

**Structural insights into the biosynthesis and biodegradation of  
cyanophycin**

Itai Sharon

Department of Biochemistry

McGill University

Montreal, Quebec, Canada

August 2022

A thesis submitted to McGill University in partial fulfilment of the requirements  
for the degree of Doctor of Philosophy.

## Table of Contents

List of figures .....	6
List of tables .....	8
Abstract .....	9
Résumé.....	11
Acknowledgements.....	13
Preface.....	14
Original contribution of knowledge.....	15
Contribution of authors .....	17
List of abbreviations .....	19
Introduction .....	20
1. Literature review.....	22
1.1. Cyanophycin – General introduction .....	22
1.2. Cyanophycin-producing bacteria .....	23
1.2.1. Uses in cyanobacteria.....	24
1.2.2. Use in non-cyanobacterial species .....	26
1.3. Cyanophycin-scavenging microorganisms.....	27
1.4. Biotechnological production and uses of cyanophycin .....	27
1.4.1. Industrial and biomedical uses.....	27
1.4.2. Biotechnological production of cyanophycin <i>in vivo</i> .....	28
1.4.3. <i>In vivo</i> production of cyanophycin variants.....	29
1.5. Enzymes involved in cyanophycin metabolism.....	30
1.5.1. Cyanophycin synthetase 1 (CphA1).....	30
1.5.2. CphA2 .....	36
1.5.3. Cyanophycinase .....	37
1.5.4. Isoaspartyl dipeptidases.....	39
Bridge to chapter 2.....	41
2. Structures and function of the amino acid polymerase cyanophycin synthetase .....	42
2.1. Abstract .....	43
2.2. Introduction .....	44
2.3. Results .....	46
2.3.1. CphA1 is a common bacterial enzyme.....	46
2.3.2. Architectures of cyanophycin synthetase .....	47
2.3.3. Structure and mutational analysis of the G domain .....	49
2.3.4. Structure and mutational analysis of the M domain.....	51
2.3.5. Structure and polymer binding function of the N domain.....	53
2.3.6. The reaction pathway for cyanophycin synthesis .....	56

2.4. Discussion .....	58
2.5. Methods .....	60
2.5.1. Cloning, protein expression and protein purification .....	60
2.5.2. Crystallography of TmCphA1 .....	62
2.5.3. Cryo-EM sample preparation and data collection.....	63
2.5.4. Cryo-EM data processing.....	63
2.5.5. Model building and refinement into cryo-EM maps.....	64
2.5.6. CphA1 activity assays .....	65
2.5.7. Protein phylogenetic tree generation.....	65
2.5.8. Synthesis of cyanophycin segments.....	65
2.5.9. Differential Scanning Fluorimetry .....	65
2.6. Data availability .....	66
2.7. Acknowledgements .....	66
2.8. Supplementary information .....	67
2.8.1. Supplementary figures .....	67
2.8.2. Supplementary note - Chemical synthesis of cyanophycin segments .....	79
2.8.3. Supplementary tables .....	84
Bridge to chapter 3.....	89
3. A cryptic third active site in cyanophycin synthetase creates primers for polymerization.....	90
3.1. Abstract .....	91
3.2. Introduction .....	92
3.3. Results .....	94
3.3.1. Primer dependence of CphA1 enzymes .....	94
3.3.2. CphA1 primer independence does not depend on the G or M domain activity .....	95
3.3.3. A cryptic N domain active site responsible for primer independence.....	97
3.3.4. The structural basis for the catalytic activity of the N domain .....	97
3.3.5. The N domain cleaves cyanophycin into primers .....	99
3.3.6. Effect of primer independence on heterologous cyanophycin synthesis .....	101
3.4. Discussion .....	103
3.5. Methods .....	106
3.5.1. Cloning, protein expression and purification.....	106
3.5.2. Cryo-EM grid preparation, data collection and processing .....	107
3.5.3. Structure refinement.....	108
3.5.4. CphA1 activity assays .....	108
3.5.5. Cyanophycin purification .....	108
3.5.6. MS analysis of cyanophycin degradation.....	109

3.5.7. SDS-PAGE analysis of cyanophycin degradation .....	109
3.5.8. Synthesis of cyanophycin segments .....	109
3.5.9. Metal analysis .....	110
3.6. Data availability .....	110
3.7. Acknowledgements .....	111
3.8. Supplementary information .....	111
3.8.1. Supplementary tables .....	111
3.8.2. Supplementary figures .....	115
Bridge to chapter 4.....	124
4. Structure and function of the $\beta$ -Asp-Arg polymerase cyanophycin synthetase 2 .....	125
4.1. Abstract .....	126
4.2. Introduction .....	127
4.3. Results and discussion.....	128
4.3.1. Cyanophycin synthesis by CphA2 .....	128
4.3.2. Oligomeric state of CphA2 enzymes .....	132
4.3.3. The crystal structure of CphA2.....	132
4.3.4. Structure and mutation of the N domain .....	134
4.3.5. Structure and mutation of the G domain .....	136
4.3.6. Model of cyanophycin biosynthesis by CphA2 .....	139
4.4. Methods .....	141
4.4.1. Molecular biology .....	141
4.4.2. Crystallography.....	141
4.4.3. Size exclusion chromatography .....	142
4.4.4. CphA2 activity assays .....	142
4.4.5. Synthesis of cyanophycin segments.....	142
4.4.6. Differential scanning fluorimetry .....	142
4.5. Accession codes .....	142
4.6. Acknowledgements .....	143
4.7. Supplementary information .....	144
Bridge to chapter 5.....	150
5. The structure of cyanophycinase in complex with a cyanophycin degradation intermediate..	151
5.1. Abstract .....	152
5.2. Introduction .....	153
5.3. Results and discussion.....	154
5.3.1. Structure of the CphB – cyanophycin covalent co-complex .....	154
5.3.2. Cyanophycin segment binding assays.....	156

5.3.3. Pseudobacteroides cellulosolvens pseudo-CphB .....	159
5.4. Material and methods .....	162
5.4.1. Cloning, protein expression and purification.....	162
5.4.2. CphB <sub>DAP</sub> expression, purification and modification .....	163
5.4.3. Protein crystallization, data collection and structure solution .....	163
5.4.4. Isothermal titration calorimetry (ITC).....	164
5.4.5. Cyanophycinase activity assay .....	164
5.5. Acknowledgements .....	165
5.6. Data availability .....	165
5.7. Supplementary information .....	165
Bridge to chapter 6.....	169
6. Bioinformatics of cyanophycin metabolism genes and characterization of promiscuous isoaspartyl dipeptidases that catalyze the final step of cyanophycin degradation.....	170
6.1. Abstract .....	171
6.2. Introduction .....	172
6.3. Results .....	174
6.3.1. Identification of cyanophycin-metabolizing gene clusters .....	174
6.3.2. IadA and IaaA from cyanophycin clusters are not specific for $\beta$ -Asp-Arg/Lys .....	176
6.4. Discussion .....	179
6.5. Materials and methods .....	181
6.5.1. Bioinformatics .....	181
6.5.2. Cloning, protein expression and purification.....	181
6.5.3. Protein crystallization, data collection, structure solution and refinement .....	182
6.5.4. Enzyme activity assays.....	183
6.6. Acknowledgements .....	183
6.7. Data Availability.....	183
Bridge to chapter 7.....	186
7. Specific cyanophycin dipeptide hydrolase enzymes suggest widespread utility of cyanophycin .....	187
7.1. Abstract .....	188
7.2. Introduction .....	189
7.3. Results .....	191
7.3.1. AbCphZ is a $\beta$ -Asp-Arg/Lys dipeptidase .....	191
7.3.2. <i>P. aeruginosa</i> AotO is a CphZ .....	193
7.3.3. The genetic context of CphZ suggests a role in dipeptide scavenging.....	194
7.3.4. The aot operon enables <i>P. aeruginosa</i> to utilize $\beta$ -Asp-Arg as a nitrogen source ....	196
7.4. Discussion .....	198

7.5. Methods .....	199
7.5.1. Cloning, protein expression and purification.....	199
7.5.2. Protein crystallization, data collection, structure solution and refinement .....	200
7.5.3. Enzyme activity assays.....	201
7.5.4. Isothermal titration calorimetry .....	201
7.5.5. Inductively coupled plasma mass spectrometry (ICP-MS) .....	201
7.5.6. Bioinformatic analysis of gene co-occurrences .....	202
7.5.7. Synthesis and purification of $\beta$ -Asp-Arg.....	202
7.5.8. <i>P. aeruginosa</i> growth assays .....	202
7.6. Acknowledgements .....	202
7.7. Supplementary information .....	203
7.7.1 Supplementary tables .....	203
7.7.2. Supplementary figures .....	204
8. Outlook and general discussion .....	208
8.1 Cyanophycin synthetases .....	208
8.1.1. CphA1 co-complexes with Asp and Arg.....	208
8.1.2. The prospects of bioengineering CphA1 .....	209
8.1.3. CphA1's N domain activity is highly conserved .....	210
8.1.4. CphA2-substrate co-complexes .....	211
8.2. Cyanophycinase .....	212
8.2.1. Is cyanophycinase unique? .....	212
8.2.2. <i>PcCphB</i> as a test case for protein structure prediction.....	213
8.3. Isoaspartyl dipeptidases.....	215
8.3.1. Unknown isoaspartyl dipeptidases likely exist .....	215
8.4. Concluding remarks .....	216
References .....	217

## List of figures

Figure 1.1.....	22
Figure 1.2.....	31
Figure 1.3.....	33
Figure 1.4.....	39
Figure 2.1.....	45
Figure 2.2.....	48
Figure 2.3.....	50
Figure 2.4.....	52
Figure 2.5.....	55
Figure 2.6.....	57
Supplementary Figure 2.1.....	67
Supplementary Figure 2.2.....	68
Supplementary Figure 2.3.....	70
Supplementary Figure 2.4.....	72
Supplementary Figure 2.5.....	74
Supplementary Figure 2.6.....	76
Supplementary Figure 2.7.....	77
Supplementary Figure 2.8.....	78
Supplementary Figure 2.9.....	84
Figure 3.1.....	93
Figure 3.2.....	95
Figure 3.3.....	99
Figure 3.4.....	101
Supplementary Figure 3.1.....	115
Supplementary Figure 3.2.....	116
Supplementary Figure 3.3.....	117
Supplementary Figure 3.4.....	118
Supplementary Figure 3.5.....	120
Supplementary Figure 3.6.....	122
Supplementary Figure 3.7.....	123
Figure 4.1.....	129
Figure 4.2.....	131
Figure 4.3.....	133
Figure 4.4.....	135
Figure 4.5.....	138
Supplementary Figure 4.1.....	144
Supplementary Figure 4.2.....	145
Supplementary Figure 4.3.....	146
Supplementary Figure 4.4.....	146
Supplementary Figure 4.5.....	147
Figure 5.1.....	155

Figure 5.2.....	157
Figure 5.3.....	160
Supplementary Figure 5.1.....	167
Supplementary Figure 5.2.....	168
Figure 6.1.....	173
Figure 6.2.....	177
Figure 6.3.....	178
Supplementary Figure 6.1.....	184
Supplementary Figure 6.2.....	184
Figure 7.1.....	189
Figure 7.2.....	192
Figure 7.3.....	195
Figure 7.4.....	197
Supplementary Figure 7.1.....	204
Supplementary Figure 7.2.....	205
Supplementary Figure 7.3.....	206
Supplementary Figure 7.4.....	207
Figure 8.1.....	209
Figure 8.2.....	214

## List of tables

Table 1.1.....	32
Supplementary Table 2.1.....	85
Supplementary Table 2.2.....	85
Supplementary Table 2.3.....	86
Supplementary Table 2.4.....	87
Supplementary Table 2.5.....	87
Supplementary Table 2.6.....	88
Table 3.1.....	102
Supplementary Table 3.1.....	111
Supplementary Table 3.2.....	112
Supplementary Table 3.3.....	113
Supplementary Table 3.4.....	114
Supplementary Table 4.1.....	149
Supplementary Table 5.1.....	165
Table 6.1.....	174
Table 6.2.....	174
Table 6.3.....	175
Supplementary Table 6.1.....	184
Table 7.1.....	196
Supplementary Table 7.1.....	203
Supplementary Table 7.2.....	204

## Abstract

Cyanophycin is a natural biopolymer consisting of a poly-L-Asp backbone with L-Arg residues attached to their  $\beta$ -carboxylate side chains by isopeptide bonds. First discovered in cyanobacteria in 1886, cyanophycin is produced by a wide range of bacteria and is important for cellular nitrogen storage. In addition to being a ubiquitous natural product, it is also studied due to its potential biotechnological applications. Cyanophycin can be synthesized by two different enzymes: cyanophycin synthetase 1 (CphA1) makes it from Asp and Arg, and cyanophycin synthetase 2 (CphA2) polymerizes  $\beta$ -Asp-Arg dipeptides. The polymer is degraded in two steps: in the first, cyanophycinase breaks it down into  $\beta$ -Asp-Arg dipeptides; in the second, enzymes with isoaspartyl dipeptidase activity degrade these dipeptides into Asp and Arg. Although cyanophycin has been known for over 100 years, many questions about its biosynthesis and biodegradation remained unanswered. This Ph.D. thesis first gives a current summary of the relevant literature about cyanophycin metabolism and its biotechnological applications. This is followed by chapters that present structural, biochemical and bioinformatic studies that show how the enzymes involved in its metabolism function.

The 2<sup>nd</sup>, 3<sup>rd</sup> and 4<sup>th</sup> chapters discuss the biosynthesis of cyanophycin. They present the X-ray crystallography and cryo-EM structures of CphA1 and CphA2, and propose models for these enzymes' activity. Chapter two presents co-complex structures of CphA1 with substrates and substrate analogs. Together with accompanying biochemical experiments, they show how this 3-domain enzyme binds its substrates and catalyzes two ATP-dependent reactions for the polymerization of cyanophycin. Chapter 3 describes the discovery that CphA1 has a third, hydrolytic active site that can cleave long cyanophycin chains into small segments that serve as primers for polymerization. Chapter 4 describes the characterization of nine different CphA2s, and illustrates the range of activity levels and oligomerization displayed by these enzymes. It also presents the crystal structure of a CphA2, highlighting differences and similarities between it and those of CphA1. The structural data, coupled with mutagenesis experiments and activity assays, show the roles of CphA2's domains and their importance for activity and stability.

The 5<sup>th</sup>, 6<sup>th</sup> and 7<sup>th</sup> chapters describe the biodegradation of cyanophycin. Chapter 5 presents the structure of a covalent enzyme-substrate intermediate of cyanophycinase with its substrate cyanophycin. The structure shows how this enzyme is able to bind and cleave cyanophycin which,

despite its peptide-like nature, is resistant to proteolytic degradation. Biochemical experiments and comparison to a structure of an inactive cyanophycinase-like protein help identify regions around the active site which are important for enzymatic activity. Chapter 6 presents a bioinformatic analysis describing the co-occurrence and clustering of genes with isoaspartyl dipeptidase activity with other cyanophycin metabolizing genes. It then describes the structural and biochemical characterization of two such enzymes. The results show that despite being clustered with *cphA1* and cyanophycinase, these enzymes retain broad substrate specificity, similarly to other isoaspartyl dipeptidases. Chapter 7 describes the identification and biochemical and structural characterization of a novel isoaspartyl dipeptidase with specific activity towards dipeptides derived from cyanophycin degradation. *In vivo* data show that this common proteobacterial enzyme allows *Pseudomonas aeruginosa* to utilize  $\beta$ -Asp-Arg as a nitrogen source.

Together, the results presented in this thesis expand our knowledge on important aspects of cyanophycin metabolism. The insights gained from these studies will hopefully promote various areas of cyanophycin research and allow us to better understand the biological and biotechnological processes in which this polymer is involved.

## Résumé

La cyanophycine est un biopolymère naturel constituée de chaînes de poly-L-aspartates, dont le  $\beta$ -carboxylate de sa chaîne latérale est liée à un résidu de L-arginine par une liaison isopeptidique. Découverte dans les cyanobactéries en 1886, la cyanophycine est produite par de nombreuses bactéries et est importante pour le stockage cellulaire de l'azote. En plus d'être un produit naturel omniprésent, elle est aussi étudiée pour son potentiel dans des applications biotechnologiques. La cyanophycine peut être synthétisée par deux enzymes différentes : la cyanophycine synthase 1 (CphA1) la fabrique à partir de l'acide aspartique et de l'arginine, tandis que la cyanophycine synthase 2 (CphA2) polymérise des dipeptides  $\beta$ -Asp-Arg. Le polymère est dégradé en deux étapes : en premier lieu, la cyanophycinase le rompt en dipeptides  $\beta$ -Asp-Arg, puis en deuxième lieu, des enzymes ayant une activité d'isoaspartyl dipeptidase dégradent ces dipeptides en acide aspartique et en arginine. Malgré le fait que la cyanophycine est connue depuis plus d'un siècle, plusieurs questions à propos de sa biosynthèse et de sa biodégradation demeurent sans réponse. Cette thèse de doctorat offre un résumé actuel de la littérature traitant sur le métabolisme de la cyanophycine et ses applications biotechnologiques. Suivront ensuite les chapitres qui couvriront des études structurales, biochimiques et bio-informatiques qui démontrent le fonctionnement des enzymes impliquées dans le métabolisme de la cyanophycine.

Les deuxième, troisième et quatrième chapitres examinent la biosynthèse de la cyanophycine. Ils présentent les structures de cristallographie aux rayons X et de cryo-EM des enzymes CphA1 et CphA2, et proposent des modèles expliquant l'activités de ses enzymes. Le chapitre 2 présente la structure de la CphA1 en complexe avec des substrats et des analogues de substrat. En conjonction avec des expériences biochimique, ses structures démontrent comment cette enzyme à trois domaines se lie à ses substrats et catalyse deux réactions nécessitant de l'ATP pour la polymérisation de la cyanophycine. Le chapitre 3 décrit la découverte d'un troisième site actif hydrolytique dans CphA1 qui clive de longues chaînes de cyanophycine en petits segments qui servent d'amorces pour la polymérisation. Le chapitre 4 décrit la caractérisation de neuf différentes CphA2 et illustre la gamme de niveaux d'activité et l'oligomérisation affichée par ses enzymes. Il présente aussi la structure cristalline de CphA2, soulignant les différences et similitudes entre celle-ci et la CphA1. Les données structurales, en conjonction avec des expériences de mutagenèse et des tests d'activité, démontrent les rôles des domaines de la CphA2 et leur importance pour l'activité et la stabilité.

Les cinquième, sixième et septième chapitres décrivent la biodégradation de la cyanophycine. Le chapitre 5 présente un complexe covalent entre enzyme, la cyanophycinase, et substrat intermédiaire, la cyanophycine. Cette structure montre comment cette enzyme est capable de se lier à et de cliver la cyanophycine qui, malgré sa nature peptidique, est résistante à la dégradation protéolytique. Des expériences biochimiques et une comparaison avec la structure d'une protéine inactive semblable à la cyanophycinase ont aidé à identifier des régions autour du site actif qui sont importantes pour l'activité enzymatique. Le chapitre 6 présente une analyse bio-informatique décrivant la cooccurrence et le regroupement de gènes dotés d'une activité d'isoaspartyl dipeptidase avec d'autres gènes métabolisant la cyanophycine. Il décrit ensuite la caractérisation structurale et biochimique de ses deux types d'enzymes. Les résultats démontrent que malgré le regroupement avec le *cphA1* et la cyanophycinase, ces enzymes conservent une vaste spécificité de substrats, similairement avec d'autres isoaspartyl dipeptidases. Le chapitre 7 décrit l'identification ainsi que la caractérisation biochimique et structurale d'une nouvelle isoaspartyl dipeptidase avec une activité spécifique envers les dipeptides dérivés de la dégradation de la cyanophycine. Des données *in vivo* ont démontré que cette enzyme protéobactérienne commune permet à *P. aeruginosa* d'utiliser la  $\beta$ -Asp-Arg comme source d'azote.

Réunis, les résultats présentés dans cette thèse élargissent nos connaissances sur des aspects importants du métabolisme de la cyanophycine. Les informations tirées de ces études permettront de, espérons-le, promouvoir divers domaines de la recherche sur la cyanophycine et nous permettront de mieux comprendre les processus biologiques et biotechnologiques dans lesquels ce polymère est impliqué.

## Acknowledgements

Firstly, I would like to thank my supervisor, Dr. Martin Schmeing, for taking me onboard the lab and offering cyanophycin metabolism as a Ph.D. project. More importantly, I would like to thank him for his continuous support and invaluable guidance and advice, all the knowledge and skills that he taught me, and for always being unbelievably patient – I'm sure it wasn't easy. I would also like to thank members of the Schmeing lab, past and present, for their help and advice. I am indebted to Dr. S. Asfarul Haque for help with cryo-EM experiments; Dr. Janice Reimer, Dr. Diego Alonzo and Dr. Frederik Hansen for their help with protein purification and crystallography; Dr. Camille Fortinez for countless hours of discussion, some of them about science; and Dr. Michael Tarry for (among many other things) always being the voice of reason.

I would also like to thank members of Bellini 4th floor for their help and advice over the years. My RAC members, Dr. Joaquin Ortega and Dr. Alba Guarné, for many useful comments and their honest and constructive criticism that made me a better scientist. Kim Munroe gave me plenty of patient and thorough advice and help, especially with ITC experiments. I am grateful to our collaborators Dr. Marcel Grogg, Dr. Donald Hilvert and Ms. Sharon Pinus, who by providing us with cyanophycin segments made most of my research possible. I thank Ms. Anne Sophie Émilie Marie Labarre for being kind enough to translate the abstract to le français. I would also like to thank the staff at FEMR, and especially Dr. Kaustuv Basu, for training and help with cryo-EM data collection; and staff at the Canadian Light Source, Advanced Photon Source and Advanced Light Source synchrotron facilities for their help with crystallographic data collection.

## Preface

This is a manuscript-based thesis which includes four published articles, one submitted manuscript and two manuscripts currently in preparation:

The literature review (chapter 1) and general discussion (chapter 8) are partially adapted from: Sharon I, Schmeing TM. Cyanophycin and its biosynthesis. *Review manuscript in preparation*.

Chapter 2 is published in: Sharon I, Haque AH, Grogg M, Lahiri I, Seebach D, Leschziner AE, Hilvert D, Schmeing TM. Structures and function of the amino acid polymerase cyanophycin synthetase. *Nature Chemical Biology* 2021 Oct; 17, 1101–1110.

Chapter 3 is published in: Sharon I, Pinus S, Grogg M, Moitessier N, Hilvert D, Schmeing TM. A cryptic third active site in cyanophycin synthetase creates primers for polymerization. *Nature Communications* 2022 Jul 7; 13:3923.

Chapter 4 is published in: Sharon I, Grogg M, Hilvert D, Schmeing TM. Structure and function of the  $\beta$ -Asp-Arg polymerase cyanophycin synthetase 2. *ACS Chemical Biology* 2022, 17, 3, 680-700.

Chapter 5 is published in: Sharon I, Grogg M, Hilvert D, Schmeing TM. The structure of cyanophycinase in complex with a cyanophycin degradation intermediate. *Biochimica et Biophysica Acta - General Subjects* 2022 November 1866(11), 130217. Online 26 July 2022.

Chapter 6 is submitted in: Sharon I, Schmeing TM. Bioinformatics of cyanophycin metabolism genes and characterization of promiscuous isoaspartyl dipeptidases that catalyze the final step of cyanophycin degradation. *Submitted to PLOS ONE*.

Chapter 7 is in preparation in: Sharon I, McKay G, Nguyen D, Schmeing TM. Specific cyanophycin dipeptide hydrolase enzymes suggest widespread utility of cyanophycin.

## Original contribution of knowledge

### Chapter 2 – “Structures and function of the amino acid polymerase cyanophycin synthetase”

- Description of *cphA1* as a widespread gene outside of the phylum *Cyanobacteria*.
- The first structures of CphA1, including structures with various substrates/substrate analogs.
- Biochemical and structural experiments show the roles of each of CphA1’s three domains, and suggest a model for the enzyme’s activity.

### Chapter 3 – “A cryptic third active site in cyanophycin synthetase creates primers for polymerization”

- Determination of the minimal length of cyanophycin that can serve as a primer for polymerization.
- Identification of a hydrolytic active site in CphA1 that allows the enzyme to degrade long cyanophycin chains into short segments that serve as primers.
- Demonstration that primer availability can be a limiting factor for cyanophycin production in heterologous hosts.

### Chapter 4 – “Structure and function of the $\beta$ -Asp-Arg polymerase cyanophycin synthetase 2”

- Characterization of nine CphA2 enzymes shows a range of activity profiles and oligomerization.
- The first structure of a CphA2.
- Identification of residues important for CphA2 activity and stability.

### Chapter 5 – “Structure of cyanophycinase in complex with a cyanophycin degradation intermediate”

- Structure of a covalent cyanophycinase-cyanophycin complex shows how the enzyme binds its substrate.
- Structure of an inactive cyanophycinase-like protein highlights conformational differences that are important for the enzyme’s activity.

Chapter 6 – “Bioinformatics of cyanophycin metabolism genes and characterization of promiscuous isoaspartyl dipeptidases that catalyze the final step of cyanophycin degradation”

- Bioinformatic analysis of the co-occurrence and clustering of *cphAI*, cyanophycinase and isoaspartyl dipeptidase genes.
- Structural and biochemical studies of two isoaspartyl dipeptidases from cyanophycin gene clusters show they retain substrate promiscuity and are not specific for cyanophycin dipeptides.

Chapter 7 – “Specific cyanophycin dipeptide hydrolase enzymes suggest widespread utility of cyanophycin”

- Identification of the only known enzyme that specifically hydrolyzes  $\beta$ -Asp-Arg/Lys.
- Structural, biochemical and bioinformatic studies show how this common proteobacterial enzyme binds and hydrolyzes its substrates.
- *In vivo* studies show how this enzyme and the operon it is part of allow *Pseudomonas aeruginosa* to utilize  $\beta$ -Asp-Arg as a nitrogen and carbon source.

## Contribution of authors

### Chapter 2

M. Grogg synthesized the cyanophycin segments and analogs under the direction of D. Seebach and D. Hilvert. Together with T. Martin Schmeing and S. A. Haque, I performed sample preparation, data collection and processing of *AbCphA1*. I performed all other biochemical, structural and bioinformatic studies. T. Martin Schmeing and I designed the experiments and wrote the manuscript together with input from A. Leschziner, I. Lahiri and D. Hilvert.

### Chapter 3

M. Grogg and S. Pinus synthesized the cyanophycin segments and analogs. I performed all other biochemical and structural studies. T. Martin Schmeing and I designed the experiments and wrote the manuscript together with input from D. Hilvert.

### Chapter 4

M. Grogg synthesized the cyanophycin segments and analogs. I performed all other biochemical and structural studies. T. Martin Schmeing and I designed the experiments and wrote the manuscript together with input from D. Hilvert.

### Chapter 5

M. Grogg synthesized the cyanophycin segments and analogs. I performed all other biochemical and structural studies. T. Martin Schmeing and I designed the experiments and wrote the manuscript together with input from D. Hilvert.

### Chapter 6

I performed all the biochemical, bioinformatic and structural studies. T. Martin Schmeing and I designed the experiments and wrote the manuscript.

## Chapter 7

G. McKay performed the *P. aeruginosa* growth assays and WGS analysis. I performed all other biochemical, bioinformatic and structural studies. G. McKay, D. Nguyen, T. Martin Schmeing and I designed the experiments. T. Martin Schmeing and I wrote the current version of the manuscript.

## List of abbreviations

ADP – Adenosine diphosphate  
ADPCP – Adenosine-5'-[( $\beta$ , $\gamma$ )-methylene]triphosphate  
AEBSF – 4-(2-aminoethyl)benzenesulfonyl fluoride  
AST – Arginine succinyltransferase pathway  
ATP – Adenosine triphosphate  
CDM – cell dry mass  
Cryo-EM – cryo-electron microscopy  
DAP – diaminopropionic acid  
ICP-MS – Inductively-coupled plasma mass spectrometry  
kDa – kilo Dalton  
LC-MS – Liquid chromatography – mass spectrometry  
PMSF – Phenylmethanesulfonyl fluoride  
SEC – size exclusion chromatography  
UDPMurNAc – uridine diphosphate-N-acetylmuramic acid

## Introduction

Understanding the processes behind the biosynthesis and biodegradation of cyanophycin is important for both biological and biotechnological research. From a biological perspective, it is important to understand cyanophycin metabolism since it is very common in nature. The ability to produce cyanophycin is widespread throughout the bacterial kingdom, and confers significant advantages in fitness. Moreover, many microorganisms from various environments that cannot produce this polymer are able to degrade it, implying that cyanophycin is common enough to make scavenging it worthwhile. From a biotechnological perspective, the industrial and pharmacological uses of cyanophycin are an active area of research. Most of this research relies on polymer that is produced and degraded enzymatically, normally *in vivo*.

For these reasons, it is important to understand the mechanisms underlying cyanophycin metabolism. Specifically, a better understanding of the enzymes that are involved in its synthesis and degradation is crucial for cyanophycin research. For example, multiple studies attempted to bioengineer cyanophycin synthetase 1 (CphA1) for optimized production of the polymer *in vivo*. However, without a good understanding of its structure and mechanism of activity it is impossible to fully understand how the enzyme works, which residues and parts of it are important for activity and how and why changes in them affect cyanophycin production. Moreover, this lack of knowledge makes it difficult to explain observed phenotypes of CphA1 mutations and build upon previously published results.

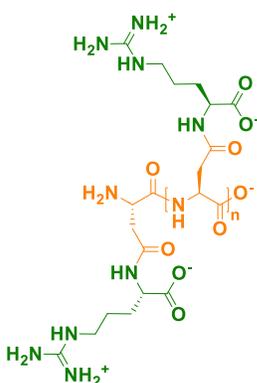
The objectives of this Ph.D. project were to elucidate the structures and mechanisms of activity of the enzymes that are involved in cyanophycin metabolism. These include: CphA1, the enzyme that makes cyanophycin from Asp and Arg; CphA2, the enzyme that polymerizes  $\beta$ -Asp-Arg dipeptides into cyanophycin; cyanophycinase, the enzyme that degrades the polymer to  $\beta$ -Asp-Arg; and isoaspartyl dipeptidases, enzymes capable of hydrolyzing these dipeptides into free amino acids. The cryo-EM and X-ray crystallography structures of these enzymes show how their active sites are arranged, how they bind their substrates and how they perform catalysis. Accompanying biochemical experiments support the structural observations and allow us to suggest models for the enzymes' activity. In addition, bioinformatic studies expand the existing knowledge about their ubiquity and biological context.

Together, these results will allow us to better understand important aspects of cyanophycin metabolism. For example: a better understanding of CphA1's active sites and conserved sequences will allow better predictions of CphA1 homologs and their activity; understanding the roles of CphA2's domains will enable better-informed bioengineering attempts; and a detailed model of the binding of cyanophycin by cyanophycinase explains how, unlike proteases, it is able to degrade this polymer. The insights that can be gained from the studies that are included in this thesis will hopefully promote various areas of cyanophycin research and allow us to better understand the biological and biotechnological processes this polymer is involved in.

## 1. Literature review

### 1.1. Cyanophycin – General introduction

About 140 years ago, Italian botanist Antonio Borzi looked at cyanobacterial cells under a microscope and noted they contained large, light refracting granules<sup>14</sup>. As the material forming these granules was still unknown it was named cyanophycin, after the cyanobacteria in which it was discovered. It would take almost 100 years until subsequent studies by Simon revealed that these granules consist of poly-aspartic acid chains with arginines attached to each of their sidechains<sup>15,16</sup> (Fig. 1.1). Following the understanding of cyanophycin's nature, it was discovered that its production in the cyanobacterium *Anabaena cylindrica* is increased, rather than being inhibited, by the addition of chloramphenicol<sup>17</sup>. As this antibiotic targets the bacterial ribosome and inhibits protein synthesis, it was concluded that cyanophycin synthesis occurs through a non-ribosomal pathway.



**Figure 1.1. The general structure of cyanophycin.** The polymer has a poly-L-Asp backbone (orange), with Arg (green) attached to each Asp side chain. Cyanophycin may occasionally contain small amounts of Lys in addition to Arg.  $n \approx 80-400$ .

Amino acid polymers of simple composition are quite rare in nature.  $\epsilon$ -Poly-lysine, and the related polymers  $\delta$ -poly-diaminobutanoic acid<sup>18</sup> and  $\gamma$ -poly-diaminopropionic acid<sup>19</sup> are made by some strains of *Streptomyces*, with  $\epsilon$ -poly-lysine finding wide use in Asia as a food preservative<sup>20</sup>. Poly-glutamate is an edible, water soluble polymer that is produced in *Bacillus* and has multiple industrial applications<sup>21</sup>. Poly-glutamate is also synthesized in mammals, not as a free molecule but as a post-translational modification on brain tubulin<sup>22</sup>. However, these are all very different in

nature to cyanophycin, being homopolymers with much shorter length (e.g. ~30 residues is typical for  $\epsilon$ -poly-lysine<sup>23</sup>), making cyanophycin a truly unique molecule.

Cyanophycin's chemical structure gives it unique properties. Despite the polymer's peptidic nature, it is resistant to proteolytic degradation by a variety of proteases<sup>16</sup>. Its high nitrogen content, 24% by mass, is higher than that of other biopolymers like proteins (~13-19%) and nucleic acids (~16%)<sup>24</sup>. Cyanophycin can sometimes contain small amounts of Lys instead of Arg, and has interesting solubility properties which depend on its exact composition. While soluble in acidic or basic solutions<sup>16,25</sup>, under physiological pH cyanophycin composed of  $\beta$ -Asp-Arg dipeptides is very insoluble<sup>16</sup>. This causes it to spontaneously precipitate and form inclusion bodies in cells<sup>26</sup>, preventing it from affecting their osmotic pressure or interfering with cellular processes. Due to these properties, bacteria often use cyanophycin as a nitrogen storage material, and its industrial-scale production and uses are an active area of research.

## **1.2. Cyanophycin-producing bacteria**

The importance of cyanophycin for the bacteria that produce it has been mostly studied in cyanobacteria, as for decades it was only been known to exist in this phylum. Initial studies characterized conditions that lead to increased accumulation of cyanophycin in cell cultures. In 1972 Ingram et al. found that following exposure to chloramphenicol and subsequent growth inhibition, *Agmenellum quadruplicatum* BG-1 cells developed granular structures of an unknown nature<sup>27</sup>. A later study by Simon<sup>17</sup>, conducted with *Anabaena cylindrica* cells, found that these granules likely consisted of cyanophycin. He showed that cells exposed to chloramphenicol at concentrations that inhibit protein synthesis stopped growing and accumulated large amounts of this polymer. Following the antibiotic's removal and resumption of protein synthesis, the polymer was degraded. Moreover, he demonstrated that cyanophycin synthesis was an energy consuming process, and correctly assumed that following its degradation, nitrogen from the polymer was used for protein synthesis<sup>28</sup>.

Following the realization that cyanophycin accumulation in cyanobacteria can be modulated by external conditions, more detailed studies were conducted in order to better understand the processes controlling it. It was discovered that in *Aphanocapsa* sp. PCC6308 cells, cyanophycin accumulation depends on the availability of sufficient carbon and fixed nitrogen (in the form of nitrate or arginine). Interestingly, several different sub-optimal growth conditions

actually led to increased accumulation of cyanophycin – low levels of light, phosphorus and sulfur all led to lower cell growth, but resulted in more relative cyanophycin accumulation (measured as % of dry weight)<sup>29</sup>. The observation that cyanobacteria may accumulate cyanophycin not during periods of plenty, but when they sense that the steady supply of nutrients is compromised, was later repeated in *Agmenellum*<sup>30</sup>, *Synechocystis* sp. PCC6308<sup>31</sup> and *Anabaena cylindrica*<sup>31</sup>. Nutrient availability is not the only stress inducing condition that can lead to increased accumulation of cyanophycin: in *Scytonema* this can also happen as a result of high salinity<sup>32</sup>, in *Aphanocapsa* PCC6308 as a result of lowered growth temperature<sup>29</sup>, and in *Fremyella diplosiphon* as a result of exposure to various antibiotics<sup>33</sup>.

### **1.2.1. Uses in cyanobacteria**

As the knowledge about the nature of cyanophycin and its accumulation in cells increased, it was realized that it has an important role in fixed nitrogen storage. To date, three cyanobacterial systems that make use of this polymer are known. These systems use the fact that cyanophycin can store fixed nitrogen as inert, stable and insoluble granules to solve challenges faced by different cyanobacterial species.

#### *1.2.1.1. Dynamic nitrogen storage*

Shortly following its characterization, cyanophycin was suggested to be a dynamic nitrogen store for nitrogen fixing cyanobacteria<sup>34</sup>, which would be very useful to these strains. Most cyanobacterial species are capable of performing photosynthesis, using water as an electron donor<sup>35</sup>. This process leads to the formation of O<sub>2</sub>, which can either be used in cellular respiration or released into the environment. Many cyanobacteria are also diazotrophic, meaning they are capable of fixing atmospheric N<sub>2</sub><sup>35</sup>. Diazotrophic bacteria are less dependent on the availability of fixed nitrogen in the environment, and so have a clear fitness advantage under nitrogen-limited conditions. However, the key enzyme required for N<sub>2</sub> fixation, nitrogenase, contains an iron-sulfur cluster which is oxidized in the presence of O<sub>2</sub>, leading to the irreversible inactivation of the enzyme<sup>36</sup>. Thus, nitrogen fixation is incompatible with photosynthesis, and must be separated from it either temporally or spatially<sup>37</sup>.

Some cyanobacteria achieve spatial separation of nitrogen fixation and photosynthesis in a remarkable way. These strains differentiate into specialized cell types: vegetative cells that perform photosynthesis and maintain high levels of cytosolic oxygen; and heterocysts that have low levels of cytosolic oxygen<sup>35</sup>, and so can perform nitrogen fixation during the day. Nitrogen

fixed in heterocysts can be used to make cyanophycin, which accumulates in their poles, close to adjacent vegetative cells. This fixed nitrogen is eventually transferred into vegetative cells<sup>38,39</sup>. Interestingly, it was found that in *Anabaena* sp. PCC7120 this nitrogen is transferred in the form of  $\beta$ -Asp-Arg dipeptides, the product of cyanophycin degradation by cyanophycinase<sup>40</sup>. Cyanophycin-producing heterocysts degrade it into dipeptides, which are shuttled to vegetative cells. The vegetative cells express high levels of isoaspartyl dipeptidase, the enzyme that degrades dipeptides into free Asp and Arg, allowing rapid funneling of cyanophycin-derived material into other metabolic processes<sup>41</sup>.

Cyanobacteria that cannot form specialized cell-types often use temporal means to separate nitrogen fixation and photosynthesis<sup>37</sup>. A 2001 study examined cyanophycin accumulation patterns in the diazotrophic, unicellular cyanobacteria *Cyanothece* sp. ATCC51142, which separate nitrogen fixation and photosynthesis by dividing them between day and night cycles<sup>42</sup>. The authors grew cells in a 12 h light / 12 h dark regime under nitrogen fixing conditions, and noted that nitrogenase activity was elevated during dark periods, when no photosynthesis occurred. They noticed that cyanophycin accumulation followed the same trend – the polymer was synthesized during dark periods and degraded in the light. A similar pattern was also observed in cells *Trichodesmium*<sup>43</sup>, but not in the non-diazotrophic strain *Synechocystis* 6803<sup>42</sup> nor in the heterocyst-forming *Gloeothece* and *Anabaena Cylindrica*<sup>44</sup>. This observed pattern is consistent with the assumption that cyanophycin serves as a dynamic reservoir of fixed nitrogen during periods of nitrogenase activity. The low solubility and reactivity of the polymer make it much better suited for this role than merely increasing cellular concentrations of  $\text{NH}_4$  or Arg, which are hard to keep localized and are more likely to interfere with cellular processes when present in high concentrations.

#### 1.2.1.2. Feast/famine nitrogen reserve

Cyanophycin can also be used to store nitrogen over long periods, as an adaptation to seasonal variations in nitrogen availability. Many species of cyanobacteria have the ability to form harmful algal blooms, a condition in which they multiply in vast quantities and dominate the phytoplanktonic community<sup>45,46</sup>. These blooms are often accompanied by the release of toxins<sup>47</sup>, leading to extensive ecological and economical damage as well as health risks to humans<sup>48</sup>. Nitrogen availability is a major factor in cyanobacteria's ability to form harmful blooms<sup>46</sup>. The bloom forming cyanobacterium *Planktothrix agardhii* changes the expression levels of

cyanophycin-metabolizing genes in response to seasonal variations in nitrogen availability<sup>49</sup>. Genes associated with cyanophycin production are expressed during periods of high nitrogen availability, while genes associated with its degradation are expressed when nitrogen levels decrease and competition for available nitrogen increases. A similar trend was observed with blooms formed by *Raphidiopsis raciborskii*, which accumulated cyanophycin during periods of fluctuation in nitrogen availability, and degraded it during periods of low nitrogen availability<sup>50</sup>.

#### 1.2.1.3. Storage in akinetes

Another way cyanobacteria can use cyanophycin is as a nutrient source for akinetes – dormant cells formed by some heterocyst-forming cyanobacterial species<sup>51</sup>. Akinetes are similar in function to spores formed by other bacteria: they are durable and capable of withstanding harsher conditions than normal cells, and are often produced when environmental conditions are unfavorable. Due to their thick wall and slow metabolism they can survive periods of elevated temperature, high salinity or low nutrient availability, and germinate once they detect suitable conditions<sup>51</sup>. Cells of *Aphanizomenon ovalisporum* accumulate cyanophycin under akinete inducing conditions. The polymer is then stored in cells that differentiate into akinetes<sup>52</sup>, a phenomenon also observed in cells of *Anabaena variabilis* ATCC29413<sup>53</sup>. This pool of cyanophycin, thanks to its stability and high nitrogen content, is a good solution for long-term storage of nutrients which can feed the akinetes once they germinate.

#### 1.2.2. Use in non-cyanobacterial species

To date, only one study has examined the possible role of cyanophycin production in non-cyanobacterial species. This study, conducted with *Clostridium perfringens* SM101, found cyanophycin to have a role in spore formation. The authors analyzed membrane associated proteins of germinated *C. perfringens* spores, and noted that they contained cyanophycinase, an enzyme required for cyanophycin degradation. Moreover, mutants deficient in cyanophycin production produced fewer and smaller spores. These results suggest that cyanophycin is involved in spore assembly<sup>54</sup>, although its exact role in this context is not yet understood. As cyanobacteria constitute only a small fraction of the bacterial species that are capable of producing cyanophycin, its role in *C. perfringens* is likely just one of several unknown functions.

### **1.3. Cyanophycin-scavenging microorganisms**

Cyanophycin is likely a common material in many environments due to reasons concerning both its synthesis and degradation. Cyanophycin can be produced by a wide range of bacteria which can inhabit a variety of environments. For example, cyanobacteria are found in both marine and fresh-water habitats; and *Acinetobacter baylyi* DSM587, a cyanophycin producing proteobacterium, was isolated from soil sample<sup>55</sup>. In addition to being widely produced, cyanophycin is also remarkably stable. Like proteins, it is composed of amide bonds. These bonds are very stable, with an estimated half-life of hundreds of years<sup>56</sup>. As a result, cyanophycin is unlikely to decompose spontaneously under normal environmental conditions and must be degraded enzymatically. However, because cyanophycin is resistant to proteases<sup>16</sup>, only organisms that express cyanophycinases can degrade it. This further decreases the likelihood that it will be degraded by non-specific pathways.

It is thus not surprising that cyanophycin-degrading bacteria have been found in samples taken from a wide variety of environments. In several different studies, Steinbüchel et al. screened samples from different environments for cyanophycin-degrading bacteria. These environments included forest soil<sup>57</sup>, aerobic<sup>58</sup> and anaerobic<sup>59</sup> pond sediments, and the gut flora of many different animals<sup>60</sup>. Remarkably, bacterial strains or consortia capable of utilizing cyanophycin both as a nitrogen and carbon source were isolated from all of these environments<sup>61</sup>. These isolates included bacteria that likely cannot produce cyanophycin themselves<sup>59</sup>, and were often found to express an extracellular version of cyanophycinase (CphE) which allowed them to degrade cyanophycin found outside of the cells. This suggests that in many places, cyanophycin is present in sufficient quantities to make scavenging for it worthwhile, and bacteria with specialized cyanophycin-scavenging operons have indeed been isolated<sup>62</sup>. Furthermore, a BLAST<sup>63</sup> search shows that versions of CphE can even be found in some fungi, which are not known to produce cyanophycin themselves but often specialize in degrading and utilizing dead matter.

### **1.4. Biotechnological production and uses of cyanophycin**

#### **1.4.1. Industrial and biomedical uses**

Cyanophycin has promising industrial and pharmacological uses, both in itself and as a source for other materials. Tseng et al. examined the ability of polyethylene glycol-conjugated cyanophycin to form self-assembling nanovesicles that can reversibly encapsulate small molecules

in a temperature and pH-dependent manner<sup>64</sup>. These could have a possible use in drug delivery, as Grogg et al. found that intravenous injection of cyanophycin to mice resulted in no observable symptoms<sup>65</sup>. Another potential use for cyanophycin in biomedicine is in wound healing, as it was found that layers of cyanophycin and hyaluronic acid or  $\gamma$ -polyglutamic acid increased cell migration in cultures and thus could be used for wound dressing<sup>66</sup>. A use of cyanophycin for the adsorption of anionic pollutants in wastewater has also been suggested<sup>67</sup>.

Cyanophycin derivatives already have commercial applications:  $\beta$ -Asp-Arg/Lys can serve as a nutritional amino acid source, and their administration as dipeptides can lead to higher availability<sup>68</sup>. Indeed, cyanophycin-derived dipeptides are already commercially available. Recently, the potential role of cyanophycin-derived dipeptides as tyrosinase inhibitors has also been evaluated<sup>69</sup>. Another promising use is as a source of polyaspartate, which is currently chemically synthesized. Polyaspartate is a biodegradable, biocompatible polymer with multiple potential biomedical<sup>70</sup> and industrial<sup>71</sup> application, for example as a green antiscalant or water softener<sup>72</sup>.

#### **1.4.2. Biotechnological production of cyanophycin *in vivo***

To realize the commercial potential of cyanophycin, it must be produced in large amounts and at low cost. The most promising way to achieve this is by *in vivo* production and subsequent purification of the polymer. Since cyanophycin can be produced by a single enzyme – CphA1 – it is relatively simple to bioengineer heterologous expression systems for its production. Many studies attempted to establish and optimize this process using a variety of hosts, CphA1 enzymes and growth conditions. An excellent review by Frommeyer et al. summarizes this field of study<sup>73</sup>.

Many organisms were tested as hosts for heterologous cyanophycin production. These include the bacteria *Escherichia coli*<sup>74</sup>, *Corynebacterium glutamicum*<sup>75,76</sup>, *Bacillus megaterium*<sup>77</sup>, *Ralstonia eutropha*<sup>77</sup>, *Sinorhizobium meliloti*<sup>78</sup> and *Pseudomonas putida*<sup>77</sup>; the fungi *Saccharomyces cerevisiae*<sup>79</sup>, *Pichia pastoris*<sup>80</sup> and *Rhizopus oryzae*<sup>81</sup>; and the plants *Nicotiana tabacum*<sup>82,83</sup> and *Solanum tuberosum*<sup>84</sup>. Generally, the use of bacteria as hosts has been the most successful, with yields of up to 48%, 43% and 38% w/w of cell dry mass (CDM) in *R. eutropha*, *P. putida* and *S. meliloti*, respectively. A recent analysis by Huckauf et al. showed that large-scale production of cyanophycin in tobacco plants, even with existing technologies, is already commercially viable<sup>83</sup>.

Native hosts can also be used as a source for cyanophycin. Many experiments attempted to optimize cyanophycin production in native bacterial hosts using both bioengineering and optimization of growth conditions. These include *A. baylyi*<sup>55</sup>, *A. cylindrica*<sup>17</sup>, *A. quadruplicatum*<sup>30</sup>, *Synechocystis* sp. PCC6308<sup>29,85</sup>, *Synechococcus* MA19<sup>86</sup> and *Synechocystis* sp. PCC6803<sup>86</sup>. Of these, metabolically engineered *A. baylyi* cells displayed the highest cyanophycin yield, at 46% CDM. Bacterial communities can also be used. A recent study examining the abundance of biopolymer-producing genes in bacterial sludge from wastewater treatment plants found a relatively high abundance of *cphA1*. Large amounts of cyanophycin could also be isolated from sludge samples, suggesting it could be used as an essentially free source of the polymer<sup>67</sup>.

As with expression hosts, the use of different CphA1s has also been explored to optimize production yields. These include *Synechocystis* sp. PCC6308<sup>77,87</sup>, *Anabaena* sp. PCC7120<sup>77,88</sup>, *Synechocystis* sp. PCC6803<sup>89,90</sup>, *Synechococcus* sp. MA19<sup>77,91</sup>, *A. baylyi*<sup>55</sup>, *A. variabilis* ATCC29413<sup>92</sup>, *Thermosynechococcus elongatus* BP-1<sup>84,93</sup>, *T. morbiroseti*<sup>90</sup> and CphA1<sub>49</sub> from a deep sea metagenome<sup>94</sup>. It is difficult to compare the yields of these different enzymes as, in addition to host and growth conditions, they may also differ in expression levels and stability under different conditions. However, the characterization of multiple CphA1s allows researchers to screen a wider range of variants in hope of finding the optimal CphA1-host combination.

An interesting aspect of cyanophycin production in heterologous hosts is that they lack a natural source of primers. Primers are short segments of cyanophycin that are extended by cyanophycin synthetases and are often required for these enzymes' activity. Nevertheless, various CphA1s display robust activity in different expression systems, suggesting the problem of primer availability can be at least partially overcome in those conditions. The ability of CphA1 to use non-cyanophycin primers, albeit with limited efficiency, is a possible explanation for this observation<sup>1,87</sup>. However, it remained unknown whether the lack of suitable primer material is a limiting factor in the production of cyanophycin in heterologous hosts.

### **1.4.3. *In vivo* production of cyanophycin variants**

The use of different systems for cyanophycin production also leads to variations in the characteristics of the produced material. While the polymer's backbone is almost always composed of Asp, as evidenced by the 50% Asp composition of analyzed samples<sup>73</sup>, the amino acid attached to their sidechains can vary. Normally, this other amino acid is Arg, although low levels of Lys are also observed in many cases<sup>73</sup>. However, by varying the CphA1, host and growth conditions,

cyanophycin-like polymers can be obtained that have high levels of other amino acids. In a notable study, Steinle et al. expressed CphA1 from *Synechocystis* sp. PCC6308 in *S. cerevisiae* strains carrying different inactivating mutations in the Arg metabolic pathway<sup>79</sup>. They found that deletion of the arginine succinate synthetase gene led to accumulation of cyanophycin with up to 20% citrulline content, and deletion of ornithine carbamoyltransferase led to polymer containing 8% ornithine. Other mutations or supplements added to the growth media also had an effect on the contents and amount of the produced polymer. Incorporation of citrulline (9% content) has also been observed when *Synechocystis* sp. PCC6308 CphA1 is expressed in *P. putida* ATCC4359<sup>95</sup>. These results highlight the potential catalytic flexibility of CphA1 under the right conditions. The composition of cyanophycin is important since it affects the polymer's properties. Cyanophycin is often purified in an "insoluble" form. This term refers to the polymer's solubility in aqueous solutions at neutral pH, although this form is highly soluble in acidic or basic conditions. However, a "soluble" form, which dissolves in water regardless of the pH level, can often be purified as well. Frommeyer et al. first reported that a major difference between soluble and insoluble forms of cyanophycin is their Lys content<sup>96</sup>. Working with different CphA1s expressed in *E. coli*, they found that soluble cyanophycin had a Lys content of at least 17%, while in the insoluble form it was up to 5%. A later study found a similar trend and also described the temperature-dependent solubility of cyanophycin with different Lys content. Generally, as the temperature and Lys content increase, cyanophycin becomes more soluble<sup>97</sup>. These properties are important as they allow for easy separation of polymer fractions with different characteristics.

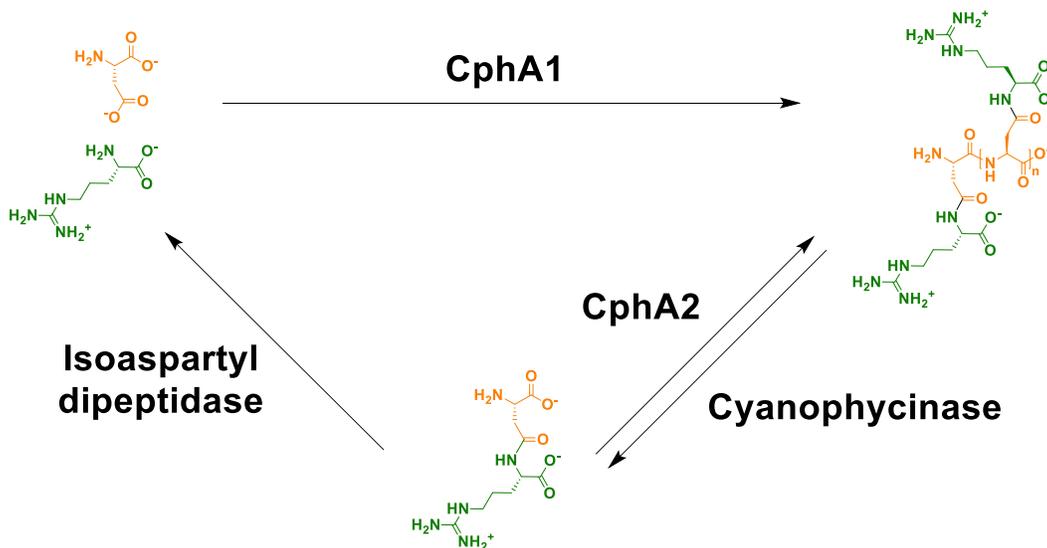
## **1.5. Enzymes involved in cyanophycin metabolism**

Cyanophycin is normally composed of L-Asp and L-Arg – two proteinogenic amino acids commonly found in cells. Thus, its metabolism is fairly simple and involves only one known non-polymer intermediate –  $\beta$ -Asp-Arg dipeptides. CphA1 makes cyanophycin from Asp and Arg and cyanophycinase degrades it to dipeptides. These dipeptides can either be further degraded to Asp and Arg, or re-polymerized into cyanophycin by CphA2. Figure 1.2 shows the known steps in cyanophycin synthesis and degradation.

### **1.5.1. Cyanophycin synthetase 1 (CphA1)**

In 1976, decades after the discovery and characterization of cyanophycin, Simon identified an enzyme capable of elongating cyanophycin chains using Asp, Arg and ATP<sup>98</sup>. As it was not yet

clear how many enzymes were involved in the full process he named it multy-L-arginyl-poly(L-aspartic acid) synthetase<sup>99</sup>. It would take more than 20 years for the gene encoding this enzyme to be identified, which enabled the enzyme to be better characterized<sup>98</sup>. This gene was first designated as cyanophycin synthetase (*cphA*), but following the discovery of other variants the more specific name *cphA1* was adopted<sup>100</sup>.



**Figure 1.2. The known steps in cyanophycin biosynthesis and biodegradation.** CphA1 polymerizes Asp and Arg into cyanophycin in an ATP-dependent manner. Cyanophycinase degrades cyanophycin to  $\beta$ -Asp-Arg dipeptides. These dipeptides can either be re-polymerized by CphA2 using ATP, or hydrolyzed into Asp and Arg by one of several different isoaspartyl dipeptidases.

CphA1 catalyzes the synthesis of cyanophycin from Asp and Arg in two ATP-dependent reactions, using two synthetic active sites. Like many other enzymes<sup>101</sup>, it requires  $K^+$  for activity, although the reasons for this are still unknown. During cyanophycin synthesis in cells of *Synechocystis* sp. PCC6803, CphA1 tends to associate with cyanophycin granules<sup>102</sup>, similar to the way polyhydroxy alkananoate (PHA) synthase is localized on the surface of PHA granules<sup>103</sup>. It then dissociates from those granules during periods of cyanophycin degradation. CphA1's tendency to bind cyanophycin has also been observed *in vitro*, and is greatly increased in the presence of  $Mg^{2+}$  ions<sup>3</sup>. CphA1s enzymes from multiple bacteria, often cyanobacteria, have been purified and characterized to some degree. Frommeyer et al. summarized most of these and the

polymer they produce *in vivo* under different growth conditions<sup>73</sup>. Table 1.1 lists studies that performed *in vitro* characterization of CphA1 and the source organism of each enzyme.

