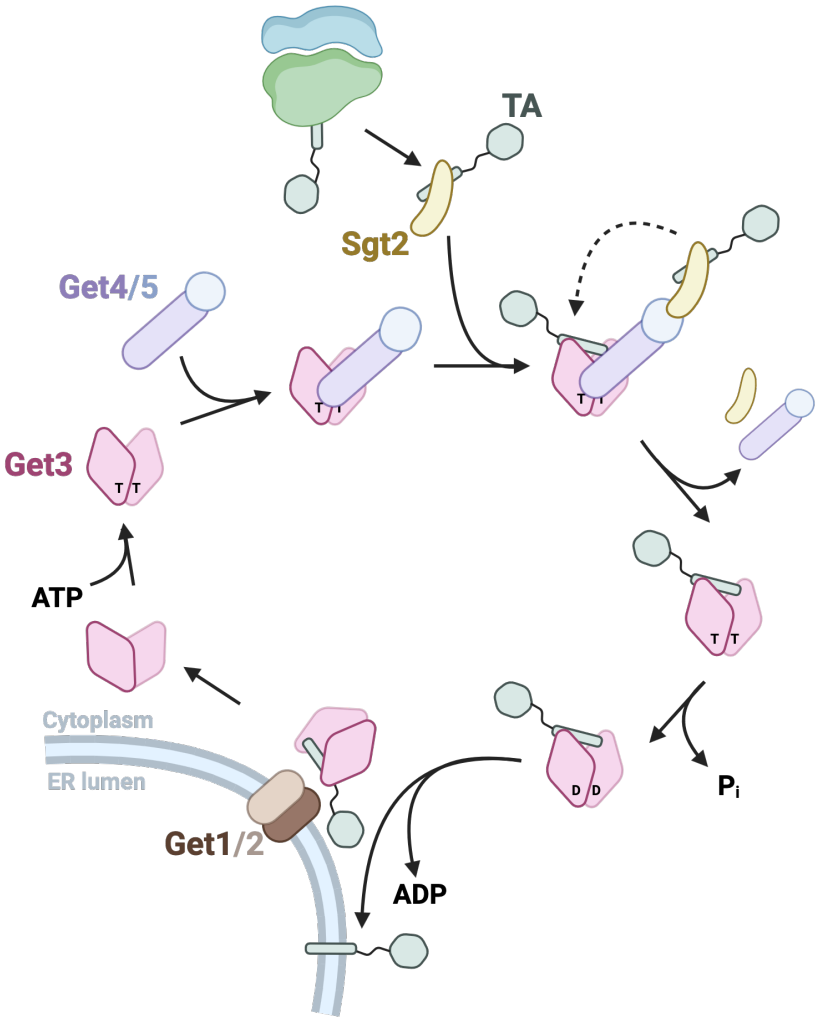


Figure 1.3: Catalytic cycle of the guided entry of tail-anchored proteins pathway in yeast. The TA protein is synthesized by the ribosome and accepted by the chaperone Sgt2 and upstream chaperones (not shown). Sgt2 binds the pre-targeting complex (Get3/4/5) and transfers the TA protein to Get3. Get3 hydrolyzes ATP and delivers the TA protein to the Get1/2 insertase complex at the ER membrane for insertion. The cycle repeats. The lipid bilayer is shown, and the cytoplasm and ER lumen are labeled. Key as follows: TA, tail-anchored protein; T, ATP; D, ADP; P_i, inorganic phosphate; ER, endoplasmic reticulum. Figure created in BioRender.com.

1.9 Tail-Anchored Protein Targeting Factor Get3

Get3 belongs to the P-loop nucleoside triphosphatase (NTPase) superfamily (*22*, *23*), a group of proteins that undergo conformational changes in response to nucleoside triphosphate binding and hydrolysis. These structural rearrangements drive essential functions, such as arsenite binding and detoxification by ArsA (*24*) or electron transfer in nitrogenase (*25*).

P-loop NTPases are defined by a conserved Walker A motif (GxxGxG**K**[ST]), which includes a lysine residue that plays a critical role in phosphate binding (*26*). Get3 and related proteins feature a variant known as the intradimeric Walker A motif (G**K**GGhGK[ST]), which includes a second conserved lysine that extends across the dimer interface, stabilizing the negative charge in the active site to promote catalysis (*27*). P-loop NTPases also include Switch I and Switch II, short loops that mediate structural changes in response to γ-phosphate binding. Switch I contains a highly conserved catalytic aspartate that coordinates a water molecule to facilitate nucleophilic attack on the γ-phosphate.

Get3 is a homodimeric ATPase that contains two domains: the nucleotide-binding domain (where ATP hydrolysis occurs) and the client-binding domain (where TA protein binds). The client-binding domain is comprised of the characteristic Get3/TRC40 insert, which forms a hydrophobic groove or chamber that directly interacts with TA protein substrates (*28*, *29*). ATP binding and hydrolysis drives a series of conformational changes in Get3, enabling the binding, delivery, and release of TA proteins into the ER membrane.

1.10 Post-Translational Membrane Protein Targeting: Alternative Pathways

A second pathway, the ER-membrane protein complex (EMC) pathway, has been shown to insert a diverse set of membrane proteins, including TA proteins and multi-pass membrane proteins, into the ER membrane (Fig. 1.2) (*30*). While the EMC was discovered over a decade ago (*31*), its role as an insertase was only more recently established (*30*, *32*, *33*). The EMC is a large oligomeric membrane protein complex consisting of ~10 subunits. Its TA protein clients tend to be less hydrophobic compared to those targeted by the GET pathway. While the specific cytosolic factors involved in targeting TA proteins to the EMC remain unclear, there is evidence suggesting that calmodulin may play a role in this process (*30*).

Recently, a third pathway, known as the SRP-independent (SND) pathway, has also been implicated in targeting TA proteins to the ER membrane (Fig. 1.2) (*34*). The

SND pathway, which includes the proteins SND1, SND2, and SND3 in yeast, was identified through studies of the TA protein Gas1, which was found to be independent of SRP and only partially dependent on the GET pathway (*35*). The SND pathway has since been shown to target a diverse array of membrane proteins, including both SA and TA proteins (*34*), and may serve as a back-up mechanism for the SRP and GET pathways.

1.11 Protein Targeting in Cyanobacteria

Cyanobacteria possess homologs of most protein targeting machinery found in their non-photosynthetic bacterial relatives, including the general secretory (Sec) pathway, the twin arginine translocation (TAT) pathway, and the SRP pathway (reviewed in (*36*)). This machinery is present in both the plasma and thylakoid membranes. The Sec and TAT pathways transport proteins across the membrane (i.e., into the periplasm and thylakoid lumen), while the SRP pathway inserts membrane proteins into the membrane (i.e., into the plasma and thylakoid membranes). Little is known about the targeting of TA proteins in cyanobacteria.

1.12 Protein Targeting into and in Chloroplasts

~95% of the proteins found in the plant chloroplast are encoded by the nuclear genome and are imported into the chloroplast from the cytosol (*37*). The vast majority of these proteins contain an N-terminal chloroplast transit peptide, which is recognized by various cytosolic factors to facilitate their delivery to the translocon of the outer chloroplast envelope (TOC) and translocon of the inner chloroplast envelope (TIC) import machinery (*38*, *39*). The TOC-TIC system works in concert to import the protein into the chloroplast stroma, where the chloroplast transit peptide is cleaved by stromal processing peptidase (*40*, *41*).

Proteins destined for the inner envelope membrane can be inserted into the membrane either during import through a stop-transfer mechanism or after full import into the stroma, though the specific factors involved remain unclear (*42*).

The majority of proteins targeted to the thylakoid membrane are recognized by the chloroplast SRP (cpSRP) (reviewed in (*43*)). Unlike the canonical SRP system, which operates co-translationally, cpSRP primarily functions post-translationally, as most of its substrates are nuclear-encoded and must be fully synthesized and imported into the stroma before targeting. A small subset of cpSRP clients are encoded by the chloroplast genome and can be targeted co-translationally via the chloroplast ribosome. Additionally, a spontaneous (or unassisted) pathway has been

suggested to target a small subset of proteins to the thylakoid membrane without requiring energy, proton motive force, or any protein factors (*44*, *45*).

Nuclear-encoded proteins destined for the thylakoid lumen contain a secondary thylakoid transit peptide, which is exposed upon cleavage of the primary chloroplast transit peptide (*46*). This thylakoid transit peptide is recognized by various factors for translocation into the thylakoid lumen via the highly conserved chloroplast Sec and TAT pathways (reviewed in (*47*)). Similar to the chloroplast transit peptide, the thylakoid transit peptide is subsequently cleaved by thylakoid processing peptidase (*46*).

Work examining chloroplast TA targeting revealed that the localization of TA proteins SecE1 (Sec translocon subunit E of the thylakoid membrane) and SecE2 (Sec translocon subunit E of the inner envelope membrane) depends solely on features of their TMDs and short C-terminal tails (*48*). While these findings imply that sequence-encoded features govern TA protein targeting, they also raise the question of whether or not distinct targeting pathways mediate this discrimination.

1.13 Tail-Anchored Proteins and Their Targeting in Plants

In plants, nearly all components of the canonical GET pathway have been identified, with the exception of Get5 and Sgt2 (Fig. 1.4) (*49–51*). A notable feature of the plant GET system is the expansion of Get3 homologs. In the model plant *Arabidopsis thaliana*, four Get3 paralogs have been identified: Get3a, Get3b, Get3c, and Get3d. Functional studies have established Get3a as the canonical cytosolic factor responsible for delivering TA proteins to the ER membrane (*52*). The functions of the other paralogs, however, are less understood.

Get3b has been proposed to participate in TA protein targeting in chloroplasts. Subcellular localization studies show that Get3b resides in the chloroplast and Get3c is localized to the mitochondrial matrix (*49*, *52*, *53*). *In vitro* pulldown experiments revealed that Get3b binds the thylakoid TA protein SecE1 but not the inner envelope membrane TA protein SecE2 (*54*), and genetic interaction studies showed that loss of *get3b* enhances the phenotype of a *cpSRP* mutant, suggesting overlapping or parallel roles (*54*). It is also suggested that the thylakoid insertases Alb3 and Alb4 may be potential Get3b interactors (*55*), although this was determined via *in vitro* pulldowns and yeast 2-hybrid assays, which may not reflect interactions in the native system. Despite these findings, no direct *in planta* evidence has confirmed the functional role of Get3b, and the other non-canonical paralogs remain even more enigmatic.

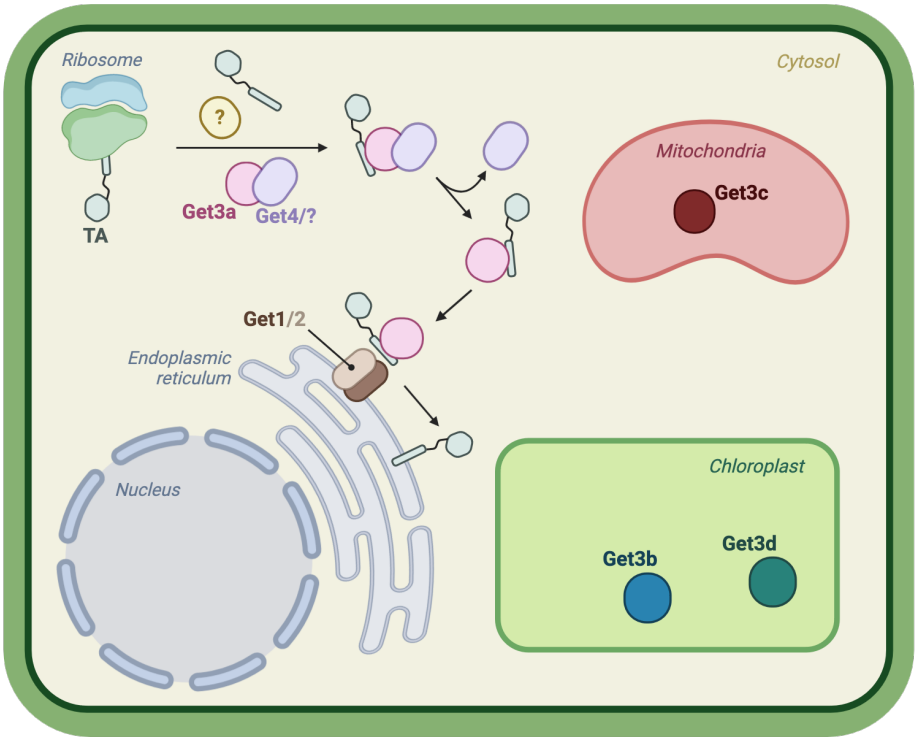


Figure 1.4: Guided entry of tail-anchored proteins (GET) pathway in plants. A cartoon representation showing the GET pathway in a plant cell. TA, Get1, Get2, Get3a, Get3b, Get3c, Get3d, and Get4 are labeled. Unidentified chaperones are depicted as a *yellow sphere* with a "?". Key as follows: TA, tail-anchored protein. The ribosome and select organelles are labeled. Figure created in BioRender.com.

Prior to the work performed here, Get3d has been the least characterized homolog in plants. Its unique domain architecture and conservation across both plants and photosynthetic bacteria suggest an evolutionarily distinct role. No previous studies have addressed its cellular function, localization, or client specificity.

1.14 Get3 Homolog Conserved in Photosynthetic Organisms – Get3d

In this work, I investigate the biochemical and functional roles of Get3d, a Get3 homolog conserved in plants and photosynthetic bacteria. Chapter 2 explores the evolutionary conservation, structure, and biochemical activity of Get3d. The protein is highly conserved across plants, cyanobacteria, and green bacteria and localizes to the chloroplast stroma. Structural studies reveal several distinctive features, including a C-terminal α-crystallin domain, an incomplete active site, and a client-binding chamber in the closed conformation. Biochemical assays demonstrate that Get3d is an active ATPase and is capable of binding tail-anchored proteins. Chapter 3

examines Get3d function in cyanobacteria, which typically encode two variants: a plant-like homolog, similar to chloroplast-localized Get3d, and a canonical homolog, which more closely resembles cytosolic Get3. The plant-like Get3d from *Synechocystis* sp. PCC 6803 is studied in detail. Spot growth assays, pigment quantification, and whole-cell proteomics reveal that this Get3d homolog supports cellular viability, potentially through membrane protein targeting and photosynthesis. The α-crystallin domain appears important for protein expression or stability. Finally, Chapter 4 summarizes these findings and proposes future directions to further elucidate Get3d's role in tail-anchored protein targeting and its broader contributions to photosynthesis.

References

(*1*) Johnson, M. P. (2016). Photosynthesis. *Essays Biochem. 60*, 255–273, DOI: 10.1042/EBC20160016.

(*2*) Sánchez-Baracaldo, P., and Cardona, T. (2020). On the origin of oxygenic photosynthesis and Cyanobacteria. *New Phytol. 225*, 1440–1446, DOI: 10.1111/nph.16249.

(*3*) Hanada, S. (2016). Anoxygenic Photosynthesis—A photochemical reaction that does not contribute to oxygen reproduction—. *Microbes Environ. 31*, 1–3, DOI: 10.1264/jsme2.ME3101rh.

(*4*) Ponce-Toledo, R. I., López-García, P., and Moreira, D. (2019). Horizontal and endosymbiotic gene transfer in early plastid evolution. *New Phytol. 224*, 618–624, DOI: 10.1111/nph.15965.

(*5*) Hoiczyk, E., and Hansel, A. (2000). Cyanobacterial cell walls: news from an unusual prokaryotic envelope. *J. Bacteriol. 182*, 1191–1199, DOI: 10.1128/jb.182.5.1191-1199.2000.

(*6*) Mareš, J., Strunecký, O., Bučinská, L., and Wiedermannová, J. (2019). Evolutionary patterns of thylakoid architecture in cyanobacteria. *Front. Microbiol. 10*, 277, DOI: 10.3389/fmicb.2019.00277.

(*7*) Kirchhoff, H. (2019). Chloroplast ultrastructure in plants. *New Phytol. 223*, 565–574, DOI: 10.1111/nph.15730.

(*8*) Zachar, I., and Boza, G. (2020). Endosymbiosis before eukaryotes: mitochondrial establishment in protoeukaryotes. *Cell. Mol. Life Sci. 77*, 3503–3523, DOI: 10.1007/s00018-020-03462-6.

(*9*) Tirincsi, A., Sicking, M., Hadzibeganovic, D., Haßdenteufel, S., and Lang, S. (2021). The molecular biodiversity of protein targeting and protein transport related to the endoplasmic reticulum. *Int. J. Mol. Sci. 23*, 143, DOI: 10.3390/ijms23010143.

(*10*) Overington, J. P., Al-Lazikani, B., and Hopkins, A. L. (2006). How many drug targets are there? *Nat. Rev. Drug Discovery 5*, 993–996, DOI: 10.1038/nrd2199.

(*11*) Li, F., Egea, P. F., Vecchio, A. J., Asial, I., Gupta, M., Paulino, J., Bajaj, R., Dickinson, M. S., Ferguson-Miller, S., Monk, B. C., et al. (2021). Highlighting membrane protein structure and function: A celebration of the Protein Data Bank. *J. Biol. Chem. 296*, 100557, DOI: 10.1016/j.jbc.2021.100557.

(*12*) von Heijne, G. (2006). Membrane-protein topology. *Nat. Rev. Mol. Cell Biol. 7*, 909–918, DOI: 10.1038/nrm2063.

(*13*) Hegde, R. S., and Keenan, R. J. (2022). The mechanisms of integral membrane protein biogenesis. *Nat. Rev. Mol. Cell Biol. 23*, 107–124, DOI: 10.1038/s41580-021-00413-2.

(*14*) Cymer, F., Von Heijne, G., and White, S. H. (2015). Mechanisms of integral membrane protein insertion and folding. *J. Mol. Biol. 427*, 999–1022, DOI: 10.1016/j.jmb.2014.09.014.

(*15*) Akopian, D., Shen, K., Zhang, X., and Shan, S.-o. (2013). Signal recognition particle: an essential protein-targeting machine. *Annu. Rev. Biochem. 82*, 693–721, DOI: 10.1146/annurev-biochem-072711-164732.

(*16*) Kutay, U., Hartmann, E., and Rapoport, T. A. (1993). A class of membrane proteins with a C-terminal anchor. *Trends Cell Biol. 3*, 72–75, DOI: 10.1016/0962-8924(93)90066-A.

(*17*) Mehlhorn, D. G., Asseck, L. Y., and Grefen, C. (2021). Looking for a safe haven: tail-anchored proteins and their membrane insertion pathways. *Plant Physiol. 187*, 1916–1928, DOI: 10.1093/plphys/kiab298.

(*18*) Chio, U. S., Cho, H., and Shan, S.-o. (2017). Mechanisms of tail-anchored membrane protein targeting and insertion. *Annu. Rev. Cell Dev. Biol. 33*, 417–438, DOI: 10.1146/annurev-cellbio-100616-060839.

(*19*) Zheng, L., Zhang, Z., Wang, H., Zheng, Z., Wang, J., Liu, H., Chen, H., Dong, C., Wang, G., Weng, Y., et al. (2023). Cryo-EM and femtosecond spectroscopic studies provide mechanistic insight into the energy transfer in CpcL-phycobilisomes. *Nat. Commun. 14*, 3961, DOI: 10.1038/s41467-023-39689-7.

(*20*) Gisriel, C. J., Wang, J., Liu, J., Flesher, D. A., Reiss, K. M., Huang, H.-L., Yang, K. R., Armstrong, W. H., Gunner, M., Batista, V. S., et al. (2022). High-resolution cryo-electron microscopy structure of photosystem II from the mesophilic cyanobacterium, Synechocystis sp. PCC 6803. *Proc. Natl. Acad. Sci. U. S. A. 119*, e2116765118, DOI: 10.1073/pnas.2116765118.

(*21*) Shan, S.-o. (2019). Guiding tail-anchored membrane proteins to the endoplasmic reticulum in a chaperone cascade. *J. Biol. Chem. 294*, 16577–16586, DOI: 10.1074/jbc.REV119.006197.

(*22*) Koonin, E. V. (1993). A superfamily of ATPases with diverse functions containing either classical or deviant ATP-binding motif. *J. Mol. Biol. 229*, 1165–1174, DOI: 10.1006/jmbi.1993.1115.

(*23*) Leipe, D. D., Wolf, Y. I., Koonin, E. V., and Aravind, L. (2002). Classification and evolution of P-loop GTPases and related ATPases. *J. Mol. Biol. 317*, 41–72, DOI: 10.1006/jmbi.2001.5378.

(*24*) Rosen, B. P., Weigel, U., Karkaria, C., and Gangola, P. (1988). Molecular characterization of an anion pump. The arsA gene product is an arsenite (antimonate)-stimulated ATPase. *J. Biol. Chem. 263*, 3067–3070, DOI: 10.1016/S0021-9258(18)69034-9.

(*25*) Einsle, O., and Rees, D. C. (2020). Structural enzymology of nitrogenase enzymes. *Chem. Rev. 120*, 4969–5004, DOI: 10.1021/acs.chemrev.0c00067.

(*26*) Walker, J. E., Saraste, M., Runswick, M. J., and Gay, N. J. (1982). Distantly related sequences in the alpha-and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J. 1*, 945–951, DOI: 10.1002/j.1460-2075.1982.tb01276.x.

(*27*) Maggiolo, A. O., Mahajan, S., Rees, D. C., and Clemons Jr., W. M. (2023). Intradimeric walker A ATPases: conserved features of a functionally diverse family. *J. Mol. Biol. 435*, 167965, DOI: 10.1016/j.jmb.2023.167965.

(*28*) Suloway, C. J., Chartron, J. W., Zaslaver, M., and Clemons Jr., W. M. (2009). Model for eukaryotic tail-anchored protein binding based on the structure of Get3. *Proceedings of the National Academy of Sciences 106*, 14849–14854, DOI: 10.1073/pnas.0907522106.

(*29*) Fry, M. Y., Najdrová, V., Maggiolo, A. O., Saladi, S. M., Doležal, P., and Clemons Jr., W. M. (2022). Structurally derived universal mechanism for the catalytic cycle of the tail-anchored targeting factor Get3. *Nat. Struct. Mol. Biol. 29*, 820–830, DOI: 10.1038/s41594-022-00798-4.

(*30*) Guna, A., Volkmar, N., Christianson, J. C., and Hegde, R. S. (2018). The ER membrane protein complex is a transmembrane domain insertase. *Science 359*, 470–473, DOI: 10.1126/science.aao3099.

(*31*) Jonikas, M. C., Collins, S. R., Denic, V., Oh, E., Quan, E. M., Schmid, V., Weibezahn, J., Schwappach, B., Walter, P., Weissman, J. S., et al. (2009). Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science 323*, 1693–1697, DOI: 10.1126/science.1167983.

(*32*) Shurtleff, M. J., Itzhak, D. N., Hussmann, J. A., Schirle Oakdale, N. T., Costa, E. A., Jonikas, M., Weibezahn, J., Popova, K. D., Jan, C. H., Sinitcyn, P., et al. (2018). The ER membrane protein complex interacts cotranslationally to enable biogenesis of multipass membrane proteins. *eLife 7*, e37018, DOI: 10.7554/eLife.37018.

(*33*) Chitwood, P. J., Juszkiewicz, S., Guna, A., Shao, S., and Hegde, R. S. (2018). EMC is required to initiate accurate membrane protein topogenesis. *Cell 175*, 1507–1519, DOI: 10.1016/j.cell.2018.10.009.

(*34*) Aviram, N., Ast, T., Costa, E. A., Arakel, E. C., Chuartzman, S. G., Jan, C. H., Haßdenteufel, S., Dudek, J., Jung, M., Schorr, S., et al. (2016). The SND proteins constitute an alternative targeting route to the endoplasmic reticulum. *Nature 540*, 134–138, DOI: 10.1038/nature20169.

(*35*) Ast, T., Cohen, G., and Schuldiner, M. (2013). A network of cytosolic factors targets SRP-independent proteins to the endoplasmic reticulum. *Cell 152*, 1134–1145, DOI: 10.1016/j.cell.2013.02.003.

(*36*) Frain, K. M., Gangl, D., Jones, A., Zedler, J. A., and Robinson, C. (2016). Protein translocation and thylakoid biogenesis in cyanobacteria. *Biochim. Biophys. Acta, Bioenerg. 1857*, 266–273, DOI: 10.1016/j.bbabio.2015.08.010.

(*37*) Dobrogojski, J., Adamiec, M., and Luciński, R. (2020). The chloroplast genome: a review. *Acta Physiol. Plant. 42*, 98, DOI: 10.1007/s11738-020-03089-x.

(*38*) Thomson, S. M., Pulido, P., and Jarvis, R. P. (2020). Protein import into chloroplasts and its regulation by the ubiquitin-proteasome system. *Biochem. Soc. Trans. 48*, 71–82, DOI: 10.1042/BST20190274.

(*39*) Jin, Z., Wan, L., Zhang, Y., Li, X., Cao, Y., Liu, H., Fan, S., Cao, D., Wang, Z., Li, X., et al. (2022). Structure of a TOC-TIC supercomplex spanning two chloroplast envelope membranes. *Cell 185*, 4788–4800, DOI: 10.1016/j.cell.2022.10.030.

(*40*) Richter, S., and Lamppa, G. K. (1998). A chloroplast processing enzyme functions as the general stromal processing peptidase. *Proc. Natl. Acad. Sci. U. S. A. 95*, 7463–7468, DOI: 10.1073/pnas.95.13.7463.

(*41*) Richter, S., and Lamppa, G. K. (1999). Stromal processing peptidase binds transit peptides and initiates their ATP-dependent turnover in chloroplasts. *J. Cell Biol. 147*, 33–44, DOI: 10.1083/jcb.147.1.33.

(*42*) Viana, A. A., Li, M., and Schnell, D. J. (2010). Determinants for stop-transfer and post-import pathways for protein targeting to the chloroplast inner envelope membrane. *J. Biol. Chem. 285*, 12948–12960, DOI: 10.1074/jbc.M110.109744.

(*43*) Ziehe, D., Dünschede, B., and Schünemann, D. (2017). From bacteria to chloroplasts: evolution of the chloroplast SRP system. *Biol. Chem. 398*, 653–661, DOI: 10.1515/hsz-2016-0292.

(*44*) Steiner, J. M., Köcher, T., Nagy, C., and Löffelhardt, W. (2002). Chloroplast SecE: evidence for spontaneous insertion into the thylakoid membrane. *Biochem. Biophys. Res. Commun. 293*, 747–752, DOI: 10.1016/S0006-291X(02)00285-1.

(*45*) Kim, S. J., Jansson, S., Hoffman, N. E., Robinson, C., and Mant, A. (1999). Distinct "assisted" and "spontaneous" mechanisms for the insertion of polytopic chlorophyll-binding proteins into the thylakoid membrane. *J. Biol. Chem. 274*, 4715–4721, DOI: 10.1074/jbc.274.8.4715.

(*46*) Shackleton, J. B., and Robinson, C. (1991). Transport of proteins into chloroplasts. The thylakoidal processing peptidase is a signal-type peptidase with stringent substrate requirements at the -3 and -1 positions. *J. Biol. Chem. 266*, 12152–12156, DOI: 10.1016/S0021-9258(18)98872-1.

(*47*) Albiniak, A. M., Baglieri, J., and Robinson, C. (2012). Targeting of lumenal proteins across the thylakoid membrane. *J. Exp. Bot. 63*, 1689–1698, DOI: 10.1093/jxb/err444.

(*48*) Anderson, S. A., Singhal, R., and Fernandez, D. E. (2019). Membrane-Specific Targeting of Tail-Anchored Proteins SECE1 and SECE2 Within Chloroplasts. *Front. Plant Sci. 10*, 1401, DOI: 10.3389/fpls.2019.01401.

(*49*) Duncan, O., van der Merwe, M. J., Daley, D. O., and Whelan, J. (2013). The outer mitochondrial membrane in higher plants. *Trends Plant Sci. 18*, 207–217, DOI: 10.1016/j.tplants.2012.12.004.

(*50*) Srivastava, R., Zalisko, B. E., Keenan, R. J., and Howell, S. H. (2017). The GET system inserts the tail-anchored protein, SYP72, into endoplasmic reticulum membranes. *Plant Physiol. 173*, 1137–1145, DOI: 10.1104/pp.16.00928.

(*51*) Asseck, L. Y., Mehlhorn, D. G., Monroy, J. R., Ricardi, M. M., Breuninger, H., Wallmeroth, N., Berendzen, K. W., Nowrousian, M., Xing, S., Schwappach, B., et al. (2021). Endoplasmic reticulum membrane receptors of the GET pathway are conserved throughout eukaryotes. *Proc. Natl. Acad. Sci. U. S. A. 118*, e2017636118, DOI: 10.1073/pnas.2017636118.

(*52*) Xing, S., Mehlhorn, D. G., Wallmeroth, N., Asseck, L. Y., Kar, R., Voss, A., Denninger, P., Schmidt, V. A. F., Schwarzländer, M., Stierhof, Y.-D., et al. (2017). Loss of GET pathway orthologs in Arabidopsis thaliana causes root hair growth defects and affects SNARE abundance. *Proc. Natl. Acad. Sci. U. S. A. 114*, E1544–E1553, DOI: 10.1073/pnas.1619525114.

(*53*) Bodensohn, U. S., Simm, S., Fischer, K., Jäschke, M., Groß, L. E., Kramer, K., Ehmann, C., Rensing, S. A., Ladig, R., and Schleiff, E. (2019). The intracellular distribution of the components of the GET system in vascular plants. *Biochim. Biophys. Acta, Mol. Cell Res. 1866*, 1650–1662, DOI: 10.1016/j.bbamcr.2019.06.012.

(*54*) Anderson, S. A., Satyanarayan, M. B., Wessendorf, R. L., Lu, Y., and Fernandez, D. E. (2021). A homolog of GuidedEntry of Tail-anchored proteins3 functions in membrane-specific protein targeting in chloroplasts of Arabidopsis. *Plant Cell 33*, 2812–2833, DOI: 10.1093/plcell/koab145.

(*55*) Bodensohn, U. S., Dünschede, B., Kuhlmann, C., Kumari, K., Ladig, R., Grefen, C., Schleiff, E., Fernandez, D., and Schünemann, D. (2025). GET3B is involved in chloroplast biogenesis and interacts with the thylakoidal ALB3 and ALB4 insertases. *Plant Cell Rep. 44*, 108, DOI: 10.1007/s00299-025-03500-2.

C h a p t e r 2

THE STRUCTURE OF THE CONSERVED GET3D

Adapted from:

Alexandra N. Barlow, M.S. Manu, Shyam M. Saladi, Paul T. Tarr, Yashpal Yadav, Aye M.M. Thinn, Yun Zhu, Arthur D. Laganowsky, William M. Clemons, Jr., and Sureshkumar Ramasamy (2023). Structures of Get3d reveal a distinct architecture associated with the emergence of photosynthesis. *J. Biol. Chem.* 299, 104752, DOI: 10.1016/j.jbc.2023.104752.

Contributions:

Alexandra N. Barlow wrote the manuscript, prepared figures, refined crystal structures, performed cloning, expressed and purified proteins for biochemical assays, performed biochemical assays, and prepared samples for mass spectrometry.

2.1 Abstract

Homologs of the protein Get3 have been identified in all domains yet remain to be fully characterized. In the eukaryotic cytoplasm, Get3 delivers tail-anchored (TA) integral membrane proteins, defined by a single transmembrane helix at their C-terminus, to the endoplasmic reticulum. While most eukaryotes have a single Get3 gene, plants are notable for having multiple Get3 paralogs. Get3d is conserved across land plants and photosynthetic bacteria and includes a distinctive C-terminal α-crystallin domain. After tracing the evolutionary origin of Get3d, we solve the *Arabidopsis thaliana* Get3d crystal structure, identify its localization to the chloroplast, and provide evidence for a role in TA protein binding. The structure is identical to that of a cyanobacterial Get3 homolog, which is further refined here. Distinct features of Get3d include an incomplete active site, a closed conformation in the apo-state, and a hydrophobic chamber. Both homologs have ATPase activity and are capable of binding TA proteins, supporting a potential role in TA protein targeting. Get3d first emerged with the development of photosynthesis and has been conserved for over 1.2 billion years, ultimately localizing to the chloroplasts of higher plants. This deep evolutionary conservation suggests a role in maintaining the homeostasis of photosynthetic machinery.

2.2 Introduction

A central problem for eukaryotes and photosynthetic microbes is the presence of multiple membranes that contain specific subsets of integral membrane proteins (IMPs). The correct targeting and insertion of IMPs to distinct and specific locations is necessary to maintain cellular homeostasis (*1*). Most IMPs span the lipid bilayer with hydrophobic α-helical transmembrane domains (TMDs) that present a challenge during biogenesis, as they must be shielded from the aqueous environment prior to insertion to avoid aggregation. For IMPs, a signal, often the first TMD, encodes the destination, and the location of the signal in the protein sequence dictates co-translational versus post-translational targeting.

A subset of IMPs, termed tail-anchored (TA) proteins, are defined by a single TMD within ~30 residues of the C-terminus (*2*), which serves as their targeting signal. Because of the position of their TMDs in the sequence, TA proteins must be fully synthesized and released by the ribosome prior to targeting (*3*). TA proteins are targeted in a post-translational chaperone-assisted mechanism primarily by the guided entry of TA proteins (GET) pathway. A central player in the GET pathway is Get3, which captures the TA signal in the cytosol and delivers the protein to a membrane insertion complex at the endoplasmic reticulum (ER), dependent on ATP hydrolysis (*4–6*). Get3 has been shown to have a conserved mechanism across eukaryotes (*5*, *6*). Many TA proteins play essential roles, for example, vesicle trafficking, protein localization, and regulation of apoptosis, and are targeted to various membranes (*7*, *8*).

Organisms that generate energy via photosynthesis require additional cellular compartments and membranes, which necessitates added complexity for the targeting of TA proteins to these membranes (*9*). Examples are chloroplasts and other plastids, multimembrane organelles, derived from a single endosymbiotic event in which a cyanobacterial ancestor was incorporated into an early eukaryote, around 1.2 billion years ago (*10*). Given this event, it is unsurprising that protein targeting in chloroplasts has features conserved from protein targeting in bacteria. While much is known about membrane protein targeting in chloroplasts (*11*, *12*), it is unclear if a distinct mechanism exists for targeting of TA proteins to either the thylakoid or inner envelope membrane of chloroplasts. Interesting chloroplastic TA protein examples are the two SecE paralogs of the distinct general secretory (Sec) translocons present on either the inner envelope (SecE2) or thylakoid (SecE1) membranes, where the respective targeting was found to be dependent solely on the TMD and flanking

C-terminal residues (*13*). Recently, the Get3 paralog *At*Get3b has been implicated in targeting SecE1 to the thylakoid membrane (*14*); however, further studies will be necessary to probe the role of *At*Get3b *in vivo*. While the protein factors and specific targeting mechanisms are not yet known, sorting mechanisms are required that distinguish the characteristics of the TA proteins (*14*, *15*).

Sequence analysis revealed that the Get3/ArsA fold family includes homologs from all three domains of life including one identified by Pfam in 2006 (*16*) and discussed first by Chartron *et al.* in 2012 (*17*) with a distinct architecture characterized by an α-crystallin domain (αCD) at its C-terminus. Unlike fungi and metazoa, which contain a single Get3 in each genome, plants and cyanobacteria have multiple Get3 genes (*8*, *17–20*). The model plant *Arabidopsis thaliana* (*A. thaliana*) contains four such genes (noted Get3a, b, c, and d). Evidence supports that *At*Get3a (UniProt ID: Q949M9) resides in the cytosol, *At*Get3b (UniProt ID: A1L4Y1) in the chloroplast stroma, and *At*Get3c (UniProt ID: Q5XF80) in the mitochondrial matrix (*18–22*). *At*Get3a has the same function as other cytoplasmic Get3 homologs in eukaryotes, targeting hydrophobic TA proteins in the cytosol to the ER membrane with knockouts of *At*Get3a resulting in distinct phenotypes (*20*, *23*). There is conflicting evidence as to whether *At*Get3b knockouts cause a phenotypic defect (*14*, *20*), and no phenotype has been found for knockouts of *At*Get3c (*20*). No experimental information is available for *At*Get3d (UniProt ID: Q6DYE4).

Here, we provide a detailed characterization of Get3d, a distinct member of the Get3 family. We demonstrate that it first evolved in photosynthetic bacteria and has been conserved in the chloroplasts of plants. We solve the atomic structure of a plant Get3d and further refine a previously deposited structure of a cyanobacterial Get3d. We identify conserved functional motifs and identify distinct features of Get3d. We then investigate these functional motifs and show that Get3d can bind TA proteins irrespective of the unique αCD at its C-terminus and can hydrolyze ATP. This work provides a comprehensive characterization of a Get3 family member that has deep evolutionary roots connected to photosynthesis.

2.3 Results

Placing Plant Paralogs within the Get3/ArsA Fold Family

As Get3/ArsA homologs span the tree of life with several distinct clade lineages, we first sought to examine the evolutionary history of Get3d by performing a thorough phylogenetic reconstruction of Get3/ArsA homologs (Figs. 2.1A and S2.1). We identified all Get3 proteins present in UniProt using the Pfam database (*16*, *22*) and then aligned them to a seed structural alignment of Get3 proteins based on solved 3D structures (*24*, *25*). A phylogenetic reconstruction was then calculated using maximum likelihood with clades collapsed at a 70% bootstrap support (*26*).

The first clade is ArsA, a soluble ATPase involved in protecting against arsenite toxicity found primarily in bacteria (*27–31*). Next is the well-studied cytoplasmic Get3 that is present across eukaryotes and some archaea and includes *At*Get3a. Another Get3 lineage is restricted to Viridiplantae and contains both *At*Get3b and *At*Get3c. The final Get3 lineage is first found in green and purple bacteria and then cyanobacteria and across Viridiplantae and contains both *At*Get3d and the Get3d homolog from the cyanobacteria *Nostoc* sp. PCC 7120 (*Nos*Get3d) with a solved structure (Figs. 2.1A and S2.1).

At the sequence level, Get3d proteins contain regions that can be aligned to notable motifs of the Get3 family, including the P-loop, switch I, switch II, Get3 motif/TRC40-insert, and A-loop (Fig. 2.1B) (*33*). The P-loop is well conserved with the so-called intradimeric (or deviant) Walker A motif (*28*, *31*, *34*). Some Get3d proteins, such as the second homolog in *Synechocystis* sp. PCC 6803 (Sll0086) and green sulfur and non-sulfur Get3d, have well-conserved switch I, switch II, and A-loop regions. More common are homologs such as *At*Get3d and *Nos*Get3d that have degenerate catalytic residues including divergent switch I loops and missing A-loops. The hydrophobic nature of the Get3 motif/TRC40-insert of Get3 homologs is conserved across the Get3d family, unlike the related ArsA proteins (*33*). The signature CXXC motif that coordinates a Zn^{2+} at the Get3 dimer interface is also missing in the Get3d family (*33*).

To better understand the origin and distribution of Get3 proteins from plants and photosynthetic bacteria, we carried out a more extensive phylogenetic reconstruction by specifically focusing on Get3 proteins from these two groups. Plant Get3a, Get3b/Get3c, as well as Get3d homologs, each form separate clades (Fig. S2.2, A–C). Get3a forms a single clade with fungal and metazoan Get3 homologs found at their root (Figs. S2.1 and S2.2A). Get3b and Get3c trace to a single common

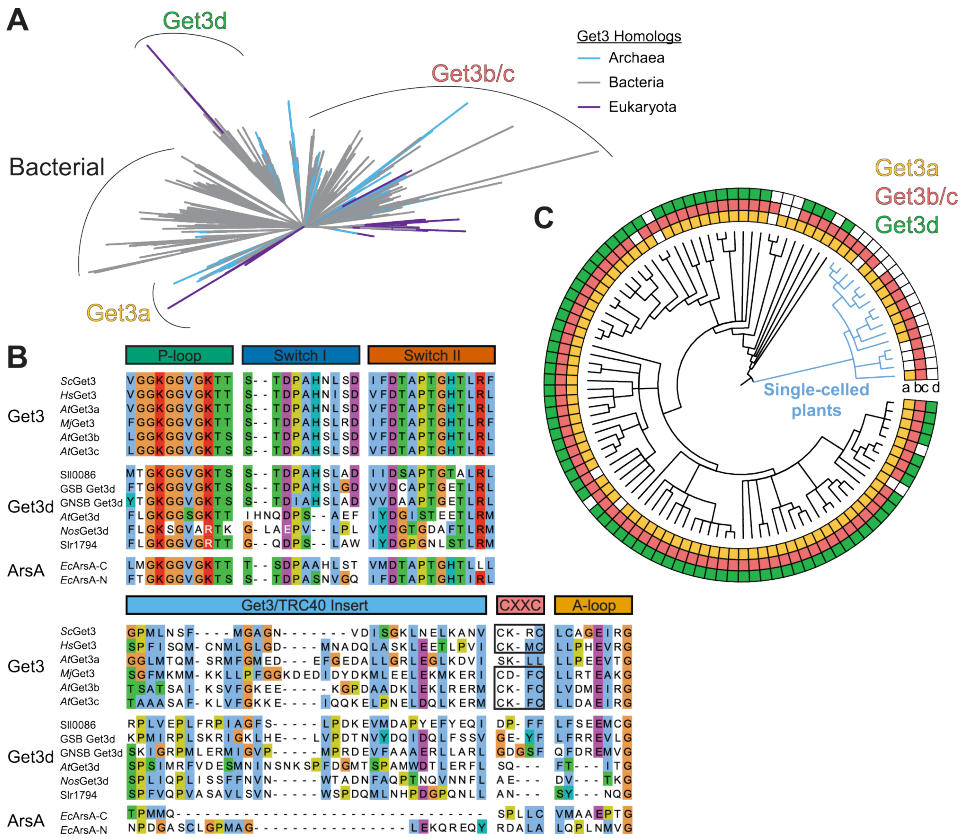


Figure 2.1: Identification and features of Get3 homologs. (A) A phylogenetic tree of Get3 homologs where branchpoints (i.e. nodes) with less than 70% bootstrap support are collapsed. Get3d is found in a separate clade. Get3a clusters with canonical cytoplasmic Get3 proteins, including yeast Get3 (Uniprot ID: Q12154). Get3b and Get3c also form a distinct clade. Superkingdom is highlighted by color archaea (*light blue*), bacteria (*grey*), & eukaryota (*purple*). Inner branches are colored where all descendants are of a single taxonomic grouping. (B) Sequence alignment of important regions from selected Get3 homologs. Features as discussed in the text are labeled *above* the respective sequences. Residues are colored per the ClustalX color scheme (*32*). Species key: *Sc*, *Saccharomyces cerevisiae*; *Hs*, *Homo sapiens*; *At*, *Arabidopsis thaliana*; *Mj*, *Methanocaldococcus jannaschii*; GSB, *Chlorobium* sp. (green sulfur bacteria); GNSB, *Chloroflexii* sp. (green non-sulfur bacteria); *Nos*, *Nostoc* sp. PCC 7120; *Ec*, *Escherichia coli*. Slr1794 and Sll0086 are Get3d homologs from *Synechocystis* sp. PCC 6803. (C) A cladogram of plants showing the presence of Get3 homologs by taxonomic genera. Filled squares show that a homolog was identified in at least one member of the genus. Clades are colored Get3a (*yellow*), Get3b/c (*red*), & Get3d (*green*). Single-celled plants are labeled and highlighted as *blue* branches in the tree. A higher resolution figure with labeled genera is provided in Fig. S2.3.

ancestor, that is, monophyletic, and the clade cannot be split into separate b and c groups (Figs. S2.1 and S2.2B). Get3d forms a single clade with cyanobacteria and the green sulfur *Chloroflexi* Get3 proteins as the nearest relatives (Figs. S2.1 and S2.2C). Get3a and Get3b/c clades trace to a more recent common ancestor before their common ancestor with Get3d (Fig. S2.1).

To investigate the conservation of Get3d in photosynthetic bacteria specifically, we quantified the number of sequenced genomes that contain Get3 proteins with the characteristic αCD in various phyla (Fig. S2.2D). Get3d is completely conserved in cyanobacteria and green sulfur and non-sulfur bacteria, which do not contain another Get3-like homolog. In purple bacteria, many families contain species that encode a cytoplasmic Get3 homolog, whereas only a single purple non–sulfur bacterium encodes a Get3d homolog. Furthermore, the number of Get3d homologs encoded by representative species from these phyla was determined (Fig. S2.2E). While some species contain no Get3d proteins, others contain multiple copies of Get3d in their genome, such as *Chlorobium chlorochromatii*, which contains five copies of Get3d. This could be due to whole genome duplication events (*35*) and horizontal gene transfers (*36*, *37*), which are integral to the evolution of protein families and homologs (*38*). As the availability of sequenced genomes from photosynthetic bacteria increases, a broader depth of information about the evolution and conservation of Get3d in photosynthetic bacteria can be learned.

We next consider the distribution of the Get3 homologs across plants. Using our phylogenetic information, we assigned each Get3 to either the a, b/c, or d group, we can correlate the taxonomic distribution of each group across Viridiplantae (Figs. 2.1C and S2.3). The tree is collapsed at the level of taxonomic genus to minimize errors resulting from uneven genome annotation; however, in some cases, poor annotation within a genus may preclude the ability to identify a given homolog. The results suggest that at least one protein from each of the Get3a and Get3b/c groups is present across all plant genomes. Get3d proteins are found across land plants (e.g., mosses, grasses, and eudicots) yet are completely missing in single-celled plants, the blue branch in Figs. 2.1C and S2.3. Given that the nearest relative of Get3d is a cyanobacteria and that plants are derived from an endosymbiotic event that led to a chloroplast (*10*, *39*, *40*), the absence of Get3d in single-celled plants suggests gene loss in the corresponding taxa.

Subcellular Localization of Get3d

The cellular localization of *At*Get3d has not been experimentally determined. Like many essential genes from the ancestral cyanobacteria, Get3d acquired a chloroplast targeting signal during the endosymbiotic transfer of genes into the plant genome (*10*, *39*, *40*). Computational methods support this predicting that *At*Get3d is localized to the chloroplast stroma (~91% likelihood) with a small probability (~5%) of it being localized to the thylakoid space (*41*).

To experimentally confirm the localization of *At*Get3d, we employed *Agrobacterium*-mediated expression of *At*Get3d in *Nicotiana benthamiana* leaves (*42*). Constructs were generated with green fluorescent protein (GFP) appended to the Get3d gene, with or without the predicted chloroplast transit peptide, as a C-terminal fusion. This allowed us to monitor the localization of expressed Get3d after *Agrobacterium*-mediated insertion into the *Nicotiana* genome. Upon infiltration, the tobacco leaves were monitored by fluorescence microscopy. For the full-length Get3d gene that contained the transit peptide, the fluorescence co-localized with the intrinsic chlorophyll autofluorescence, indicating that *At*Get3d localizes to the chloroplast (Fig. 2.2). Localization to the stroma versus thylakoid lumen could not be distinguished here. For Get3d lacking the transit peptide, the GFP signal gave a pattern of distinct puncta not associated with a clear subcellular structure in the mesophyll cells. The puncta are not consistent with the typical localization of GFP, the cytosol and nucleus, in these cells (*43*). These results confirm the predictions of a functioning chloroplast targeting signal.

Crystal Structures of Photosynthesis-Associated Get3d

The significant differences in Get3 sequences hint at unique structural features of the photosynthetic homologs, motivating us to solve the crystal structure of *At*Get3d (Figs. 2.3A and S2.4A). We generated a construct without the chloroplast transit peptide ($\Delta 1 - 57$) that purified as a single peak via size-exclusion chromatography, consistent with a homodimer. While the sequence suggested missing residues for ATP binding, only crystals grown in the presence of ADP resulted in diffraction. Final crystals, grown by sitting drop in 50 mM sodium cacodylate (pH 5.47), 50 mM lithium sulfate, and 30% PEG-4000, were frozen with the addition of 30% glycerol as a cryoprotectant, and a complete native dataset was collected to 2.0 Å resolution in the space group $P\ 1\ 2_1\ 1$. The closest homolog, with 31% sequence identity, in the Protein Data Bank (PDB) is the cyanobacterial *Nos*Get3d (PDB ID: 3IGF) (*44*, *45*) from *Nostoc* sp. PCC 7120 (also referred to as *Anabaena* sp. PCC

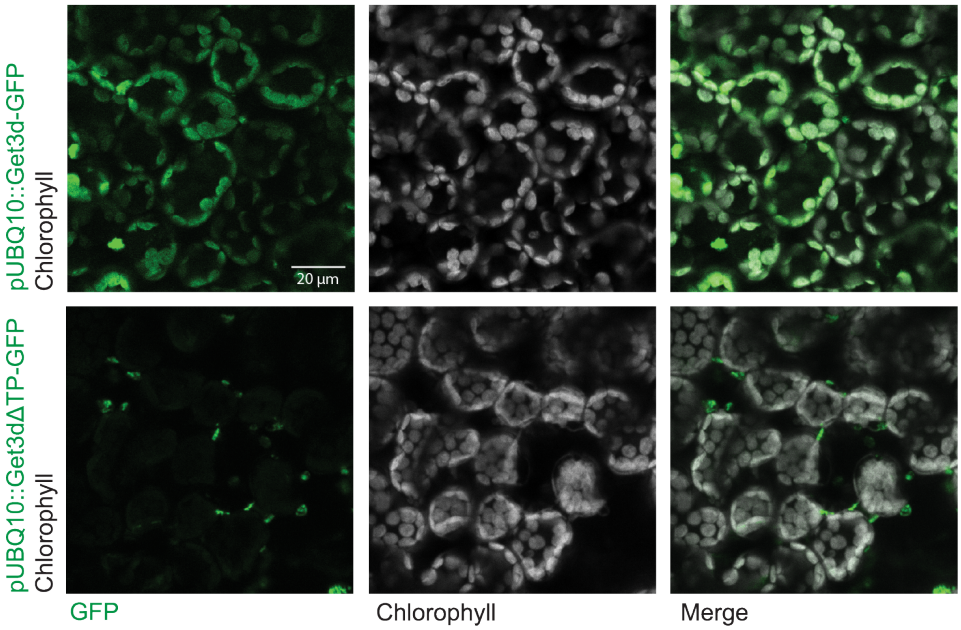


Figure 2.2: Cellular localization of *At*Get3d. Confocal microscopy images of *Nicotiana benthamiana* leaf mesophyll infiltrated with *Agrobacterium tumefaciens* harboring T-DNA plasmids containing *At*Get3d-GFP with (*top*) and without (*bottom*) the chloroplast transit peptide (TP). Get3d is expressed under the control of the pUBQ10 promoter. Scale shown.

7120). The *Nos*Get3d homodimer was used as a search model to obtain phases by molecular replacement. As anticipated, the structure contained a homodimer in the asymmetric unit. The structure was refined to an R-factor of 0.22 and free-R-factor of 0.26 (crystallographic statistics in Table S2.1). No density for nucleotide was visible in the putative active site; however, there was clear density for an inorganic phosphate and a Mg^{2+} ion (Fig. S2.4B). Residues 250 to 260, 330 to 331, and 378 to 382 in monomer A and residues 252 to 261 and 380 to 384 in monomer B could not be resolved in the density.

Like structures of Get3 homologs, *At*Get3d is a homodimer with a core nucleotide-binding domain and an α-helical client-binding domain (CBD) (Figs. 2.3A and S2.4A) (*33*, *46*). A structural alignment suggests that, after *Nos*Get3d (Fig. 2.3B), *At*Get3d is most similar to the closed conformation of yeast Get3 (Fig. 2.3C) (*46*). Consistent with our bioinformatic analysis, *At*Get3d has an αCD at its C-terminus, which is unique to the Get3d clade (Fig. S2.4A). The P-loop, switch I, and switch II that define the nucleotide-binding domain are conserved in this subgroup, whereas the A-loop, which recognizes the adenosine of the substrate ATP, is missing, as predicted from the sequence (Figs. 2.1B and S2.5, A–C). *At*Get3d also lacks the helix that would contain the CXXC motif and, expectedly, no bound Zn^{2+} is observed

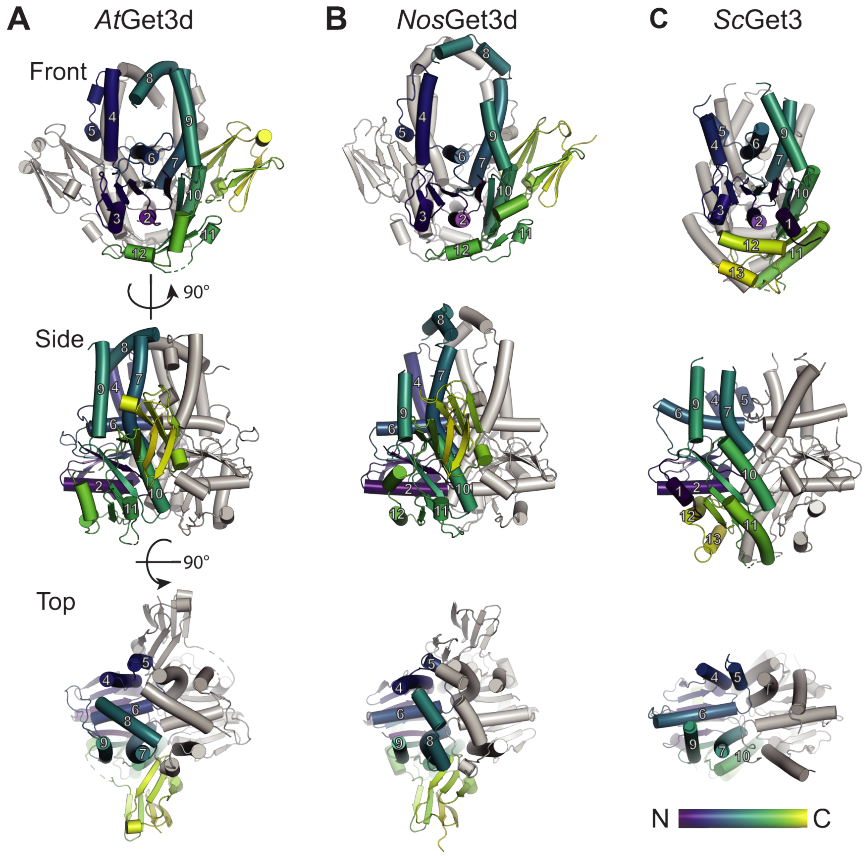


Figure 2.3: Structures of *At*Get3d and *Nos*Get3d. *Front*, *side*, and *top* view of the structure of (A) *At*Get3d (PDB ID: 8ELF), (B) *Nos*Get3d (PDB ID: 8EGK), and (C) the closed conformation of yeast Get3 (*Sc*Get3, PDB ID: 2WOJ). For each, one monomer is shown in *viridis*, and the other is shown in *grey*. TMDs are numbered from N- to C-terminus for reference based on the fungal Get3 structure.

(Figs. 2.1B and S2.5D). As opposed to the groove seen in fungal Get3 structures, the CBD of *At*Get3d is a chamber (Fig. S2.4A) (*33*, *46*).

Upon inspection, additional density was visualized in the chamber (Fig. S2.6A) that could not be explained by protein or solvent. Native mass spectrometry was performed, and analysis confirmed that *At*Get3d is a dimer and that it copurifies with small molecules (~750 ± 140 Da) (Fig. S2.6, B and C). Although the mass of these adducts is consistent with phospholipids (*47*), it is not possible to assign the exact lipid given the mass and associated error. Considering these results, the density, and the properties of the binding site, a phosphatidic acid was built into this density (Fig. S2.6, A and D). The aliphatic chains fit the two tubes of density, whereas the phosphate head group forms a salt bridge with R176 and R182 along with a possible H-bond with Q136 (Fig. S2.6A). The general features are found in the *Nos*Get3d structure, yet it is unlikely that this is primarily a lipid-binding site,

as the charged residues are not conserved (Fig. S2.6E).

A Refined Structure of *Nostoc* sp. Get3d

In *At*Get3d, the α-helical CBD is enclosed by additional helices not seen in the deposited Get3d structure from *Nostoc* sp. (Figs. 2.3A and S2.7A). We viewed the electron density for the *Nostoc* sp. homolog using the deposited structure factors for PDB ID: 3IGF (*44*). With this map, we could clearly identify additional density consistent with the helices of *At*Get3d that enclosed the CBD. We built into this density adding 41 residues total and further refined the *Nos*Get3d structure (Fig. S2.7, B–D) with refinement statistics in Table S2.1.

Overall, *Nos*Get3d is structurally very similar to the *At*Get3d structure (backbone root mean square deviation = 2.33 Å). As with *At*Get3d, *Nos*Get3d has some unidentified densities in the hydrophobic chamber (Fig. S2.6F). One prominent difference is that the helices enclosing the CBD in *Nos*Get3d are further from the bottom of the chamber than in *At*Get3d, resulting in a larger hydrophobic chamber (Fig. 2.3, A and B).

The α-Crystallin Domain

A distinctive feature of Get3d is the αCD at its C-terminus (Fig. 2.4A). This domain has the hallmark α-crystallin fold, a compact β-sandwich composed of seven antiparallel β-strands (Fig. 2.4, A and B) (*48*, *49*). While the fold is conserved, this domain has low sequence similarity to other αCDs and the related small heat shock proteins (sHSPs) and is missing typical features important to αCD dimerization and oligomerization, such as the loop containing β6 found in most plant, yeast, and bacterial αCD/sHSPs (Fig. S2.8). We have found no evidence of either the αCD dimers or higher order oligomers that are found for most αCDs.

In both structures, the αCD sits at the interface of the two Get3 monomers with electrostatic and hydrophobic interactions to both monomers (Figs. 2.4C and S2.9A). The interface buries ~1270 Å^2 in *At*Get3d and ~1280 Å^2 in *Nos*Get3d. The αCD occupies the same binding site as Get4 on cytoplasmic Get3 with a comparable buried surface (~1320 Å^2, PDB ID: 4PWX) (Fig. 2.4D) (*50*). Overall, the two Get3d interfaces have similar properties, yet the specific electrostatic and hydrophobic interactions are different. Some residues are conserved, such as *At*Get3d R209 and D425, which form a salt bridge across the interface (Fig. 2.4C). In *At*Get3d, the loop connecting the N-terminus of the αCD to the rest of the protein is partially disordered (Fig. 2.4A). Disruption of the αCD interface would likely result in the αCD

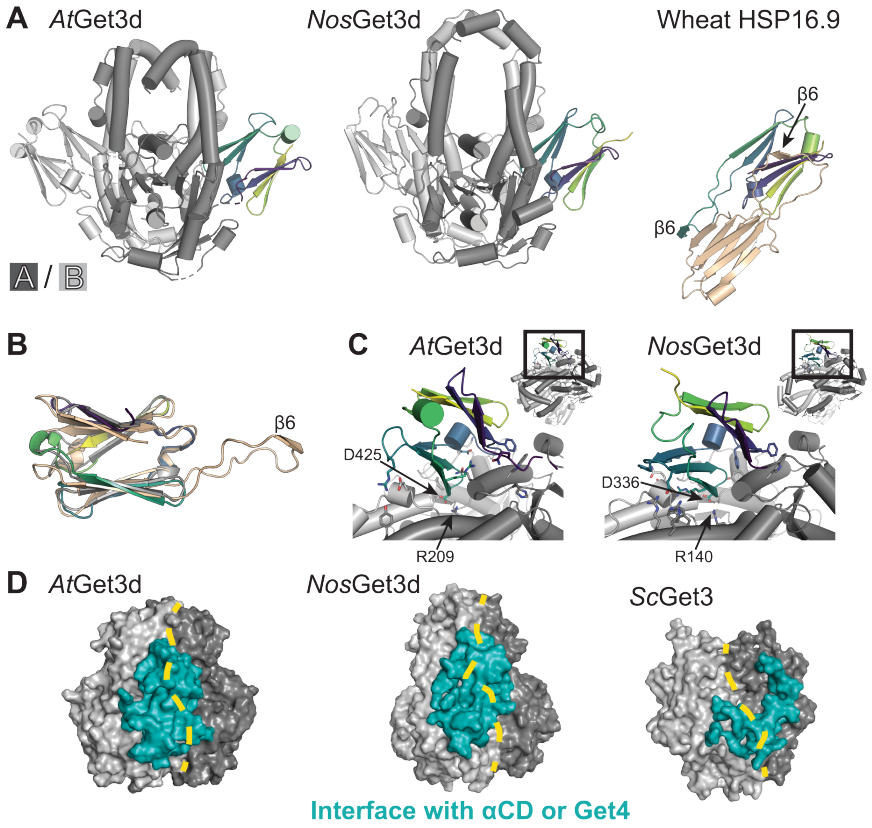


Figure 2.4: Structural analysis of the α-crystallin domain (αCD) of Get3d. (A) Front view of the structure of *At*Get3d and *Nos*Get3d and the αCD of wheat HSP16.9 (PDB ID: 1GME) (from *left* to *right*). Get3d colored with monomer A αCD (*viridis*), monomer A Get3 domain (*dark grey*), and monomer B (*light grey*). HSP16.9 colored monomer A (*viridis*) and monomer B (*wheat*). (B) Alignment of the αCD of *At*Get3d (*viridis*), *Nos*Get3d (*grey*), and wheat HSP16.9 (*wheat*). (C) Residues involved in electrostatic and hydrophobic interactions at the interface of the αCD of *At*Get3d (*left*) and *Nos*Get3d (*right*) shown as sticks with discussed residues labeled. For each, the region highlighted is shown in the full structure *above*. Colored as in (A). (D) The interaction surface of the αCD of *At*Get3d and *Nos*Get3d with the Get3 domain and yeast Get3 (*Sc*Get3) with Get4/5 (PDB ID: 4PWX) (from *left* to *right*) showing the Get3 domains (*dark and light grey*), interaction surface (*dark teal*), and interface between the two Get3 domains (*dotted yellow line*).

being loosely associated with the rest of the protein. It is interesting to speculate that under some conditions, this domain could be exposed.

There are a few additional features of interest related to the αCD interface with the rest of Get3d. First, in *At*Get3d, a salt bridge formed between K107 in the Get3 domain to E404 in the αCD of the opposite monomer is reminiscent of a salt bridge between the homologous yeast Get3 K69 to D74 of Get4 (Fig. S2.9B). In the cytoplasm, this interaction regulates Get3 ATPase activity (*50*). Here, this interaction could be important for regulating communication between the αCD and the active site, although it is not conserved in *Nos*Get3d. A second surprising feature is that both Get3d homologs contain a conserved cis-proline in the loop before helix α11 (340 in *At*Get3d and 265 in *Nos*Get3d), which is not present in cytoplasmic Get3 proteins (Fig. 2.3).

Get3d as an ATPase

Both *At*Get3d and *Nos*Get3d retain components required for ATPase activity including the P-loop (Walker A motif) that recognizes the α- and β-phosphates of the substrate ATP, as found in fungal Get3 (Fig. 2.5, A and B) (*33*, *46*). Get3 belongs to the Mrp/MinD subfamily of the SIMIBI class of NTPases, which are characterized by having an intradimeric (or deviant) Walker A motif (*28*, *31*, *34*, *51*). The canonical Walker A motif contains a conserved lysine (GxxGxGK[ST]) that mediates phosphate binding. The intradimeric Walker A motif contains a second conserved lysine (GKGGhGK[ST]) that reaches across the dimer interface when ATP is bound to facilitate catalysis. In cytoplasmic Get3 proteins, the intradimeric Walker A lysine stabilizes the accumulation of negative charge that builds up in the active site during the water-mediated nucleophilic attack on the γ-phosphate (*33*, *46*). This lysine, and presumably its catalytic role, is conserved in Get3d (Fig. 2.5, A and B). The canonical Walker A lysine, which is broadly conserved in P-loop NTPases (*28*) including Get3 (*33*, *46*), points toward the β-phosphate within the same monomer. In the *At*Get3d crystal structure, the bound inorganic phosphate correlates to the β-phosphate in the ATP-bound Get3 structures and is partially coordinated by this lysine (Fig. 2.5, A and B) (*5*, *46*). While the lysine is conserved in *At*Get3d, it is unexpectedly an arginine in *Nos*Get3d (Figs. 2.1B and 2.5, A and B), which is distinct from most cyanobacterial Get3d homologs and would suggest that in this organism there has been significant evolutionary drift.

For the catalytic switch loops, switch I is structurally conserved, whereas switch II

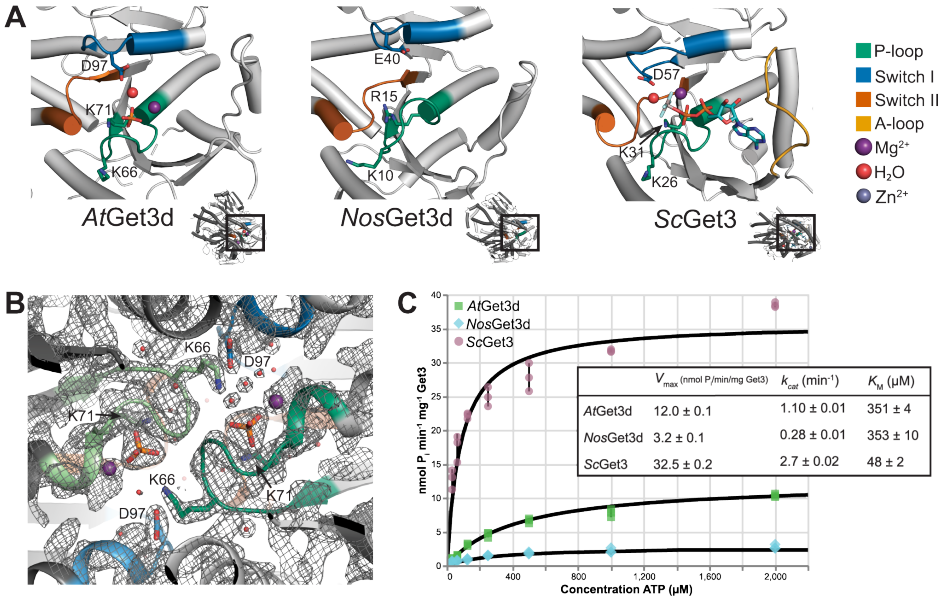


Figure 2.5: ATPase activity of Get3d. (A) The active site and signature Get3 features of a monomer of *At*Get3d, *Nos*Get3d, and the closed conformation of yeast Get3 (*Sc*Get3, PDB ID: 2WOJ) (from *left* to *right*). Signature Get3 features are colored P-loop (*green*), Switch I (*blue*), Switch II (*orange*), A-loop (*yellow*), Mg^{2+} (*purple sphere*), Zn^{2+} (*grey sphere*), and H_2O (*red sphere*). ATP and P_i are shown as *sticks*. For each, the region highlighted is shown on the *right* in the full structure. (B) $2F_O$-F_C electron density (*light grey mesh*) in the active site of *At*Get3d contoured at 1.5σ. Discussed residues are shown as *sticks*. Colored as in (A) showing both P-loop (*dark and light green*) and Switch I motifs (*dark and light blue*). (C) ATPase activity of *At*Get3d, *Nos*Get3d, and yeast Get3 (*Sc*Get3) in nmol P_i min^{-1} mg^{-1} Get3 versus concentration ATP (μM) determined with the EnzCheck Phosphate Assay. Analyzed using ICEKAT (*52*) with kinetic constants reported. Standard deviation shown as error bars. For *Sc*Get3, previously reported k_{cat} is 1.3 ± 0.4 min^{-1} and K_M is 37 ± 6.7 μM (*6*).

is conserved at both the sequence and structural levels (Figs. 2.1B and 2.5, A and B) (*33*, *46*). In P-loop NTPases, switch I and switch II couple structural rearrangements to the presence of the ATP γ-phosphate (*28*). In cytoplasmic Get3 proteins, it has been shown that the highly conserved aspartate in switch I coordinates a water, helping to align the water for nucleophilic attack on the γ-phosphate (Fig. 2.5A) (*46*). This aspartate is conserved in the Get3d family, with *Nos*Get3d uniquely having a glutamate at this position, further supporting its evolutionary drift (Figs. 2.1B and 2.5A). Notably, in *At*Get3d, the water is in a slightly different position likely because of the slightly shorter switch I and the presence of an inorganic phosphate instead of a nucleotide (Fig. 2.5, A and B). The sequence and structural conservation of these features suggests that some nucleotide-dependent structural rearrangements could occur in Get3d.

Both *At*Get3d and *Nos*Get3d are missing features that select for the adenosine nucleoside. Rather than an asparagine in strand β7, which specifically selects for the adenine base (*33*, *46*), both *At*Get3d and *Nos*Get3d have an isoleucine (Fig. S2.5, A–C). Importantly, the A-loop present in cytoplasmic Get3 proteins, which interacts with both the adenine and the ribose, is completely missing in the Get3d structures described here (Figs. 2.1B, 2.5, A and B, and S2.5, A–C). This suggests that Get3d may not be specific for ATP.

In addition, both *At*Get3d and *Nos*Get3d are missing the Zn^{2+}-coordinating CXXC helix (Figs. 2.1B and S2.5D). The Zn^{2+} acts as a rotation point for the conformational changes coordinated with the ATPase cycle. While there are some examples of Get3 proteins missing the CXXC motif (Fig. 2.1B) (*20*, *23*, *53*, *54*), it is not known in these cases how conformational changes are coupled to ATP hydrolysis, which is critical in the targeting cycle (*5*). As Get3d adopts a closed conformation in the absence of nucleotide, there are likely unique conformational changes associated with these proteins.

As Get3d retains most of the components required for nucleotide hydrolysis, it is important to establish that Get3d is an NTPase. To investigate this, the ATPase activities of *At*Get3d and *Nos*Get3d were determined by monitoring the phosphate produced by Get3d in a spectrophotometric assay (Fig. 2.5C). Both *At*Get3d and *Nos*Get3d were found to have ATPase activity, with a V_{max} of 12.0 ± 0.1 and 3.2 ± 0.1 nmol P_i/min/mg Get3 and k_{cat} of 1.10 ± 0.01 and 0.28 ± 0.01 min^{-1}, respectively (Fig. 2.5C). Notably, *At*Get3d and *Nos*Get3d have similar affinities for ATP with a K_M of 351 ± 4 and 353 ± 10 μM, respectively (Fig. 2.5C). To compare this, the ATPase activity of yeast Get3 was determined, a V_{max} of 32.5 ± 0.2 nmol P_i/min/mg Get3, k_{cat} of 2.7 ± 0.02 min^{-1}, and a K_M of 48 ± 2 μM (Fig. 2.5C), consistent with values previously observed (*6*, *46*). Both *At*Get3d and *Nos*Get3d have a lower affinity for ATP and a slower maximum velocity compared with yeast Get3, likely because of the structural differences in the active site and the lack of the A-loop, which facilitates nucleotide binding.

The Hydrophobic Chamber and Binding to a Tail-Anchored Protein Client

The CBD of *At*Get3d is comprised of 10 amphipathic helices (Figs. 2.3A and S2.4A), which form a hydrophobic chamber (Fig. 2.6, A and B). Two crossing helices (α6 using the same numbering convention described previously (*46*)) form the bottom of the chamber, the sides are formed by four helices (α4, α5, α7, and

α9), and the chamber is enclosed on the top by two additional helices (α8) (Fig. 2.3A). The inside of the chamber is characterized by hydrophobic and uncharged amino acids, which would be expected for a binding site of a TMD (Fig. 2.6, A and B). There is sufficient volume in the chamber to accommodate a TMD of ~22 amino acids. To visualize this, we aligned Get3d to the structure of *Giardia intestinalis* Get3 (*Gi*Get3) in complex with the yeast TA protein Bos1 (PDB ID: 7SQ0) (*5*). Here, the TMD of Bos1 fits easily into the chamber (Fig. 2.6C). The chamber of *Nos*Get3d is similar, albeit slightly larger (Figs. 2.6, B and C and S2.10A). The chamber is a unique feature of Get3d, in stark contrast to the hydrophobic groove seen in closed fungal Get3 structures (Figs. 2.3C and S2.10, B and C) and demonstrates a mechanism for closing reminiscent of the *Giardia* Get3–TA complex (*5*, *46*, *55*). As we know that this chamber can accommodate a TA protein, it will be interesting to see how this domain rearranges.

To determine the TA protein binding capability of *At*Get3d and *Nos*Get3d, His_6-tagged Get3d constructs were expressed alone or coexpressed with a maltose-binding protein (MBP)–tagged yeast TA protein, Sbh1. Capture assays were performed by passing the lysate over immobilized metal affinity chromatography followed by elution with imidazole. Stable complexes were observed by Coomassie-stained SDS-PAGE (Fig. 2.6D). Both constructs were able to capture the TA protein, confirmed by western blot (Fig. S2.10D). The specificity of the Get3d and TA interaction was confirmed by repeating the procedure with MBP-TA alone, and as expected, no binding of MBP-TA was observed (Fig. S2.10E). The ability of Get3d to form a complex with a TA protein suggests that Get3d may be involved in TA protein targeting, or, at a minimum, binds TA proteins to protect them from the aqueous environment as a general chaperone. As αCDs have been shown to bind to unfolded proteins (*48*, *49*, *56*), we wanted to see if the αCD was required for TA protein complex formation. Capture assays were performed in a similar manner for both constructs with the αCD removed. These were also able to capture the TA protein (Figs. 2.6D and S2.10D), revealing that the αCD is not required for TA protein binding.

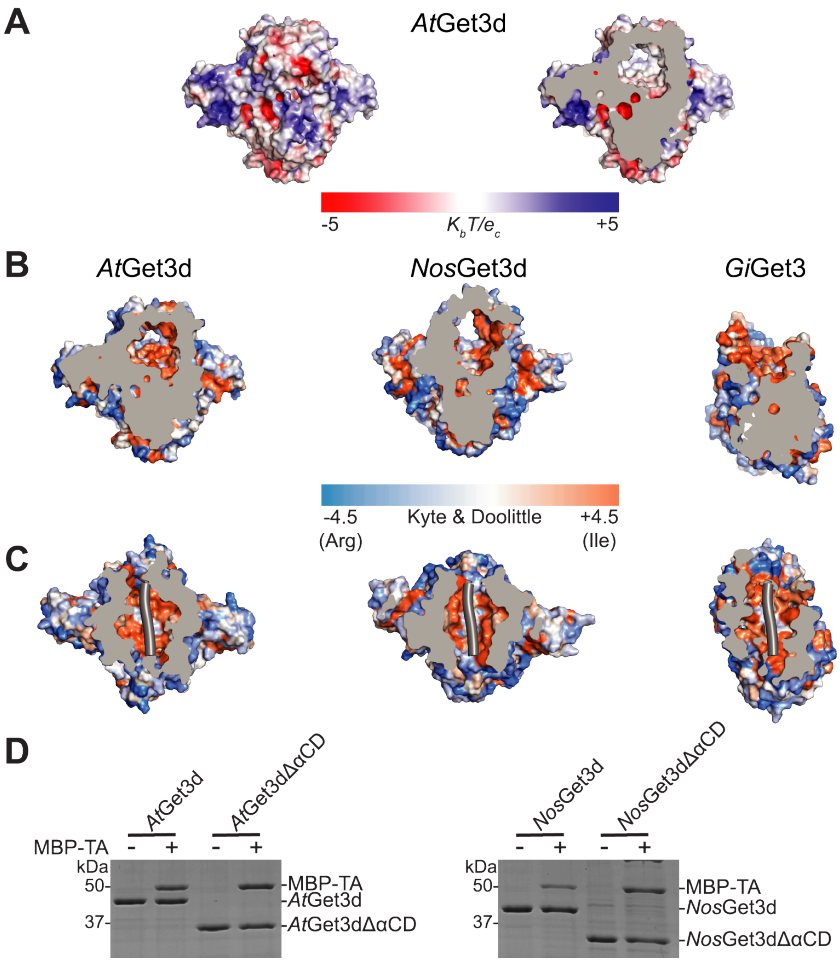


Figure 2.6: Hydrophobic chamber of Get3d. (A) Full (*left*) and slice view (*right*) surface electrostatic potential of *At*Get3d as in Fig. S2.10B. (B) Slabbed view of accessible surface of *At*Get3d, *Nos*Get3d, and *Gi*Get3 (PDB ID: 7SQ0) (from *left* to *right*) colored by hydrophobicity using the Kyte and Doolittle scale. (C) Top-down slice view of the surface hydrophobicity of *At*Get3d, *Nos*Get3d, and *Gi*Get3 in complex with the TMD of the yeast TA protein Bos1 (PDB ID: 7SQ0) (from *left* to *right*). After aligning the three structures, Bos1 is overlaid on the *At*Get3d and *Nos*Get3d structures. Scale as in (B). (D) *In vitro* TA protein capture assays in which His-tagged *At*Get3d and *Nos*Get3d with and without the αCD are expressed in the absence or presence of a maltose binding protein (MBP)-tagged yeast TA protein Sbh1. After purification by Nickel affinity chromatography, the eluate is analyzed by Coomassie stained SDS-PAGE. Molecular weight marker (kDa) shown.