Most CphA1s have been described as being primer dependent<sup>98</sup>, meaning they can only extend existing chains of cyanophycin rather than start polymerization *de novo*. The minimal known length of cyanophycin that can serve as a primer is ( $\beta$ -Asp-Arg)<sub>3</sub><sup>1</sup>, although no reports excluded the possibility that shorter segments may also be sufficient. Some molecules other than cyanophycin, such as *N*-acetylglucosamine, can also serve as primers, albeit with low efficiency<sup>104</sup>. So far, only one CphA1 with robust primer-independent activity has been described<sup>105</sup> – that from the thermophilic cyanobacterium *Thermosynechococcus elongatus* BP-1. Neither the authors of this study nor subsequent analyses<sup>73</sup> were able to identify the source of this difference in primer dependence.

Organism	Purification	Reported Oligomerization	Study, Year
<i>Anabaena cylindrica</i>	yes	n.r	Simon, 1976 <sup>99</sup>
<i>Anabaena variabilis</i>	yes	dimer	Ziegler et al., 1998 <sup>98</sup>
<i>Synechococcus</i> sp. MA19	yes	n.r	Hai et al., 1999 <sup>86</sup>
<i>Synechocystis</i> sp. PCC6803	partial	n.r	Aboulmagd et al., 2000 <sup>106</sup>
<i>Anabaena variabilis</i> ATCC29413	yes	n.r	Berg et al., 2000 <sup>1</sup>
<i>Synechocystis</i> sp. PCC6308	yes	dimer	Aboulmagd et al., 2000 <sup>87</sup>
<i>Synechococcus</i> sp. MA19	yes	n.r	Hai et al., 2002 <sup>104</sup>
<i>Acinetobacter baylyi</i> DSM587	no	n.r	Krehenbrink et al., 2002 <sup>74</sup>
<i>Desulfitobacterium hafniense</i>	no	n.r	Ziegler et al., 2002 <sup>107</sup>
<i>Acinetobacter baylyi</i> DSM587	yes	n.r	Krehenbrink et al., 2004 <sup>3</sup>
<i>Anabaena</i> sp. PCC7120	no	n.r	Voss et al., 2004 <sup>77</sup>
<i>Nostoc ellipsosporum</i>	no	n.r	Hai et al., 2006 <sup>88</sup>
<i>Thermosynechococcus elongatus</i> BP-1	yes	tetramer	Arai et al., 2008 <sup>105</sup>
<i>Nostoc ellipsosporum</i>	yes	dimer	Hai et al., 2008 <sup>108</sup>
Unknown cyanobacterium 49	yes	n.r	Du et al., 2013 <sup>94</sup>

**Table 1.1. Studies that performed *in vitro* characterization of CphA1s.** n.r – not reported.

CphA1s are normally ~100 kDa in mass, and have been reported to form either dimers<sup>87,98</sup> or tetramers<sup>105</sup> in solution. Sequence alignment led to the identification of 2 distinct domains common to all CphA1s<sup>98,107</sup>: an ATP-grasp like domain approximately 300 residues in length (residues ~160-470 in *Synechocystis* sp. UTEX2470 CphA1); and a C-terminal, Mur-ligase like domain approximately 400 residues in length<sup>73</sup> (residues ~470-870). Both domains include ATP-binding sites, and it is assumed that there is an active site in each<sup>73,100</sup>. Both ATP-grasp enzymes<sup>109</sup> and Mur-ligases<sup>110</sup> often form dimers in solution, so CphA1 oligomers were thought to form in an analogous way to at least one of those families.

In addition to the two aforementioned domains, CphA1 also has a conserved N-terminal region. This part of the enzyme consists of the first ~160 residues. It does not align well with other known proteins, and displays high sequence variability in many positions. This N-terminal region has received little attention in the published literature. Only one review, relying on unpublished data, reports that the mutations C59A and C133A in *Synechocystis* sp. PCC6308 led to reduced activity in *E. coli*<sup>111</sup>. However, no attempt was made to explain these phenotypes. As this region is always present and quite large, it likely forms an important part of the enzyme.

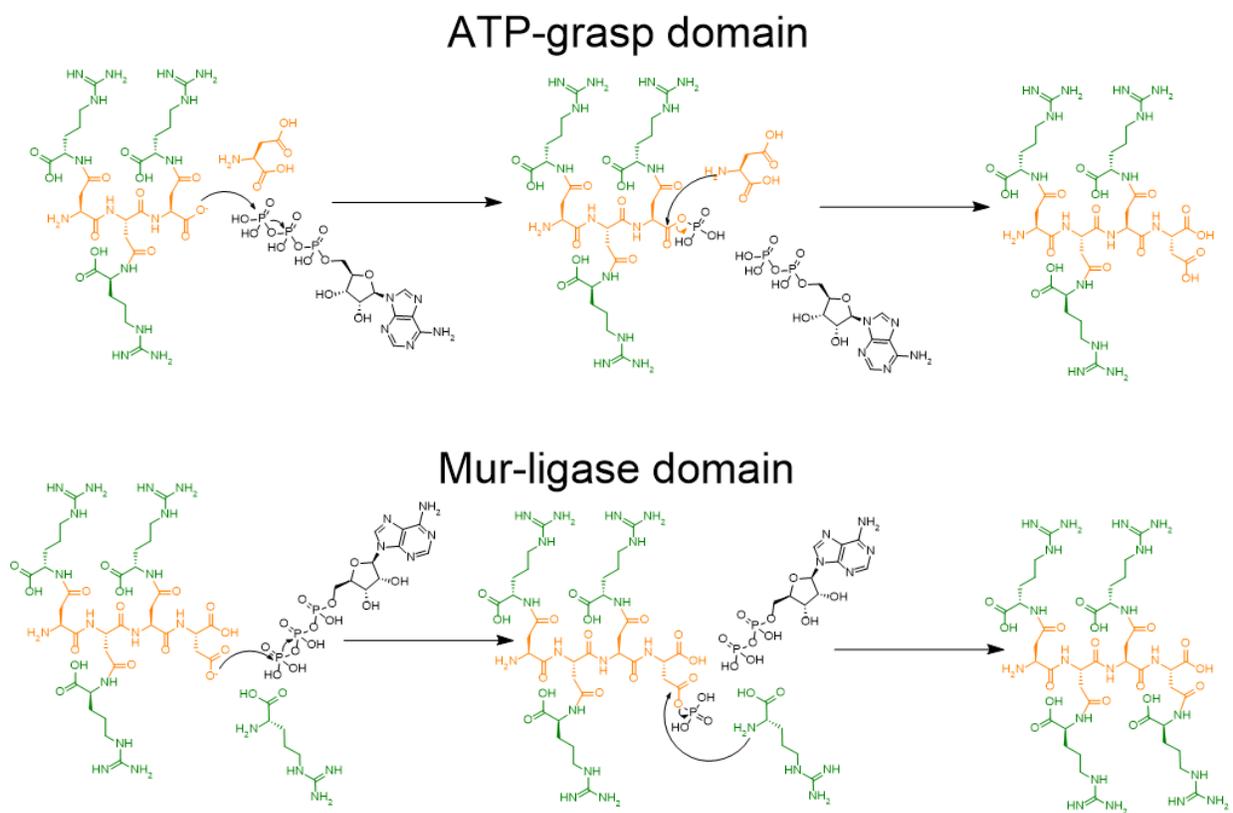
#### *1.5.1.1. The ATP-grasp like domain*

Residues ~160-470 of CphA1 shows sequence similarity to ATP-grasp enzymes<sup>100</sup>. This is a superfamily that includes some important enzymes, such as glutathione synthetase, biotin carboxylase and D-ala-D-ala ligase<sup>112</sup>. ATP-grasp enzymes normally ligate two substrates by catalyzing two sequential reactions: in the first, a carboxylate moiety of one substrate is phosphorylated using ATP, forming an acyl-phosphate intermediate; in the second, this intermediate is attacked by a nucleophile, leading to formation of the ligated product and release of phosphate and ADP (Fig. 1.3). Two exceptions are known which only catalyze the first part of the reaction, making those enzymes kinases rather than ligases<sup>113,114</sup>. Despite having overall low sequence similarity, ATP grasp enzymes display an overall similar fold. They have a stable core domain (composed of the A and C1 domains, as named in a previous review<sup>112</sup>) and a flexible lid domain (sometimes called B domain<sup>112</sup>). The lid forms part of the ATP-binding pocket and contains the flexible P-loop, which is important for interaction with the polyphosphate portion of ATP and thus for catalysis<sup>115</sup>. The core domain, in addition to forming the binding site of the phosphorylated substrate, also contains the “large loop<sup>116</sup>” which is important for selection and binding of the second substrate molecule<sup>117</sup>. In some enzymes, this loop is part of a larger “omega” subdomain<sup>7</sup>.

Sequence alignment shows that the ATP-grasp like portion of CphA1 has equivalents for the P and large loops<sup>98,107</sup>, and most of their ATP-binding residues of this family’s members are also conserved<sup>100</sup>. Interestingly, CphA1 requires K<sup>+</sup> for activity<sup>99</sup>, a feature shared with other ATP grasp enzymes<sup>118</sup>. This, coupled with the fact that, like ATP-grasp enzymes, CphA1 releases P<sub>i</sub> during catalysis<sup>98</sup>, makes it likely that this part of the enzyme has an active site that functions in a similar way to this family of enzymes.

### 1.5.1.2. The Mur-ligase like domain

The C-terminal ~400 residues of CphA1 show sequence homology to Mur-ligases. These are a family of four enzymes (MurC-F) that catalyze steps in the biosynthesis of peptidoglycan, the main component of the bacterial cell wall. A key step in the formation of peptidoglycan is the formation of UDP-*N*-acetylmuramic acid (UDPMurNAc) – a sugar moiety ligated to a short peptide. This peptide normally has the sequence L-Ala-D-Glu-X-D-Ala-D-Ala, with X being either *meso*-diaminopimelate (*m*DAP) or L-Lys. Following the formation of UDPMurNAc, the peptide portion is sequentially synthesized by the ligation of L-Ala (MurC), L-Glu (MurD), *m*DAP/L-Lys (MurE) and D-Ala-D-Ala (MurF). These four enzymes all share a 3-domain



**Figure 1.3. The catalytic mechanisms of the synthetic reactions catalyzed by CphA1.**

The ATP-grasp like domain phosphorylates the backbone carboxyl of cyanophycin<sup>1</sup>. The acyl-phosphate intermediate then undergoes nucleophilic attack by Asp, resulting in extension of the polymer's backbone. This product is then released from the first active site and binds the Mur-ligase like domain's active site, where the side chain of the previously added Asp is phosphorylated by ATP. Similar to the previous reaction, the intermediate is attacked by Arg, thus completing the catalytic cycle.

architecture with N-terminal (domain 1 in a previous review<sup>110</sup>), core (domain 2) and lid (domain 3) domains. The peptidoglycan substrate is bound by domains 1 and 2, while the flexible domain 3 functions as a lid and forms part of the ATP binding site<sup>110</sup>.

All Mur-ligases catalyze the ATP-dependent synthesis of amide bonds by formation of an acyl-phosphate intermediate, and as a result release P<sub>i</sub>. These features, as well as the presence of known ATP-binding residues<sup>100</sup> being conserved in CphA1, support the hypothesis that CphA1 has an active site in the C-terminal, Mur-ligase like portion. The conserved Mur-ligase like region of CphA1 is often followed by a shorter, less conserved section. Beyond residue ~870, CphA1s display high variability in sequence and length<sup>88</sup>. Two studies examined the role of this part of the enzyme through mutagenesis and truncation experiments of the 901-residue long CphA1 from *N. elliposporum* PCC7120. It was found that removal of the last 31 C-terminal residues increased the enzyme's activity more than 2-fold, both *in vitro* and *in vivo*<sup>88</sup>. Truncation of up to 45 residues retained enzymatic activity, but removal or mutagenesis of one more residue (E856) led to complete inactivation<sup>108</sup>. Subsequent experiments suggested these results to be caused by differences in enzyme stability, although a possible role for E856 in catalysis was not excluded<sup>108</sup>.

#### 1.5.1.3. The proposed mechanism for cyanophycin biosynthesis by CphA1

Berg et al. were the first to propose a detailed mechanism for cyanophycin synthesis by CphA1 using studies with synthetic primers<sup>1</sup>. This model, illustrated in Figure 1.3, includes two steps – one catalyzed by each of the two putative active sites. In the first step, the polymer backbone is phosphorylated using ATP, forming an acyl-phosphate intermediate. This intermediate is then attacked by Asp, resulting in the formation of an amide bond and the extension of the polymer backbone by one residue. In the second step, the side chain of the recently added Asp is phosphorylated with ATP and the intermediate attacked by Arg, completing the extension of the polymer by one dipeptide unit. Mutagenesis experiments showed that the N-terminal active site is responsible for incorporation of Asp into the polymer, and so it is assumed that Arg is incorporated by the C-terminal site<sup>111</sup>. Subsequent biochemical experiments found the K<sub>m</sub> values for Asp to be between 240-500 μM, and those of Arg to be between 15-50 μM. Two different K<sub>m</sub> values for ATP could be calculated (38 and 210 μM), supporting the proposed two-active site model<sup>3</sup>.

### 1.5.2. CphA2

CphA2 is a cyanobacterial enzyme first identified in *Anabaena* sp. PCC7120 in a *cphA2-cphB* cluster, which the strain has in addition to the more common *cphA1-cphB* cluster<sup>119</sup>. *cphA2* and the adjacent *cphB*, which are found in opposite orientations on the genome, are expressed monocistronically in the presence of N<sub>2</sub>, ammonium or nitrate. The operon's expression is higher in the absence of ammonium, suggesting it is increased by low availability of reduced nitrogen. A similar trend was observed in the toxic-bloom forming cyanobacteria *planktothrix*, which express higher levels of *cphA2* during seasons with low nitrogen availability, and *cphA1* when ammonium is abundant<sup>49</sup>.

CphA2 is only found in diazotrophic cyanobacteria that also have CphA1, suggesting its activity depends on the presence of CphA1<sup>12,120</sup>. Knockout experiments showed that cells lacking the *cphA2* gene accumulated 10-20% less cyanophycin under N<sub>2</sub>-fixing conditions<sup>12,119</sup>. These results suggest that CphA2 contributes less to cyanophycin accumulation than CphA1. However, CphA2 is nevertheless important for cell fitness and cannot be completely replaced by CphA1, with experiments in *A. variabilis* ATCC29413 showing that disruption of the *cphA2* gene lead to impaired growth under N<sub>2</sub>-fixing conditions<sup>12</sup>.

Sequence alignment shows that CphA2 is clearly related to CphA1, with equivalents to the ATP-grasp and Mur-ligase domains, as well as the N-terminal region. However, at the sequence level there are two main differences between the enzymes. First, the N-terminal region of CphA2, while similar in length to that of CphA1, displays very low sequence similarity to it. Second, the C-terminal domain of CphA2 is ~250 residues shorter than that of CphA1, being truncated at the C-terminus. This truncation removes most of the Mur-ligase lid domain, which is conserved and important for activity<sup>110</sup>. This is very likely to render this domain catalytically inactive, leaving CphA2 with only the ATP-grasp like active site intact<sup>12</sup>.

CphA2 catalyzes a single reaction – the ATP-dependent polymerization of β-Asp-Arg dipeptides to form cyanophycin<sup>12</sup>. It was presumed to do so using the ATP-grasp like active site, which evolved to accept β-Asp-Arg as substrate instead of Asp as in CphA1. This difference in substrate specificity may be due to differences in the sequence of the ATP-grasp “large loop”, which is believed to be important for substrate specificity in these enzymes<sup>117</sup>.

The CphA2 enzymes from two different cyanobacteria - *A. variabilis* ATCC29413 and *Cyanothece* sp. PCC7425 – have been purified and characterized. These enzymes are 75 and 73

kDa in mass, respectively. Size exclusion chromatography (SEC) experiments suggest that they form either trimers or tetramers in solution. Activity assays showed that they are active in the presence of high  $\beta$ -Asp-Arg concentration (100 mM) in absence of primer, and their activity levels increase in the presence of cyanophycin primer<sup>12</sup>. Like CphA1, CphA2 also requires K<sup>+</sup> for activity.

### 1.5.3. Cyanophycinase

Cyanophycinase degrades cyanophycin to  $\beta$ -Asp-Arg dipeptides by sequentially cleaving the C-terminal dipeptide from a cyanophycin chain<sup>121</sup>. It is the key enzyme in cyanophycin degradation, as the polymer is resistant to common proteases. Consequently, it is found in the genomes of many cyanophycin-producing bacteria<sup>100</sup>, although the existence of some strains with *cphB* and not *cphA1* suggests that other enzymes can play a similar role<sup>100</sup>. First isolated in 1999 from *Synechocystis* sp. PCC6803<sup>121</sup>, cyanophycinase displays sequence similarity to peptidase E from *Salmonella enterica* serovar Typhimurium (EC 3.4.13.21, PepE), which is an aspartyl dipeptidase. Cyanophycinase exists in several different forms: CphB is a ~30 kDa protein that forms dimers in solution and is used for intracellular degradation of cyanophycin<sup>121</sup>; CphI is a pseudodimeric version of CphB that only has one active site<sup>100</sup>; and CphE is a ~45kDa, extracellular cyanophycinase used for cyanophycin scavenging<sup>61</sup>. Bacteria typically have only one version of cyanophycinase.

*cphB* and *cphI* are often found in genomes that also have *cphA1*, and tend to be adjacent to it. Despite this, and as expected for two enzymes with catabolic and anabolic functions respectively on a common molecule, cyanophycinase and CphA1 are not always co-transcribed. Studies conducted in *Anabaena* sp. PCC7120 showed that *cphA1* and *cphB* can be co-transcribed from 3 putative promoters, but two promoters in the intergenic region lead to monocistronic expression of only *cphA1*<sup>119</sup>. This suggests that the cyanophycin metabolism pathway can be controlled in various ways, likely in response to different extra- and intracellular conditions<sup>29,30,122</sup>. *cphE*, however, is found in bacteria (and some fungi) that do not make cyanophycin. Little is known about the mechanism controlling its expression, but it is interesting that extracellular cyanophycinase activity has been detected in many bacterial isolates from a variety of environments<sup>57,60,61</sup>.

Cyanophycinase is a serine-protease like enzyme, with a conserved E-H-S catalytic triad (E201, H174 and S132 in *Synechocystis* sp. PCC6803 CphB), similar to PepE<sup>121</sup>. As such, like

serine proteases, it is inhibited by phenylmethanesulfonyl fluoride (PMSF) and 4-(2-aminoethyl)benzenesulfonyl fluoride (AEBSF) but not by aspartic protease inhibitors<sup>121</sup>. It is also highly specific towards cyanophycin, and displays almost no proteolytic activity<sup>121</sup>. As with other serine-proteases, the catalytic mechanism of cyanophycinase involves formation of a covalent enzyme-substrate intermediate through an ester bond between the active site serine and the backbone C-terminus of cyanophycin. The formation of this intermediate leads to the release of the P1' dipeptide from the C-terminus of a cyanophycin chain. The covalent intermediate is then attacked by a water molecule, releasing the product cyanophycin chain from the enzyme.

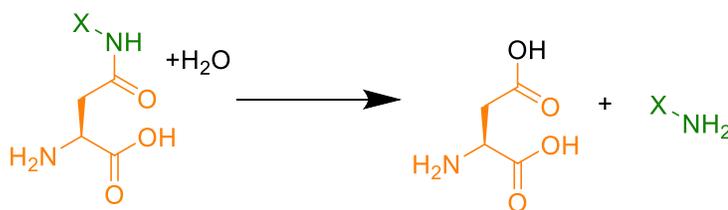
This catalytic mechanism makes it possible to study covalent complexes of serine-proteases with substrates or inhibitors. One system which allows for such studies is the introduction of the unnatural amino acid diaminopropionic acid (DAP) in place of the active site Ser using TAG codon suppression and an orthogonal tRNA synthetase<sup>123</sup>. This replaces the nucleophilic moiety of the active site from oxygen to nitrogen. Consequently, the covalent enzyme-substrate intermediate is formed through an amide bond rather than a labile ester bond. This makes the resulting complexes similar in structure to the native ones, but much more stable and thus suitable for structural studies.

The structure of *Synechocystis* sp. PCC6803 CphB without cyanophycin substrate/product has been determined, and provides insight into the enzyme's active site and possible mechanisms for substrate recognition and binding<sup>124</sup>. Proteases often use deep binding pockets that accommodate specific side chains. Cyanophycin is a peptide-like material, but the  $\beta$ -Asp-Arg residues give it much bulkier sidechains than those of proteinogenic amino acids. This prevents cyanophycin from fitting in the binding pockets of standard proteases, making it resistant to proteolytic degradation by enzymes such as trypsin. In contrast to proteases, cyanophycinase is thought to bind cyanophycin using very shallow binding pockets on both sides of the active site. The P1' dipeptide is presumably bound in the shallow S1' pocket formed by the conserved E16, K18, Y56 and T131 (in *Synechocystis* sp. PCC6803 CphB), and the P1 dipeptide is bound in the S1 groove formed by the conserved Q101, Q173, R178, R180 and R183<sup>124</sup>. Three conserved Arg residues (R178, R180 and R183 in *Synechocystis* sp. PCC6803 CphB) form a basic patch in the S1 groove, and are important for the enzyme's activity. Based on their distance from the catalytic S132 residue and substrate-binding models, they are probably important for substrate binding.

Mutagenesis experiments showed that all the aforementioned residues are indeed important for the enzymatic activity of CphB.

#### 1.5.4. Isoaspartyl dipeptidases

Enzymes with isoaspartyl dipeptidase activity are common in nature and found in a variety of organisms, since  $\beta$ -aspartyl dipeptides can form by spontaneous rearrangements in proteins, and are a major form of damaged proteins<sup>125</sup>. These dipeptides cannot be degraded by proteases and peptidases, as these are typically specific to  $\alpha$ -peptides, and so can potentially accumulate to toxic levels in cells<sup>126,127</sup>. To prevent this, organisms have pathways to either fix damaged proteins<sup>127</sup> or degrade the  $\beta$ -aspartyl dipeptides left after proteolytic degradation to amino acids<sup>128</sup> (Fig. 1.4). In bacteria, two enzymes are known which display this activity – isoaspartyl dipeptidase (IadA<sup>128</sup>) and isoaspartyl aminopeptidase (IaaA<sup>129</sup>). IadA is a ~40 kDa bacterial metalloenzyme related to dihydroorotases and imidases. It forms octamers in solution and has a binuclear  $Zn^{2+}$  active site. IaaA is a common enzyme found in plants and animals as well as bacteria. It is a Ntn family enzyme that is expressed as a proenzyme and is converted into the active form by auto-proteolytic cleavage into  $\alpha$  and  $\beta$  subunits. A conserved Thr residue is critical for both the maturation and catalytic activity of this enzyme. Both IadA and IaaA have broad substrate specificity and hydrolyze a variety of  $\beta$ -aspartyl dipeptides, as their substrate recognition and binding mostly rely on the Asp portion of the dipeptide.



**Figure 1.4. The general reaction mechanism catalyzed by isoaspartyl amidohydrolases.** When X is any amino acid, the enzyme is said to have isoaspartyl

The only known pathway for cyanophycin degradation is through cyanophycinase, which makes  $\beta$ -Asp-Arg/Lys dipeptides. Thus, enzymes with isoaspartyl dipeptidase activity are required for the final degradation step of cyanophycin into Asp and Arg/Lys. Research conducted in cyanobacteria suggests this step is carried out by IaaA. The importance of this enzyme for cyanophycin degradation was first studied in detail by Lockau et al.<sup>129</sup>, who found that the IaaA enzymes from *Synechocystis* sp. PCC6803 and *Anabaena* sp. PCC7120 displayed isoaspartyl

dipeptidase activity. The authors noted that IaaA can degrade  $\beta$ -Asp-Arg/Lys dipeptides, and suggested that it can be a key step in cyanophycin degradation. A later study by Flores et al.<sup>40</sup> found that in the heterocyst-forming cyanobacteria *Anabaena* sp. PCC7120, heterocysts release  $\beta$ -Asp-Arg and vegetative cells express high levels of IaaA. This result suggested that the cyanophycin-producing heterocysts use  $\beta$ -Asp-Arg dipeptides to shuttle fixed nitrogen to vegetative cells. The vegetative cells express IaaA to hydrolyze those dipeptides into free amino acids which can be used by other metabolic processes. Other studies suggested that other enzymes might also be used for  $\beta$ -Asp-Arg degradation, but no experimental results support these hypothesis<sup>100,121</sup>. Specifically, no biochemical study has shown that IadA is involved in cyanophycin degradation.

In 2007, Füsler and Steinbüchel analyzed 48 genomes that have at least one of CphA1 and cyanophycinase, and found that 12 of them contained an *iadA* and 15 an *iaaA* copy. This left 22 genomes with a known cyanophycin metabolizing gene and no apparent means of completing its degradation. The authors hypothesized that at least some of these bacteria have uncharacterized enzymes with isoaspartyl dipeptidase activity. Other possible explanations might be that some bacteria degrade cyanophycin without cyanophycinase, use cyanophycin for purposes that do not require its degradation to amino acids, or rely on other members of the bacterial community to provide the necessary isoaspartyl dipeptidase activity<sup>58</sup>.

Thus, since the discovery of cyanophycin over a century ago, much had been learned about its metabolism and uses. However, many questions still existed, and our structural knowledge of cyanophycin synthetase was gravely lacking.

## Bridge to chapter 2

Cyanophycin synthetase 1 is the most common enzyme that makes cyanophycin, and has been known for decades. Consequently, many studies sought to characterize its activity and optimize it for biotechnological applications. These studies made important discoveries: Enzymes with sequence homology to CphA1 were isolated, the enzyme's substrate specificity was explored, favorable conditions for CphA1 activity were fine-tuned and mutants with desirable phenotypes were identified. However, as the structure of CphA1 remained undetermined, researchers were left guessing as to the causes for many of the obtained results. It was unclear, for example, why the enzyme required certain reaction conditions, what determined its substrate specificity or why some mutations had the effects that they did. Even seemingly basic questions, like the identity and function of the enzyme's active sites or the role of its 160-residue long N-terminus, remained at least partially unanswered. Therefore, I undertook structure determination of CphA1.

## 2. Structures and function of the amino acid polymerase cyanophycin synthetase

Published in: Sharon I, Haque AH, Grogg M, Lahiri I, Seebach D, Leschziner AE, Hilvert D, Schmeing TM. *Nature Chemical Biology* 2021 Oct; 17, 1101–1110.

## **2.1. Abstract**

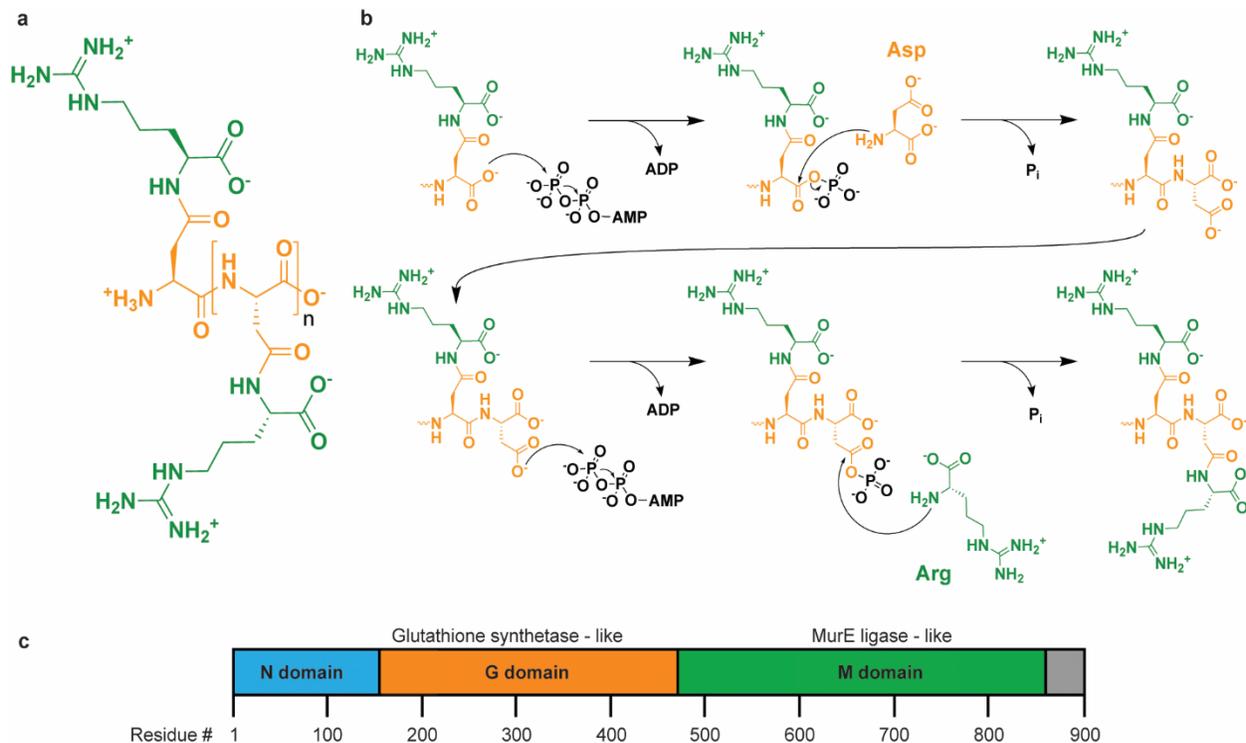
Cyanophycin is a natural biopolymer consisting of a chain of poly-L-Asp residues with L-Arg residues attached to the  $\beta$ -carboxylate side chains by isopeptide bonds. First discovered in cyanobacteria in 1886, cyanophycin is produced by a wide range of bacteria and is important for cellular nitrogen storage. Cyanophycin is synthesized from ATP, aspartic acid and arginine by a homooligomeric enzyme called cyanophycin synthetase (CphA1). CphA1 has domains that are homologous to glutathione synthetases and muramyl ligases, but no other structural information has been available, and many questions regarding its structure and mechanism of action were outstanding. Here, we present cryo-electron microscopy and X-ray crystallography structures of cyanophycin synthetases from three different bacteria, including co-complex structures of CphA1 with ATP and cyanophycin polymer analogs at 2.6 Å resolution. These structures reveal two distinct tetrameric architectures, show the configuration of active sites and polymer-binding regions, indicate dynamic conformational changes, and afford insight into catalytic mechanism. Accompanying biochemical interrogation of substrate binding sites, catalytic centers and oligomerization interfaces combine with the structures to provide a holistic understanding of cyanophycin biosynthesis.

## **2.2. Introduction**

Cyanophycin was described over 130 years ago as a light-scattering granule in cyanobacterial cells<sup>14</sup>. It is a biopolymer of a poly-L-Asp backbone with L-Arg residues attached via isopeptide bonds to the  $\beta$ -carboxylates of each Asp side chain, ranging in length from 80 to 400 dipeptides ( $(\beta\text{-Asp-Arg})_{\sim 80-400}$ )<sup>15,99</sup> (Fig. 2.1a). The high nitrogen content of cyanophycin makes it good for storage of fixed nitrogen<sup>130</sup>. With 24% nitrogen by mass, it does so more efficiently than proteins (~13-19%) and nucleic acids (~16%)<sup>24</sup>, whereas glycogen and fat contain no nitrogen. Cyanophycin is useful for bacteria that keep nitrogen and carbon fixation separated, either spatially or temporally<sup>42</sup>, because nitrogenase enzymes are inactivated in aerobic environments<sup>131</sup>. In single-cell cyanobacteria, cyanophycin is synthesized during periods of low light, when aerobic photosynthesis does not occur, and consumed during periods of high light<sup>42</sup>. In multicellular cyanobacteria communities, cyanophycin metabolism can be performed in specialized heterocysts, dedicated cells for fixing nitrogen<sup>132</sup> or spore-like akinetes<sup>52</sup>. The heterocysts generate and store cyanophycin, which is degraded when needed and transferred to vegetative cells that cannot fix nitrogen<sup>40</sup>. Cyanophycin can also be used for carbon and energy storage<sup>133,134</sup>, as a scavenged nutrient source<sup>61</sup>, for other metabolic processes<sup>30,133</sup>, in spore assembly<sup>54</sup>, and in plant-symbiont relationships<sup>135,136</sup>.

Commercial interest in cyanophycin has led to its heterologous production in hosts from bacteria to tobacco<sup>75,82,93</sup>. Potential commercial applications include processing to poly-Asp for use as a biodegradable antiscalant, water softener, and super swelling material<sup>137</sup>. Cyanophycin-type polymers are also interesting for biotechnological applications, like formation of heat-sensitive nanovesicles<sup>64</sup>.

Cyanophycin is synthesized in bacteria by cyanophycin synthetase (CphA1) using ATP, aspartate and arginine<sup>98</sup> (Fig. 2.1b). CphA1 is a dimer<sup>104</sup> or tetramer<sup>105</sup> of ~900 residue monomers. The N-terminal ~160 residues ("N domain") show no similarity to other proteins, the middle ~300 residues ("G domain") are homologous to glutathione synthetase<sup>1,138</sup>, and the C-terminal ~400 residues ("M domain") have homology to MurE-like muramyl ligases<sup>139</sup>(Fig. 2.1c). glutathione synthetase and MurE both catalyze formation of single amide bonds by activating carboxylates via ATP-dependent phosphorylation, but are structurally unrelated. The amide bond forming functions of glutathione synthetase and MurE appear to have been co-opted by CphA1 for cyanophycin



**Figure 2.1. Cyanophycin and cyanophycin synthetase.** (a) The chemical structure of cyanophycin. The backbone is made of L-Asp residues and each Asp side chain is linked through an isopeptide bond to an L-Arg residue.  $n=80-400$ . (b) The reactions catalyzed by CphA1. Top: First, the terminal carboxylate of the cyanophycin chain is phosphorylated and extended by one L-Asp residue. Bottom: Then, the side chain of the newly added L-Asp residue is phosphorylated and decorated with an L-Arg residue. Error bars represent the standard deviation of the replicates. (c) Analysis of the amino acid sequence of CphA1 reveals three major domains: an N-terminal domain (blue) with no known protein homologue, a middle G domain (orange) homologous to bacterial glutathione synthetase (and other ATP-grasp domain enzymes), and a C-terminal M domain (green) homologous to MurE ligase.

polymerization (Fig. 2.1b): CphA1 adds one L-Asp to the growing polymer's backbone, then ligates an L-Arg to the side chain of that Asp, with each reaction releasing ADP and phosphate<sup>98,138,139</sup>. The G domain has been shown to extend the Asp backbone, so the M domain is assumed to attach Arg to the Asp side chains<sup>98,138,139</sup>. Cyanophycin synthesis is usually “primer

dependent”<sup>1</sup>, where CphA1 only extends an existing segment of cyanophycin or, less efficiently, another biopolymer<sup>104,105</sup>.

Many studies sought to dissect, characterize and exploit CphA1, and its overall activity and substrate specificity is established<sup>1,88,98,99,105,140</sup>. However, without structural information, the results could not all be rationalized, and a holistic understanding of cyanophycin synthetase function has been lacking. Critically, it was not understood how CphA1 combines the activity of its active sites to achieve the combined, iterative process of cyanophycin synthesis.

We have determined structures of CphA1 from three bacterial species, including high resolution structures with substrate analogs. These structures and accompanying biochemical experiments provide an overall understanding of cyanophycin synthesis, including how the constituent domains work together to make cyanophycin.

## **2.3. Results**

### **2.3.1. CphA1 is a common bacterial enzyme**

Research on cyanophycin has largely focused on cyanobacteria, because of its discovery in that phylum. In 2007, Füsler and Steinbüchel reported *cphA1* genes in 44 of 946 bacterial genomes analyzed, including in non-cyanobacterial species<sup>100</sup>. In the current NCBI non-redundant protein data sequence bank, we found >4000 *cphA1* sequences. Strikingly, only 18% of species encoding CphA1 are cyanobacterial (Supplementary Fig. 2.1). CphA1 is found in most bacterial phyla, including groups like *Rhizobiales*, which form symbiotic relationships with legumes<sup>141</sup>; *Nitrosomonas* and *Nitrospira*, which are important for the nitrogen cycle<sup>142</sup> and wastewater treatment<sup>143</sup>; and *Clostridia*, including pathogens *C. botulinum* and *C. tetani*. Like many secondary metabolite genes, *cphA1* is not conserved in every strain of a species or every member of a clade despite its ability to confer a fitness advantage<sup>102,144</sup>. There is evidence for both ancient and recent horizontal gene transfer and repeated loss of *cphA1* (Supplementary Fig. 2.1).

Cyanophycin synthetases from fourteen species were selected: six firmicutes, four cyanobacteria, two gammaproteobacteria, one betaproteobacteria, and one alphaproteobacteria. Three could be expressed in *E. coli* and purified as robust samples: cyanobacterial *Synechocystis* sp. UTEX2470 (*Su*CphA1), and gammaproteobacterial *Acinetobacter baylyi* DSM587 (*Ab*CphA1) and *Tatumella morbirosei* DSM23827 (*Tm*CphA1) (Supplementary Fig. 2.1). These CphA1s

produced cyanophycin *in vitro* from Asp, Arg, ATP and cyanophycin primer with different kinetics (Fig. 2.2a,b), within the range of previously-reported rates<sup>3</sup>.

We determined structures of all three enzymes: *Su*CphA1 to 2.6 Å resolution by cryo-EM, *Ab*CphA1 to 4.4 Å by cryo-EM and *Tm*CphA1 to 3.1 Å by X-ray crystallography (Fig. 2.2c,d, Supplementary Tables 2.1-3). The structural and solution data indicate that all three enzymes are tetramers, assembled as dimers of dimers (Fig. 2.2c,d, Supplementary Fig. 2.2). All CphA1 monomers and dimers are similar to each other, but they form two distinct tetrameric architectures (Fig. 2.2, Supplementary Fig. 2.2a-d).

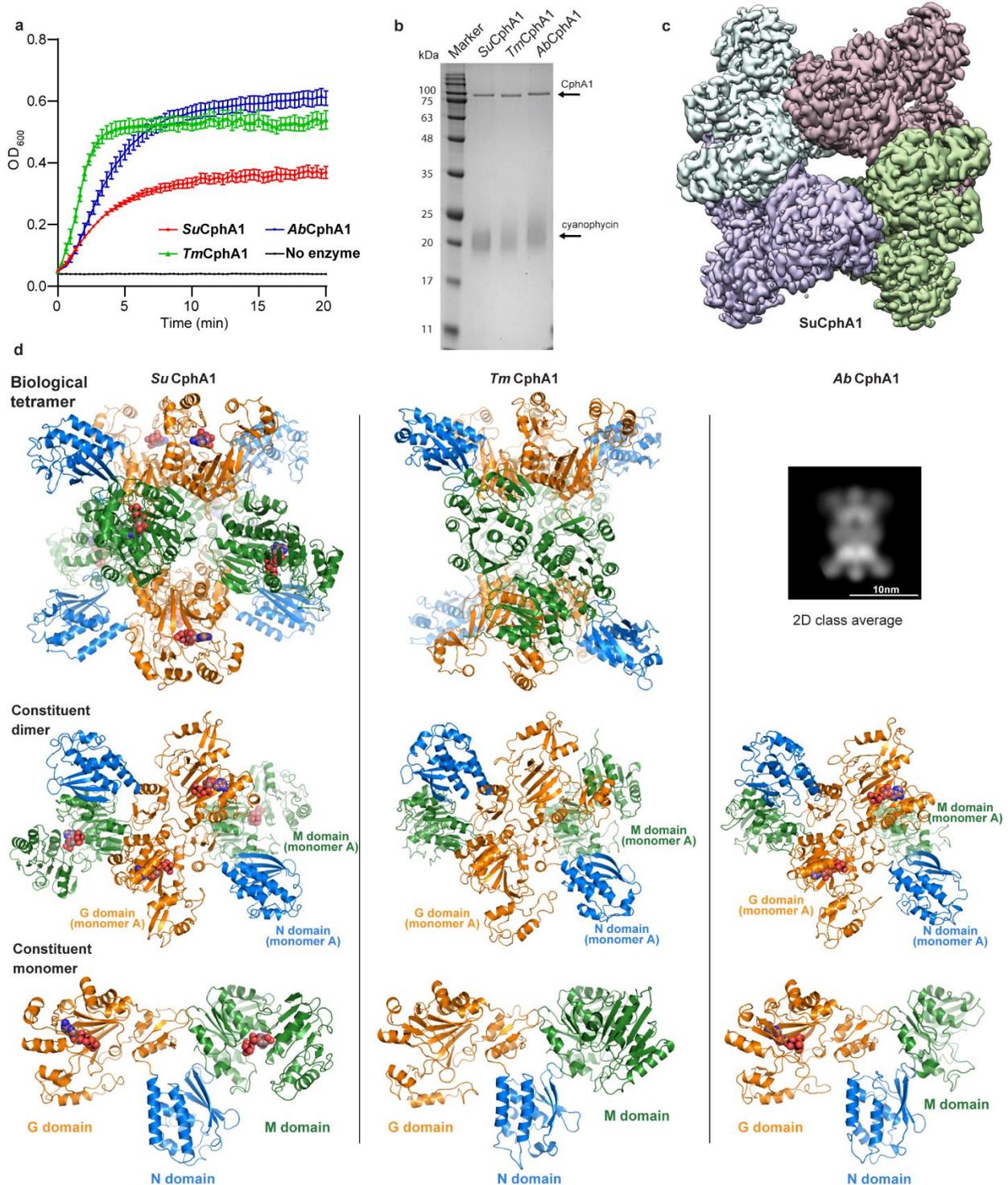
### 2.3.2. Architectures of cyanophycin synthetase

CphA1 monomers are tri-lobed, with each lobe corresponding to one of the three domains (Fig. 2.2d – bottom). The central lobe is the N domain (*Su*CphA1 residues 1-161), flanked on one side by the G domain (162-470) and on the other by the M domain (490-875). The G and M domain active sites face approximately the same side of the monomer, but are ~60 Å apart. The CphA1s have similar rotationally symmetric dimers, with extensive dimer interfaces burying ~1800 Å<sup>2</sup> of surface area (Fig. 2.2d – middle).

The tetramer architectures differ between gammaproteobacterial and cyanobacterial CphA1 (Fig. 2.2d – top, Supplementary Fig. 2.2). In *Su*CphA1, each tetramer interface buries only ~450 Å<sup>2</sup> of surface area, through M domain residue W672 inserting into a pocket near R470 of the G domain of the adjacent dimer (Supplementary Fig. 2.2e). The back sides of G and M

domains pack pseudo-symmetrically, and the N, G and M domain bodies radiate out so the *Su*CphA1 tetramer takes a “spiky ball” / “morning star” shape with a large central ovoid cavity of ~28 to 54 Å diameter (Fig. 2.2d, Supplementary Movie 1, Supplementary Figure 2.2a-d). In *Tm*CphA1, the tetramer interface is very different (Supplementary Movie 2). Relative to *Su*CphA1, one *Tm*CphA1 dimer is shifted by ~20° and ~10 Å, allowing M domains to form an interface of 1810 Å<sup>2</sup> of buried surface area. Remarkably, this shift means that different monomers of the dimer make the tetramer interfaces (*Su*CphA1 molecules A and C vs. *Tm*CphA1 molecules A and D; Supplementary Fig. 2.2a-d). Although D2 symmetry is maintained, this gives a distinctively different shape to *Tm*CphA1, of a “spiky ring” with a large central cavity of ~40x45x50 Å (Fig. 2.2d). *Ab*CphA1, *Tm*CphA1 and *Su*CphA1 are all tetrameric in solution (Supplementary Fig. 2.2f), but most *Ab*CphA1 tetramers dissociate into dimers on the EM grid. The 4.4 Å reconstruction of *Ab*CphA1 is this dimer, but some class averages clearly show a tetramer similar to *Tm*CphA1 (Fig.

2.2d, top right). The key *Su*CphA1 tetramerization residue, W672, is conserved in cyanobacteria and *Bacteroidetes*, but not elsewhere (Supplementary Fig. 2.2). These groups represent ~30% of CphA1s, and they likely all have the morning star shape.



**Figure 2.2. Overall structure and activity of CphA1.** (a) Cyanophycin-synthesis activity of the three homologs used in this study. Cyanophycin polymer formed in the reaction scatters light, causing an increase in OD<sub>600</sub>. The activity rates determined for each homolog are: *Su*CphA1 – 149U, *Tm*CphA1 – 460U, *Ab*CphA1 – 249U, where 1 U is defined as the incorporation of 1nmol ( $\beta$ -Asp-Arg)/min<sup>3</sup>. Data points represent the mean value of 3 measurements and the error bars show SD values. (b) Non-quantitative SDS-PAGE of the reaction mixtures of all three homologs show CphA1 (~100kDa) show similar size of cyanophycin product (~20kDa). (c) Cryo-EM map of tetrameric *Su*CphA1 complexed with ATP at 2.6 Å resolution, segmented by monomer. (d) Top: The tetrameric architecture of the three homologs used in this study. Middle: The constituent dimers that make up the tetramers. A region near the beginning of the G domain (181-232) contributes most (~1100 Å<sup>2</sup>) of the buried surface area. The 4.4 Å reconstruction of *Ab*CphA1 (middle right) is a dimer because most of the particles dissociate to dimers when applied to cryo-EM grids. *Ab*CphA1 is tetrameric in solution (Supplementary Fig. 2.2) and some particles remain as intact tetramers, as exemplified by the 2D class average shown. This 2D class average clearly shows that *Ab*CphA1 has a similar tetramer architecture as *Tm*CphA1. The constituent dimers (middle) and monomers (bottom) of *Su*CphA1, *Tm*CphA1 and *Ab*CphA1 are similar. ATP is shown in spheres to mark the active site of the G domain of *Su*CphA1 and *Tm*CphA1, and of the M domain of *Su*CphA1.

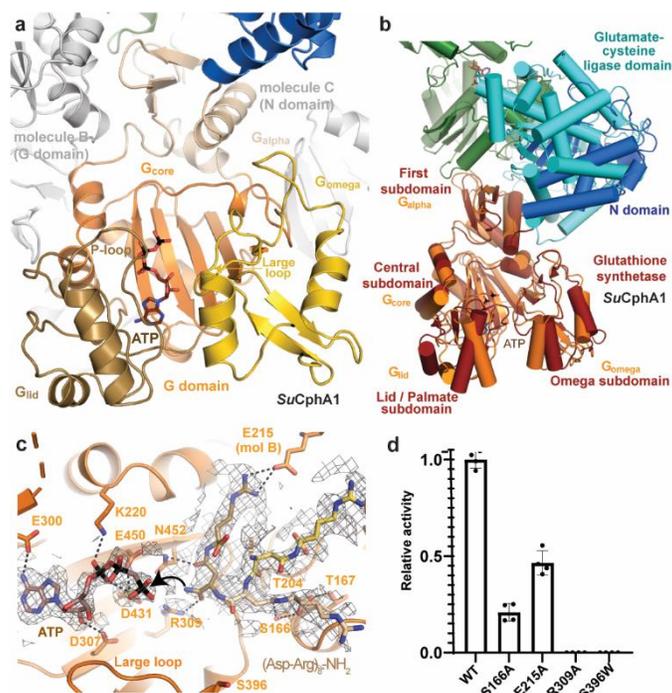
Because CphA1 has domains with distinct functions and evolutionary origins, we first analyze the domains separately and then analyze how individual activities combine to achieve cyanophycin synthesis.

### 2.3.3. Structure and mutational analysis of the G domain

The G domain catalyzes ATP-dependent addition of Asp to the C-terminus of cyanophycin polymer<sup>98,138,139</sup> (Fig. 2.1b). The active site of the G domain is located between the body of the G domain (G<sub>core</sub>) and two subdomains, G<sub>lid</sub> (*Su*CphA1 residues 235-305) and G<sub>omega</sub> (residues 325-399) (Fig. 2.2d, 2.3a). G<sub>core</sub> and G<sub>lid</sub> are also present in bacterial glutathione synthetases<sup>138</sup> and *D*-alanine-*D*-alanine ligases<sup>145</sup> (Fig. 2.3b). The G<sub>omega</sub> subdomain incorporates the “large loop”<sup>116</sup> of glutathione synthetase (Fig. 2.3a), and was previously only seen in a fused glutathione synthetase/glutamate-cysteine ligase, which also shares a modified ATP-grasp topology with

CphA1 (Fig. 2.3b)<sup>7</sup>. The overall binding of ATP is similar to that of ATP-grasp enzymes (Supplementary Fig. 2.3a), and has a partially ordered “P-loop” (residues 263-269)<sup>138,146</sup> covering part of the active site (Fig. 2.3a). *Su*CphA1 P-loop residue H267 is conserved as His or Gln in CphA1, but is Gly, Ser or Thr in other ATP-grasp enzymes<sup>147</sup>.

Cryo-EM datasets of *Su*CphA1 in the presence of the Asp, ADPCP and cyanophycin analog ( $\beta$ -Asp-Arg)<sub>8</sub>-NH<sub>2</sub> (Supplementary note section 2.3.5, Supplementary Fig. 2.3f) gave a 2.6 Å resolution map with clear signal for ATP and cyanophycin analogs at the active site (Fig. 2.3c).



**Figure 2.3. Structure and mutagenesis of the G domain.** (a) The overall structure of *Su*CphA1 G domain, colored by subdomain. (b) Overlay of the *Su*CphA1 G domain and glutathione synthetase-cysteine ligase from *S. agalactiae*<sup>7</sup> showing the overall structure, including G<sub>core</sub>, G<sub>lid</sub> and G<sub>omega</sub>. (c) The structure of the *Su*CphA1 G domain complexed with (Asp-Arg)<sub>8</sub>-NH<sub>2</sub> and ADPCP. The Cryo-EM map was carved 2 Å around the substrates at level 4.5. (d) Activity assays of *Su*CphA1 G domain mutants. S166, E215, and R309 bind cyanophycin close to the active site. S396W is assumed to lose activity by blocking the incoming Asp binding site. All measurements were performed in quadruplets. n=4 independent experiments. Data are presented as individual measurements and mean value, error bars represent SD values.

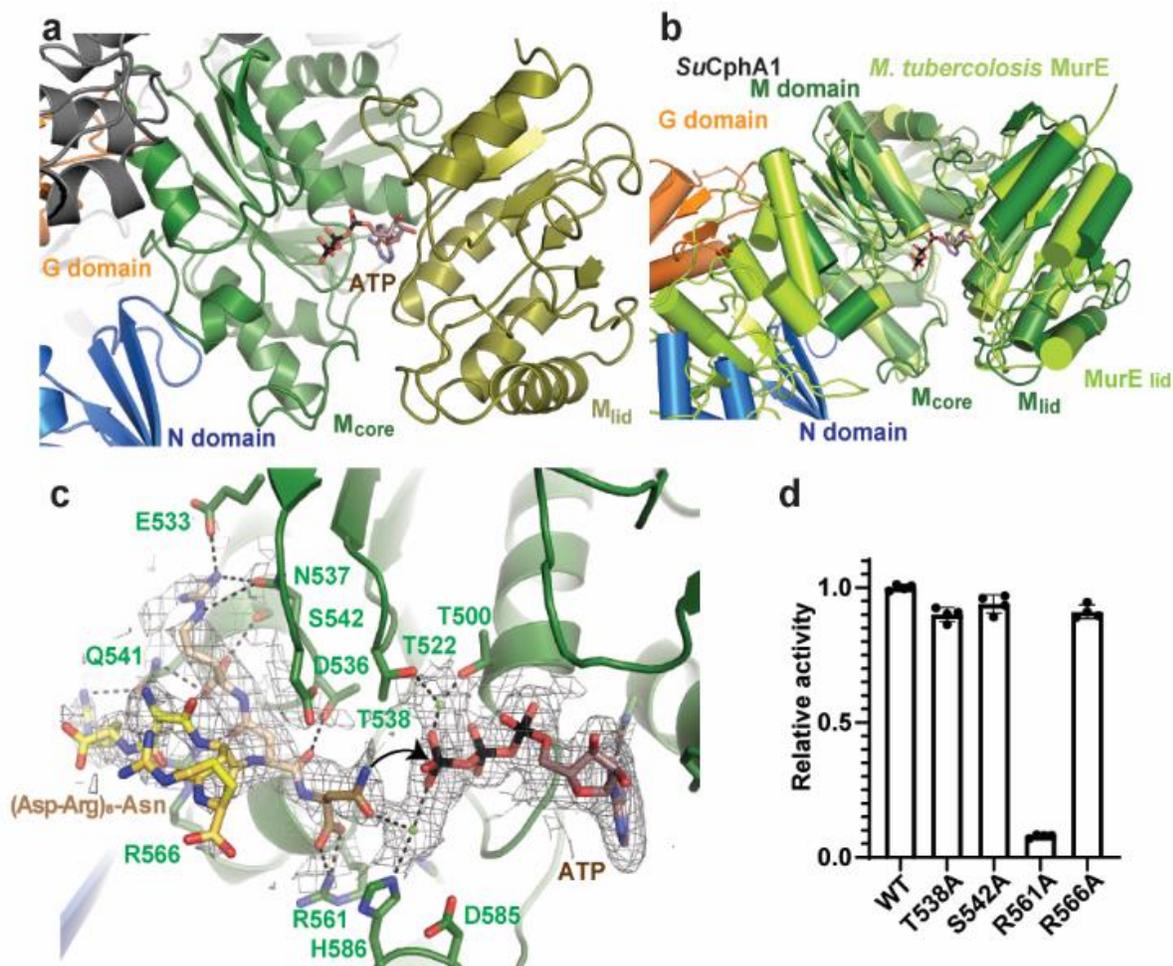
Three  $\beta$ -Asp-Arg dipeptides are visible, with the most C terminal dipeptide near the ATP analog. The two adjacent dipeptides extend from the G domain active site along  $G_{\text{core}}/G_{\text{alpha}}$ , away from  $G_{\text{lid}}$  and  $G_{\text{omega}}$ , and toward the N domain. The polymer makes several interactions with the active site and surroundings, including with conserved residues S166, R309 and E215. Mutation of these residues to alanine reduced or eliminated activity (Fig. 2.3d). The terminal amide group is  $\sim 6 \text{ \AA}$  from the  $\gamma$ -phosphate of ADPCP (Fig. 2.3c) in the average conformation represented by the EM map, just out of range for nucleophilic attack for phosphorylation.

$G_{\text{lid}}$  and most of  $G_{\text{omega}}$  show clear evidence of mobility (Supplementary Fig. 2.4a). The range of motion of  $G_{\text{lid}}$  is demonstrated by the *TmCphA1* structure, where crystal contacts hold  $G_{\text{lid}}$  of molecule B in a conformation rotated by  $32^\circ$ , which does not allow ATP binding (Supplementary Fig. 2.4b). However, the average  $G_{\text{lid}}$  position does not change markedly upon polymer binding to *SuCphA1*, as there is a maximal  $\sim 2 \text{ \AA}$  shift in positions with and without polymer. Three-dimensional variability analysis with CryoSPARC<sup>148</sup> reveals distinct modes of movement for both  $G_{\text{lid}}$  and  $G_{\text{omega}}$  (Supplementary Movie 3) that are likely the result of simple thermal motion. The most closed position of the G domain should place the reactive carboxylate within reaction distance of the  $\gamma$ -phosphate of ATP, bridging the  $\sim 6 \text{ \AA}$  gap we see, and allowing the G domain to catalyze its first reaction (Fig. 2.1b) to produce main-chain phosphorylated cyanophycin.

The large loop of  $G_{\text{omega}}$  is well ordered and contributes to a shallow pocket. Despite not observing aspartate in the co-complex structures (as is common in studies of glutathione synthetase<sup>7</sup>), geometry suggests this pocket is a likely binding site for aspartate, the substrate of the second G domain reaction. The large loop is highly conserved among CphA1s, and the large loop of *G. max* homogluthathione synthetase is important for selection of its amino acid substrate<sup>117</sup>. To assess the importance of the pocket formed by the large loop for activity, we introduced a bulky S396W mutation that should partially block access to this region. As expected, this mutation abolished activity (Fig. 2.3d).

#### 2.3.4. Structure and mutational analysis of the M domain

The M domain is assumed to catalyze isopeptide bond formation between the Asp side chains and Arg<sup>98,138,139</sup> (Fig. 2.1b). The active site of the M domain in *SuCphA1* is between the central body ( $M_{\text{core}}$ ; *SuCphA1* 488-723) and a large subdomain ( $M_{\text{lid}}$ ; 724-875) (Fig. 2.2d, 2.4a).



**Figure 2.4. Structure and mutagenesis of the M domain.** (a) The overall structure of the *Su*CphA1 M domain, colored by subdomains. (b) Overlay of the *Su*CphA1 M domain and MurE ligase from *M. tuberculosis*<sup>9</sup> showing the similar overall structure, including M<sub>lid</sub> in the closed conformation. (c) The structure of the *Su*CphA1 M domain with (Asp-Arg)<sub>8</sub>-Asn and ATP. The Cryo-EM map was carved 3Å around the substrates at level 5. (d) Activity assays of *Su*CphA1 M domain mutants. R561 binds the main-chain carboxylate of the Asp residue to which Arg is attached. T538, S542, and R566 bind cyanophycin β-Asp-Arg dipeptides close to the active site. Since each dipeptide is bound by several residues, mutation of T538, S542 or R566 individually to Ala does not significantly reduce activity. n=4 independent experiments. Data are presented as individual measurements and mean value, error bars represent SD values.