2.4 Discussion

In eukaryotes, the efficient and precise insertion of membrane proteins is an imperative step for their accurate function in various organelles (*1*, *57*, *58*). Errors in targeting may lead to mislocalization of these proteins, which can result in unfavorable cellular effects. Recent work explores the guided entry of tail-anchored proteins (GET) pathway in plants, with all components of the GET pathway excluding Get5 and Sgt2 having been identified (*18–20*, *23*, *59*). However, a striking difference in plants compared with other eukaryotes is the presence of multiple paralogs of Get3 (*8*, *17*, *19*). In *A. thaliana*, four paralogs of Get3 exist, termed Get3a–d. This study is the first to characterize Get3d, a distinct homolog that is conserved across a few billion years of the evolution of photosynthesis from bacteria to plants.

Overall, some Get3d proteins are more similar in sequence to the cytoplasmic Get3 proteins, whereas others are more diverged (Fig. 2.1B) suggesting distinct evolutionary paths. For example, Get3d proteins from representative green sulfur and green non-sulfur bacteria, along with *Synechocystis* sp. Sll0086, contain a highly conserved switch I and A-loop, whereas *At*Get3d, *Nos*Get3d, and *Synechocystis* sp. Slr1794 do not. It will be interesting to investigate the differences between the two *Synechocystis* sp. Get3d proteins as one is more similar to *At*Get3d and *Nos*Get3d than its paralog, which has features closer to cytoplasmic Get3 proteins, implying distinct roles.

Catalytic residues show some interesting variability in Get3d. The canonical Walker A lysine, which coordinates the nucleotide β-phosphates, is broadly conserved in P-loop NTPases (*28*). For Get3d, this lysine is conserved throughout angiosperms, such as *A. thaliana*. Green sulfur (*Chlorobiaceae*) and green non-sulfur (*Chloroflexi*) bacteria also have a lysine at this position (Fig. 2.1B). Certain cyanobacterial Get3d proteins have replaced this with an arginine, which is unique to Get3d proteins that appear early in the cyanobacterial lineage, including simple filamentous cyanobacteria, such as some *Pseudanabaena* species, *Leptolyngbya*, *Halomicronema*, and several clades of synechococcalean cyanobacteria (*39*, *60*). Based on this evidence, the Walker A lysine may have mutated early in the cyanobacterial lineage, whereas green sulfur and non-sulfur bacteria and angiosperms retained the lysine. The catalytic aspartate in Get3, D57 in yeast, coordinates water and primes it for nucleophilic attack of the γ-phosphate of ATP (*33*, *46*). This residue is highly conserved in Get3d proteins; however, it is a glutamate in *Nos*Get3d, which may contribute to its lower V_{max} (Figs. 2.1B and 2.5, A and C).

When examining the structure of Get3d in detail, it is unclear if it should be able to hydrolyze ATP. The variations in Get3d around the active site, such as the slightly shorter loop in switch I and the absence of the A-loop, may preclude Get3d from hydrolyzing ATP and producing the conformational changes that are coupled to TA protein targeting in fungal Get3 (*33*, *46*). Our data reveal that both *At*Get3d and *Nos*Get3d are active ATPases (Fig. 2.5C). The decreased K_{M} of Get3d for ATP when compared with yeast Get3 is likely due to the absence of the A-loop (Figs. 2.5A and S2.5, A–C).

Notably, the ATPase cycle and conformational changes of Get3, which regulate protein targeting, are inextricably linked (*4–6*). As Get3d can hydrolyze ATP (Fig. 2.5C), Get3d likely adopts additional conformations relative to the structures presented here. The structures here are most similar to the closed conformation of Get3, yet they do not have either ATP or an ATP analog bound. The closed conformation of Get3 depends on a bound ATP or ATP analog (e.g., ADP–aluminum fluoride closed yeast dimer [PDB ID: 2WOJ] (*46*)); therefore, the missing nucleotide in the Get3d structures is surprising. For the *At*Get3d structure, the requirement for ADP in the crystallization condition suggests that the phosphate bound in the active site may support a conformational change and could act in a regulatory manner similar to that of ATP binding and hydrolysis. Studies have shown that the concentration of inorganic phosphate in the chloroplast stroma changes drastically with changes in light conditions across the day/night cycle (*61*), and it is conceivable that this fluctuation may regulate Get3d in some manner.

An interesting feature of Get3d is the presence of a hydrophobic chamber in the closed state instead of a client-binding groove as seen in closed fungal Get3 structures, suggesting this may be the default state for Get3d (Figs. 2.6, A–C and S2.10, A–C). Density in the hydrophobic chamber of *At*Get3d consistent with a phospholipid (Fig. S2.6, A and D) is unsurprising as the stable hydrophobic chamber may nonspecifically carry the lipid through purification, although it is possible that Get3d may have a role in lipid binding *in vivo*. Phospholipids are major components of the cytoplasmic membrane in both cyanobacteria and *Escherichia coli* (*62*). The head group of the modeled phosphatidic acid forms specific interactions (Fig. S2.6A); however, only the positive charge of the *At*Get3d R182 is conserved in Get3d proteins overall (Fig. S2.6E). The acyl chain lining the bottom of the hydrophobic chamber is more ordered than the other, which extends into the chamber, likely because of strong interactions with the hydrophobic residues of the groove (Fig. S2.6, A and

D). Further studies will be necessary to determine if there is a physiological role for lipid binding.

The most parsimonious model is that Get3d plays a role in TA protein targeting similar to the cytoplasmic Get3 proteins (*6*, *17*, *46*). In our TA protein capture assays, both *At*Get3d and *Nos*Get3d were able to form a stable complex with a transmembrane domain (Figs. 2.6D and S2.10D), consistent with the structure. Further studies will be necessary to demonstrate a direct role in TA protein targeting. If Get3d does participate in TA protein targeting, it would necessitate new partners as no other GET pathway components have been identified in the chloroplast (*18*, *20*, *23*, *59*). These Get3d partners would likely be conserved in photosynthetic bacteria as well.

The presence of an α-crystallin domain appended to the C-terminus is a unique feature of Get3d proteins (Fig. 2.4A). Many α-crystallins/sHSPs act as ATP-independent chaperones by binding to unfolded proteins to protect cells from damage because of protein aggregation (*56*). Because the αCD of Get3d is not required for TA protein binding (Figs. 2.6D and S2.10D), the role of the αCD is unclear. The presence of the αCD may suggest that *At*Get3d acts as a general chaperone in a manner similar to α-crystallins/sHSPs. While the αCD of Get3d maintains the overall fold of α-crystallins/sHSPs, it lacks the features characteristic of oligomerization. Thus, if the αCD of Get3d does oligomerize, it would require novel architectures. As the interface between the Get3 and αCD of Get3d is similar to that of yeast Get3 and Get4 (Fig. 2.4D), it is possible that the αCD of Get3d stabilizes the closed conformation and/or acts in a regulatory manner similarly to how yeast Get4 regulates ATP hydrolysis by Get3 (*6*, *50*). Another possible function could be that the αCD binds and stabilizes the N-terminal soluble domains of specific TA protein clients; however, not all proteins with αCDs act as chaperones, thus, care must be taken when classifying new αCD-containing proteins (*48*, *49*, *63*). Further investigation is needed to shed light on the significance of αCD of Get3d.

In Get3d, the absence of the CXXC motif and its coordinated Zn^{2+} ion is an important distinction (Figs. 2.1B and S2.5D). While conserved in most Get3 proteins, its presence is not necessary for Get3 activity in all cases (*20*, *23*, *53*). Of note, *At*Get3a, the cytoplasmic Get3 that targets TA proteins to the ER in *A. thaliana*, also lacks the CXXC motif (Fig. 2.1B) (*20*, *23*). In addition to a role in dimerization, the CXXC motif has also been implicated in modulating a secondary function of Get3 as a general chaperone regulated by oxidation (*64*). Get3d may bypass this

requirement to act as a general chaperone.

Evidence supports that *A. thaliana* has two Get3 paralogs localized to the chloroplast: Get3b and Get3d (*14*, *20*). While *At*Get3d appeared early in the evolution of photosynthesis, *At*Get3b is first found with the appearance of chloroplasts, suggesting it was a newly acquired role in plants. A possible role is that *At*Get3b and *At*Get3d are both involved in TA protein targeting with different substrate specificities or different destination membranes (thylakoid versus inner envelope membrane). This is an exciting hypothesis, as *At*Get3b was shown to interact with the thylakoid membrane protein *At*SecE1 but not the inner envelope membrane *At*SecE2 (*14*). While previous work has shown that *At*Get3b localizes to the chloroplast stroma specifically (*14*), we were unable to distinguish between the stroma and thylakoid lumen here. Thus, the possibility that *At*Get3b functions in the stroma and *At*Get3d functions in the thylakoid lumen cannot be ruled out. *At*Get3b and *At*Get3d may also act in different tissues, in different plastid types (e.g., chloroplast, leucoplast, and chromoplast) (*65*), or during different stages of development. As they are both conserved across plants, it will be necessary to determine the roles of *At*Get3b and *At*Get3d.

If Get3d plays a role in TA protein targeting in chloroplasts, possible clients include several essential chloroplast-encoded proteins such as multiple photosystem I and II reaction center components and several cytochrome b_6f proteins (*66*). The pool of substrates may also include nuclear-encoded TA proteins such as SecE1 and SecE2. Many of these proteins are conserved throughout photosynthesis and represent an interesting pool of possible Get3d substrates.

The deep evolutionary connection between the Get3d fold and photosynthesis, while correlative, does not address the function of Get3d. Clearly, the preservation of its presumed function during the endosymbiotic event that created chloroplasts provides evidence that the role is critical for the conversion of light into chemical energy. This function remains to be determined, but as is seen for cytoplasmic Get3 proteins, it is likely either involved in TA protein targeting, acts as a chaperone, or perhaps both. A completely new role is possible, and future work to identify phenotypes and interaction partners should illuminate this puzzle. The breadth of Get3-like proteins scattered across kingdoms leads to questions about where this fold first appeared and what that role may have been.

2.5 Experimental Procedures

A Reference Alignment and Phylogeny of the Get3/ArsA Family

All proteins from UniProt version 2020_06 with an annotation as Get3/ArsA were pulled down (*IPR027542*, *IPR016300*, and *IPR025723*) along with all other InterPro domains identified (*22*, *67*, *68*). Get3/ArsA domains were then identified by searching against hidden Markov models for monomeric Get3/ArsA domains from solved crystal structures, split by hand for pseudodimers. The resulting domains were then searched (jackhammer, three iterations (*69*)) against the preliminary set of UniProt proteins to find additional monomeric representatives. Each hit that covered 90% of the best scoring query was considered complete. Pseudodimers were split midway between the end of the first domain and the start of the second domain on the parent sequence.

A reference alignment of Get3/ArsA domains was then created by clustering domains using mmseqs at 65% sequence identity (*70*). Clusters were then aligned using MAFFT, version 7.471 (*25*), with a seed alignment given by a structural alignment (STAMP) of Get3/ArsA proteins with solved structures (genafpair, maxiterate 1000, and retree 20) (*24*). A maximum-likelihood phylogeny was computed using RAxML, version 8.2.12 (*26*): automatic model assignment using machine learning criterion, best scoring model LG with empirical base frequencies; and rapid bootstrap search complete after 400 replicates (-I autoMRE) (*71*).

Plant, Photosynthetic, and αCD-Containing Get3/ArsA Proteins

Get3/ArsA proteins with αCD domains were identified by a match to Pfam PF17886 (*ArsA_HSP20*) as annotated by UniProt. Organisms were defined as putatively photosynthetic if a UniProt proteome contained more than 10 proteins assigned to the Gene Ontology term GO:0015979 (photosynthesis) (*72*, *73*). Because of uneven sequencing coverage across genomes, analysis of the presence of αCD-containing Get3 proteins in photosynthetic versus nonphotosynthetic organisms was carried out for proteomes annotated by UniProt as reference, nonredundant, or with a BUSCO completeness >75% (*32*).

Get3 proteins were separated into subsets to be clustered (mmseqs easycluster) into representative sequences for phylogenetic analysis with specific minimum sequence identity levels for each group: Get3 proteins from plants (85%), Get3 proteins with αCD domains from photosynthetic bacteria (70%), Get3 proteins without αCD domains from photosynthetic bacteria (65%), and all other Get3/ArsA proteins

(60%). Where one of a pair of a pseudodimer was indicated as a representative, the other half was kept as well. Get3 proteins from *A. thaliana* and those with solved crystal structures were also included. The resulting sequences were then aligned using MAFFT (genafpair, maxiterate 1000, and retree 20) seeded by the structure-based alignment specified previously.

A maximum-likelihood phylogeny was computed using RAxML: automatic model assignment using ML criterion, best scoring model VT with fixed base frequencies; rapid bootstrap search complete after 100 replicates (-I autoMRE); N-terminal only, Get3-only, and C-terminal portion only tested as subpartitions. Nonrepresentative sequences were then added into the alignment using MAFFT (–add, –keeplength) and then placed onto the tree using the RAxML Evolutionary Placement Analysis (-f v) (*74*).

The resulting tree is post facto rooted using the most recent common ancestor of *E. coli* ArsA and *Mj*Get3. Plant proteins are assigned to Get3a, b/c, or d by the presence of the corresponding *A. thaliana* protein in that clade. Trees are manipulated and drawn using phyloseq (*75*), phytools (*76*), treeio (*77*), ggtree (*78*, *79*), tidytree (*79*), and ggplot2 (*80*) packages in R (*81*) and the tidyverse (*82*).

Reagents

All reagents were purchased from Sigma-Aldrich unless otherwise specified.

Cloning

Constructs and primers used in this study are given in Table S2.2.

The *At*Get3d purification and crystallization construct was prepared by inserting *At*Get3dΔ1–57 into the NdeI–XhoI cut sites of pET22b(+) using standard restriction enzyme cloning methods. Based on signal/chloroplast targeting peptide prediction (*41*) and MAFFT sequence alignments (*25*), the first 57 amino acids were truncated from *At*Get3d.

Get3d-GFP fusion constructs were generated by isothermal assembly following standard procedures (*83*). Each assembled construct was in the pENTR plasmid with L1 and L2 sites that allowed for recombination of the Get3d-GFP construct into a destination vector with the UBQ10 promoter regulatory sequence and an OCS terminator sequence (pMOA pUBQ10-GW-OCS).

The *At*Get3d ATPase assay construct was prepared using standard Gibson cloning protocols (*83*) by inserting *At*Get3dΔ1–57 into a pET22b(+) vector containing an

N-terminal His_6 tag and human rhinovirus 3C protease cut site. The pET33b-His_6-TEV-*Nos*Get3d ATPase assay construct was prepared using standard Gibson cloning methods (*83*), and pET33b-His_6-TEV-*Sc*Get3 was prepared as described previously (*33*).

The Get3d TA protein capture assay constructs were prepared using standard Gibson cloning methods (*83*). The MBP-Sbh1 TA protein was prepared as described previously (*84*).

Plant Material

Nicotiana benthamiana seeds were germinated on Sunshine Mix 5 with perlite and vermiculite added at a ratio of 3:1:1, respectively. After seedlings germinated and the first true leaves appeared, plants were transplanted and allowed to grow for 14 days in 16:8 light:dark hr cycle.

***Agrobacteria* Transformation and Tobacco Infiltration of Get3d-GFP Variants**

For transient expression, plasmids pUBQ10::Get3d-GFP and pUBQ10::Get3dΔTP-GFP were introduced into the *Agrobacterium* strain GV3101 by triparental mating. *Agrobacteria* strains were grown in 2xYT media with gentamycin (30 μg/ml), rifampicin (50 μg/ml), and spectinomycin (100 μg/ml) and adjusted to an absorbance of 0.1 in 10 mM $MgCl_2$ and 150 μM acetosyringone. About 1 ml of each *Agrobacterium* sample was infiltrated into each leaf using a 1 ml syringe. Three to four leaves were infiltrated per construct.

Confocal Microscopy

After 48 hrs, tobacco leaf samples were imaged on an upright Zeiss 780 Confocal Laser Scanning Microscope. The 488 nm laser line was used to excite both GFP and induce chlorophyll autofluorescence. Standard excitation and emission windows were used for GFP and chlorophyll b. Microscopy images were processed in ImageJ (*85*, *86*).

***At*Get3dΔ1–57 Expression, Purification, and Crystallization**

*At*Get3dΔ1–57 was expressed in *E. coli* BL21(DE3) Star. Briefly, 100 ml of LB with ampicillin was inoculated with overnight culture (1%) and grown at 37°C. At an absorbance of ~1.0 at 600 nm, protein expression was induced by adding 0.4 mM IPTG, and the culture was incubated overnight at 16°C. After overnight induction, cells were harvested by centrifugation, resuspended in lysis buffer (25 mM

Tris–HCl [pH 7.5], 500 mM NaCl, 30 mM imidazole, 2 mM β-mercaptoethanol, 0.1% Triton X-100, 2 mM $MgCl_2$, and 10% glycerol), and sonicated for 5 min with 5 s on/off pulse and 45% amplitude. Cell debris was pelleted by centrifugation at 13,000 g for 30 min, and the clarified lysate was passed over a pre-equilibrated nickel–nitrilotriacetic acid column (Qiagen). The column was washed with 10 column volumes wash buffer (25 mM Tris–HCl [pH 7.5], 500 mM NaCl, 30 mM imidazole, 2 mM β-mercaptoethanol, 2 mM $MgCl_2$, and 10% glycerol) and eluted by raising to 250 mM imidazole. Elution fractions were pooled and concentrated using Amicon 30k molecular weight cutoff (MWCO) concentrator (Sigma-Aldrich). The concentrated protein was further purified through a pre-equilibrated (25 mM Tris–HCl [pH 7.5], 150 mM NaCl, 2 mM DTT, 1 mM $MgCl_2$, and 10% glycerol) Superdex 200 10/300 GL (Cytiva) size-exclusion column.

Fractions were pooled and further concentrated to 10 mg/ml for crystallization trials by sitting drop vapor diffusion. Crystals were grown at room temperature by mixing equal volume of protein solution containing 2 mM ADP/adenylyl imidodiphosphate with reservoir solution containing 50 mM sodium cacodylate (pH 5.47), 50 mM lithium sulfate, and 30% PEG-4000. Crystals were cryoprotected in the mother liquid supplemented with 30% glycerol before flash-freezing in liquid nitrogen.

Data Collection, Structure Solution, and Refinement

*At*Get3d data collection was done at European Synchrotron Radiation Facility beamline BM-14 at 100 K and 0.97625 Å. The data were integrated with XDS and scaled with Aimless (CCP4 suite) in space group $P\ 1\ 2_1\ 1$ to a resolution of 2.0 Å (*87–89*). Data collection and refinement statistics are listed in Table S2.1.

The *At*Get3d structure was determined by molecular replacement with PHASER using the Get3d homodimer from *Nostoc* sp. PCC 7120 (PDB ID: 3IGF) as the search model (*44*, *90*). Several rounds of model building and refinement were carried out with phenix.refine, CCP4/Refmac, and COOT (*91–95*). A single molecule of an *At*Get3d dimer was found in the asymmetric unit. Side-chain density was generally weak in the α-helical subdomains, and density was missing for residues 250 to 260, 330 to 331, and 378 to 382 in monomer A and residues 252 to 261 and 380 to 384 in monomer B. For *At*Get3d, residue numbering includes the chloroplast targeting peptide.

The *Nostoc* sp. PCC 7120 Get3d structure (PDB ID: 3IGF) was refined using phenix.refine, and manual building was performed in COOT (*91–93*). Residues 160

to 182 from monomer A and 88 to 89 and 167 to 182 from monomer B were added in this study. Residue 195 from monomer A was removed in our refinement because of poor density. Density was missing for residues 183 to 195 in monomer A and 183 to 192 in monomer B. *Nos*Get3d data collection information is available at PDB ID: 3IGF (*44*). Data collection and refinement statistics are listed in Table S2.1.

Structural figures were generated using PyMOL (https://pymol.org/) (*96*).

Native Mass Spectrometry

His_8-SUMO-GSx2-*At*Get3dΔ1–57 was expressed in NiCo21(DE3) *E. coli* in 2xYT. At an absorbance of ~0.7 at 600 nm, protein expression was induced by adding 0.4 mM IPTG, and the culture was incubated for 4 hrs at 37°C. Cells were harvested and lysed in 50 mM Tris–HCl (pH 7.5), 300 mM NaCl, 20 mM imidazole, 10 mM β-mercaptoethanol, 1 mM PMSF, and 1 mM benzamidine in a Microfluidizer (Microfluidic). After cell debris was pelleted, the clarified lysate was passed over 500 μl per 1 L culture of pre-equilibrated nickel–nitrilotriacetic acid resin, washed with 100 column volume wash buffer (50 mM Tris–HCl [pH 7.5], 300 mM NaCl, 20 mM imidazole, 10 mM β-mercaptoethanol, 1 mM PMSF, and 1 mM benzamidine), and eluted with elution buffer (20 mM Tris–HCl [pH 7.5], 150 mM NaCl, 300 mM imidazole, and 10 mM β-mercaptoethanol). The nickel affinity chromatography elution fractions were pooled and dialyzed with 0.02 mg/ml Ulp1 against Get3 buffer (50 mM Hepes [pH 7.5], 150 mM KOAc, 5 mM $Mg(OAc)_2$, and 10 mM β-mercaptoethanol) in 30 kDa MWCO Snakeskin Dialysis Tubing (Thermo Fisher Scientific) overnight at 4°C. Protein was concentrated in an Amicon 30k MWCO concentrator (Sigma-Aldrich) and then purified by size-exclusion chromatography using a Superdex 200 10/300 GL size-exclusion column (Cytiva) in Get3 buffer. Fractions were pooled and concentrated to 50 μM for mass spectrometry analysis.

Samples were prepared as previously described (*97*). In brief, protein samples were buffer exchanged into MS buffer (200 mM ammonium acetate [pH 7.4]) with a centrifugal desalting column (Micro Bio-Spin 6 Columns; Bio-Rad). Mass spectra were collected on a Q Exactive UHMR Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Fisher Scientific). Samples were loaded into a gold-coated borosilicate glass capillary prepared in-house and were introduced into the mass spectrometer via nano electrospray ionization. The mass spectrometry conditions were set as follows: in-source trapping off, high m/z setting for ion transfer target range, injection flatapole DC 5 V, inter flatapole lens 4 V, bent flatapole DC, and transfer multipole

DC were both 0 V. The higher-energy collisional dissociation (HCD) events were set as follows: HCD time 3 ms, purge time 20 ms, HCD field gradient 200 V, and trapping gas pressure was set to 5. Native mass spectra were collected with an m/z range from 500 to 15,000, resolution at 12,500, one microscan, spray voltage of 1.6 kV, capillary temperature of 100°C, and a maximum inject time of 500 ms. The CE varied from 50 V to 200 V. The raw spectra were processed and deconvoluted using UniDec (*98*).

ATPase Activity Assay

His_6-3C-*At*Get3dΔ1–57, His_6-TEV-*Nos*Get3d, or His_6-TEV-*Sc*Get3 (yeast Get3) were expressed and purified as described previously with 1 mM IPTG. Subsequent experiments were carried out in the same fashion for each construct. The nickel affinity chromatography elution fractions were run on 15% SDS-PAGE, pooled fractions were concentrated in an Amicon 30k MWCO concentrator (Sigma-Aldrich), and protein concentration was determined using a NanoDrop spectrophotometer (Thermo Fisher Scientific).

ATPase activity was determined at 37°C using the EnzChek Phosphate Assay Kit (Thermo Fisher Scientific), a microplate spectrophotometric assay that couples inorganic phosphate production to the enzymatic conversion of 2-amino-6-mercapto-7-methyl-purine riboside to ribose 1-phosphate and 2-amino-6-mercapto-7-methylpurine by purine nucleoside phosphorylase (PNP) (*99*). 100 μl reactions were carried out with Get3 (either 15 μM *At*Get3d, 15 μM *Nos*Get3d, or 2.5 μM *Sc*Get3), 0.2 mM 2-amino-6-mercapto-7-methyl-purine riboside, 1 U/l PNP, 5 mM $MgCl_2$, and 0 μM, 37.25 μM, 62.5 μM, 125 μM, 250 μM, 500 μM, 1 mM, or 2 mM ATP in Get3 buffer (described previously). Reactions were initiated with the addition of Get3d or Get3, and the absorbance was measured at 360 nm every 20 sec for a total of 6.67 min using a Tecan Infinite M Nano+ plate reader in 96-well plates (Corning Costar 96-Well Plate; Thermo Fisher Scientific). The method was programmed using Magellan, version 7.2 software (Tecan). Measurements were taken in triplicate at each concentration.

As a control, reactions were performed as described previously with 0 μM, 1.56 μM, 3.13 μM, 6.25 μM, 12.5 μM, 25 μM, 50 μM, or 100 μM P_i without Get3 or ATP. Reactions were initiated with the addition of P_i. Maximal absorbance at 5 min was plotted against concentration P_i (μM), fit with a linear trendline (A_{360} = 0.0020198 × [P_i (μM)] + 0.0043, R^2 = 0.9937), and utilized to determine A_{360}/nmol P_i.

The resultant data were corrected for background absorbance (A_{360} at 0 μM ATP), analyzed using ICEKAT (*52*), and plotted with the Altair Python package (*100*).

Coexpression and Pulldown of TA Protein by *At*Get3d and *Nos*Get3d with and without αCD

pET33b-His_6-TEV-tagged *At*Get3dΔ1–57, *At*Get3dΔ1–57ΔαCD, *Nos*Get3d, or *Nos*Get3dΔαCD were expressed in *E. coli* BL21(DE3) Star with or without the pACYCDuet-MBP-tagged yeast TA protein Sbh1 (*33*, *84*) and purified as described previously with the following changes: 0.2 mM IPTG was used, and cultures were incubated overnight at 18°C. Subsequent experiments were carried out in the same fashion for each construct. The nickel affinity chromatography elution fractions were run on 15% SDS-PAGE, normalizing for the amount of Get3d in each sample by integrating the Get3d band intensity in ImageJ (*85*, *86*). Identity of the bands was confirmed by western blot with the samples run on a 15% SDS-PAGE and then transferred to 0.45 μm nitrocellulose membranes using the Transblot Turbo System (Bio-Rad). The membranes were blocked using 5% dry milk in TTBS (20 mM Tris–HCl [pH 7.5], 150 mM NaCl, 0.1% Tween-20) at room temperature for 1 hr followed by incubation with either an anti-His_5 antibody for Get3d (from mouse) (Sigma-Aldrich; catalog no.: SAB1305538) or an anti-MBP antibody for the TA protein (from mouse) (New England Biolabs; catalog no.: E8032) in 5% dry milk in TTBS overnight at 4°C. Bands were visualized after rinsing with TTBS (3×5 min) and then incubating with anti-mouse antibody (from rabbit) conjugated to Alkaline Phosphatase (Rockland Immunochemicals; catalog no.: 610-4512) (in 5% dry milk in TTBS) for 3 hrs at room temperature. The membranes were rinsed with TTBS (3×5 min) and AP Developing Buffer (100 mM Tris–HCl [pH 9.5], 100 mM NaCl, and 5 mM $MgCl_2$, 1×5 min), developed using AP Substrate (0.33 mg/ml NBT and 0.165 mg/ml BCIP) in AP Developing Buffer, and imaged using a ChemiDoc MP Imaging System (Bio-Rad).

2.6 Acknowledgments

We are grateful to Elliot Meyerowitz for providing access to reagents and equipment for the plant work along with critical advice. We thank Ailiena Maggiolo and Jens Kaiser for technical assistance with structure refinement. We thank members of the Clemons laboratory for discussion and comments. Crystallography data for *At*Get3d were collected at the European Synchrotron Radiation Facility beamline 14. European Synchroton Radiation Facility operations at BM14 were supported

jointly by the European Molecular Biology Laboratory and the Indian Department of Biotechnology.

Work at Caltech was supported by the National Institutes of Health (NIH) grants R01GM097572, R01GM125063, and DP1GM105385 and a Weston Havens Foundation grant to W.M.C. Work at CSIR-NCL was supported by the Indian Department of Biotechnology grant BT/PR11737/BPA/118/32/2014. Work at Texas A&M was supported by NIH grant R01GM139876 to A.D.L. S.M.S. was supported by National Science Foundation Graduate Research fellowship under grant 11444469. A.N.B. and S.M.S. were supported by an NIH/National Research Service Award Training Grant T32GM07616. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

(*1*) Hegde, R. S., and Keenan, R. J. (2022). The mechanisms of integral membrane protein biogenesis. *Nat. Rev. Mol. Cell Biol. 23*, 107–124, DOI: 10.1038/s41580-021-00413-2.

(*2*) Kutay, U., Hartmann, E., and Rapoport, T. A. (1993). A class of membrane proteins with a C-terminal anchor. *Trends Cell Biol. 3*, 72–75, DOI: 10.1016/0962-8924(93)90066-A.

(*3*) Shan, S.-o. (2019). Guiding tail-anchored membrane proteins to the endoplasmic reticulum in a chaperone cascade. *J. Biol. Chem. 294*, 16577–16586, DOI: 10.1074/jbc.REV119.006197.

(*4*) Chio, U. S., Cho, H., and Shan, S.-o. (2017). Mechanisms of tail-anchored membrane protein targeting and insertion. *Annu. Rev. Cell Dev. Biol. 33*, 417–438, DOI: 10.1146/annurev-cellbio-100616-060839.

(*5*) Fry, M. Y., Najdrová, V., Maggiolo, A. O., Saladi, S. M., Doležal, P., and Clemons Jr., W. M. (2022). Structurally derived universal mechanism for the catalytic cycle of the tail-anchored targeting factor Get3. *Nat. Struct. Mol. Biol. 29*, 820–830, DOI: 10.1038/s41594-022-00798-4.

(*6*) Rome, M. E., Rao, M., Clemons Jr., W. M., and Shan, S.-o. (2013). Precise timing of ATPase activation drives targeting of tail-anchored proteins. *Proc. Natl. Acad. Sci. U. S. A. 110*, 7666–7671, DOI: 10.1073/pnas.1222054110.

(*7*) Abell, B. M., and Mullen, R. T. (2011). Tail-anchored membrane proteins: exploring the complex diversity of tail-anchored-protein targeting in plant cells. *Plant Cell Rep. 30*, 137–151, DOI: 10.1007/s00299-010-0925-6.

(*8*) Manu, M., Ghosh, D., Chaudhari, B. P., and Ramasamy, S. (2018). Analysis of tail-anchored protein translocation pathway in plants. *Biochem. Biophys. Rep. 14*, 161–167, DOI: 10.1016/j.bbrep.2018.05.001.

(*9*) Frain, K. M., Gangl, D., Jones, A., Zedler, J. A., and Robinson, C. (2016). Protein translocation and thylakoid biogenesis in cyanobacteria. *Biochim. Biophys. Acta, Bioenerg. 1857*, 266–273, DOI: 10.1016/j.bbabio.2015.08.010.

(*10*) Oborník, M. (2019). Endosymbiotic evolution of algae, secondary heterotrophy and parasitism. *Biomolecules 9*, 266, DOI: 10.3390/biom9070266.

(*11*) Ries, F., Herkt, C., and Willmund, F. (2020). Co-translational protein folding and sorting in chloroplasts. *Plants 9*, 214, DOI: 10.3390/plants9020214.

(*12*) Lee, D. W., Lee, J., and Hwang, I. (2017). Sorting of nuclear-encoded chloroplast membrane proteins. *Curr. Opin. Plant Biol. 40*, 1–7, DOI: 10.1016/j.pbi.2017.06.011.

(*13*) Anderson, S. A., Singhal, R., and Fernandez, D. E. (2019). Membrane-Specific Targeting of Tail-Anchored Proteins SECE1 and SECE2 Within Chloroplasts. *Front. Plant Sci. 10*, 1401, DOI: 10.3389/fpls.2019.01401.

(*14*) Anderson, S. A., Satyanarayan, M. B., Wessendorf, R. L., Lu, Y., and Fernandez, D. E. (2021). A homolog of GuidedEntry of Tail-anchored proteins3 functions in membrane-specific protein targeting in chloroplasts of Arabidopsis. *Plant Cell 33*, 2812–2833, DOI: 10.1093/plcell/koab145.

(*15*) Fry, M. Y., Saladi, S. M., Cunha, A., and Clemons Jr., W. M. (2021). Sequence-based features that are determinant for tail-anchored membrane protein sorting in eukaryotes. *Traffic 22*, 306–318, DOI: 10.1111/tra.12809.

(*16*) El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res. 47*, D427–D432, DOI: 10.1093/nar/gky995.

(*17*) Chartron, J. W., Clemons Jr., W. M., and Suloway, C. J. (2012). The complex process of GETting tail-anchored membrane proteins to the ER. *Current Opinion in Structural Biology 22*, 217–224, DOI: 10.1016/j.sbi.2012.03.001.

(*18*) Bodensohn, U. S., Simm, S., Fischer, K., Jäschke, M., Groß, L. E., Kramer, K., Ehmann, C., Rensing, S. A., Ladig, R., and Schleiff, E. (2019). The intracellular distribution of the components of the GET system in vascular plants. *Biochim. Biophys. Acta, Mol. Cell Res. 1866*, 1650–1662, DOI: 10.1016/j.bbamcr.2019.06.012.

(*19*) Duncan, O., van der Merwe, M. J., Daley, D. O., and Whelan, J. (2013). The outer mitochondrial membrane in higher plants. *Trends Plant Sci. 18*, 207–217, DOI: 10.1016/j.tplants.2012.12.004.

(*20*) Xing, S., Mehlhorn, D. G., Wallmeroth, N., Asseck, L. Y., Kar, R., Voss, A., Denninger, P., Schmidt, V. A. F., Schwarzländer, M., Stierhof, Y.-D., et al. (2017). Loss of GET pathway orthologs in Arabidopsis thaliana causes root hair growth defects and affects SNARE abundance. *Proc. Natl. Acad. Sci. U. S. A. 114*, E1544–E1553, DOI: 10.1073/pnas.1619525114.

(*21*) Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., and Huala, E. (2015). The Arabidopsis information resource: making and mining the "gold standard" annotated reference plant genome. *Genesis 53*, 474–485, DOI: 10.1002/dvg.22877.

(*22*) The UniProt Consortium (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res. 49*, D480–D489, DOI: 10.1093/nar/gkaa1100.

(*23*) Srivastava, R., Zalisko, B. E., Keenan, R. J., and Howell, S. H. (2017). The GET system inserts the tail-anchored protein, SYP72, into endoplasmic reticulum membranes. *Plant Physiol. 173*, 1137–1145, DOI: 10.1104/pp.16.00928.

(*24*) Russell, R. B., and Barton, G. J. (1992). Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. *Proteins: Struct., Funct., Bioinf. 14*, 309–323, DOI: 10.1002/prot.340140216.

(*25*) Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol. 30*, 772–780, DOI: 10.1093/molbev/mst010.

(*26*) Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics 30*, 1312–1313, DOI: 10.1093/bioinformatics/btu033.

(*27*) Chen, C.-m., Misra, T. K., Silver, S., and Rosen, B. P. (1986). Nucleotide sequence of the structural genes for an anion pump. The plasmid-encoded arsenical resistance operon. *J. Biol. Chem. 261*, 15030–15038, DOI: 10.1016/S0021-9258(18)66824-3.

(*28*) Leipe, D. D., Wolf, Y. I., Koonin, E. V., and Aravind, L. (2002). Classification and evolution of P-loop GTPases and related ATPases. *J. Mol. Biol. 317*, 41–72, DOI: 10.1006/jmbi.2001.5378.

(*29*) Rosen, B. P., Weigel, U., Karkaria, C., and Gangola, P. (1988). Molecular characterization of an anion pump. The arsA gene product is an arsenite (antimonate)-stimulated ATPase. *J. Biol. Chem. 263*, 3067–3070, DOI: 10.1016/S0021-9258(18)69034-9.

(*30*) Zhou, T., Radaev, S., Rosen, B. P., and Gatti, D. L. (2000). Structure of the ArsA ATPase: the catalytic subunit of a heavy metal resistance pump. *The EMBO Journal*, DOI: 10.1093/emboj/19.17.4838.

(*31*) Koonin, E. V. (1993). A superfamily of ATPases with diverse functions containing either classical or deviant ATP-binding motif. *J. Mol. Biol. 229*, 1165–1174, DOI: 10.1006/jmbi.1993.1115.

(*32*) Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics 25*, 1189–1191, DOI: 10.1093/bioinformatics/btp033.