M<sub>core</sub> and M<sub>lid</sub> are conserved with Mur-ligases, which additionally have an N-terminal lobe not present in CphA1 (Fig. 2.4b, Supplementary Fig. 2.3c)<sup>149</sup>. M<sub>lid</sub> displays a large range of motion,

also observed in Mur ligases<sup>9,110</sup>. In the *Tm*CphA1 crystal structure, M<sub>lid</sub> is disordered in one monomer and held far from an ATP-binding conformation by crystal contacts in the other (Supplementary Fig. 2.4b). In the EM map of *Ab*CphA1, no sign of M<sub>lid</sub> is present, even at low threshold, despite the presence of ATP in the sample. M<sub>lid</sub> is resolved in EM maps of *Su*CphA1, though it appears weaker than other portions of the map.

A cryo-EM map of *Su*CphA1 incubated with Arg, ATP and a cyanophycin analog ( $\beta$ -Asp-Arg)<sub>8</sub>-Asn (Supplementary Fig. 2.3f) shows clear signal for ATP and the cyanophycin analog. The reactive end of the cyanophycin analog interacts with R561, close to ATP. It then extends away towards the other two domains (Fig. 2.4c). The dipeptide adjacent to the terminal residue makes many interactions, most notably E533, N537, T538, Q541, S542 and R566. The next dipeptide is less ordered and reaches into solvent, whereas the third is better ordered and interacts with Q541, near the N domain (Supplementary Fig. 2.3e). Thereafter, signal is visible only at low contour, and extends toward the N domain (Supplementary Fig. 2.4c). The polymer makes more extensive contact with the M domain than with the G domain. Accordingly, mutation of T538, S542 or R566 individually to alanine resulted in subtle changes in activity, consistent with redundancy in binding interactions. In contrast, mutation of R561, which interacts with the terminal Asp, to alanine had a more drastic effect on activity (Fig. 2.4c,d). R561 orients the Asp (Asn in our analog), so its reactive side chain is 4.2 Å from the ATP  $\gamma$ -phosphate, in a good pre-attack conformation. Two Mg<sup>2+</sup> coordinating the  $\beta$  and  $\gamma$ -phosphates of ATP are resolved (Fig. 2.4c). The structures and 3D variability analysis (Supplementary Movie 3) all indicate flexibility of M<sub>lid</sub>, in which, as in Mur ligases<sup>110</sup>, a closing motion is likely important for transition from the observed pre-reaction state to the phosphorylation reaction. The Arg substrate of the second reaction likely binds in the crevice between M<sub>core</sub> and M<sub>lid</sub> for isopeptide bond formation, but we could not unambiguously identify it in the maps.

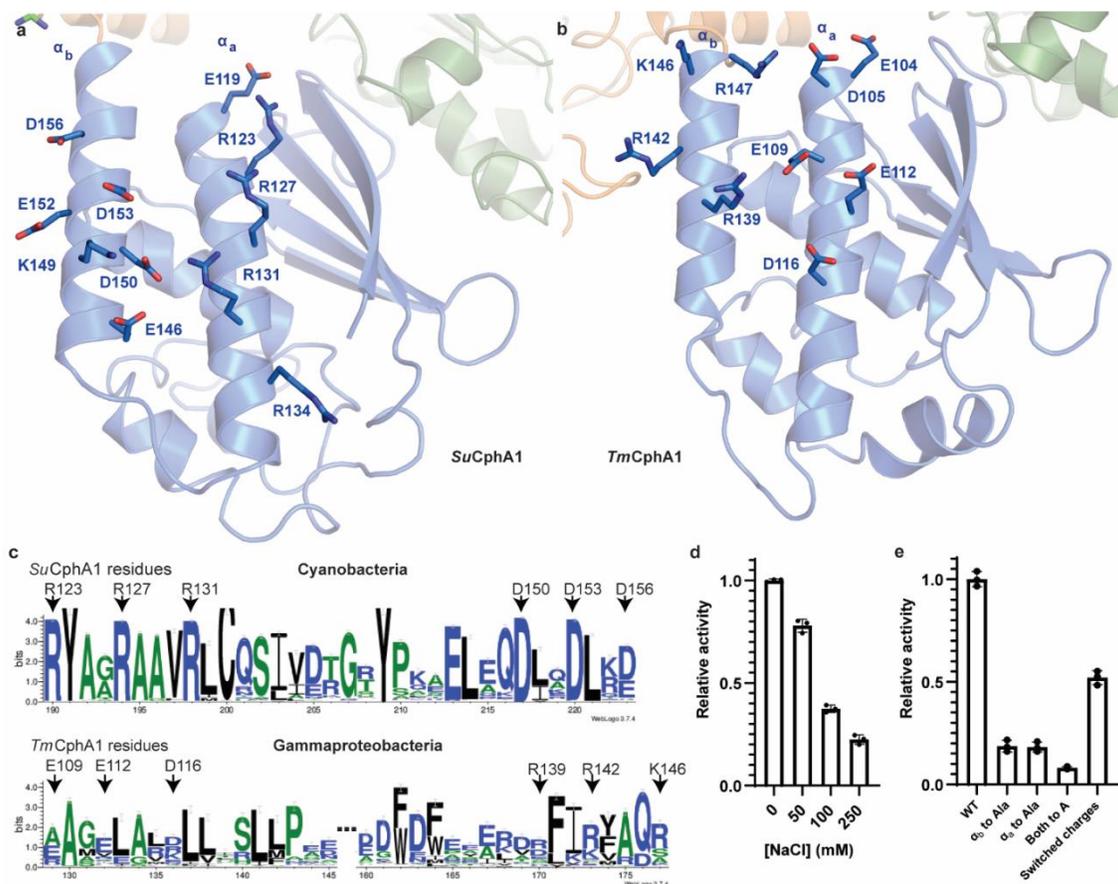
### 2.3.5. Structure and polymer binding function of the N domain

The N domain is not homologous to other proteins and was not expected to contain any catalytic residues. The CphA1 structures reveal the hitherto unknown fold of the N domain to be a 4-stranded, antiparallel  $\beta$  sheet backed by two long, antiparallel helices, a long helix at 45° to those, and two or three shorter helices (Fig. 2.5a,b). Structure similarity searches shows good matches only for the four-stranded  $\beta$  sheet and two backing helices, e.g. with parts of *E. coli* RNA-

polymerase  $\alpha$ -subunit (Supplementary Fig. 2.5a). The N domain sits between G and M domains, but only packs tightly with the M domain.

The N domain features two notable charged patches, one negative and one positive, along the two long anti-parallel  $\alpha$ -helices (Fig. 2.5a,b, Supplementary Fig. 2.4d). *Su*CphA1  $\alpha_a$  (residues 119-138) includes four surface-exposed arginines, and  $\alpha_b$  (144-161) includes six surface-exposed aspartates or glutamates (Fig. 2.5a). This trend is reversed in *Ab*CphA1 and *Tm*CphA1, where  $\alpha_a$  is more negative, with five aspartates or glutamates, and  $\alpha_b$  is more positive, with four arginines or lysines (Fig. 2.5b). In all three enzymes,  $\alpha_a$  and  $\alpha_b$  contribute two of several charged patches on the front side of CphA1, while the backside (central cavity / central channel of the tetramers) is far less charged (Supplementary Fig. 2.4d). This pattern suggests that these charged patches, including helices  $\alpha_a$  and  $\alpha_b$ , could be involved in cyanophycin binding. Indeed, comparison of the unsharpened EM maps of *Su*CphA1 without a cyanophycin analog to those with  $(\beta\text{-Asp-Arg})_8\text{-Asn}$  or  $(\beta\text{-Asp-Arg})_8\text{-NH}_2$  clearly shows additional features when a cyanophycin analog is present (Supplementary Fig. 2.4c). The map features are ill-defined and likely represent an ensemble of cyanophycin polymer. Lowering the contour level shows that the features lead from the N domain either directly to the M domain active site or toward a negatively charged patch on the back of  $G_{\text{omega}}$  near the G domain active site (Supplementary Movie 3). Cryo-EM maps of *Su*CphA1 in the presence of  $(\beta\text{-Asp-Arg})_{16}$  are nearly identical to that of the  $(\beta\text{-Asp-Arg})_8\text{-NH}_2$  complex, suggesting that analogs of 8 or more dipeptides should be representative of how cyanophycin binds to CphA1. Importantly, signal in the EM maps representing binding to  $\alpha_a$  and  $\alpha_b$  is repeatedly observed.

The structures imply that cyanophycin binds CphA1 through loose anchoring to the N domain via salt bridges with  $\alpha_a$  and  $\alpha_b$ . We evaluated this by altering ionic conditions and by mutagenesis. *Su*CphA1 shows a clear decrease in activity with increasing ionic strength, consistent with ionic interaction (Fig. 2.5d). Mutagenesis of charged residues on  $\alpha_a$  and  $\alpha_b$  also support this binding mode: The triple  $\alpha_a$  mutant R123A-R127A-R131A and quadruple  $\alpha_b$  mutant D150A-E152A-D153A-D156A each display decreased activity, and the combined  $\alpha_a$ - $\alpha_b$  septuple alanine mutations further reduced cyanophycin synthesis (Fig. 2.5e). However, an  $\alpha_a$ - $\alpha_b$  septuple charge-swap mutant (R123E-R127E-R131E-D150R-E152K-D153R-D156R) restores 50% of WT activity. Ser mutants of the equivalent residues of *Tm*CphA1 displayed similar results (Supplementary Fig. 2.5c), and DSF of all mutants showed them to have a  $T_m$  similar to that of



**Figure 2.5. Structure, conservation and mutagenesis of the N domain.** (a,b) The charged residues on  $\alpha_a$  and  $\alpha_b$  of *Su*CphA1 (a) and *Tm*CphA1 (b) form patches of positive and negative charges on their surface. (c) Weblogo<sup>8</sup> analysis of the region covering  $\alpha_a$  and  $\alpha_b$  of *Su*CphA1 (top) and *Tm*CphA1 (bottom). While the distribution of charged residues is conserved in cyanobacterial CphA1 enzymes, gammaproteobacterial sequences show high variability at the equivalent positions. (d) *Su*CphA1 activity decreases with increasing sodium chloride concentration in the reaction buffer, consistent with cyanophycin binding the CphA1 through salt bridges with charged residues.  $n=4$  independent experiments. Data are presented as individual measurements and mean value, error bars represent SD values. (e) Activity assays of *Su*CphA1 with mutation of the charged residues on  $\alpha_a$  (R123A-R127A-R131A) and on  $\alpha_b$  (D150A-E152A-D153A-D156A). Mutation of both helices together (R123A-R127A-R131A-D150A-E152A-D153A-D156A) resulted in decreased activity compared to mutation of either alone, while reversal of the charges on both helices (R123D-R127D-R131D-D150R-E152R-D153R-D156R) restored activity to 50%.  $n=3$  independent experiments. Data are presented as individual measurements and mean value, error bars represent SD values.

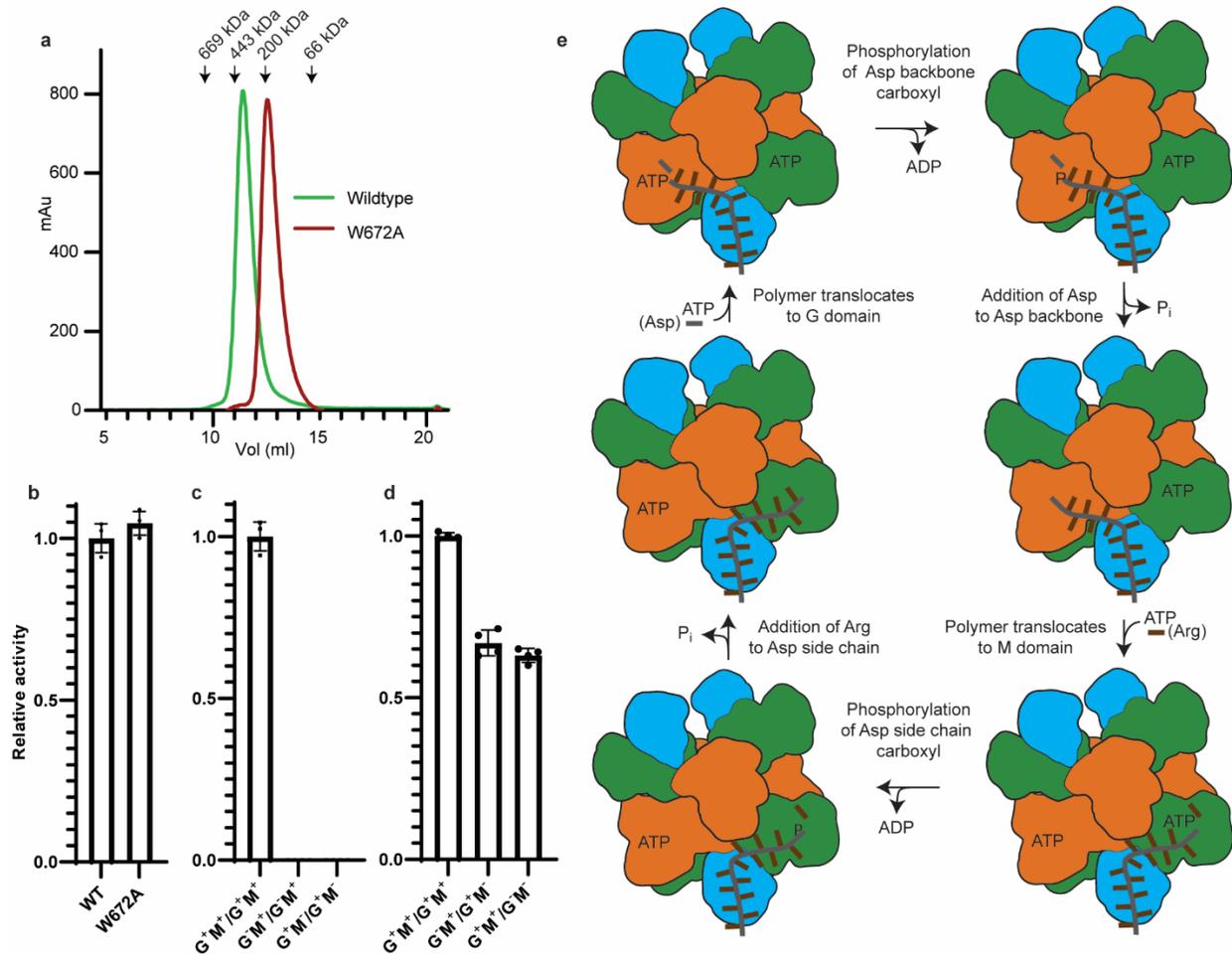
WT enzymes (Supplementary Fig. 2.5d). The clear effect of mutating residues located so far from the active sites ( $\sim 32 - 62$  Å) strongly suggests that loose, N-domain anchoring is a key contributing feature of efficient cyanophycin synthesis.

### 2.3.6. The reaction pathway for cyanophycin synthesis

The CphA1 structures allow experiments to provide direct insights into how the different active sites cooperate to produce cyanophycin. Tetrameric CphA1 contains eight active sites (4 G domain, 4 M domain), as well as four  $\alpha_a/\alpha_b$  helix pairs important for cyanophycin binding. In a CphA1 monomer, the G and M active sites are  $\sim 60$  Å apart, with the  $\alpha_a/\alpha_b$  helices completing a functional G-M-N triangle. In a CphA1 dimer, the length of the unobstructed path between a G domain of one monomer and the M domain of the other is comparable to the distance between those domains within a single monomer. In contrast, although the G domain active site is  $\sim 60$  Å and  $\sim 80$  Å (in *Su*CphA) or  $\sim 70$  Å and  $\sim 90$  Å (in *Tm*CphA1) away from the two M domain active sites of the other dimer within the tetramer, they are on the opposite sides of assemblies, meaning the unobstructed paths are much, much longer ( $>130$  Å). This suggests it would be simplest for a single cyanophycin polymer to be iteratively extended and decorated in the active sites of a single monomer or those within the dimer.

To investigate how active sites coordinate cyanophycin synthesis, we compared the activity of tetrameric CphA1 with an enforced dimeric CphA1. Because the tetramer interface of *Su*CphA1 appeared completely reliant on W672 (Supplementary Fig. 2.2e, 2.6a), we mutated it to alanine, and indeed saw *Su*CphA1(W672A) to be dimeric (Fig. 2.6a). Interestingly, dimeric *Su*CphA1 displayed very similar cyanophycin synthesis activity in *in vitro* assays (Fig. 2.6b), indicating that tetramerization does not impart an obvious catalytic advantage, at least *in vitro*.

Active dimeric *Su*CphA1 allows examination of whether the two active sites within a single monomer are responsible for iteratively synthesizing a particular cyanophycin polymer chain (“monomer-peptide exclusivity”), or whether the G domain of one monomer can alternate action with the either M domain in the dimer (“monomer-peptide promiscuity”). We constructed inactivating mutations for each active site: H267A for the G domain (G-) and D585A-H586A for the M domain (M-) (Fig. 2.6c, Supplementary Fig. 2.6b). These mutations were introduced into expression vectors featuring either a poly-histidine or a calmodulin binding peptide tag. Co-expression of *Su*CphA1(W672A) from both vectors in the same *E. coli* cells, and sequential nickel affinity and calmodulin affinity chromatography, allows specific purification of a heterodimer



**Figure 2.6. Dimeric *Su*CphA1 mutants and model of cyanophycin synthesis.** (a) Gel filtration chromatograms of WT and W672A *Su*CphA1 show that W672A converts *Su*CphA1 to a dimer in solution. (b) WT and W672A *Su*CphA1 displayed similar activity, calculated as described in the Methods section. (c) Dimeric *Su*CphA1 with both constituent monomer harbouring either a G domain active site mutation (H267A =  $G^-$ ) or M domain active site mutations (D585A-H586A =  $M^-$ ) are completely inactive. (d) Dimeric *Su*CphA1 which contained one native G domain active site and one native M domain active site retained over half its activity, and had comparable activity independent of whether the active site mutations were in the same monomer ( $G^+M^+/G^-M^-$ ), or spread between the two monomers ( $G^-M^+/G^+M^-$ ).  $n=4$  independent experiments. Data are presented as individual measurements and mean value, error bars represent SD values. (e) Proposed model of cyanophycin synthesis by CphA1: The windshield wiper model model of elongation of cyanophycin is consistent with all available data. See Supplementary Fig. 2.8 for the schematic of cyanophycin synthesis by  $G^-M^+/G^+M^-$ .

comprised of one monomer encoded by each plasmid. We assessed combinations of dimeric *Su*CphA1(W672A): G<sup>+</sup>M<sup>+</sup>/G<sup>+</sup>M<sup>+</sup>; G<sup>+</sup>M<sup>+</sup>/G<sup>-</sup>M<sup>-</sup>; G<sup>-</sup>M<sup>+</sup>/G<sup>+</sup>M<sup>-</sup>, as well as G<sup>-</sup>M<sup>+</sup>/G<sup>-</sup>M<sup>+</sup> and G<sup>+</sup>M<sup>-</sup>/G<sup>+</sup>M<sup>-</sup> negative controls (Fig. 2.6c,d). As expected, G<sup>+</sup>M<sup>+</sup>/G<sup>-</sup>M<sup>-</sup> and G<sup>-</sup>M<sup>+</sup>/G<sup>+</sup>M<sup>-</sup> have reduced activity compared to G<sup>+</sup>M<sup>+</sup>/G<sup>+</sup>M<sup>+</sup>, because of the reduced number of wildtype active sites. Their activity is somewhat higher than 50%, perhaps because of advantage gained by a second N domain maintaining higher local concentration of cyanophycin. Notably, there was not a substantial difference in activity between the dimers with inactivating mutations in both active sites of one monomer (G<sup>+</sup>M<sup>+</sup>/G<sup>-</sup>M<sup>-</sup>) compared to inactivating mutations of one active site in each monomer (G<sup>-</sup>M<sup>+</sup>/G<sup>+</sup>M<sup>-</sup>) (Fig. 2.6d). Altering ionic strength did not change this result (Supplementary Fig. 2.6c). These data clearly indicate that synthesis with monomer-peptide exclusivity and promiscuity are both possible and comparable in catalytic efficiency within the dimer. Furthermore, this result combined with the comparable activities of dimer and tetramer suggests dimer-peptide exclusivity in cyanophycin synthesis.

## **2.4. Discussion**

Structures of three cyanophycin synthetases reveal two distinct, elegant architectures. The sphere or ring shapes are created by the core, immobile portions of the domains (G<sub>core</sub>/G<sub>alpha</sub>, M<sub>core</sub>), with spiky projections formed by the N domain and mobile subdomains adjacent to the active sites (G<sub>lid</sub>, G<sub>omega</sub>, M<sub>lid</sub>). CphA1s throughout phylogeny share all of these elements. They diverge by up to ~35% sequence identity and have modest changes in size other than a variable, dispensable C-terminus extension of up to ~100 residues. Removal of this region from *Ab*CphA1 and *N. ellipsoforum* CphA1 increased thermal stability and activity *in vivo*<sup>88,108</sup>. *Su*CphA lacks this extension, while in *Tm*CphA1, it is ~40 residues, but not visible in maps. Our structures explain why the truncation of the *N. ellipsoforum* CphA1 by 31 residues did not inhibit activity, while truncation by 59 residues led to complete inactivation<sup>88</sup>: The former removes only the variable C-terminus while the latter also removes part of M<sub>lid</sub>, including the central  $\beta$ -strand, undoubtedly resulting in improper folding.

In evolving cyanophycin synthetase, nature has elegantly co-opted and fused two enzymes which perform the same basic amide bond forming reaction and repurposed them for amino acid polymerization. The binding mode of cyanophycin to the G domain can explain two fundamental properties of cyanophycin synthesis: primer dependence and lack of poly-Asp polymerase activity.

Clear signal for three ordered dipeptides in the active site means that the ideal primer would be at least 3-4 dipeptides long to allow for the strongest binding. This is consistent with the report that  $(\beta\text{-Asp-Arg})_3$  is a suitable primer<sup>1</sup>. In contrast, a poly-Asp should be a poor substrate for elongation since most interactions between the G domain and cyanophycin involve the arginine appendages, lacking in poly-Asp. In the M domain, extensive hydrogen bonding with the dipeptide adjacent to the reactive Asp residue, as well as a more distal dipeptide residue, further underscores primer dependence. The first reaction to make cyanophycin from free amino acids would require phosphorylation of the  $\beta$ -carboxylate of Asp by the M domain, but the structures clearly indicate that free Asp would make only a small fraction of the observed interactions for  $(\beta\text{-Asp-Arg})_3\text{-Asp}$ .

The N domain is central to CphA1 function, literally and figuratively, being physically between G and M domains and key for cyanophycin synthesis. Although not possessing catalytic activity itself, its cyanophycin binding role allows the enzyme to combine the G and M domains activities. The N domain binds the growing cyanophycin polymer through electrostatic interactions and acts as a soft anchor-point to help feed the growing end into the catalytic sites: The ill-defined polymer density above the positive and negative patches along  $\alpha_a$  and  $\alpha_b$  and the activity are consistent with an ensemble of overlapping registers, which would be advantageous to allow sliding during polymer growth and movement of the C-terminus between active sites. Modelling of cyanophycin polymer with  $\beta$  strand backbone angles and Asp side chain  $\chi_2$  angles of  $\sim 150^\circ$  positions the positive Arg guanidiniums in alignment with  $\alpha_b$  and the negative Arg  $\alpha$ -carboxylates in alignment with  $\alpha_a$  (Supplementary Fig. 2.5b). Switching the  $\chi_2$  angles to  $\sim 30^\circ$  allows cyanophycin's positive charges to interact with  $\alpha_a$  and its negative charges with  $\alpha_b$ . This would enable CphA1s with either helix-charge pattern (Fig. 2.5a,b), or a mixed pattern, to interact with cyanophycin. This plasticity also results in cryptic conservation of these patches. In cyanobacteria,  $\alpha_a$  is largely positive and  $\alpha_b$  largely negative (Fig. 2.5a,c), but this pattern is not conserved among other groups. In *TmCphA1* and *AbCphA1*,  $\alpha_a$  is largely negative and  $\alpha_b$  positive (Fig. 2.5b), and other gammaproteobacterial CphA1s display other charge distribution patterns (Fig. 2.5c). Thus, no conservation is shown on alignment of all CphA1 sequences, so bioinformatics did not reveal the importance of this region.

The structures, mutagenesis and previous data come together to support an overall model of cyanophycin synthesis (Supplementary Movie 4): In early steps of synthesis, primers and short strands of cyanophycin must diffuse randomly between active sites, relying on the specific binding

interactions with G and M domains such as those observed with *Su*CphA1. Once the polymer is of sufficient length, its soft anchoring with the N domain would keep it engaged with CphA1, but allow sliding. Sliding could enable iterative insertion of the polymer's C-terminus into the two different active sites, for processive cyanophycin synthesis via a “windshield wiper”-like movement between the G and M domains (Fig. 2.6e, Supplementary Fig. 2.8, Supplementary Movie 4). The sigmoidal shape of cyanophycin synthesis is consistent with distinct initiation and elongation phases of synthesis (Fig. 2.2a). A cyanophycin molecule anchored to a particular N domain would be able to access the G and M domains in the same polypeptide chain as well as the M domain from the other subunit of the dimer (Supplementary Fig. 2.8). However, the position of the N domain in the tetramers (Fig. 2.2d) seemingly precludes interactions between the growing peptide and other active sites in the tetramer, consistent with the observation that the *Su*CphA1(W672A) dimer is as active as the wildtype tetramer (Fig. 2.6b). The mechanism of polymer length determination and termination is subject of ongoing study. The model, in combination with the knowledge gained on precise substrate binding, active site conservation, primer dependence and overall architecture, provides a greater understanding of cyanophycin synthesis.

## **2.5. Methods**

### **2.5.1. Cloning, protein expression and protein purification**

Fourteen CphA1 genes from 5 different phyla were cloned: 6 from firmicutes (*D. hafniense* DSM10664, *S. thermosulfidooxidans* DSM9293, *A. californiensis* DSM14826, *C. acetigignens* DSM18802, and two homologs from *P. cellulosolvens* DSM2933), 4 from cyanobacteria (*T. elongates* BP-1, *Synechococcus* sp. MA-19, *Synechocystis* sp. UTEX2470, *Anabaena* sp. UTEX2576), 2 from gammaproteobacteria (*A. baylyi* DSM587, *T. morbirosei* DSM23827), 1 from betaproteobacteria (*B. cepacia* DSM7288), and 1 from alphaproteobacteria (*P. soli* DSM 19599). Genes were inserted into pJ411-derived plasmids and small scale expression trials were performed with each. *E. coli* BL21(DE3) or *E. coli* BL21(DE3) Rosetta2 cells harboring these plasmids were grown in LB or TB media supplemented with 100ug/ml kanamycin at 37 °C until they reached an OD<sub>600</sub> of ~0.5, at which time protein expression was induced with 0.1 – 1 mM isopropyl β-d-1-thiogalactopyranoside (IPTG) and the growth temperature was shifted to between 16 and 37 °C and incubated for a further ~4 - 20 hours before harvesting. Cells were lysed by repeated freeze-

thaw, separated into soluble and insoluble fractions by centrifugation and analysed by SDS-PAGE. Only *Su*CphA1, *Ab*CphA1 and *Tm*CphA1 gave robust soluble expression.

The genes encoding *Su*CphA1 (from genome CP007542.1, encoding protein WP\_028947105.1) and *Ab*CphA1 CphA1 (from genome CR543861.1, encoding protein WP\_004925893.1) were cloned from genomic DNA (purchased from University of Texas (UTEX) and DSMZ culture collections, respectively), and the gene encoding *Tm*CphA1 (WP\_038021094.1) was codon optimized for expression in *E. coli* and synthesized by the US Department of Energy Joint Genome Institute. Genes were inserted into pJ411-derived plasmids encoding C-terminal tobacco etch virus (TEV) protease recognition sites and an octahistidine affinity or calmodulin binding peptide sequence. All cloning and mutagenesis were performed by transforming DH5- $\alpha$  *E. coli* cells with PCR fragments containing overlapping ends. Proteins were heterologously expressed in *E. coli* BL21(DE3) (*Ab*CphA1, *Tm*CphA1) or *E. coli* BL21(DE3) Rosetta2 (*Su*CphA1). Cells were grown in LB media supplemented with 100ug/ml kanamycin (and 25ug/ml chloramphenicol in the case of *Su*CphA1) at 37 °C until they reached an OD<sub>600</sub> of ~0.5, at which time protein expression was induced with 0.5mM (*Ab*CphA1, *Tm*CphA1) or 0.2mM (*Su*CphA1) IPTG and the growth temperature was shifted to 22 °C and incubated for a further ~20 hours before harvesting. All protein purification steps were carried out at 4°C. After centrifugation, the cells were resuspended in buffer A (250mM NaCl, 50mM Tris pH8, 10mM imidazole, 2mM  $\beta$ -mercaptoethanol) supplemented with a few crystals of lysozyme, lysed by sonication and the lysate was clarified by centrifugation at 40,000xg. The lysate was then loaded onto a HisTrap HP column (Cytiva), washed extensively with buffer B (buffer A with 30mM imidazole) and eluted with buffer C (buffer A with 250mM imidazole). In the case of *Tm*CphA1 and *Su*CphA1, protein was incubated with TEV protease for removal of the octahistidine tag while being dialyzed overnight against buffer D (250mM NaCl, 20mM Tris pH 8, 5mM  $\beta$ -mercaptoethanol) prior to application to a HisTrap column. The flow through was collected and loaded onto a MonoQ 16/10 column (GE Healthcare) equilibrated in buffer E (100mM NaCl, 20mM Tris pH 8, 5mM  $\beta$ -mercaptoethanol), washed with several column volumes of buffer E, then eluted using a NaCl gradient of 100-500mM over 160ml. Pooled, purified sample was concentrated and applied to a Superdex200 16/60 column (GE Healthcare) equilibrated in buffer F (100mM NaCl, 20mM Tris pH8, 1mM dithiothreitol). The tag of *Ab*CphA1 was not cleaved, and following elution from the HisTrap column the protein was concentrated and loaded onto a Superdex200 16/60 column

equilibrated in buffer G (500mM NaCl, 20mM Tris pH8, 1mM dithiothreitol). Following gel filtration, fractions with the highest purity were pooled and concentrated to 12mg/ml by 30 kDa molecular weight cut off Amicon centrifugation concentrators (EMD Millipore). Glycerol was added to a final of 10% and sample was flash frozen and stored at -80 °C until use.

Selenomethionine-labeled *TmCphA1* was expressed in *E. coli* B834(DE3) in SelenoMet medium (Molecular Dimensions) supplemented with 50mg/L selenomethionine. Cell growth and protein purification procedures as described above.

For dimer mutant-combination assays, *E. coli* BL21(DE3) Rosetta2 cells were co-transformed with plasmid pCDF-UTEX2470-CphA1-CBP (harbouring spectinomycin resistance) and plasmid pBacIT-UTEX2470-CphA1-8xHis (harbouring kanamycin resistance). Cells were grown in LB media supplemented with 100 µg/ml kanamycin, 25 µg/ml chloramphenicol and 50 µg/ml spectinomycin until reaching an OD<sub>600</sub> of ~0.5, at which point protein expression was induced by addition of 0.2mM IPTG. Growth temperature was shifted to 22°C and culture was grown for an additional 40 hours prior to harvesting. Cells were lysed and nickel affinity chromatography was performed as described above. Pooled fractions were mixed with CaCl<sub>2</sub> to a final concentration of 2mM and applied to a column of calmodulin-sepharose (Agilent) equilibrated with buffer H (250mM NaCl, 50mM Tris pH 8, 2mM CaCl<sub>2</sub>, 2mM β-mercaptoethanol), washed with buffer H and eluted with buffer I (250mM NaCl, 50mM Tris pH8, 2mM EGTA, 2mM β-mercaptoethanol). Fractions with the highest purity were pooled and concentrated to 12mg/ml by 30 kDa molecular weight cut off Amicon centrifugation concentrators (EMD Millipore). Glycerol was added to a final of 10% and sample was flash frozen and stored at -80 °C until use.

### **2.5.2. Crystallography of *TmCphA1***

Selenomethionine-labelled *TmCphA1* was crystallized using the sitting drop vapour diffusion method. Drop solution of 2µl of *TmCphA1* at 6mg/ml in buffer F were mixed with 2µl of well solution (13.25% PEG3350, 320mM sodium formate, 1% glycerol, 100mM sodium/potassium phosphate pH6.8) and was equilibrated against a reservoir of 400µl of well solution at 4 °C. After three weeks, crystals reached their full size and were dehydrated by replacing the well solution with a dehydration solution of 20% PEG3350, 320mM sodium formate, 16% glycerol, 100mM sodium/potassium phosphate pH 6.8 and equilibrating for 24 hours. Crystals were looped and flash vitrified in liquid nitrogen, and diffraction data was collected on

APS beamline 24-ID-E. Diffraction data were collected using RAPD and indexed using DIALS<sup>150</sup> and then data from 6 crystals were analyzed using BLEND<sup>151</sup> and scaled and merged together using AIMLESS. The structure was solved in CCP4i2 using a combination of single wavelength anomalous dispersion and molecule replacement using a pseudoatom representation of the cryo-EM map of *AbCphA1* and models of residues 1-150 and 715-850 generated by Rosetta<sup>152</sup> as search models. The model was manually re-built and completed in Coot<sup>153</sup> and refined using REFMAC<sup>154</sup>, LORESTR and Rosetta. Crystallography data statistics are listed in Supplementary Table 2.1.

### 2.5.3. Cryo-EM sample preparation and data collection

*SuCphA1* in buffer F was mixed with substrate to give the following 4 samples: (1) No polymer sample - 2mg/ml *SuCphA1*, 10mM MgCl<sub>2</sub>, 2mM ATP, 20mM Asp and 20mM Arg (dataset,); (2) G domain substrate analog sample - 2mg/ml *SuCphA1*, 10mM MgCl<sub>2</sub>, 2mM AMPPCP, 20mM Asp, and 5mM (β-Asp-Arg)<sub>8</sub>-NH<sub>2</sub>; (3) Long G domain substrate analog sample - 3.5 mg/ml *SuCphA*, 10mM MgCl<sub>2</sub>, 2mM ATP, and 1mM (β-Asp-Arg)<sub>16</sub>-OH; and (4) M domain substrate analog sample - 3.5 mg/ml *SuCphA*, 10mM MgCl<sub>2</sub>, 2mM ATP, 20mM Arg, and 5mM (β-Asp-Arg)<sub>8</sub>-Asn. Octyl β-D-glucopyranoside was added directly before vitrification to a final concentration of 0.09%. For vitrification, 3μl of protein sample was applied to glow discharged C-flat 200 or 300 mesh 1.2/1.3 Cu holey carbon grids, blotted for 2-3 seconds at 4°C and 90% humidity using a Vitrobot IV (FEI) and plunge-frozen into liquid ethane. Data were collected at the McGill Facility for EM Research (FEMR) using an FEI Titan Krios TEM operating at 300kV with a Gatan K3 DED and a Gatan GIF BioQuantum LS. Movies were collected in counting mode using SerialEM, with a total dose of 55-65e/Å<sup>2</sup> and defocus range of -0.75 to -2.5μm at a nominal magnification of 105,000, resulting in a pixel size of 0.855Å<sup>2</sup>. For *AbCphA1*, protein in buffer G was mixed with 10mM MgCl<sub>2</sub>, 20mM KCl, and 2mM ATP, final protein concentration 0.42mg/ml. Samples of 3μl were applied to grids and blotted in the same way as *SuCphA1*. Data was collected at the University of California, San Diego using a Talos Arctica TEM operating at 200kV with a Gatan K2 Summit DED. Movies of 60 frames were collected at super-resolution mode with a total dose of 57e/Å<sup>2</sup> at a nominal magnification of 30 thousand resulting in an unbinned pixel size of 0.58Å<sup>2</sup>. Data collection details are listed in Supplementary Table 2.2.

### 2.5.4. Cryo-EM data processing

*SuCphA1* micrographs were motion corrected using Relion3.1<sup>155</sup>. The micrographs were imported to CryoSPARC2<sup>148</sup> for patch-CTF estimation and particle picking. One thousand

particles were manually picked and subjected to 2D classification in order to generate templates for auto-picking. After picking particles from all good micrographs, particles were extracted using a box size of 400 pixels and several rounds of 2D classification and one round of 3D classification were performed to remove undesirable particles. The resulting particle set was used to generate an initial model using *ab-initio* reconstruction, and a map was calculated using homogenous refinement with per-particle defocus and high-order CTF parameters optimization. The particles were then exported to Relion3.1 for two rounds of Bayesian polishing, and the polished particles were used to generate the final reconstruction using CryoSPARC2. Local resolution estimation followed by local filtering were then performed in CryoSPARC2, and the locally filtered maps were used for model building. *AbCphA1* data micrographs were processed in CryoSPARC2 unless otherwise stated. Patch motion corrected micrographs were CTF estimated using GCTF<sup>156</sup>. Particles were picked and extracted with a box size of 720 pixels and binned by 2, resulting in a pixel size of 1.16 Å. Several rounds of 2D classification were performed to remove junk particles. An initial model generated in using *ab initio* reconstruction, and 3D reconstruction was then performed using non-uniform refinement in CryoSPARC3.

Conformational heterogeneity was analyzed using 3D variability analysis in CryoSPARC2. Particles were first down-sampled to 200 pixels and symmetry expanded. The analysis was performed with a mask around one monomer with a 4 Å low-pass filter applied. The resulting reconstructions used for movies were generated using 3D variability display in intermediates mode, with 10 frames, min/max percentile of 3%, filter resolution of 4 Å, and real-space cropping to 160 pixels.

### **2.5.5. Model building and refinement into cryo-EM maps**

The map of *Su*CphA1 with only ATP was used to build an initial model using Buccaneer implemented in CCP-EM 1.4, followed by manual model building in Coot. Since signal for the  $M_{lid}$  was not of sufficient quality for *ab initio* model building, the structure of this lobe from the *Tm*CphA1 crystals structure was used as an initial model. Several rounds of refinement using Rosetta followed by manual fitting in Coot were performed, assisted by symmetry expansion and model validation in CCP-EM. Signal for  $M_{lid}$  of *Su*CphA1 was better in the no-polymer dataset, and so that map was used as an initial model for the other *Su*CphA1 maps. Each model was refined into its own map using Rosetta and manually modelled in Coot. Finally, ligands were fitted manually in Coot. The structures were separately validated using Molprobit. All models and

conformational constraints of substrates were generated using eLBOW5 as implemented in Phenix. Figures were generated using Pymol and UCSF Chimera.

### **2.5.6. CphA1 activity assays**

Reactions contained 20µg purified CphA1, 100mM HEPES pH8.2, 20mM KCl, 10mM MgCl<sub>2</sub>, 2mM each of L-Asp and L-Arg, 4mM ATP, and 50uM synthetic cyanophycin 12mer as primer. Sodium chloride was also added in some experiments as indicated. The reactions were carried out in triplicates or quadruplets at 23 °C, in 96-well plates with a total reaction volume of 100µl. Optical density at 600 nm was monitored using a SpectraMax Paradigm spectrophotometer (Molecular Devices), with 5 second linear shaking between reads. Typical reaction times were 30-60 minutes. Data were analyzed using GraphPad Prism. To calculate activity rates, the maximum of the first derivative of each OD<sub>600</sub> curve was taken. The derivatives curves were smoothed with a 2<sup>nd</sup> order polynomial in order to reduce noise in measurements. The lag phases of each reaction were not considered in this analysis, because they represent pre-steady state. The values displayed in the graphs are the mean maximal values of the first derivatives of all replicates normalized to the WT mean value, and the error bars represent the standard deviation of the mean. A standard curve was used to determine the dependence of OD<sub>600</sub> on cyanophycin concentration, allowing us to determine specific activity in comparable units to those used by previous studies<sup>3</sup>.

### **2.5.7. Protein phylogenetic tree generation**

A P-BLAST search was performed using *Su*CphA1 as a subject. Only sequences with at least 40% identity and 85% coverage were considered as hits. The resulting list of protein sequences was aligned using ClustalW. The phylogenetic tree was calculated using IQ-TREE<sup>157</sup> and displayed and manually annotated using iTOL<sup>158</sup>.

### **2.5.8. Synthesis of cyanophycin segments**

Solid phase synthesis was used for the synthesis of all molecules using Fmoc-(β-Asp-Arg)(OtBu)-OH as building blocks in a manner similar to that previously described<sup>65,159</sup>. Full synthesis procedures are detailed in Supplementary note.

### **2.5.9. Differential Scanning Fluorimetry**

DSF assays were performed with 0.5mg/ml protein in a buffer containing 50mM HEPES pH 8.2, 100mM NaCl, 1mM DTT and 5x Sypro<sup>TM</sup> Orange in a total reaction volume of 20ul. The temperature was ramped from 5° C to 95° C over 2 hours and readings taken using a One Step Plus RT-PCR (Applied Biosystems).

## **2.6. Data availability**

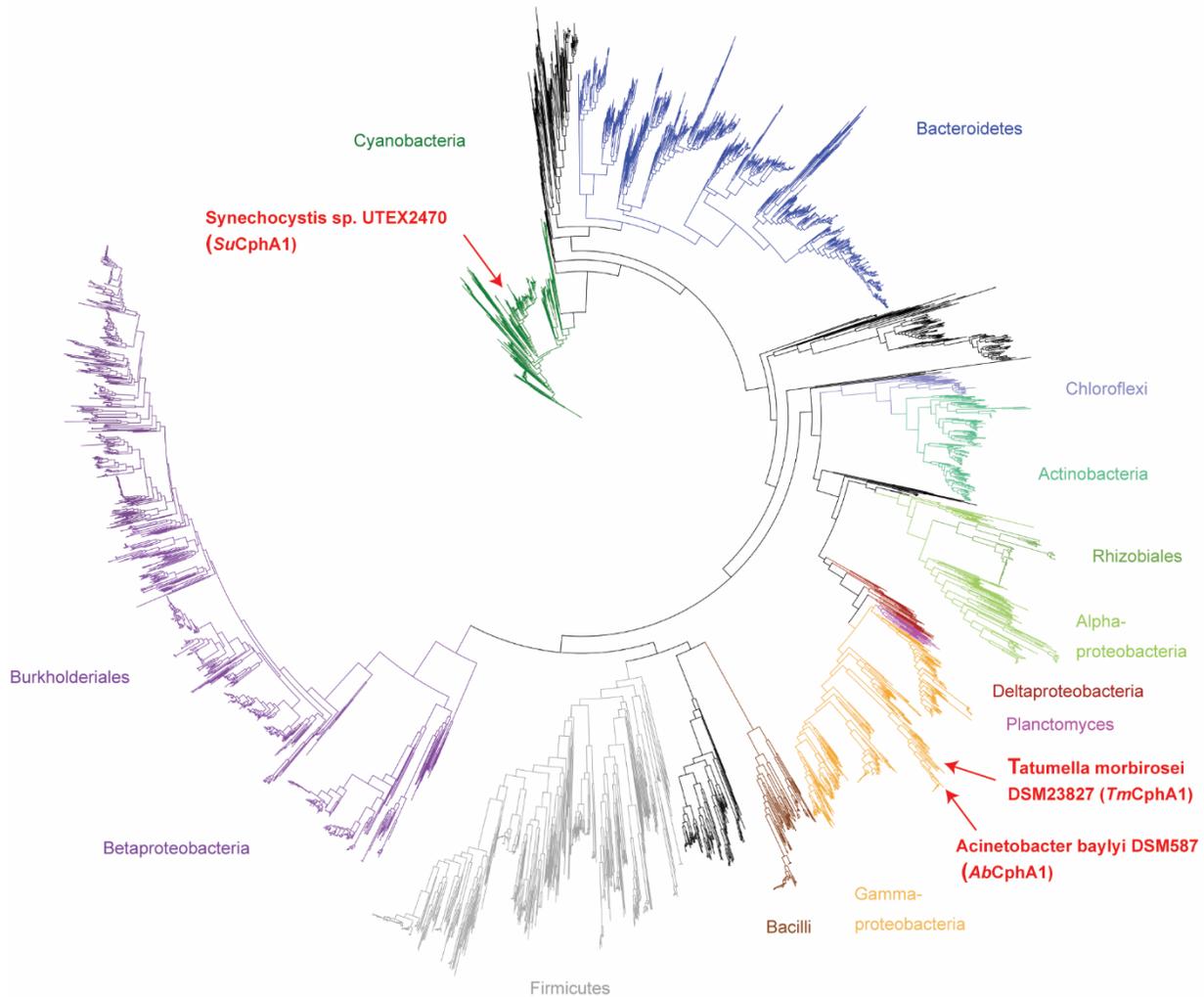
The cryo-EM maps created in this study have been deposited to the EMDB: *Su*CphA1 bound with ATP (EMD-23311), *Su*CphA1 bound with ADPCP and ( $\beta$ -Asp-Arg)<sub>8</sub>-NH<sub>2</sub> (EMD-23325), *Su*CphA1 bound with ATP and ( $\beta$ -Asp-Arg)<sub>8</sub>-Asn (EMD-23328), *Su*CphA1 with ATP and ( $\beta$ -Asp-Arg)<sub>16</sub> (EMD-23326), and *Ab*CphA1 with ATP (EMD-23327). The protein structures solved in this study have been deposited to the PDB: *Su*CphA1 with ATP (7LG5), *Su*CphA1 with ADPCP and ( $\beta$ -Asp-Arg)<sub>8</sub>-NH<sub>2</sub> (7LGJ), *Su*CphA1 with ATP and ( $\beta$ -Asp-Arg)<sub>8</sub>-Asn (7LGQ), *Su*CphA1 with ATP, *Ab*CphA1 with ATP (7LGM), and *Tm*CphA1 (7LGN).

## **2.7. Acknowledgements**

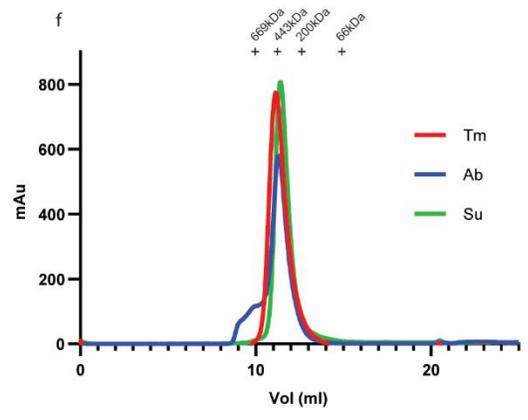
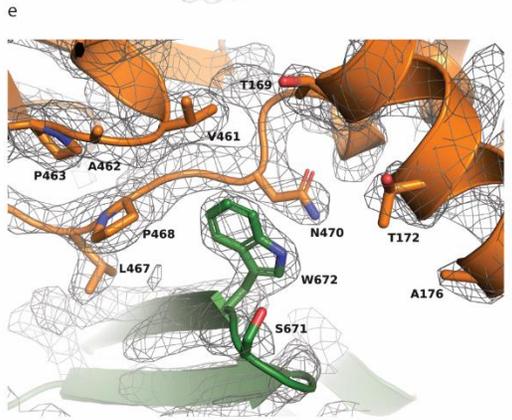
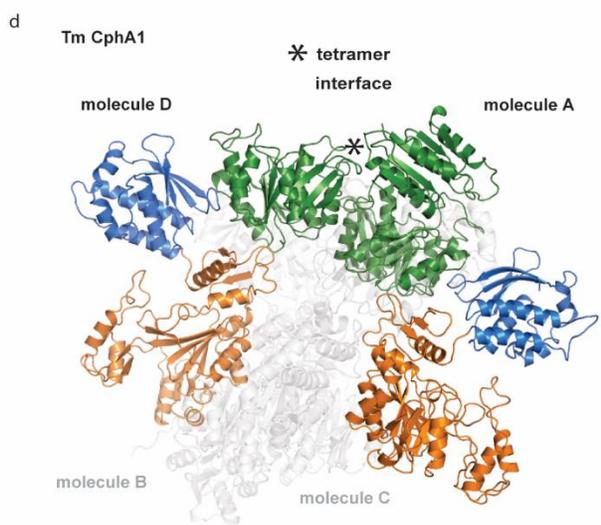
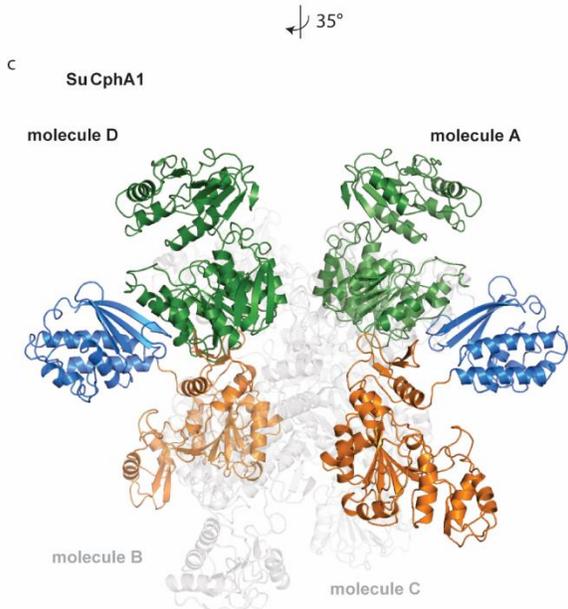
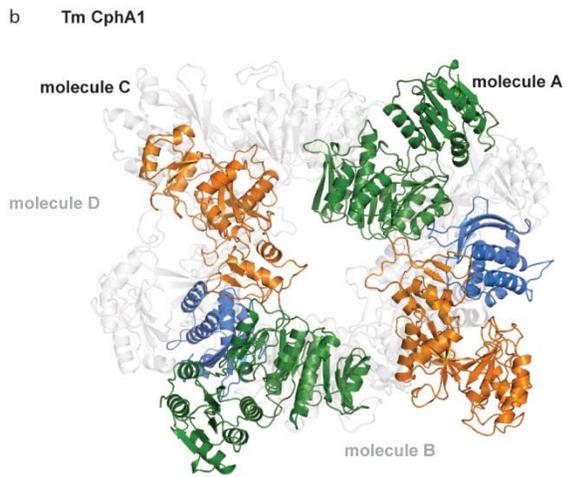
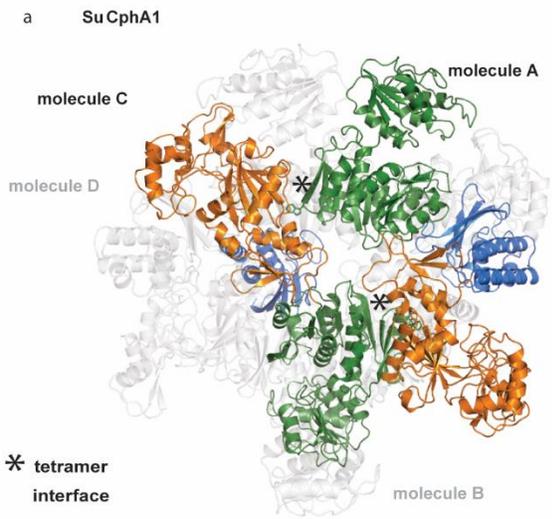
We thank members of the Schmeing lab for helpful advice and discussion, Nancy Rogerson for proofreading, staff at McGill Facility of EM Research (Mike Strauss, Kaustuv Basu and Kelly Sears) and APS (Frank Murphy and Surajit Banarjee) for support during data collection. We thank the UC San Diego Cryo-Electron Microscopy Facility, which was supported in part by NIH grants to Dr. Timothy S. Baker and a gift from the Agouron Institute to UCSD. This work was supported by a Canada Research Chair and NSERC Discovery Grant 418420 to TMS, and by funding from the Swiss National Science Foundation and the ETH Zurich to DH. We thank Paul Emsley and Rob Nicholls for help with restraints file generations. This work includes research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the National Institutes of Health (P30 GM124165). The Eiger 16M detector on the 24-ID-E beam line is funded by a NIH-ORIP HEI grant (S10OD021527). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. Gene synthesis of *Tm*CphA1 was conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, which is supported under Contract No. DE-AC02-05CH11231, as part of JGI Grant 503632 to TMS.

## 2.8. Supplementary information

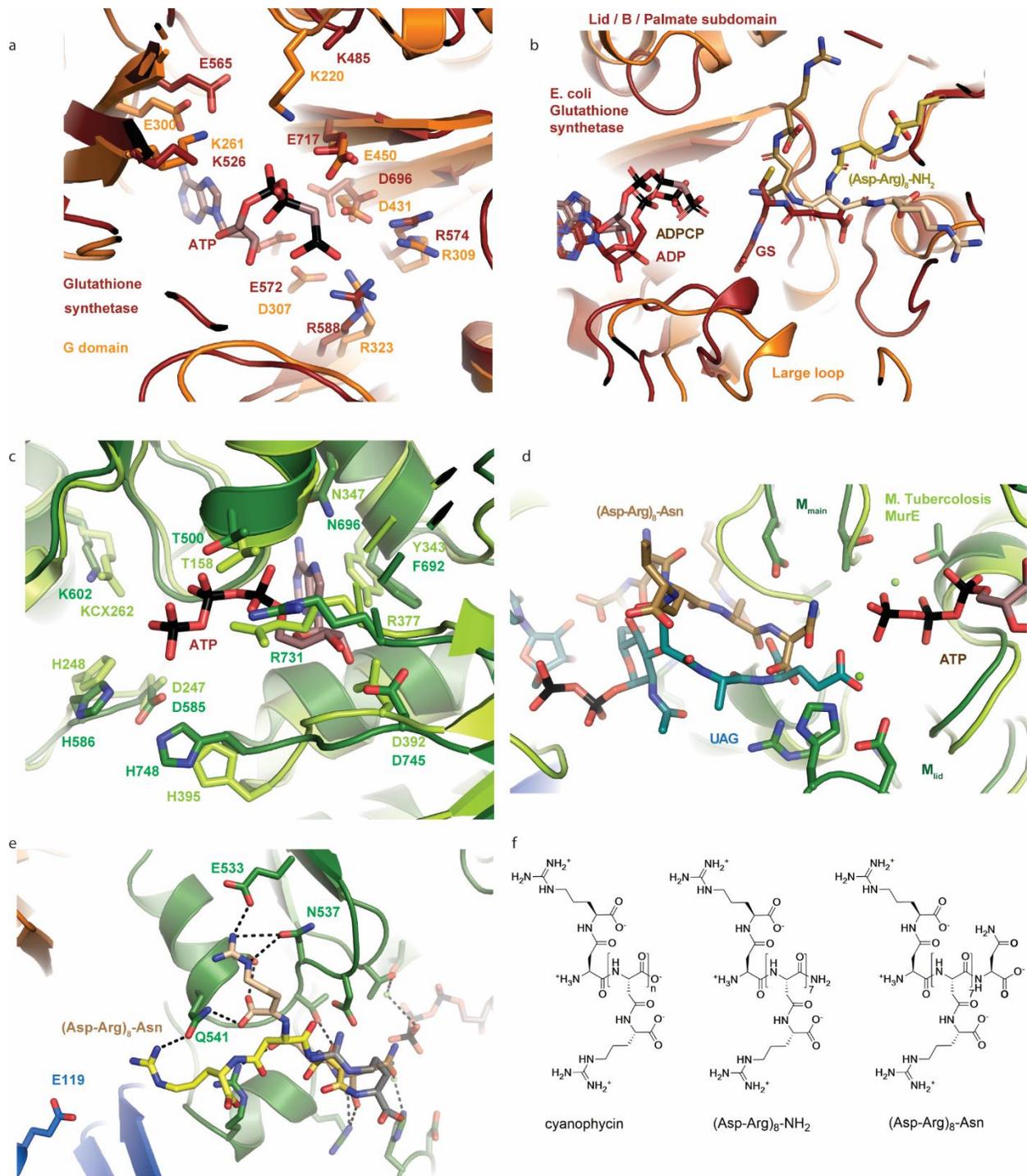
### 2.8.1. Supplementary figures



**Supplementary Figure. 2.1.** *CphA1* distribution. Phylogenetic tree of CphA1 sequences. A BLAST search found over 4000 CphA1-encoding gene sequences. Analysis of these sequences revealed that they are spread across most major bacterial phyla. Specific clades of particular interest were manually annotated and colored. The homologs used in this study are labeled in red. Gammaproteobacterial *Tm*CphA1 and *Ab*CphA1 share ~41% identity with cyanobacterial *Su*CphA1. There is evidence for both ancient horizontal gene transfer (alpha-, delta- and gammaproteobacterial CphA1s cluster together, but apart from betaproteobacteria) and more recent transfer in the unlabeled, black clusters of CphA1s that are from several different bacterial groups.