(*33*) Suloway, C. J., Chartron, J. W., Zaslaver, M., and Clemons Jr., W. M. (2009). Model for eukaryotic tail-anchored protein binding based on the structure of Get3. *Proceedings of the National Academy of Sciences 106*, 14849–14854, DOI: 10.1073/pnas.0907522106.

(*34*) Maggiolo, A. O., Mahajan, S., Rees, D. C., and Clemons Jr., W. M. (2023). Intradimeric walker A ATPases: conserved features of a functionally diverse family. *J. Mol. Biol. 435*, 167965, DOI: 10.1016/j.jmb.2023.167965.

(*35*) Serres, M. H., Kerr, A. R., McCormack, T. J., and Riley, M. (2009). Evolution by leaps: gene duplication in bacteria. *Biol. Direct 4*, 1–17, DOI: 10.1186/1745-6150-4-46.

(*36*) Raymond, J., and Blankenship, R. E. (2003). Horizontal gene transfer in eukaryotic algal evolution. *Proc. Natl. Acad. Sci. U. S. A. 100*, 7419–7420, DOI: 10.1073/pnas.1533212100.

(*37*) Zhaxybayeva, O., Gogarten, J. P., Charlebois, R. L., Doolittle, W. F., and Papke, R. T. (2006). Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. *Genome Res. 16*, 1099–1108, DOI: 10.1101/gr.5322306.

(*38*) Shi, T., and Falkowski, P. G. (2008). Genome evolution in cyanobacteria: the stable core and the variable shell. *Proc. Natl. Acad. Sci. U. S. A. 105*, 2510–2515, DOI: 10.1073/pnas.0711165105.

(*39*) Moore, K. R., Magnabosco, C., Momper, L., Gold, D. A., Bosak, T., and Fournier, G. P. (2019). An expanded ribosomal phylogeny of cyanobacteria supports a deep placement of plastids. *Front. Microbiol. 10*, 1612, DOI: 10.3389/fmicb.2019.01612.

(*40*) Zimorski, V., Ku, C., Martin, W. F., and Gould, S. B. (2014). Endosymbiotic theory for organelle origins. *Curr. Opin. Microbiol. 22*, 38–48, DOI: 10.1016/j.mib.2014.09.008.

(*41*) Armenteros, J. J. A., Salvatore, M., Emanuelsson, O., Winther, O., Von Heijne, G., Elofsson, A., and Nielsen, H. (2019). Detecting sequence signals in targeting peptides using deep learning. *Life Sci. Alliance 2*, DOI: 10.26508/lsa.201900429.

(*42*) Gelvin, S. B. (2017). Integration of Agrobacterium T-DNA into the plant genome. *Annu. Rev. Genet. 51*, 195–217, DOI: 10.1146/annurev-genet-120215-035320.

(*43*) Köhler, R. H. (1998). GFP for in vivo imaging of subcellular structures in plant cells. *Trends Plant Sci. 3*, 317–320, DOI: 10.1016/S1360-1385(98)01276-X.

(*44*) Forouhar, F., Abashidze, M., Seetharaman, J., Mao, M., Xiao, R., Ciccosanti, C., and et al. Crystal Structure of the All4481 protein from Nostoc sp. PCC 7120, Northeast Structural Genomics Consortium Target NsR300, 2009, DOI: 10.2210/pdb3IGF/pdb.

(*45*) Fujisawa, T., Narikawa, R., Maeda, S.-i., Watanabe, S., Kanesaki, Y., Kobayashi, K., Nomata, J., Hanaoka, M., Watanabe, M., Ehira, S., et al. (2017). CyanoBase: a large-scale update on its 20th anniversary. *Nucleic Acids Res. 45*, D551–D554, DOI: 10.1093/nar/gkw1131.

(*46*) Mateja, A., Szlachcic, A., Downing, M. E., Dobosz, M., Mariappan, M., Hegde, R. S., and Keenan, R. J. (2009). The structural basis of tail-anchored membrane protein recognition by Get3. *Nature 461*, 361–366, DOI: 10.1038/nature08319.

(*47*) Fahy, E., Sud, M., Cotter, D., and Subramaniam, S. (2007). LIPID MAPS online tools for lipid research. *Nucleic Acids Res. 35*, W606–W612, DOI: 10.1093/nar/gkm324.

(*48*) Basha, E., O'Neill, H., and Vierling, E. (2012). Small heat shock proteins and α-crystallins: dynamic proteins with flexible functions. *Trends Biochem. Sci. 37*, 106–117, DOI: 10.1016/j.tibs.2011.11.005.

(*49*) Waters, E. R., and Vierling, E. (2020). Plant small heat shock proteins-evolutionary and functional diversity. *New Phytol. 227*, 24–37, DOI: 10.1111/nph.16536.

(*50*) Gristick, H. B., Rao, M., Chartron, J. W., Rome, M. E., Shan, S.-o., and Clemons Jr., W. M. (2014). Crystal structure of ATP-bound Get3-Get4-Get5 complex reveals regulation of Get3 by Get4. *Nat. Struct. Mol. Biol. 21*, 437–442, DOI: 10.1038/nsmb.2813.

(*51*) Bange, G., and Sinning, I. (2013). SIMIBI twins in protein targeting and localization. *Nat. Struct. Mol. Biol. 20*, 776–780, DOI: 10.1038/nsmb.2605.

(*52*) Olp, M. D., Kalous, K. S., and Smith, B. C. (2020). ICEKAT: an interactive online tool for calculating initial rates from continuous enzyme kinetic traces. *BMC Bioinf. 21*, 1–12, DOI: 10.1186/s12859-020-3513-y.

(*53*) Suloway, C. J., Rome, M. E., and Clemons Jr., W. M. (2012). Tail-anchor targeting by a Get3 tetramer: the structure of an archaeal homologue. *The EMBO Journal 31*, 707–719, DOI: 10.1038/emboj.2011.433.

(*54*) Farkas, Á., De Laurentiis, E. I., and Schwappach, B. (2019). The natural history of Get3-like chaperones. *Traffic 20*, 311–324, DOI: 10.1111/tra.12643.

(*55*) Mateja, A., Paduch, M., Chang, H.-Y., Szydlowska, A., Kossiakoff, A. A., Hegde, R. S., and Keenan, R. J. (2015). Structure of the Get3 targeting factor in complex with its membrane protein cargo. *Science 347*, 1152–1155, DOI: 10.1126/science.1261671.

(*56*) Haslbeck, M., and Vierling, E. (2015). A first line of stress defense: small heat shock proteins and their function in protein homeostasis. *J. Mol. Biol. 427*, 1537–1548, DOI: 10.1016/j.jmb.2015.02.002.

(*57*) Fry, M. Y., and Clemons Jr., W. M. (2018). Complexity in targeting membrane proteins. *Science 359*, 390–391, DOI: 10.1126/science.aar5992.

(*58*) Kunze, M., and Berger, J. (2015). The similarity between N-terminal targeting signals for protein import into different organelles and its evolutionary relevance. *Front. Physiol. 6*, 259, DOI: 10.3389/fphys.2015.00259.

(*59*) Asseck, L. Y., Mehlhorn, D. G., Monroy, J. R., Ricardi, M. M., Breuninger, H., Wallmeroth, N., Berendzen, K. W., Nowrousian, M., Xing, S., Schwappach, B., et al. (2021). Endoplasmic reticulum membrane receptors of the GET pathway are conserved throughout eukaryotes. *Proc. Natl. Acad. Sci. U. S. A. 118*, e2017636118, DOI: 10.1073/pnas.2017636118.

(*60*) Mareš, J., Struneckỳ, O., Bučinská, L., and Wiedermannová, J. (2019). Evolutionary patterns of thylakoid architecture in cyanobacteria. *Front. Microbiol. 10*, 277, DOI: 10.3389/fmicb.2019.00277.

(*61*) Santarius, K. A., and Heber, U. (1965). Changes in the intracellular levels of ATP, ADP, AMP and Pi and regulatory function of the adenylate system in leaf cells during photosynthesis. *Biochim. Biophys. Acta, Biophysics including Photosynthesis 102*, 39–54, DOI: 10.1016/0926-6585(65)90201-3.

(*62*) Hölzl, G., and Dörmann, P. (2019). Chloroplast lipids and their biosynthesis. *Annu. Rev. Plant Biol. 70*, 51–81, DOI: 10.1146/annurev-arplant-050718-100202.

(*63*) Scharf, K.-D., Siddique, M., and Vierling, E. (2001). The expanding family of Arabidopsis thaliana small heat stress proteins and a new family of proteins containing α-crystallin domains (Acd proteins). *Cell Stress Chaperones 6*, 225, DOI: 10.1379/1466-1268(2001)006<0225:tefoat>2.0.co;2.

(*64*) Voth, W., Schick, M., Gates, S., Li, S., Vilardi, F., Gostimskaya, I., Southworth, D. R., Schwappach, B., and Jakob, U. (2014). The protein targeting factor Get3 functions as ATP-independent chaperone under oxidative stress conditions. *Mol. Cell 56*, 116–127, DOI: 10.1016/j.molcel.2014.08.017.

(*65*) Choi, H., Yi, T., and Ha, S.-H. (2021). Diversity of plastid types and their interconversions. *Front. Plant Sci. 12*, 692024, DOI: 10.3389/fpls.2021.692024.

(*66*) Dobrogojski, J., Adamiec, M., and Luciński, R. (2020). The chloroplast genome: a review. *Acta Physiol. Plant. 42*, 98, DOI: 10.1007/s11738-020-03089-x.

(*67*) Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., et al. (2021). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res. 49*, D344–D354, DOI: 10.1093/nar/gkaa977.

(*68*) Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L., Tosatto, S. C., Paladin, L., Raj, S., Richardson, L. J., et al. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Res. 49*, D412–D419, DOI: 10.1093/nar/gkaa913.

(*69*) Eddy, S. HMMER User's Guide, available at http://hmmer.wustl.edu/; HHMI / Washington University School of Medicine, Saint Louis, Missouri, 2020.

(*70*) Steinegger, M., and Söding, J. (2018). Clustering huge protein sequence sets in linear time. *Nat. Commun. 9*, 2542, DOI: 10.1038/s41467-018-04964-5.

(*71*) Pattengale, N. D., Alipour, M., Bininda-Emonds, O. R., Moret, B. M., and Stamatakis, A. (2010). How many bootstrap replicates are necessary? *J. Comput. Biol. 17*, 337–354, DOI: 10.1089/cmb.2009.0179.

(*72*) Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., et al. (2000). Gene ontology: tool for the unification of biology. *Nat. Genet. 25*, 25–29, DOI: 10.1038/75556.

(*73*) The Gene Ontology Consortium (2021). The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res. 49*, D325–D334.

(*74*) Berger, S. A., Krompass, D., and Stamatakis, A. (2011). Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. *Syst. Biol. 60*, 291–302, DOI: 10.1093/sysbio/syr010.

(*75*) McMurdie, P. J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One 8*, e61217, DOI: 10.1371/journal.pone.0061217.

(*76*) Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.*, 217–223, DOI: 10.1111/j.2041-210X.2011.00169.x.

(*77*) Wang, L.-G., Lam, T. T.-Y., Xu, S., Dai, Z., Zhou, L., Feng, T., Guo, P., Dunn, C. W., Jones, B. R., Bradley, T., et al. (2020). Treeio: an R package for phylogenetic tree input and output with richly annotated and associated data. *Mol. Biol. Evol. 37*, 599–603, DOI: 10.1093/molbev/msz240.

(*78*) Yu, G. (2020). Using ggtree to visualize data on tree-like structures. *Curr. Protoc. Bioinf. 69*, e96, DOI: 10.1002/cpbi.96.

(*79*) Yu, G., *Data Integration, Manipulation and Visualization of Phylogenetic Trees*, 1st edition, available at https://yulab-smu.top/treedata-book/; Chapman and Hall: Boca Raton, Florida, 2022.

(*80*) Wickham, H., Danielle, N., and Pedrsen, T. L., *ggplot2: Elegant Graphics for Data Analysis*, 3rd edition, available at https://ggplot2.tidyverse.org/; Springer: New York City, New York, 2016.

(*81*) R Core Team. R: A Language and Environment for Statistical Computing, available at https://www.R-project.org/; R Foundation for Statistical Computing, Vienna, Austria, 2021.

(*82*) Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the Tidyverse. *J. Open Source Softw. 4*, 1686, DOI: 10.21105/joss.01686.

(*83*) Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison III, C. A., and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods 6*, 343–345, DOI: 10.1038/nmeth.1318.

(*84*) Lin, K.-F., Fry, M. Y., Saladi, S. M., and Clemons Jr., W. M. (2021). Molecular basis of tail-anchored integral membrane protein recognition by the cochaperone Sgt2. *J. Biol. Chem. 296*, DOI: 10.1016/j.jbc.2021.100441.

(*85*) Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods 9*, 676–682, DOI: 10.1038/nmeth.2019.

(*86*) Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods 9*, 671–675, DOI: 10.1038/nmeth.2089.

(*87*) Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 62*, 72–82, DOI: 10.1107/S0907444905036693.

(*88*) Evans, P. R., and Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallogr., Sect. D: Biol. Crystallogr. 69*, 1204–1214, DOI: 10.1107/S0907444913000061.

(*89*) Kabsch, W. (2010). XDS. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 66*, 125–132, DOI: 10.1107/S0907444909047337.

(*90*) McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007). Phaser crystallographic software. *Appl. Crystallogr. 40*, 658–674, DOI: 10.1107/S0021889807021206.

(*91*) Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012). Towards automated crystallographic structure refinement with phenix. refine. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 68*, 352–367, DOI: 10.1107/S0907444912001308.

(*92*) Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 60*, 2126–2132, DOI: 10.1107/S0907444904019158.

(*93*) Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 66*, 486–501, DOI: 10.1107/S0907444910007493.

(*94*) Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 67*, 355–367, DOI: 10.1107/S0907444911001314.

(*95*) Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 67*, 235–242, DOI: 10.1107/S0907444910045749.

(*96*) Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 2.5.0, available at http://pymol.org/, 2021.

(*97*) Laganowsky, A., Reading, E., Hopper, J. T., and Robinson, C. V. (2013). Mass spectrometry of intact membrane protein complexes. *Nat. Protoc. 8*, 639–651, DOI: 10.1038/nprot.2013.024.

(*98*) Marty, M. T., Baldwin, A. J., Marklund, E. G., Hochberg, G. K., Benesch, J. L., and Robinson, C. V. (2015). Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Anal. Chem. 87*, 4370–4376, DOI: 10.1021/acs.analchem.5b00140.

(*99*) Webb, M. R. (1992). A continuous spectrophotometric assay for inorganic phosphate and for measuring phosphate release kinetics in biological systems. *Proc. Natl. Acad. Sci. U. S. A. 89*, 4884–4887, DOI: 10.1073/pnas.89.11.4884.

(*100*) VanderPlas, J., Granger, B., Heer, J., Moritz, D., Wongsuphasawat, K., Satyanarayan, A., Lees, E., Timofeev, I., Welsh, B., and Sievert, S. (2018). Altair: interactive statistical visualizations for Python. *J. Open Source Softw. 3*, 1057, DOI: 10.21105/joss.01057.

2.7 Supporting Information

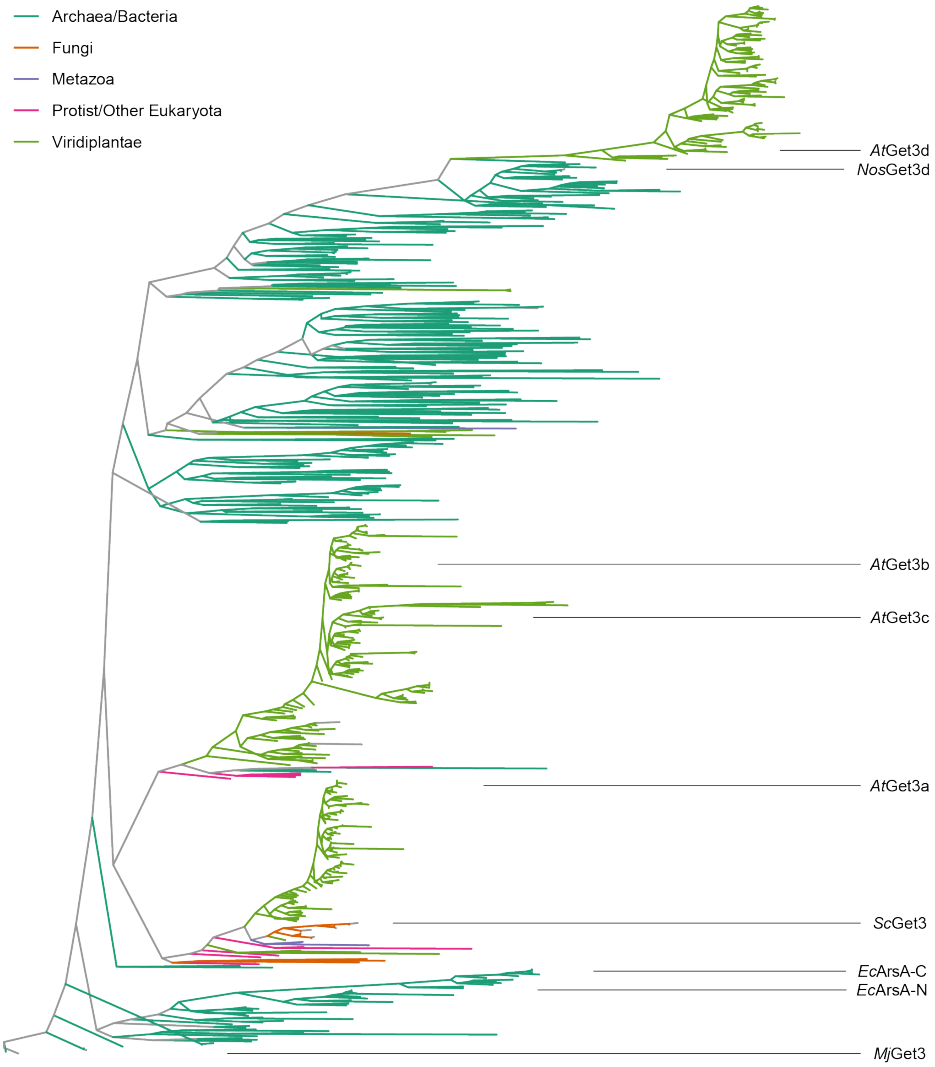
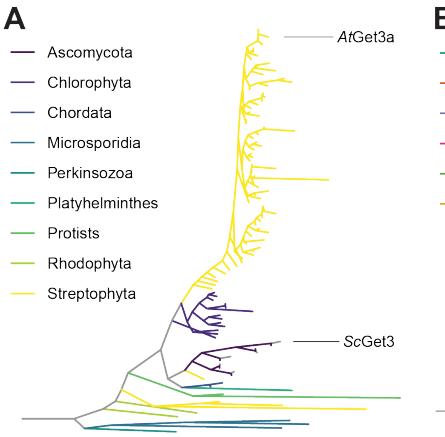
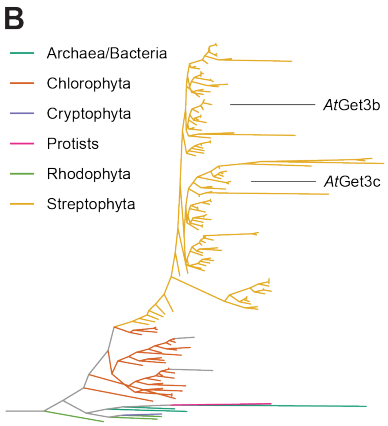
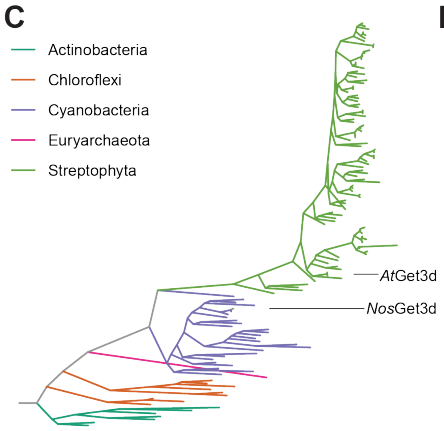


Figure S2.1: Comprehensive phylogeny of Get3 homologs. Get3 proteins from notable species are labeled to the *right* with species abbreviations as in Fig. 2.1. Branches are colored by taxonomic grouping: Archaea/Bacteria (*blue-green*), Fungi (*orange*), Metazoa (*purple*), protists/other Eukaryota (*pink*), Viridiplantae (*green*). Inner branches are colored where all descendants are of a single taxonomic grouping.







E

	Number of Get3/ArsA Proteins	
	without αCD	with αCD
Chlorobium chlorochromatii	0	5
Roseiflexus sp.	0	2
Gloeobacter violaceus	0	2
Nostoc punctiforme	1	2
Trichodesmium erythraeum	1	2
Prochlorococcus marinus	0	0
Synechococcus sp.	0	2
Gemmatimonas phototrophica	1	0
Acidithiobacillus ferrooxidans	1	0
Rhodopseudomonas palustris	0	0
Roseateles depolymerans	0	0
Allochromatium warmingii	0	0

D

		Number of Genomes with a Get3/ArsA		
Phylum	Family	without αCD	with αCD	
Chlorobi	*Chlorobiaceae*	0	2	Green sulfur
Chloroflexi	*Chloroflexia*	0	1	Green non-sulfur
Cyanobacteria	*Chroococcidiopsidales*	0	2	
	Gloeobacteria	0	2	
	Gloeoemargaritales	0	1	
	Nostocales	0	33	
	Oscillatoriophycideae	0	37	
	Pleurocapsales	0	4	
	Synechococcales	0	24	
Proteobacteria	*Acidithiobacillia*	1	0	
	Alphaproteobacteria	8	1	Purple non-sulfur
	Betaproteobacteria	13	0	
	Gammaproteobacteria	7	0	Purple sulfur
Gemmatimonadetes	*Gemmatimonadales*	1	0	

Figure S2.2: Detailed phylogenetic analysis of Get3 homologs. Get3 proteins from notable species are labeled to the *right* with species abbreviations as in Fig. 2.1. Phylogenetic tree of (A) Get3a, (B) Get3b/c, and (C) Get3d paralogs, colored by taxonomy. Phyla are colored as follows: (A) Ascomycota (*dark purple*), Chlorophyta (*purple*), Chordata (*blue*), Microsporidia (*dark blue-green*), Perkinsozoa (*light blue-green*), Platyhelminthes (*dark green*), protists (*green*), Rhodophyta (*light green*), Streptophyta (*yellow*); (B) Archaea/Bacteria (*dark green*), Chlorophyta (*dark orange*), Cryptophyta (*purple*), protists (*pink*), Rhodophyta (*green*), Streptophyta (*orange*); (C) Actinobacteria (*blue-green*), Chloroflexi (*orange*), Cyanobacteria (*light purple*), Euryarchaeota (*pink*), Streptophyta (*green*). Inner branches are colored where all descendants are of a single taxonomic grouping. (D) Number of reference/complete genomes that contain Get3/ArsA homologs with and without the characteristic αCD in various phyla. (E) Number of Get3/ArsA homologs with and without the αCD encoded by representative species from the phyla in (D).

Table S2.1: Comparison of data collection and refinement statistics for *At*Get3d and *Nos*Get3d crystal structures.

	***At*Get3d PDB ID: 8ELF**	***Nos*Get3d PDB ID: 8EGK†**	***Nos*Get3d PDB ID: 3IGF††**
Data Collection			
Space Group	$P\ 1\ 2_1\ 1$		$C\ 1\ 2\ 1$
Cell dimensions			
a, b, c (Å)	59.25, 67.05, 99.40		124.17, 55.55, 122.36
α, β, γ (°)	90.00, 97.81, 90.00		90.00, 98.87, 90.00
Resolution (Å)	58.70 - 2.00 (2.05 - 2.00)*		30.00 - 2.00 (2.07 - 2.00)
R_{merge}	0.061 (1.863)		0.063 (0.288)
I/σ	10.3 (0.6)		18.8 (3.90)
Completeness (%)	99.42 (99.25)		95.22 (91.54) **
Redundancy	4.2 (4.2)		3.3 (3.1)
Refinement			
Resolution (Å)	32.83 - 2.00	27.52 - 1.98	19.96 - 2.00
No. reflections	52,063	54,162	53,378 **
R_{work} / R_{free}	0.215 / 0.258	0.189 / 0.221	0.189 / 0.233
No. non-hydrogen atoms	6112	5873	5680
Protein	5955	5528	5207
Ligand	116	6	
Solvent	100	339	473
Avg. B factors	59.52	42.26	36.76
Protein	59.63	42.55	36.46
Ligand	75.78	41.34	
Solvent	43.66	37.47	40.08
R.m.s. deviations			
Bond length (Å)	0.004	0.007	0.017
Bond angles (°)	0.62	0.76	1.71
Validation			
MolProbity Score	1.35	1.15	2.49
Rotamer outliers (%)	0.92	0.33	8.13
Cβ outliers	0	0	1
Ramachandran outliers (%)	0.00	0.00	0.15
Ramachandran favored (%)	97.47	97.86	96.97
Ramachandran allowed (%)	2.53	2.42	2.88
Rama-Z	-1.21 ± 0.27	-0.90 ± 0.30	-0.97 ± 0.30

*Note: At*Get3d dataset was collected from a single crystal and structure was determined by molecular replacement using PDB ID: 3IGF as the search model. †*Nos*Get3d re-refinement and validation statistics from this study. ††*Nos*Get3d data collection, refinement, and validation statistics from PDB ID: 3IGF. *Values in parentheses are for the highest resolution shell. **Values determined using Phenix.

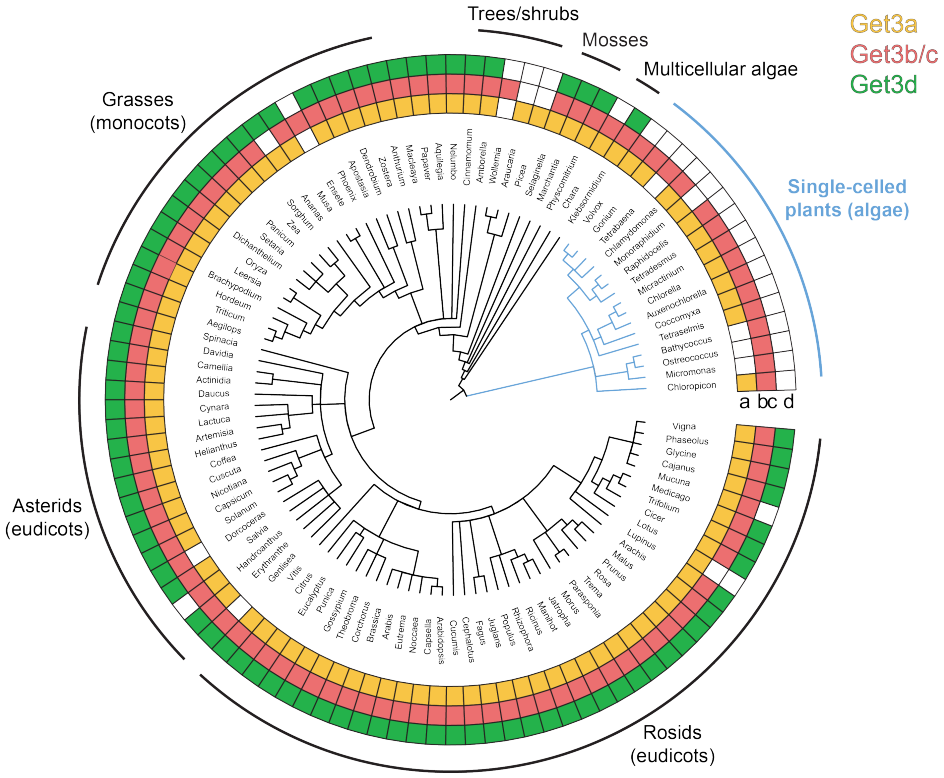


Figure S2.3: Detailed cladogram of Get3 homologs in plants. The presence of Get3 homologs is shown as in Fig. 2.1C. Genera are labeled on the interior of the circle. Selected taxa/clades are labeled on the exterior of the circle.

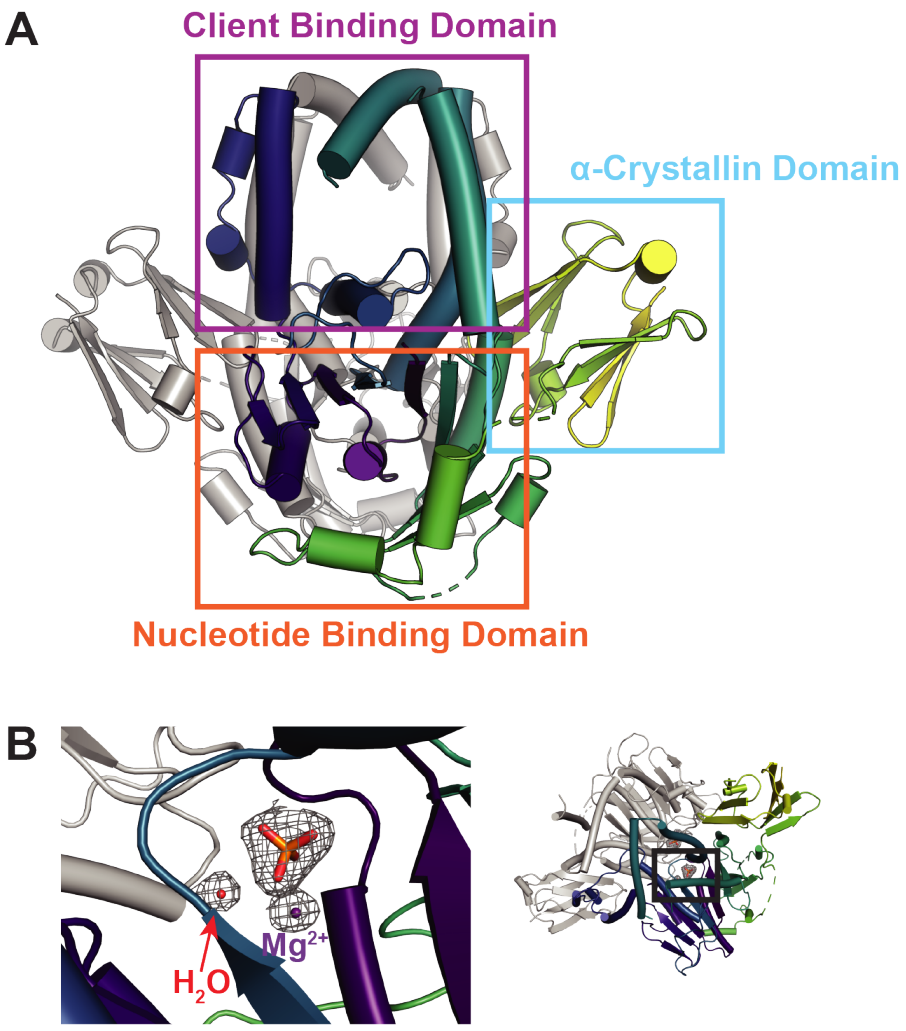


Figure S2.4: Features of the *At*Get3d structure. (A) View from the front of *At*Get3d. The client binding, nucleotide binding, and α-crystallin domains are indicated. (B) Active site of *At*Get3d showing $2F_O$-F_C electron density at 1σ (*grey mesh*) for water (*red sphere*), Mg^{2+} (*purple sphere*), and phosphate (*sticks*). The region highlighted is shown on the *right* in the full structure. *At*Get3d colored as in Fig. 2.3.

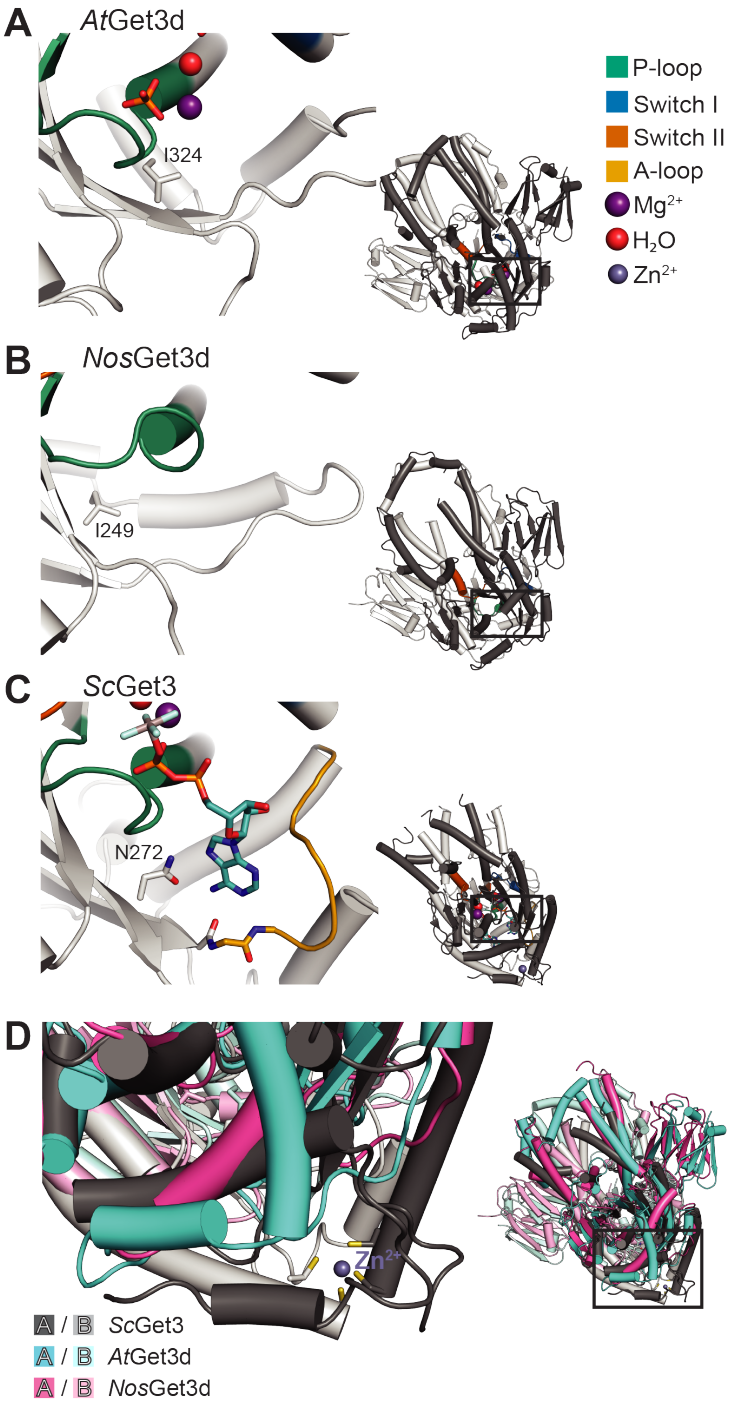


Figure S2.5: View of the missing A-loop and CXXC motif. Structure of the active site of a monomer of (A) *At*Get3d, (B) *Nos*Get3d, and (C) the closed conformation of yeast Get3 (*Sc*Get3, PDB ID: 2WOJ) showing the A-loop and asparagine missing in Get3d. Colored as in Fig. 2.5. (D) Structure of the CXXC motif of the closed conformation of *Sc*Get3 (PDB ID: 2WOJ, *dark and light grey*), with cysteine residues (*sticks*) coordinating a Zn^{2+} ion (*grey sphere*). The structures of *At*Get3d (*dark and light aquamarine*) and *Nos*Get3d (*dark and light pink*) are aligned. The region highlighted is shown on the *right* in the full alignment.

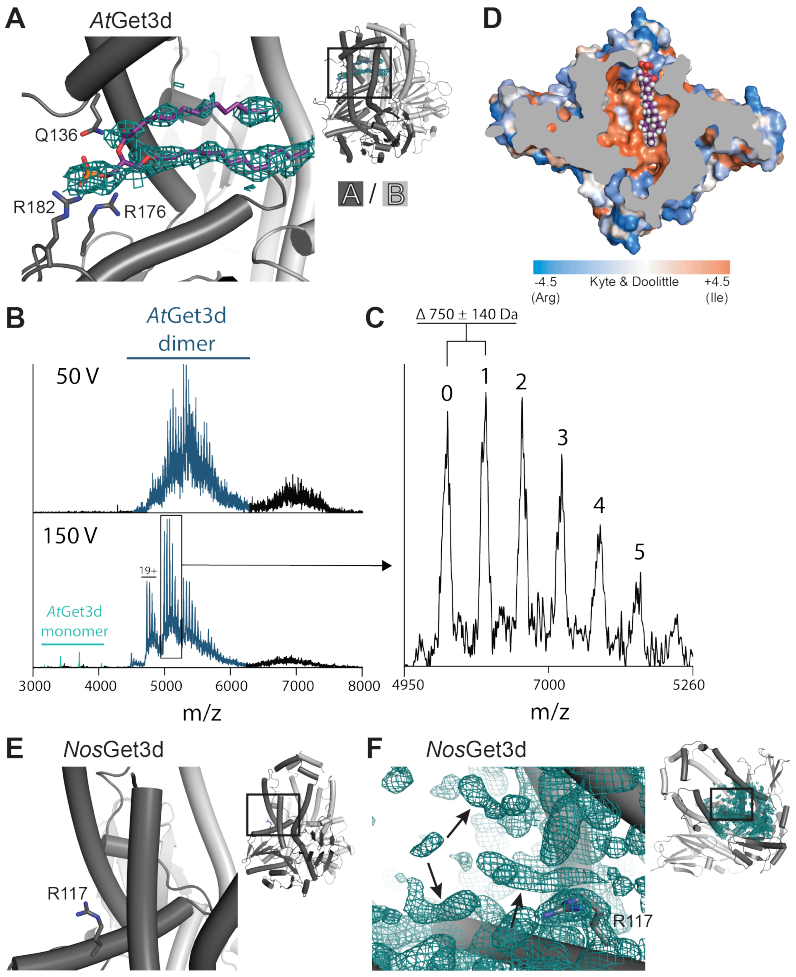


Figure S2.6: Possible lipids in the hydrophobic chamber of Get3d. (A) $2F_O$-F_C electron density (*teal mesh*) in the bottom of the hydrophobic chamber of *At*Get3d near the refined phosphatidic acid (*sticks* with two fourteen carbon saturated acyl chains with *purple* carbons) contoured at 1σ. Monomer A and B are shown in *grey* and *light grey*, respectively. Side chains for residues that coordinate the putative phosphate head group are shown as *sticks*. The region highlighted is shown on the *right* in the full structure. (B) Native mass spectra of *At*Get3d acquired under different activation energies. The theoretical and measured dimer masses are 88,878 Da and 89,944 Da, respectively. (C) Deconvolution of the mass spectrum shown in (B) acquired at CE of 150V. Multiple adducts are bound to the protein with a mass of 750±140 Da. (D) Slice through view of the *At*Get3d accessible surface in the bottom of the hydrophobic chamber colored by hydrophobicity using the Kyte and Doolittle scale with the phosphatidic acid shown as *spheres*. (E) Structure of *Nos*Get3d showing the conserved arginine in *sticks*. The region highlighted is shown on the *right* in the full structure. (F) $2F_O$-F_C electron density map (*teal mesh*) in the bottom of the hydrophobic chamber of *Nos*Get3d contoured at 1σ. Arrows point to densities that were unidentified in our refinement. (E) and (F) colored as in (A).