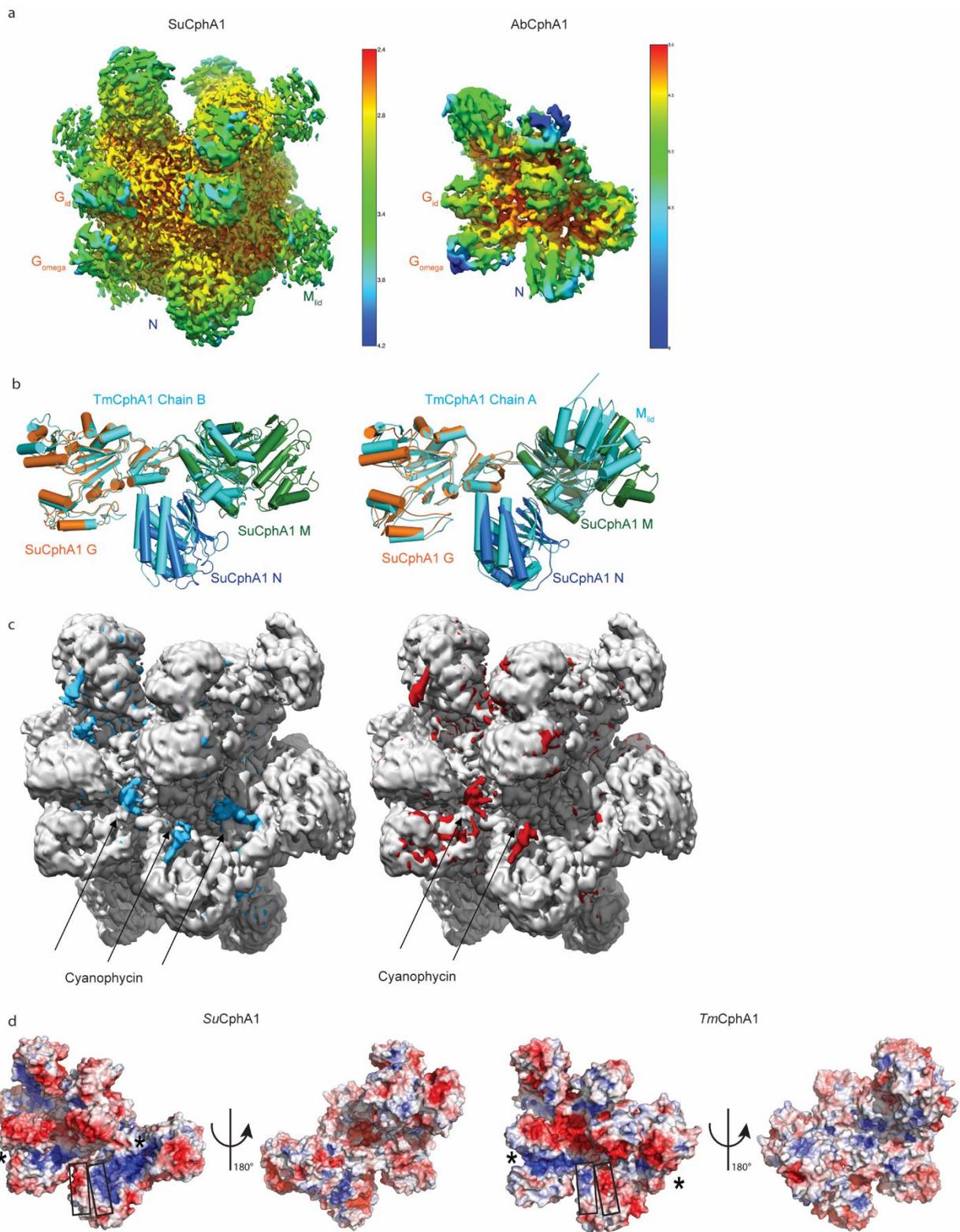


**Supplementary Figure 2.2.** CphA1 tetramerization. **a**, *Su*CphA1 (a, c) and *Tm*CphA1 (b, d) display different tetramer architectures, in which different monomers are responsible for tetramer-forming interactions. (e) The EM map and structure of *Su*CphA1 showing the tetramer interface, which is centered on W672. (f) Gel filtration chromatograms of all three CphA1 homologs used in this study, show they all form tetramers in solution.

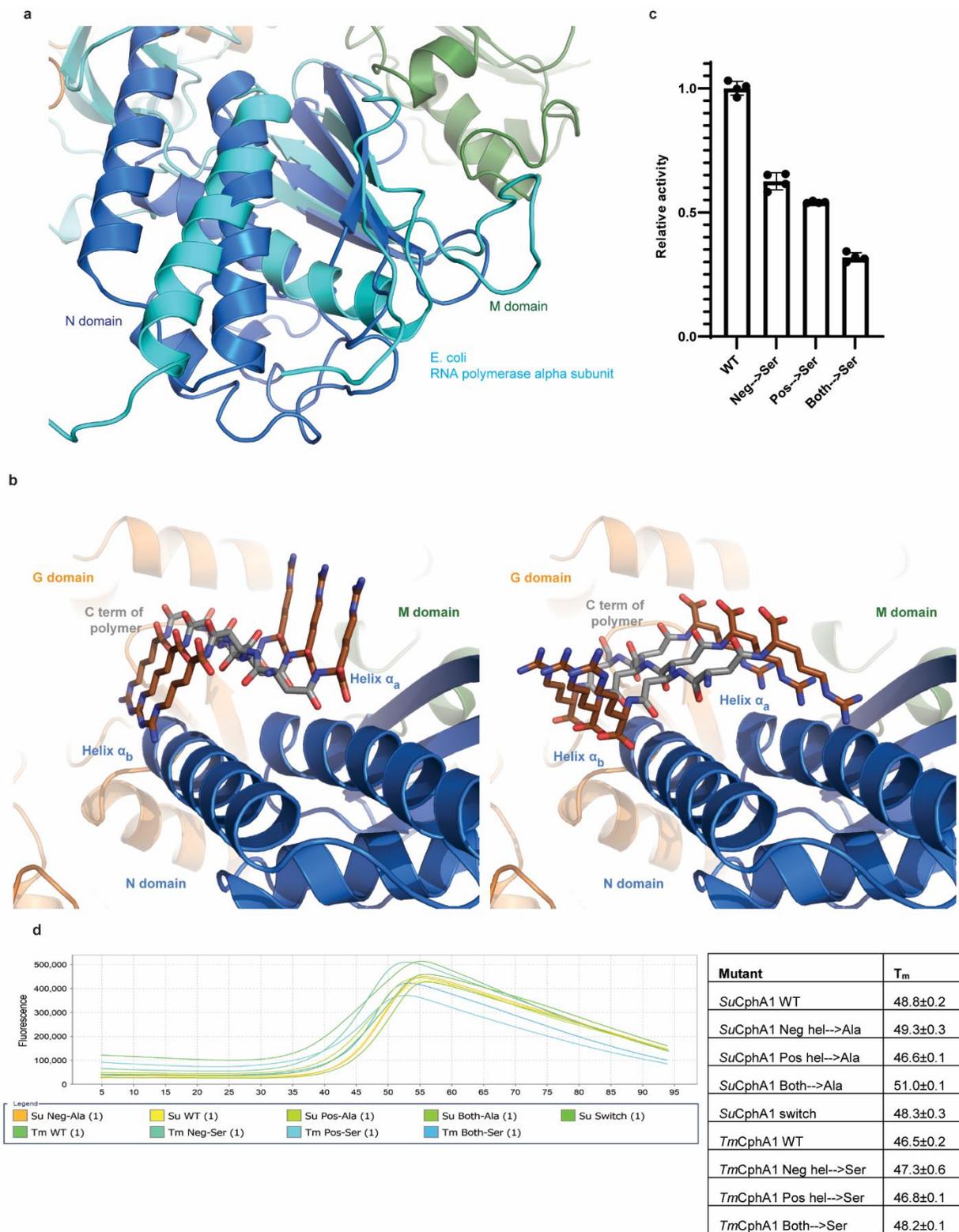


**Supplementary Figure 2.3. Comparison of CphA1 active sites to homologous enzymes.** (a), Overlay of *Su*CphA1 G domain and bifunctional glutathione synthetase from *S. agalactiae* (PDB code 3LN6) showing the similar ATP binding mode and conserved residues. (b) Overlay of *Su*CphA1 G domain and glutathione synthetase from *E. coli* (PDB code 1GSA) showing the

similar substrate orientation and overall structure. (c) Overlay of *Su*CphA1 M domain and MurE ligase from *M. tuberculosis* residues (PDB code 2WTZ) showing the similar ATP binding mode and conserved, and (d) similar substrate orientation. (e) The interactions made by cyanophycin with residues in the M domain of *Su*CphA1. (f) The three versions of cyanophycin and cyanophycin analogs used for the determined structures of *Su*CphA1 presented in this study.

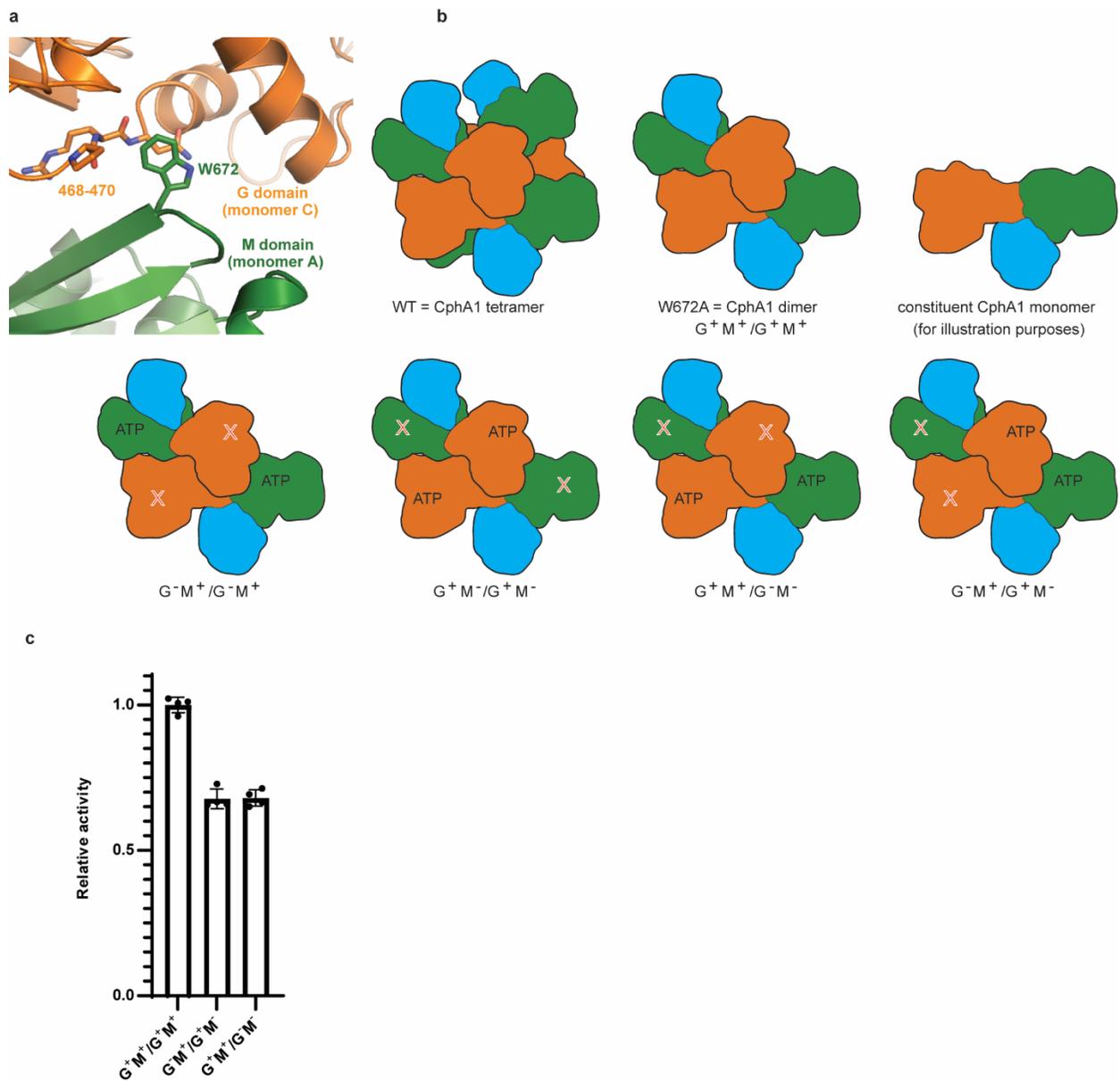


**Supplementary Figure 2.4. Flexibility, substrate binding and surface electrostatic potential of CphA1.** (a) Local resolution estimates of the cryo-EM maps of tetrameric *Su*CphA1 (left) and dimeric *Ab*CphA1 (right). (b) Overlay of the two chains in the crystal structure of *Tm*CphA1 (light blue) on the cryo-EM structure of *Su*CphA1 (colored), showing the different conformation adopted by  $M_{\text{lid}}$  of the crystal structure chain A and  $G_{\text{lid}}$  in chain B.  $M_{\text{lid}}$  is not visible in chain B. (c) Overlay of the unsharpened maps of *Su*CphA1 without cyanophycin substrate analogs (gray), and with (Asp-Arg)<sub>8</sub>-NH<sub>2</sub> (red, right) and (Asp-Arg)<sub>8</sub>-Asn (blue, left). Clear extra density is visible in the maps calculated in the presence of substrate analogs, mostly near the active sites and the N domain. (Asp-Arg)<sub>8</sub>-Asn is also seen as product in the G domain active site, but no density is visible for the terminal Asn residue. (d) Surface electrostatic potential maps of *Su*CphA1 and *Tm*CphA1 dimers showing how the side that faces the active sites is lined with negatively and positively charged patches. The side facing the inner cavity, which is opposite the active sites, is mostly neutral. Active sites are marked with \*,  $\alpha_a$  and  $\alpha_b$  are marked with rectangles.



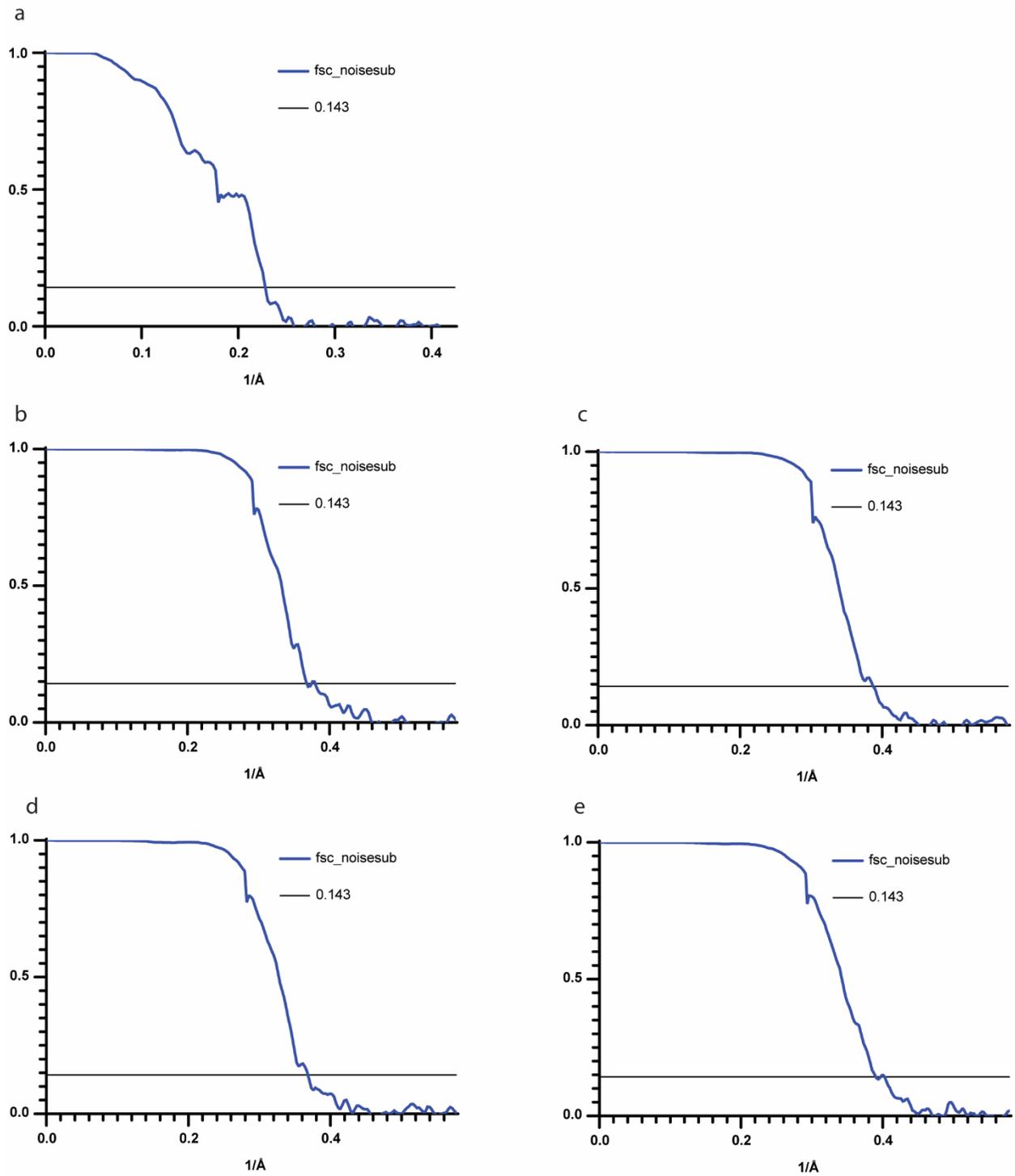
**Supplementary Figure 2.5. CphA1 N domain structural homology, cyanophycin binding mode and mutants analysis.** (a) Structure overlay of CphA1 with *E. coli* RNA polymerase alpha

subunit (PDB code 4JK1), showing similarity in parts of their structures. (b) Possible arrangement of cyanophycin that allows either its positive charges or negative charges to interact with  $\alpha$ a or  $\alpha$ b. (c) Activity assays of *TmCphA1* N domain mutants showing similar results to those observed for the equivalent *SuCphA1* mutants (displayed in Fig. 2.5). n = 4 independent experiments. Data are presented as individual measurements and mean value, error bars represent SD values. (d) Differential scanning fluorimetry melting curves and protein T<sub>m</sub> values of CphA1 N domain mutants. The similarity of T<sub>m</sub> values between wildtype and N domain mutants suggests that the observed differences in activity are a result of differences in interaction with cyanophycin rather than differences in protein stability. Additionally, the gel filtration profiles of the proteins during the purification process were all similar, again suggesting similar stability of wildtype and mutants. The values in the table represent the mean and SD of 3 independent measurements.



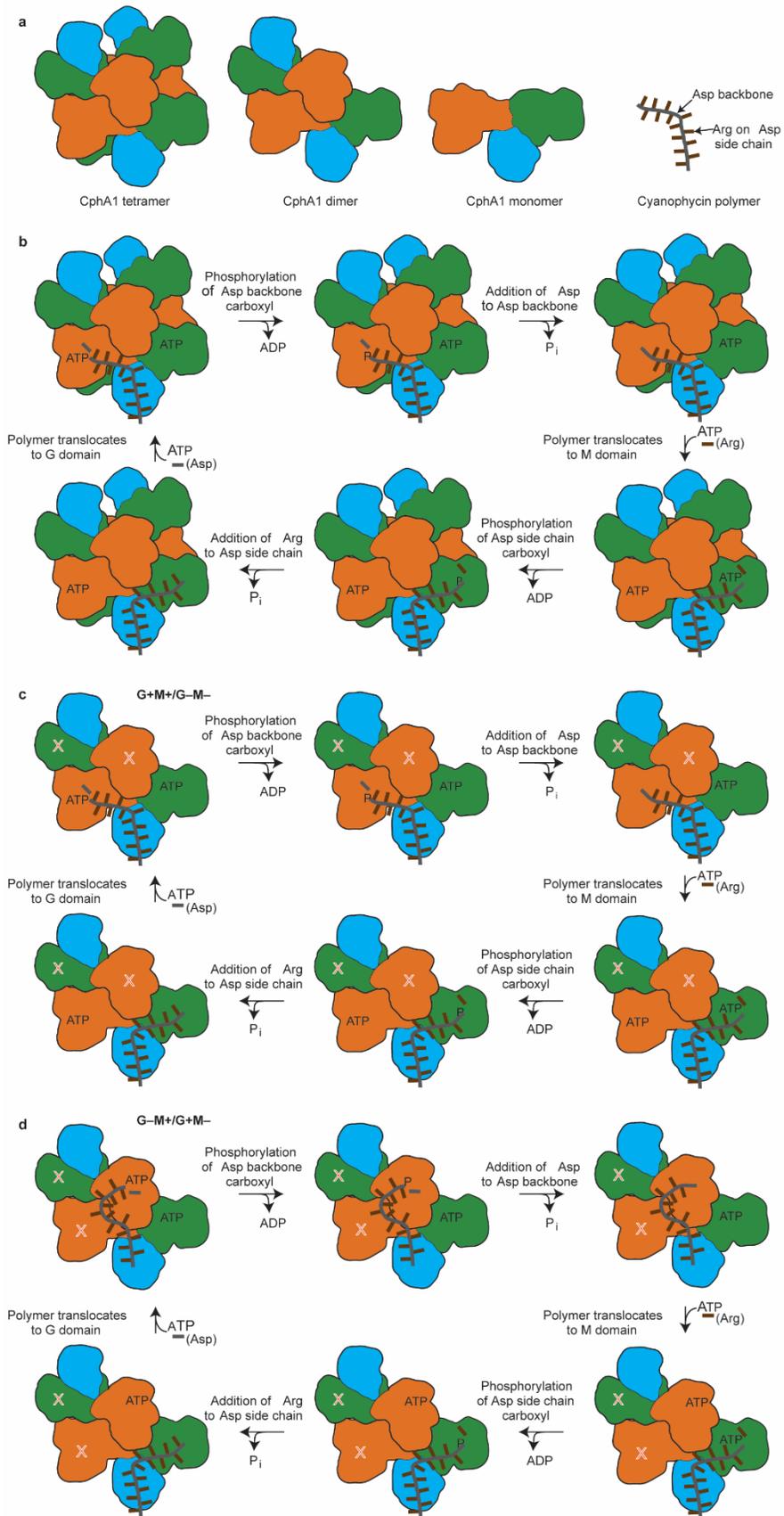
**Supplementary Figure 2.6. Dimeric CphA1 mutation schematics and activity in 100 mM sodium chloride.**

(a) The tetramerization interface, between W672 and residues 468–470, is disrupted in the obligate dimer W672A mutants. (b) Cartoon representation of the CphA1 mutants. (c) Both mutant combinations of dimer mutants ( $G^+M^-/G^-M^-$  and  $G^+M^+/G^-M^+$ ) show the same decrease in activity level relative to wildtype. The ratio between the activity rate of the wildtype CphA1 and mutants is similar to that observed with no sodium chloride in the reaction buffer.  $n = 4$  independent experiments. Data are presented as individual measurements and mean value, error bars represent SD values.



- A - *AbCphA1* with ATP (EMD-23327)
- B - *SuCphA1* bound with ATP (EMD-23311)
- C - *SuCphA1* bound with ADPCP and  $(\beta\text{-Asp-Arg})_8\text{-NH}_2$  (EMD-23325)
- D - *SuCphA1* bound with ATP and  $(\beta\text{-Asp-Arg})_6\text{-Asn}$  (EMD-23328)
- E - *SuCphA1* bound with ATP and  $(\beta\text{-Asp-Arg})_{16}$  (EMD-23326)

**Supplementary Figure 2.7. Fourier shell correlation for EM datasets.** FSC plots for all EM maps determined in this study.



**Supplementary Figure 2.8. Model of cyanophycin synthesis within wildtype and mutant CphA1.** Models of cyanophycin synthesis by WT CphA1 and the active site mutants used in the study.

## **2.8.2. Supplementary note - Chemical synthesis of cyanophycin segments**

### *2.8.2.1. General Information*

#### 2.8.2.1.1 Peptide synthesis

The resins were purchased from Biotage (Uppsala, Sweden; HMPB-ChemMatrix) or Bachem (Bubendorf, CH; PAL). All classical side chain protected amino acids were purchased from Bachem. HATU was purchased from Aapptec (Louisville, USA). All other chemicals were purchased from Sigma Aldrich (Merck KGaA, Darmstadt, Germany). All reagents were used as received and solvents were technical grade.

Fmoc peptide couplings were carried out in ISOLUTE Double fritted filtration columns, 15 or 25 mL (*reaction vessel*, Biotage) with orbital shaking at 600-700 rpm at RT. The building blocks for peptide synthesis were activated in 20 mL brown glass storage vials (27x57 mm, *activation vessel*, Infochroma ag) closed with Teflon lined caps, or in 4 mL screw vials (45x14.77 mm, *activation vessel*, BGB) closed with Teflon lined caps (Thermo Scientific) at RT.

MALDI-TOF mass spectra (MALDI-TOF MS) were recorded on a Bruker microflex benchtop MALDI-TOF system. High-resolution mass spectrometry (HRMS) was performed on a Bruker maXis UHR-TOF by electrospray ionization (ESI) or a Bruker solariX by matrix-assisted laser desorption/ionization (MALDI) by the Molecular and Biomolecular Analysis Service (MoBiAS) of the LOC at ETH Zurich.

#### 2.8.2.1.2 Analytical HPLC conditions for peptide analysis

All analytical HPLC runs were performed with a Dionex Ultimate 3000 system equipped with a 3000 pump-module, a 3000 Autosampler, a 3000 RS Variable Wavelength Detector, and a Xbridge C18 3.5  $\mu$ m 150x4.6 mm column (Waters). The eluent system was a mixture of H<sub>2</sub>O and ACN containing 0.1% TFA. A summary of the method used can be found in Supplementary Table 4.

#### 2.8.2.1.3 Semi-preparative HPLC conditions

All semi-preparative HPLC runs were performed with a Waters preparative 150 LC system equipped with a 2545 quaternary gradient module, a 2489 UV/visible detector, a Fraction Collector

III and a Reprosil Gold 120 C18 5  $\mu\text{m}$  250x20 mm column (Morvay Analytik). The eluent system was  $\text{H}_2\text{O}$  and ACN, both supplemented with 0.1% TFA, and the flow rate was set to 15 mL/min.

Lyophilized samples were always dissolved in a solvent mixture identical to that at the beginning of the purification gradient. If the lyophilized peptide was insoluble under these conditions, ACN up to 20% was added over the original amount. In some cases, drops of MeOH or DMF were added to help solubilization. Every sample was filtered through a 0.22  $\mu\text{m}$  filter before injection on the semi-preparative column. For the exact separation conditions refer to the detailed synthesis protocol of the individual peptides. Different gradients were used. They are listed in Supplementary Table 5.

2.8.2.2. *Cyanophycin building block synthesis ( $N^2$ -(((9H-Fluoren-9-yl)methoxy)carbonyl)- $N^4$ -(1-(tert-butoxy)-1-oxo-5-(3-((2,2,4,6,7-pentamethyl-2,3-dihydrobenzofuran-5-yl)sulfonyl)guanidino)pentan-2-yl)asparagine; Fmoc-( $\beta$ -Asp-Arg)(OtBu)-OH; Fmoc-( $\beta$ -Asp-Arg))*

The building block Fmoc-( $\beta$ -Asp-Arg)(OtBu)-OH for the synthesis of cyanophycin segments and its intermediate H-Arg(Pbf)-OH for the *in situ* generation of ( $\beta$ -Asp-Arg) on acidic linker were synthesized according to the procedure described<sup>65,159</sup>.

2.8.2.3. *Solid phase peptide synthesis of cyanophycin oligomers.*

2.8.2.3.1 GP1. General protocol for coupling the first amino acid (0.1 mmol scale)

In a 25 mL *reaction vessel* equipped with a valve and attached to a suction system, a resin with free amide linker (0.1 mmol) was shaken in DMF for 30 min. The resin was then treated with 20% (vol/vol) piperidine-DMF (5 mL) for 5 min and washed with DMF (5x 1 min, 5 mL each). In a 10 mL *activation vessel*, the Fmoc-protected amino acid of choice (0.6 mmol, 6 eq.) was dissolved in 0.4 M HATU in DMF (1360  $\mu\text{L}$ , 0.54 mmol, 5.4 eq.) and DIPEA (187  $\mu\text{L}$ , 10.8 mmol, 10.8 eq.), and activated for 2 min at RT. Afterwards, the solution was transferred to the *reaction vessel* and DMAP in DMF (10 mg/mL, 1 mL) was added. The reaction was shaken overnight at RT. Finally, the resin was washed with DMF (5x 1 min, 5 mL each), treated for 5 min with DMF-Ac<sub>2</sub>O-DIPEA (10 mL 8:1:1) and washed again with DMF (5x 1 min, 5 mL each). The resin was then stored at 4 °C until further use (usually within 2 weeks).

#### 2.8.2.3.2 GP2. General protocol for *in situ* synthesis of Fmoc protect ( $\beta$ -Asp-Arg)(OtBu) on HMPB-Chemmatrix (0.1 mmol scale)

In a 25 mL *reaction vessel* equipped with a valve and attached to a suction system, HMPB-Chemmatrix (0.1 mmol) was shaken in DMF for 30 min. The resin was then treated with 20% (vol/vol) piperidine-DMF (5 mL) for 5 min and washed with DMF (5x 1 min, 5 mL each). In a 10 mL *activation vessel*, Fmoc-Asp(OAll)-OH (237 mg, 0.6 mmol, 6 eq.) was dissolved in 0.4 M HATU in DMF (1360  $\mu$ L, 0.54 mmol, 5.4 eq.) and DIPEA (187  $\mu$ L, 10.8 mmol, 10.8 eq.) and activated for 2 min at RT. Afterwards, the solution was transferred to the *reaction vessel* and DMAP in DMF (10 mg/mL, 1 mL) was added. The reaction was shaken overnight at RT. Finally, the resin was washed with DMF (5x 1 min, 5 mL each), treated for 5 min with DMF-Ac<sub>2</sub>O-DIPEA (10 mL, 8:1:1) and washed again with DMF (5x 1 min, 5 mL each).

Afterwards, the resin was washed with DMF (5x 1 min, 5 mL each) followed by dry CH<sub>2</sub>Cl<sub>2</sub> (5x 1 min, 5 mL each). For all steps involving dry CH<sub>2</sub>Cl<sub>2</sub>, the resin was not shaken but bubbled with a stream of N<sub>2</sub>. Small amounts of fresh, dry CH<sub>2</sub>Cl<sub>2</sub> had to be added every once in a while, in order to compensate for evaporation of the solvent. In order to remove the Allyl protecting group on the side chain of Asp, Pd(PPh<sub>3</sub>)<sub>4</sub> (35 mg, 30  $\mu$ mol, 0.2 eq.) and PhSiH (250  $\mu$ L, 2 mmol, 13.3 eq.) predissolved in dry CH<sub>2</sub>Cl<sub>2</sub> (5 mL) were added to the *reaction vessel*, and the mixture was bubbled with N<sub>2</sub> for 30 min at RT. Finally, the resin was washed with dry CH<sub>2</sub>Cl<sub>2</sub> (5x 1 min, 5 mL each)). The Pd treatment and the subsequent CH<sub>2</sub>Cl<sub>2</sub> washing steps were repeated twice. After the last CH<sub>2</sub>Cl<sub>2</sub> wash, the resin was treated with 5% (w/vol) sodium diethyldithiocarbamate in DMF (2x min, 5 mL each) and washed with DMF (5x 1 min, 5 mL each).

The resin was then treated with a mixture of 0.4 M HATU (450  $\mu$ L, 0.18 mmol) and DIPEA (62  $\mu$ L, 3.6 mmol) in DMF (1000  $\mu$ L) for 3 min and H-Arg(Pbf)-OtBu (289 mg, 0.6 mmol) predissolved in DMF (1000  $\mu$ L) was added. The *reaction vessel* was then shaken for 5 h at RT. Finally, the resin was washed with DMF (5x 1 min, 5 mL each) and stored at 4 °C until further use (usually within 2 weeks).

#### 2.8.2.3.3 GP3. General protocol for manual Fmoc SPPS with 3 eq. amino acid (0.03 mmol scale)

In a 15 mL *reaction vessel* equipped with a valve and attached to a suction system, the resin loaded with the first Fmoc-protected C-terminal amino acid (0.03 mmol) was shaken in DMF for 30 min. Then, the resin was treated with 20% (vol/vol) piperidine-DMF (5 mL) for 5 min and washed with DMF (5x 1 min, 5 mL each). Peptide chain assembly took place as described in

Supplementary Table 6. The only exception was for the first coupling, which was carried out for 4h. Finally, the resin was treated with CH<sub>2</sub>Cl<sub>2</sub> (5x 1 min washes, 5 mL each), dried under suction for 25 min and transferred to a flask suitable for cleavage. The dried resin was either cleaved directly or stored at -20 °C.

#### 2.8.2.3.4 GP4. General cleavage protocol

In a round-bottom flask, the dried resin was swollen in TFA-ddH<sub>2</sub>O-TIS (95:2.5:2.5; 20 mL/g dry resin). The mixture was stirred at RT for at least 4h. Afterwards, the resin was filtered off using a fritted glass filter in a 50 mL falcon tube and washed with 2-3 mL neat TFA. The TFA was evaporated with N<sub>2</sub> flow until some material started to precipitate. Then, 4 °C Et<sub>2</sub>O was added to the falcon tube, the resulting cloudy suspension was centrifuged at 7000 g and 4 °C for 10 min and the supernatant was discarded. The centrifugation step was repeated two to three times with fresh 4 °C Et<sub>2</sub>O. Finally, to solubilize the peptidic material, the precipitate was resuspended in H<sub>2</sub>O-ACN (1:1) + 0.1% TFA (30-40 mL). The falcon tube was frozen, lyophilized and stored at 4 °C before purification via semi-preparative HPLC.

#### 2.8.2.3.5 (β-Asp-Arg)<sub>8</sub>-NH<sub>2</sub>

(β-Asp-Arg)<sub>8</sub>-NH<sub>2</sub> was synthesized on a 0.03 mmol scale on PAL resin (111 mg, 0.28 mmol/g). Fmoc-(β-Asp-Arg)(OtBu)-OH was used for every coupling. The first coupling was performed according to **GP1**. The rest of the synthesis followed **GP3** but using 2 equivalents of building block instead of 3. The cleavage was carried out as described in **GP4**. The crude product was purified with gradient 1 and gradient 2 in two steps. Fractions were analyzed with MALDI-TOF MS and analytical HPLC, combined according to purity and lyophilized. Pure (β-Asp-Arg)<sub>8</sub>-NH<sub>2</sub> was obtained as a white powder. The corresponding analytical chromatogram is shown in Supplementary Figure 2.9a. HRMS analysis of the produced peptide delivered the expected mass.

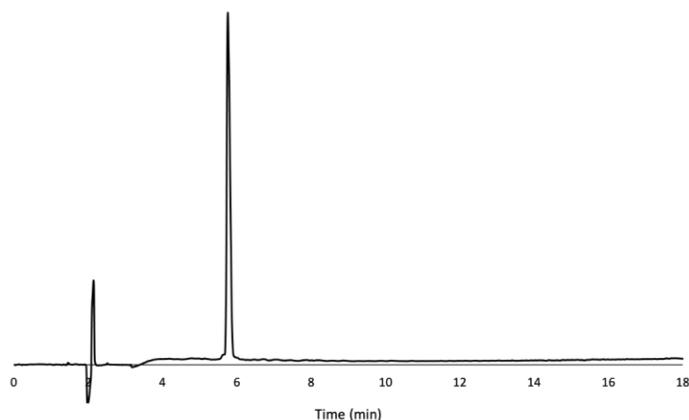
#### 2.8.2.3.6 (β-Asp-Arg)<sub>8</sub>-Asn

(β-Asp-Arg)<sub>8</sub>-Asn was synthesized on a 0.015 mmol scale on HMPB Chemmatrix (34 mg, 0.44 mmol/g). The first coupling was performed according to **GP1** with Fmoc-Asn(Trt)-OH. The rest of the synthesis followed **GP3** (adapted for 0.015 mmol scale) with Fmoc-(β-Asp-Arg)(OtBu)-OH as building block. The cleavage was carried out as described in **GP4**. The crude product was purified with gradient 3. Fractions were analyzed with MALDI-TOF MS and analytical HPLC, combined according to purity and lyophilized. Pure (β-Asp-Arg)<sub>8</sub>-Asn was obtained as a white

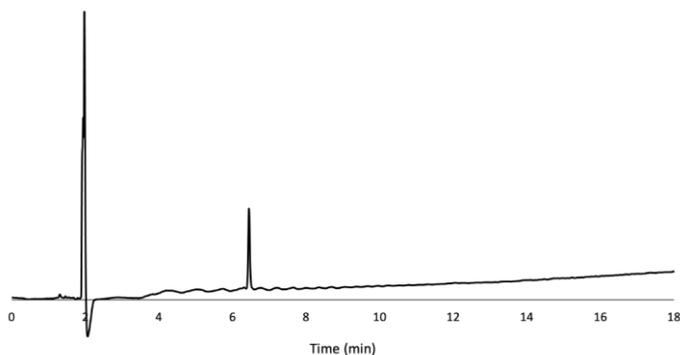
powder. The corresponding analytical chromatogram is shown in Supplementary Figure 2.9b. HRMS analysis of the produced peptide delivered the expected mass.

#### 2.8.2.3.7 ( $\beta$ -Asp-Arg)<sub>16</sub>

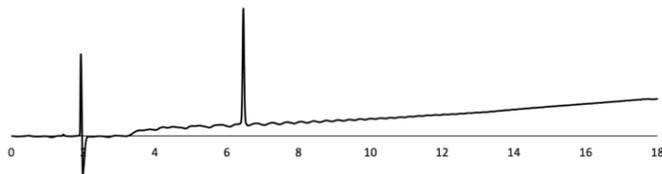
( $\beta$ -Asp-Arg)<sub>16</sub> was synthesized on a 0.03 mmol scale on HMPB-Chemmatrix (111 mg, 0.44 mmol/g). The first addition of the first (Asn-Asp) dipeptide unit to the resin was performed according to **GP2**. The rest of the synthesis followed **GP3** with Fmoc-( $\beta$ -Asp-Arg)(OtBu)-OH as building block. The cleavage was carried out as described in **GP4**. The crude product was purified with gradient 3 and gradient 4 in two steps. Fractions were analyzed with MALDI-TOF MS and analytical HPLC, combined according to purity and lyophilized. Pure ( $\beta$ -Asp-Arg)<sub>16</sub> was obtained as a white powder. The corresponding analytical chromatogram is shown in Supplementary Figure 2.9c. HRMS analysis of the produced peptide delivered the expected mass.



**a.** ( $\beta$ -Asp-Arg)<sub>8</sub>-NH<sub>2</sub>. HRMS (ESI):  $m/z$  [M+H]<sup>3+</sup> calculated at 729.6909, found at 729.6922.



**b.** ( $\beta$ -Asp-Arg)<sub>8</sub>-Asn. (ESI):  $m/z$  [M+H]<sup>3+</sup> calculated at 768.0333, found at 768.0341.



c. ( $\beta$ -Asp-Arg)<sub>16</sub>-OH. HRMS (MALDI):  $m/z$  [M+H]<sup>+</sup> calculated at 4357.0672, found at 4357.0851.

**Supplementary Figure 2.9. Analytical HPLC chromatogram of synthesized cyanophycin segments monitored at 220 nm.**

### 2.8.3. Supplementary tables

<i>TmCphA1</i> (PDB 7LGN)	
<b>Data collection</b>	
Space group	I4 <sub>1</sub>
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	222.6, 222.6, 105.2
$\alpha$ , $\beta$ , $\gamma$ (°)	90.0, 90.0, 90.0
Resolution (Å)	78.7-3.1 (3.2-3.1)
<i>R</i> <sub>merge</sub>	0.5479 (4.269)
<i>I</i> / $\sigma$ <i>I</i>	28.53 (0.92)
Completeness (%)	99.92 (99.87)
Redundancy	17.6 (17.8)
CC <sub>1/2</sub>	0.995 (0.768)
<b>Refinement</b>	
Resolution (Å)	78.7-3.1
No. reflections	46773
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.2711/0.3264
No. atoms	11761
Protein	11761
Ligand/ion	0
Water	0
<i>B</i> -factors	
Protein	131.9

R.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.59

**Supplementary Table 2.1.** X-ray data collection and refinement statistics.

\*Data from 6 crystals was used in this dataset. Ramachandran favored 94.93%, Allowed 5.01%, outliers 0.0%. Clash score is 4.48.

	<i>AbCphA1</i> + ATP (EMDB 23327) (PDB 7LGM)	<i>SuCphA1</i> + ATP (EMDB 23321) (PDB 7LG5)	<i>SuCphA1</i> + ATP + (Asp-Arg) <sub>8</sub> -Asn (EMDB 23328) (PDB 7LGQ)	<i>SuCphA1</i> + AMPPCP + (Asp- Arg) <sub>8</sub> -NH <sub>2</sub> (EMDB 23325) (PDB 7LGJ)	<i>SuCphA1</i> + ATP + (Asp-Arg) <sub>16</sub> (EMDB 23326)
<b>Data collection and processing</b>					
Magnification	30,000x	105,000x	105,000x	105,000x	105,000x
Voltage (kV)	200	300	300	300	300
Electron exposure (e- /Å <sup>2</sup> )	57	60	65	61	55
Defocus range (µm)	-2.0- to -3.50	-0.75 to -2.5	-0.75 to -2.5	-0.75 to -2.5	-0.75 to -2.5
Pixel size (Å)	0.58 (unbinned)	0.855	0.855	0.855	0.855
Symmetry imposed	C2	D2	D2	D2	D2
Initial particle images (no.)					
Final particle images (no.)	296,574	687,356	368,761	311,861	135,110
Map resolution (Å) FSC threshold 0.143	4.4	2.6	2.7	2.6	2.6
Map resolution range (Å)	3.6-9.1	2.2-7.9	2.3-8.4	2.4-8.8	2.4-9.0
<b>Refinement</b>					
Model resolution (Å) FSC threshold 0.143	4.4	2.6	2.7	2.6	
Model resolution range (Å)	3.6-9.1	2.2-7.9	2.3-8.4	2.4-8.8	
Map sharpening <i>B</i> factor (Å <sup>2</sup> )	-241	-105	-95	-103	-105
<b>Model composition</b>					
Non-hydrogen atoms	10954	26892	27756	27432	
Protein residues	10892	26632	26548	26584	

Ligands	62	260	1208	848
<i>B</i> factors (Å <sup>2</sup> )				
Protein	-241	-105	-95	-103
Ligand	-241	-105	-95	-103
R.m.s. deviations				
Bond lengths (Å)	0.018	0.018	0.019	0.018
Bond angles (°)	1.846	1.866	1.750	1.773
Validation				
MolProbity score	1.5	1.07	1.11	1.11
Clashscore	2.68	0.94	1.14	1.09
Poor rotamers (%)	0.0	0.14	0.0	0.0
Ramachandran plot				
Favored (%)	93.03	95.88	95.82	95.76
Allowed (%)	6.97	4.12	4.18	4.24
Disallowed (%)	0.0	0.0	0.0	0.0

**Supplementary Table 2.2.** Cryo-EM data collection, refinement and validation statistics.

Phylum/Class	Organism	Strain
Firmicutes	<i>D. hafniense</i>	DSM10664
Firmicutes	<i>S. thermosulfidooxidans</i>	DSM9293
Firmicutes	<i>A. californiensis</i>	DSM14826
Firmicutes	<i>C. acetigignens</i>	DSM18802
Firmicutes	<i>P. cellulosolvens</i>	DSM2933 (two homologs)
Cyanobacteria	<i>T. elongates</i>	BP-1
Cyanobacteria	<i>Synechococcus</i> sp.	MA-19
Cyanobacteria	<i>Synechocystis</i> sp.	UTEX2470
Cyanobacteria	<i>Anabaena</i> sp.	UTEX2576
Gammaproteobacteria	<i>A. baylyi</i>	DSM587
Gammaproteobacteria	<i>T. morbirosei</i>	DSM23827
Betaproteobacteria	<i>B. cepacia</i>	DSM7288
Alphaproteobacteria	<i>P. soli</i>	DSM19599

**Supplementary Table 2.3.** CphA1 homologs tested for soluble expression in this study.

Condition	Column	% ACN	Time (min)
1 mL/min 25 °C	Xbridge C18 3.5 μm 150x4.6 mm (Waters)	5	0 -> 1
		5 -> 50	1 -> 16
		50	16 -> 19
		50 -> 95	19 -> 20
		95	20 -> 23
		95 -> 5	23 -> 24
		5	24 -> 28

**Supplementary Table 2.4. Summary of the condition used for analytical HPLC**

Gradient	% ACN	Time (min)
Gradient 1	5	0 -> 5
	5 -> 28	5 -> 40
	28 -> 90	40 -> 45
	90	45 -> 50
Gradient 2	0	0 -> 5
	0-> 30	5 -> 45
	30 -> 90	45 -> 50
	90	50 -> 60
Gradient 3	2	0 -> 5
	2 -> 25	5 -> 40
	25 -> 90	40 -> 45
	90	45 -> 50
Gradient 4	10	0 -> 5
	10 -> 50	5 -> 45
	50 -> 90	45 -> 46
	90	46 -> 60

**Supplementary Table 2.5. Summary of the gradients used for peptide purification**

Step	Procedure	Reagent	Conditions
1	Activation	<i>In activation vessel</i> 0.09 mmol ( $\beta$ -Asp-Arg)(OtBu)-OH (3 eq.) 210 $\mu$ L HATU (0.4 M in DMF, 83 $\mu$ mol, 2.8 eq.) 27.6 $\mu$ L DIPEA (0.16 mmol, 5.3 eq.) 600 $\mu$ L DMF	3 min
2	Coupling	<i>In reaction vessel</i> 0.09 mmol activated amino acid 600 $\mu$ L of DMF	2 h
3	Washing	DMF	5 x 1 min washes
4	Deprotection	20 % piperidine (vol/vol) in DMF	1 x 2 min and 2 x 8 min treatments
5	Washing	DMF	5 x 1 min washes

**Supplementary Table 2.6. Chain assembly of peptides by Fmoc SPPS on a 0.03 mmol scale with 3 eq. of amino acid.**

## Bridge to chapter 3

CphA1 had mostly been described as having primer-dependent activity, meaning it can only extend existing cyanophycin chains, rather than start polymerization *de novo*. One exception was reported: CphA1 from *Thermosynechococcus elongatus* BP-1 displays robust activity *in vitro* even when no primer is added to the reaction. While performing the research described in the first chapter, I noticed that the CphA1 from *Synechocystis* sp. UTEX2470 is also capable of primer-independent activity. This prompted me to examine primer independence more closely. Specifically, I wanted to determine what length of cyanophycin can serve as primer and identify the structural determinants of CphA1 primer-dependence and primer-independence. I hoped to use this information to study the importance of primer availability for cyanophycin production in heterologous hosts.

### 3. A cryptic third active site in cyanophycin synthetase creates primers for polymerization

Published in: Sharon I, Pinus S, Grogg M, Moitessier N, Hilvert D, Schmeing TM. *Nature Communications* 2022 Jul 7; 13:3923.

### **3.1. Abstract**

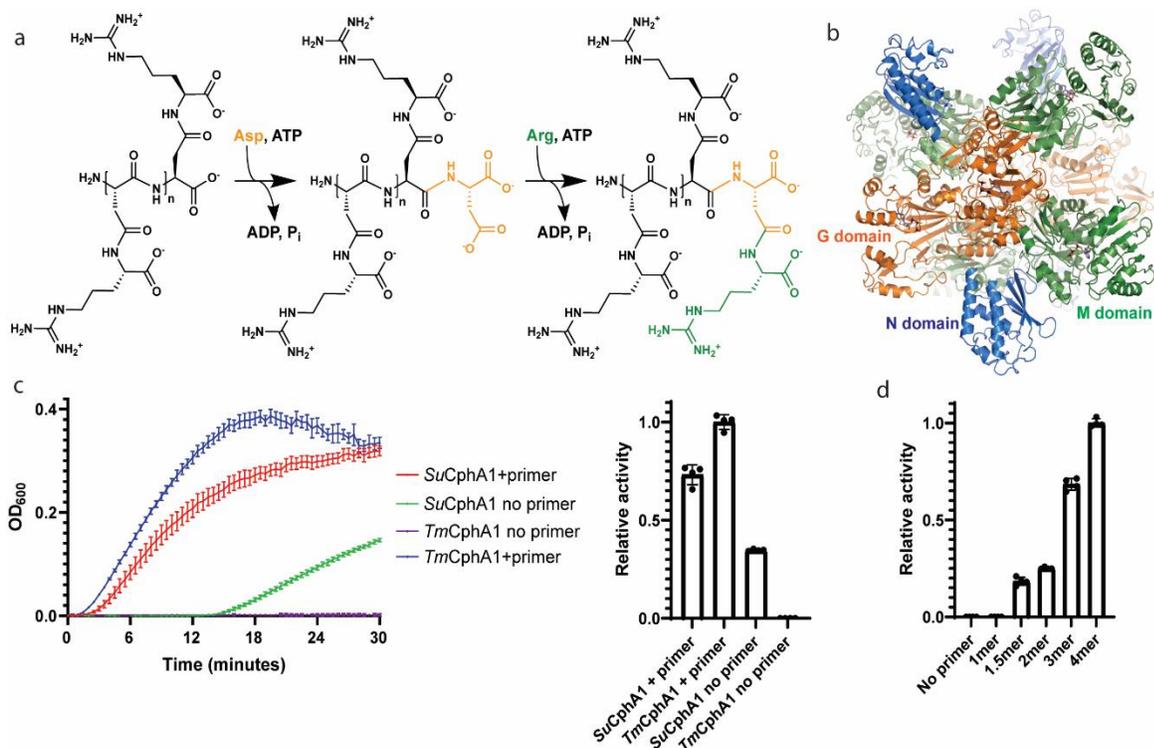
Cyanophycin is nitrogen reserve biopolymer in many bacteria that has promising industrial applications. Made by cyanophycin synthetase 1 (CphA1), it has a poly-L-Asp backbone with L-Arg residues attached to each aspartate sidechain. CphA1s are thought to typically require existing segments of cyanophycin to act as primers for cyanophycin polymerization. In this study, we show that most CphA1s will not require exogenous primers and discover the surprising cause of primer independence: CphA1 can make minute quantities of cyanophycin without primer, and an unexpected, cryptic metallopeptidase-like active site in the N-terminal domain of many CphA1s digests these into primers, solving the problem of primer availability. We present co-complex cryo-EM structures, make mutations that transition CphA1s between primer dependence and independence, and demonstrate that primer dependence can be a limiting factor for cyanophycin production in heterologous hosts. In CphA1, domains with opposite catalytic activities combine into a remarkable, self-sufficient, biosynthetic nanomachine.

### **3.2. Introduction**

Cyanophycin is a natural biopolymer discovered over 130 years ago as large, insoluble granules within cyanobacterial cells<sup>14</sup>. Cyanophycin chains consist of a poly-L-Asp backbone with L-Arg residues attached to each Asp side chain through isopeptide bonds<sup>16</sup>. Chains are typically ~80-400  $\beta$ -Asp-Arg dipeptides in length ( $(\beta\text{-Asp-Arg})_{\sim 80-400}$ )<sup>15,99</sup>. The high nitrogen content and inert nature of cyanophycin make it ideal for storing fixed nitrogen<sup>130</sup>, as well as carbon and energy<sup>133,134</sup>. Cyanophycin is especially useful for nitrogen-fixing cyanobacteria which separate aerobic photosynthesis and anaerobic nitrogen fixation either spatially or temporally<sup>40,42</sup>, has been shown to enhance the efficiency of nitrogen assimilation in non-diazotrophic strains<sup>102</sup>, and is also produced by many other bacteria across the kingdom<sup>2</sup>. Cyanophycin has promising potential commercial applications, from use as bandage material<sup>66</sup> to providing a source of poly-Asp, a biodegradable antiscalant, water softener, and super-swelling material<sup>137</sup>. Nevertheless, production yields of cyanophycin are currently too low for commercial viability and many studies have sought to increase them<sup>80,82,160,161</sup>.

Cyanophycin synthetase 1 (CphA1) catalyzes polymerization of Asp and Arg into cyanophycin in two iterative, ATP-dependent reactions (Fig. 3.1a). In the first reaction, the C-terminus of a cyanophycin chain is activated by phosphorylation and then elongated by peptide bond formation with aspartic acid<sup>1,98</sup>. In the second reaction, the side chain carboxylate of the newly added Asp residue is phosphorylated and then reacts with arginine to form an isopeptide bond. CphA1 contains dedicated domains for each reaction (Fig. 3.1b): the ATP-grasp family G domain ligates the Asp to the main chain, and the Mur ligase-like M domain adds the Arg to the Asp side chain. All CphA1 enzymes also contain an N-terminal domain (N domain), whose function was previously unknown. We recently showed that the N domain aids polymerization by loosely binding cyanophycin through charged patches, which helps the growing cyanophycin polymer alternate binding to G and M domain active sites<sup>2,98</sup>. Our study also visualized two separate tetrameric architectures for CphA1<sup>2,98</sup>.

CphA1s have most often been described as possessing primer-dependent activity<sup>3,98,99,106</sup>. It is widely accepted that primer-dependent CphA1s cannot synthesize cyanophycin *de novo* from only Asp, Arg and ATP, but require existing chains to extend. Only CphA1 from *Thermosynechococcus elongatus* BP-1 has been shown to display robust primer-independent



**Figure 3.1. CphA1 structure and activity.** (a) Schematic diagram of the biosynthetic reactions catalyzed by the G and M domains of CphA1. (b) The overall structure of tetrameric CphA1<sup>2</sup> from *Synechocystis* sp. UTEX2470 (*Su*CphA1, PDB code 7LG5). ATP molecules mark the positions of the G (orange) and M (green) domain active sites. The N domain is colored in blue. (c) Cyanophycin biosynthesis plots and rate comparison of synthesis by *Su*CphA1 and *Tm*CphA1 with and without primer. *Tm*CphA1 is completely inactive in the absence of primer. n=4 independent experiments. Data are presented as individual measurements and mean value, error bars represent SD values. (d) Activity levels of *Tm*CphA1 in the presence of various cyanophycin primers: 1mer ( $\beta$ -Asp-Arg)<sub>1</sub>, 1.5mer is ( $\beta$ -Asp-Arg)-Asp, 2mer ( $\beta$ -Asp-Arg)<sub>2</sub>, 3mer ( $\beta$ -Asp-Arg)<sub>3</sub>, 4mer ( $\beta$ -Asp-Arg)<sub>4</sub>, etc. n=4 independent experiments. Data are presented as individual measurements and mean value, error bars represent SD values.

activity<sup>105</sup>. Primer-dependent CphA1s are known to use long cyanophycin chains<sup>106</sup>, trimer dipeptide segments ( $(\beta$ -Asp-Arg)<sub>3</sub>)<sup>1</sup>, and (albeit with low efficiency) other biomaterials<sup>104</sup> as primers, but characteristics of minimal and optimal primers are not established. It was also completely unknown what determines whether a CphA1 enzyme is primer dependent or primer independent. When heterologously expressed, CphA1 is catalytically active and produces

cyanophycin within host cells<sup>3,74,86,94,98,99,106</sup>, so understanding the nature of primers and primer independence could be important for bioproduction yields from these hosts.

Here, we report the discovery that the key to primer independence is a cryptic metallopeptidase-like active site in the N domain. We use a combination of cryo-EM, mass spectrometry, mutagenesis and biochemical assays to characterize and manipulate primer-independent CphA1 activity. The results show how the N domain enables biosynthesis by CphA1 without exogenous primers and demonstrate the implications of primer independence for *in vivo* cyanophycin production.