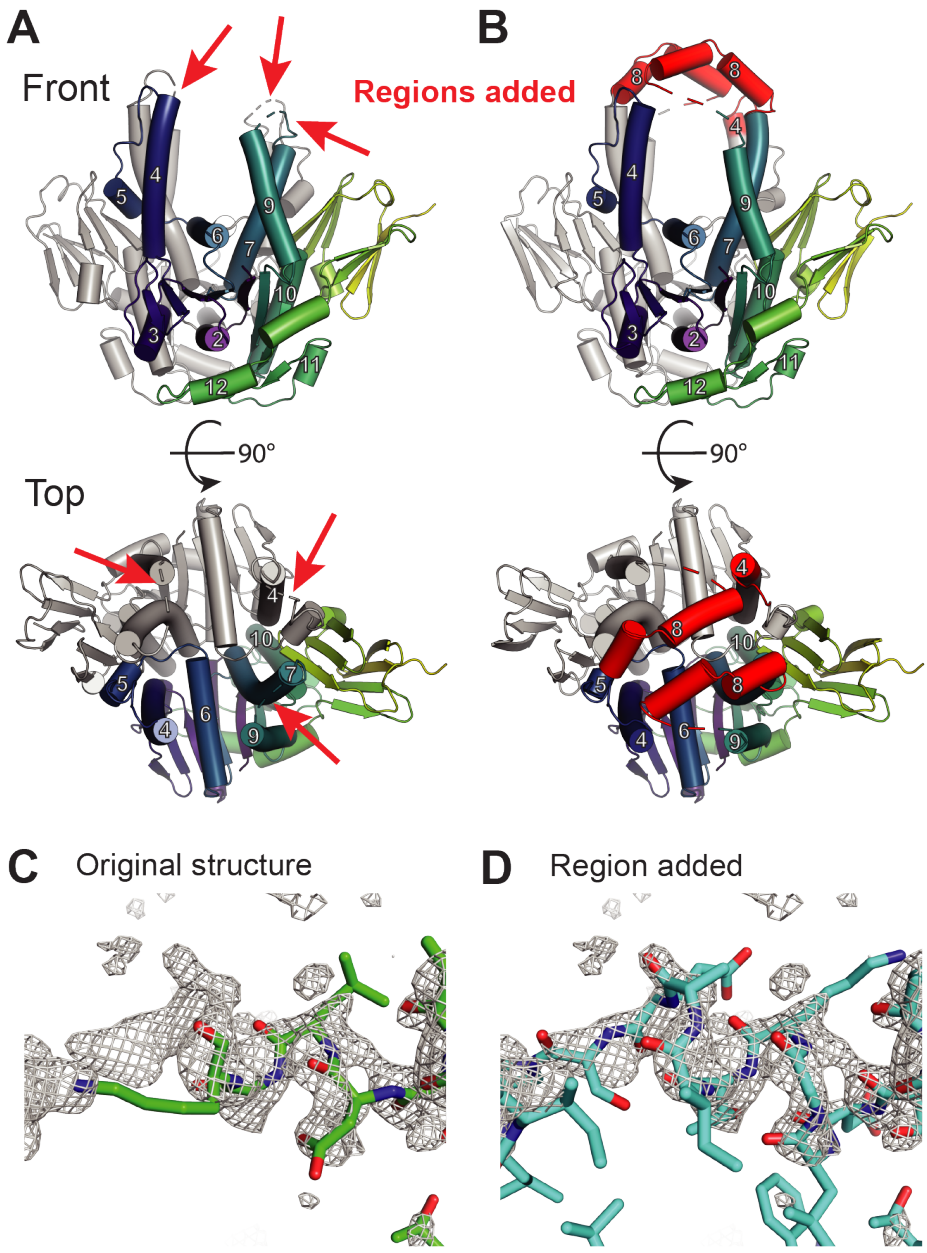


Figure S2.7: New refinement of *Nos*Get3d. (A) Original deposited structure of *Nos*Get3d (PDB ID: 3IGF) and (B) structure after refinement here. The *front* and *top* views are shown. Highlighted *red* regions were added in this refinement. Color and TMD numbering as in Fig. 2.3. (C) $2F_O$-F_C electron density map (*light grey mesh*) and model of the original deposited structure of *Nos*Get3d (*green sticks*) and (D) the structure after refinement (*blue sticks*), showing the improved fit and residues added. $2F_O$-F_C electron density map contoured at 1σ.

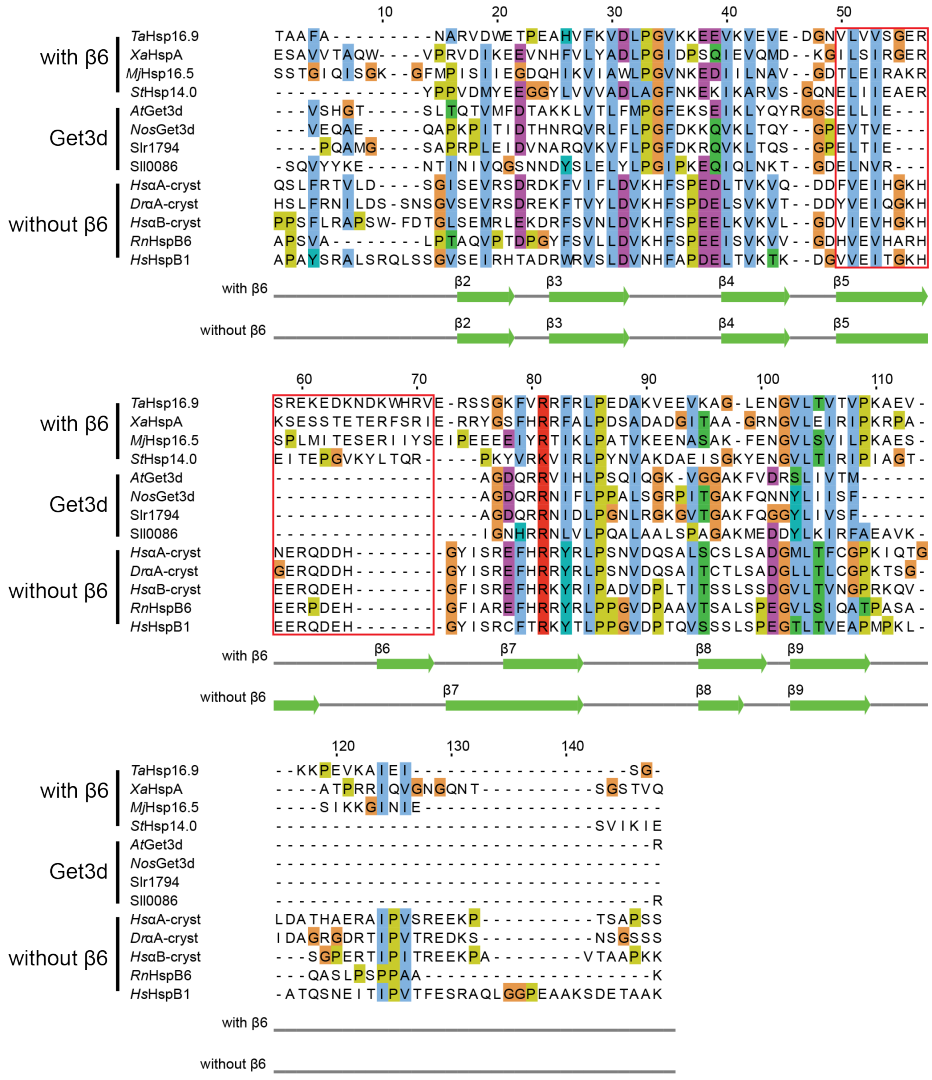


Figure S2.8: Sequence alignment of selected α-crystallin domains (αCDs). Predicted secondary structure is shown *below* the sequences for αCD with and without the loop containing β6. Sequences are classified as either αCD with or without the loop containing β6 or αCD of Get3d. Discussed regions are boxed in *red*. Species key: *Ta*, *Triticum aestivum*; *Xa*, *Xanthomonas axonopodis*; *St*, *Sulfurisphaera tokodaii*; *Dr*, *Danio rerio*; *Rn*, *Rattus norvegicus*. Coloring and other species as in Fig. 2.1B.

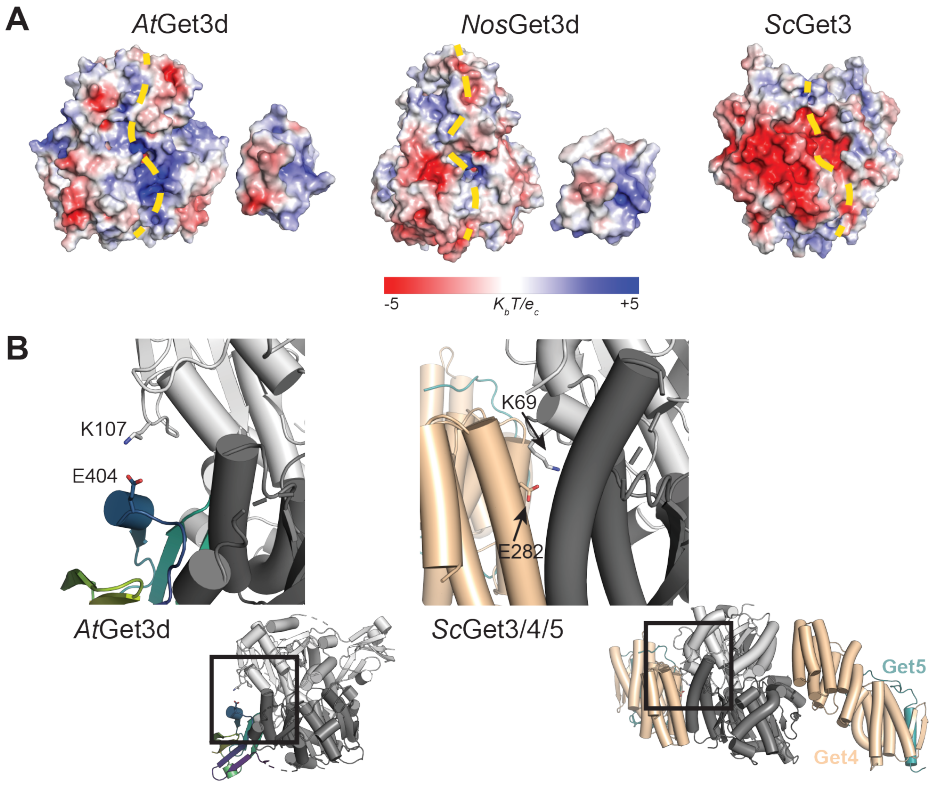


Figure S2.9: Interactions at the α-crystallin domain (αCD) interface. (A) Electrostatic potential showing the interactions between the αCD of *At*Get3d (*left*), *Nos*Get3d (*middle*), and yeast Get3 (*Sc*Get3) with Get4/5 (*right*). Orientation of the Get3 domain (*left*) as in Fig. 2.4D with αCD reflected (*right*) to show the surface that interacts with the Get3 domain. The electrostatic surface was calculated using PDB2PQR and Adaptive Poisson-Boltzmann Solver (APBS) in Pymol, scale shown. The *dotted yellow line* traces the interface between the two monomers of the Get3 domain. (B) Structure of *At*Get3d (*left*) showing a salt bridge between the Get3 domain and the αCD of the opposite monomer. Colored as in Fig. 2.4A. Structure of yeast Get3 (*right*, *dark and light grey*) in complex with Get4/5 (*wheat* and *teal*, respectively) (*Sc*Get3/4/5) showing a salt bridge formed between Get3 and Get4. Discussed residues are shown as *sticks*. For each, the region highlighted is shown *below* in the full structure.

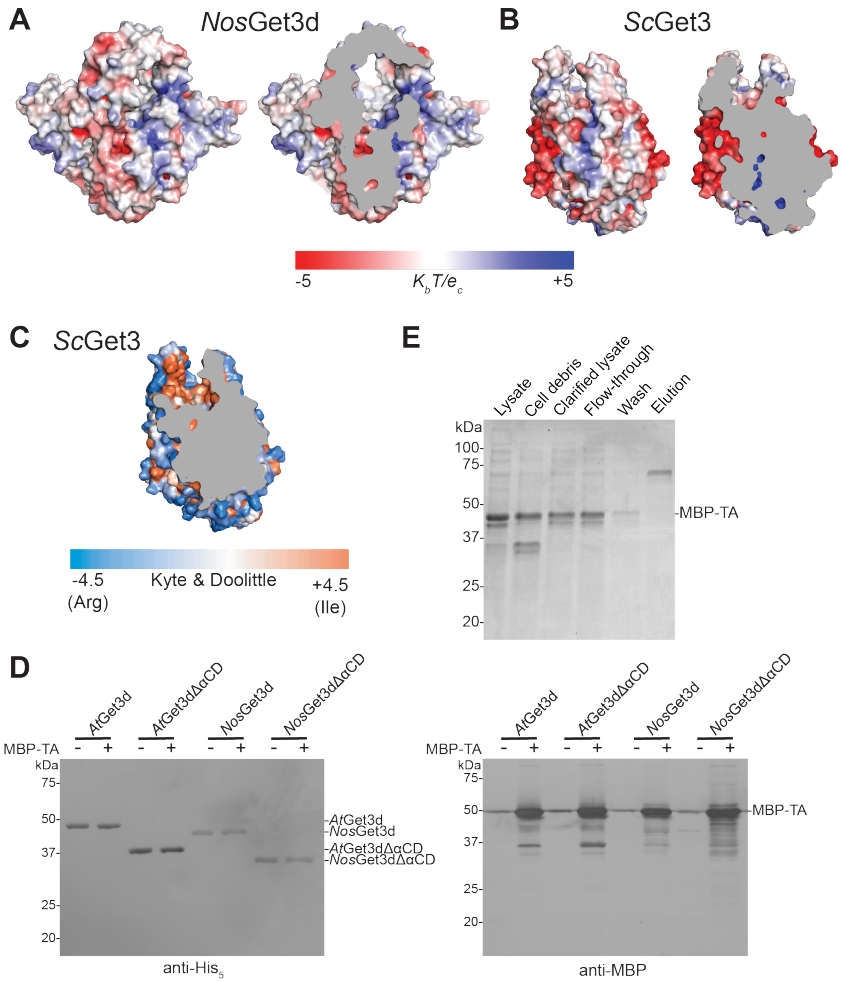


Figure S2.10: Additional features of the hydrophobic chamber of Get3d. Full (*left*) and slice view (*right*) surface electrostatic potential of (A) *Nos*Get3d and (B) the closed conformation of yeast Get3 (*Sc*Get3, PDB ID: 4XTR) as in Fig. 2.6A. (C) Slice view of the surface hydrophobicity of the closed conformation of yeast Get3 (*Sc*Get3, PDB ID: 4XTR) as in Fig. 2.6B. (D) Anti-His_5 (*left*) and anti-MBP (*right*) western blots of samples from Fig. 2.6D. Get3d is identified using an anti-His_5 antibody, and TA protein is identified using an anti-MBP antibody. Molecular weight marker (kDa) shown. (E) Coomassie stained SDS-PAGE of MBP-TA without Get3d treated as in Fig. 2.6D. Molecular weight marker (kDa) shown.

Table S2.2: Constructs and primers used in this study.

Construct	**Primer**	**Notes**
pET22b-*At*Get3d Δ1–57-His_6	F: GCTTCATATGACCAAATTCGTCACCTTTCTCGG R: CGGAAGCTTCCGCATTGTGACGATGAGAC	NdeI/XhoI
pUBQ10-gGet3d-EGFP	F: TTATAATGCCAACTTTGTACAAAAAAGCA-GAAAAAAATGGTGTCTTTGGTCAATTC R: AGATCCAGCAGATCCCCGCATTGTGACGATGAG F: AAAAAAATGGTGTCTTTGGTCAATTCTTCT R: CCGCATTGTGACGATGAGACT	*A. thaliana* genomic DNA amplification
pUBQ10-gGet3dΔ1–34-EGFP	F: TTATAATGCCAACTTTGTACAAAAAAGCA-GAAAAAAATGGTGGCAGCTTATG R: AGATCCAGCAGATCCCCGCATTGTGACGATGAG F: AAAAAAATGGTGGCAGCTTATGTGGCGGCTAC R: CCGCATTGTGACGATGAGACT	*A. thaliana* genomic DNA amplification
pUBQ10-gGet3d-EGFP & Δ1–34-EGFP	F: CAGCTTTCTTGTACAAAGTTG R: CTGCTTTTTTGTACAAAGTTG F: ATCGTCACAATGCGGGGATCTGCTGGATCTGCTGCTG-GATCTGGAGAATTTATGGTGAGCAAGGGCGAG R: TTATAATGCCAACTTTGTACAAGAAAGCTGTTACTTG-TACAGCTCGTCC F: AAAAAATGGTGGCAGCTTATGTGGCGGCTAC R: CATAAGCTGCCACCATTTTTTTCTGCTTTTTTGTA-CAAAGTTG	pENTR and GFP amplification; sequencing

Continued on next page

Table S2.2: *(Continued)*

Construct	**Primer**	**Notes**
pET22b-His_6-3C-*At*Get3dΔ1–57	Vector F: GTCACAATGCGGTAGTCTCGAGCACCACCAC Vector R: GACGAATTTGGTACTTCCGCTGCCTGGTC Insert F: ACCAAATTCGTCACCTTTCTCG Insert R: CCGCATTGTGACGATGAGAC	Gibson cloning
pET33b-His_6-TEV-*At*Get3dΔ1–57	F: CAGAGCGTCGACACCAAATTCGTCAC R: TCCCATATGCTACCGCATTGTGACG	Amplification for Gibson cloning
pET33b-His_6-TEV-*At*Get3dΔ1–57,377–485	F: CAGAGCGTCGACACCAAATTCGTCAC R: TTCCCATATGCTAAGTTTCAGAGAGAAGTTC	Amplification for Gibson cloning
pET33b-His_6-TEV-*Nos*Get3d	F: CAGAGCGTCGACGCCCTGATATTGAC R: TCCCATATGCTACTCGAGGAAAGAAATGATC	*Nos*Get3d amplification for Gibson cloning
pET33b-His_6-TEV-*Nos*Get3dΔ292–366	F: CAGAGCGTCGACGCCCTGATATTGAC R: TCCCATATGCTACGCTTGTTCGGCTTG	Amplification for Gibson cloning
pET33b-His_6-TEV vector	F: TAGCATATGGGAATTCGAAGCTTGCGG R: GTCGACGCTCTGGAAGTACAGGTTTTC	Vector backbone amplification for Gibson cloning
pET33b-His_6-TEV-*Sc*Get3	—	See Suloway 2009
pACYCDuet-MBP-*Sc*Sbh1(52–82)	—	See Lin 2021

Note: "F" = forward primer; "R" = reverse primer. All primer sequences are written 5' to 3'.

Chapter 3

CHARACTERIZATION OF THE PLANT-LIKE GET3D HOMOLOG IN CYANOBACTERIA

3.1 Abstract

The ATPase Get3 is the central targeting factor in the guided entry of tail-anchored membrane proteins pathway, responsible for delivering tail-anchored membrane proteins to the endoplasmic reticulum membrane. A distinct homolog of Get3, termed Get3d, was recently identified in plants and photosynthetic bacteria. Its conservation among photosynthetic organisms suggests a specialized role in photosynthesis. This work investigates the function of the plant-like Get3d paralog (*slr1794* gene, Slr1794 protein) in the model cyanobacterium *Synechocystis* sp. PCC 6803. We explore the sequence features and predicted structure of the Slr1794, demonstrating its similarity to other Get3d homlogs. Growth assays reveal that *slr1794* is essential for cellular growth, while its ATPase activity is dispensable. Pigment analyses reveal that *slr1794* is required for proper chlorophyll *a* and carotenoid biosynthesis. Whole-cell proteomic analysis further supports a role in photosynthesis by demonstrating the requirement of Slr1794's ATPase activity to maintain normal expression of photosynthesis-related proteins, including components of the photosystems and light-harvesting complexes. Putative interaction partners are identified and shown to be enriched in membrane and photosynthesis-related proteins. Together, these findings indicate that the plant-like Get3d paralog in *Synechocystis* sp. PCC 6803 is a critical factor for maintaining cellular health and may function in the biogenesis or maintenance of membrane proteins and photosynthetic machinery.

3.2 Introduction

Tail-anchored (TA) proteins are a class of membrane proteins with a single transmembrane domain (TMD) within ~30 amino acids of their C-terminus (*1*). TA proteins play many important roles in the cell such as in vesicle trafficking (reviewed in (*2*)), protein translocation across membranes (reviewed in (*3*)), and regulation of apoptosis (*4*). TA proteins, which make up ~1–2% of the total proteome, are present in all cellular membranes, including the endoplasmic reticulum (ER), mitochondria, and chloroplast (reviewed in (*5*)).

A challenge arises when targeting TA proteins to their destination membranes. For most membrane proteins, the first TMD is recognized by targeting factors during translation by the ribosome, allowing for co-translational targeting (reviewed in (*6*)). For TA proteins, the TMD is not accessible to targeting factors until they have been completely synthesized and released from the ribosome. This presents a problem for the cell, as the hydrophobic TMD must be shielded from the hydrophilic environment of the cytosol to maintain protein stability. To overcome this, multiple membrane protein targeting pathways have evolved, including the guided entry of tail-anchored proteins (GET) pathway, which delivers TA proteins to the ER membrane (reviewed in (*5*, *7*)).

In the canonical yeast system, the GET pathway is composed of six proteins: the chaperone Sgt2 and Get1 through Get5. Upon the release of a TA protein from the ribosome, Sgt2 binds the TA protein and delivers it to the preassembled Get3/Get4/Get5 pre-targeting complex. The TA protein is then transferred from Sgt2 to the homodimeric ATPase Get3, the central targeting factor, which delivers the TA protein to the Get1/Get2 insertase at the ER membrane (*8*). The GET pathway has primarily been characterized in opisthokonts (including animals and fungi), which account for only a small fraction of the diversity of life.

The GET pathway in plants is of particular interest, as a deeper understanding of the fundamental biology of plants will be required to continue to provide the planet with food, oxygen, and energy (reviewed in (*9*)). Plants are unique in that they have multiple paralogs of Get3, termed Get3a–d (*8*, *10*). Get3a serves as the cytosolic TA protein targeting factor (*11*, *12*). Get3b may be involved in TA protein targeting in the chloroplast stroma, but more evidence is needed to support this claim (*13*, *14*). Get3c is known to localize to the mitochondrial matrix, although there is no evidence suggesting its functional role (*10*, *11*, *15*). Our group recently identified the fourth paralog of Get3, which we termed Get3d, that is conserved in plants and

photosynthetic bacteria (*8*).

We have previously shown that Get3d from the model plant *Arabidopsis thaliana* localizes to the chloroplast and that Get3d is an active ATPase that can bind TA protein *in vitro* (*16*). Additionally, we determined the structure of Get3d using X-ray crystallography and found a unique structural fold with an α-crystallin domain (αCD) and a closed client binding chamber (PDBID: 8ELF, 8EGK). These studies suggest that Get3d may play a role similar to TA protein targeting in the plant chloroplast and photosynthetic bacteria.

This work seeks to characterize a subgroup of Get3d homologs using the model cyanobacterium *Synechocystis* sp. PCC 6803. We identify two subgroups of Get3d proteins: those that are similar to the *Arabidopsis thaliana* Get3d, termed "plant-like" Get3d, and others that retain more features of the cytosolic Get3, termed "canonical" Get3d. We prepare a *Synechocystis* sp. PCC 6803 strain with the plant-like Get3d gene *slr1794* deleted and various complementation strains to probe the characteristic features of Get3d. The *slr1794* knockout has a significant growth defect that is dependent on the αCD but not on ATPase activity. Chlorophyll *a* and carotenoid content are determined and shown to be dependent on *slr1794*, with significantly reduced pigment levels in the knockout strain. Whole-cell proteomic analysis suggests that Slr1794 plays a role in redox homeostasis and photosynthesis. Finally, putative interaction partners are identified by co-immunoprecipitation coupled to mass spectrometry and are enriched in both photosynthesis-related and membrane proteins. Overall, this work reveals that the plant-like Get3d paralog of the model cyanobacterium *Synechocystis* sp. PCC 6803 plays a key role in the cell and suggests that it may function in membrane protein targeting and/or photosynthesis.

3.3 Results

Sequence Features of Get3d Homologs

To investigate the conservation of Get3d sequence features, the sequences of Get3/ArsA homologs from selected species were aligned using multiple alignment using fast Fourier transform (MAFFT) (*17*), visualized in Jalview (*18*), and colored per the ClustalX color scheme (*19*). The sequence alignments of notable motifs are shown in Fig. 3.1A. Get3d paralogs contain the same motifs as other Get3/ArsA homologs, which includes the P-loop, Switch I, Switch II, Get3/TRC40 insert, and A-loop (*16*). All Get3d homologs are missing the CXXC motif.

Some Get3d homologs, including Get3d from *A. thaliana* (*At*Get3d), Get3d from *Nostoc* sp. PCC 7120 (*Nos*Get3d), and Slr1794 from *Synechocystis* sp. PCC 6803 (PCC 6803), have low conservation in the P-loop, Switch I, and Switch II motifs and are completely missing the A-loop (Fig. 3.1A). They are also missing the pivot helix, which serves as a rotation point for structural rearrangements that occur throughout the ATPase cycle of Get3 (*8*). These Get3d paralogs are termed "plant-like" since they are similar to the Get3d homolog from the plant *A. thaliana* (*At*Get3d). Other Get3d homologs, including Sll0086 from PCC 6803, Get3d from the green sulfur bacteria (GSB) *Chlorobium* sp., and Get3d from the green non-sulfur bacteria (GNSB) *Chloroflexii* sp., have high conservation in the P-loop, Switch I, and Switch II motifs (Fig. 3.1A) and retain the A-loop and the pivot helix. These Get3d paralogs are termed "canonical," as they retain more of the canonical Get3 motifs.

To further probe the conservation of sequence features in Get3d paralogs, representative species from each order of cyanobacteria (Chroococcales, Gloeobacterales, Nostocales, Oscillatoriales, Pleurocapsales, Spirulinales, and Synechococcales (*20*)) were selected, and Get3d sequences were identified by Protein BLAST using Slr1794 (the plant-like Get3d protein encoded by the *slr1794* gene) and Sll0086 (the canonical Get3d protein encoded by the *sll0086* gene) as the query sequences (*21*). The sequences were aligned, visualized, and colored as above (Fig. S3.3). The sequence motifs are highly conserved within the plant-like Get3d proteins and the canonical Get3d proteins.

To quantify the conservation of sequence features in select Get3d homologs, sequence identity and similarity were calculated using the Sequence Manipulation Suite (*22*). Slr1794 has the highest sequence identity and similarity to *Nos*Get3d (52% identity and 73% similarity), which is expected as they are both plant-like Get3d proteins from cyanobacteria. Slr1794 has lower sequence identity and similarity to *At*Get3d (27% identity and 42% similarity), due to their more distant evolutionary relationship, and Sll0086 (26% identity and 44% similarity) because it is a canonical Get3d protein and not a plant-like Get3d protein. Finally, yeast Get3 (*Sc*Get3) has the lowest sequence identity and similarity compared to Slr1794 (18% identity and 34% similarity) since it is a canonical Get3 protein and not a Get3d protein.

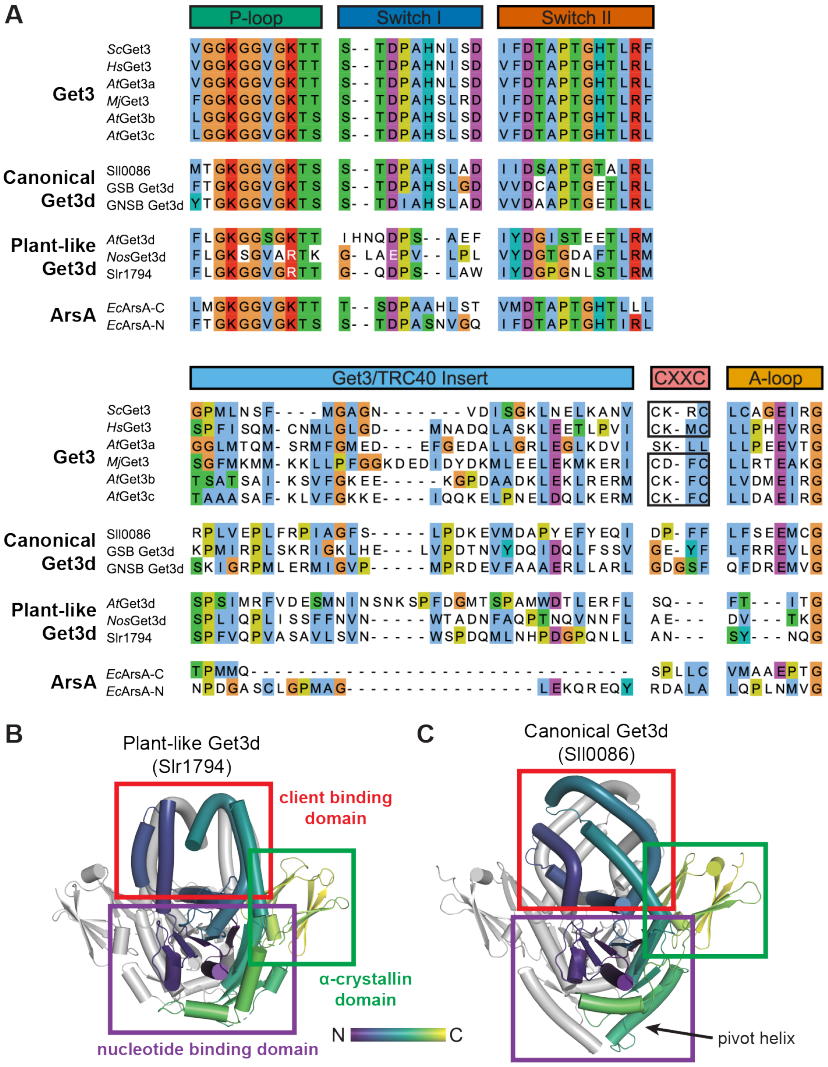


Figure 3.1: Features of Get3d homologs. (A) Sequence alignment of selected Get3/Get3d homologs was prepared using multiple alignment using fast Fourier transform (MAFFT) (*17*, *23*). Important sequence regions were visualized in Jalview (*18*) and colored per the ClustalX color scheme (*19*). Features as discussed in the text are labeled *above* the respective sequences. Species key: *Sc*, *Saccharomyces cerevisiae*; *Hs*, *Homo sapiens*; *At*, *Arabidopsis thaliana*; *Mj*, *Methanocaldococcus jannaschii*; GSB, *Chlorobium* sp. (green sulfur bacteria); GNSB, *Chloroflexii* sp. (green non-sulfur bacteria); *Nos*, *Nostoc* sp. PCC 7120; *Ec*, *Escherichia coli*. Slr1794 and Sll0086 are Get3d homologs from *Synechocystis* sp. PCC 6803. (B) Predicted structure of Slr1794 and (C) predicted structure of Sll0086. Structures were predicted using AlphaFold3 with coloring per modified predicted local distance difference test (pLDDT) coloring scheme shown in Fig. S3.2 (*24*). For each, monomer A is shown in *viridis* (scale shown) and monomer B is shown in *grey*. The following domains are boxed and labeled: client binding domain, nucleotide binding domain, α-crystallin domain. The pivot helix of Sll0086 is labeled.

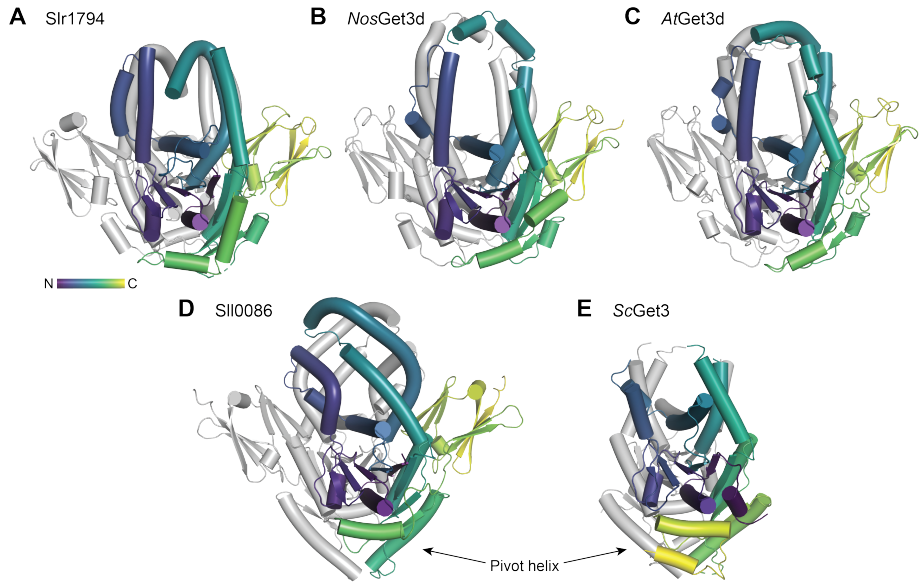


Figure 3.2: Structural comparison of Get3d homologs. Predicted structure of (A) Slr1794, (B) X-ray crystal structure of *Nos*Get3d (PDB ID: 8EGK) (*16*), (C) X-ray crystal structure of *At*Get3d (PDB ID: 8ELF) (*16*), (D) predicted structure of Sll0086, and (E) X-ray crystal structure of closed yeast Get3 (*Sc*Get3, PDB ID: 2WOJ) (*26*). For each, monomer A is shown in *viridis* (scale shown) and monomer B is shown in *grey*. The pivot helix of Sll0086 and *Sc*Get3 is labeled. Structures of Slr1794 and Sll0086 dimers were predicted using AlphaFold3 (*24*) with coloring per modified predicted local distance difference test (pLDDT) coloring scheme shown in Fig. S3.2.

Structural Predictions of Get3d Homologs

To investigate Get3d in more detail, AlphaFold3 was utilized to predict structures of the Slr1794 dimer (Figs. 3.1B and S3.2A) and the Sll0086 dimer (Figs. 3.1C and S3.2B) (*24*). The characteristic nucleotide binding domain (NBD) and client binding domain (CBD) of all Get3 proteins and the α-crystallin domain (αCD) characteristic of Get3d proteins are noted (Figs. 3.1B, C).

Consistent with the sequence identity and similarity, the predicted structure of Slr1794 (Figs. 3.2A and S3.2A) (*24*) is most similar to the published X-ray crystal structures of the plant-like Get3d homologs *Nos*Get3d (Fig. 3.2B, PDB ID: 8EGK) and *At*Get3d (Fig. 3.2C, PDB ID: 8ELF) (*16*) with root mean square deviation (RMSD) values of 1.38 Å and 2.57 Å, respectively (*25*). The predicted structure of Slr1794 (Figs. 3.2A and S3.2A) is less similar to the predicted structure of Sll0086 (Figs. 3.2D and S3.2B) (*24*) and the published X-ray crystal structure of the closed conformation of *Sc*Get3 (Fig. 3.2E, PDBID: 2WOJ) (*26*) with RMSD values of 3.05 Å and 5.86 Å, respectively (*25*).

Plasmid Design and *Synechocystis* sp. PCC 6803 Transformation

To probe the functional role of Slr1794, the plant-like Get3d protein from PCC 6803, various PCC 6803 strains were prepared to knockout and rescue the *slr1794* gene and to append an affinity-tag for co-immunoprecipitation coupled to mass spectrometry (co-IP/MS^2). The strains prepared in this study are reported in Table 3.5. All plasmids were designed using the backbone from pAM1579 for propagation in *Escherichia coli* (*27*). PCC 6803 was transformed per standard protocols (*28*) and serially streaked with appropriate antibiotic(s) until complete segregation was reached (i.e., each copy of the genome contained the construct of interest).

The knockout (KO) strain was prepared by replacing the native *slr1794* gene with a kanamycin resistance cassette. The rescue strain was prepared by complementing the KO strain with the wild-type *slr1794* gene at the *psbA2* locus (WT-rescue). Two variant rescue strains were prepared by complementing with an ATPase-deficient mutant (D39N-rescue) and a C-terminal αCD truncation ($\Delta\alpha$CD-rescue). Finally, a strain with a C-terminal 3xFLAG tag appended to the native *slr1794* gene (Slr1794-3xFLAG) was prepared for co-IP/MS^2.

In the rescue strains, the *psbA2* gene was replaced with the *slr1794* gene, resulting in *slr1794* gene expression under the control of the *psbA2* promoter. It has been shown that the deletion of *psbA2* does not impair photosynthesis, as *psbA3* provides compensatory D1 expression (*29*). The *psbA2* locus and its promoter region have been extensively used to constitutively express PCC 6803 genes of interest (*30*, *31*).