### **3.3. Results**

#### **3.3.1. Primer dependence of CphA1 enzymes**

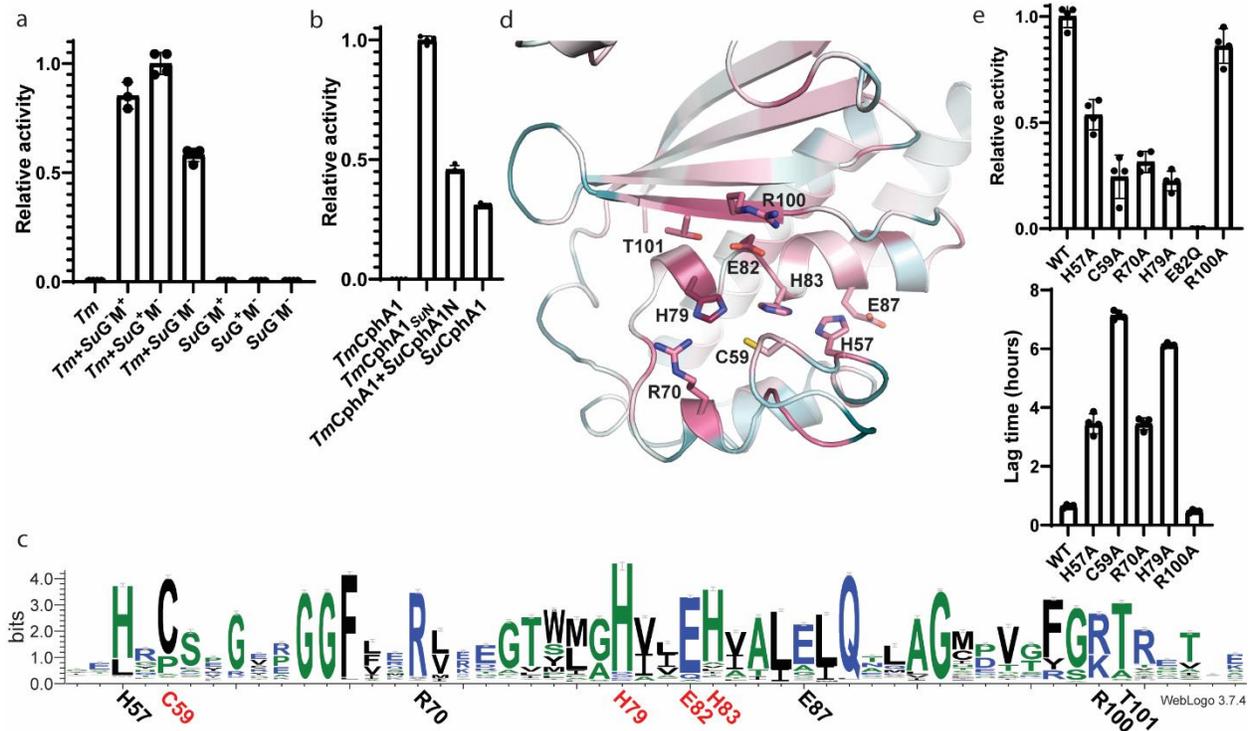
Cyanophycin synthetases from *Synechocystis sp.* UTEX2470 (*Su*CphA1) and *Tatumella morbirosei* DSM23827 (*Tm*CphA1) show high activity in the presence of a ( $\beta$ -Asp-Arg)<sub>3</sub> primer<sup>1,2,98</sup> (Fig. 3.1c). Notably, in the absence of primer, *Su*CphA1 displays a lag phase of ~15 minutes, followed by robust cyanophycin synthesis. This result is surprising, because only one other CphA1 enzyme has ever been reported to be primer independent<sup>105</sup>. In contrast to *Su*CphA1, no primer-independent activity is observed for *Tm*CphA1, even at high protein concentrations and incubation with substrates over several days (Supplementary Fig. 3.1a). These results led us to ask what properties of the primer and enzyme control primer dependent and independent cyanophycin synthesis.

We first sought to define the minimal length of cyanophycin that can serve as a primer for CphA1 enzymes (Fig. 3.1d). *Tm*CphA1 could not perform cyanophycin synthesis in the presence of  $\beta$ -Asp-Arg, but synthesis was observed with ( $\beta$ -Asp-Arg)<sub>2</sub>, and the observed rate increases progressively in reactions with ( $\beta$ -Asp-Arg)<sub>3</sub> and ( $\beta$ -Asp-Arg)<sub>4</sub> (Fig. 3.1d). We then tested ( $\beta$ -Asp-Arg)-Asp, the CphA1 biosynthetic intermediate between  $\beta$ -Asp-Arg and ( $\beta$ -Asp-Arg)<sub>2</sub>, and saw similar activity as with ( $\beta$ -Asp-Arg)<sub>2</sub> (Fig. 3.1d). Assays with the primer-independent *Su*CphA1 showed similar results: Addition of ( $\beta$ -Asp-Arg)-Asp and longer fragments shortens the lag phase, and (Asp-Arg)<sub>3</sub> and ( $\beta$ -Asp-Arg)<sub>4</sub> give the highest rates of synthesis (Supplementary Fig. 3.1b). Together, these results define ( $\beta$ -Asp-Arg)-Asp as the minimal primer for CphA1.

### 3.3.2. CphA1 primer independence does not depend on the G or M domain activity

We next sought to identify the source of primer independence by comparing primer-independent *Su*CphA1 to primer-dependent *Tm*CphA1. We hypothesized that affinity differences for short cyanophycin segments at the G and/or M domain active sites might dictate primer (in)dependence<sup>13</sup>. *Su*CphA1 has several polar residues at the G and M domain active sites that could increase its affinity to nascent cyanophycin segments relative to *Tm*CphA1, which has hydrophobic residues at the analogous positions<sup>2,98</sup> (Supplementary Fig. 3.2a,b). However, experiments with 15 different mutants of *Tm*CphA1 in which one or more of these *Su*CphA1 hydrophilic residues were introduced into *Tm*CphA1's G domain, M domain, or G and M domains did not lead to primer independent activity (Supplementary Fig. 3.2c).

We then reasoned that if the G or M domains of *Su*CphA1 were individually responsible for primer independence, providing the necessary domain *in trans* to a reaction including *Tm*CphA1 would lead to primer independent synthesis. Therefore, we prepared *Su*CphA1 with inactivating mutations<sup>2</sup> in the G domain (H267A), the M domain (D585A H586A), or in both (H267A; D585A H586A) (Supplementary Fig. 3.2a,b). To our surprise, all three of these constructs enabled robust primer-independent activity when added to *Tm*CphA1 (Fig. 3.2a). This



**Figure 3.2. The N domain of CphA1 is responsible for primer-independent activity.** (a)

Primer-independent activity of *TmCphA1* in the presence of *SuCphA1* constructs with inactivating mutation as the G and/or M domains<sup>2</sup>. Active G and M domains are not required for activity, suggesting a different part of *SuCphA1* is responsible for primer-independence. n=4 independent experiments. Data are presented as individual measurements and mean value, error bars represent SD values. (b) *SuCphA1* N domain confers primer independence. *TmCphA1* is inactive in the absence of primers. A chimera of *TmCphA1* G and M domains and *SuCphA1* N domain (*TmCphA1<sub>SuN</sub>*) is active. Reactions including *TmCphA1* and 5  $\mu$ M of a construct of the extruded N domain of *SuCphA1* harboring solubilizing mutations Y14S and I17T shows cyanophycin synthesis in the absence of primers. Note that the extruded *SuCphA1* N domain was completely insoluble before introducing mutations into a hydrophobic loop that, in the intact enzyme, interacts with the M domain (Supplementary Fig. 2d). n=4 independent experiments. Data are presented as individual measurements and mean value, error bars represent SD values. (c) Weblogo<sup>8</sup> diagram showing the conserved CX<sub>19</sub>HxxEH motif. This Weblogo was constructed from sequence alignments of all CphA1 enzymes generated by ClustalOmega<sup>11</sup>, and excludes cyanophycin synthetase 2 (CphA2) sequences. CphA2s are specialized cyanobacterial enzymes that polymerize  $\beta$ -Asp-Arg dipeptides recovered from degraded cyanophycin<sup>12,13</sup>. CphA2 N domains share low sequence identity to CphA1 N domains and the N domain active site motif is absent from CphA2 sequences. (d) The N domain of *SuCphA1* colored by per-residue conservation. The residues around the CX<sub>19</sub>HxxEH motif are conserved (purple). The conservation was calculated by ConSurf<sup>6</sup> using 500 randomly chosen CphA1 sequences. (e) Activity rate and lag time (time until Vmax is reached) of *SuCphA1* N domain mutants without added primer. All mutants except R100A displayed varying levels of reduced activity rate and longer lag phases than the WT enzyme. E82Q was inactive and its lag time is not shown. n=4 independent experiments. Data are presented as individual measurements and mean value, error bars represent SD values.

result suggested that *SuCphA1* has a third, previously unsuspected active site that is responsible for primer independence.

### 3.3.3. A cryptic N domain active site responsible for primer independence

Because our results indicated that the G and M domain active sites are not important for primer independence, we investigated whether the N domain is. We created a chimeric CphA1 (*TmCphA1<sub>SuN</sub>*) comprising the N domain of *SuCphA1* and the G and M domains of *TmCphA1*. Intriguingly, the *TmCphA1<sub>SuN</sub>* chimera displayed robust primer-independent activity, suggesting that the N domain is responsible for primer independence (Fig. 3.2b). To verify this conclusion, we added the isolated *SuCphA1* N domain *in trans* to a reaction including *TmCphA1*. Again, primer-independent activity was observed, proving that the N domain is vital for primer independence in CphA1 (Fig. 3.2b, Supplementary Fig. 3.2d).

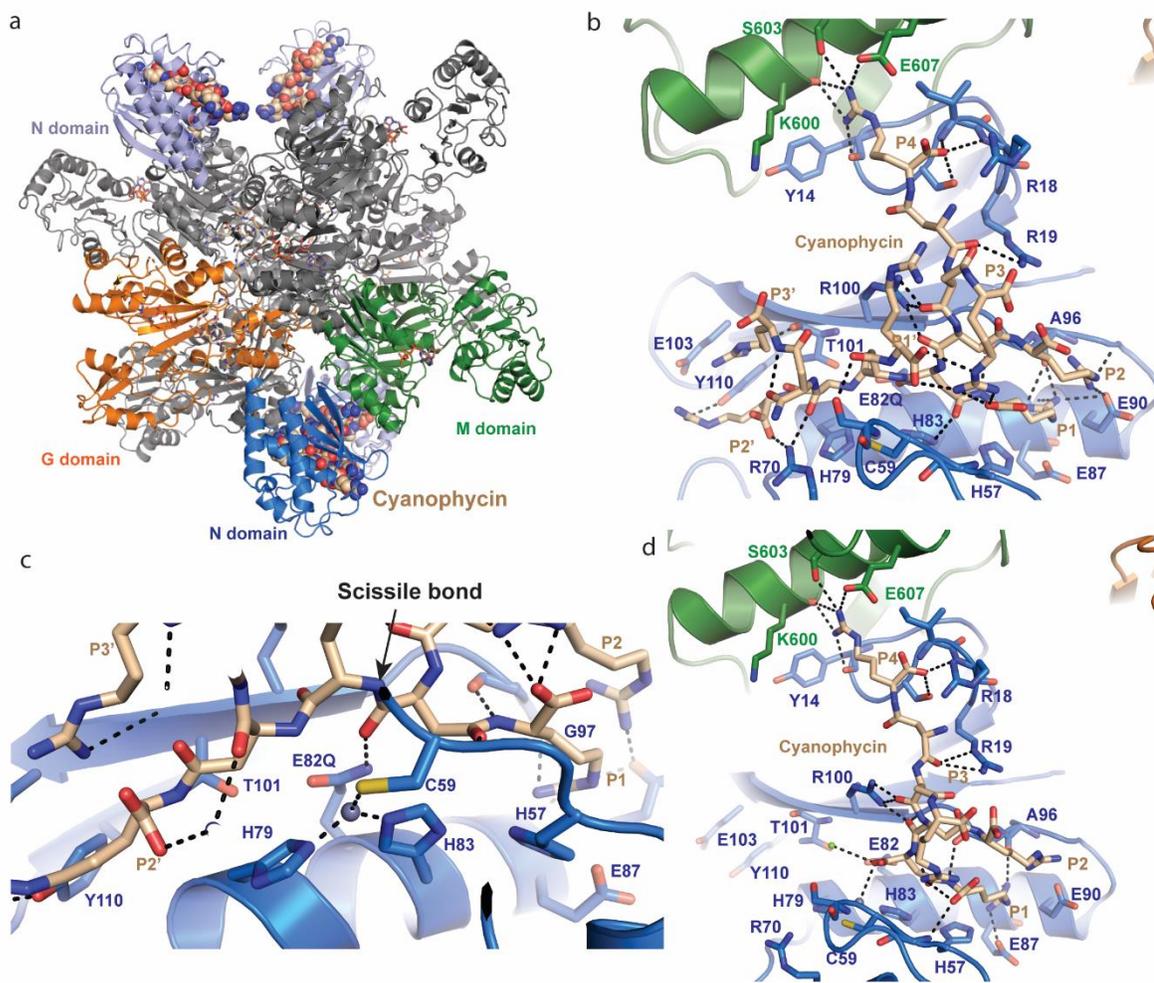
A catalytic role for the N domain has never been suggested before, so this discovery led us to re-examine the sequence conservation of N domains to search for a cryptic active site. No putative catalytic residues are 100% conserved, which is expected since some proportion of CphA1 enzymes will be primer dependent like *TmCphA1*. However, an HxxEH motif<sup>162</sup> can be seen by careful inspection of sequence alignments (Fig. 3.2c, Supplementary Fig. 3.3a). The N domain of our existing *SuCphA1* structures<sup>2</sup> shows the motif residues H79, E83 and H83 cluster together with C59, and are surrounded by several conserved arginine and histidine residues (Fig. 3.2d). All four residues of the C<sub>X19</sub>HxxEH motif are present in the primer-independent *SuCphA1* and none of the four are present in the primer-dependent *TmCphA1* (Supplementary Fig. 3.3b). Of CphA1s with a non-redundancy of 70% (nr70), the C<sub>X19</sub>HxxEH motif is fully present in 83% of sequences. Mutations in and near this putative N domain active site (H57A, C59A, R70A, H79A, E82Q<sup>163</sup> and R100A) did not greatly affect primer-dependent activity (Supplementary Fig. 3.3c), but all except R100A reduced or abolished primer-independent activity (Fig. 3.2e), confirming that this N domain active site is responsible for primer independence. Notably, HxxEH is the active site motif for inverted zinc metallopeptidases<sup>162</sup>, suggesting the N domain may have peptidase activity important for primer generation from cyanophycin polymer.

### 3.3.4. The structural basis for the catalytic activity of the N domain

To structurally characterize this cryptic N domain active site and its binding to cyanophycin, we turned to cryo-EM. We determined a structure of the inactivated<sup>163</sup> *SuCphA1*(E82Q) in complex with (β-Asp-Arg)<sub>16</sub> at 2.7 Å resolution by cryo-EM (Fig. 3.3a,b, Supplementary Fig. 3.4a,b, Supplementary Table 3.1). The maps showed clear signal for a chain

of seven  $\beta$ -Asp-Arg dipeptide residues centered on the conserved region of the N domain. This region harbors a metallopeptidase-like active site<sup>163</sup> (Fig. 3.3c): H79 and H83 from the active site helix (residues 77-92) ligate an ion. We have tentatively assigned this ion as  $Zn^{2+}$  because it is by far the most abundant metal detected in inductively coupled plasma mass spectrometry (ICP-MS) analyses of *Su*CphA1 (and not of *Tm*CphA1; Supplementary Table 3.2), and because of similarities to inverted zinc metallopeptidase<sup>162</sup> and peptide deformylase<sup>164</sup> active sites, both of which can bind  $Zn^{2+}$  tightly. The two histidine side chains bind  $Zn^{2+}$  with their  $N_\epsilon$  atoms. H83, which forms  $\pi$ -stacking interactions with the conserved F67, also forms a hydrogen bond network with the conserved H57 and E87, stabilizing the tautomeric form in which its  $N_\epsilon$  lone pair electrons face the  $Zn^{2+}$  site<sup>165</sup>. C59, present in a loop region, serves as the third metal ligand, a role typically played by a glutamate, aspartate or histidine from the metallopeptidase “glutamate helix”<sup>163</sup>. Q82 (taking the place of the general base E82 in this *Su*CphA1(E82Q) construct) sits above the metal-binding histidines, as in metallopeptidases. The cyanophycin polymer makes an extensive hydrogen bonding network with itself and *Su*CphA1 residues (Fig. 3.3b), including the backbones of Y14, C59, A96, G97 and T101, and the side chains of E90, R70, Y110, S60 and M domain residues S603 and E607. These interactions position four visualized dipeptides upstream of the zinc ion and three downstream. The main chain carboxyl oxygen of the fourth dipeptide is 2.6 Å from Q82 and 3.6 Å from the  $Zn^{2+}$  ion, in good pre-cleavage position.

We had also previously calculated a cryo-EM map of wildtype (WT) *Su*CphA1 in the presence of  $(\beta\text{-Asp-Arg})_{16}$ <sup>2</sup>. The new results described here led us to re-examine it. Signal consistent with cyanophycin bound to the N domain is visible in that map, although not quite as strong as that seen with *Su*CphA1(E82Q) (Fig. 3.3d, Supplementary Fig. 3.4c, Supplementary Table 3.1). This signal is not present in maps of *Su*CphA1 which was not incubated with cyanophycin segments<sup>2</sup>. Interestingly, we were able to fit the first four  $\beta$ -Asp-Arg dipeptide residues into this map in the same conformation as seen bound to *Su*CphA1(E82Q), but there is no signal for any dipeptide residues C-terminal to the N domain active site. The C-terminal Asp carboxyl group is positioned directly next to the  $Zn^{2+}$  ion, indicating this represents an N domain product complex, derived from *in situ* cleavage of  $(\beta\text{-Asp-Arg})_{16}$ .



**Figure 3.3. The structure of SuCphA1 N domain with bound cyanophycin.** (a) Structure of *Su*CphA1 in complex with cyanophycin substrate in both the G domain active site (sticks) and N domain active site (spheres). (b) The *Su*CphA1 N domain in complex with  $(\beta\text{-Asp-Arg})_{12}$  as a substrate. Seven dipeptide residues are visible. Polymer binding residues and their interactions are highlighted. P4, P3, P2, P1, P1', P2', P3' denote  $\beta\text{-Asp-Arg}$  dipeptides numbered relative the cleavage point. (c) Close up view of the structure of *Su*CphA1 in complex with cyanophycin substrate. (d) The structure of *Su*CphA1 in complex with *in situ* cleaved cyanophycin. Four dipeptide residues are visible. Polymer binding residues and their interactions are highlighted.

### 3.3.5. The N domain cleaves cyanophycin into primers

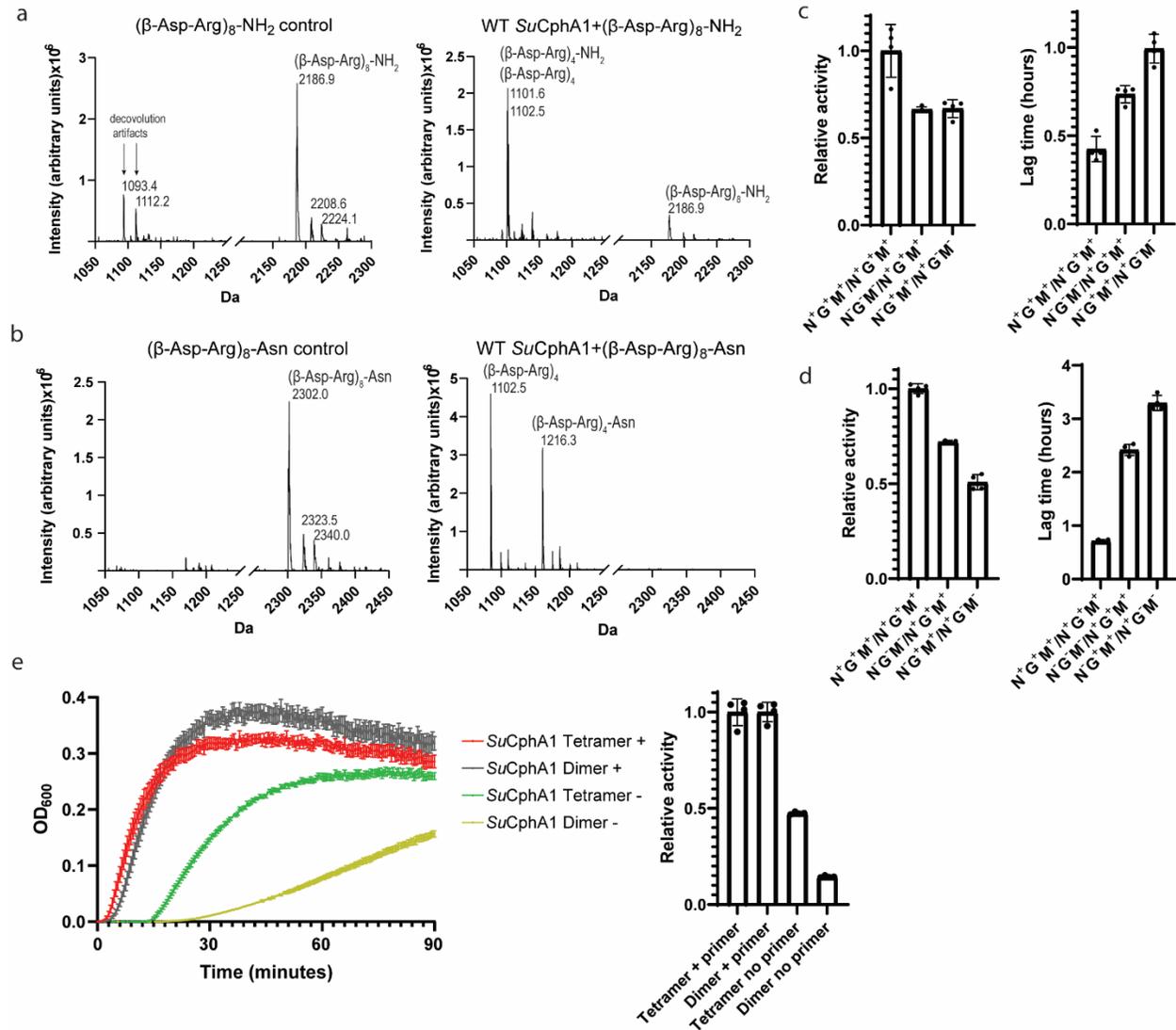
The structures suggest that the N domain possesses endo-cyanophycinase activity. To directly observe this activity, we incubated purified cyanophycin with *Su*CphA1 and examined the

reaction with SDS-PAGE (Supplementary Fig. 3.5a, 3.7). A slow but clear decrease in cyanophycin is observed over several days, especially in the molecular weight range of ~15-20 kDa. We next performed mass spectrometry-based cyanophycin cleavage assays. Incubation of *Su*CphA1 with  $(\beta\text{-Asp-Arg})_8\text{-NH}_2$  over several hours led to the gradual formation of species with mass values corresponding to  $(\beta\text{-Asp-Arg})_4\text{-NH}_2$  and  $(\beta\text{-Asp-Arg})_4$  (Fig. 3.4a). Likewise,  $(\beta\text{-Asp-Arg})_8\text{-Asn}$  is converted by *Su*CphA1 to species with masses corresponding to  $(\beta\text{-Asp-Arg})_4\text{-Asn}$  and  $(\beta\text{-Asp-Arg})_4$  (Fig. 3.4b). Control reactions with primer-dependent *Tm*CphA1 or the N domain *Su*CphA1(E82Q) variant with  $(\beta\text{-Asp-Arg})_8\text{-NH}_2$  or  $(\beta\text{-Asp-Arg})_8\text{-Asn}$  did not result in the appearance of the product peaks (Supplementary Fig. 3.5b). *Su*CphA1 cleaved  $(\beta\text{-Asp-Arg})_{12}$  into major products  $(\beta\text{-Asp-Arg})_8$  and  $(\beta\text{-Asp-Arg})_4$ , and minor products  $(\beta\text{-Asp-Arg})_5$  and  $(\beta\text{-Asp-Arg})_7$  (Supplementary Fig. 3.5c,d,e). Thus, the N domain of *Su*CphA1 is indeed a cryptic primer-generating endo-cyanophycinase that possesses a low catalytic rate and, at least with the cyanophycin segments used in these experiments, preferentially yields  $(\beta\text{-Asp-Arg})_4$  fragments as products.

The relative geometry of the CphA1 active sites is important for cyanophycin biosynthesis<sup>2</sup>, so we interrogated whether it is also important for primer-generating cleavage. We combined the W672A mutation (which forces *Su*CphA1 to be dimeric instead of tetrameric<sup>2</sup>) with mutations that abolish activity of each active site (E82Q = N<sup>-</sup>; H267A = G<sup>-</sup>; D585A H586A = M<sup>-</sup>) and used orthogonal affinity tags to purify different *Su*CphA1 heterodimers<sup>2</sup>. *Su*CphA1 N<sup>+</sup>G<sup>+</sup>M<sup>+</sup>/N<sup>-</sup>G<sup>-</sup>M<sup>-</sup> (with all wildtype active sites on the same protomer) and *Su*CphA1 N<sup>-</sup>G<sup>+</sup>M<sup>+</sup>/N<sup>+</sup>G<sup>-</sup>M<sup>-</sup> (with the wildtype N domain on the opposite protomer as wildtype G and M domains) have similar activity in the presence of primer (Fig. 3.4c). However, N<sup>-</sup>G<sup>+</sup>M<sup>+</sup>/N<sup>+</sup>G<sup>-</sup>M<sup>-</sup> is somewhat less active in the absence of primer (Fig. 3.4d). This suggests that the proximity of the hydrolytic active site to the biosynthetic active sites is beneficial for primer-independent activity. Interestingly, *Su*CphA1 dimers<sup>2,98</sup> exhibit the same primer-dependent activity rate as tetramers, but a lower primer-independent activity rate (Fig. 3.4e), indicating that the tetramer architecture of *Su*CphA1 is also beneficial for primer-independent activity.

### 3.3.6. Effect of primer independence on heterologous cyanophycin synthesis

Understanding the basis of primer dependence in cyanophycin biosynthesis and obtaining primer-independent and primer-dependent variants of the same CphA1 enzymes allowed us to examine the importance of primer availability for cyanophycin accumulation in a heterologous host. To that end, we separately expressed primer-independent *Su*CphA1 and primer-dependent *Su*CphA1(E82Q) in *E. coli* BL21(DE3) and quantified the amount of polymer produced *in vivo* by each variant. *E. coli* harboring wildtype *Su*CphA1 produced on average 2.3-fold more cyanophycin than the E82Q mutant, measured as milligrams of cyanophycin per liter of growth culture (Table 3.1). The total wet cell mass was 12% lower for cells expressing the WT enzyme, suggesting they divert more resources to cyanophycin synthesis from cell growth compared to those harboring the primer-dependent mutant. Similarly, in experiments with primer-dependent *Tm*CphA1 and the



**Figure 3.4. Catalytic activity of the N domain and cyanophycin synthesis by dimeric CphA1.** (a) Mass spectra traces of  $(\beta\text{-Asp-Arg})_8\text{-NH}_2$ , an 8mer cyanophycin segment in which the terminal carboxylate is replaced by an amide, before (left) and after (right) incubation with WT *Su*CphA1. After incubation with enzyme, the peak corresponding to  $(\beta\text{-Asp-Arg})_8$  (expected at 2187.0 Da) is reduced, and two peaks with sizes matching to  $(\beta\text{-Asp-Arg})_4$  (expected at 1102.5 Da) and  $(\beta\text{-Asp-Arg})_4\text{-NH}_2$  (expected at 1101.5 Da) appear. Peaks corresponding to  $\text{Na}^+$  and  $\text{K}^+$  adducts are also labeled. Peaks at 1093.4 Da and 1112.2 Da are deconvolution artifacts (Supplementary Fig. 3.5). (b) Mass spectra of  $(\beta\text{-Asp-Arg})_8\text{-Asn}$  before (left) and after (right) incubation with WT *Su*CphA1. After incubation with enzyme, the peak corresponding to  $(\beta\text{-Asp-Arg})_8\text{-Asn}$  (expected at 2301.1 Da) disappears and two peaks with sizes matching to  $(\beta\text{-Asp-Arg})_4$  (expected at 1102.5 Da) and  $(\beta\text{-Asp-Arg})_4\text{-Asn}$  (expected at 1216.6 Da) appear. (c,d) Activity rate and lag time of *Su*CphA1 dimer complementation assays with (c) and without (d) primer. In presence of exogenous primer, the two mutant combinations  $\text{G}^-\text{M}^-\text{N}^-/\text{G}^+\text{M}^+\text{N}^+$  and  $\text{G}^-\text{M}^+\text{N}^+/\text{G}^+\text{M}^-\text{N}^-$  display a similar activity rate, although the  $\text{G}^-\text{M}^+\text{N}^+/\text{G}^+\text{M}^-\text{N}^-$  combination displays a somewhat longer lag time. In the absence of primer, the  $\text{G}^-\text{M}^-\text{N}^-/\text{G}^+\text{M}^+\text{N}^+$  combination displays a somewhat higher activity rate and lower lag time.  $n=4$  independent experiments. Data are presented as individual measurements and mean value, error bars represent SD values. (e) Activity assays and activity rate comparison of WT *Su*CphA1 (tetramer) and the W672A mutant (dimer) with (+) and without (-) primer. The two enzymes display similar primer-dependent activity, but the dimer has lower primer-independent activity.  $n=4$  independent experiments. Data are presented as individual measurements and mean value, error bars

primer-independent chimera *Tm*CphA1<sub>SuN</sub>, the chimera produced 2-fold more cyanophycin than the primer-dependent WT enzyme, in a lower total wet cell mass (Table 3.1).

Enzyme	Cell pellet (mg)	Polymer (mg)	% of cell mass (w/w)
<i>Su</i> CphA1 WT	4730±160	413±35	8.7
<i>Su</i> CphA1 E82Q	5390±140	177±28	3.2
<i>Tm</i> CphA1 WT	5570±380	247±18	4.4
<i>Tm</i> CphA1 <sub>SuN</sub>	4880±170	486±52	9.9

**Table 3.1. Heterologous cyanophycin production in *E. coli*.**

### **3.4. Discussion**

The sequence, structure and activity data all indicate that the N domain of CphA1 has cryptic metallopeptidases activity. The endo-cyanophycinase activity of the N domain presents an explanation for the primer dependence and independence in CphA1 enzymes: We propose that all CphA1 enzymes possess very low levels of true primer-independent activity for the first steps of cyanophycin synthesis, e.g., ligating Asp and Arg to  $\beta$ -Asp-Arg, and ligating  $\beta$ -Asp-Arg and Asp to ( $\beta$ -Asp-Arg)-Asp. In the next steps of elongation of these intermediates, the rate of polymerization increases and a long chain of cyanophycin is made. In CphA1 enzymes with N domain metallopeptidase activity, the chain is cleaved to generate cyanophycin segments such as ( $\beta$ -Asp-Arg)<sub>4</sub> that act efficiently as primers. This leads to more long chains and more primers, and thus rapid accumulation of cyanophycin after the initial lag phase we observe. CphA1 enzymes that lack active N domains make long cyanophycin chains as well, but because they are limited by the very slow initial rates in absence of primers, they make so few as to be undetectable in light-scattering or ATP hydrolysis assays<sup>3,98,99,106</sup> (Supplementary Fig. 3.1a).

*In vivo*, these CphA1 enzymes that do not have N domains active sites likely use remnant strands of cyanophycin left over from the last round of catabolism, or other cellular small molecules<sup>104</sup> as primer. The maximum rates we observe *in vitro* indicate that polymerization of cyanophycin is several fold faster than hydrolytic cleavage, but it is difficult to relate these rates to the situation *in vivo*, where cellular conditions are not constant and availability of cyanophycin chains will change as molecules aggregate into granules. However, the accumulation of large amounts of cyanophycin in native bacteria and heterologous hosts clearly indicates the relative *in vivo* rates of polymerization and hydrolytic primer production are well tuned for cyanophycin biosynthesis.

Intriguingly, our experiments with dimeric CphA1 mutants (Fig. 3.4) suggest that a nascent chain could be polymerized at one end while being cleaved near the other end. In the absence of exogenous primer, mutant heterodimeric *Su*CphA1 displays higher synthesis rates when all three intact active sites are in the same protomer, hinting the increased rate is from cleavage *in cis*. A cyanophycin chain being elongated at its C terminus by the G and M domains with soft anchoring on the N domain helices  $\alpha_a$  and  $\alpha_b$  could intermittently wrap around to be cleaved at that same N domain's active site (Supplementary Fig. 3.6a). Similarly, geometrical considerations can

rationalize why dimeric *Su*CphA1 is less active than tetrameric *Su*CphA1 in the absence of exogenous primer: In the tetramer, two N domain active sites face each other and are 55 Å apart (protomers A and C; Supplementary Fig. 3.6a). It is possible that after cleavage, the new N-terminus of a cyanophycin chain experiences increased local concentration of N domain active sites, facilitating binding and increasing the rate of hydrolytic primer production. We note that the primer-dependent *Tm*CphA1 has a different tetramer architecture<sup>2</sup>, in which the equivalent N domain positions are ~80 Å apart (Supplementary Fig. 3.6b).

The N domain appears distantly related to the M16 peptidase family<sup>166,167</sup>, which includes endopeptidases such as pitrilysin<sup>163</sup> and insulin degrading enzyme<sup>168</sup>. The family is also known as inverzincins<sup>169</sup> because the active site motif HxxEH is inverted from the HExxH of the canonical mononuclear metallopeptidase motif. Active CphA1 N domains share the inverzincin<sup>169</sup> HxxEH motif, as well as three structural elements: an active site helix, an adjacent β sheet and the “backing helix”<sup>169</sup> (Supplementary Fig. 3.6c). The CphA1 backing helix doubles as the α<sub>a</sub> helix, which binds nascent cyanophycin chains through its surface-exposed side during biosynthesis<sup>2</sup>. Substrate binding in CphA1 and pitrilysin<sup>163</sup> is similar, with the scissile peptide bond in analogous positions (Supplementary Fig. 3.6d). However, CphA1 N domains are clearly distinct from known inverzincins: CphA1 N domains are much smaller (~160 residues vs. up to 1000 residues), the structural similarity is modest and confined to the region around the active site, and crucially, the third metal-binding residue in CphA1 is a Cys upstream in sequence from the histidines in a C-H-H metal binding triad, rather than a downstream Asp, Glu or His in a M16 peptidase H-H-D/E/H triad<sup>165,169,170</sup>.

Two or more Cys ligands are common in structural Zn<sup>2+</sup>-binding motifs, but Cys as a Zn<sup>2+</sup> ligand in an active metallopeptidase is rare<sup>169,171-173</sup>. The best-known example for a metallopeptidase with C-H-H metal coordinating residues is peptide deformylase (PDF), a ubiquitous enzyme responsible for deformylation of N-terminal fMet residues<sup>164</sup>. CphA1 N domains and PDFs share very little structural similarity, and their active site helices are in opposite orientations (Supplementary Fig. 3.6e), but the geometry of the metal binding residues of the two enzymes are remarkably similar (Supplementary Fig. 3.6f)<sup>174</sup>. PDF can bind Zn<sup>2+</sup> tightly, but has a lower catalytic rate when bound with zinc than when bound with cobalt, nickel<sup>175</sup> or iron<sup>176</sup>. The peptidase activity of CphA1 needs to be properly tuned so hydrolysis can generate primers but not

efficiently compete with polymerization in this biosynthetic enzyme tasked with making long cyanophycin chains for storage. Because the biosynthetic and hydrolytic activities are both encoded into the same enzyme, the balance of these activities cannot be regulated by protein expression levels. Other features of the N domain active site which may temper rate of hydrolysis are the lack of a residue for transition state stabilization (inverted zinc metallopeptidase<sup>162</sup> have a Tyr; peptide deformylase<sup>177</sup> have a Gln), or the lack of an active site residue accepting a hydrogen bond from N<sub>δ</sub> of H79. The latter would promote the N<sub>ε</sub> lone-pair electrons facing the metal binding site which can be important for activity<sup>165,178</sup>. H83 has such an interaction, but it is common for it to be seen for both active site histidines<sup>165,178</sup>.

Long strands of cyanophycin precipitate into granules for storage<sup>15,96</sup>. This precipitation may also serve to sequester cyanophycin from CphA1's hydrolytic activity, since these chains of cyanophycin are largely stable both *in vivo* and *in vitro* in the presence of CphA1s which have active N domains<sup>96,98,105</sup> (Supplementary Fig. 3.5a). Similarly, sequestration from the polymerizing G and M domains by precipitation into multistrand granules may be involved in determining cyanophycin chain length, which varies with CphA1, host and other factors<sup>86,96,98,106,179</sup>. The exo-cyanophycinase CphB has a high V<sub>max</sub> and an active site that is shallow and accessible<sup>121,124</sup>, allowing rapid degradation of strands in granules when needed.

The *in vivo* experiments we performed show that primer dependence can be a limiting factor for cyanophycin production in heterologous hosts. This understanding can help guide future efforts for more efficient polymer production *in vivo*, for example by prioritizing primer-independent enzymes. With ~80% of CphA1s having the CX<sub>19</sub>HXXEH motif, primer-independent CphA1s are more common than previously realized, and primer independence can be conferred by using N domain chimeras like *TmCphA1*<sub>SuN</sub>. Of the four constructs we assayed here, the chimera that introduces an active N domain into a primer-dependent CphA1 produced the highest cyanophycin yields.

Although cryptic active sites are not unheard of<sup>180</sup>, it is unusual to discover a cryptic active site in an enzyme that has been studied for decades. However, it was completely unexpected that cyanophycin synthetase would generate its own primers or have hydrolytic activity for any reason, given its biosynthetic role. Furthermore, the active site motif was obscured in sequence alignments (Fig. 3.2c, Supplementary Fig. 3.3a) by the ~20% of primer-dependent CphA1 enzymes that do

not have the active site, and the structural similarity to metalloproteases is so modest that they do not appear in the top 100 results of DALI<sup>181</sup> searches. Only the observation that the N domain confers primer independence led to the discovery of the N domain cyanophycinase site, thus showing that evolution combined three different enzymes into one elegant macromolecular machine. We are not aware of any other polymerase that has a dedicated active site to create primers needed for its biosynthetic cycle, making CphA1 a truly remarkable, multifunctional enzyme.

### **3.5. Methods**

#### **3.5.1. Cloning, protein expression and purification**

The genes encoding *Su*CphA1 (protein WP\_028947105.1) and *Tm*CphA1 (protein WP\_004925893.1) were cloned into pJ411-derived plasmids in a previous study<sup>2</sup>. Point mutants and (sub)domain chimeras used in this study were generated by transforming DH5- $\alpha$  *E. coli* cells with PCR fragments containing overlapping ends. Phusion® DNA polymerase (New England Biolabs) was used for all PCR reactions. Sequences of DNA primers used for cloning are listed in Supplementary Table 3.3. Proteins were expressed in *E. coli* BL21(DE3). Cells were grown in LB media supplemented with 100  $\mu$ g/ml kanamycin at 37 °C until OD600 reached ~0.5. The growth temperature was then reduced to 22 °C and protein expression induced with 0.25 mM isopropyl  $\beta$ -d-1-thiogalactopyranoside (IPTG) for ~20 hours. Following harvesting by centrifugation, the cells were resuspended in buffer A (250 mM NaCl, 50 mM Tris pH 8.0, 10 mM imidazole, 2 mM  $\beta$ -mercaptoethanol) supplemented with a few crystals of lysozyme and DNase I, and lysed by sonication at 0 °C. The lysate was clarified by centrifugation at 40,000g, then loaded onto a HisTrap HP column (Cytiva), washed extensively with buffer B (buffer A with 30 mM imidazole) and eluted with buffer C (buffer A with 250 mM imidazole). For structural studies, the proteins were incubated with TEV protease for removal of the affinity tag while being dialyzed overnight against buffer D (250 mM NaCl, 20 mM Tris pH 8, 5 mM  $\beta$ -mercaptoethanol) and then applied again to a HisTrap column. Resulting samples were concentrated using Amicon centrifugation concentrators (EMD Millipore) and loaded onto a Superdex200 16/60 column (GE Healthcare) equilibrated in buffer E (100 mM NaCl, 20 mM Tris pH 8.0, 1 mM dithiothreitol). Following gel filtration, fractions with the highest purity were pooled and concentrated to ~20 mg/ml. Glycerol was added to a final concentration of 10% v/v, and the samples were flash frozen and stored at -80 °C until use.

For dimer complementation experiments, *Su*CphA1\_W672A carrying the desired active-site mutations was cloned into pCDF-derived plasmids with a C-terminal calmodulin binding protein (CBP) tag. *E. coli* BL21(DE3) cells were co-transformed with a pJ411-derived plasmids (for a His-tagged version) and a pCDF-derived plasmid and grown in LB media supplemented with 100 µg/ml kanamycin and 100 µg/ml spectinomycin as described above. All purification steps were similar to those already described up to the elution from the HisTrap HP column. Following elution, the protein was mixed with CaCl<sub>2</sub> to a final concentration of 2 mM and loaded onto a column of calmodulin-sepharose (Agilent) equilibrated with buffer F (250 mM NaCl, 50 mM Tris pH 8.0, 2 mM CaCl<sub>2</sub>, 2 mM β-mercaptoethanol), washed with buffer F and eluted with buffer G (250 mM NaCl, 50 mM Tris pH 8.0, 2 mM EGTA, 2 mM β-mercaptoethanol). The eluted protein was buffer exchanged into buffer E, concentrated and frozen.

### **3.5.2. Cryo-EM grid preparation, data collection and processing**

*Su*CphA1(E82Q) (3.5 mg/ml) was mixed with 2 mM ATP, 10 mM MgCl<sub>2</sub>, 1 mM (β-Asp-Arg)<sub>16</sub> and 0.09% octyl β-D-glucopyranoside. Three microliters of this sample were applied to glow-discharged C-flat 300 mesh 1.2/1.3 Cu holey carbon grids, blotted for 3 seconds at 4 °C and 90% humidity using a Vitrobot IV (FEI) and plunge-frozen into liquid ethane. Data were collected at the McGill Facility for EM Research (FEMR) using a FEI Titan Krios TEM operating at 300 kV with a Gatan K3 DED and a Gatan GIF BioQuantum LS. Movies were collected in counting mode using SerialEM, with a total dose of 60 e/Å<sup>2</sup> over 30 frames and a set defocus range of -1.0 to -2.0 µm at a nominal magnification of 105,000, resulting in a pixel size of 0.855 Å<sup>2</sup>. Micrographs were motion corrected using Relion3.1<sup>155</sup>. The motion-corrected micrographs were imported to CryoSPARC2<sup>148</sup> for patch-CTF estimation, particle picking and several rounds of 2D and 3D classification to remove junk particles. The particles were then exported to Relion3.1<sup>155</sup> for 3D refinement followed by two rounds of Bayesian polishing and CTF refinement. The polished particles were then exported to CryoSPARC2 and 3D refined using homogenous refinement with defocus and high-order aberrations refinement. Local resolution estimation followed by local filtering was then performed in CryoSPARC2, and the locally filtered map used for model building. The map of WT *Su*CphA1 with ATP and (β-Asp-Arg)<sub>16</sub> was calculated in a previous study<sup>2</sup> and deposited as EMDB-23326.

### 3.5.3. Structure refinement

The previously determined structure of *Su*CphA1 with ATP (PDB 7LG5) was used as a starting model for the two structures. The model was manually docked into the maps using UCSF Chimera<sup>182</sup> and refined using Rosetta<sup>152</sup>. Further refinement of the protein and positioning of the substrate molecules were done manually in Coot<sup>153</sup>, using the model validation feature in CCP-EM 1.4<sup>183</sup> for guidance. Conformational constraints of substrates were generated in CCP4i2<sup>184</sup>. Figures were generated using PyMOL.

### 3.5.4. CphA1 activity assays

CphA1 activity was monitored by following scattering of light by cyanophycin at neutral pH as previously described<sup>2</sup>. Unless stated otherwise, reactions contained 700 nM purified CphA1, 100 mM HEPES pH 8.2, 20 mM KCl, 10 mM MgCl<sub>2</sub>, 2 mM each L-Asp and L-Arg, 4 mM ATP and 50  $\mu$ M synthetic cyanophycin primer as indicated. The reaction volume was 100  $\mu$ l and reactions were performed in quadruplicate. OD600 was monitored using a SpectraMax Paradigm spectrophotometer running SoftMax Pro 5.4.1 (Molecular Devices), with 5 second linear shaking between reads. Data were analyzed using GraphPad Prism. To calculate maximal rates, the maximum of the first derivative of each OD600 curve was taken. The derivatives curves were smoothed with a 2nd order polynomial to reduce noise in measurements. Lag time to maximal rate is the time when the first derivative reaches its maximal value.

### 3.5.5. Cyanophycin purification

*E. coli* BL21(DE3) cells were transformed with the same plasmids used for protein expression and plated on LB plates supplemented with 50  $\mu$ g/ml kanamycin. The next day, single colonies were picked and used to inoculate 10 ml of LB supplemented with 100  $\mu$ g/ml kanamycin. The starter culture was grown overnight with shaking at 37 °C and then used to inoculate 1 L of LB supplemented with 100  $\mu$ g/ml kanamycin. One-liter cultures were grown with shaking at 37 °C until OD600 reached 0.5, and then the temperature was reduced to 25 °C. After 1 hour, protein expression was induced with 0.25 mM IPTG and the cultures grown for another 20 hours. The next day, cells were harvested by centrifugation, resuspended in 1 ml ddH<sub>2</sub>O for every 0.2 gram of cell pellet, and lysed by sonication at room temperature. The lysates were acidified to pH 0.9 using concentrated HCl and clarified by centrifugation at 3500g for 20 minutes. The pH of the clarified lysate was then neutralized using 2 M NaOH. Following centrifugation at 3500g for 10

minutes, the pellets contained insoluble cyanophycin and the lysate contained soluble cyanophycin. The pellets were resuspended in 0.1 M HCl, centrifuged at 3500g for 10 minutes and the resulting pellets were discarded. The pH of the liquid phase was neutralized with 2 M NaOH and centrifuged for 10 minutes at 3500g. The resulting pellets, consisting of purified insoluble cyanophycin, were lyophilized and weighed. The lysate containing soluble cyanophycin was mixed with 1 volume of 95% EtOH and centrifuged for 10 minutes at 3500g. The resulting pellet was resuspended in ddH<sub>2</sub>O, mixed with 1 volume of 95% EtOH, and centrifuged for 10 minutes at 3500g. The resulting pellet, consisting of purified soluble cyanophycin, was lyophilized and weighed. The reported amounts of purified cyanophycin are the sum of the soluble and insoluble polymer from each culture.

### **3.5.6. MS analysis of cyanophycin degradation**

Synthetic cyanophycin segments were digested in 100  $\mu$ l reactions containing 1  $\mu$ M purified CphA1, 100 mM (NH<sub>4</sub>)<sub>2</sub>CO<sub>3</sub>, 20 mM KCl and 5 mM MgCl<sub>2</sub>, and 2 mM cyanophycin segments. Samples of 10  $\mu$ l were taken at specific time points and diluted into 90  $\mu$ l of 100 mM (NH<sub>4</sub>)<sub>2</sub>CO<sub>3</sub>, then directly injected for 2 minutes at 40  $\mu$ l/min into a Bruker amaZon speed ETD ion trap mass spectrometer operating at positive ionization mode. The resulting spectra were deconvoluted using the max entropy method.

### **3.5.7. SDS-PAGE analysis of cyanophycin degradation**

Reactions contained 20  $\mu$ M purified CphA1, 50 mM (NH<sub>4</sub>)<sub>2</sub>CO<sub>3</sub>, 20 mM KCl and 5 mM MgCl<sub>2</sub>, 5 mg/ml cyanophycin purified from *E. coli*, 0.01% NaN<sub>3</sub> and 200  $\mu$ M phenylmethylsulfonyl fluoride. The reactions were incubated at room temperature. Samples of 20  $\mu$ l were mixed with 10  $\mu$ l of 5x loading buffer, boiled for 1 minute and analyzed on a 17% polyacrylamide gel.

### **3.5.8. Synthesis of cyanophycin segments**

$\beta$ -Asp-Arg dipeptides were made from purified cyanophycin made *in vitro* in a primer-independent reaction by *Su*CphA1. The produced polymer was isolated by centrifugation at 3500g for 10 minutes, washed with ddH<sub>2</sub>O and resuspended in 50 mM (NH<sub>4</sub>)<sub>2</sub>CO<sub>3</sub>. The polymer suspension was digested with purified cyanophycinase from *Synechocystis sp.* PCC6803<sup>124</sup> until the suspension became clear, then filtered using a 3 kDa molecular weight cut-off Amicon centrifugation concentrator (EMD Millipore) and lyophilized.

All other cyanophycin segments were prepared by manual Fmoc solid-phase peptide synthesis (SPPS) as previously described<sup>2,65,159</sup>. Briefly,  $(\beta\text{-Asp-Arg})_n$ , where  $n = 2, 3, 4, 8$  or  $16$ , were synthesized on an HMPB-ChemMatrix resin (Biotage) on a  $0.01 - 0.03$  mmol scale using Fmoc- $(\beta\text{-Asp-Arg})(OtBu,Pbf)\text{-OH}$  as the building block. Fmoc groups on the growing chains were removed with piperidine in DMF, and coupling was carried out with HATU/DIPEA in DMF. Cleavage of the peptides from the resin and removal of the *OtBu* and *Pbf* protecting groups were achieved with TFA- $\text{H}_2\text{O}$ - $i\text{PrSiH}$  (95:2.5:2.5).  $(\beta\text{-Asp-Arg})\text{-Asp}$  and  $(\beta\text{-Asp-Arg})_n\text{-Asn}$  ( $n = 4$  and  $8$ ) were prepared analogously, but using Fmoc-Asp (*OtBu*)-OH or Fmoc-Asn(*Trt*)-OH rather than Fmoc- $(\beta\text{-Asp-Arg})(OtBu,Pbf)\text{-OH}$  for the first coupling to the resin; the allyl protecting group was removed with Pd( $\text{PPh}_3$ )<sub>4</sub> and PhSiH<sub>3</sub> in the final deprotection step. The C-terminal amides  $(\beta\text{-Asp-Arg})_n\text{-NH}_2$  ( $n = 4$  and  $8$ ) were synthesized by manual Fmoc-SPPS on an N-alkylated PAL resin (Bachem); couplings and peptide release from the resin were otherwise the same as for the other derivatives. All products were purified by reverse phase preparative HPLC and analyzed by high-resolution mass spectrometry<sup>2</sup> (HRMS) (Supplementary Table 3.4).

### **3.5.9. Metal analysis**

For metal analysis, purified protein samples of *Su*CphA1, its extruded N domain and *Tm*CphA1 were buffer-exchanged into 100 mM  $(\text{NH}_4)_2\text{CO}_3$  by performing gel filtration with a Superdex S200 10/300 column equilibrated with that buffer. Protein containing fractions were concentrated to 100  $\mu\text{M}$  and analyzed by ICP-MS at the Center for Applied Isotope Studies, University of Georgia. A sample of the buffer eluted from the column was used as a control.

### **3.6. Data availability**

The structural models and maps (Supplementary Fig. 3.4, Supplementary Table 1) generated in this study are available in the Protein Data Bank database under accession codes 7TXU and 7TXV and the Electron Microscopy Data Bank under accession code EMD-26161. The biochemical data (Fig. 3.1c, 3.1d, 3.2a, 3.2b, 3.2e, 3.4, Table 3.1, Supplementary Fig. 3.1, 3.2c, 3.3c, 3.5a, 3.5b, 3.5c, 3.5d, 3.5e, Supplementary Table 3.4) generated in this study are provided in the source data file.

### **3.7. Acknowledgements**

We thank all the members of the Schmeing lab for advice and ongoing discussions on this project, J.F. Trempe for advice with mass spectrometry, Christopher Thibodeaux and Kenneth Johnson for advice with data interpretation, N. Rogerson for proofreading, staff at McGill Facility of EM Research (Kaustuv Basu and Kelly Sears) for support during data collection and the Plasma Chemistry Laboratory at the Center for Applied Isotope Studies, University of Georgia for ICP-MS. The work (10.46936/10.25585/60001153) conducted by the U.S. Department of Energy Joint Genome Institute (<https://ror.org/04xm1d337>), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231. This work was funded by CIHR Project Grant 178084 and a Canada Research Chair to TMS, and the Schweizerischer Nationalfonds and ETH Zurich to DH.

### **3.8. Supplementary information**

#### **3.8.1. Supplementary tables**

**Supplementary Table 3.1. Cryo-EM data collection, refinement and validation statistics.**

	<i>Su</i> CphA1 E82Q + ATP + ( $\beta$ -Asp-Arg) <sub>16</sub> (EMD-26161) (PDB 7TXV)	<i>Su</i> CphA1 WT + ATP + ( $\beta$ -Asp-Arg) <sub>16</sub> (EMDB-23326*) (PDB 7TXU)
<b>Data collection and processing</b>		
Magnification	105,000x	
Voltage (kV)	300	
Electron exposure (e <sup>-</sup> /Å <sup>2</sup> )	60	
Defocus range (μm)	-1.0 to -2.0	
Pixel size (Å)	0.855	
Symmetry imposed	D2	
Initial particle images (no.)		
Final particle images (no.)	318,594	
Map resolution (Å) FSC threshold 0.143	2.7	
Map resolution range (Å)	2.4-5.5	
<b>Refinement</b>		

Model resolution (Å) FSC threshold 0.143	2.7	2.6
Model resolution range (Å)	2.4-5.5	2.4-9.0
Map sharpening <i>B</i> factor (Å <sup>2</sup> )	-106	-105
Model composition		
Non-hydrogen atoms	27656	27408
Protein residues	26584	26584
Ligands	1072	824
<i>B</i> factors (Å <sup>2</sup> )		
Protein	70	45
Ligand	61	40
R.m.s. deviations		
Bond lengths (Å)	0.022	0.020
Bond angles (°)	1.959	1.913
Validation		
MolProbity score	1.34	1.28
Clashscore	2.57	1.81
Poor rotamers (%)	0.70	0.14
Ramachandran plot		
Favored (%)	95.73	95.03
Allowed (%)	4.16	4.97
Disallowed (%)	0.12	0.0

\*The collection statistics for EMDB-23326 were previously reported<sup>2</sup>.

**Supplementary Table 3.2. ICP-MS analysis of *Su*CphA1, *Su*CphA1 N domain and a buffer control.** All units are in µg/kg. BQ = below limit of quantitation.

Sample	<sup>55</sup> Mn	<sup>56</sup> Fe	<sup>59</sup> Co	<sup>60</sup> Ni	<sup>65</sup> Cu	<sup>66</sup> Zn
<b>Buffer control</b>	BQ < 0.88	BQ < 26.1	BQ < 0.46	BQ < 2.05	4.48	BQ < 6.85
<b><i>Su</i>CphA1</b>	1.14	44.2	BQ < 0.46	424	989	5382
<b><i>Su</i> N domain</b>	BQ < 0.89	49.2	0.77	1501	322	3747
<b><i>Tm</i>CphA1</b>	BQ < 1.03	108	BQ < 0.70	82.7	218	109

**Supplementary Table 3.3. DNA primers used in this study for cloning.**

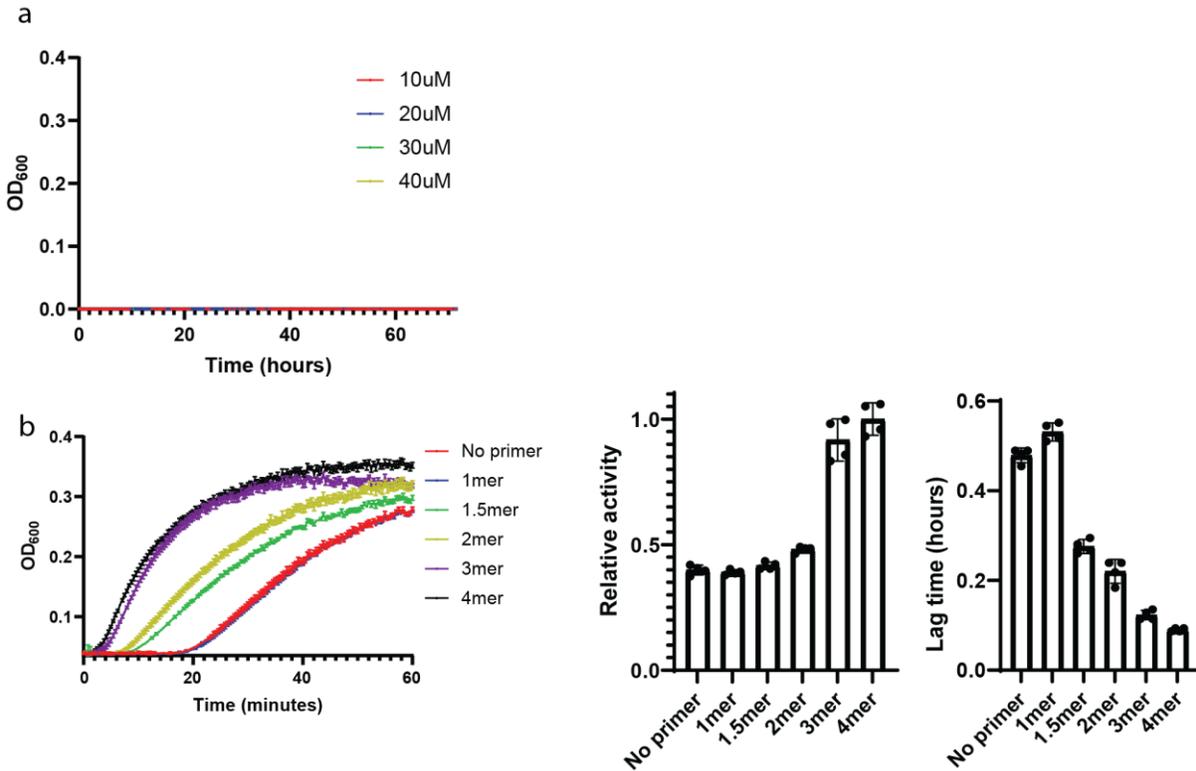
Sequence	Name
CATATGTTTTTACCTCCTTAAAAGTTAAAC	R. pBacIT
GAGAATTTGTACTTCCAAGGTCACC	F. pBacPtandem reverse2
TTAACTTTTAAGGAGGTAAAAACATATGaaaattcttaaaactctgactctccg	F. UTEX2470 cphA into pBacIT
TGGTGACCTTGGGAAGTACAAATTCTCaccaatgggattgeggacc	R. UTEX2470 cphA into pBacIT
gattaaaccctggatggcaacGCtggccggggcatcacc	F. UTEX2470 H267A
ggtgatgccccggccaGCgttccatccaggggttaac	R. UTEX2470 H267A
ctacctgtcaattttgagggttccGCgactctggggtggagcag	F. UTEX2470 W672A
ctgctccaccgcagagtcGCggaacctccaaaattgacaggtag	R. UTEX2470 W672A
gctgaatgtggcggcgCtGCcctgggctgggggatattg	F. UTEX2470 D585A H586A
caatatccccagccccaggGCaGccgcgccacattcagc	R. UTEX2470 D585A H586A
GTTAGTAGCAGCAGCGCGCCGTGTGccagccatgtgactgg	F. UTEX2470 omega into DSM23827
TAACATCTACAGCGGTCCCACCTGTgcttaggttagcgggtggc	R. UTEX2470 omega into DSM23827
GGTGGGACCGCTGTAGATG	F. DSM23827 change omega
ACGGCGCGCTGCTGC	R. DSM23827 change omega
GGGCGATCTCGGTCTTCTGTACCTcggggcaccaccatcc	F. 2470 lid into 23827
CGTTGACAACCAGTAAGCGATGGTCactgcctctagtaacgctc	R. 2470 lid into 23827
GACCATCGCTTACTGGTTG	F. 23827 change lid
AGGTACAGGAAGACCGAG	R. 23827 change lid
TCACTAATTCCATCGCAGACGGCCCcagggccgagttagcc	R. UTEX2470 CphA N into 23827 CphA
GGGCCGTCTGCGATGG	F. DSM23827 CphA change N
GATCCGCTCAGGATTACGCACGTTTaccacatcagcagaacag	F. UTEX2470 Mlid into 23827 CphA
GGTGACCTTGGGAAGTACAAATTCTCaccaatgggattgegg	R. UTEX2470 Mlid into 23827 CphA
AAACGTGCGTAATCCTGAG	R. 23827 CphA change Mlid
cttatatgggtGCtattgtcgagc	F. UTEX2470 CphA1 H79A
gctcgacaataGCacccatataag	R. UTEX2470 CphA1 H79A
gtcatattgtcgCgGCtgtggccctgg	F. UTEX2470 E82A H83A
ccagggccacaGCcGgacaatatgac	R. UTEX2470 E82A H83A
ggtcatattgtcCagcatgtggc	F. UTEX2470 E82Q
gccacatgtGgacaatatgacc	R. UTEX2470 E82Q
GGTGACCTTGGGAAGTACAAATTCTCtaaateccgcaaatcttcc	R. UTEX2470 N domain
ggccctaactCttggagtaCtcgagcgaag	F. UTEX2470 Y14S II7T
cttgctcgaGtactccaaGagttagggcc	R. UTEX2470 Y14S II7T
ggaacacttcGCctcgccggg	F. UTEX2470 C59A
cccggcgagGCgaagtgtcc	R. UTEX2470 C59A
gggtttggcGCcaccagggaaac	F. UTEX2470 R100A
gtttccctggtGCgcaaaaacc	R. UTEX2470 R100A
ggattttggaaGCggtaaaagaaggg	F. UTEX2470 R70A
cccttctttaccGCtccaaaaatcc	R. UTEX2470 R70A

ccctggtgaaGCcttctgctcg	F. UTEX2470 H57A
gcgagcagaagGCttccaccagg	R. UTEX2470 H57A
gggattttggaaGCggtaaaagaaggac	F. UTEX2470 R70A 2
gtccctctttaccGCttcaaaaatccc	R. UTEX2470 R70A 2
cctggtgaaGCcttctgctcg	F. UTEX2470 H57A 2
cgagcagaagGCttccaccagg	R. UTEX2470 H57A 2
aGAAAGGTGACAacACGGGCCCCGctgTCGGCTAAAATGGTTCTTCG	F. DSM23827 V516E M520N V524L
cagCGGGCCCGTgtTGTCACCTTTCtCGGTTACATTACCATCGATCAG	R. DSM23827 V516E M520N V524L
GCTGTTGAGATTGCAiCAGATAAGAATATGTG	F. DSM23827 CphA A204S
CACATATTCTTATCTGaTGCAATCTCAACAGC	R. DSM23827 CphA A204S
CGACTTCAGTGTTCgTGGCCCGAATACG	F. DSM23827 CphA V10R
CGTATTCGGGCCAcgAAACTGAAGTCG	R. DSM23827 CphA V10R
GGTATCGCATGAATGATcgtTCTCTTATCCAGGTAGG	F. DSM23827 CphA A173R
CCTACCTGGATAAGAGAacgATCATTATCGGATACC	R. DSM23827 CphA A173R
GGATTGAGGCCaCTTTGACCTC	F. DSM23827 CphA A190T
GAGGTCAAAGtGGCCTCAATCC	R. DSM23827 CphA A190T
GTTGAGATTGCAtgcGATAAGAATATGTG	F. DSM23827 CphA A204C
CACATATTCTTATCgcaTGCAATCTCAAC	R. DSM23827 CphA A204C
GATTAATGCGGcGCCTGGTCTG	F. DSM23827 CphA G437A
CAGACCAGGCgCCGCATTAATC	R. DSM23827 CphA G437A
GGcGCCTGGTtTtCGCATGCATgTCTCTCCGTC	F. DSM23827 L440F I444V
GACGGAGAGAcATGCATGCgAaACCAGGCgCC	R. DSM23827 L440F I444V

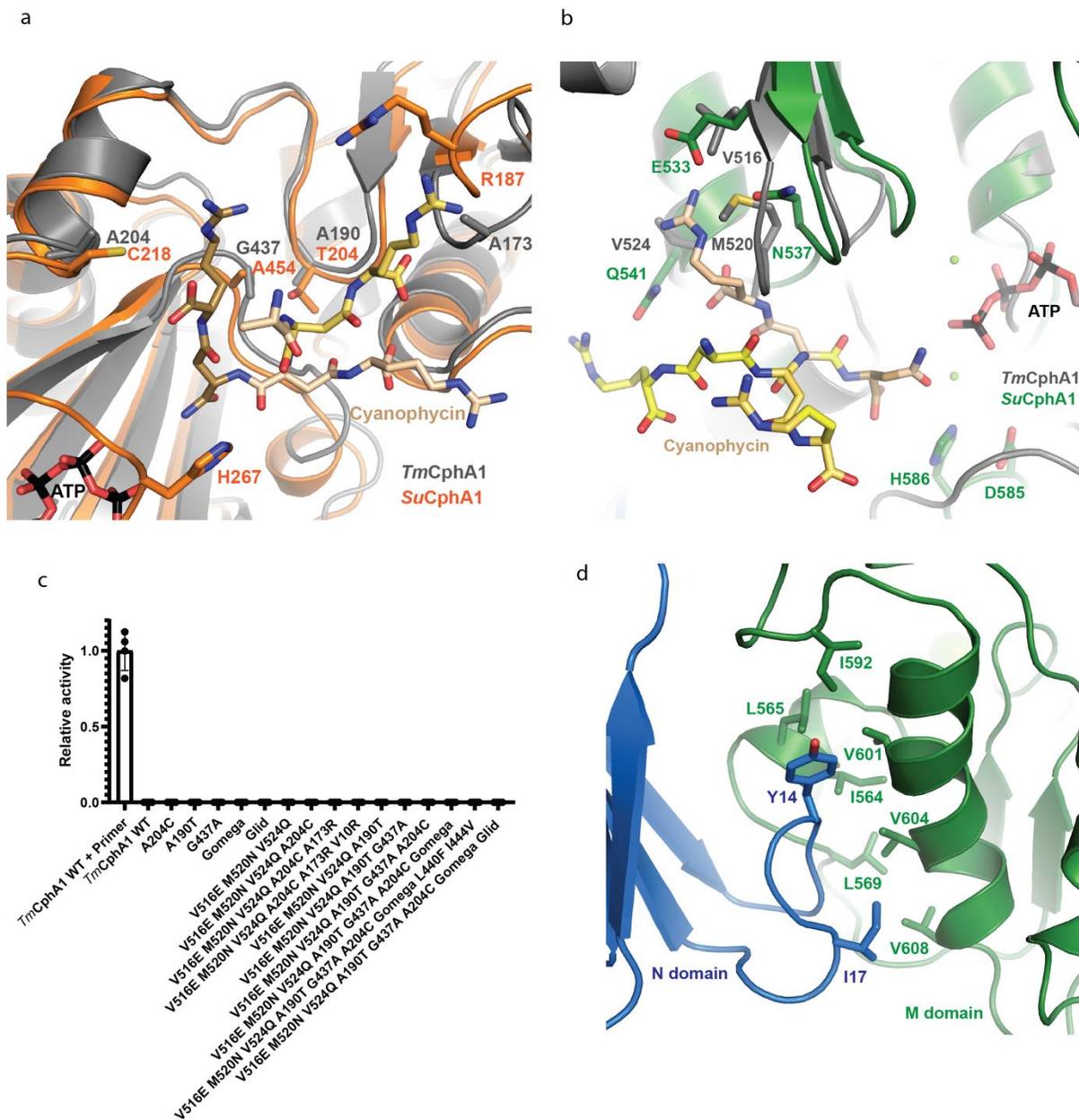
**Supplementary Table 3.4. HR-MS analysis of the cyanophycin segments used in this study.**

Molecule	Charge	Expected m/z	Observed m/z
$\beta$ -Asp-Arg	+1	290.14590	290.14560
( $\beta$ -Asp-Arg)-Asp	+1	405.17284	405.17276
( $\beta$ -Asp-Arg) <sub>2</sub>	+1	561.27395	561.27327
( $\beta$ -Asp-Arg) <sub>3</sub>	+1	832.40066	832.40062
( $\beta$ -Asp-Arg) <sub>4</sub>	0	1103.53006	1103.52799
( $\beta$ -Asp-Arg) <sub>8</sub> -NH <sub>2</sub> (reference <sup>2</sup> )	0	2187.05826	2187.06290
( $\beta$ -Asp-Arg) <sub>8</sub> -Asn (reference <sup>2</sup> )	+2	1151.54624	1151.54538
( $\beta$ -Asp-Arg) <sub>12</sub> (reference <sup>2</sup> )	+3	1091.52301	1091.52134

### 3.8.2. Supplementary figures

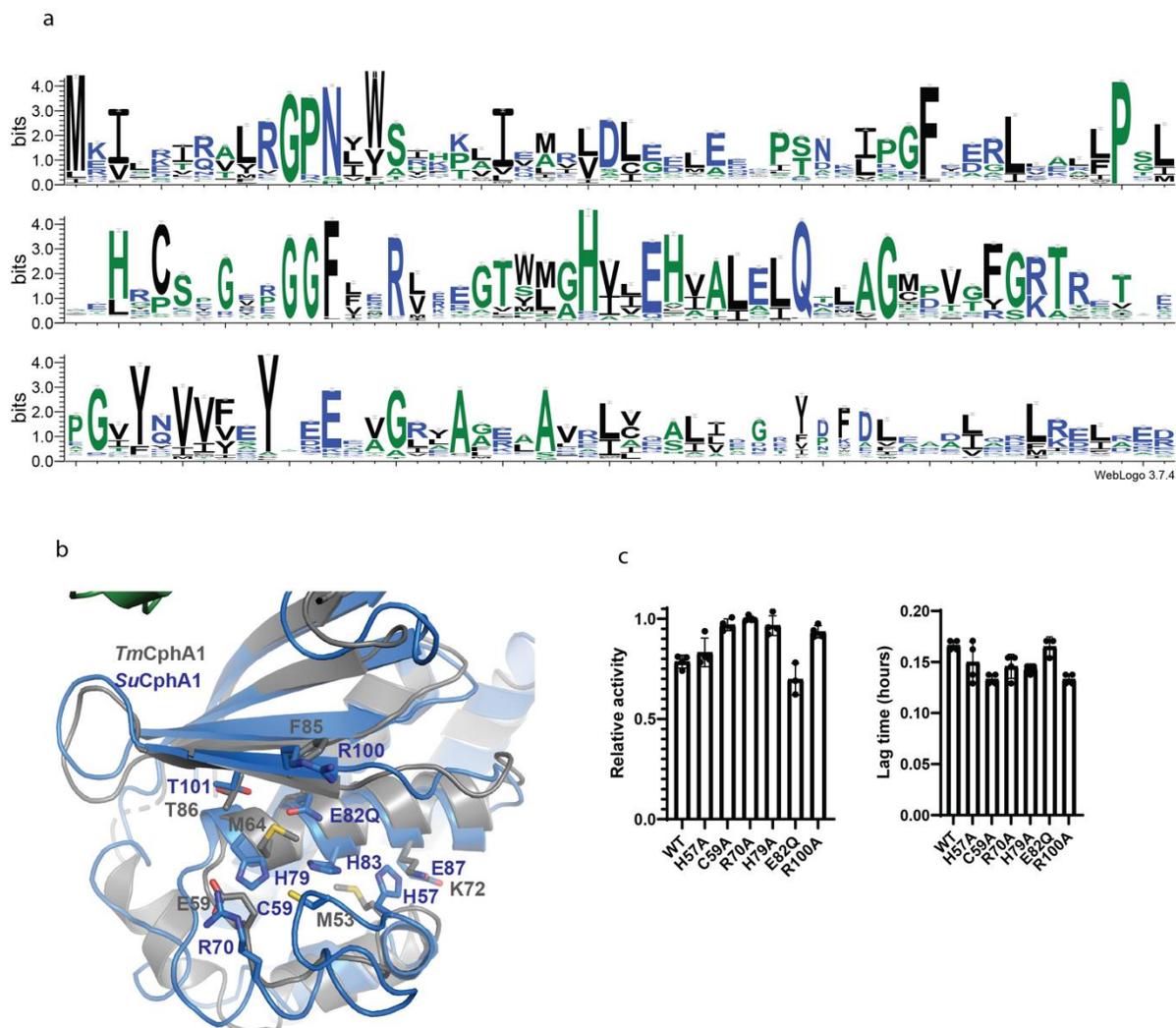


**Supplementary Figure 3.1. *SuCphA1* primer-dependence.** (a) Activity assay plots of different concentrations of *TmCphA1* in the absence of primer. Without primer, the enzyme displays no activity after 3 days, even at high concentrations. n=4 independent experiments. Data are presented as mean value. (b) Activity assay plots, activity rate values and lag time of *SuCphA1* in the presence of various cyanophycin primers. Asp, Arg and  $\beta$ -Asp-Arg display similar activity profiles, suggesting  $\beta$ -Asp-Arg are not used as primers. ( $\beta$ -Asp-Arg)-Asp and ( $\beta$ -Asp-Arg)<sub>2</sub> both shorten the lag phase before onset of activity, suggesting they are moderately good primers. ( $\beta$ -Asp-Arg)<sub>3</sub> and ( $\beta$ -Asp-Arg)<sub>4</sub> are both equally good primers for this enzyme. n=4 independent experiments. Data are presented as individual measurements and mean value, error bars represent SD values.



**Supplementary Figure 3.2. CphA1 mutations.** (a) Overlay of *Su*CphA1 (orange) and *Tm*CphA1 (gray) G domain active sites. Polymer binding residues which are different in these enzymes are labeled. (b) Overlay of *Su*CphA1 (green) and *Tm*CphA1 (gray) M domain active sites. Polymer binding residues which are different in these enzymes are labeled. (c) Attempts to mutate *Tm*CphA1 G and M active sites to make them more similar to those of *Su*CphA1 did not result in primer-independent activity, suggesting these active sites are not responsible for primer-independent activity. n=4 independent experiments. Data are presented as individual

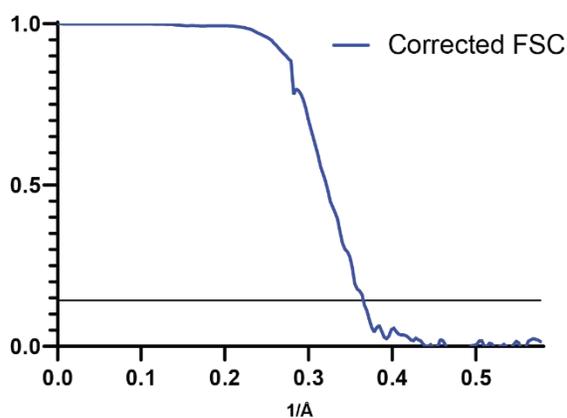
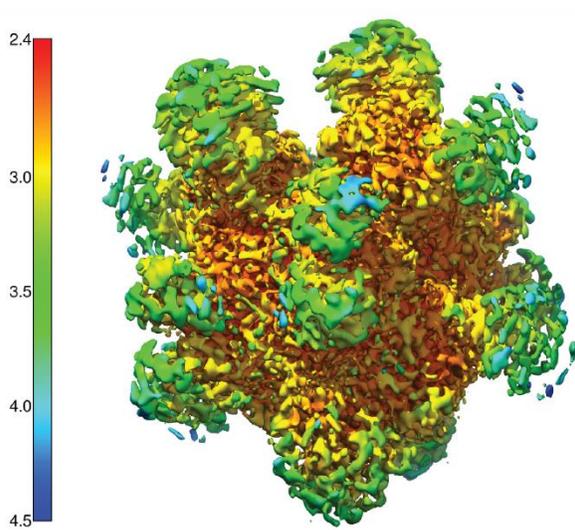
measurements and mean value, error bars represent SD values. (d) The loop containing *Su*CphA1 Y14 and I17 interacts with a hydrophobic patch on the M domain, thus burying those hydrophobic residues. Double mutations Y14S I17T allowed soluble expression and purification of the excised N domain.



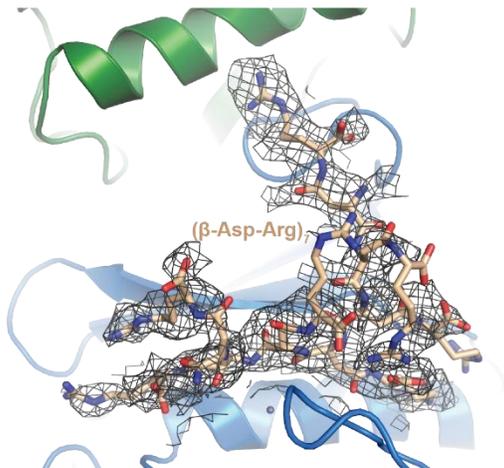
**Supplementary Figure 3.3. Important residues for N domain activity.** (a) Weblogo<sup>8</sup> diagram covering the entire N domain of CphA1. This Weblogo was constructed from sequence alignments of CphA1 enzymes using ClustalW<sup>11</sup>, and excludes cyanophycin synthetase 2 (CphA2) sequences. CphA2s are specialized cyanobacterial enzymes that polymerize  $\beta$ -Asp-Arg dipeptides recovered from degraded cyanophycin<sup>12,13</sup>. CphA2 N domains share low sequence identity to CphA1 N domains and the N domain active site motif is absent from CphA2 sequences. (b) Overlay of the N domains of *Tm*CphA1 (PDB code 7LGN [<https://www.rcsb.org/structure/7LGN>], gray) and

*Su*CphA1 (blue). The domains have a similar overall structure, but with key differences in sequence. Conserved residues found in *Su*CphA1 and the residues in equivalent positions in *Tm*CphA1 are labeled. (c) Activity rate and lag time of *Su*CphA1 N domain mutants with  $(\beta\text{-Asp-Arg})_3$  as primer. The mutants displayed similar or slightly higher activity rates to those of the WT enzyme. n=4 independent experiments. Data are presented as individual measurements and mean value, error bars represent SD values.

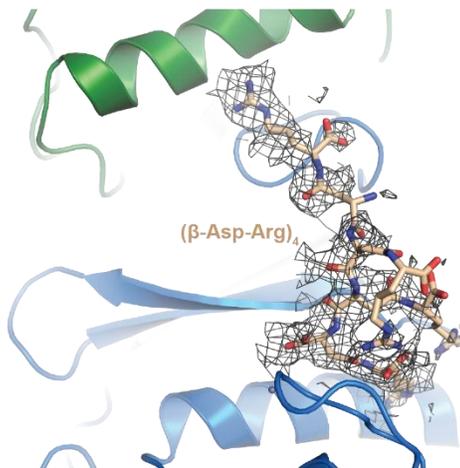
a



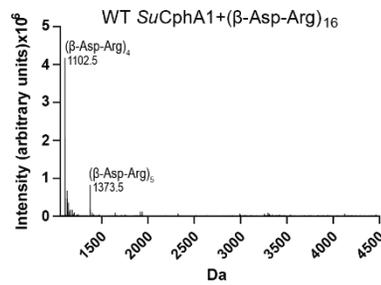
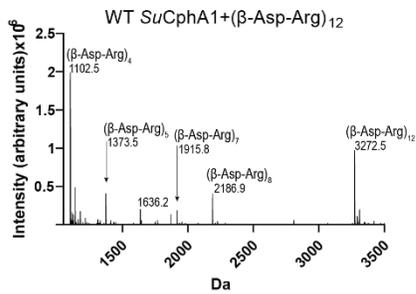
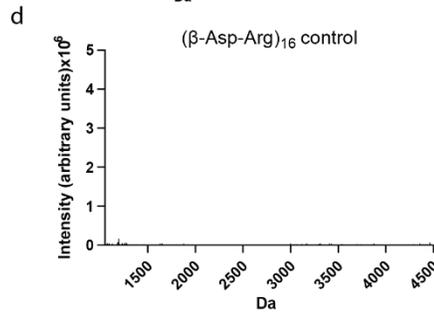
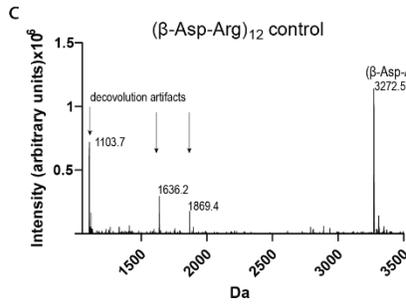
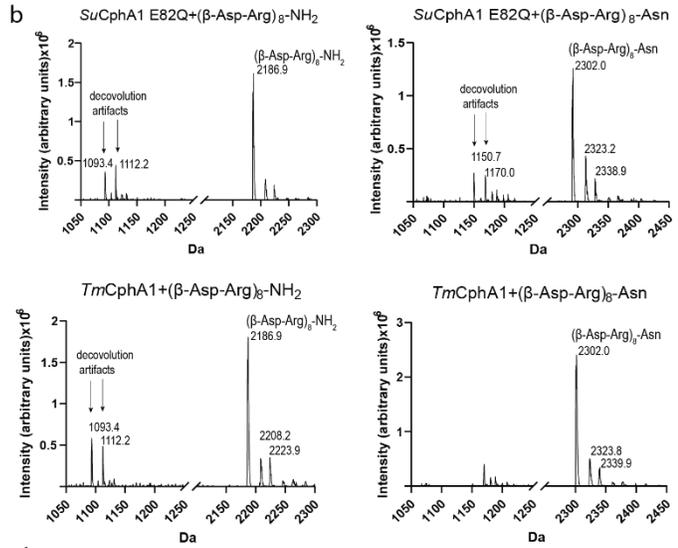
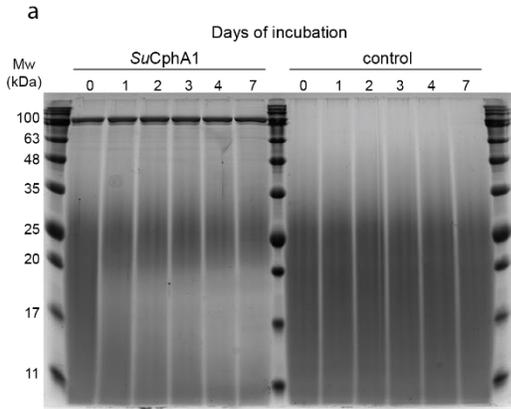
b



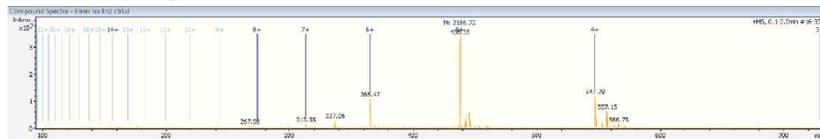
c



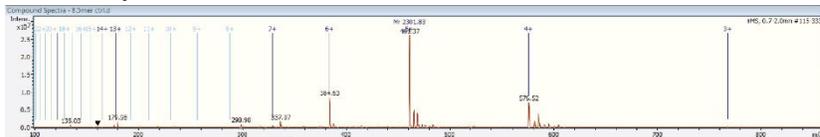
**Supplementary Figure 3.4. Cryo-EM of SuCphA1 with cyanophycin.** (a) The cryo-EM map of *Su*CphA1 E82Q colored by local resolution and the phase-randomized FSC curve of the map. (b) The map of *Su*CphA1 E82Q with ATP and ( $\beta$ -Asp-Arg)<sub>16</sub>, with ( $\beta$ -Asp-Arg)<sub>7</sub> fit in the map near the N domain active site. The map is displayed at a contour level of 5 with 2.5 Å carving around the ligand. (c) Cryo-EM map<sup>2</sup> of *Su*CphA incubated with ( $\beta$ -Asp-Arg)<sub>16</sub>. Signal for a chain of four dipeptide residues is visible in this map. Similar signal is seen in maps of complexes containing ( $\beta$ -Asp-Arg)<sub>8</sub>-NH<sub>2</sub> or ( $\beta$ -Asp-Arg)<sub>8</sub>-Asn, but not in maps of complexes containing only ATP, Arg and Asp<sup>2</sup>. The map is displayed at a contour levels of 5 with 2.5 Å carving around the ligand.



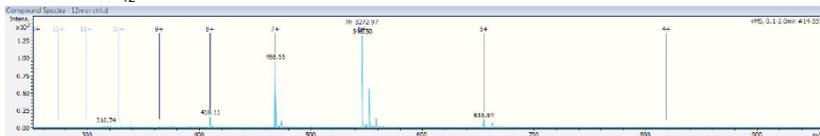
**e** ( $\beta$ -Asp-Arg)<sub>8</sub>-NH<sub>2</sub> control



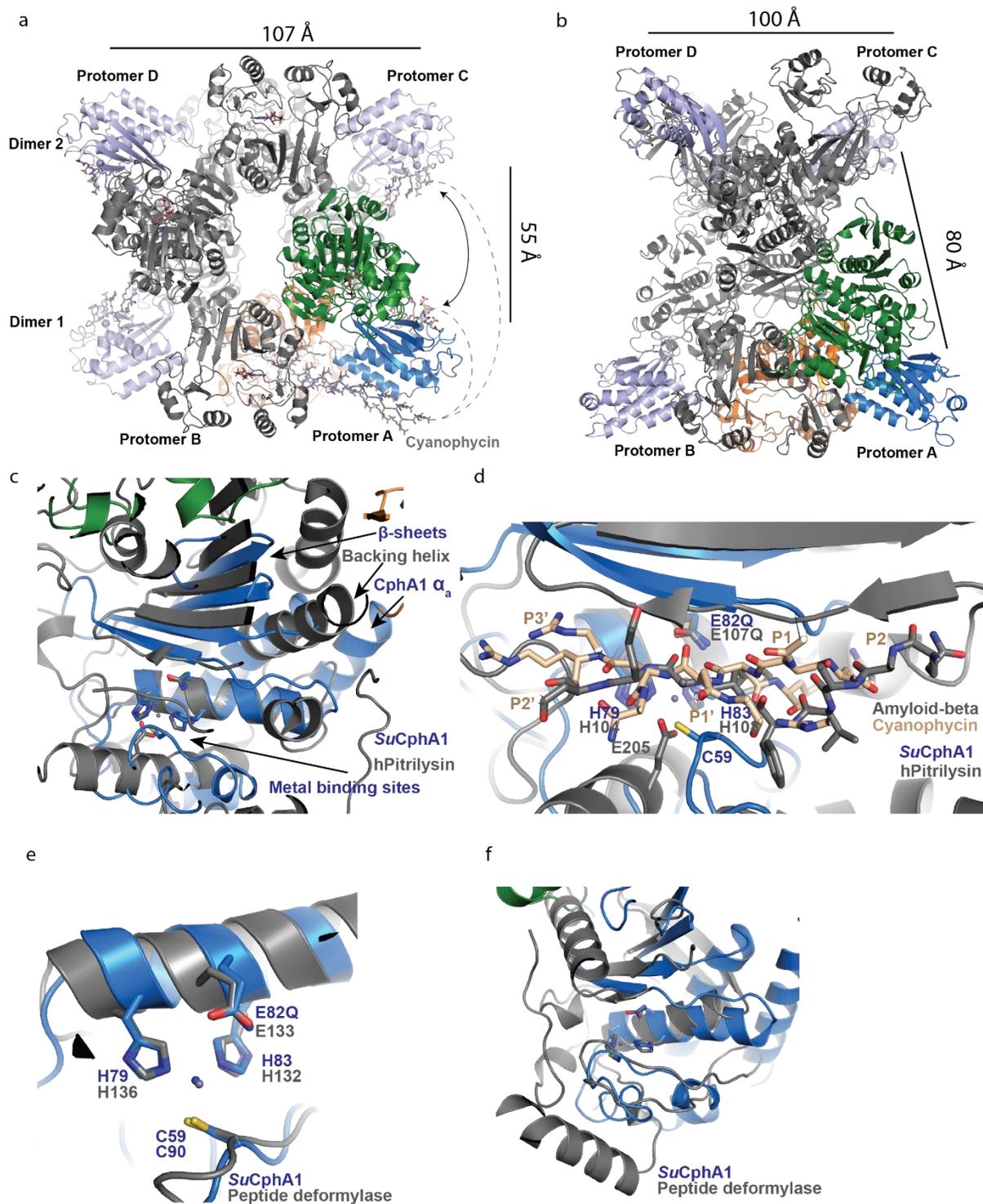
( $\beta$ -Asp-Arg)<sub>8</sub>-Asn control



( $\beta$ -Asp-Arg)<sub>12</sub> control

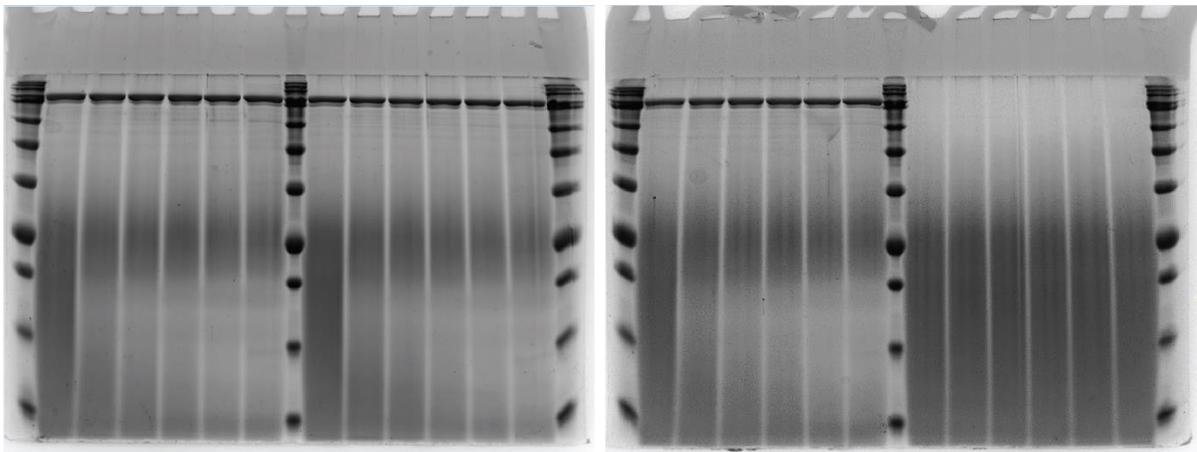


**Supplementary Figure 3.5. Mass spectra of cyanophycin degradation assays.** (a) SDS-PAGE analysis of cyanophycin degradation assays. Polymer purified from *E. coli* expressing *TmCphA1* was incubated with and without *SuCphA1* over several days. The gradual decrease in smear intensity, especially between the 11 and 20 kDa markers, shows cyanophycin is being slowly degraded in the presence of *SuCphA1*. n=3 independent experiments. The uncropped gel images are presented in Supplementary Fig. 3.7. (b) MS traces of  $(\beta\text{-Asp-Arg})_8\text{-NH}_2$  and  $(\beta\text{-Asp-Arg})_8\text{-Asn}$  after incubation with *SuCphA1* E82Q or WT *TmCphA1*. (c) MS traces of  $(\beta\text{-Asp-Arg})_{12}$  before (top) and after (bottom) incubation with WT *SuCphA1*. After incubation with enzyme, the peak matching  $(\beta\text{-Asp-Arg})_{12}$  (expected at 3271.9 Da) is reduced and peaks corresponding to  $(\beta\text{-Asp-Arg})_4$  (expected at 1102.5 Da),  $(\beta\text{-Asp-Arg})_5$  (expected at 1373.6 Da),  $(\beta\text{-Asp-Arg})_7$  (expected at 1915.9 Da) and  $(\beta\text{-Asp-Arg})_8$  (expected at 2187.0 Da) appear. (d) MS traces of  $(\beta\text{-Asp-Arg})_{16}$  before (top) and after (bottom) incubation with WT *SuCphA1*. No signal was observed for  $(\beta\text{-Asp-Arg})_{16}$  under these conditions, presumably because the used MS conditions result in low signal for long cyanophycin chains. After incubation with enzyme, the peak matching  $(\beta\text{-Asp-Arg})_4$  (expected at 1102.5 Da) and  $(\beta\text{-Asp-Arg})_5$  (expected at 1373.6 Da) appear. (e) Representative raw MS spectra of cyanophycin controls and degradation products. The charged-state rulers show that all major peaks can be accounted for. These spectra were deconvoluted to produce the data in figures 3.4a-c.



**Supplementary Figure 3.6. N domain orientation in CphA1 tetramers and homology to M16 family members.** (a) *Su*CphA1 with modelled cyanophycin bound to the charged patches on the N domain. Dashed lines show a possible route of cyanophycin to the N domain active sites. The

dimer architecture (chains A+B, C+D) positions N domain active sites (measured as the distance between the two Zn ions) 107 Å apart and facing away from each other. In a tetramer, however, N domains from adjacent dimers (chains A+C, B+D) face each other and are only 55 Å apart. (b) The tetramer architecture of *TmCphA1* (PDB code 7LGN) leads to an increased distance of ~80 Å between N domains from different dimers (chains A+C, B+D). However, as these N domains lack catalytic activity. (c) Alignment of *SuCphA1* N domain (blue) and human pitrilysin<sup>163</sup> (PDB code 4NGE, gray) shows moderate structural conservation of the metal binding site helix (pitrilysin residues 101-113), core β-sheet (pitrilysin residues 138-150, 86-93, 263-269) and backing helix (pitrilysin residues 153-168). (d) Alignment of the active sites of *SuCphA1* N domain (blue) and human pitrilysin (PDB code 4NGE [<https://www.rcsb.org/structure/4NGE>], gray) shows structural conservation of the metal binding residues and substrate positioning. (e) Alignment of the active sites of *SuCphA1* N domain (blue) and *E. coli* peptide deformylase<sup>174</sup> (PDB code 1DFF, gray) shows high structural similarity of the C-H-H metal binding triad. (f) Alignment of the overall structures of *SuCphA1* N domain (blue) and *E. coli* peptide deformylase<sup>174</sup> (PDB code 1DFF, gray) shows the two enzymes share little structural similarity.



**Supplementary Figure 3.7. Non-cropped gels.** Gels associated with supplementary figure 3.5a.

## Bridge to chapter 4

Our analysis in chapter 1 shows that CphA1 is a common enzyme found in members of most bacterial phyla. However, some cyanobacteria have, in addition to CphA1, a closely related enzyme called CphA2. This enzyme provides them with another pathway for cyanophycin synthesis – the polymerization of  $\beta$ -Asp-Arg dipeptides. Studies showed that these two enzymes do not have completely overlapping roles. To fully understand how cyanophycin metabolism works, then, it is necessary to understand the activity of CphA2 as well as that of CphA1. From a biotechnological perspective, bioengineering of cyanophycin synthetases to produce cyanophycin-like polymers is an ongoing subject of research. An earlier study found that CphA2 only has one active site, potentially making it more amenable to bioengineering efforts than CphA1. I decided to structurally and biochemically characterize CphA2 in the hope of providing data that will better explain its activity and facilitate its bioengineering.

#### 4. Structure and function of the $\beta$ -Asp-Arg polymerase cyanophycin synthetase 2

Published in: Sharon I, Grogg M, Hilvert D, Schmeing TM. *ACS Chemical Biology* 2022, 17, 3, 680-700.

#### **4.1. Abstract**

Cyanophycin is a biopolymer composed of long chains of  $\beta$ -Asp-Arg. It is widespread in nature, being synthesized by many clades of bacteria, which use it as a cellular reservoir of nitrogen, carbon and energy. Two enzymes are known to produce cyanophycin: cyanophycin synthetase 1 (CphA1), which builds cyanophycin from the amino acids Asp and Arg by alternating between two separate reactions for backbone extension and side chain modification; and cyanophycin synthetase 2 (CphA2), which polymerizes  $\beta$ -Asp-Arg dipeptides. CphA2 is evolutionarily related to CphA1, but questions about CphA2's altered structure and function remain unresolved. Cyanophycin and related molecules have drawn interest as green biopolymers. Because it only has a single active site, CphA2 could be more useful than CphA1 for biotechnological applications seeking to produce modified cyanophycin. In this study, we report biochemical assays on nine cyanobacterial CphA2 enzymes and report the crystal structure of CphA2 from *Gloeotheca citriformis* at 3.0 Å resolution. The structure reveals a homodimeric, 3-domain architecture. One domain harbors the polymerization active site and the two other domains have structural roles. The structure and biochemical assays explain how CphA2 binds and polymerizes  $\beta$ -Asp-Arg and highlights differences in *in vitro* oligomerization and activity between CphA2 enzymes. Using the structure and distinct activity profile as a guide, we introduced a single point mutation that converted *Gloeotheca citriformis* CphA2 from a primer-dependent enzyme into a primer-independent enzyme.

## **4.2. Introduction**

Cyanophycin, first discovered in cyanobacteria almost 140 years ago<sup>14</sup>, is a biopolymer produced by many bacterial species<sup>2,100</sup>. It has a poly-L-Asp backbone with an L-Arg attached to each sidechain through an isopeptide bond<sup>99</sup> (Fig. 4.1a). This composition gives it a high nitrogen content of 24% by mass, making it especially valuable for nitrogen storage<sup>130</sup>, although it can also be useful for storing carbon and energy<sup>133,134</sup>. Cyanophycin's function as a nitrogen reservoir is especially beneficial in nitrogen-fixing cyanobacteria<sup>102</sup>: nitrogenase is inhibited by oxygen<sup>131</sup>, so (aerobic) photosynthesis and nitrogen fixation must be separated, either temporally (in a day/night cycle)<sup>40</sup> or spatially (e.g. in heterocyst and vegetative cell types)<sup>42</sup>. Cyanophycin biosynthesis is coordinated with nitrogen fixation to produce reserves of excess fixed nitrogen. Subsequent cyanophycin degradation allows this store of nitrogen to be utilized on demand in aerobic periods or locations<sup>40,42</sup>.

Cyanophycin also has attractive biotechnological potential, with applications ranging from a source of poly-Asp (a water softener and super swelling material and a biodegradable alternative to poly-acrylic acid) to a material for bandages<sup>66</sup>. The commercial use of cyanophycin has been limited because yields of polymer to date are too low for commercial viability. Many studies have targeted combinations of enzyme, mutations, and host system to maximize heterologous expression<sup>80,82,160</sup>. Deeper understanding of cyanophycin biosynthesis could benefit these efforts.

Cyanophycin is polymerized by one of two enzymes, which use different substrates and reaction pathways. CphA1 alternately adds an aspartate to the polymer backbone and then an arginine to the Asp side chain, in two separate ATP-dependent reactions at two different active sites<sup>1</sup>. In contrast, CphA2 links  $\beta$ -Asp-Arg dipeptides together by a repetitive, ATP-dependent polymerization reaction at a single active site<sup>12,119</sup> (Fig. 4.1a).  $\beta$ -Asp-Arg dipeptides used by CphA2 are generated by previous degradation of cyanophycin by cyanophycinase<sup>61</sup>. CphA1 is found throughout bacteria<sup>2,100</sup>, but CphA2 evolved from CphA1 fairly recently in cyanobacteria<sup>12,119</sup>. The two enzymes seem to have partially overlapping roles *in vivo*, as the functions of both are important for maximal cyanophycin production in *Anabaena*<sup>12,119</sup>. Together with cyanophycinase and isoaspartyl dipeptidases, CphA1 and CphA2 balance cyanophycin biosynthesis and degradation, allowing cyanobacteria to respond to variations in nitrogen availability.

Cyanophycin synthesis can be primer dependent or primer independent<sup>99</sup>. In primer dependent synthesis, cyanophycin synthetase can extend an existing chain of cyanophycin, but cannot perform *de novo* cyanophycin production from ATP, Asp and Arg (for CphA1), or from ATP and  $\beta$ -Asp-Arg (for CphA2). CphA1 enzymes are largely primer dependent<sup>105</sup>. CphA2 has been reported to be primer-independent *in vitro*, but the reactions were performed with a very high  $\beta$ -Asp-Arg concentration<sup>12</sup>.

Recently, we revealed the architecture of CphA1 and showed how it makes cyanophycin by combining the functions of three different domains. The ATP-grasp-like G domain adds Asp to the terminal backbone carboxylate of cyanophycin; the Mur-ligase family-like M domain adds Arg to the side chain of the newly-added Asp; and the N domain loosely binds nascent cyanophycin to allow the end of the polymer chain to move more efficiently from one active site to the other.

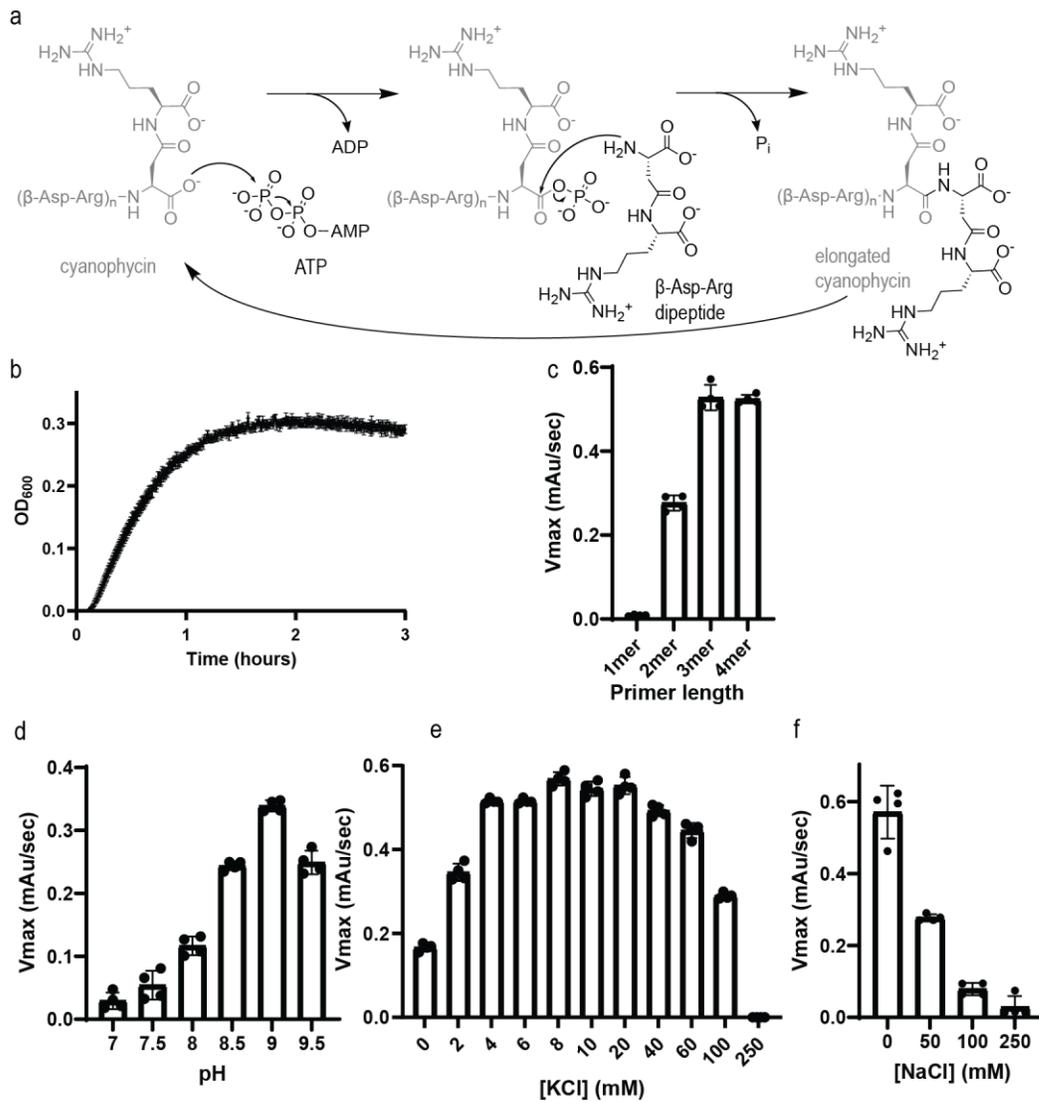
Biochemical characterization of CphA2<sup>12,119</sup> and the structures of the related CphA1 enzymes<sup>2</sup> are informative for understanding CphA2, but unanswered questions about its function and structure remain. Sequence analysis shows that CphA2 has a region similar in length to the N domain, an intact G domain and a C-terminal truncation that likely compromises the active site of the M domain<sup>2,12</sup> (Supplementary Fig. 4.1a). Moreover, the overall architecture of CphA2 is unknown and the oligomeric state uncertain<sup>12</sup>. It is assumed that the single intact (G domain) active site of CphA2 catalyzes a similar reaction to that of CphA1, but it is not known what differences it has acquired. Likewise, it is unknown whether the N domain of CphA2 has a role in polymer binding. It is also unclear whether CphA2 enzymes are primer dependent at physiological substrate concentrations, and whether primer (in)dependency is shared by all CphA2s from different bacteria or varies from enzyme to enzyme.

In this study, we characterized 9 CphA2 enzymes and solved the structure of one, gaining insight into the mechanism of substrate recognition and activity. The results, and comparison with CphA1, provide insights into the roles of the individual domains of CphA2 and help explain their contribution to enzyme activity.

## **4.3. Results and discussion**

### **4.3.1. Cyanophycin synthesis by CphA2**

The  $\beta$ -Asp-Arg polymerization reaction performed by CphA2 is thought to proceed in a two-step manner, analogous to Asp ligation by the CphA1 G domain and amide bond formation

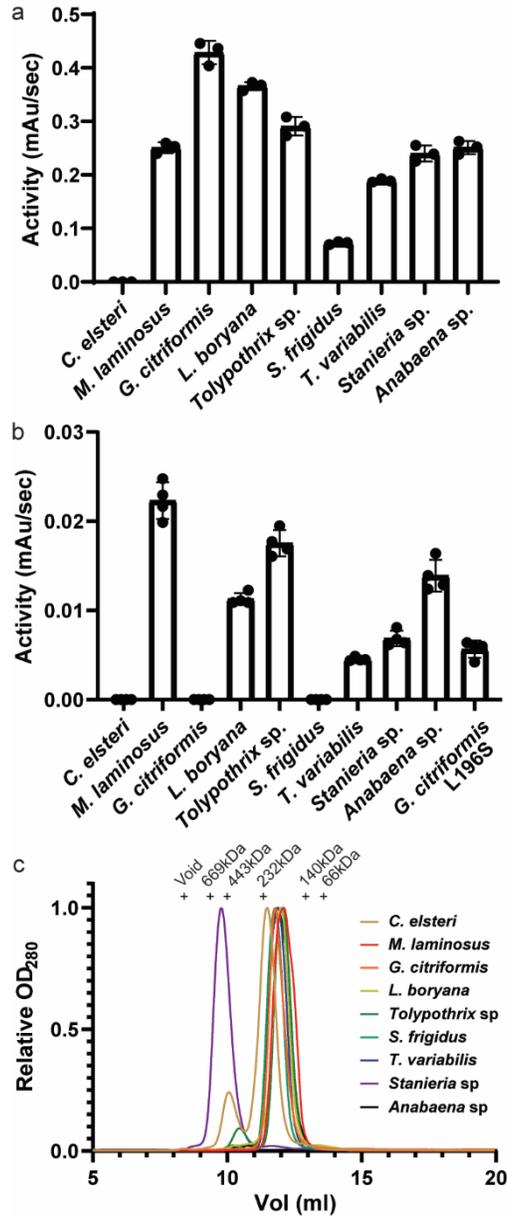


**Figure 4.1. Cyanophycin synthesis by CphA2.** (a) Cyanophycin synthesis catalyzed by CphA2. The terminal backbone carboxylate of a cyanophycin chain is phosphorylated, prior to attacked by the  $\alpha$ -amino group of an incoming  $\beta$ -Asp-Arg dipeptide, thus extending the cyanophycin polymer. (b) Cyanophycin synthesis by *G. citriformis* CphA2 in the presence of 50  $\mu$ M  $(\beta$ -Asp-Arg)<sub>3</sub> primer. (c) *G. citriformis* CphA2 activity with different primer lengths. (d) pH dependence of *G. citriformis* CphA2 activity. (e) Dependence of *G. citriformis* CphA2 activity on KCl concentration. (f) Dependence of *G. citriformis* CphA2 activity on NaCl concentration. Increasing ionic strength decreases activity, presumably because the high salt concentration interferes with the binding of cyanophycin, which is predominantly electrostatic. Reactions contain 40 mM KCl in addition to the NaCl concentrations indicated. All experiments were carried out in quadruplicate. Bar height represents the mean value of the maximal activity, and error bars show the standard deviation.

by other ATP-grasp enzymes. The terminal carboxylate of the cyanophycin polymer is first phosphorylated using ATP and then the resulting acylphosphate intermediate undergoes nucleophilic attack by the  $\alpha$ -amino group of  $\beta$ -Asp-Arg, extending the polymer chain by one dipeptide (Fig. 4.1a). Published *in vitro* assay conditions for *A. variabilis* and *Cyanothece sp.* CphA2 include 100 mM substrate  $\beta$ -Asp-Arg<sup>12</sup>, a very high concentration which is unlikely to be common in the cell. Although a cyanophycin primer was not necessary, the authors stated that the presence of purified soluble cyanophycin enhanced *in vitro* activity<sup>12</sup>.

For more detailed biochemical characterization of how primers influence CphA2-catalyzed cyanophycin synthesis, we cloned, expressed and purified CphA2 from *Gloeotheca citriformis* PCC7424 (Sup. Fig 4.1c). Cyanophycin formation was monitored by an increase in the OD<sub>600</sub> caused by scattering of light by insoluble cyanophycin<sup>2</sup>. Although no activity was observed with 2 mM  $\beta$ -Asp-Arg alone, addition of 50  $\mu$ M ( $\beta$ -Asp-Arg)<sub>3</sub> as a primer enabled robust polymerization of the dipeptide (Fig. 4.1b). The shorter ( $\beta$ -Asp-Arg)<sub>2</sub> peptide could also prime the reaction but was less effective than ( $\beta$ -Asp-Arg)<sub>3</sub>, whereas increasing the length of the primer to ( $\beta$ -Asp-Arg)<sub>4</sub> did not result in a further increase in rate (Fig. 4.1c). Thus, ( $\beta$ -Asp-Arg)<sub>3</sub> appears to be an optimal length to prime cyanophycin polymerization by CphA2 and was used in all subsequent experiments. Variation of the buffer conditions showed that *G. citriformis* CphA2 exhibited the highest activity at pH 9 (Fig. 4.1d) and moderate KCl concentrations (Fig. 4.1e). Potassium is required for CphA1 and CphA2 activity, and cannot be replaced by sodium<sup>12</sup>. However, high ionic strength was inhibitory, as activity decreased sharply with increasing KCl or NaCl concentrations (Fig. 4.1e,f).

Having established reaction conditions for one CphA2 enzyme, we expanded our study to eight additional CphA2 enzymes, from *Anabaena variabilis* PCC7120, *Anabaena sp.* UTEX2576, *Calothrix elsteri* CCALA953, *Leptolyngbya boryana* NIES2135, *Stenomitos frigidus* ULC18, *Mastigocladus laminosus* UU774, *Stanieria sp.* NIES3757 and *Tolypothrix sp.* NIES4075 (Supplementary Fig. 4.1c). These homologs have 51–97% identity to each other (Supplementary Fig. 4.1b), and come from cyanobacterial sections I-IV<sup>185</sup>. Like *G. citriformis* CphA2, seven of the new enzymes displayed cyanophycin synthesis activity when provided with the ( $\beta$ -Asp-Arg)<sub>3</sub> primer (Fig. 4.2a, Supplementary Fig. 4.1d). However, they exhibited substantial differences in maximal observed activity and in lag time before detectable cyanophycin synthesis. *C. elsteri* CphA2 did not display any activity under the conditions tested.



**Figure 4.2. Activity and size exclusion profile of nine CphA2 enzymes.** (a) Cyanophycin synthesis activity of nine different CphA2 enzymes in the presence of 50  $\mu\text{M}$   $(\beta\text{-Asp-Arg})_3$  primer. (b) Cyanophycin synthesis activity of the nine CphA2 enzymes in the absence of primer. (c) SEC chromatograms of the 9 homologs. All peaks were normalized to the maximal peak height.

We next asked if any of the new CphA2 enzymes could synthesize cyanophycin in the absence of primer. Primer-independent activity has been reported for one CphA1 enzyme<sup>105</sup> and, at very high substrate concentration, for *A. variabilis* and *Cyanothece sp.* CphA2<sup>12</sup>. We performed

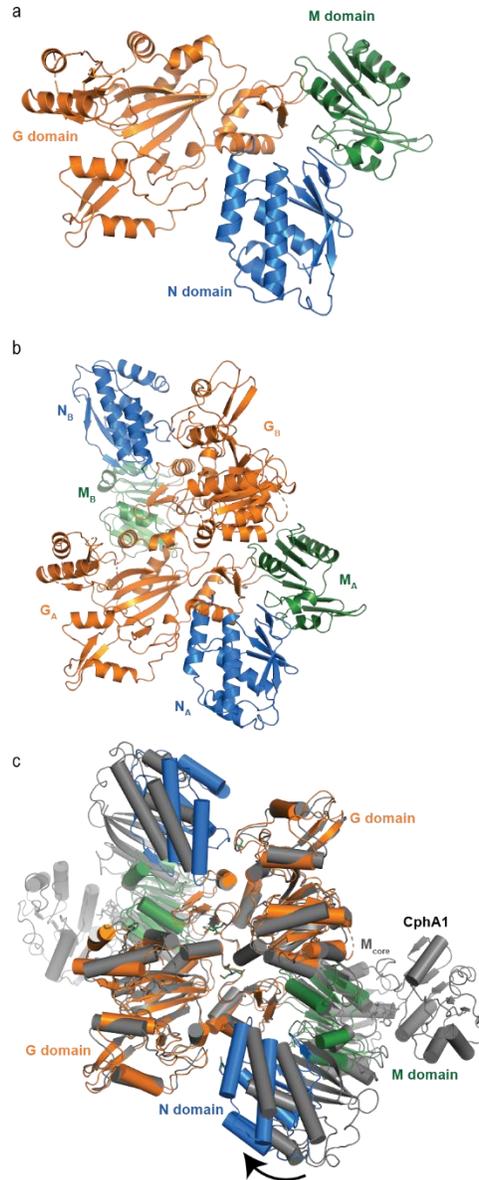
assays with the purified CphA2 enzymes using 2 mM  $\beta$ -Asp-Arg and no primer (Fig. 4.2b, Supplementary Fig. 4.1e). CphA2 from *S. frigidus*, like the *G. citriformis* enzyme, produced no cyanophycin in the absence of primer, despite being active in primer-dependent cyanophycin synthesis. However, the other six CphA2 enzymes able to perform primer-dependent synthesis also made cyanophycin in the absence of primer. For each enzyme, the primer-independent rate we observed was much lower than the primer-dependent rate, and the lag time was much longer (Supplementary Fig. 4.1e). We found no correlation between the relative maximal rates of the different enzymes' primer-dependent and primer independent activity. For example, *G. citriformis* CphA2 displayed the highest primer-dependent activity but had no measurable primer-independent activity, whereas *M. laminosus* CphA2 gave the highest primer-independent activity yet ranked fifth for the primer-dependent activity (Fig. 4.2a,b).