Confirmation of Transformation and Segregation

To confirm the transformation and segregation of each strain, the loci of interest were amplified with appropriate primers, DNA gel electrophoresis was performed, and bands were compared to expected fragment sizes for both the native and manipulated loci (Fig. S3.1). The PCR products for the native *slr1794* and *psbA2* loci are expected to be 1.4 kb and 1.3 kb, respectively. The *slr1794* locus in the *slr1794* knockout strain is expected to be 1.0 kb, and *psbA2* locus in the rescue strains is expected to be ~2.8-3.0 kb. For each strain, bands corresponding to the expected native and manipulated loci were present. In the *slr1794* knockout and rescue strains, no native *slr1794* was present at the *slr1794* locus, and in the rescue strains, no native *psbA2* was present at the *psbA2* locus, confirming full transformation and segregation at these sites.

Whole-Cell Proteomic Analysis

Whole-cell proteomic analysis was performed to explore the global proteomic differences upon *slr1794* gene deletion. For each of the five strains (WT, KO, WT-rescue, D39N-rescue, and $\Delta\alpha$CD-rescue), 2 μg peptides from whole-cell lysate were analyzed by mass spectrometry. Raw abundance values and metadata are reported in Table S3.6. Differential expression analysis was performed (Table S3.7). To validate Slr1794 expression in the five strains, the levels of expression of Slr1794 were analyzed (Table 3.1). As expected, Slr1794 was nearly undetectable in the KO (–5.53 $\log_2$FC, $p = 0.0036$), while the WT-rescue strain showed significant overexpression relative to WT (+1.30 $\log_2$FC, $p = 0.0285$), likely due to the strong *psbA2* promoter. The D39N-rescue strain displayed moderately reduced expression (–0.60 $\log_2$FC), although the reduction was not statistically significant ($p = 0.178$). In contrast, the $\Delta\alpha$CD-rescue strain exhibited significantly reduced expression (–4.73 $\log_2$FC, $p = 0.0056$), suggesting instability or impaired expression of the truncated protein.

Table 3.1: Expression of Slr1794 across *Synechocystis* sp. PCC 6803 strains.

Comparison	**Mean (Group)**	**$\log_2$FC vs. WT**	**p-value**
WT (baseline)	2.75×10^8	–	–
KO vs. WT	5.96×10^6	–5.53	0.0036
WT-resc vs. WT	6.80×10^8	+1.30	0.0285
D39N-resc vs. WT	1.81×10^8	–0.60	0.1780
$\Delta\alpha$CD-resc vs. WT	1.03×10^7	–4.73	0.0056

Note: Quantitative analysis of Slr1794 protein abundance in wild-type (WT) *Synechocystis* sp. PCC 6803, *slr1794* knockout (KO), and the three complementation strains (WT-resc, D39N-resc, and $\Delta\alpha$CD-resc). Values represent mean precursor intensity, $\log_2$(fold-change) ($\log_2$(FC)) relative to WT, and p-value (unpaired t-test).

Principal component analysis (PCA) was performed to assess global proteomic variation across the strains. Log_2 protein abundance values were calculated and were median normalized to reduce technical variability, and PCA was applied to the resulting matrix. The first two principal components (PC1 and PC2) captured the dominant sources of variance in the dataset and were used to visualize sample clustering by strain (Fig. S3.6). PC1 and PC2 accounted for 18.2% and 14.5% of the total variance, respectively. The PCA shows clear separation between the WT and KO clusters. This suggests that the deletion of the *slr1794* gene results in widespread changes in the proteome. The WT-rescue strain clusters between WT and KO, suggesting partial restoration of the WT proteome. The D39N-rescue strain

clusters closer to KO, which suggests that ATPase activity is required for proper function. As expected, the αCD-rescue strain also clusters closer to KO, likely due to the significantly reduced levels of Slr1794 in the $\Delta\alpha$CD-rescue strain.

Volcano plots were prepared to visualize up- and down-regulation of proteins in the various strains (Fig. 3.3 and Table S3.7). First, the up- and down-regulation of proteins in KO versus WT was probed to determine the global proteomic effects of the deletion of the *slr1794* gene. In the KO versus WT analysis, 104 proteins were significantly up-regulated and 61 proteins were significantly down-regulated ($\log_2$(fold-change) > 0.58 and $p < 0.05$) for a total of 165 differentially expressed proteins. This is indicative of strong cellular disruptions with loss of Slr1794. In the WT-rescue versus KO analysis, there are only 53 differentially expressed proteins, a marked reduction compared to KO versus WT. This decreased number of differentially expressed proteins is indicative that the reintroduction of the Slr1794 complements the loss, restoring at least some of the changes that result from the loss of the *slr1794* gene. The D39N-rescue versus KO analysis reveals 95 total differentially expressed proteins, suggesting that the ATPase-deficient mutant is unable to fully rescue the KO phenotype. This is indicative of the ATPase activity being functionally important for Slr1794. The $\Delta\alpha$CD-rescue versus KO has 136 differentially expressed proteins, the highest number among the rescue strains and approaching KO levels. Due to the significantly decreased levels of Slr1794 in the $\Delta\alpha$CD-rescue strain, the large number of differentially expressed proteins is expected, as the levels of expression of Slr1794 are actually approaching that of the KO.

To further characterize the differential expression in the various strains, the Database for Annotation, Visualization and Integrated Discovery (DAVID) was used to analyze enrichment clusters to identify the functional clusters up- and down-regulated in the various strains (Table S3.8). In the KO versus WT, DAVID enrichment analysis of up-regulated proteins revealed a strong functional cluster (enrichment score 2.82) enriched for DNA-binding domains, including the Cro/C1-type helix-turn-helix (HTH) motif (IPR001387), lambda-like DNA-binding domains (IPR010982), and GO:0003677 DNA binding. These findings suggest transcriptional regulators are up-regulated in the absence of Slr1794, indicating a broad compensatory response to proteomic imbalance. Among proteins down-regulated in the KO strain, DAVID enrichment analysis revealed a cluster (enrichment score 1.44) enriched for thioredoxin-related domains, including IPR013766 (Thioredoxin domain) and

IPR036249 (Thioredoxin-like superfamily). This suggests a loss of redox buffer capabilities in the KO strain, suggesting a role of Slr1794 in redox stress response, either directly or indirectly.

In the enrichment analysis of the WT-rescue strain versus KO, there are no significantly enriched clusters identified (enrichment score > 1.3). This, along with the relatively lower number of differentially expressed proteins, indicates that the reintroduction of Slr1794 restores proteomic homeostasis overall. In the D39N-rescue strain, functional annotation clustering of proteins down-regulated relative to KO identified a significant enrichment of photosynthesis-related terms (enrichment score = 1.78). This cluster included GO:0015979 (photosynthesis), GO:0030096 (photosystem II), and UniProt keywords such as KW-0602 (photosynthesis), KW-0604 (photosystem II), and KW-0605 (phycobilisome) (Table 3.2). These results indicate that ATPase-deficient Slr1794 is unable to restore the expression of key components of the thylakoid membrane and photosynthetic machinery, highlighting the requirement of ATPase activity in Slr1794's function. Finally, in the $\Delta\alpha$CD-rescue strain, DAVID analysis revealed a strongly enriched functional cluster (enrichment score = 2.37) down-regulated compared to KO. The enriched terms in this cluster are involved in CRISPR-related and antiviral defense pathways, including KW-0051 (antiviral defense), IPR005537 (RAMP3 family), and GO:0051607 (defense response to virus). Since levels of Slr1794 are unexpectedly low in the $\Delta\alpha$CD-rescue strain, approaching KO levels, this enrichment does not likely represent a true functional distinction between these strains, but instead represents a modest suppression of stress-response pathways in the $\Delta\alpha$CD-rescue strain that is strongly activated in the complete KO.

Spot Growth Experiment

To investigate the impact of the loss of the *slr1794* gene, spot growth assays were conducted with each *Synechocystis* sp. PCC 6803 strain (WT, KO, WT-rescue, D39N-rescue, and $\Delta\alpha$CD-rescue) (Fig. 3.4). 3 μL spots of a 1:5 dilution series (starting OD_{730} = 0.5) were plated on a BG-11 agar plate without antibiotics and grown under standard conditions.

The KO resulted in a significant growth defect, suggesting that the *slr1794* gene is vital for cellular fitness. The growth defect was rescued by the introduction of the WT gene at a neutral site (WT-rescue), confirming that this growth defect is in fact due to the loss of the *slr1794* gene and is not an artifact of genetic manipulations.

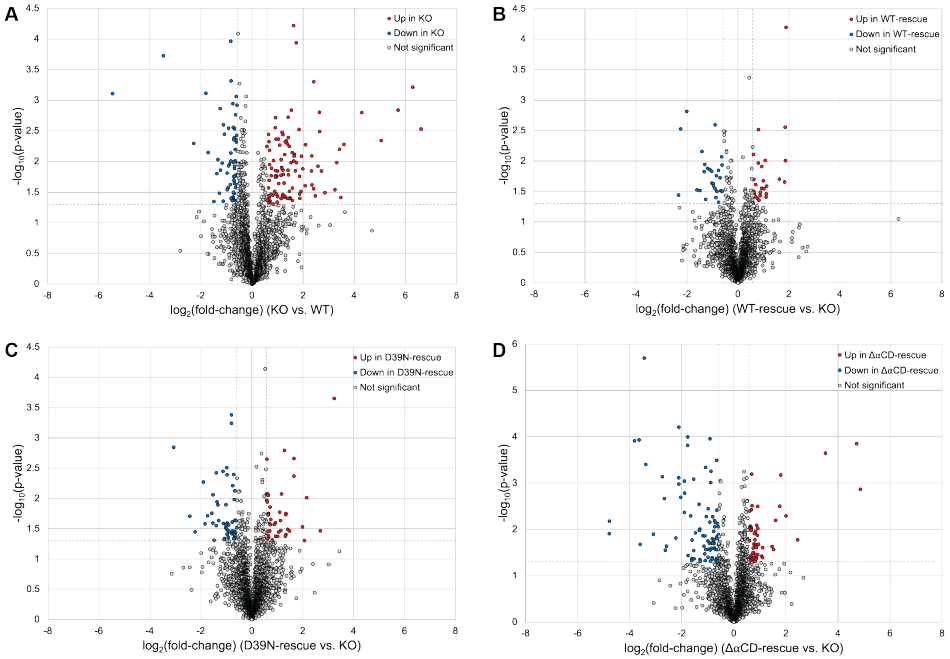


Figure 3.3: Volcano plots of differential protein abundance in whole-cell mass proteomes of *Synechocystis* sp. PCC 6803. Comparisons shown: (A) *slr1794* knockout (KO) versus wild-type (WT) (104 up, 61 down), (B) WT-rescue versus KO (26 up, 27 down), (C) D39N-rescue versus KO (44 up, 51 down), and (D) ΔαCD-rescue versus KO (52 up, 84 down). Log_2(fold-change) versus -$\log_{10}$(adjusted [adj.] p-value) is plotted. *Red* and *blue points* denote up- and down-regulation, respectively ($\log_2$(fold-change) > 0.58 and $p < 0.05$). *Grey points* denote no significance. Vertical and horizontal dashed lines indicate the fold-change and significance cutoffs. Abundance and metadata are reported in Table S3.6, and differential protein abundance comparisons are reported in Table S3.7.

Table 3.2: Photosynthesis-related proteins significantly down-regulated in the ATPase-deficient complementation strain compared to the *slr1794* knockout strain.

UniProt ID	**Protein Name**	**Gene**
P05429	Photosystem II CP47 reaction center protein	*psbB*
Q01950	Phycobilisome 7.8 kDa linker polypeptide, allophycocyanin-associated, core	*apcC*
Q01951	Allophycocyanin alpha chain	*apcA*
P74551	Allophycocyanin subunit beta-18	*apcF*
P27181	ATP synthase subunit b	*atpF*
P73676	Photosystem II reaction center protein Y	*psbY*
P09190	Cytochrome b559 subunit alpha	*psbE*
P27589	Cytochrome b6-f complex subunit 4	*petD*

Note: Proteins were identified from Cluster 1 of the Database for Annotation, Visualization and Integrated Discovery (DAVID) functional annotation enrichment (Enrichment Score = 1.78) and include components of photosystems I and II, phycobilisomes, and electron transport complexes (Table S3.8). The Uniprot ID, protein name, and gene name (if annotated) are reported.

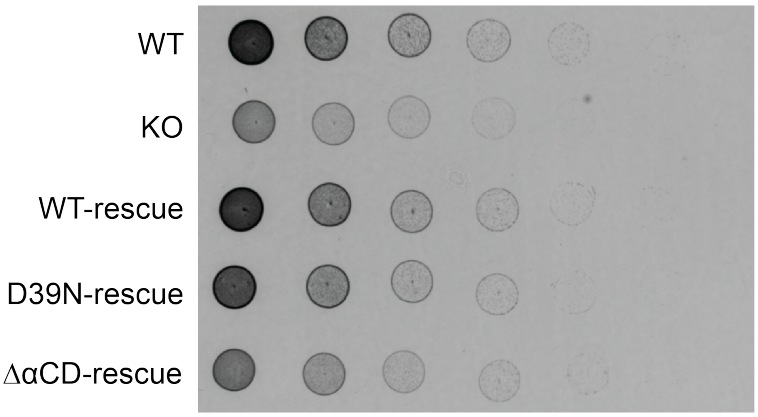


Figure 3.4: Spot growth assay of *slr1794* knockout and complementation strains. 1 into 5 dilution series of equal OD_{730} *Synechocystis* sp. PCC 6803 strains prepared and plated on BG-11 agar. Strains per the following: wild-type (WT), *slr1794* knockout (KO), and complementation strains (WT-rescue, D39N-rescue, and $\Delta\alpha$CD-rescue).

The reintroduction of an ATPase-deficient mutant at a neutral site (D39N-rescue) also fully restores this growth defect. This suggests that the ATPase activity of Slr1794 is *not* essential for its vital function. This is unexpected, as functioning ATPase activity required for other Get3 homologs to fulfill their cellular functions (*26*, *32*).

The reintroduction of *slr1794* with a C-terminal αCD deletion ($\Delta\alpha$CD-rescue) does not fully rescue the growth defect. This suggests that the αCD of Slr1794 is essential. However, whole-cell proteomic analysis revealed that the Slr1794 $\Delta\alpha$CD mutant ($\Delta\alpha$CD-rescue) is present at much lower levels than both the WT (WT-rescue) and ATPase-deficient mutant (D39N-rescue) (Tables 3.1 and S3.6). Therefore, the lack of rescue of the growth defect may simply be due to the lower levels of the *slr1794* gene product and not due to the loss of the αCD in particular.

This difference in protein levels is unexpected, as each rescue is integrated at the same *psbA2* locus, a neutral site with a high-expressing promoter, and should therefore be expressed at similar levels (*30*, *31*). The reduced protein levels suggest that either the corresponding $\Delta\alpha$CD mRNA or Slr1794 $\Delta\alpha$CD protein is less stable than both the full-length and ATPase-deficient mutants, suggesting that this domain is necessary for stable expression of the *slr1794* gene product.

Pigment Analysis

To further characterize the loss of the *slr1794* gene, pigments were extracted per standard methanol extraction protocols and quantified (*33*). For each strain (WT, KO,

WT-rescue, D39N-rescue, and ΔαCD-rescue), three independent cultures cultures were grown, and pigments were extracted from equal OD_{730} cells. The absorbance at 470 nm, 665 nm, and 720 nm were measured (raw data in Table S3.2). The concentration of chlorophyll *a* and carotenoids were calculated with Equations 3.1–3.2 and plotted (Fig. 3.5).

To assess differences in chlorophyll *a* and carotenoid content across the various strains, statistical analyses were performed using one-way analysis of variance (ANOVA) followed by Tukey's honest significant difference (HSD) post hoc test (Tables S3.3 and S3.4). A significant effect was observed in chlorophyll *a* content ($F(4, 10) = 23.72$, $p = 4.4 \times 10^{-5}$). Pairwise comparisons revealed that chlorophyll *a* levels were significantly reduced in the KO, D39N-rescue, and ΔαCD-rescue strains compared to WT. The WT-rescue strain exhibited partial rescue in chlorophyll *a* content ($p = 0.048$ vs. WT). It is important to consider that the deficiency of chlorophyll *a* in the ΔαCD-rescue strain may simply be due to the low levels of Slr1794 in this strain (Tables 3.1 and S3.6). A significant effect was also observed in carotenoid content ($F(4, 10) = 4.15$, $p = 0.030$). Pairwise comparisons showed that carotenoid levels were significantly lower in the KO strain compared to WT ($p = 0.0418$ vs. WT), while all other pairwise comparisons were not statistically significant, including the WT-rescue, D39N-rescue, and ΔαCD-rescue strains, which restored carotenoids to near-WT levels.

Co-Immunoprecipitation to Identify Putative Interaction Partners of Slr1794

To identify putative interaction partners of Slr1794, co-immunoprecipitation coupled to mass spectrometry (co-IP/MS^2) experiments were performed. Both wild-type PCC 6803 and cells expressing Slr1794-3xFLAG under the native promoter were subjected to anti-FLAG co-IP (3 independent cultures each). Western blotting was performed to validate the success of the co-IP (Fig. 3.6A).

A Volcano plot was prepared (Fig. 3.6B) to visualize the putative interaction partners. All proteins identified by mass spectrometry are reported in Table S3.9. For each identified protein, abundance ratios for Slr1794-3xFLAG versus wild-type PCC 6803 were calculated. Proteins that were significantly enriched in the Slr1794-3xFLAG sample ($\log_2$(fold-change) > 0.585 and adjusted p-value < 0.05) are shown in Table 3.3. These 120 proteins are putative interaction partners.

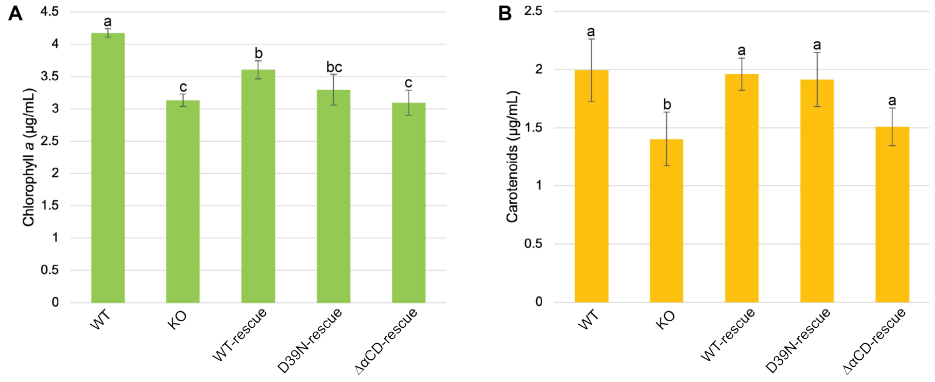


Figure 3.5: Pigment content of *slr1794* knockout and complementation strains. (A) Chlorophyll *a* and (B) carotenoid concentrations (μg/mL) determined using methanol extraction and spectrophotometry per standard procedures (*33*). For each strain, three independent cultures were grown, and pigments were extracted from whole cells at equal OD_{730}. Raw data is shown in Table S3.2. Error bars represent one standard deviation. Letters *above* each bar correspond to significance as calculated by Tukey's honest significant difference (HSD) post hoc test (Tables S3.3 and S3.4), where the same letter denotes no statistically significant difference between samples. Strains per the following: wild-type (WT), *slr1794* knockout (KO), and complementation strains (WT-rescue, D39N-rescue, and $\Delta\alpha$CD-rescue).

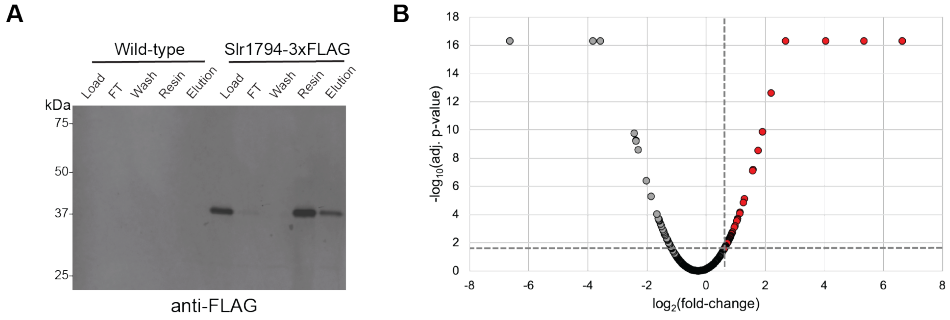


Figure 3.6: Identification of putative interaction partners of Slr1794. Slr1794-3xFLAG and wild-type *Synechocystis* sp. PCC 6803 (PCC 6803) were subjected to co-immunoprecipitation coupled to mass spectrometry (co-IP/MS^2). (A) Anti-FLAG western blot of co-IP samples. Load, flow-through (FT), wash, resin, and elution fractions were analyzed. Molecular weight marker (kDa) is shown to the *left*. (B) Volcano plot analysis of the co-IP/MS^2 of Slr1794-3xFLAG versus wild-type PCC 6803. $\log_2$(fold-change) versus -$\log_{10}$(adjusted [adj.] p-value) is shown. Enrichment and significance cuts offs are shown with dotted lines. Significantly enriched proteins are plotted in *red*, and all other identified proteins are plotted in *grey*. Information for all the identified proteins is reported in Table S3.9, and the significantly enriched proteins are reported in Table 3.3.

Table 3.3: Proteins significantly enriched in the Slr1794-3xFLAG sample.

Uniprot ID	Description	$\log_2$(fold-change)	Adjusted p-value
P72799	Slr1794 protein	6.64	5.09e-17
Q55873	Slr0103 protein	6.64	5.09e-17
P73936	Sll1429 protein	6.64	5.09e-17
P74524	Slr1420 protein	6.64	5.09e-17
P73390	Poly(3-hydroxyalkanoate) polymerase subunit PhaC	6.64	5.09e-17
P74555	Slr1462 protein	6.64	5.09e-17
P72863	Slr0971 protein	6.64	5.09e-17
P74307	Slr0941 protein	6.64	5.09e-17
P54984	Probable hydrolase sll0100	6.64	5.09e-17
P73492	Probable glutaredoxin ssr2061	6.64	5.09e-17
P73146	Slr1035 protein	6.64	5.09e-17
P74016	Processing protease	6.64	5.09e-17
P74444	Slr0145 protein	6.64	5.09e-17
P73770	Slr1243 protein	6.64	5.09e-17
Q55711	Slr0634 protein	6.64	5.09e-17
Q55694	tRNA uridine 5-carboxymethylaminomethyl modification enzyme MnmG	6.64	5.09e-17
P72637	UPF0182 protein sll1060	6.64	5.09e-17
P73549	Sll1268 protein	6.64	5.09e-17
P72884	ABC transporter	6.64	5.09e-17
P72765	Slr1772 protein	6.64	5.09e-17
P74100	Extracytoplasmic solute receptor protein	6.64	5.09e-17
P74370	Dioxygenase	6.64	5.09e-17
Q55769	ComE ORF1	6.64	5.09e-17
P73865	Sensory transduction histidine kinase	6.64	5.09e-17
Q55637	Ribonuclease 3 2	6.64	5.09e-17
P74368	Ribonuclease 3 1	6.64	5.09e-17
Q55836	Slr0514 protein	6.64	5.09e-17
P74218	Hydrogenase maturation factor HypB	6.64	5.09e-17
Q55853	Sll0595 protein	6.64	5.09e-17
P72957	Sll0641 protein	6.64	5.09e-17
Q55773	Sll0183 protein	6.64	5.09e-17
P74300	Slr0935 protein	6.64	5.09e-17
Q6ZEW9	UPF0235 protein ssr5011	6.64	5.09e-17
Q55190	High-affinity Na(+)/H(+) antiporter NhaS3	6.64	5.09e-17
P73585	N-acetylmuramic acid 6-phosphate etherase	6.64	5.09e-17
P73207	Slr1704 protein	6.64	5.09e-17
Q55546	Sll0294 protein	6.64	5.09e-17
P73700	Slr1815 protein	6.64	5.09e-17
P74282	Slr1686 protein	6.64	5.09e-17
P74453	Sll0148 protein	6.64	5.09e-17
P73213	Ssr2857 protein	6.64	5.09e-17
P74242	Slr1166 protein	6.64	5.09e-17
P74564	Photosystem I reaction center subunit PsaK 2	6.64	5.09e-17
Q55744	Sll0381 protein	6.64	5.09e-17
P73014	50S ribosomal protein L32	6.64	5.09e-17

Continued on next page

Table 3.3: *(Continued)*

Uniprot ID	Description	$\log_2$(fold-change)	Adjusted p-value
P53383	Iron-sulfur cluster carrier protein	6.64	5.09e-17
P72909	Slr1073 protein	6.64	5.09e-17
Q55440	Sucrose-phosphate synthase	6.64	5.09e-17
P74791	Ssr0692 protein	6.64	5.09e-17
Q55455	Prolycopene isomerase	6.64	5.09e-17
Q55034	Glucosylglycerol-phosphate phosphatase	6.64	5.09e-17
P29273	Phytoene dehydrogenase	6.64	5.09e-17
Q55503	ATP phosphoribosyltransferase	6.64	5.09e-17
P73656	Ssr3189 protein	6.64	5.09e-17
P74537	Sll1334 protein	6.64	5.09e-17
P73200	Slr1702 protein	6.64	5.09e-17
Q55424	Putative amidase	6.64	5.09e-17
P74573	Beta-fructofuranosidase	6.64	5.09e-17
P73380	Slr2077 protein	6.64	5.09e-17
Q55869	Chorismate mutase AroH	6.64	5.09e-17
Q55392	Slr0565 protein	6.64	5.09e-17
Q55875	RuBisCO accumulation factor 1	6.64	5.09e-17
P73235	Slr2015 protein	6.64	5.09e-17
P73449	Nitrate transport protein NrtD	6.64	5.09e-17
Q55983	Queuine tRNA-ribosyltransferase	6.64	5.09e-17
Q55354	Photosystem II reaction center protein L	6.64	5.09e-17
Q6ZEA8	DUF3692 domain-containing protein	6.64	5.09e-17
Q55413	Slr0825 protein	6.64	5.09e-17
P73716	Slr1732 protein	6.64	5.09e-17
P74455	High-affinity branched-chain amino acid transport protein BraE	6.64	5.09e-17
P52965	Putative pyruvate-flavodoxin oxidoreductase	6.64	5.09e-17
Q55158	Riboflavin biosynthesis protein RibD	6.64	5.09e-17
P73765	Slr0869 protein	6.64	5.09e-17
P72843	Regulatory components of sensory transduction system	6.64	5.09e-17
P74483	Slr1931 protein	6.64	5.09e-17
P73388	Ferredoxin	6.64	5.09e-17
Q55004	50S ribosomal protein L34	5.34	5.09e-17
P73300	50S ribosomal protein L36	4.05	5.09e-17
P74135	Sll1873 protein	2.69	5.09e-17
P72769	Slr1774 protein	2.20	2.42e-13
Q6ZE90	Sll8002 protein	1.91	1.38e-10
Q55386	Slr0924 protein	1.76	2.97e-09
Q55622	Transcriptional repressor NrdR	1.59	6.71e-08
P73565	Sll0872 protein	1.58	7.91e-08
P72907	Slr1071 protein	1.30	7.70e-06
P73351	Slr1201 protein	1.26	1.45e-05
P73903	Glutaminase	1.15	6.87e-05
P74625	Photosystem I-associated linker protein CpcL	1.14	8.39e-05
P74187	Slr1275 protein	1.07	1.99e-04
Q55625	Ribosome-binding factor A	1.06	2.36e-04
Q55451	Sirohydrochlorin cobaltochelatase	1.04	2.96e-04

Continued on next page

Table 3.3: *(Continued)*

Uniprot ID	Description	$\log_2$(fold-change)	Adjusted p-value
P73638	Phycocyanobilin lyase subunit alpha	0.98	6.46e-04
P74261	Uncharacterized tRNA/rRNA methyltransferase slr1673	0.96	8.01e-04
P54224	Uroporphyrinogen decarboxylase	0.89	1.79e-03
P73320	50S ribosomal protein L3	0.89	1.81e-03
P74267	50S ribosomal protein L27	0.87	2.29e-03
Q55876	Slr0105 protein	0.86	2.46e-03
P73939	Slr1503 protein	0.84	3.11e-03
P73293	30S ribosomal protein S9	0.84	3.20e-03
P48949	30S ribosomal protein S21	0.83	3.38e-03
P72621	dTDP-4-dehydrorhamnose reductase	0.82	3.86e-03
P73315	50S ribosomal protein L22	0.81	4.43e-03
P48054	Imidazoleglycerol-phosphate dehydratase	0.79	5.21e-03
P73306	50S ribosomal protein L6	0.78	5.92e-03
Q55332	Photosystem II 12 kDa extrinsic protein	0.78	6.04e-03
P73313	50S ribosomal protein L16	0.76	6.90e-03
P72752	Thylakoid membrane protein ssl2009	0.75	7.65e-03
P73953	Slr1512 protein	0.74	8.93e-03
P74609	Slr1576 protein	0.74	9.09e-03
Q55972	Slr0708 protein	0.73	9.35e-03
P72958	Uncharacterized transporter sll0640	0.72	1.05e-02
P72914	Ssr1766 protein	0.67	1.67e-02
P36239	50S ribosomal protein L19	0.67	1.71e-02
P73947	Slr1507 protein	0.66	1.94e-02
P72675	Slr0731 protein	0.65	2.08e-02
P73152	Uncharacterized thylakoid-associated protein sll0982	0.64	2.17e-02
P74554	Sll1365 protein	0.64	2.26e-02
P72737	Slr1094 protein	0.62	2.55e-02
Q55862	Sll0588 protein	0.62	2.66e-02
P73702	Sll1696 protein	0.61	3.00e-02

Note: Uniprot ID, protein description, $\log_2$(fold-change), and adjusted p-value for proteins significantly enriched ($\log_2$(fold-change) > 0.585 and adjusted p-value < 0.05) in the FLAG co-immunoprecipitation coupled to mass spectrometry analysis of Slr1794-3xFLAG versus wild-type PCC 6803.

The Database for Annotation, Visualization and Integrated Discovery (DAVID) (*34*) was utilized for functional annotation clustering of proteins enriched in the Slr1794-3xFLAG co-IP/MS^2 pulldown compared to wild-type PCC 6803 (Table S3.10). This analysis revealed a strong enrichment of ribosome-associated proteins, including structural ribosomal components (Cluster 1, enrichment score = 3.04), rRNA-binding proteins (Cluster 2, enrichment score = 1.36), and nucleases (Cluster 3, enrichment score = 0.58). Additional enrichment was observed for membrane-associated and thylakoid-localized proteins (Clusters 4–5), consistent with a potential role for Slr1794 in interactions near the photosynthetic membrane.

A smaller subset of ATP-binding and NTPase-domain-containing proteins was also identified (Cluster 6), suggesting possible association with transport proteins or protein targeting factors.

Of particular interest are photosynthesis-related and membrane proteins, as these proteins are promising Get3d clients. The photosynthesis-related proteins identified are PsaK2, RuBisCO accumulation factor 1 (Raf1), PsbL, PsbU, and Photosystem I-associated linker protein CpcL (Tables 3.3 and S3.9). To further probe the putative membrane protein interaction partners, the list of 120 enriched proteins was cross-referenced against all *Synechocystis* sp. PCC 6803 proteins annotated with one or more TMDs in Uniprot (*35*). The membrane proteins significantly enriched in the FLAG co-IP/MS^2 are reported in Table 3.4 with their number of TMDs and localization (if annotated) reported. Of the photosynthesis-related proteins, PsaK2 has two TMDs. PsbL is a short single TMD membrane protein and CpcL is a canonical TA protein, making them ideal clients for TA protein targeting factor machinery. Of the other membrane proteins identified, the final TA protein is the Nitrate transport protein NrtD.

3.4 Discussion

The central targeting factor in the guided entry of tail-anchored (TA) proteins (GET) pathway is the homodimeric P-loop ATPase Get3 (reviewed in (*5*, *7*)). We recently identified a novel homolog of Get3, which we termed Get3d (*16*). Get3d is conserved in plants and photosynthetic bacteria, suggesting it is important in the evolution of photosynthesis. To date, the function of Get3d is unknown. This work seeks to characterize the "plant-like" Get3d paralog in the model cyanobacterium *Synechocystis* sp. PCC 6803 to gain insights into the function of plant-like Get3d homologs.

All Get3d proteins contain the C-terminal α-crystallin domain (αCD) with unknown function (*16*). The conservation of the other important sequence features varies across Get3d homologs. Here, we investigated the conservation of the functional motifs in Get3d, showing that there are two subclasses of Get3d proteins, which we termed "plant-like" and "canonical" Get3d (Figs. 3.1A and S3.3). The plant-like Get3d homologs have divergent active site motifs and lack key features of canonical Get3, including the A-loop and the pivot helix, which serves as a rotation point for the conformational changes during the ATPase cycle (*26*, *36*). The canonical Get3d homologs have highly conserved active site motifs and retain the A-loop and pivot helix (*26*, *36*). The difference in the conservation of these features suggests the two

Table 3.4: Membrane proteins significantly enriched in the Slr1794-3xFLAG sample.

Uniprot ID	Description	No. TMD	Localization
P72637	UPF0182 protein sll1060	9	Cell
P72765	Slr1772 protein	1 – SA	Outer
P72863	Slr0971 protein	1	
P72957	Sll0641 protein	9	
P73235	Slr2015 protein	1 – SA	
Q55354	Photosystem II reaction center protein L	1 – short	Thylakoid
P73380	Slr2077 protein	1	Cell
P73449	Nitrate transport protein NrtD	1 – TA	Cell
P73549	Sll1268 protein	1 – SA	
P73865	Sensory transduction histidine kinase	3	
P74300	Slr0935 protein	2	
P74455	High-affinity branched-chain amino acid transport protein BraE	9	Cell
P74483	Slr1931 protein	1 – SA	
P74555	Slr1462 protein	2	
P74564	Photosystem I reaction center subunit PsaK2	2	Thylakoid
Q55190	High-affinity Na(+)/H(+) antiporter NhaS3	11	Thylakoid
Q55392	Slr0565 protein	5	
Q55769	ComE ORF1	1 – SA	
P73351	Slr1201 protein	8	Cell
P74625	Photosystem I-associated linker protein CpcL	1 – TA	Thylakoid
P74187	Slr1275 protein	1 – SA	
P72752	Thylakoid membrane protein ssl2009	1 – SA	Thylakoid
P73953	Slr1512 protein	8	
P72958	Uncharacterized transporter sll0640	12	Cell
P74554	Sll1365 protein	2	

Note: Uniprot ID, protein description, number (No.) transmembrane domains (TMD), and annotated membrane localization (if any) of membrane proteins enriched in the FLAG co-immunoprecipitation coupled to mass spectrometry. Membrane proteins were identified by searching Uniprot for proteins with one or more TMD in *Synechocystis* sp. PCC 6803 (*35*). Signal-anchored (SA) proteins were manually identified as those whose TMD is within 30 residues of their N-terminus. Tail-anchored (TA) proteins were manually identified as those whose TMD is within 30 residues of their C-terminus. Short proteins were manually identified as those whose TMD is within 30 residues of their N-terminus and 30 residues of their C-terminus.

subgroups of Get3d homologs may have divergent functions.

AlphaFold3 was utilized to predict the structures of the two Get3d paralogs in PCC 6803: Slr1794, the plant-like Get3d, and Sll0086, the canonical Get3d (Figs. 3.1B, C and S3.2) (*24*). These structures share the same overall fold with a the nucleotide binding domain (NBD), client binding domain (CBD), the αCD. The confidence of these predictions is high, particularly in the NBDs, αCDs, and the CBD of Slr1794 (Fig. S3.2). As expected, the Slr1794 predicted structure matches published plant-like Get3d structures, while the Sll0086 predicted structure has more features of cytosolic Get3 (Fig. 3.2).

A PCC 6803 strain was prepared with a *slr1794* gene deletion (knockout, KO) to investigate the proteomic and phenotypic changes that occur upon the loss of the *slr1794* gene. Whole-cell proteomic analysis revealed 165 differentially expressed proteins in the KO strain (Fig. 3.3A, Tables S3.6 and S3.7). Functional cluster enrichment analysis revealed significant up-regulation of proteins with DNA-binding domains in the KO strain, suggesting that transcriptional regulators are up-regulated in the absence of Slr1794 (Table S3.8). This indicates a broad compensatory response to the proteomic imbalance resulting from the deletion of the *slr1794* gene. A functional cluster of proteins containing thioredoxin-related domains was identified in the down-regulated proteins, suggesting a loss of redox balancing mechanisms upon *slr1794* deletion (Table S3.8). Together, these results suggest that the absence of Slr1794 causes broad proteomic pertubations and reveal that Slr1794 may function in redox regulation.

No significantly enriched clusters were identified in the WT-rescue strain (KO compensated with WT *slr1794*) versus the KO (Table S3.8). This, along with the low number of differentially expressed proteins (Fig. 3.3B, Tables S3.6 and S3.7), shows that the proteomic balance is restored in this strain. Most interestingly, the D39N-rescue strain (KO complemented with ATPase-deficient mutant) compared to the KO reveals significant enrichment in down-regulated proteins annotated with photosynthesis-related terms (Table S3.8). This suggests that the ATP-deficient mutant is unable to restore the expression of key proteins involved in photosynthesis, including multiple photosystem II, phycobilosome, and cytochrome b_6f components (Table 3.2). The $\Delta\alpha$CD-rescue (KO complemented with the C-terminal αCD truncation) showed 136 differentially expressed proteins, which is approaching KO levels of dysregulation (Fig. 3.3D, Tables S3.6 and S3.7). Quantification of Slr1794 levels between the various strains reveals very low levels of Slr1794 in the $\Delta\alpha$CD-rescue

strain (~250× less than WT levels) (Tables 3.1, S3.6, and S3.7). This could be due to many factors such as poor mRNA or protein stability, suggesting that the αCD is vital to maintaining the proper levels of functioning Slr1794 in the cell. Of the proteins differentially expressed in the $\Delta\alpha$CD-rescue strain compared to the KO, there is significant down-regulation of CRISPR-related and anti-viral defense pathway proteins (Tables S3.8, S3.6, and S3.7). As the levels of Slr1794 in this strain are so low, this enrichment probably does not represent a true biological function but rather a modest suppression of stress-response pathways in the $\Delta\alpha$CD-rescue strain that is highly up-regulated in the KO strain.

Overall, our whole-cell proteomic analysis reveals that the deletion of the *slr1794* gene leads to widespread changes in protein abundance (Fig. S3.6), including up-regulation of DNA-binding transcriptional regulators and down-regulation of thioredoxin-related redox proteins (Table S3.8). This is consistent with cellular stress and impaired homeostasis. Complementation with WT *slr1794* restored protein expression profiles to near-WT levels, indicating successful rescue (Fig. S3.6). In contrast, rescue with the ATPase-deficient D39N mutant failed to restore proteomic homeostasis (Fig. S3.6), resulting in substantial down-regulation of photosynthesis-related proteins (Table S3.8). These results demonstrate that the ATPase activity is essential for Slr1794 function and suggests that Slr1794 serves in maintaining redox balance and photosynthesis.

The phenotypic analysis of these strains revealed that *slr1794* is essential for cellular health. Spot growth assays revealed a significant growth defect in the *slr1794* deletion strain that is recovered with complementation with both wild-type *slr1794* and the ATPase-deficient mutant, revealing the ATPase activity of Slr1794 is not required for overall cellular health (Fig. 3.4). On the other hand, complementation with the C-terminal αCD deletion does not rescue this growth defect, likely due to the decreased levels of Slr1794 in this strain overall.

Pigments are responsible for absorbing light energy, with different pigments absorbing different wavelengths of light (*37*). Chlorophyll *a* is the primary pigment in PCC 6803 and absorbs light in the red and blue regions of the visible light spectrum. Carotenoids are a secondary pigment in PCC 6803 that absorb light in the blue-green and violet regions of the spectrum. Carotenoids also play an important role in photoprotection, as they dissipate excess energy as heat to prevent photoinhibition (*38*). To probe the functional relevance of the *slr1794* gene on pigment levels, methanol extraction and spectrophotometry were used to analyze pigment content

in the various strains (Fig. 3.5, Tables S3.2, S3.3, and S3.4). The levels of chlorophyll *a* and carotenoids were significantly lessened upon *slr1794* gene deletion (Fig. 3.5A). The decrease in chlorophyll *a* levels is partially rescued with the reintroduction of the WT *slr1794* gene and not rescued in the ATPase-deficient rescue mutant (D39N-rescue) or the αCD deletion rescue mutant (αCD-rescue). Carotenoid levels were maintained at WT levels in all of the rescue strains (Fig. 3.5B). These results suggest that Slr1794 may play a role in pigment production or photosynthesis and that its ATPase activity is required for the maintenance of chlorophyll *a* levels.

Co-immunoprecipitation coupled to mass spectrometry (co-IP/MS^2) identified 120 putative interaction partners (Fig. 3.6 and Table S3.9). Of the 120 proteins, functional cluster enrichment analysis revealed enrichment in photosynthesis-related and membrane proteins, supporting the hypothesis that Slr1794 functions in membrane protein targeting, particularly of photosynthesis-related proteins (Table S3.10). Of the enriched proteins, 25 are membrane proteins, two are TA proteins, and one is a short (single TMD) protein that is a candidate for TA protein targeting pathways (Table S3.6). Structural predictions were performed to probe interactions between Slr1794 and the photosynthesis-related proteins and TA proteins. Structures were predicted using AlphaFold3 of the Slr1794 dimer with the following proteins (and Mg^{2+}): PsaK2, Raf1, PsbL, PsbU, CpcL, and NrtD (Figs. S3.4 and S3.5). Each of these proteins, excluding Raf1, were predicted to interact with Slr1794 in the putative TA protein binding chamber.

It is important to consider that there are many limitations in protein structure prediction. For each prediction, AlphaFold3 also performs a predicted local distance difference test (pLDDT), which gives a measure of the confidence of the prediction (*24*). For each of these predicted complex structures, the confidence in the placement of the TMD (or predicted lumenal targeting signal in the case of PsbU) is relatively low (Fig. S3.5). While these predictions suggest that a real biological interaction may occur, the overall inaccuracy of protein structure predictions (particularly in the case of protein complexes) and the low confidence in the placement of the putative interaction partner helices necessitate a skeptical analysis of these predictions.

This work investigated the functional role of Slr1794, the plant-like Get3d paralog, in the model cyanobacterium *Synechocystis* sp. PCC 6803. Through a combination of whole-cell proteomics, spot growth assays, and pigment analysis, the data support a role for Slr1794 in maintaining redox balance and supporting photosynthetic function. Co-immunoprecipitation coupled to mass spectrometry identified putative

Slr1794 interaction partners enriched in membrane-associated and photosynthesis-related proteins. Whole-cell proteomic analysis further demonstrated that Slr1794's ATPase activity is required to sustain normal expression of photosynthetic complexes, while the C-terminal α-crystallin domain appears essential for maintaining stable levels of Slr1794 itself. This study provides the first functional characterization of a plant-like Get3d homolog *in vivo*. While further work will be needed to fully elucidate its mechanistic role, these findings suggest that Slr1794 contributes to core physiological processes central to redox homeostasis and photosynthesis in cyanobacteria.

3.5 Methods

Note: All reagents were purchased from Sigma-Aldrich unless otherwise specified.

Sequence Analysis

Protein sequences were obtained from Uniprot (*35*, *39*). Get3d homologs from representative species from each order of cyanobacteria were obtained using Protein BLAST with the query sequences Slr1794 (Uniprot ID: P72799) and Sll0086 (Uniprot ID: Q55794) (*21*). Sequence alignments were prepared with multiple alignment using fast Fourier transform (MAFFT) (*17*). Sequence alignments were visualized in Jalview (*18*) and colored per the ClustalX color scheme (*19*). Sequence homology was analyzed using the Sequence Manipulation Suite (*22*).

Structural Prediction and Analysis

Protein structures were predicted using AlphaFold3 (*24*), and predicted local distance difference test (pLDDT) results were colored per a modified coloring scheme. Structural figures were generated using PyMOL (https://pymol.org/) (*25*). Structural alignments, including calculations of root mean square deviation (RMSD), were prepared using PyMOL.

Plasmids and Cloning

Table 3.5 shows a list of constructs used in this study. All plasmids used the backbone from the pAM1579 plasmid (a gift from Susan Golden, Addgene plasmid #40240, based on the pBR322 replicon), which contains an ampicillin resistance cassette for propagation in *Escherichia coli* (*E. coli*) and a kanamycin resistance cassette for selection in *Synechocystis* sp. PCC 6803.

The *slr1794* and *psbA2* genes and flanking regions were obtained from CyanoBase

Table 3.5: Strains used in this study.

Strain	Description	Antibiotic resistance(s)
WT	Wild-type *Synechocystis* sp. PCC 6803	None
KO	*slr1794* knockout (Kan^R)	Kan^R
WT-rescue	KO complemented with WT *slr1794* at the *psbA2* locus	Kan^R/Cm^R
D39N-rescue	KO complemented with ATPase-deficient mutant (D39N) at the *psbA2* locus	Kan^R/Cm^R
$\Delta\alpha$CD-rescue	KO complemented with C-terminal αCD truncation ($\Delta\alpha$CD) at the *psbA2* locus	Kan^R/Cm^R
Slr1794-3xFLAG	*slr1794* with C-terminal 3xFLAG tag at the *slr1794* locus	Kan^R

(`http://genome.microbedb.jp/cyanobase/`) and amplified from *Synechocystis* sp. PCC 6803 genomic DNA (ATCC #27184D-5). The 3xFLAG gene was synthesized by Millipore Sigma. The chloramphenicol resistance cassette was obtained from the pEERM4 plasmid (a gift from Pia Lindberg, Addgene plasmid #64026). All primers were purchased from Integrated DNA Technologies (IDT).

Primers for Gibson assembly were designed using NEBuilder (`https://nebuilder.neb.com/`), and primers for deletions, insertions shorter than 50 bp, and point mutations were designed using NEBaseChanger (`https://nebasechanger.neb.com/`). PCR reactions were performed with 2X Q5 High Fidelity Master Mix (New England Biolabs, NEB) per the following protocol: $1 \times 98°C$ for 30 sec; $35 \times$ (a) 98°C for 10 sec, (b) respective annealing temperature for 30 sec, and (c) 72°C for 30 sec per kilobase (kb); $1 \times 72°C$ for 2 mins (if shorter than 2 kb) or 5 mins (if longer than 2 kb); hold at 4°C. Annealing temperatures were determined during primer design. PCR products were analyzed agarose gel electrophoresis was performed.

Fragments for Gibson assembly were incubated with DpnI (NEB) for 1 to 3 hrs at 37°C and purified via QIAquick PCR Purification or QIAquick Gel Extraction (Qiagen). Gibson assembly reactions were performed using 2X HiFi Master Mix (NEB) at 50°C for 15 mins or 1 hr following standard protocols. Required insert DNA mass was calculated using NEBioCalculator (`https://nebiocalculator.neb.com/`). For deletions, insertions shorter than 50 bp, and point mutations, Kinase, Ligase, and DpnI (KLD) reactions were performed using 10X KLD Enzyme Mix (NEB) at room temperature for 15 mins or 1 hr.

Resultant products were transformed into NEB10β *E. coli* (NEB) or Top10 *E.*

coli (Thermo Fisher Scientific) per standard protocols. Plasmids were sent to Plasmidsaurus Labs for full plasmid sequencing.

Synechocystis sp. PCC 6803 Growth

Synechocystis sp. PCC 6803 was a gift from Dianne Newman. BG-11 agar plates and liquid media were prepared per standard protocols using 100x BG-11 concentrate (Millipore Sigma #73816) and 1000x Trace Metal Mix A5 with Co (Millipore Sigma #92949) with 10 mM TES pH ~8.2 (for agar plates) or 50 mM HEPES pH ~8.2 (for liquid media) with or without antibiotic(s) (*28*). Antibiotic concentrations per the following: 25 μg/mL kanamycin sulfate, 20 μg/mL chloramphenicol (Research Products International).

Plates were streaked and incubated at 32°C under constant illumination (~ 5μmol/m^2/s) by 20W Full Spectrum (380 nm-780 nm) LED Grow Lights (DOMMIA). Liquid cultures were inoculated from fresh colonies or thawed 15% glycerol stocks stored at –80°C and incubated at 32°C and 170 rpm under constant illumination (~100 μmol/m^2/s).

Synechocystis sp. PCC 6803 Transformation and Segregation

Transformations were performed per (*28*) for integration by homologous recombination. Briefly, cultures were used for transformations at OD_{730} = 0.3 – 0.5. ~1 μg plasmid was transformed per strain, and 200 μL cells were plated on Whatman Nuclepore Hydrophilic Membranes (0.4 μm pore size, 47 mm circle, Cytiva #10417112) on BG-11 agar plates without antibiotics. After 12 hrs, membranes were transferred to BG-11 agar with 1/2 concentration antibiotic(s) (12.5 μg/mL kanamycin sulfate or 10 μg/mL chloramphenicol), and after 2–3 days, membranes were transferred to BG-11 agar with full concentration antibiotic(s) (25 μg/mL kanamycin sulfate or 20 μg/mL chloramphenicol). After ~2 weeks, colonies were streaked onto BG-11 agar plates with antibiotic(s) followed by weekly streaking to allow for full segregation.

To confirm successful transformation and complete segregation, PCR was performed to amplify the *slr1794* and *psbA2* loci of single colonies resuspended in sterile water and liquid cultures for each PCC 6803 strain. The primers used to ampilfy the loci are reported in Table S3.5. Reactions were performed per the 2X Q5 High Fidelity Master Mix protocol described above or the 2X OneTaq Master Mix (NEB) protocol as follows: 1 × 94°C for 30 sec; 35 × (a) 94°C for 20 sec, (b) respective annealing temperature for 30 sec, and (c) 68°C for 1 min per kb; 1 ×

68°C for 5 mins; hold at 4°C, followed by agarose gel electrophoresis. Annealing temperatures were determined during primer design or using NEB Tm Calculator (`https://tmcalculator.neb.com/`).

Whole-Cell Mass Spectrometry and Analysis

For each strain, three 50 mL independent cultures (OD_{730} ~0.75) were pelleted and resuspended in lysis buffer (50 mM triethylammonium bicarbonate (TEAB) [pH 8.5], 5% SDS). Cells were lysed by sonication with a Misonix S-4000 sonicator (2 × 2 s on (80% amplitude), 2 mins off (on ice)) using the MicroTip. Samples were prepared following the S-trap micro spin column digestion protocol (ProtiFi).

Samples were analyzed using the Orbitrap Eclipse Tribid Mass Spectrometer coupled to Vanquish Neo UHPLC by the Proteome Exploration Laboratory in Beckman Institute at Caltech. 2 μg peptides from each sample was injected and separated on an Aurora UHPLC Column (60 cm × 75 μm, 1.7 μm C18, AUR3-60075C18-TS, Ion Opticks) with a flow rate of 0.30 μL/min for a total duration of 2 hrs and ionized at 1.8 kV in the positive ion mode. Raw data were searched against the *Synechocystis* sp. (strain PCC 6803 / Kazusa) database using the Proteome Discover 3.1 software based on the Sequest HT algorithm. Oxidation / +15.995 Da (M), Deamidated / +0.984 Da (N), protein N-terminal Met-loss / -131.040 Da and protein N-terminal acetylation / +42.011 Da were set as dynamic modifications; carbamidomethylation / +57.021 Da(C) was set as fixed modification. The precursor mass tolerance was set to 10 ppm, whereas fragment mass tolerance was set to 0.05 Da. The maximum false peptide discovery rate was specified as 0.01 using the Percolator Node. Post-analysis was performed using Tidyproteomics 1.8.3 (`jeffsocal.github.io/tidyproteomics/index.html`).

Functional annotation enrichment was performed using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (*34*). UniProt IDs from differentially expressed protein sets were submitted to the DAVID Functional Annotation Clustering tool using default settings. Enriched Gene Ontology (GO), UniProt Keywords (KW), and InterPro (IPR) terms were grouped into annotation clusters. Clusters with an enrichment score greater than 1.3 (equivalent to a geometric mean p-value < 0.05) were considered significant.

***Synechocystis* sp. PCC 6803 Spot Arowth assays**

Synechocystis sp. PCC 6803 cells at OD_{730} < 1.0 were diluted to an OD_{730} = 0.5 in BG-11 media, and a 1:5 dilution series was prepared in BG-11 media. 3 μL spots

were plated on BG-11 agar without antibiotics. Plates were incubated at 32°C under constant illumination as above.

Pigment Content Analysis

Chlorophyll *a* and carotenoids were extracted and quantified per standard methanol extraction protocols (*33*). Briefly, three independent cultures for each strain were grown to OD_{730} = 0.5 – 1.0, and 5 mL each culture was harvested and resuspended in 1.25 mL BG-11 media followed by dilution to an equal OD_{730} in BG-11 media.

Briefly, 1 mL each culture (at an equal OD_{730}) was centrifuged at 15,000 xg for 7 mins at room temperature. Pellets were carefully resuspended in ice-cold methanol then incubated at 4°C for 20 mins in the dark. Samples were centrifuged at 15,000 xg for 7 mins at 4°C. The absorbance (in arbitrary units) at 470 nm, 665 nm, and 720 nm of each supernatant was measured with the GENESYS 180 UV-Vis Spectrophotometer (Thermo Scientific), blanking with methanol (Table S3.2).

Pigment concentrations were calculated according to the following equations:

$$\text{Chl } a \ [\mu\text{g/mL}] = 12.9447 \times (A_{665} - A_{720}) \tag{3.1}$$

$$\text{Carotenoids } [\mu\text{g/mL}] = \frac{1{,}000 \times (A_{470} - A_{720}) - 2.86 \times \text{Chl } a \ [\mu\text{g/mL}]}{221} \tag{3.2}$$

Statistical analyses were performed in Python using the `statsmodels` package (*40*). A one-way analysis of variance (ANOVA) was conducted for each pigment type (chlorophyll *a* and carotenoids) to compare pigment levels across the strains. Tukey's honest significant difference (HSD) test was applied as a post hoc analysis to evaluate pairwise differences between groups while controlling the family-wise error rate (Tables S3.3 and S3.4). Adjusted p-values (p-adj) were reported, and differences were considered statistically significant at $p < 0.05$.

Co-Immunoprecipitation Sample Preparation

For both WT PCC 6803 and Slr1794-3xFLAG, three 50 mL independent cultures (OD_{730} ~0.75) were pelleted and resuspended in ~1.5 mL lysis buffer (50 mM Tris-HCl [pH 7.5], 150 mM NaCl, 1 mM EDTA, 1% Triton X-100) with 1xHALT Protease and Phosphatase Inhibitor, EDTA-Free (Thermo Fisher Scientific) to an equal OD_{730}. Cells were lysed by bead beating in a MP Biomedicals FastPrep-24 with 0.1 mm silica beads (MP Biomedicals) at 7 × 30 sec on at 6 M/s and 5 min off on ice. Cell debris was pelleted, and clarified lysate was removed.

Co-immunoprecipitation (co-IP) was performed following manufacturer's instructions with ANTI-FLAG M2 Magnetic Beads (Millipore Sigma #M8823). Briefly, 20 μL packed gel volume was equilibrated with 1xTBS (50 mM Tris [pH 7.4], 150 mM NaCl) and incubated with 0.9 mL clarified lysate overnight at 4°C on a roller shaker. Beads were isolated and washed in 1xTBS and either eluted with sample buffer for analysis by SDS-PAGE and western blotting or treated with protease for on-bead digestion.

For analysis by western blotting, co-IP samples were run on 15% SDS-PAGE and transferred to a 0.2 μm nitrocellulose membrane using the Transblot Turbo System (Bio-Rad). The membrane was blocked with 5% dry milk in TTBS (20 mM Tris–HCl [pH 7.5], 150 mM NaCl, 0.1% Tween-20) at room temperature for 1 hr and incubated with anti-FLAG antibody (from rabbit) (Rockland Immunochemicals #600-401-383) in 5% dry milk in TTBS overnight at 4°C. The membrane was rinsed with TTBS (3 $\times$ 5 min) and incubated with anti-rabbit antibody (from goat) conjugated to Alkaline Phosphatase (Rockland Immunochemicals #611-1502) in 5% dry milk in TTBS for 3 hrs at room temperature then rinsed with TTBS (3 $\times$ 5 min) and AP Developing Buffer (100 mM Tris–HCl [pH 9.5], 100 mM NaCl, and 5 mM $MgCl_2$, 1 $\times$ 5 min), developed using AP Substrate (0.33 mg/ml NBT and 0.165 mg/ml BCIP) in AP Developing Buffer, and imaged using a ChemiDoc MP Imaging System (Bio-Rad).

Co-Immunoprecipitation Mass Spectrometry and Analysis

Washed beads were treated with endoproteinase Lys-C and trypsin protease. Peptides were collected, and samples were desalted with Pierce C18 spin columns (Thermo Fisher Scientific). Samples were analyzed using the Orbitrap Eclipse Tribid Mass Spectrometer coupled to Vanquish Neo UHPLC by the Proteome Exploration Laboratory in Beckman Institute at Caltech. Peptides were separated on an Aurora UHPLC Column (25 cm $\times$ 75 μm, 1.7 μm C18, AUR3-25075C18-TS, Ion Opticks) with a flow rate of 0.35 μL/min for a total duration of 1 hr and ionized at 1.6 kV in the positive ion mode. Raw data were searched against the *Synechocystis* sp. (strain PCC 6803 / Kazusa) database using the Proteome Discover 3.0 software based on the Sequest HT algorithm. Oxidation / +15.995 Da (M), Deamidated / +0.984 Da (N), protein N-terminal Met-loss / -131.040 Da and protein N-terminal acetylation / +42.011 Da were set as dynamic modifications; carbamidomethylation / +57.021 Da(C) was set as fixed modification. The precursor mass tolerance was set to 10 ppm, whereas fragment mass tolerance was

set to 0.05 Da. The maximum false peptide discovery rate was specified as 0.01 using the Percolator Node. Post-analysis was performed using Tidyproteomics 1.7.3 (`jeffsocal.github.io/tidyproteomics/index.html`). Functional annotation enrichment was performed using DAVID as above.

References

(*1*) Kutay, U., Hartmann, E., and Rapoport, T. A. (1993). A class of membrane proteins with a C-terminal anchor. *Trends Cell Biol. 3*, 72–75, DOI: 10.1016/0962-8924(93)90066-A.

(*2*) Han, J., Pluhackova, K., and Böckmann, R. A. (2017). The multifaceted role of SNARE proteins in membrane fusion. *Frontiers in Physiology 8*, 244875, DOI: 10.3389/fphys.2017.00005.

(*3*) Mehlhorn, D. G., Asseck, L. Y., and Grefen, C. (2021). Looking for a safe haven: tail-anchored proteins and their membrane insertion pathways. *Plant Physiol. 187*, 1916–1928, DOI: 10.1093/plphys/kiab298.

(*4*) Valentijn, A. J., Upton, J.-P., Bates, N., and Gilmore, A. P. (2008). Bax targeting to mitochondria occurs via both tail anchor-dependent and-independent mechanisms. *Cell Death and Differentiation 15*, 1243–1254, DOI: 10.1038/cdd.2008.39.

(*5*) Chio, U. S., Cho, H., and Shan, S.-o. (2017). Mechanisms of tail-anchored membrane protein targeting and insertion. *Annu. Rev. Cell Dev. Biol. 33*, 417–438, DOI: 10.1146/annurev-cellbio-100616-060839.

(*6*) Hegde, R. S., and Keenan, R. J. (2022). The mechanisms of integral membrane protein biogenesis. *Nat. Rev. Mol. Cell Biol. 23*, 107–124, DOI: 10.1038/s41580-021-00413-2.

(*7*) Shan, S.-o. (2019). Guiding tail-anchored membrane proteins to the endoplasmic reticulum in a chaperone cascade. *J. Biol. Chem. 294*, 16577–16586, DOI: 10.1074/jbc.REV119.006197.

(*8*) Chartron, J. W., Clemons Jr., W. M., and Suloway, C. J. (2012). The complex process of GETting tail-anchored membrane proteins to the ER. *Current Opinion in Structural Biology 22*, 217–224, DOI: 10.1016/j.sbi.2012.03.001.

(*9*) Raven, P. H. Plants make our existence possible, 2021, DOI: 10.1002/ppp3.10173.

(*10*) Duncan, O., van der Merwe, M. J., Daley, D. O., and Whelan, J. (2013). The outer mitochondrial membrane in higher plants. *Trends Plant Sci. 18*, 207–217, DOI: 10.1016/j.tplants.2012.12.004.

(*11*) Xing, S., Mehlhorn, D. G., Wallmeroth, N., Asseck, L. Y., Kar, R., Voss, A., Denninger, P., Schmidt, V. A. F., Schwarzländer, M., Stierhof, Y.-D., et al. (2017). Loss of GET pathway orthologs in Arabidopsis thaliana causes root hair growth defects and affects SNARE abundance. *Proc. Natl. Acad. Sci. U. S. A. 114*, E1544–E1553, DOI: 10.1073/pnas.1619525114.

(*12*) Asseck, L. Y., Mehlhorn, D. G., Monroy, J. R., Ricardi, M. M., Breuninger, H., Wallmeroth, N., Berendzen, K. W., Nowrousian, M., Xing, S., Schwappach, B., et al. (2021). Endoplasmic reticulum membrane receptors of the GET pathway are conserved throughout eukaryotes. *Proc. Natl. Acad. Sci. U. S. A. 118*, e2017636118, DOI: 10.1073/pnas.2017636118.

(*13*) Bodensohn, U. S., Dünschede, B., Kuhlmann, C., Kumari, K., Ladig, R., Grefen, C., Schleiff, E., Fernandez, D., and Schünemann, D. (2025). GET3B is involved in chloroplast biogenesis and interacts with the thylakoidal ALB3 and ALB4 insertases. *Plant Cell Rep. 44*, 108, DOI: 10.1007/s00299-025-03500-2.

(*14*) Anderson, S. A., Satyanarayan, M. B., Wessendorf, R. L., Lu, Y., and Fernandez, D. E. (2021). A homolog of GuidedEntry of Tail-anchored proteins3 functions in membrane-specific protein targeting in chloroplasts of Arabidopsis. *Plant Cell 33*, 2812–2833, DOI: 10.1093/plcell/koab145.

(*15*) Bodensohn, U. S., Simm, S., Fischer, K., Jäschke, M., Groß, L. E., Kramer, K., Ehmann, C., Rensing, S. A., Ladig, R., and Schleiff, E. (2019). The intracellular distribution of the components of the GET system in vascular plants. *Biochim. Biophys. Acta, Mol. Cell Res. 1866*, 1650–1662, DOI: 10.1016/j.bbamcr.2019.06.012.

(*16*) Barlow, A. N., Manu, M. S., Saladi, S. M., Tarr, P. T., Yadav, Y., Thinn, A. M. M., Zhu, Y., Laganowsky, A. D., Clemons Jr., W. M., and Ramasamy, S. (2023). Structures of Get3d reveal a distinct architecture associated with the emergence of photosynthesis. *J. Biol. Chem. 299*, DOI: 10.1016/j.jbc.2023.104752,

(*17*) Katoh, K., Rozewicki, J., and Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in bioinformatics 20*, 1160–1166, DOI: 10.1093/bib/bbx108.

(*18*) Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics 25*, 1189–1191, DOI: 10.1093/bioinformatics/btp033.

(*19*) Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics 23*, 2947–2948, DOI: 10.1093/bioinformatics/btm404.

(*20*) Mehdizadeh Allaf, M., and Peerhossaini, H. (2022). Cyanobacteria: model microorganisms and beyond. *Microorganisms 10*, 696, DOI: 10.3390/microorganisms10040696.

(*21*) Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol. 215*, 403–410, DOI: 10.1016/S0022-2836(05)80360-2.

(*22*) Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques 28*, 1102–1104, DOI: 10.2144/00286ir01.

(*23*) Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol. 30*, 772–780, DOI: 10.1093/molbev/mst010.

(*24*) Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 1–3, DOI: 10.1038/s41586-024-07487-w.

(*25*) Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 2.5.0, available at http://pymol.org/, 2021.

(*26*) Mateja, A., Szlachcic, A., Downing, M. E., Dobosz, M., Mariappan, M., Hegde, R. S., and Keenan, R. J. (2009). The structural basis of tail-anchored membrane protein recognition by Get3. *Nature 461*, 361–366, DOI: 10.1038/nature08319.

(*27*) Andersson, C. R., Tsinoremas, N. F., Shelton, J., Lebedeva, N. V., Yarrow, J., Min, H., and Golden, S. S. In *Methods in Enzymology, Vol. 305*, Ziegler, M. M., and Baldwin, T. O., Eds.; Academic Press: Cambridge, MA, 2000; Chapter 37, pp 527–542, DOI: 10.1016/s0076-6879(00)05511-7.

(*28*) Eaton-Rye, J. J. In *Photosynthesis Research Protocols*, Carpentier, R., Ed.; Humana Press: Totowa, NJ, 2011, pp 295–312, DOI: 10.1007/978-1-60761-925-3_22.

(*29*) Mohamed, A., and Jansson, C. (1989). Influence of light on accumulation of photosynthesis-specific transcripts in the cyanobacterium Synechocystis 6803. *Plant Mol. Biol. 13*, 693–700, DOI: 10.1007/BF00016024.

(*30*) Englund, E., Liang, F., and Lindberg, P. (2016). Evaluation of promoters and ribosome binding sites for biotechnological applications in the unicellular cyanobacterium Synechocystis sp. PCC 6803. *Scientific Reports 6*, 36640, DOI: 10.3389/fmicb.2013.00246.

(*31*) Englund, E., Andersen-Ranberg, J., Miao, R., Hamberger, B., and Lindberg, P. (2015). Metabolic engineering of Synechocystis sp. PCC 6803 for production of the plant diterpenoid manoyl oxide. *ACS Synth. Biol. 4*, 1270–1278, DOI: 10.1021/acssynbio.5b00070.

(*32*) Bhattacharjee, H., Choudhury, R., and Rosen, B. P. (2008). Role of conserved aspartates in the ArsA ATPase. *Biochemistry 47*, 7218–7227, DOI: 10.1021/bi800715h.

(*33*) Zavřel, T., Sinetova, M. A., and Červeny̌, J. (2015). Measurement of chlorophyll a and carotenoids concentration in cyanobacteria. *Bio-Protoc. 5*, e1467–e1467, DOI: 10.21769/BioProtoc.1467.

(*34*) Sherman, B. T., Hao, M., Qiu, J., Jiao, X., Baseler, M. W., Lane, H. C., Imamichi, T., and Chang, W. (2022). DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic acids research 50*, W216–W221, DOI: 10.1093/nar/gkac194.

(*35*) Consortium, U. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research 47*, D506–D515, DOI: 10.1093/nar/gky1049.

(*36*) Suloway, C. J., Chartron, J. W., Zaslaver, M., and Clemons Jr., W. M. (2009). Model for eukaryotic tail-anchored protein binding based on the structure of Get3. *Proceedings of the National Academy of Sciences 106*, 14849–14854, DOI: 10.1073/pnas.0907522106.

(*37*) Saini, D. K., Pabbi, S., and Shukla, P. (2018). Cyanobacterial pigments: Perspectives and biotechnological approaches. *Food Chem. Toxicol. 120*, 616–624, DOI: 10.1016/j.fct.2018.08.002.

(*38*) Maoka, T. (2020). Carotenoids as natural functional pigments. *Journal of natural medicines 74*, 1–16, DOI: 10.1007/s11418-019-01364-x.

(*39*) The UniProt Consortium (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res. 49*, D480–D489, DOI: 10.1093/nar/gkaa1100.

(*40*) Seabold, S., and Perktold, J. (2010). Statsmodels: econometric and statistical modeling with python. *SciPy 7*, 92–96, DOI: 10.25080/majora-92bf1922-011.

3.6 Supporting Information

Table S3.1: Tail-anchored proteins in *Synechocystis* sp. PCC 6803.

Uniprot ID	Protein	Function	Length (aa)	TMD (aa)	Membrane
P72803	DnaJ protein	Chaperone	174	135–158	
P26287	Cytochrome f	Electron transport	328	296–313	Thylakoid
P72717	Cytochrome b_6f complex subunit 8 (PetN)	Electron transport	29	3–23	Thylakoid
P74149	Cytochrome b_6f complex subunit 5 (PetG)	Electron transport	38	5–25	Thylakoid
P74810	Cytochrome b_6f complex subunit 7 (PetM)	Electron transport	36	9–29	Thylakoid
P72632	UDP-N-acetyl-D-manno-saminuronic acid transferase	Glycosylation	251	229–245	
P74200	Penicillin-binding protein 4	Glycosylation	414	383–408	
Q6ZEL7	Phosphomannomutase	Glycosylation	96	74–91	
Q6ZEX1	Haemolysin XhlA	Infection, nutrient uptake	79	54–75	
P73449	Nitrate transport protein NrtD	Nutrient uptake	332	293–315	
P14835	Photosystem II reaction center protein H (6 kDa phosphoprotein)	Photosynthesis	64	29–49	Thylakoid
Q55354	Photosystem II reaction center protein L (PSII-L)	Photosynthesis	39	18–38	Thylakoid
P09191	Cytochrome b559 subunit beta (PSII subunit VI)	Photosynthesis	44	19–35	Thylakoid
P15819	Photosystem II reaction center protein K (PSII-K)	Photosynthesis	45	24–44	Thylakoid
P73070	Photosystem II reaction center protein J (PSII-J)	Photosynthesis	39	7–27	Thylakoid
P74625	PSI-associated linker protein CpcL (CpcG2)	Photosynthesis	249	223–247	Thylakoid
P26286	Protein PsbN	Photosynthesis	43	4–24	Thylakoid
P72575	Photosystem II reaction center X protein	Photosynthesis	39	10–30	Thylakoid
P72701	Photosystem II reaction center protein M (PSII-M)	Photosynthesis	35	7–27	Thylakoid
P73676	Photosystem II protein Y	Photosynthesis	39	5–23	Thylakoid
Q54697	Photosystem II reaction center protein I (Ycf12)	Photosynthesis	38	6–26	Thylakoid
P72986	Photosystem I reaction center subunit XII (PSI-M)	Photosynthesis	31	6–25	Thylakoid
Q55329	Photosystem I reaction center subunit IX	Photosynthesis	40	7–27	Thylakoid
Q55330	Photosystem I reaction center subunit VIII	Photosynthesis	40	12–32	Thylakoid
Q55438	Photosystem II reaction center protein Ycf12	Photosynthesis	39	12–32	Thylakoid
P74787	Photosystem II reaction center protein T (PSII-T)	Photosynthesis	31	3–23	Thylakoid
P73563	High light inducible protein HliC	Photosynthesis	47	15–36	Thylakoid

Continued on next page

Table S3.1: *(Continued)*

Uniprot ID	Protein	Function	Length (aa)	TMD (aa)	Membrane
P38382	Protein translocase subunit SecE	Protein transport	81	50–70	Plasma & Thylakoid
P74508	Protein-export membrane protein SecG	Protein transport	77	55–76	Plasma
P74614	NADH dehydrogenase	Respiration	524	484–504	
P72932	High light inducible protein		57	31–56	Thylakoid
P73183	High light-inducible protein		70	44–67	Thylakoid
P73429	High light inducible protein		70	44–68	Thylakoid
Q55115	UPF0754 thylakoid membrane protein sll0412		419	394–414	Thylakoid
P73517	Thylakoid membrane protein ssr2422		84	61–81	Thylakoid
P72574	Ssr0109 protein		62	40-60	
P72713	Ssl0410 protein		90	66-86	
P72716	Sgl0002 protein		49	12-33	
P72829	Ssl2384 protein		57	35-54	
P72838	Ssr2153 protein		65	20-41	
P73017	Slr1052 protein		170	151-169	
P73064	Ssr3129 protein		61	38-60	
P73430	Slr1544 protein		103	77-96	
P73442	Ssr2611 protein		65	37-57	
P73616	Slr1866 protein		333	313-331	
P73856	Slr1724 protein		264	225-248	
P73857	Ssl3127 protein		75	42-59	Plasma
P73919	Ssl3769 protein		74	59-50	
P74067	Ssl1498 protein		61	35-59	Thylakoid
P74145	Sll1390 protein		249	221-243	Thylakoid
P74486	Sll1862 protein		143	119-140	
P74776	Ssl1552 protein		87	39-62	
P74797	Ssr0332 protein		70	43-67	
Q55186	Slr0496 protein		109	63-86	

Continued on next page

Table S3.1: *(Continued)*

Uniprot ID	Protein	Function	Length (aa)	TMD (aa)	Membrane
Q55359	Sll0178 protein		1319	1273-1296	
Q55726	Sll0602 protein		691	649-669	
Q55934	Slr0793 protein		215	182-203	Outer
Q6YRS7	Ssl6018 protein (Ssl6077 protein)		71	46-67	
Q6ZEC4	Ssl7074 protein		98	72-92	
Q6ZEC5	Slr7073 protein		114	91-111	
Q6ZEI0	Ssr7018 protein		73	37-65	
Q6ZEK9	Ssr5121 protein		71	46-67	
Q6ZEL6	DUF4164 family protein		94	47-66	
Q6ZES9	DUF4164 family protein		130	107-127	

Note: Tail-anchored (TA) proteins were identified by searching UniProt for membrane proteins with single transmembrane domains in *Synechocystis* sp. PCC 6803 (*1*). TA proteins were identified as those whose transmembrane domain was within 30 residues of their C-terminus. UniProt ID, protein name, function (if annotated), length in amino acids (aa), transmembrane domain (TMD) position in amino acids (aa), and membrane localization (if annotated) are reported.

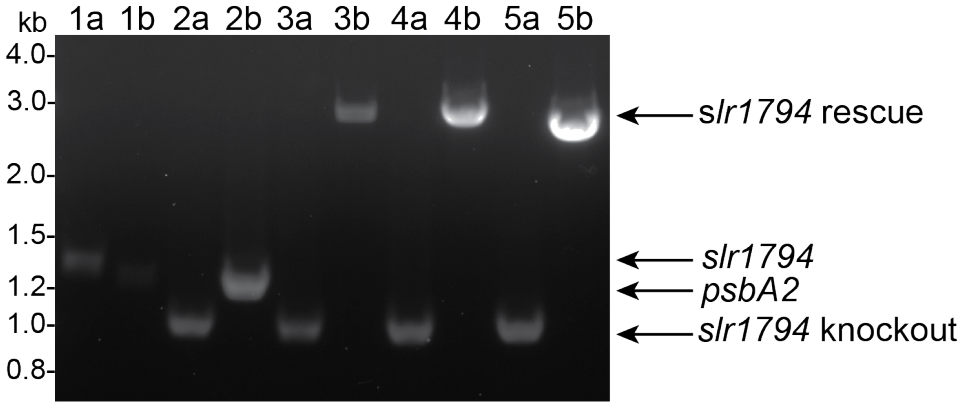


Figure S3.1: Validation of *Synechocystis* sp. PCC 6803 transformation. PCR was performed to amplify the (a) *slr1794* and (b) *psbA2* loci for each strain followed by analysis by agarose gel electrophoresis. Strains as follows: 1, wild-type; 2, *slr1794* knockout; 3, WT-rescue; 4, D39N-rescue; 5, ΔαCD-rescue. Molecular weight marker (kb) is shown to the *left*, and bands are identified on the *right*. Primer sequences per Table S3.5.