#### 4.3.2. Oligomeric state of CphA2 enzymes

Oligomerization of CphA2 would be expected, as CphA1 typically exists as a tetramer<sup>2</sup>, and ATP-grasp enzymes, to which the G domain is related, form dimers<sup>186</sup> or tetramers<sup>146</sup>. In fact, CphA2 was previously reported to exist as a multimer, although its oligomeric state was not clear. Klemke *et. al.* showed that the peak in size exclusion chromatography (SEC) chromatograms of *A. variabilis* and *Cyanothece sp.* CphA2 corresponded to a trimer or tetramer<sup>12</sup>. To determine whether CphA2 has a conserved oligomerization state, we performed SEC experiments with the 9 enzymes we purified in this study (Fig. 4.2c). Although the molecular mass of each CphA2 protomer is similar (71-74 kDa), there were differences in the SEC elution profiles. Seven enzymes eluted as species with masses in the range 156-186 kDa. Although these masses are somewhat higher than the expected ~146 kDa for a dimer, higher oligomeric states appear unlikely for these variants. In contrast, *C. elsteri* CphA2, the only enzyme that was inactive in our biochemical assays, eluted as a major peak (~217 kDa) and a minor peak (~460 kDa). It is possible that these peaks represent a trimer (expected at 219 kDa) and hexamer (438 kDa) and that these oligomeric states are the reason no activity was observed from this enzyme. *Stanieria sp.* CphA2, which is active, eluted as a single peak corresponding to ~536 kDa, between the sizes of a heptamer and an octamer (511 kDa and 584 kDa, Fig. 4.2c).

#### 4.3.3. The crystal structure of CphA2

All nine CphA2 enzymes used in the study could be crystallized under multiple conditions, and datasets extending to between 4 and 2 Å resolution were collected from crystals of six of them.



**Figure 4.3. The crystal structure of *G. citriformis* CphA2.** (a) View of the CphA2 protomer, composed of the N domain (blue), G domain (orange) and M domain (green). (b) The biological dimer of CphA2. (c) CphA2 is similar in architecture to the constituent dimer of a tetrameric CphA1.

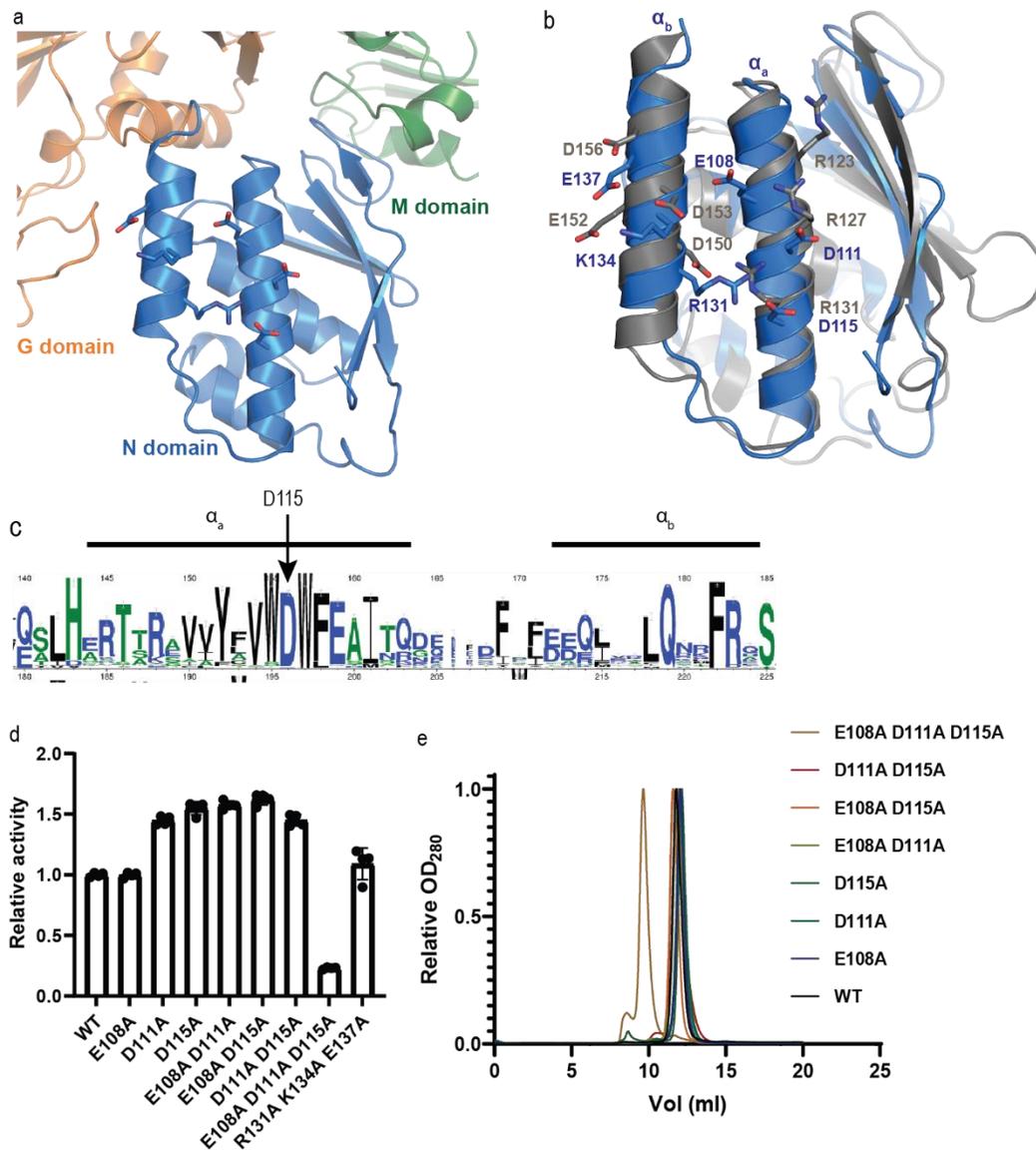
Unfortunately, analysis of the data sets almost always revealed major pathologies, the most common being severe twinning, which precluded structure determination and/or model refinement. Eventually, suitable crystals of the *G. citriformis* CphA2 enzyme were obtained. Combining data from two crystals gave a dataset with which we could phase, model build and refine the structure at 3.0 Å resolution (Supplementary Table 4.1).

CphA2 is, as expected, a three-domain protein, consisting of an N domain (residues 1-143), a G domain (144-470), and an M domain (471-616). The N domain is nestled between the G and M domains (Fig. 4.3a). The protein is dimeric, consistent with the SEC results for *G. citriformis* CphA2 and most other CphA2s (Fig. 4.2c, 4.3b). The asymmetric unit contains one CphA2 protomer (Fig. 4.3a), with a symmetry mate completing the dimer (Fig. 4.3b). The dimer interface, which is mainly formed by two G domain helices (189-213) and a loop (166-170), buries 1526 Å<sup>2</sup> of surface area.

The structure of CphA2 shows clear similarities to, and differences from, CphA1<sup>2</sup>. The individual domains are arranged in a similar way, with the N domain shifted by ~24 degrees (Fig. 4.3c). CphA2 lacks part of CphA1's M<sub>core</sub> (residues 644-723 in *Synechocystis sp.* UTEX2470 CphA1 (*Su*CphA1)<sup>2</sup>) and the entire M<sub>lid</sub> lobe (724 to the C terminus). Because of these deletions, the ATP binding site of the M domain is completely missing. In addition, many residues that are important for cyanophycin binding to the M domain of CphA1 (e.g. T538, E533, R561 and S542 in *Su*CphA1) are not conserved in CphA2. This is consistent with CphA2 requiring only one active site, the G domain's, for dipeptide polymerization, whereas CphA1 requires the M domain active site to ligate Arg to the main-chain Asp residue. The dimer interface in CphA2 is similar to that in CphA1, which is also largely formed by G domains. However, both possible tetramerization interfaces observed in CphA1 enzymes<sup>2</sup> are missing in CphA2: CphA1 enzymes adopt two different tetramer arrangements, with both tetramer interfaces involving M domain residues. Because its M domain is truncated, CphA2 lacks these residues, and is unable to form these contacts.

#### 4.3.4. Structure and mutation of the N domain

Since the CphA2 structure confirms that the M domain was altered and inactivated over the course of evolution, we focused on the contributions of the other domains to polymerase activity. The N domain of CphA2 has the same overall fold as those in CphA1<sup>2</sup>. Each have central 5-stranded beta sheet backed by two long and three short helices, although the lengths of the strands and positions of the smallest helices and loops differ (Fig. 4.4a,b). In CphA2, the two long anti-parallel helices are  $\alpha_a$  (103-122) and  $\alpha_b$  (128-141) (Fig. 4.4a). Unlike the corresponding helices in CphA1 N domains, the surface of CphA2  $\alpha_a$  and  $\alpha_b$  does not contain many conserved charged residues (Fig. 4.4c). The conserved charged patches in CphA1 are important for binding the growing cyanophycin polymer and their mutation drastically decreased CphA1 activity<sup>2</sup>, so we



**Figure 4.4. The N domain of CphA2.** (a) The CphA2 N domain has two long antiparallel helices,  $\alpha_a$  and  $\alpha_b$ , supported by a central  $\beta$ -sheet and smaller helices. Several non-conserved charged residues on the surface of these two helices are highlighted. (b) Overlay of CphA2 (blue) and *Su*CphA1 (gray) N domains. The two domains have the same  $\beta$ - $\beta$ - $\alpha$ - $\alpha$ - $\beta$ - $\beta$ - $\alpha$ - $\alpha$  fold. The conserved charged residues of CphA1 and non-conserved charged residues of  $\alpha_a$  and  $\alpha_b$  CphA2 are highlighted. (c) Weblogo<sup>8</sup> showing conservation of CphA2  $\alpha_a$  and  $\alpha_b$ . (d) Activity assays of *G. citriformis* N domain mutants. Experiments were carried out in quadruplicate. Bar height represents the mean value of the maximal activity rate and error bars show the standard deviation. (e) SEC chromatograms of *G. citriformis* CphA2 N domain mutants. All peaks were normalized to the maximal peak height.

wondered whether these helices play a similar role in CphA2. *G. citriformis* CphA2 does have several charged residues at positions roughly corresponding to those in CphA1, i.e. E108, D111 and D115 on  $\alpha_a$  and R131, K134 and E137 on  $\alpha_b$  (Fig. 4.4b), but only D115 seems to be conserved (Fig. 4.4c). To determine whether these residues were important for CphA2 activity, we created the two triple mutants E108A-D111A-D115A (CphA2 $\alpha_a$ mut) and R131A-K134A-E137A (CphA2 $\alpha_b$ mut). CphA2 $\alpha_b$ mut displayed similar activity to that of wildtype (WT) CphA2, while CphA2 $\alpha_a$ mut displayed substantially reduced activity (Figure 4.4d). However, the latter result is unlikely to be the direct consequence of abrogated cyanophycin binding. The SEC elution profiles of CphA2 $\alpha_b$ mut and WT CphA2 were similar, but CphA2 $\alpha_a$ mut eluted as a single peak corresponding to a size of 573 kDa, close to the expected mass of an octamer (584 kDa) (Fig. 4.4e). This dramatic change in oligomerization state could be responsible for the observed loss in activity, for example by steric occlusion of the active site. We therefore introduced single (E108A, D111A and D115A) and double (E108A-D111A, E108A-D115A, D111A-D115A) mutations in helix  $\alpha_a$  to obtain CphA2  $\alpha_a$  variants that oligomerized normally (Fig. 4.4e). Activity assays revealed that E108A had similar activity to that of wildtype. However, to our surprise, every single or double mutant that contained D111A or D115A displayed ~50% higher activity than WT (Fig. 4.4d). All enzymes have similar melting temperatures, suggesting no significant difference in stability (Supplementary Fig. 4.2a). It is unusual that we could identify mutants of the N domain that increased activity and but none (other than the octameric CphA2 $\alpha_a$ mut) that reduced activity. This shows that if the N domain is involved in polymer binding, it does so in a very different way to CphA1's N domain. Another possibility is that the N domain is important for solubility. Consistent with this hypothesis, attempts to express CphA2 variants of *G. citriformis* and *Tolypothrix* sp. which lack the N domain (CphA2 $\Delta$ N) did not give protein. Conversely, constructs including only the N domains from these enzymes yielded well behaved, soluble proteins (Supplementary Fig. 4.2b).

#### 4.3.5. Structure and mutation of the G domain

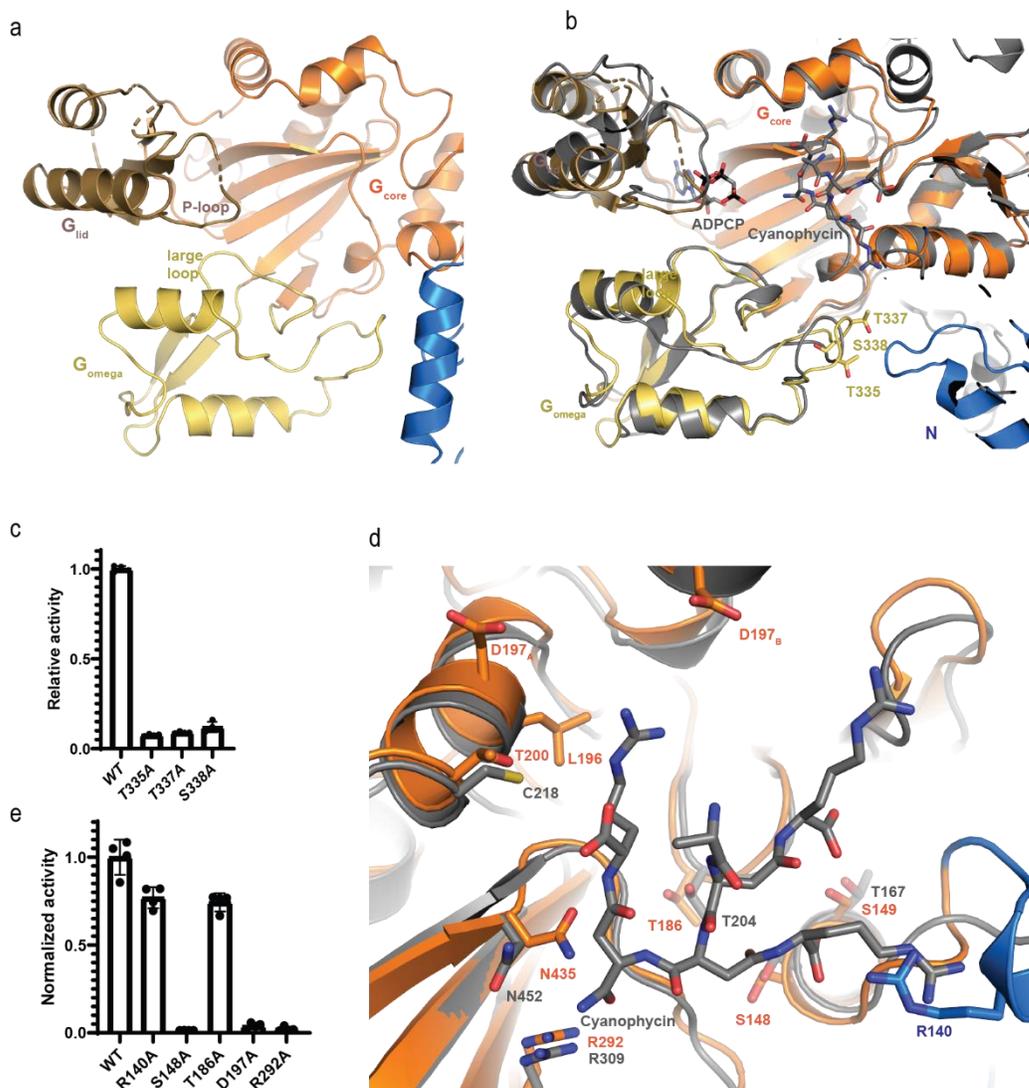
The G domain of CphA2 is composed of a main body (G<sub>core</sub>; 144-218, 284-307, 383-470) and two flexible lobes: G<sub>lid</sub> (219-283) and G<sub>omega</sub> (308-382) (Fig. 4.5a). G<sub>lid</sub> does not make crystal contacts and has poorer electron density and higher B-factors than other parts of the enzyme. This G<sub>lid</sub> flexibility is conserved among ATP-grasp enzymes and is believed to be important for activity<sup>147</sup>. G<sub>lid</sub> contains the flexible P-loop present in all ATP-grasp enzymes<sup>138,146</sup> (Fig. 4.5a), and

the conserved H247, which is also present in most CphA1s, is centrally positioned in the CphA2 P-loop.  $G_{\text{omega}}$  has only been previously observed in the structures of CphA1 and a bifunctional glutathione synthetase<sup>7</sup>, and contains the "large loop"<sup>116</sup> typical for ATP-grasp enzymes (Fig. 4.5a). The density and B-factors of  $G_{\text{omega}}$  also suggest the presence of flexibility, although not as extensive as in  $G_{\text{jid}}$ .

$G_{\text{omega}}$  and its large loop, are thought to be important for binding and recognition of the incoming substrate<sup>117</sup>. The large loop (CphA2 374-381) and an adjacent loop (CphA2 328-340) are present in both CphA1 and CphA2, but differ in sequence and structure (Fig. 4.5b), in accordance with differing substrate identity (Asp vs  $\beta$ -Asp-Arg; Fig. 4.1a, Supplementary Fig. 4.4a). The large loop forms a putative substrate binding pocket near the G domain active site that is capped by the adjacent loop (Fig. 4.5b). Mutation to alanine of the highly conserved T335, T337 or S338 all lead to large decreases in activity (Fig. 4.5c), consistent with a role in  $\beta$ -Asp-Arg recognition for this region of  $G_{\text{omega}}$ . We then asked if transplanting  $G_{\text{omega}}$  from CphA2 onto a CphA1 would confer  $\beta$ -Asp-Arg recognition capabilities to the CphA1 enzyme. We created a chimeric protein containing  $G_{\text{omega}}$  from *G. citriformis* CphA2 and all other (sub)domains from *Tatumella morbirosei* CphA1. However, while the resulting chimera, *TmCphA1*<sub>Gcit-omega</sub>, could be produced and purified, it displayed neither CphA1 nor CphA2 activity.

To attempt to visualize how CphA2 binds cyanophycin, we crystallized the enzyme in the presence of short polymer segments. However, datasets collected from crystals that were either soaked or co-crystallized with 5 mM ADP or ATP and 1 mM of ( $\beta$ -Asp-Arg)<sub>3</sub> or ( $\beta$ -Asp-Arg)<sub>4</sub> did not show extra density near the active site that we could confidently attribute to cyanophycin. This is likely because tartrate was present at 200 mM concentration in the crystallization buffer. Indeed, strong density was visible in maps of the unliganded CphA2 near the G domain active site, where cyanophycin is expected to bind. Two tartrate molecules were fit into this density (Supplementary Fig. 4.3a).

Despite being unable to co-crystallize CphA2 with substrate, insight into cyanophycin binding could be gained by superimposing CphA2 with *Su*CphA1 in complex with ADPCP and ( $\beta$ -Asp-Arg)<sub>8</sub>-NH<sub>2</sub> (Fig. 4.5b,d) and mutagenizing presumed contacts. In CphA1, recognition of cyanophycin by the G domain is achieved through interactions with several conserved residues on the surface of  $G_{\text{core}}$ . The structure and sequence show that the conserved residues involved in cyanophycin binding by G domain of CphA1 are conserved in CphA2 as well. *G. citriformis*



**Figure 4.5. The G domain of CphA2.** (a) The structure of the G domain: G<sub>core</sub> (orange), G<sub>lid</sub> (brown) and G<sub>omega</sub> (yellow). (b) Overlay of CphA2 (colored) and ADPCP- and cyanophycin-bound *Su*CphA1 (gray) G domains. The overall structures are very similar, with the main differences arising from conformational variations between the flexible lobes G<sub>lid</sub> and G<sub>omega</sub>, which are likely influenced by crystal packing. Conserved loop residues are shown as yellow sticks. (c) Mutation of conserved loop residues decrease CphA2 activity. (d) A close-up view of the G domain active site of CphA2 (colored) and *Su*CphA1 with bound cyanophycin (gray). Conserved cyanophycin binding residues are highlighted. These residues are very similar in CphA1 and CphA2, suggesting that they bind cyanophycin in a similar way. (e) Activity assays of *G. citriformis* CphA2 G domain mutants.

CphA2 should thus bind the reactive C-terminal carboxylate of cyanophycin through interaction with R292 (*Su*CphA1 R309), the first  $\beta$ -Asp-Arg dipeptide with T200 and N435 (*Su*CphA1 C218 and N452), the second  $\beta$ -Asp-Arg dipeptide with S148 and T186 (*Su*CphA1 S166 and T204), and the third with S149 (*Su*CphA1 T167, Fig. 4.5d, Supplementary Fig. 4.5a). Indeed, mutations of these residues reduced (R140A, T186A) or eliminated (S148A, D197A, R292A) activity (Fig. 4.5e), supporting the hypothesis that CphA2 binds cyanophycin much like CphA1 does. The deleterious effects of increasing ionic strength (Fig. 4.1e,f) underscore the importance of electrostatic interactions for substrate binding to CphA2.

Finally, we used the structure to better understand CphA2's primer dependence. *G. citriformis* CphA2 displays no primer-independent activity despite having the highest primer-dependent activity (Fig. 4.2a). We reasoned that increased interactions with the primer terminus could enable a single dipeptide to be used for initiation of synthesis. We therefore examined the residues around the active site of the G domain that could contact a primer, and looked for differences between *G. citriformis* and other CphA2 enzymes that displayed primer independent activity. Hydrophobic residue L196 of *G. citriformis* CphA2 is adjacent to where first dipeptide of a cyanophycin primer would bind (Fig. 4.5d), and this position is occupied by a serine in CphA2 enzymes that were capable of primer-independent synthesis. This serine might form a hydrogen bond with the guanidinium group of the first  $\beta$ -Asp-Arg dipeptide, which could stabilize binding of that  $\beta$ -Asp-Arg dipeptide and allow proper positioning for the first amide bond synthesis. Indeed, introduction of the L196S mutation into *G. citriformis* CphA2 afforded primer-independent activity similar to that of other CphA2 enzymes (Fig. 4.2b), demonstrating the importance of this interaction for the binding of  $\beta$ -Asp-Arg as the carboxylate donor in the absence of longer primers.

#### **4.3.6. Model of cyanophycin biosynthesis by CphA2**

The structural and biochemical data suggest that cyanophycin polymerization by CphA2 proceeds by a simple process. An existing polymer chain binds at the G domain active site through its three C-terminal dipeptidyl residues. The C-terminal carboxylate is phosphorylated and then extended by one dipeptide in the same way other ATP-grasp enzymes perform ligation<sup>112</sup>. The elongated polymer likely then dissociates from this product position and re-associates, placing the newly added  $\beta$ -Asp-Arg in the substrate position. A dissociation/re-association mechanism appears more likely than a smooth shift or slide of the polymer, because the transition between

substrate and product binding modes would require extensive rearrangements of the conformations of the terminal three dipeptide residues, and it is not clear that polymer binds to other parts of CphA2. It is possible that general electrostatic interactions help keep the polymer in the vicinity of CphA2 during dissociation and reassociation.

The N domain of CphA1 tethers the polymer to the enzyme, but our data are not definitive as to whether the N domain of CphA2 does so. Two mutations in the CphA2 N domain increased rate of synthesis, but we found no mutation in the N domain that decreases activity by directly disrupting an N domain - cyanophycin interaction (i.e. without changing oligomeric state). If the CphA2 N domain is involved in polymer binding, its mode seems to be distinct from that in CphA1. The CphA1 N domain allows the end of the nascent polymer to transition between the two active sites; perhaps such a role is less important in an enzyme with one active site. Our data do indicate that the N domain is important for the stability of CphA2, and its mutation can influence activity by altering its oligomeric state, perhaps by blocking access to the active site within higher-order oligomers. This is hinted at by twinned structures of homologs that could be phased but not refined to a  $R_{\text{free}}$  below ~40%, where the G domain active site is blocked by an interaction with an N domain from a symmetry mate. It is conceivable that cyanobacteria could take advantage of the link between activity and oligomeric state to regulate CphA2 activity *in vivo*.

CphA2 is clearly most active when a primer is provided, but some CphA2 enzymes can synthesize cyanophycin in the absence of primer (or, more precisely, using a single  $\beta$ -Asp-Arg dipeptide as primer). We found no correlation between the relative rates of the primer-dependent and primer-independent activities of different CphAs, suggesting that the mechanisms controlling primer binding and rate of catalytic activity are not identical. Differences in binding of incoming substrate dipeptide or in sequence and flexibility of the in  $G_{\text{omega}}$  and  $G_{\text{lid}}$  could influence rate of polymerization in CphA2 enzymes. Nonetheless, that a single mutation in a primer-binding residue was sufficient to allow *G. citriformis* CphA2 to synthesize cyanophycin in the absence of primer, confirms that affinity for  $\beta$ -Asp-Arg dipeptide “primer” can be limiting for synthesis *in vitro*. Because the primer-independent activity is low (or absent) in all enzymes assayed, and because CphA2 will likely have existing cyanophycin chains to use as primer *in vivo*, primer independence is likely not important for the enzyme's activity in cyanobacteria, but it could be useful for biotechnological applications.

## **4.4. Methods**

### **4.4.1. Molecular biology**

*Anabaena* sp. UTEX2576 *cphA2* was cloned from genomic DNA, and the other *cphA2* genes were codon optimized for *E. coli* and synthesized by BioBasic or the DOE Joint Genome Institute. Genes were cloned into pJ411-derived plasmids encoding C-terminal tobacco etch virus (TEV) protease recognition sites and 8xHis affinity tags by transforming DH5- $\alpha$  *E. coli* cells with PCR fragments containing overlapping ends. CphAs were expressed in BL21(DE3) *E. coli* in TB media plus 150  $\mu$ g/ml kanamycin. Cells were grown at 37 °C to OD<sub>600</sub> ~1, before the temperature was lowered to 18 °C and protein expression was induced with 0.2mM isopropyl- $\beta$ -d-1-thiogalactopyranoside (IPTG). Cultures were incubated for ~20 hours before harvesting. Following centrifugation, cells were resuspended in buffer A (250 mM NaCl, 50 mM Tris pH 8, 10 mM imidazole, 2 mM  $\beta$ -mercaptoethanol) supplemented with a few lysozyme crystals and DNaseI, and lysed by sonication. The lysate was clarified by centrifugation at 40,000 g, and loaded onto a 5 ml HisTrap HP column (Cytiva), washed with 30 column volumes of buffer B (buffer A with 30 mM imidazole) and eluted with buffer C (buffer A with 250 mM imidazole). The proteins were then incubated with TEV protease (1:10 by mass) for tag removal during overnight dialysis against buffer D (250 mM NaCl, 20 mM Tris-HCl pH 8, 3 mM  $\beta$ -mercaptoethanol), prior to re-application to the HisTrap column. The flowthrough was concentrated using 100 kDa molecular weight cut off Amicon centrifugation concentrators (EMD Millipore) and applied to a Superdex 200 16/60 column (Cytiva) equilibrated in buffer E (100 mM NaCl, 20 mM Tris-HCl pH 8, 1 mM dithiothreitol). Purified protein was concentrated to 20 mg ml<sup>-1</sup> and, following the addition of glycerol to 10% v/v, flash frozen in liquid nitrogen and stored at -80 °C.

### **4.4.2. Crystallography**

Crystals were grown using the sitting drop method at 22 °C. *G. citriformis* CphA2 (12.5 mg ml<sup>-1</sup>) in buffer E was spiked with 1:5 v/v of 15 mM FOS-choline-12, before mixing 2  $\mu$ l samples with 1.6  $\mu$ l of well solution (0.1 M Tris-bicine pH 8.5, 8.8% w/v PEG8000, 17.6% v/v ethylene glycol, 60 mM MgCl<sub>2</sub>, 200 mM Na/K tartrate, 5 mM betaine and 3% v/v 2,2,2-trifluoroethanol) and 0.4  $\mu$ l 100 mM NiCl<sub>2</sub>, and equilibrating against well solution. Fully grown crystals were dehydrated for 24–48 hours by replacing the well solution with 0.1 M Tris-bicine pH 8.5, 16% w/v PEG8000, 32% v/v ethylene glycol, 60 mM MgCl<sub>2</sub>, 200 mM Na/K tartrate, 5 mM

betaine and 3% v/v 2,2,2-trifluoroethanol prior to looping and flash cooling in liquid nitrogen. Data were collected at the Canadian Light Source beamline CMCF-BM and processed using DIALS<sup>150</sup>. The structure was solved using PHASER implemented in CCP4i2<sup>184</sup> with an ensemble of 3 CphA1 structures<sup>2</sup> as a search model, and refined using Rosetta<sup>152</sup>, LORESTR<sup>187</sup>, Phenix<sup>188</sup> and Coot<sup>153</sup>.

#### **4.4.3. Size exclusion chromatography**

SEC was performed using a Superdex 200 increase 10/300 gl column (Cytiva) equilibrated in buffer E, run at 0.5 ml min<sup>-1</sup> and an injection volume of 150 µl. The column was calibrated with HMW standard (Cytiva).

#### **4.4.4. CphA2 activity assays**

Reactions contained 2 µM CphA2, 100 mM HEPES pH 8.2, 40 mM KCl, 10 mM MgCl<sub>2</sub>, 2 mM β-Asp-Arg, 2 mM ATP, and 50 µM cyanophycin primer. NaCl was added in some experiments as indicated. Bis-tris propane replaced HEPES in the pH dependency assays. Reactions were carried out in quadruplicate at 23 °C, in 96-well plates with volumes of 100 µl. OD<sub>600</sub> was monitored using a SpectraMax Paradigm spectrophotometer (Molecular Devices), with 5 seconds shaking between reads. Data were analyzed using GraphPad Prism. Rates were calculated using the maximum of the first derivative of OD<sub>600</sub> curves, with derivatives curves smoothed using a 2<sup>nd</sup> order polynomial to reduce noise.

#### **4.4.5. Synthesis of cyanophycin segments**

Solid phase synthesis using Fmoc-(β-Asp-Arg)(OtBu)-OH building blocks was performed as previously described<sup>2,65,159</sup>.

#### **4.4.6. Differential scanning fluorimetry**

CphA2 (20 µl of 0.5 mg ml<sup>-1</sup>) in buffer E and 5x Sypro<sup>TM</sup> Orange was heated from 5–95 °C over 2 hours in a One Step Plus RT-PCR (Applied Biosystems).

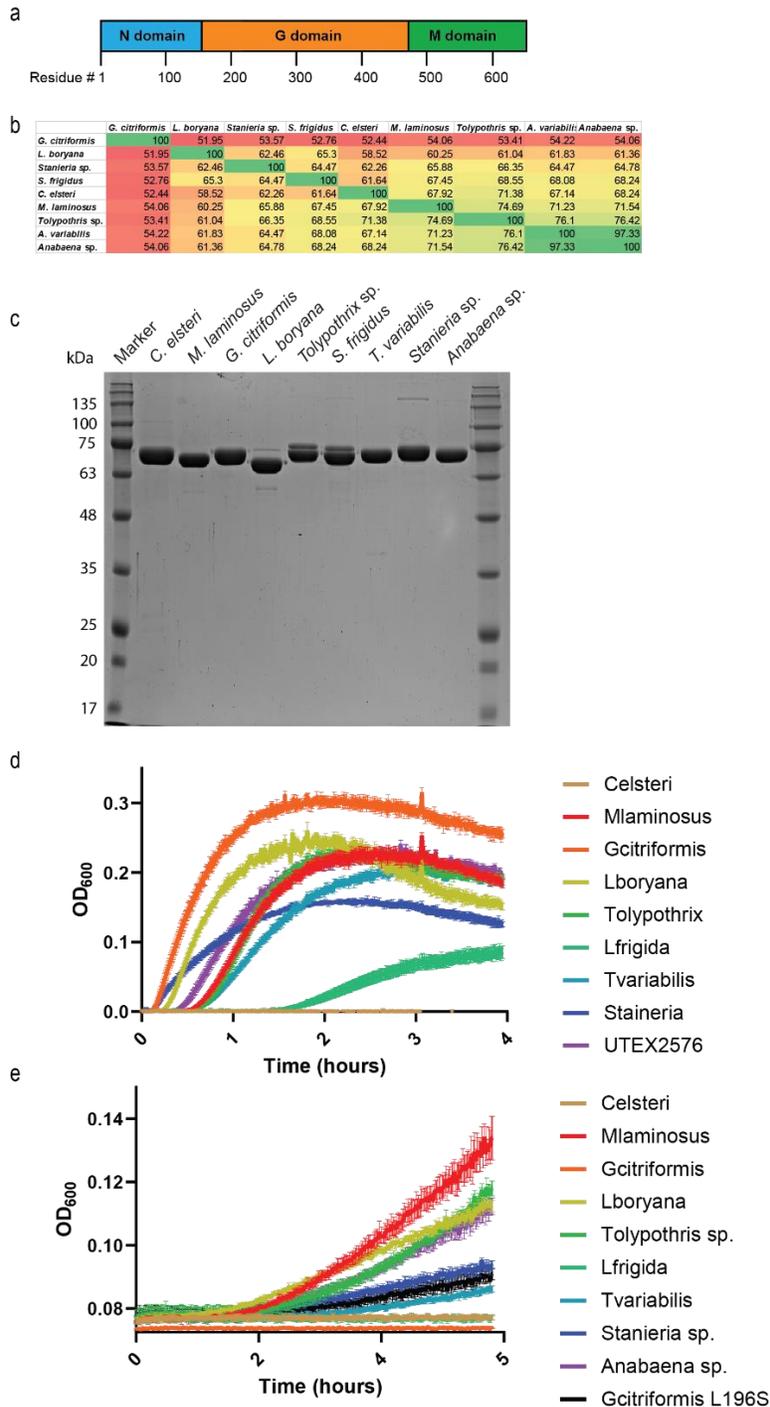
### **4.5. Accession codes**

Atomic coordinates and structure factors for CphA2 been deposited in the Protein Data Bank under accession code 7TA5.

#### **4.6. Acknowledgements**

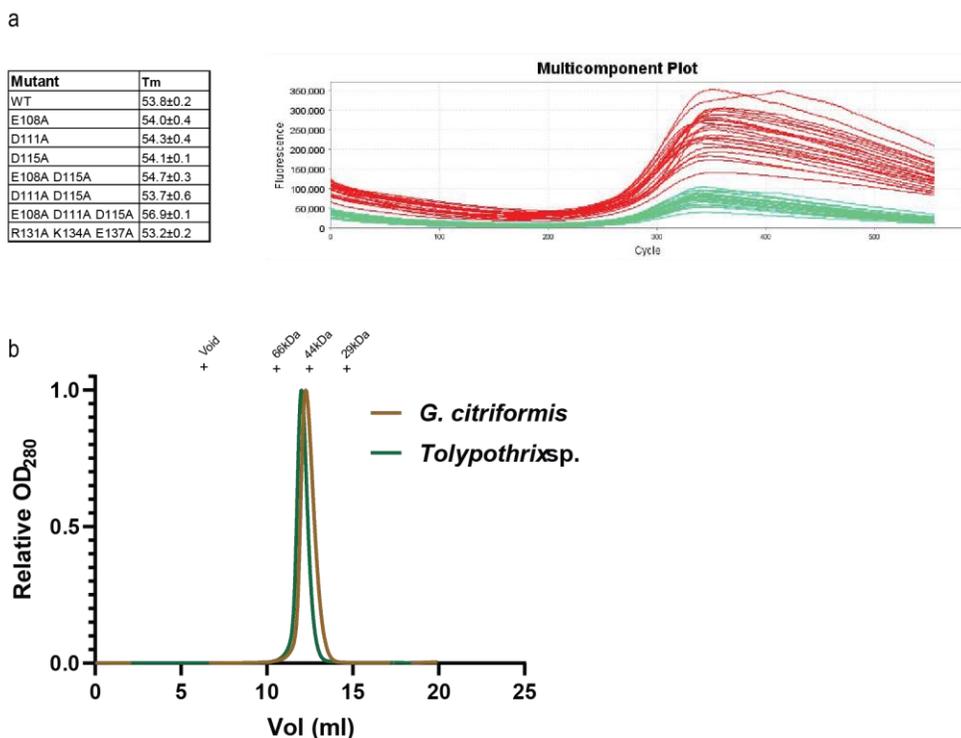
We thank all the members of the Schmeing lab for important advice and ongoing discussions on this project and synchrotron staff S. Labiuk, K. Janzen and K. Mundboth (Canadian Light Source) for facilitating remote collection of diffraction datasets. Data was collected using beamline 08ID-1 at the Canadian Light Source, a national research facility at the University of Saskatchewan which is supported by the Canada Foundation for Innovation (CFI), the Natural Sciences and Engineering Research Council (NSERC), the National Research Council (NRC), the Canadian Institutes of Health Research (CIHR), the Government of Saskatchewan, and the University of Saskatchewan.

## 4.7. Supplementary information

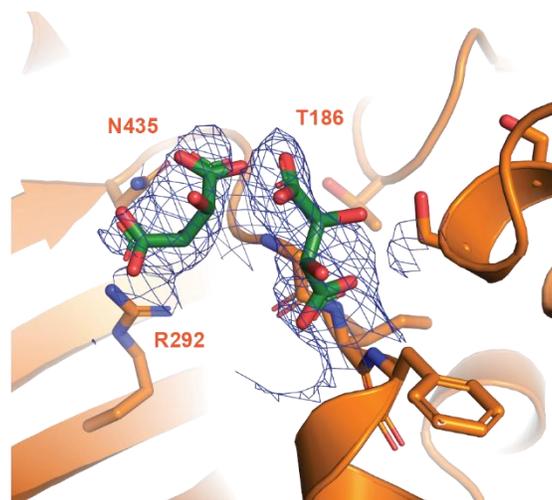


**Supplementary Figure 4.1. Biochemical characterization of the nine homologs used in this study.** (a) Domain organization of CphA2. (b) Pairwise sequence identity matrix of the 9 CphA2 enzymes used in this study. (c) An SDS-PAGE of the 9 CphA2 enzymes used in this study. All enzymes have similar molecular masses, ranging between 71-74 kDa. (d) Activity assay plots of

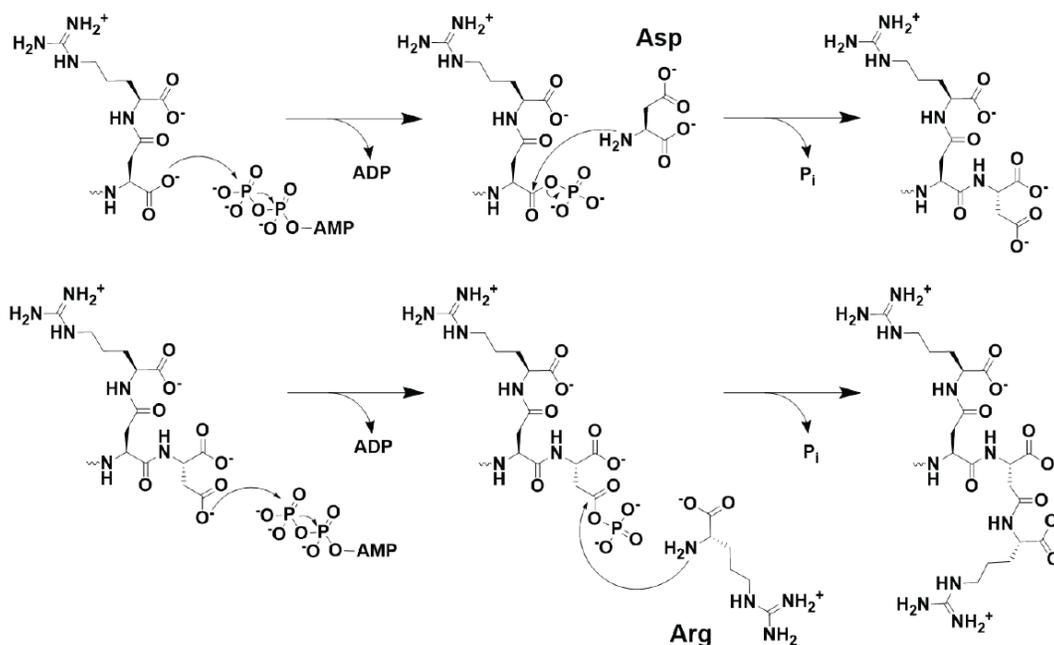
reactions of the 9 CphA2 enzymes in the presence of 50  $\mu$ M ( $\beta$ -Asp-Arg)<sub>3</sub> as primer. The maximal rates of these curves were used to generate the data in figure 4.2a. (e) Activity assay plots of reactions of the 9 CphA2 enzymes and the *G. citriformis* mutant L196S in the absence of primer. The maximal rates of these curves were used to generate the data in figure 4.2c. All curves are an average of 4 independent measurements. Error bars represent standard deviations.



**Supplementary Figure 4.2. Biochemical characterization of *G. citriformis* N-domain mutants and purified CphA2 N domain.** (a) Average T<sub>m</sub> values (table) and melting curves of *G. citriformis* WT and N domain mutants, obtained by DSF. All enzymes displayed similar T<sub>m</sub> values, suggesting that the observed differences in activity between them are not the result of differences in protein stability. (b) SEC chromatograms of N domain constructs of *G. citriformis* and *Tolypothrix* sp. Both constructs migrate as a single peak and show no sign of aggregation, suggesting they are stable in solution. In contrast, constructs of the *G. citriformis* and *Tolypothrix* sp enzymes lacking the N domain could not be expressed in soluble form.



**Supplementary Figure 4.3. The two tartrate molecules observed in the active site of *G. citriformis* CphA2.** The high tartrate concentration in the crystallization condition (200 mM) prevented the  $(\beta\text{-Asp-Arg})_3$  (or  $\beta\text{-Asp-Arg})_4$  primers used in co-complex experiments from binding to the enzyme. Co-crystallization and crystal soaking with 1 mM substrate did not result in density for cyanophycin primer in the calculated maps. The figure was made using the 2Fo-Fc map of crystals with no added cyanophycin, contoured at a level of  $1\ \sigma$  and carved  $3\ \text{\AA}$  around the tartrate molecules.



**Supplementary Figure 4.4. Cyanophycin synthesis by CphA1.** First, the G domain adds an Asp residue to the polymer's backbone, in a reaction similar to that catalyzed by the G domain of CphA2. Then, the M domain decorated the side chain of this Asp with an Arg residue.



**Supplementary Table 4.1. X-ray data collection and refinement statistics.**

---

<i>G. citriformis</i> CphA2	
<b>Data collection</b>	
Space group	P6 <sub>4</sub> 22
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	99.8, 99.8, 348.5
$\alpha$ , $\beta$ , $\gamma$ (°)	90.0, 90.0, 120.0
Resolution (Å)	348.5-3.0 (3.1-3.0)
<i>R</i> <sub>merge</sub>	0.04381 (0.6688)
<i>R</i> <sub>pim</sub>	0.04381 (0.6688)
<i>I</i> / $\sigma$ <i>I</i>	18.3 (0.5)
CC <sub>1/2</sub>	0.999 (0.464)
Completeness (%)	99.94 (100.00)
Redundancy	105.6 (105.0)
<b>Refinement</b>	
Resolution (Å)	86.41-3.0
No. reflections	43186
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.2536/0.2833
No. atoms	4746
Protein	4696
Ligand/ion	39
Solvent	11
<i>B</i> -factors	
Protein	99.8
Clashscore	2.55
Molprobity score	1.17
R.m.s. deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.66
<b>Deposition</b>	
PDB ID code	7TA5

---

## Bridge to chapter 5

Chapters 2 to 4 of this thesis described the biosynthesis of cyanophycin. However, a complete picture of its metabolism also requires information about the degradative steps involved. Despite its peptide-like nature, cyanophycin is resistant to proteolytic degradation. The only known pathway for cyanophycin degradation to amino acids starts with cleavage of the polymer's backbone peptide bonds, which results in the release of  $\beta$ -Asp-Arg dipeptides. This is done by cyanophycinase – a specialized serine-protease like enzyme found in many bacteria and some fungi. Many studies isolated and characterized cyanophycinases, and one even described the enzyme's structure. However, no co-complex structure of cyanophycinase with cyanophycin was available, and our understanding of how the enzyme binds its substrate relied on modelling and mutagenesis experiments. In our lab, we have a system for the incorporation of the unnatural amino acid diaminopropionic acid (DAP) into proteins. This presented me with a great opportunity to closely examine cyanophycinase's activity. By replacing the active site Ser of cyanophycinase with DAP, I was able to trap the covalent enzyme-substrate intermediate complex and use it for crystallographic studies.

5. The structure of cyanophycinase in complex with a cyanophycin degradation intermediate

Published in: Sharon I, Grogg M, Hilvert D, Schmeing TM. *Biochimica et Biophysica Acta - General Subjects*. 2022 November 1866(11), 130217. Online 26 July 2022.

## **5.1. Abstract**

Cyanophycinases are serine protease family enzymes which are required for the metabolism of cyanophycin, the natural polymer multi-L-arginyl-poly(L-aspartic acid). Cyanophycinases degrade cyanophycin to  $\beta$ -Asp-Arg dipeptides, which enables use of this important store of fixed nitrogen. We used genetic code expansion to incorporate diaminopropionic acid into cyanophycinase in place of the active site serine, and determined a high-resolution structure of the covalent acyl-enzyme intermediate resulting from attack of cyanophycinase on a short cyanophycin segment. The structure indicates that cyanophycin dipeptide residues P1 and P1' bind shallow pockets adjacent to the catalytic residues. We observe many cyanophycinase - P1 dipeptide interactions in the co-complex structure. Calorimetry measurements show that at least two cyanophycin dipeptides are needed for high affinity binding to cyanophycinase. We also characterized a putative cyanophycinase which we found to be structurally very similar but that shows no activity and could not be activated by mutation of its active site. Despite its peptidic structure, cyanophycin is resistant to degradation by peptidases and other proteases. Our results help show how cyanophycinase can specifically bind and degrade this important polymer.

## **5.2. Introduction**

Cyanophycin is a biopolymer produced by many bacteria <sup>2,85,189</sup>. It is composed of a polypeptide backbone of L-aspartic acid residues, with L-arginine attached to each Asp side chain through an isopeptide bond (Fig. 5.1a). Cyanophycin polymer chains are ~80-400 dipeptides in length and aggregate to form membrane-less insoluble granules within the bacterial cells that produce them. The granules are used as a store of nitrogen, carbon and energy <sup>40,42</sup>, which are accessed upon cyanophycin degradation. Cyanophycin has many potential industrial applications. For example, it is a source of the biodegradable polymer poly-Asp (used as a water softener and super-swelling material) <sup>71</sup> and of the  $\beta$ -dipeptide  $\beta$ -Asp-Arg, which can be used as a nutrition supplement <sup>190</sup>.

Cyanophycin biosynthesis is accomplished by one of two enzymes: Cyanophycin synthetase 1 (CphA1) makes cyanophycin by polymerizing Asp and Arg in iterative cycles of backbone elongation and side chain modification <sup>2,98</sup>, while cyanophycin synthetase 2 (CphA2) polymerizes  $\beta$ -Asp-Arg dipeptides <sup>12,13</sup>. Despite having a peptidic structure, cyanophycin is resistant to degradation by conventional proteases <sup>16</sup>. Instead, its degradation occurs in two sequential reactions. First, cyanophycin is hydrolyzed to  $\beta$ -Asp-Arg dipeptides by the enzyme cyanophycinase (Fig. 5.1a) <sup>121</sup>. The dipeptides are then further hydrolyzed to Asp and Arg by a variety of enzymes with isoaspartyl dipeptidase activity <sup>129</sup>. Enzymes capable of performing the latter step are present in most bacteria and can generally cleave a variety of  $\beta$ -aspartyl substrates <sup>129</sup>, but cyanophycinase is specialized for cyanophycin degradation. Heterologous co-expression of cyanophycin synthetase and cyanophycinase is being pursued for production of  $\beta$ -Asp-Arg dipeptides for agricultural feed supplementation <sup>190-193</sup>.

Cyanophycinases are members of the serine protease family of enzymes. They can exist as intracellular dimers (CphB) <sup>121</sup>, intracellular pseudodimers (CphI) <sup>100</sup> and secreted, extracellular monomers (CphE) <sup>61</sup>. CphB from *Synechocystis sp.* PCC6803 and CphE from *Pseudomonas anguilliseptica* BI have been shown to display C-terminal exocyanophycinase activity, hydrolyzing  $\beta$ -Asp-Arg dipeptides from cyanophycin <sup>61,121</sup>. They are highly specific and display low or no activity toward other substrates that contain peptide or isopeptide bonds <sup>61,121</sup>. While *cphB* is found in bacteria that have *cphA1*, *cphE* is also present in bacteria and fungi that do not appear to have cyanophycin synthesis capabilities, suggesting that these microbes scavenge

cyanophycin<sup>58-60</sup>. Indeed, bacterial species and consortia can be isolated that can use extracellular cyanophycin as their sole carbon and nitrogen source<sup>57-59,61</sup>.

Law et al.<sup>124</sup> determined the crystal structure of CphB from *Synechocystis sp.* PCC6803 and used mutagenesis to identify residues that are important for P1' substrate recognition and enzymatic activity. Their results suggested a mechanism for cyanophycin recognition by CphB. However, they were unable to crystallize the enzyme in the presence of cyanophycin and relied on modeling to predict how CphB binds substrate. While their predictions were reasonable and supported by the accompanying biochemical experiments, there is value in determining experimental co-complex structures for direct insight into the enzyme's function.

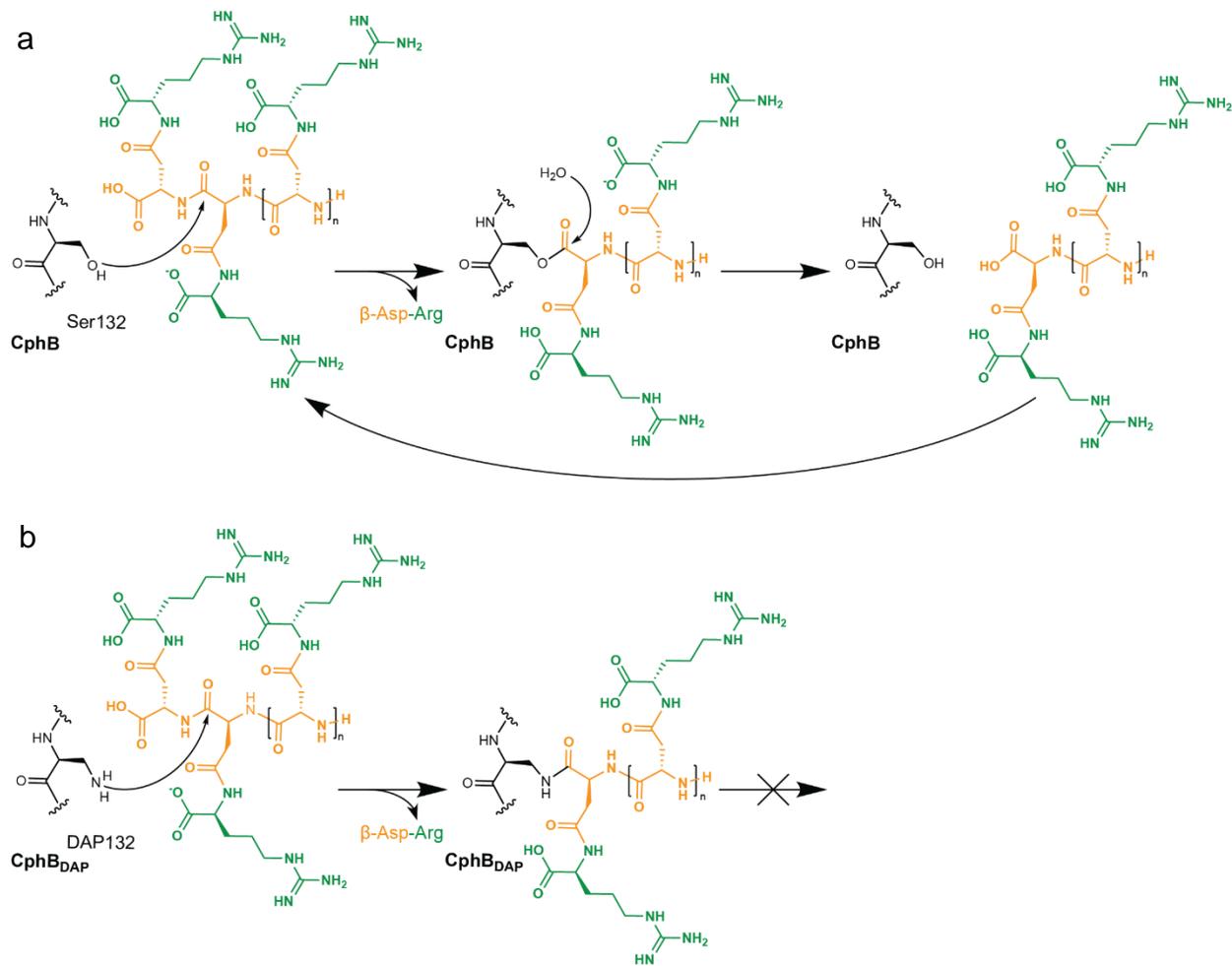
In this study, we used a system to incorporate the unnatural amino acid diaminopropionic acid (DAP)<sup>123</sup> into *Synechocystis sp.* PCC6803 CphB (SyCphB), in place of the catalytic nucleophile S132. This allowed us to form a stable SyCphB<sub>DAP</sub>-cyanophycin intermediate complex suitable for structure determination and thus directly visualize the complex that forms during degradation. We also solved the structure of an inactive CphB-like protein from *Pseudobacteroides cellulosolvens* (PcCphB), and by comparing the two structures were able to gain further insights into important structural aspects of the CphB active site. We used isothermal titration calorimetry (ITC) and *in vitro* activity assays to support the conclusions drawn from the structures. The new data give a more complete understanding of the way cyanophycinase recognizes its substrate and cleaves cyanophycin, one  $\beta$ -Asp-Arg dipeptide at a time.

### **5.3. Results and discussion**

#### **5.3.1. Structure of the CphB – cyanophycin covalent co-complex**

Law et al.<sup>124</sup> determined a high quality crystal structure of apo CphB. Unfortunately, their attempts to crystallize the enzyme in the presence of  $\beta$ -Asp-Arg dipeptides yielded no density for these molecules<sup>124</sup>. This is not overly surprising, as the  $\beta$ -Asp-Arg dipeptide product may have low affinity for CphB and the crystallization conditions may not be favorable for its binding to the enzyme. To overcome these obstacles, we sought to form a stable CphB co-complex for structural studies. CphB uses the classic serine protease catalytic mechanism with two half-reactions<sup>194</sup>: First, the active site serine attacks the scissile amide of the substrate, lysing the peptide bond and forming an acyl-enzyme intermediate. Then, in a second step, the intermediate is resolved by hydrolysis of the labile ester bond (Fig. 5.1a). Replacement of the active site serine with the

unnatural amino acid DAP enables trapping of the intermediate<sup>123</sup>: The amine of DAP attacks the peptide nucleophilically, but the resulting adduct is covalently linked to the enzyme via an amide bond, resulting in a stable complex amenable to structural study (Fig. 5.1b).



**Figure 5.1. CphB reaction and inhibition mechanisms.** (a) The reaction catalyzed by cyanophycinase. Catalytic S132 attacks the scissile peptide bond, forming a covalent CphB-cyanophycin intermediate, which is then hydrolyzed by a water molecule. (b) The mechanism of SyCphB<sub>DAP</sub>-cyanophycin complex formation. The DAP residue attacks the scissile amide bond, forming a covalent enzyme-substrate intermediate linked by an amide bond. The stability of this bond protects it from hydrolysis.