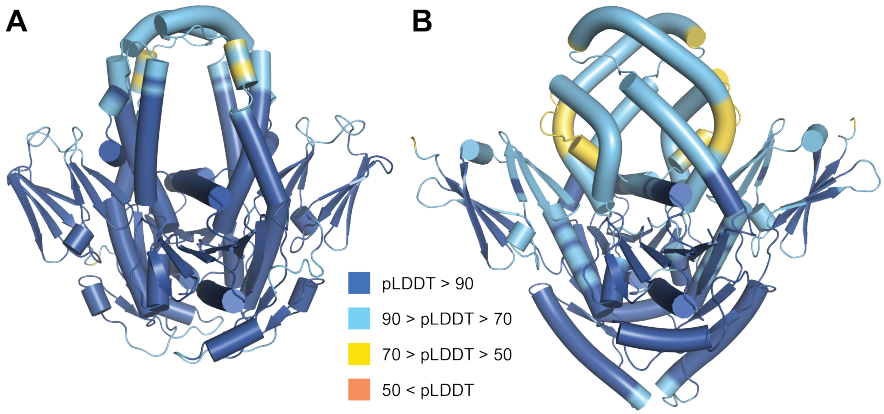


Figure S3.2: Predicted local distance difference test (pLDDT) for structural predictions of Get3d. Structures of (A) the Slr1794 dimer (Fig. 3.2A) and (B) the Sll0086 dimer (Fig. 3.2D) were predicted using AlphaFold3 (*2*) and colored per modified pLDDT coloring scheme (scale shown).

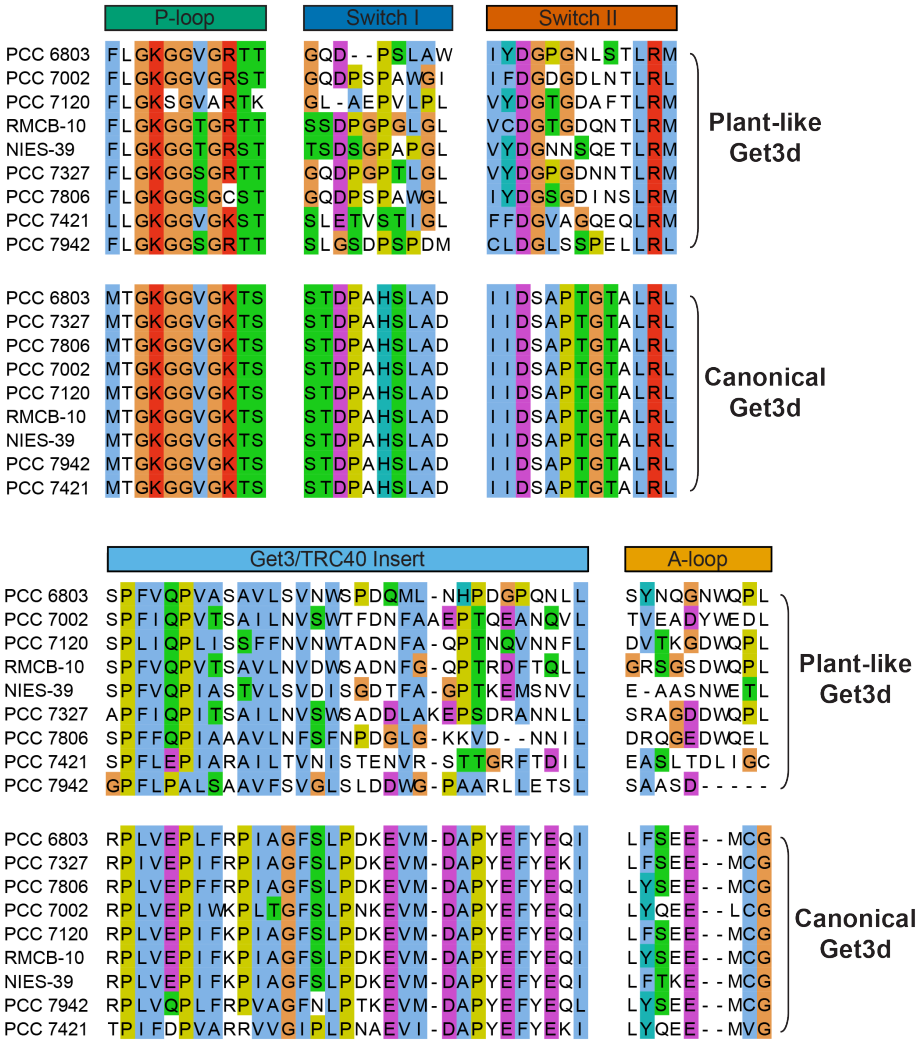


Figure S3.3: Sequence features of cyanobacterial Get3d paralogs. Sequences of plant-like Get3d and canonical Get3d paralogs from selected cyanobacterial species were aligned using multiple alignment using fast Fourier transform (MAFFT) (*3*), visualized in Jalview (*4*), and colored per the ClustalX color scheme (*5*). Sequences are classified as plant-like Get3d or canonical Get3d on the *right*. Characteristic Get3 features are labeled *above* the sequences. Species key: PCC 6803, *Synechocystis* sp. PCC 6803; PCC 7002, *Synechococcus* sp. PCC 7002; PCC 7120, *Nostoc* sp. PCC 7120; RMCB-10, *Oscillatoria princeps*; NIES-39, *Arthrospira platensis*; PCC 7327, *Pleurocapsa* sp. PCC 7327; PCC 7806, *Microcystis aeruginosa* PCC 7806; PCC 7421, *Gloeobacter violaceus* PCC 7421; PCC 7942, *Synechococcus elongatus* PCC 7942.

Table S3.2: Absorbance measurements of pigment extraction.

Sample	$\mathbf{A_{470}}$	$\mathbf{A_{665}}$	$\mathbf{A_{720}}$
WT-1	0.481 55	0.303 76	0.004 94
WT-2	0.437 65	0.309 15	0.003 45
WT-3	0.370 17	0.301 35	0.004 73
KO-1	0.246 61	0.224 81	0.001 52
KO-2	0.339 15	0.222 59	0.002 32
KO-3	0.321 23	0.236 54	0.003 29
WT-rescue-1	0.408 38	0.269 25	0.002 24
WT-rescue-2	0.452 82	0.251 30	0.003 15
WT-rescue-3	0.413 46	0.280 62	0.017 14
D39N-rescue-1	0.364 92	0.255 33	0.004 99
D39N-rescue-2	0.476 17	0.237 58	0.019 72
D39N-rescue-3	0.409 17	0.247 40	0.004 04
$\Delta\alpha$CD-rescue-1	0.321 03	0.231 56	0.007 06
$\Delta\alpha$CD-rescue-2	0.298 95	0.241 14	0.005 45
$\Delta\alpha$CD-rescue-3	0.364 11	0.213 89	0.005 65

Note: Chlorophyll *a* and carotenoids were extracted from *Synechosystis* sp. PCC 6803. Strains as follows: wild-type (WT), *slr1794* knockout (KO), and complementation strains (WT-rescue, D39N-rescue, and $\Delta\alpha$CD-rescue). Absorbance at 470 nm, 665 nm, and 720 nm (in arbitrary units) was measured by spectrophotometry. "-1", "-2", and "-3" correspond to measurements for three independent cultures.

Table S3.3: Tukey's honestly significant difference post hoc test for chlorophyll *a* concentrations.

Group 1	**Group 2**	**Mean Diff**	**p-adj**	**Lower CI**	**Upper CI**	**Significant**
D39N-rescue	WT	0.8784	0.0004	0.4519	1.3049	Yes
KO	WT	1.0394	0.0001	0.6129	1.4659	Yes
WT-rescue	WT	0.5676	0.0094	0.1411	0.9941	Yes
$\Delta\alpha$CD-rescue	WT	1.0782	0.0001	0.6517	1.5047	Yes
D39N-rescue	KO	-0.1610	0.7290	-0.5875	0.2655	No
D39N-rescue	WT-rescue	0.3108	0.1928	-0.1157	0.7373	No
D39N-rescue	$\Delta\alpha$CD-rescue	-0.1998	0.5610	-0.6263	0.2267	No
KO	WT-rescue	0.4718	0.0291	0.0453	0.8983	Yes
KO	$\Delta\alpha$CD-rescue	-0.0388	0.9979	-0.4653	0.3877	No
WT-rescue	$\Delta\alpha$CD-rescue	-0.5106	0.0183	-0.9371	-0.0841	Yes

Note: Tukey's honestly significant difference (HSD) post hoc test to evaluate pairwise differences between concentrations of chlorophyll *a* extracted from *Synechocystis* sp. PCC 6803. Strains as follows: wild-type (WT), *slr1794* knockout (KO), and complementation strains (WT-rescue, D39N-rescue, and $\Delta\alpha$CD-rescue). Mean difference (mean diff), adjusted p-value (p-adj), lower confidence interval (CI), upper CI, and significance conclusion (yes/no) are reported. Differences are considered statistically significant at $p < 0.05$. Raw absorbance data is reported in Table S3.2.

Table S3.4: Tukey's honestly significant difference post hoc test for carotenoid concentrations.

Group 1	Group 2	Mean Diff	p-adj	Lower CI	Upper CI	Significant
D39N-rescue	WT	0.0784	0.9900	-0.4914	0.6481	No
KO	WT	0.5897	0.0418	0.0200	1.1595	Yes
WT-rescue	WT	0.0327	0.9997	-0.5371	0.6024	No
$\Delta\alpha$CD-rescue	WT	0.4839	0.1072	-0.0858	1.0537	No
D39N-rescue	KO	-0.5114	0.0842	-1.0811	0.0584	No
D39N-rescue	WT-rescue	0.0457	0.9987	-0.5241	0.6154	No
D39N-rescue	$\Delta\alpha$CD-rescue	-0.4056	0.2086	-0.9753	0.1642	No
KO	WT-rescue	0.5570	0.0560	-0.0127	1.1268	No
KO	$\Delta\alpha$CD-rescue	0.1058	0.9700	-0.4640	0.6755	No
WT-rescue	$\Delta\alpha$CD-rescue	-0.4513	0.1423	-1.0210	0.1185	No

Note: Tukey's honestly significant difference (HSD) post hoc test to evaluate pairwise differences between concentrations of carotenoids extracted from *Synechosystis* sp. PCC 6803. Strains as follows: wild-type (WT), *slr1794* knockout (KO), and complementation strains (WT-rescue, D39N-rescue, and $\Delta\alpha$CD-rescue). Mean difference (mean diff), adjusted p-value (p-adj), lower confidence interval (CI), upper CI, and significance conclusion (yes/no) are reported. Differences are considered statistically significant at $p < 0.05$. Raw absorbance data is reported in Table S3.2.

Table S3.5: Primer sequences used in this study.

Name	Sequence (5' → 3')
slr1794-fwd	CACAGCCTAGAAAGTTATGCC
slr1794-rev	CACAGCCTAGAAAGTTATGCC
psbA2-fwd	GCTTCGTGTATATTAACTTCCTG
psbA2-rev	GATCGATGGCAATCAAGATCAG

Note: slr1794-fwd and slr1794-rev amplify the *slr1794* locus, and psbA2-fwd and psbA2-rev amplify the *psbA2* locus.

Table S3.6: Whole-cell mass spectrometry protein abundance and metadata of *Synechocystis* sp. PCC 6803. Each column represents a biological replicate from one of five strains: wild-type (WT), *slr1794* knockout (KO), and complementation strains (WT-resc, D39N-resc, $\Delta\alpha$CD-resc). Protein abundances are reported as summed precursor intensities from Proteome Discoverer 3.0. Metadata columns include sequence coverage, peptide counts, identification confidence, and gene annotation. These values are untransformed and unnormalized. Complete data available in attached Excel file: `barlow_chapter3_whole_cell_abundance_supp.xlsx`.

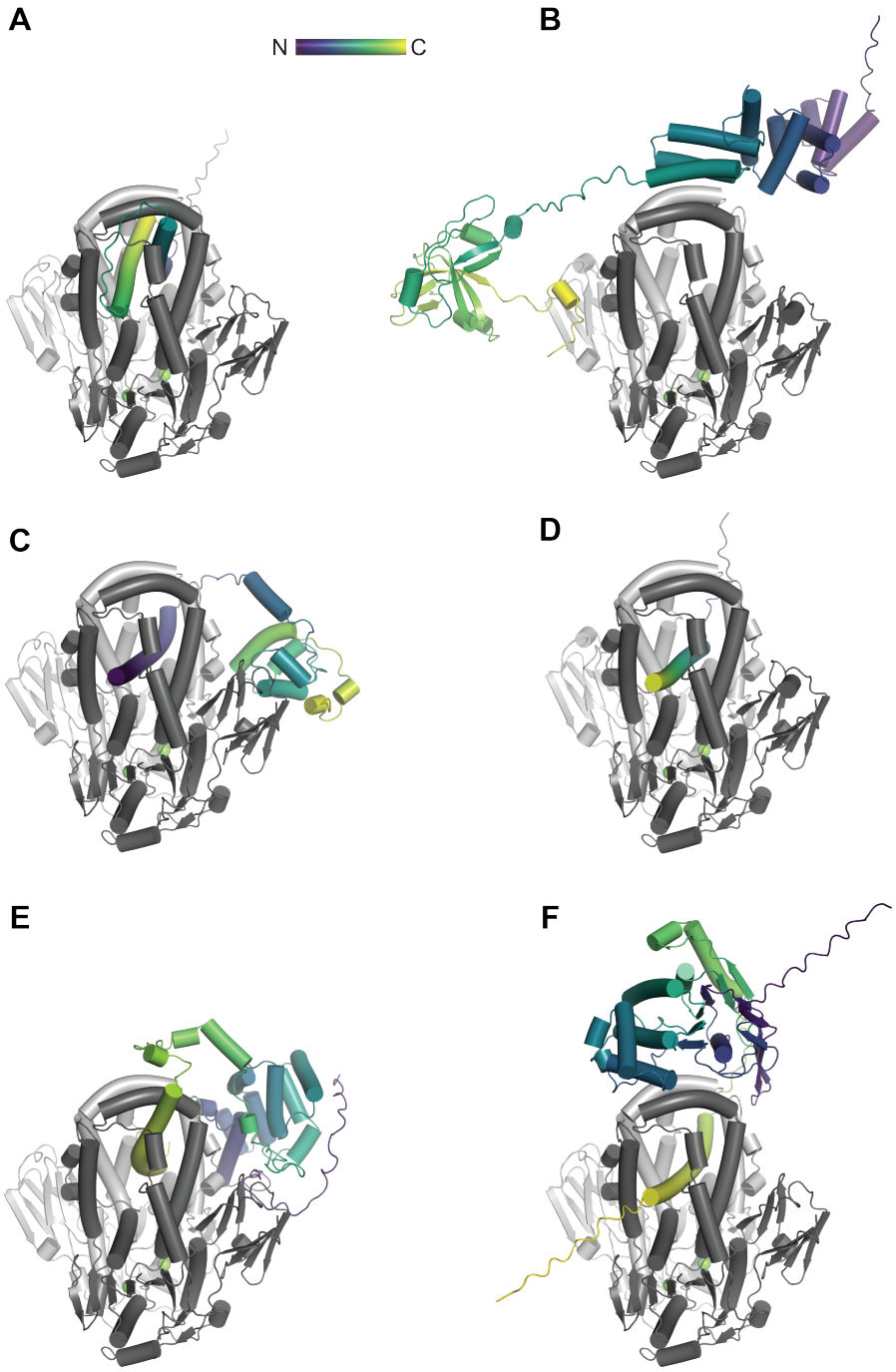


Figure S3.4: Structural predictions of Slr1794 with putative interaction partners. Structures of the Slr1794 dimer with (A) PsaK2, (B) Raf1, (C) PsbL, (D) PsbU, (E) CpcL, and (F) NrtD were predicted using AlphaFold3 (*2*). For each, one Slr1794 monomer is shown in *dark grey* and the other is shown in *light grey*. The putative interaction partner is shown in *viridis* (scale shown). Mg^{2+} ions are shown as *spheres*. Coloring per modified predicted local distance difference test (pLDDT) coloring scheme shown in Fig. S3.5.

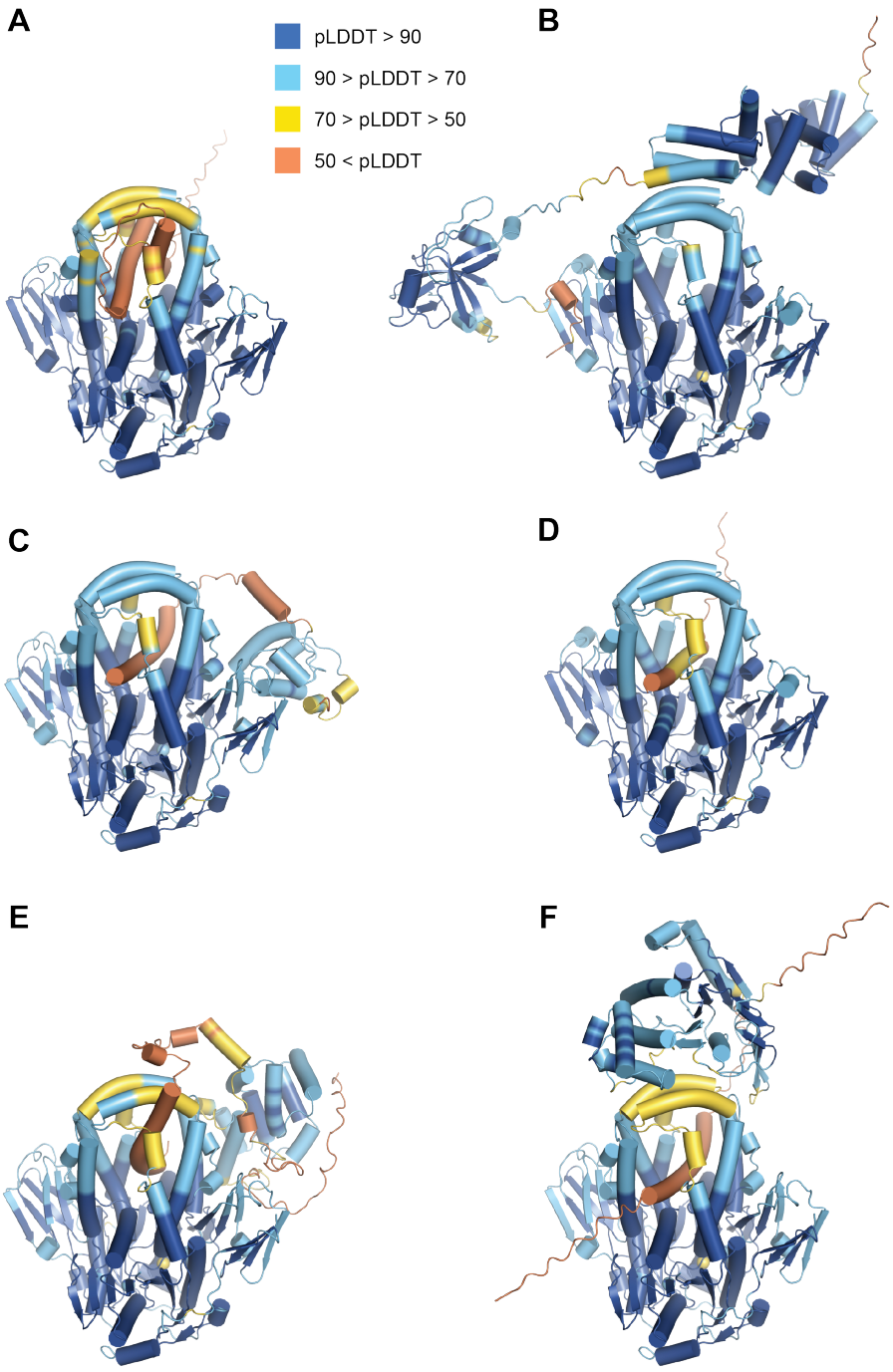


Figure S3.5: Predicted local distance difference test (pLDDT) for structural predictions of Slr1794 with putative interaction partners. Structures of the Slr1794 dimer with (A) PsaK2, (B) Raf1, (C) PsbL, (D) PsbU, (E) CpcL, and (F) NrtD were predicted using AlphaFold3 (*2*) and colored per modified pLDDT coloring scheme (scale shown). Mg^{2+} ions are shown as *spheres*.

Table S3.7: Differential protein abundance comparisons in whole-cell proteomic analysis of *Synechocystis* sp. PCC 6803. Comparisons include knockout (KO) versus wild-type (WT) and complementation strains (WT-resc, D39N-resc, and ΔαCD-resc) versus KO. Reported values include $\log_2$(fold-change) ($\log_2$FC), p-values, adjusted p-values (Benjamini-Hochberg false discovery rate [FDR]), $-\log_{10}$(p-values), and group mean abundances (normalized, $\log_2$ scale). Complete data available in attached Excel file: `barlow_chapter3_whole_cell_enrichment_supp.xlsx`.

Table S3.8: Functional enrichment analysis of annotation clusters derived from proteins significantly up- or down-regulated across strains in the whole-cell proteomic analysis. Output from the Database for Annotation, Visualization, and Integrated Discovery (DAVID) functional annotation clustering. Comparisons include *slr1794* knockout (KO) versus wild-type (WT) *Synechocystis* sp. PCC 6803 and complementation strains (WT-rescue, D39N-rescue, and ΔαCD-rescue) versus KO. Clusters with an enrichment score > 0.1 are reported. Each row represents a significantly enriched category within a given cluster, including GO terms or UniProt keywords, enrichment statistics, and lists of corresponding UniProt IDs. Complete data available in attached Excel file: `barlow_chapter3_whole_cell_DAVID_supp.xlsx`.

Table S3.9: Protein abundance, enrichment, and metadata for the FLAG co-immunoprecipitation coupled to mass spectrometry. Samples include wild-type (WT) *Synechocystis* sp. PCC 6803 and cells expressing Slr1794-3xFLAG (Slr1794-FLAG). Protein abundances are reported as summed precursor intensities from Proteome Discoverer 3.0. Metadata includes peptide counts, Sequest HT identification scores, false discovery rate [FDR] confidence levels, and gene annotations. Enrichment analysis includes $\log_2$ abundance ratios, associated p-values, adjusted p-values, and replicate variability. Complete data available in attached Excel file: `barlow_chapter3_coIP_abundance_enrichment_supp.xlsx`.

Table S3.10: Functional enrichment analysis of annotation clusters derived from proteins significantly enriched in the Slr1794-3xFLAG sample.

Cluster	Enrich. Score	Category	Term	Count	p-value	Fold Enrich.	Adj. p-value	Proteins (UniProt IDs)
1	3.04	GOTERM_MF_DIRECT	GO:0003735 structural constituent of ribosome	11	0.00003	5.14	0.00230	P48949, P73293, P73300, P74267, P73014, P73320, P73315, P73313, P36239, P73306, Q55004
1	3.04	UP_KW_MOLECULAR_FUNCTION	KW-0689 Ribosomal protein	11	0.00007	4.46	0.00095	P48949, P73293, P73300, P74267, P73014, P73320, P73315, P73313, P36239, P73306, Q55004
1	3.04	UP_KW_MOLECULAR_FUNCTION	KW-0687 Ribonucleoprotein	11	0.00008	4.38	0.00095	P48949, P73293, P73300, P74267, P73014, P73320, P73315, P73313, P36239, P73306, Q55004
1	3.04	GOTERM_BP_DIRECT	GO:0006412 translation	10	0.00054	3.97	0.03600	P48949, P73293, P73300, P74267, P73014, P73320, P73315, P73313, P36239, Q55004
1	3.04	GOTERM_CC_DIRECT	GO:0022625 cytosolic large ribosomal subunit	5	0.00680	6.19	0.12900	P74267, P73320, P73315, P36239, P73306
1	3.04	GOTERM_CC_DIRECT	GO:1990904 ribonucleoprotein complex	4	0.01940	6.60	0.18400	P48949, P73300, P73313, Q55004
1	3.04	GOTERM_CC_DIRECT	GO:0005840 ribosome	4	0.04780	4.72	0.30200	P48949, P73300, P73313, Q55004
2	1.64	GOTERM_CC_DIRECT	GO:0022625 cytosolic large ribosomal subunit	5	0.00680	6.19	0.12900	P74267, P73320, P73315, P36239, P73306
2	1.64	GOTERM_MF_DIRECT	GO:0019843 rRNA binding	6	0.01130	4.24	0.47900	Q55637, P73320, P73315, P74368, P73313, P73306
2	1.64	UP_KW_MOLECULAR_FUNCTION	KW-0699 rRNA-binding	6	0.03720	3.09	0.21400	Q55637, P73320, P73315, P74368, P73313, P73306
2	1.64	UP_KW_MOLECULAR_FUNCTION	KW-0694 RNA-binding	7	0.09840	2.12	0.45200	P72914, Q55637, P73320, P73315, P74368, P73313, P73306

Continued on next page

Table S3.10: *(Continued)*

Cluster	Enrich. Score	Category	Term	Count	p-value	Fold Enrich.	Adj. p-value	Proteins (UniProt IDs)
3	0.58	UP_KW_MOLECULAR_FUNCTION	KW-0255 Endonuclease	3	0.32700	2.52	1.00000	P72914, Q55637, P74368
3	0.58	UP_KW_MOLECULAR_FUNCTION	KW-0540 Nuclease	3	0.55100	1.62	1.00000	P72914, Q55637, P74368
4	0.37	GOTERM_CC_DIRECT	GO:0031676 plasma membrane-derived thylakoid membrane	7	0.35600	1.42	1.00000	P73152, P74564, P72752, Q55190, Q55354, Q55332, P74625
4	0.37	UP_KW_CELLULAR_COMPONENT	KW-0793 Thylakoid	7	0.37000	1.39	1.00000	P73152, P74564, P72752, Q55190, Q55354, Q55332, P74625
4	0.37	GOTERM_BP_DIRECT	GO:0015979 photosynthesis	5	0.41100	1.51	1.00000	P74564, Q55875, Q55354, Q55332, P74625
4	0.37	UP_KW_BIOLOGICAL_PROCESS	KW-0602 Photosynthesis	4	0.63000	1.22	1.00000	P74564, Q55875, Q55354, Q55332
5	0.28	UP_KW_CELLULAR_COMPONENT	KW-0472 Membrane	33	0.47100	1.03	1.00000	P74187, P73152, P73351, P74483, P74564, P74300, Q55392, P72863, P72765, P73953, P29273, P73235, Q55190, P72884, P74625, Q55769, P73380, P74554, P74455, P74453, P73765, P73865, P72752, P74555, P73549, P72637, P73449, P72957, P72958, Q55354, Q55332, Q55972, Q55455
5	0.28	UP_SEQ_FEATURE	TRANSMEM:Helical	26	0.47700	1.06	1.00000	P74187, P73351, P74483, P74564, P74300, Q55392, P72863, P72765, P73953, P73235, Q55190, P74625, Q55769, P73380, P74554, P74455, P73865, P72752, P74555, P73549, P72637, P73449, P72957, P72958, Q55354, Q55972

Continued on next page

Table S3.10: *(Continued)*

Cluster	Enrich. Score	Category	Term	Count	p-value	Fold Enrich.	Adj. p-value	Proteins (UniProt IDs)
5	0.28	UP_KW_DOMAIN	KW-0812 Transmembrane	27	0.53900	1.02	1.00000	P74187, P73351, P74483, P74564, P74300, Q55392, P72863, P72765, P73953, P73235, Q55190, P72884, P74625, Q55769, P73380, P74554, P74455, P73865, P72752, P74555, P73549, P72637, P73449, P72957, P72958, Q55354, Q55972
5	0.28	GOTERM_CC_DIRECT	GO:0016020 membrane	18	0.55500	1.04	1.00000	P74187, P74483, P74554, P74453, Q55392, P72863, P72765, P73953, P73865, P73235, P72884, P74555, P73549, P72957, Q55769, P73380, Q55972, Q55455
5	0.28	UP_KW_DOMAIN	KW-1133 Transmembrane helix	22	0.56300	1.02	1.00000	P74187, P74483, P74554, P74564, P72863, P72765, P73953, P73865, P73235, P72752, Q55190, P74555, P73549, P72637, P74625, P73449, P72957, P72958, Q55769, Q55354, P73380, Q55972
6	0.14	INTERPRO	IPR027417:P-loop NTPase	9	0.44300	1.23	1.00000	P53383, P73770, P73765, Q55876, P74555, P72884, P72799, P73449, P74218
6	0.14	GOTERM_MF_DIRECT	GO:0016887 ATP hydrolysis activity	4	0.87000	0.82	1.00000	P53383, Q55876, P72884, P73449
6	0.14	GOTERM_MF_DIRECT	GO:0005524 ATP binding	7	0.98100	0.60	1.00000	P53383, P73770, Q55622, Q55876, Q55503, P72884, P73449

Note: Output from the Database for Annotation, Visualization, and Integrated Discovery (DAVID) functional annotation clustering based on proteins significantly enriched in the Slr1794-3xFLAG sample versus wild-type *Synechocystis* sp. PCC 6803 determined by co-immunoprecipitation coupled to mass spectrometry. Clusters with an enrichment (enrich.) score > 0.1 are reported. Each row represents a significantly enriched category within a given cluster, including GO terms or UniProt keywords, enrichment statistics, and lists of corresponding UniProt IDs.

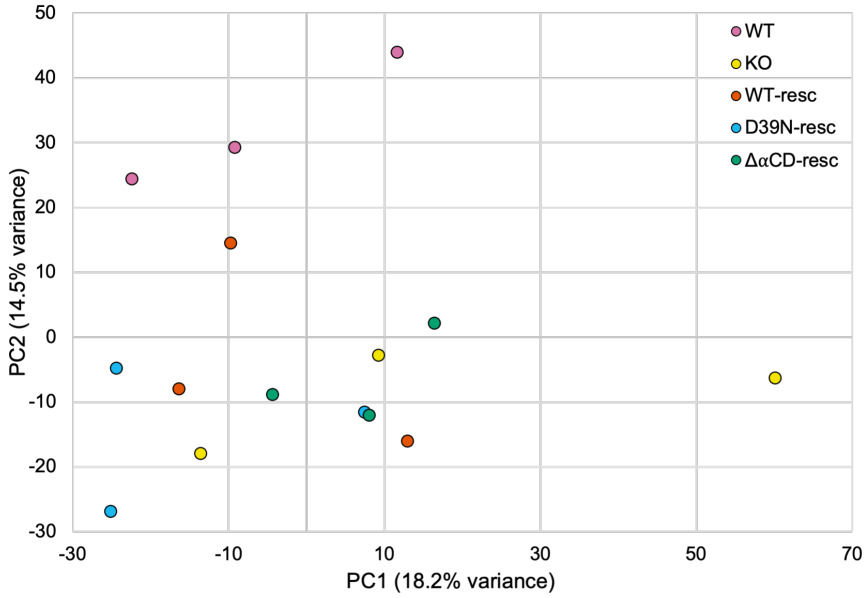


Figure S3.6: Principal component analysis (PCA) of whole-cell proteomes from *Synechocystis* sp. PCC 6803. PCA was performed on normalized, $\log_2$-transformed protein abundance values across 15 samples. Samples cluster by genotype: wild-type (WT, *pink circles*), *slr1794* knockout (KO, *yellow circles*), and complementation strains (WT-resc, *red circles*; D39N-resc, *light blue circles*; ΔαCD-resc, *green circles*). The first two principal components (PC1 and PC2) capture the major variance in the dataset (18.2% and 14.5%, respectively).

Chapter 4

CONCLUSIONS AND PERSPECTIVES ON GET3D FUNCTION IN PHOTOSYNTHETIC ORGANISMS

4.1 Conclusions

This work investigated the structure, biochemical activity, and cellular function of Get3d, a unique homolog of the tail-anchored (TA) protein targeting factor Get3. Through detailed phylogenetic analysis, we found that Get3d paralogs are conserved across plants and photosynthetic bacteria, suggesting an evolutionarily conserved role in photosynthetic organisms. We determined the crystal structure of a plant Get3d and further refined the structure of a cyanobacterial homolog, enabling comparative analysis of their conserved and unique features. *In vitro*, we demonstrated that Get3d is an active ATPase capable of binding TA protein clients, independent of its C-terminal α-crystallin domain.

To explore Get3d function *in vivo*, we characterized the plant-like paralog *slr1794* in the model cyanobacterium *Synechocystis* sp. PCC 6803. Proteomic and phenotypic analyses of deletion and complementation strains revealed that loss of *slr1794* disrupts proteome homeostasis, including up-regulation of transcriptional regulators and down-regulation of redox-associated proteins. Complementation with an ATPase-inactive mutant failed to restore levels of key photosynthesis-related proteins, including components of the photosystems, phycobilisome, and cytochrome complexes, highlighting the importance of Get3d ATPase activity for photosynthetic regulation. In contrast, the α-crystallin domain deletion mutant exhibited strongly reduced Slr1794 protein levels, suggesting that this domain contributes to Get3d's biogenesis or stability.

We observed a significant growth defect in the *slr1794* deletion strain, which was rescued by the ATPase-inactive mutant, indicating that ATP hydrolysis is dispensable for general cell viability. However, pigment content was significantly decreased in deletion strains, supporting a role for Slr1794 in maintaining photosynthetic capacity. Co-immunoprecipitation coupled to mass spectrometry identified putative interaction partners enriched in membrane-associated and photosynthetic proteins. Together, these findings support a functional role for Get3d in redox balance and regulation of photosynthetic machinery.

These findings contribute to a broader understanding of the guided entry of tail-anchored proteins (GET) pathway, particularly in photosynthetic organisms, where this pathway remains poorly characterized. In plants, the GET pathway diverges significantly from the canonical model described in yeast and mammals. Notably, multiple Get3 paralogs have been identified in plants, including organellar variants, a feature unique to this lineage. Get3d is especially intriguing due to its conservation

in the chloroplasts of plants and in photosynthetic bacteria, strongly implicating it in the maintenance of photosynthetic function. As the first biochemical and functional study of this Get3 subfamily, these results provide a significant contribution to the field of TA protein targeting.

Our understanding of TA protein targeting in plant chloroplasts and cyanobacteria remains limited. Currently, no comprehensive models describe TA protein targeting in cyanobacteria, and little *in vivo* work has addressed the targeting machinery in chloroplasts. The absence of canonical GET pathway components in these compartments suggests that novel interaction partners and mechanisms must be at play. By identifying the functional consequences of *slr1794* deletion and putative Slr1794 interaction partners, this study begins to define a distinct targeting system operating in photosynthetic membranes.

Moreover, this work reveals a previously unrecognized role for Get3-like proteins in the regulation of photosynthesis, a function not associated with canonical Get3 homologs. These findings broaden our understanding of TA protein targeting beyond the classical endoplasmic reticulum context and suggest that specialized GET-like systems have evolved to support the unique demands of photosynthetic cells.

4.2 Limitations

There are limitations to this study. A substantial portion of the biochemical characterization of Get3d was performed *in vitro*, which may not fully reflect its behavior in a cellular context. While the ATPase activity and TA protein binding were clearly demonstrated, the physiological relevance of these activities *in vivo* remains to be validated. Similarly, functional insights derived from whole-cell proteomics and phenotypic analysis, though informative, are inherently indirect. The interpretation of proteomic changes relies on correlation, not direct evidence of substrate-client relationships.

Additionally, identifying true client proteins or pathway-specific functions remains challenging due to the complexity and noise inherent to mass spectrometry data. While co-immunoprecipitation coupled to mass spectrometry provides a valuable starting point for identifying interaction partners, distinguishing specific substrates from background binders requires further targeted validation.

4.3 Future Directions

Future studies employing crosslinking approaches, proximity labeling, structural co-complex determination, or genetic interaction screens will be necessary to confirm and refine the mechanistic model proposed here. Building on the current findings, further work should aim to identify bona fide client proteins of Get3d and to define the precise mechanisms of TA protein targeting in cyanobacteria and chloroplasts.

In higher plants, phenotypic characterization of Get3d knockout lines may provide additional insight into its physiological roles. Comparative studies of Get3d interaction partners in chloroplasts, particularly those conserved with cyanobacterial Slr1794, may help elucidate core components of the targeting pathway. Such studies could clarify whether conserved interactors represent functionally relevant substrates or accessory factors.

A significant gap remains in understanding the ATPase cycle of Get3d. The available crystal structures capture Get3d in a closed, apo state, and it is unclear whether Get3d undergoes the same nucleotide-driven conformational transitions as canonical Get3 proteins. Notably, the plant-like Get3d paralogs lack the conserved pivot helix, raising questions about how, or if such conformational dynamics are retained. Additional structural studies, including cryo-electron microscopy or X-ray crystallography of different nucleotide states, would provide insight into the full catalytic cycle.

The role of the C-terminal α-crystallin domain (αCD) is also of particular interest. Although dispensable for TA protein binding *in vitro*, the αCD appears important for maintaining stable protein levels *in vivo*. Biophysical studies such as thermal shift assays and circular dichroism could assess the domain's contribution to protein stability, while chaperone activity assays may test for functional similarity to small heat shock proteins. Domain-swapping experiments and targeted truncations could determine whether the αCD plays a specific regulatory role or serves as a general stabilization element.

Finally, Get3d is only one of several Get3 paralogs present in plants, and its relationship to other chloroplast-localized variants, such as Get3b, remains unresolved. It is currently unclear whether these paralogs are functionally redundant, cooperate within a shared pathway, or fulfill distinct roles in chloroplast biology. Clarifying the interplay among these Get3-like proteins will be critical to fully understanding the diversity of TA protein targeting mechanisms and the specialized adaptations that support redox homeostasis and photosynthetic function in photosynthetic organisms.

By uncovering the structural and functional properties of Get3d, this work not only expands our understanding of TA protein targeting in non-canonical systems, but also highlights the evolutionary innovation required to maintain photosynthetic and redox balance in complex cellular environments. These findings open new avenues for exploring the adaptive versatility of the GET pathway in the context of photosynthetic life.