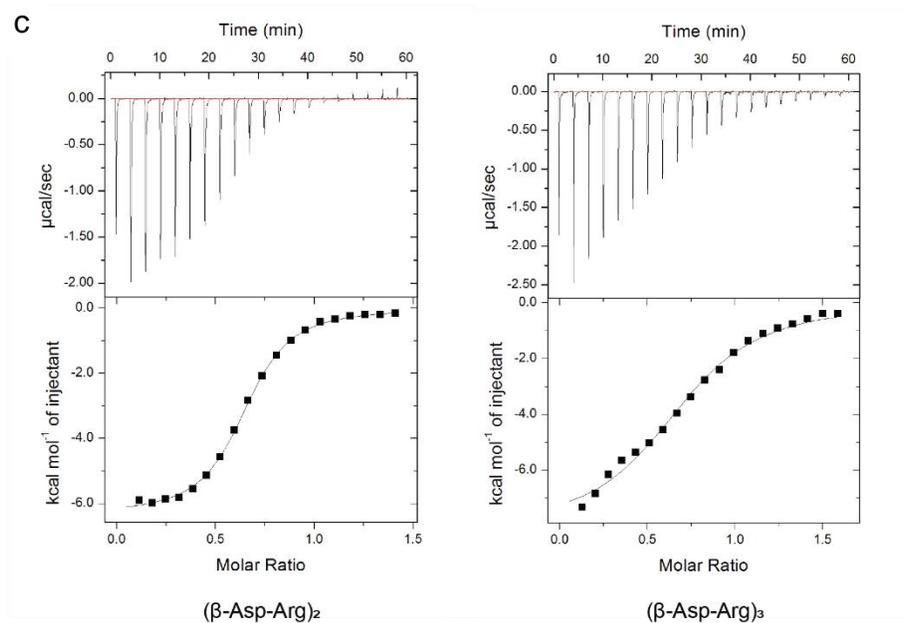
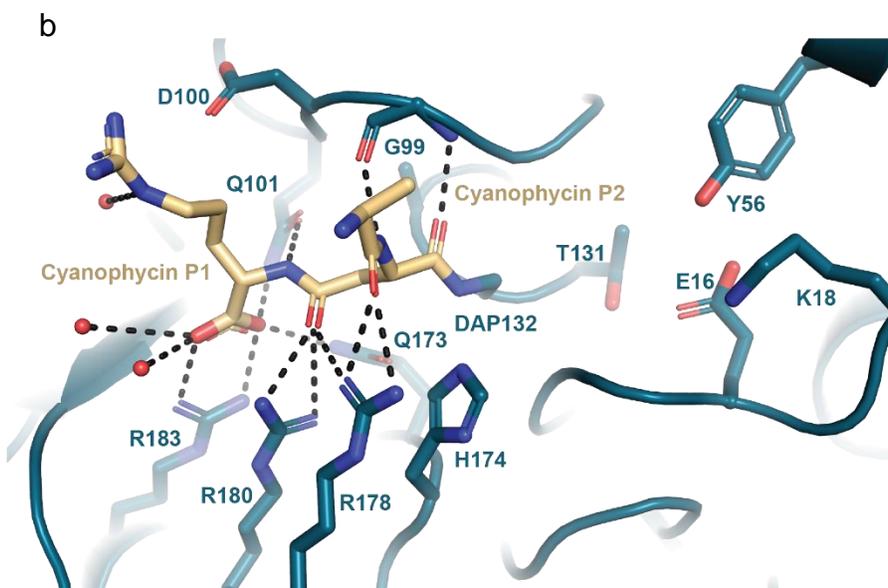
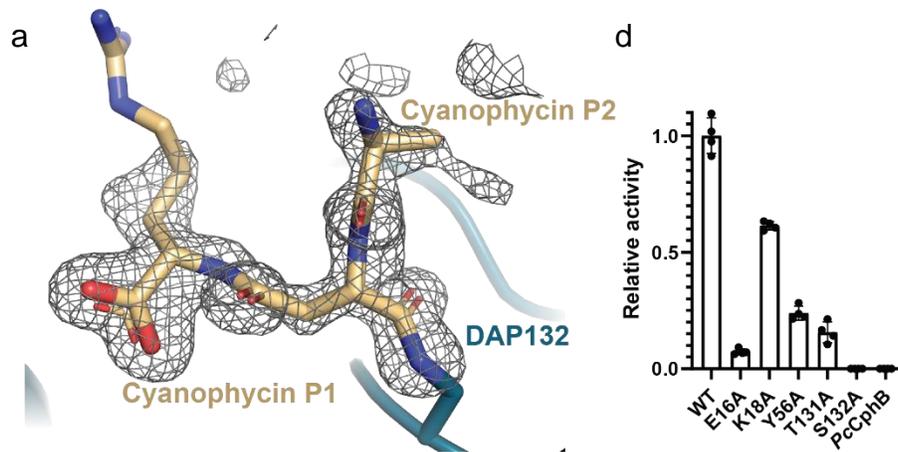
We introduced DAP into *Synechocystis sp.* PCC6803 CphB (SyCphB) by amber suppression of a TAG codon encoding residue 132 with a previously-reported DAP-incorporation system<sup>123</sup>. We then incubated the resulting SyCphB<sub>DAP</sub> with (β-Asp-Arg)<sub>3</sub>. Intact protein MS showed the formation of a peak corresponding to the combined masses of SyCphB<sub>DAP</sub> and (β-Asp-

Arg)<sub>2</sub> (Supplementary Fig. 5.1a), consistent with the formation of a covalent acyl-enzyme intermediate following cleavage of the C-terminal P1' dipeptide (Fig. 5.1b). We crystallized the complex in similar conditions to those previously reported<sup>124</sup> and solved the structure in the same space group, to 1.5 Å resolution (Supplementary Table 5.1).

There are three SyCphB<sub>DAP</sub> monomers in the asymmetric unit in near-identical conformations, with a C<sub>α</sub> RMSD of 0.70-0.79 Å. Unbiased maps showed extra density extending from the nitrogen of the active site DAP (Fig. 5.2a). We fit the P1 β-Asp-Arg dipeptide of cyanophycin and some of the backbone atoms of the P2 dipeptide into that density. Weak density is present for additional P2 dipeptide atoms, but it is not definitive enough to enable modeling. No clear density was observed for the cleaved P1' dipeptide. The structure of SyCphB<sub>DAP</sub>, including H174 and E201, which complete the catalytic triad with S132, is similar to the wildtype enzyme (RMSD of 0.42 Å, Supplementary Fig. 5.1b,c). The attached segment of cyanophycin is bound in a broadly similar way to that shown in the prediction figure of Law et al.<sup>124</sup>. Arg residues R178, R180 and R183 interact with one face of the P1 dipeptide of cyanophycin through its carbonyl oxygens, with additional hydrogen bonding through G99, Q101, Q172, and G198 (Fig. 5.2b). The strength of the density decreases along the Arg moiety of the P1 dipeptide, indicating that the Arg carboxylate is more important for binding CphB than its guanidinium moiety. Cyanophycin can contain up to 25% Lys residues in place of Arg residues<sup>96</sup>, so stringent specificity for Arg over Lys would not be desired. The backbone carbonyl of the P2 dipeptide interacts with R178, but this dipeptide residue is otherwise mostly disordered.

### 5.3.2. Cyanophycin segment binding assays

The structural results suggest that binding of cyanophycin relies largely on the P1 dipeptide. We performed isothermal titration calorimetry (ITC) using the inactive S132A mutant and ligands β-Asp-Arg, (β-Asp-Arg)<sub>2</sub> and (β-Asp-Arg)<sub>3</sub> (Fig. 5.2c). No measurable signal was observed with the β-Asp-Arg dipeptide, which is the CphB product. In contrast, a strong signal was observed for (β-Asp-Arg)<sub>2</sub> ( $\Delta H = -3789 \pm 170$  cal/mol,  $\Delta S = 10.3 \pm 0.8$  cal/mol/deg) and the K<sub>d</sub> was measured to be  $9.3 \pm 0.5$  μM (Fig. 5.2c). With (β-Asp-Arg)<sub>3</sub>, there is clear binding, with a consistent shape of the thermogram which suggests that there are two binding modes. However, the K<sub>d</sub> values of these two modes are very similar, so we could not reliably fit a two binding site model to the data and used a one binding site model. This gave a calculated K<sub>d</sub> of  $25.1 \pm 2.1$  μM,



**Figure 5.2. The binding of cyanophycin by CphB.** (a) Polder map calculated around the  $(\beta\text{-Asp-Arg})_2$  product covalently bound to  $\text{SyCphB}_{\text{DAP}}$ . The map is calculated to 1.5 Å resolution, shown at 3 sigma and within 3 Å of the cyanophycin molecule. Most of the observed density corresponds to the P1 dipeptide. (b) The active site of  $\text{SyCphB}_{\text{DAP}}$  showing interactions between the enzyme and the covalently bound  $(\beta\text{-Asp-Arg})_2$ . Five hydrogen bond donors are observed within 2.7 - 3.1 Å of the P1 carboxylate, which over-satisfies the carboxylate, but such over-satisfaction is not altogether uncommon <sup>5</sup>. (c) Plots of ITC experiments conducted with  $\text{SyCphB S132A}$  and  $(\beta\text{-Asp-Arg})_2$  (left) or  $(\beta\text{-Asp-Arg})_3$  (right). Calculated  $K_d$  values were  $9.3 \pm 0.5 \mu\text{M}$  for  $(\beta\text{-Asp-Arg})_2$  and  $25.1 \pm 2.1 \mu\text{M}$  for  $(\beta\text{-Asp-Arg})_3$ . (d) Cyanophycinase activity of  $\text{SyCphB}$ ,  $\text{SyCphB P1}'$ -binding mutants, and  $\text{PcCphB}$ . All  $\text{SyCphB}$  mutants displayed reduced activity compared to the wildtype, and  $\text{PcCphB}$  (and mutants thereof) displayed no cyanophycinase activity.

somewhat higher than that for  $(\beta\text{-Asp-Arg})_2$ , although more heat was released upon its binding to CphB ( $\Delta H = -5837 \pm 171 \text{ cal/mol}$ ,  $\Delta S = 1.5 \pm 0.9 \text{ cal/mol/deg}$ ; Fig. 5.2c). We also performed ITC experiments with  $(\beta\text{-Asp-Arg})_4$ , which unlike  $(\beta\text{-Asp-Arg})_3$  exhibited only one binding mode. However, the compound exhibited poor solubility (evidenced by calculated stoichiometry  $n \sim 0.25$ ), so we did not include these data in our analysis.

CphB residues involved in binding the P1 dipeptide were previously interrogated by mutagenesis <sup>124</sup>, and D17, D100, Q101, D158, D172, Q173, R178, R180, R183 and D202 were all found to be important for CphB's activity. However, a potential binding site for the cleaved P1' dipeptide had not been studied, and we sought to determine whether it made any significant contribution to the enzyme's activity. To that end, we performed activity assays with wildtype  $\text{SyCphB}$  and several mutants targeting conserved residues that form a shallow pocket opposite to where the P1 dipeptide binds: E16A, K18A, Y56A and T131A (Fig. 5.2c). All four mutants displayed reduced activity, to between ~5% and ~60% of wildtype (Fig. 5.2d). Based on the distance between these residues and the scissile amide bond, it is possible that the mutated residues contact the amide and carboxylate moieties of the P1' dipeptide.

The lack of any density for the cleaved P1' dipeptide in the  $\text{SyCphB}_{\text{DAP}}$ -cyanophycin map and the lack of observed binding of  $\beta\text{-Asp-Arg}$  in ITC indicates that the dipeptide product does not have high affinity for CphB. However, both the P1 and the P1' dipeptide binding pocket are

clearly important for productive binding because mutations of residues lining those sites reduce cyanophycinase activity (Fig 5.2c and <sup>124</sup>). It is very likely that the high-affinity binding of ( $\beta$ -Asp-Arg)<sub>2</sub> seen in ITC represents the substrate state (binding to P1-P1' pockets) rather than the product state (binding to P2-P1 pockets, where the P2 “pocket” is only an interaction with R178). Neither the P1 nor P1' pocket provides high affinity on its own (at least in the S132A mutant used in ITC), but together they form the productive cyanophycinase substrate site, positioning the scissile bond above Ser132 for cleavage.

### 5.3.3. *Pseudobacteroides cellulosolvans* pseudo-CphB

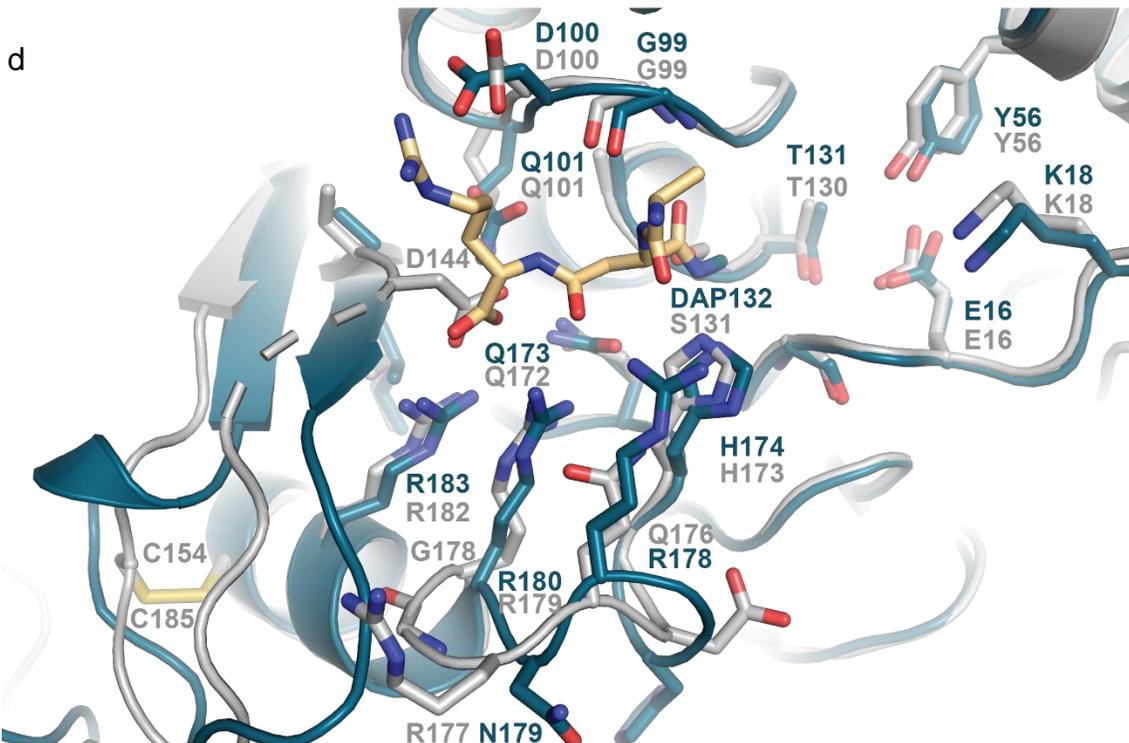
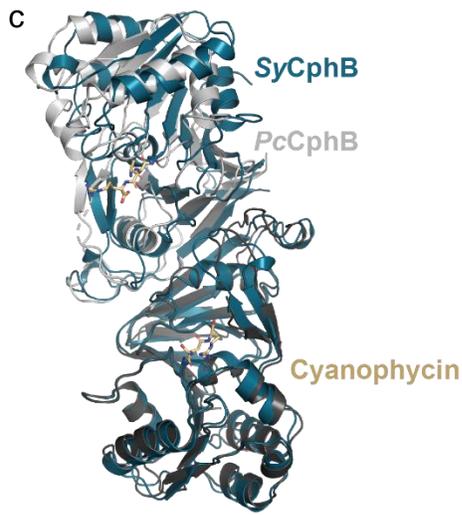
We sought to determine the structure of a second CphB enzyme for additional insight. The genome of *Pseudobacteroides cellulosolvans* DSM2933 includes two *cphA1* genes and a gene annotated as *cphB* (encoding *PcCphB*). This gene is adjacent to a *cphA1* gene, and *PcCphB* has 33% identity and 57% similarity to *SyCphB*, with most active site residues conserved between the sequences (Supplementary Fig. 5.2). We cloned, expressed and purified *PcCphB* for biochemical and structural studies. To our surprise, the enzyme displayed no cyanophycinase activity (Fig. 5.2d).

To understand the structural basis for this inactivity, we solved the crystal structure of *PcCphB* to 2.4 Å resolution (Fig. 5.3a, Supplementary Table 5.1). The asymmetric unit is composed of two dimers, and protomers within each dimer are arranged in a similar way to that of *SyCphB* (Fig. 5.3b). The main difference between the dimer architectures of the two proteins is that the angle between two monomers of *PcCphB* is ~10° larger than that of *SyCphB* (Fig. 5.3b). This leads to a slightly more “open” conformation, but this does not seem to affect the accessibility of the active site. The protomers of the two proteins have similar structures overall, with an RMSD of 2.8 Å between subunits A of each protein (Fig. 5.3b). As predicted by sequence alignment, the structure of the active sites of the two proteins is very similar (Fig. 5.3c, Supplementary Fig. 5.2). Almost all the residues known to be involved in substrate binding and catalysis are present in both proteins, and the structure of the active site and surrounding areas is much the same (Fig. 5.3c).

This high similarity between the active and inactive proteins led us to consider whether two structural differences between them might be the source of *PcCphB*'s inactivity. First, the 175-178 loop in *PcCphB* (sequence DQRG) differs from its counterpart in *SyCphB*, 176-179 (HNRN). Consequently, cyanophycin binding residue *SyCphB* R178 is structurally replaced by Q176 in *PcCphB* (Fig. 5.3c). The shorter Q176 is unlikely to be able to contact cyanophycin. We

**a**

PCC6803	MELSSQPALLIIGGAE <b>EDK</b> VHGREILQTFWRSRGGNDIIGIIPASREPLLIGER <b>Y</b> QTIF 60
DSM2933	MEEKSKGNLVIIGGAE <b>EDK</b> KGESKILKKVAEIAFGDMEFIVLTATEHPVEVGNE <b>Y</b> LNVF 60
	* :*:***** :**... :* .* : : :*:...*: :*:.* :*:*
PCC6803	SDMGVKELKVLDIRDRAQGGDDSGYRLFVEQCTGIFMTG <b>GDQ</b> LRLCGLLADTFLMDRIRQR 120
DSM2933	QRLGINNIEVLDISTRDANNEENYKIVNSGGVFTMG <b>GDQ</b> LRITGILGGTKVFNALIEA 120
	. :*:::**** * :... : : . :*****: :*.* : : : :
PCC6803	VHNGEISLAG <b>TS</b> GAGAAVMGHMIAGSSGGEWENRALVDMAVGLSIVPEIVD <b>QH</b> FHNR <b>R</b> 180
DSM2933	YLKG-VVIAG <b>TS</b> GASVMSNTMIVDGNNDPARKCTLKMASGLGLEEAI <b>DQH</b> FD <b>QR</b> 179
	* : :*****:*: :*..*.*: :. :.* *****: * :*****:*
PCC6803	MARLLSAISTHPELLGL <b>IDE</b> DTCAMFERDGSVKVIGQGTVSFVDARDMSYTNAAVLGAN 240
DSM2933	FG <b>RL</b> LLCGVAENPHMLG <b>IDE</b> DTAIRVYFDAHFEVVGSYAVTIIDGKSIVSSNVSELKPD 239
	:.****. :*:**:*..... :*..**.*: :*:...*.: :*: : :
PCC6803	APLSLHNLRLNLLVHGEVIHQVQR <b>AF</b> PRV- 270
DSM2933	EILAIANVTVHVLPEGY <b>GF</b> DMK-RREVLRLH 269
	*: :* :*:.* :. :* :* :



**Figure 5.3. The structure of *PcCphB*.** (a) The crystal structure of *PcCphB*. (b) Structural alignment of chains B of *PcCphB* (gray) and *SyCphB*<sub>DAP</sub>-cyanophycin (teal) shows that the two enzymes have very similar overall folds, with an RMSD of 2.8 Å between chains A of each. (c) Structural alignment of the active sites of *PcCphB* (gray) and *SyCphB*<sub>DAP</sub>-cyanophycin (teal) shows that most of their polymer-binding residues are conserved. However, the loop containing *PcCphB* R177 adopts a different conformation from that of *SyCphB*, preventing this residue from binding cyanophycin. In addition, the disordered loop following *PcCphB* D144 pushes it into the binding pocket of the P1 dipeptide, perhaps preventing cyanophycin binding. *PcCphB* that were mutated in an attempt to make the protein active are labeled (\*).

therefore introduced two mutations into *PcCphB* (D175H and G178N) to make the loop more like that of *SyCphB*, but the double mutant remained inactive.

A second structural difference is in the *PcCphB* loop M141-S160 (*SyCphB* M142-V161), which is between two strands of a central  $\beta$  sheet (Fig. 5.3c). This region, located ~6-25 Å from the catalytic serine, has lower sequence similarity than the active site residues and includes a C154-C185 disulfide bond in *PcCphB* (Fig. 5.3c). Perhaps because of these differences, the sidechain of *PcCphB* D144 is positioned where it could clash with the incoming cyanophycin (Fig. 5.3c). We constructed *PcCphB* mutants C154A (to break the disulfide), D144G and D144A (to resolve the clash), as well D144G\_C154A\_D175H\_G178N (combining these with the active site mutations), but none of these substitutions imparted activity to *PcCphB*.

The inactivity of *PcCphB* is interesting, as sequence alignment with *CphB* gives no clue to other plausible reasons for its inactivity. It is possible the enzyme is actually active but very sensitive to reaction conditions, and we have not found appropriate solutions. More likely, the accumulation of structural mutations renders the active site incapable of binding cyanophycin, even though the catalytic residues and most substrate binding residues are maintained. We performed ITC experiments of binding between wildtype *PcCphB* and ( $\beta$ -Asp-Arg)<sub>2</sub>, and could not detect substantial binding, supporting this conclusion. This protein's inactivity raises the question of whether *P. cellulosolvens* DSM2933 is indeed active in cyanophycin biosynthesis, and if the two *cphA1* copies in *P. cellulosolvens* DSM2933 are also pseudogenes, despite no telltale signs of inactivation in their sequence.

In all, this study contributes insight into the atypical serine protease cyanophycinase through co-complex structural biology, cyanophycin hydrolysis assays, mutagenetic interrogation and binding assays. The substrate specificity of typical serine proteases is determined by binding pockets in the active site that accommodate the side chains of peptide substrates. As pointed out by Law et al. <sup>124</sup>, these pockets are typically too small to allow binding of the  $\beta$ -Asp-Arg dipeptides, which are larger than canonical proteinogenic amino acid residues, making cyanophycin resistant to proteolytic degradation <sup>16</sup>. Cyanophycinase solves this problem by binding the P1' and P1 dipeptides in two very shallow pockets on either side of the catalytic triad (Supplementary Fig. 5.1d). This positions the dipeptides in a conformation that exposes the scissile bond and allows the catalytic serine to access it despite the bulky dipeptides surrounding it, enabling robust cyanophycin hydrolysis.

## **5.4. Material and methods**

### **5.4.1. Cloning, protein expression and purification**

The *Synechocystis* sp. PCC6803 *cphB* gene (protein WP\_010872518.1) was codon optimized for expression in *E. coli* and synthesized (BioBasic). *P. cellulosolvens* DSM2933 *cphB* (protein WP\_036936401.1) was PCR-amplified from genomic DNA (DSMZ). Both genes were cloned into a pCDF-derived plasmid with a C-terminal TEV cleavage site followed by an octa-His tag. All cloning and mutagenesis steps were performed by transforming *E. coli* DH5- $\alpha$  cells with PCR fragments containing overlapping ends. Expression of protein that did not contain DAP was carried out in *E. coli* BL21(DE3) cells in terrific broth (TB) media. One liter of media supplemented with 100  $\mu$ g/ml spectinomycin was inoculated with 10 ml of an overnight culture and grown at 37 °C until OD600 reached ~1.0. The temperature was lowered to 16 °C, and protein expression was induced with 0.2 mM isopropyl  $\beta$ -d-1-thiogalactopyranoside (IPTG) for 16 hours. All purification steps were carried out at 4 °C. Following centrifugation, the cells were resuspended in buffer A (250 mM NaCl, 50 mM Tris-HCl pH 8.0, 10 mM imidazole, 2 mM  $\beta$ -mercaptoethanol) supplemented with a few crystals of lysozyme and DNase I and lysed by sonication. The lysate was clarified by centrifugation at 40,000 g, then loaded onto a HisTrap HP column (Cytiva), washed with 40-50 column volumes of buffer B (buffer A plus 30 mM imidazole) and eluted with buffer C (buffer A plus 250 mM imidazole). Protein samples for crystallization experiments were then mixed with TEV protease for tag removal and dialyzed overnight at 4 °C

against buffer D (250 mM NaCl, 20 mM Tris-HCl pH 8.0, 5 mM  $\beta$ -mercaptoethanol). The samples were then loaded onto a HisTrap HP column, and the flow-through was collected. All protein samples were then concentrated by 30 kDa molecular weight cut-off Amicon centrifugation concentrators (EMD Millipore) and applied to a Superdex75 16/600 column (Cytiva) equilibrated in buffer E (100 mM NaCl, 20 mM Tris-HCl pH 8.0, 1 mM dithiothreitol). Fractions with the highest protein purity were pooled, concentrated, flash frozen with 10% v/v glycerol and stored at -80 °C until use.

#### **5.4.2. CphB<sub>DAP</sub> expression, purification and modification**

For SyCphB<sub>DAP</sub> expression, the codon for S132 in pCDF-PCC6803\_cphB was replaced with a TAG codon to produce plasmid pCDF-PCC6803DAP. BL21(DE3) cells were co-transformed with pCDF-PCC6803DAP and pSFDAPRS-PylT, which carries the orthogonal DAP tRNA synthetase system. A single colony was used to inoculate 100 ml of LB media supplemented with 100  $\mu$ g/ml kanamycin and 100  $\mu$ g/ml spectinomycin and grown overnight at 37 °C. The next day, 10 ml of starter culture were used to inoculate 1 L of 2YT media supplemented with 100  $\mu$ g/ml kanamycin, 100  $\mu$ g/ml spectinomycin and 50 mg photolabile-protected DAP<sup>123</sup> ((2S)-2-amino-3-(((2-((1-(6-nitrobenzo[d][1,3]dioxol-5-yl)ethyl)thio)ethoxy)carbonyl)amino)propanoic acid; Sapala Organics). Cultures were grown to OD<sub>600</sub> ~0.15 and protein expression was induced by the addition of 0.2 mM IPTG. The temperature was lowered to 18 °C and the cells were grown for another 48 hours prior to harvest. Purification was performed as with the wildtype enzyme, with the exception that following the first elution from the HisTrap column, the sample was irradiated with UV light (365 nm, 35 mW/cm<sup>2</sup>, 1 min) to remove the protecting group from the incorporated DAP. By LC-MS, we observed no evidence of CphB with residues other than DAP at position 132 in these preparations. Pure protein was concentrated to 5.5 mg/ml in buffer E and ( $\beta$ -Asp-Arg)<sub>3</sub> was added to a final concentration of 1 mM. The enzyme and substrate were incubated together for 48 hours at RT before LC-MS analysis and crystallization. Nearly complete modification of the enzyme with ( $\beta$ -Asp-Arg)<sub>2</sub> was verified by protein LC-MS using a Bruker amaZon speed ETD ion trap mass spectrometer (Supplementary Fig. 5.1a).

#### **5.4.3. Protein crystallization, data collection and structure solution**

Proteins were crystallized using the sitting drop method. Protein (2  $\mu$ l) in buffer E was mixed with 2  $\mu$ l of well solution and allowed to equilibrate against 500  $\mu$ l of well solution at 22 °C. For the SyCphB<sub>DAP</sub>-( $\beta$ -Asp-Arg)<sub>2</sub> complex, modified protein at 5 mg/ml was used with a well

solution containing 6% dioxane, 0.1 M MES pH 6.6, 1.65 M ammonium sulfate and 4% formamide. The crystals were cryo-protected by dipping in well solution supplemented with 20% glycerol and flash frozen in liquid nitrogen. Data were collected at the Advanced Photon Source (APS) beamline 24-ID-C and processed in XDS<sup>195</sup> and AIMLESS<sup>196</sup> in space group C222<sub>1</sub>. The structure was solved using 3ENO<sup>124</sup> as a search model and refined with Coot<sup>153</sup> and REFMAC5<sup>187</sup> implemented in CCP4i2<sup>184</sup>. The resulting maps are clearly high quality, though the R<sub>free</sub> and R values are higher than expected for this resolution, likely because of translational noncrystallographic symmetry or partial twinning. *PcCphB* crystals were obtained with protein concentrated to 20 mg/ml and well solution of 0.3 M magnesium formate. Crystals were cryo-protected with well solution supplemented with 30% 2-methyl-2,4-pentanediol (MPD) and flash frozen in liquid nitrogen. Data were collected at the Canadian Light Source (CLS) beamline CMCF-BM and processed with DIALS<sup>150</sup> and AIMLESS<sup>196</sup> in space group P3<sub>1</sub>21. The structure was solved using 3ENO<sup>124</sup> as a search model and refined with REFMAC5<sup>187</sup> implemented in CCP4i2<sup>184</sup>, Phenix<sup>188</sup>, Rosetta<sup>152</sup>, and Coot<sup>153</sup>. Figures were prepared using PyMOL (Schrödinger).

#### 5.4.4. Isothermal titration calorimetry (ITC)

ITC experiments were carried out using a MicroCal iTC200 (GE Healthcare) instrument at 25 °C. Both the enzyme and substrate were dissolved in buffer F (100 mM NaCl, 20 mM Tris pH 8.0, 5 mM β-mercaptoethanol). The cell contained 200 μM β-Asp-Arg or (β-Asp-Arg)<sub>3</sub> or 300 μM (β-Asp-Arg)<sub>2</sub>. The syringe contained 2 mM of SyCphB S132A or 1 mM wildtype *PcCphB*. Nineteen 2 μl injections were interspaced by 180 seconds each. Data were analyzed using Microcal Origin 7.0 (OriginLab, Northhampton, Massachusetts) with a binding model stoichiometry of 1:1 with the “protein in cell” option. All measurements were performed in triplicate.

#### 5.4.5. Cyanophycinase activity assay

Because cyanophycin is insoluble and scatters light at near-neutral pH, but β-Asp-Arg is soluble and does not scatter light, cyanophycinase activity was measured by monitoring the decrease in scattering of 600 nm light over time. Cyanophycin, purified from *E. coli* cells<sup>2,90</sup> expressing CphA1 using previously described methods<sup>3</sup>, was resuspended in 0.1 M HCl to a stock concentration of 25 mg/ml. Cyanophycinase assays contained 625 μg cyanophycin, 25 mM NaOH, 100 mM Tris-HCl pH 8.0, and 1.5 μM enzyme in a total reaction volume of 100 μl. Assays were performed at 25 °C. OD600 was monitored using a SpectraMax Paradigm spectrophotometer

(Molecular Devices), with 8 second linear shaking between reads, and the resulting data were analyzed using GraphPad Prism. To calculate activity rates, the minimum of the first derivative of each curve was taken. The derivative curves were smoothed with a 2nd order polynomial to reduce noise in the measurements. All measurements were performed in quadruplicate.

## **5.5. Acknowledgements**

We thank all the members of the Schmeing lab for important advice and ongoing discussions on this project, Kim Munro for help with ITC, Nancy Rogerson for proofreading and synchrotron staff S. Labiuk and K. Janzen (Canadian Light Source) and David Neau (Advanced Photon Source) for facilitating remote collection of diffraction datasets. This study includes work based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the National Institutes of Health (P30 GM124165). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. Part of the research described in this paper was performed using beamline CMCF-BM at the Canadian Light Source, a national research facility of the University of Saskatchewan, which is supported by the Canada Foundation for Innovation (CFI), the Natural Sciences and Engineering Research Council (NSERC), the National Research Council (NRC), the Canadian Institutes of Health Research (CIHR), the Government of Saskatchewan, and the University of Saskatchewan.

## **5.6. Data availability**

The protein structures solved in this study have been deposited to the PDB: SyCphB<sub>DAP</sub> (PDB 7UQW), PcCphB (PDB 7UQV).

## **5.7. Supplementary information**

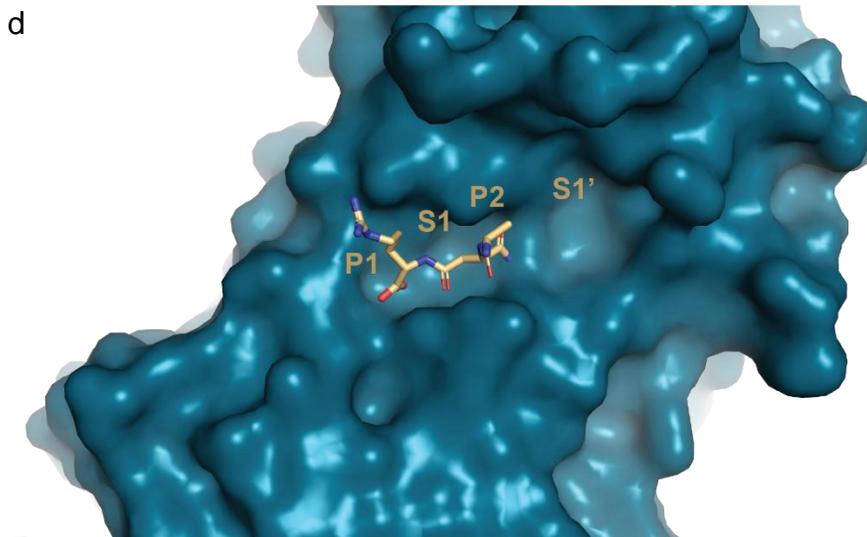
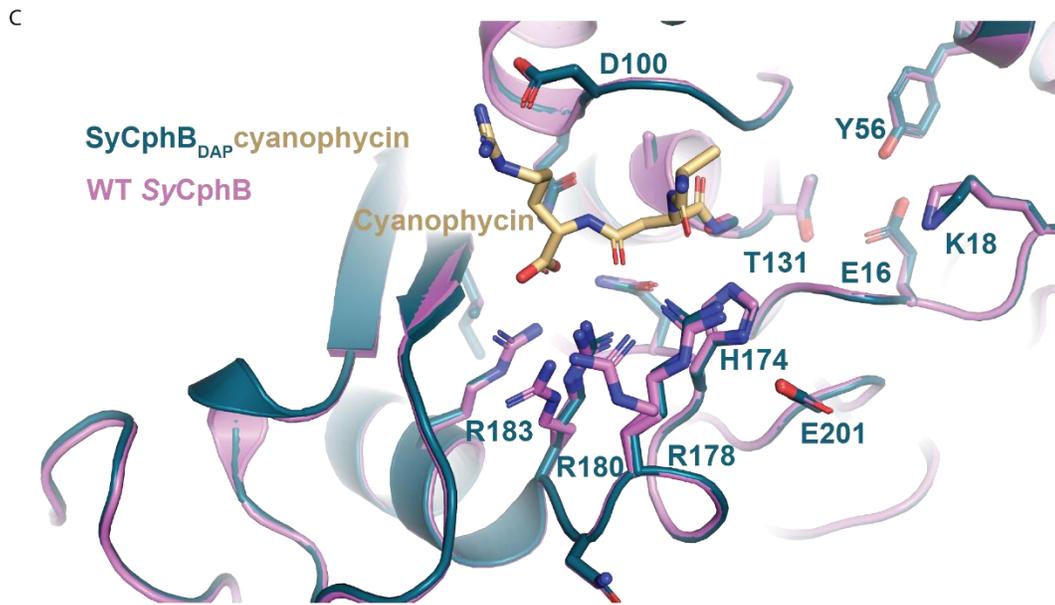
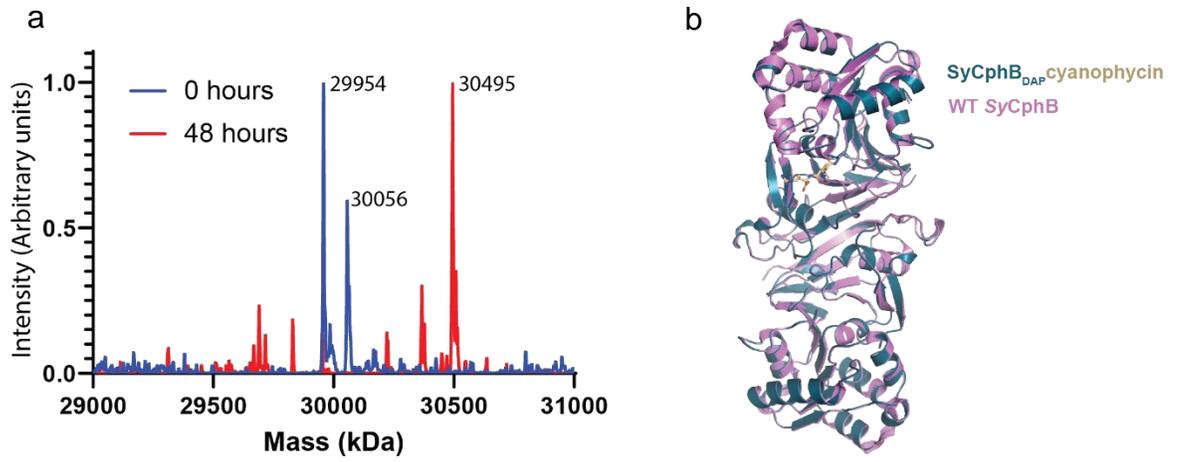
### **Supplementary Table 5.1. Crystallography statistics**

	<b>SyCphB<sub>DAP</sub> (PDB 7UQW)</b>	<b>PcCphB (PDB 7UQV)</b>
<b>Data collection</b>		
Space group	C222 <sub>1</sub>	P3 <sub>1</sub> 21
Cell dimensions (a, b, c; Å)	76.4, 132.9, 164.1	172.8, 172.8, 116.9

Cell angles ( $\alpha$ , $\beta$ , $\gamma$ ; °)	90, 90, 90	90, 90, 120
Resolution (Å)	61.6 - 1.5 (1.53 - 1.50)	149.65 - 2.4 (2.44 - 2.40)
R <sub>merge</sub>	0.1382 (1.573)	0.01288 (0.1396)
R <sub>pim</sub>	0.03978 (0.4458)	0.01288 (0.1396)
I/ $\sigma$ I	15.18 (3.10)	14.96 (0.27)
CC <sub>1/2</sub>	0.997 (0.866)	1 (0.961)
Completeness (%)	99.92 (99.98)	97.81 (78.99)
Redundancy	12.9 (13.0)	18.1 (17.9)

### Refinement

Resolution (Å)	61.6 - 1.5 (1.53 - 1.50)	86.4 - 2.4 (2.44 - 2.40)
No. reflections	133142 (13155)	76981 (6145)
R <sub>work</sub> / R <sub>free</sub>	0.1819 / 0.1999	0.2001 / 0.2340
No. atoms	6659	7864
Protein	6023	7864
Ligand/ion	95	0
Solvent	541	0
B factors		
Protein	24.44	67.1
Ligands	32.83	-
Ramachandran favored	97.45 %	97.55 %
Ramachandran outliers	0 %	0 %
Rotamer outliers	0.16%	0 %
Clash score	2.96	3.3
Mol probity score	1.20	1.21
R.M.S. deviations		
Bond lengths (Å)	0.015	0.007
Bond angles (°)	1.79	1.13



**Supplementary Figure 5.1.** Modification of SyCphB<sub>DAP</sub> with cyanophycin. (a) MS spectra of SyCphB<sub>DAP</sub> incubated with (β-Asp-Arg)<sub>3</sub> after 0 and 48 hours show the disappearance of the unliganded, de-protected SyCphB<sub>DAP</sub> (expected mass 29951 Da) and SyCphB<sub>DAP</sub> liganded with the transient thiocarbamate DAP protection intermediate (expected mass 30056 Da), and the formation of a peak with a mass corresponding to SyCphB<sub>DAP</sub> with (β-Asp-Arg)<sub>2</sub> covalently attached to it (expected mass 30495 Da). (b) Structural alignment of SyCphB (PDB 3EN0; violet) and SyCphB<sub>DAP</sub> (orange) shows very similar overall structures. (c) Structural alignment of the active sites of SyCphB (PDB 3EN0; violet) and SyCphB<sub>DAP</sub> (orange) shows very similar conformations in liganded and unliganded form. (d) Surface representation of SyCphB<sub>DAP</sub>-cyanophycin showing the two shallow pockets (S1 and S1') used for binding of the P1 and P1' dipeptides of cyanophycin substrates.

```

PCC6803      MPLSSQPAILIIGGAEEDKVHGREILQTFWRSRSGNDAIIGIIPSASREPLLIGERYQTIF 60
DSM2933      MEEKSKGNLVIIGGAEEDKKGESKILKKVAEIAGFGDMEFIVLTTATEHPVEVGNEYLNVF 60
*   .*:   ::*****   :*:... . :* .* : :: :*...*: :*...* ..*

PCC6803      SDMGVKELKVLDIRDRAQGDSGYRLFVEQCTGIFMTGGDQLRLCGLLADTPLMDRIRQR 120
DSM2933      QRLGINNIEVLDISTREDANNEENYYKIVNSGGVFMTGGDQLRITSILGGTKVFNALIEA 120
.   :*:...:**** * :... . : . * :*****: ..*..* : : :

PCC6803      VHNGEISLAGTSSAGAAVMGHHMIAGGSSGEWPNRALVDMAVGLGIVPEIVVDQHFHNRNR 180
DSM2933      YLKG-VVIAGTSSAGASVMSNTMIVDGNSNDPARKCTLKMASGLGLLEEAIIDQHFDQRGR 179
   :* : :*****:*. : **..*.*. : . . .** ***: * :*****:*. *

PCC6803      MARLLSAISTHPELLGLGIDEDTCAMFERDGSVKVIGQGTVSFVDARDMSYTNAALVGAN 240
DSM2933      FGRLLCGVAENPHMLGIGIDEDTAIRVYPDAHFEVVGSYAVTIIDGKSIVSSNVSELKPD 239
: .***. . . :*.:**:*..... . * . .**.* :*...*... :*.. :

PCC6803      APLSLHNLRLNLVHGEVYHQVKQRAFPRV- 270
DSM2933      EILAIANVTVHVLPEGYGFDMK-RREVLRLH 269
*:: * : ::* .* : . :* . *

```

**Supplementary Figure 5.2.** Sequence alignment of PcCphB and SyCphB showing that virtually all the cyanophycin-binding residues of SyCphB are also present in PcCphB. Sixteen residues which form the active site and surroundings are highlighted in yellow.

## Bridge to chapter 6

Chapter 5 focused on the first step of cyanophycin biodegradation. The second step of cyanophycin catabolism – degradation of  $\beta$ -Asp-Arg dipeptides into free Asp and Arg – has mostly been regarded as fairly trivial. Two well-known bacterial enzymes are capable of degrading a wide range of  $\beta$ -aspartyl dipeptides: isoaspartyl aminopeptidase (IaaA) and isoaspartyl dipeptidase (IadA). These enzymes are involved in other cellular processes and so their activity and structures have been extensively studied outside of a cyanophycin context. It is widely assumed that these enzymes are also responsible for the final step of cyanophycin degradation. However, upon examination of the existing literature I found that only IaaA in cyanobacteria has been directly shown to be involved in cyanophycin metabolism. In Chapter 6, I describe experiments designed to validate or disprove the common notion about  $\beta$ -Asp-Arg degradation.

6. Bioinformatics of cyanophycin metabolism genes and characterization of promiscuous isoaspartyl dipeptidases that catalyze the final step of cyanophycin degradation

Also described in: Sharon I, Schmeing TM. *Submitted to PLOS ONE*.

## **6.1. Abstract**

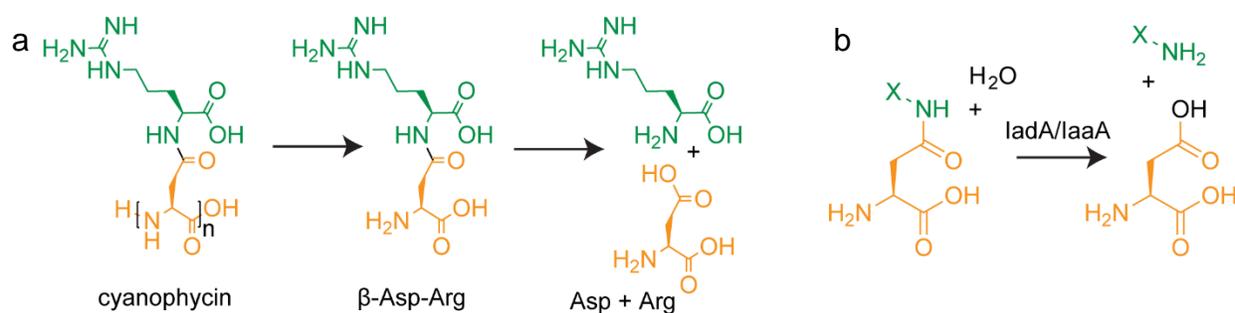
Cyanophycin is a bacterial biopolymer used for storage of fixed nitrogen. It is composed of a backbone of L-aspartate residues with L-arginines attached to each of their side chains. Cyanophycin is produced by cyanophycin synthetase 1 (CphA1) using Arg, Asp and ATP, and is degraded in two steps. First, cyanophycinase breaks down the backbone peptide bonds, releasing  $\beta$ -Asp-Arg dipeptides. Then, these dipeptides are broken down into free Asp and Arg by enzymes with isoaspartyl dipeptidase activity. Two bacterial enzymes are known to possess promiscuous isoaspartyl dipeptidase activity: isoaspartyl dipeptidase (IadA) and isoaspartyl aminopeptidase (IaaA). We performed a bioinformatic analysis to investigate whether genes for cyanophycin metabolism enzymes cluster together or are spread around the microbial genomes. Many genomes showed incomplete contingents of known cyanophycin metabolizing genes. Cyanophycin synthetase and cyanophycinase are usually clustered together when recognizable genes for each are found within a genome. Cyanophycinase and isoaspartyl dipeptidase genes typically cluster within genomes lacking *cphA1*. About one-third of genomes with genes for CphA1, cyanophycinase and IaaA show these genes clustered together, while the proportion is around one-sixth for CphA1, cyanophycinase and IadA. We used X-ray crystallography and biochemical studies to characterize an IadA and an IaaA from two such clusters. The enzymes retained their promiscuous nature, showing that being associated with cyanophycin-related genes did not make them specific for  $\beta$ -Asp-Arg dipeptides derived from cyanophycin degradation.

## **6.2. Introduction**

Cyanophycin is a biopolymer first described over 100 years ago as large, light scattering granules observed in cyanobacterial cells<sup>14</sup>. These granules are composed of chains with backbones of L-aspartate residues with L-arginine attached to each Asp side chain<sup>15</sup> (Fig. 6.1a). Cyanophycin contains 26% nitrogen content by mass, which, along with its inert nature and low solubility, makes it useful for nitrogen, carbon and energy storage<sup>130,133,134</sup>. Cyanophycin can be produced by a wide variety of bacteria<sup>2,100</sup>, but research in a biological context has mostly focused on cyanobacteria<sup>31,40,85,99,119</sup>. Cyanophycin is known to be especially useful for nitrogen-fixing cyanobacteria, which separate anaerobic nitrogen fixing from oxygen-producing photosynthesis either spatially in different cell types<sup>40</sup> or temporally in a day/night cycle<sup>42</sup>.

Cyanophycin is made by cyanophycin synthetase 1 (CphA1)<sup>98</sup> or 2 (CphA2)<sup>12</sup> (Supplementary Fig. 6.1). CphA1 is a widespread enzyme that catalyzes two ATP-dependent reactions<sup>98,197</sup>: it first adds Asp to the polymer backbone and then attaches Arg to the side chain of that Asp residue through an isopeptide bond<sup>2</sup>. Some CphA1 enzymes can also incorporate lysine into cyanophycin in place of arginine, though at lower efficiency<sup>96</sup>. CphA2, a cyanobacterial enzyme related to CphA1, uses a single active site to catalyze the ATP-dependent repolymerization of  $\beta$ -Asp-Arg dipeptides into cyanophycin<sup>12,13</sup>.

To access the nitrogen, carbon and energy stored in cyanophycin<sup>40,42</sup>, bacteria degrade it into free amino acids. This is done in two steps (Fig. 6.1a, Supplementary Fig. 6.1): First, cyanophycin is hydrolyzed into  $\beta$ -Asp-Arg dipeptides by a specialized exo-cyanophycinase enzyme, either the intracellular CphB<sup>121</sup> or CphI<sup>100</sup>, or the extracellular CphE<sup>121</sup>. Then the  $\beta$ -Asp-Arg dipeptides are hydrolyzed into Asp and Arg by enzymes that possess isoaspartyl-dipeptidase activity<sup>129</sup> (Fig. 6.1b). The two degradation steps occur within the same cells in cyanobacterial species that have day/night regulation of cyanophycin metabolism<sup>102</sup>, while in cyanobacterial communities with cyanophycin-synthesizing heterocysts, dipeptides are shuttled to vegetative cells for hydrolysis<sup>40</sup>. Many bacterial communities capable of using exogenous cyanophycin as a carbon and nitrogen source have been identified<sup>59,61</sup>. These communities can be found in a variety of environments, such as animal gut flora<sup>60</sup>, soil<sup>57</sup> and fresh-water sediments<sup>58</sup>, suggesting



**Figure 6.1. The structure and degradation of cyanophycin.** (a) Long polymer chains (typically  $n=80-400$ ) are degraded by cyanophycinase into  $\beta$ -Asp-Arg dipeptides, which are then hydrolyzed by isoaspartyl dipeptidases, resulting in free Asp and Arg. (b) The general reaction is catalyzed by isoaspartyl dipeptidases. X=any amino acid.

cyanophycin is commonly found in these environments. There is evidence that the two steps of cyanophycin degradation are sometimes split between members of a bacterial consortium, where some members express cyanophycinase and others degrade the  $\beta$ -aspartyl dipeptides<sup>59</sup>.

Enzymes capable of degrading  $\beta$ -aspartyl dipeptides are very common, because  $\beta$ -aspartyl residues can form spontaneously from intramolecular rearrangement of Asp and Asn residues in proteins<sup>198</sup>. The resulting  $\beta$ -aspartyl dipeptides, if not degraded, can accumulate to pathological levels in cells<sup>127</sup>. In bacteria, these  $\beta$ -aspartyl residues can either be repaired by L-isoaspartyl O-methyltransferase enzymes (E.C 2.1.1.77)<sup>126</sup> or be hydrolyzed into their amino acid constituents<sup>125</sup>. Two bacterial enzymes are known to have significant  $\beta$ -aspartyl dipeptidase activity: isoaspartyl dipeptidase (IadA)<sup>4,128</sup>, a bacterial zinc metallopeptidase; and isoaspartyl aminopeptidase (IaaA, also called plant-type asparaginase, EcAIII and IadC)<sup>129,199-201</sup>, a common Ntn-family enzyme with known plant and animal homologs. IadA and IaaA are evolutionarily unrelated and have different catalytic mechanisms, but both have broad substrate specificity because damage to proteins can lead to the attachment of different amino acids to Asp/Asn side chains<sup>4,129,202</sup>. Accordingly, they are also capable of degrading  $\beta$ -Asp-Arg/Lys, so it is assumed that  $\beta$ -Asp-Arg/Lys dipeptides derived from cyanophycin are degraded by general isoaspartyl dipeptidases<sup>100,121,124,129</sup>. In addition, several other enzymes, such as glycosylasparaginases, catalyze similar reactions and can display low levels of  $\beta$ -aspartyl dipeptidase activity<sup>203</sup>.

In this study, we analyzed the genomes in the NCBI RefSeq database<sup>204</sup> to investigate the tendency of cyanophycin metabolism genes to co-occur and cluster together in the genome. We observe moderate levels of co-occurrence of *cphA1*, cyanophycinase and an isoaspartyl

dipeptidase genes within these genomes. The rates of clustering of various combinations of the genes are well above random, ranging from moderate (e.g., 37 of 231 genomes containing *cphA1*, a cyanophycinase gene and *iadA* show all three genes to cluster) to high (e.g., 30 of 32 genomes with a cyanophycinase gene and *iaaA*, but without *cphA1* genes show clustering). Characterization of the activity and structures of representative IadA and IaaA enzymes which cluster with cyanophycin synthetase and cyanophycinase genes revealed that they have not become specific for  $\beta$ -Asp-Arg dipeptides.

### 6.3. Results

#### 6.3.1. Identification of cyanophycin-metabolizing gene clusters

To quantify the occurrence and clustering tendency of cyanophycin-metabolizing genes, we analyzed the presence and genomic localization of *cphA1*, cyanophycinase (*cphB*, *cphI* or *cphE*) and isoaspartyl dipeptidase (*iaaA*<sup>129</sup> or *iadA*<sup>128</sup>) in all 27,349 non-redundant, complete bacterial genomes in the NCBI RefSeq database<sup>204</sup>. Isoaspartyl dipeptidases are common (found in 11,814 genomes, 43.2%), which is expected, as they have roles other than cyanophycin metabolism. Cyanophycin synthetase 1 is found in 1,614 genomes (6%, Table 6.1), and a recognizable cyanophycinase gene is present in 739 genomes (3%, Table 6.2).

	<b>CphA1 + cyanophycinase and/or IaaA and/or IadA</b>	<b>CphA1 + cyano- phycinase</b>	<b>CphA1 + IaaA</b>	<b>CphA1 + IadA</b>	<b>CphA1 + IaaA and/or IadA</b>	<b>CphA1 + IaaA + cyano- phycinase</b>	<b>CphA1 + IadA + cyano- phycinase</b>	<b>CphA1 + cyanophycin- ase + IaaA and/or IadA</b>
In genome	1614	658	968	232	1181	153	231	366
Clustered	538	535	51	38	88	49	37	85

**Table 6.1. Analysis of genomes which encode CphA1.**

	<b>cyanophycinase + CphA1 and/or IaaA and/or IadA</b>	<b>cyanophycinase + IaaA</b>	<b>cyanophycinase + IadA</b>	<b>cyanophycinase + IaaA and/or IadA</b>
In genome	739	185	251	418
Clustered	578	79	52	130

**Table 6.2. Analysis of genomes which encode a cyanophycinase.**

Next, we examined the tendency of cyanophycin-metabolizing genes to co-occur and cluster together. We defined co-occurrence as at two genes present in the same genome, and clustering as genes separated by not more than a 5 kilobase pair (kbp) intergenic region. Of the genomes that have *cphA1*, 658 (41%) also have a recognizable cyanophycinase. These genes are clustered in most (535; 82%) of the genomes that have both. Genes for IaaA or IadA are found in 1181 (73%) *cphA1*-containing genomes, with 968 (60%) of those genomes having *iaaA* and 232 (14%) having *iadA*. However, in contrast to cyanophycinase genes, isoaspartyl dipeptidase genes generally do not cluster with *cphA1*, being proximal in only 88 (7.5%) of genomes that have both (Table 6.1).

Interestingly, clustering of *cphA1* and isoaspartyl dipeptidase is more common in genomes that have genes encoding all three steps of cyanophycin metabolism. There are 366 such genomes in the RefSeq database. In genomes that have *cphA1*, a cyanophycin gene, and *iaaA*, 49 of 153 show clustering. In the case of *iadA*, 37 of 231 genomes with *cphA1*, a cyanophycin gene and *iadA* show these three clustered.

Ben Hania et al. have described the utility and occurrence of a “cyanophycin utilization locus” which includes cyanophycinase genes, *iadA* and a transporter so a microbe can scavenge cyanophycin from the environment<sup>62</sup>. This observation also holds for *iaaA*: Searches of the NCBI RefSeq database returned 52 genomes that contain a cyanophycinase gene and *iaaA* or *iadA* but not *cphA1*, and 45 of them had cyanophycinase and isoaspartyl-dipeptidase genes clustered (Table 6.3).

	<b>cyanophycinase + IaaA (no cphA)</b>	<b>cyanophycinase + IadA (no cphA)</b>	<b>cyanophycinase + IaaA and/or IadA (no cphA)</b>
In genome	32	20	52
Clustered	30	15	45

**Table 6.3. Analysis of genomes which encode a cyanophycinase and isoaspartyl dipeptidase but not CphA1.**

The rate of each of the above clustering is above random chance: As a control, we detected 955 genomes with *cphA1* and dihydrofolate reductase (*folA*), a common housekeeping gene

unrelated to cyanophycin metabolism. None of these genomes had the two genes clustered together.

### 6.3.2. IadA and IaaA from cyanophycin clusters are not specific for $\beta$ -Asp-Arg/Lys

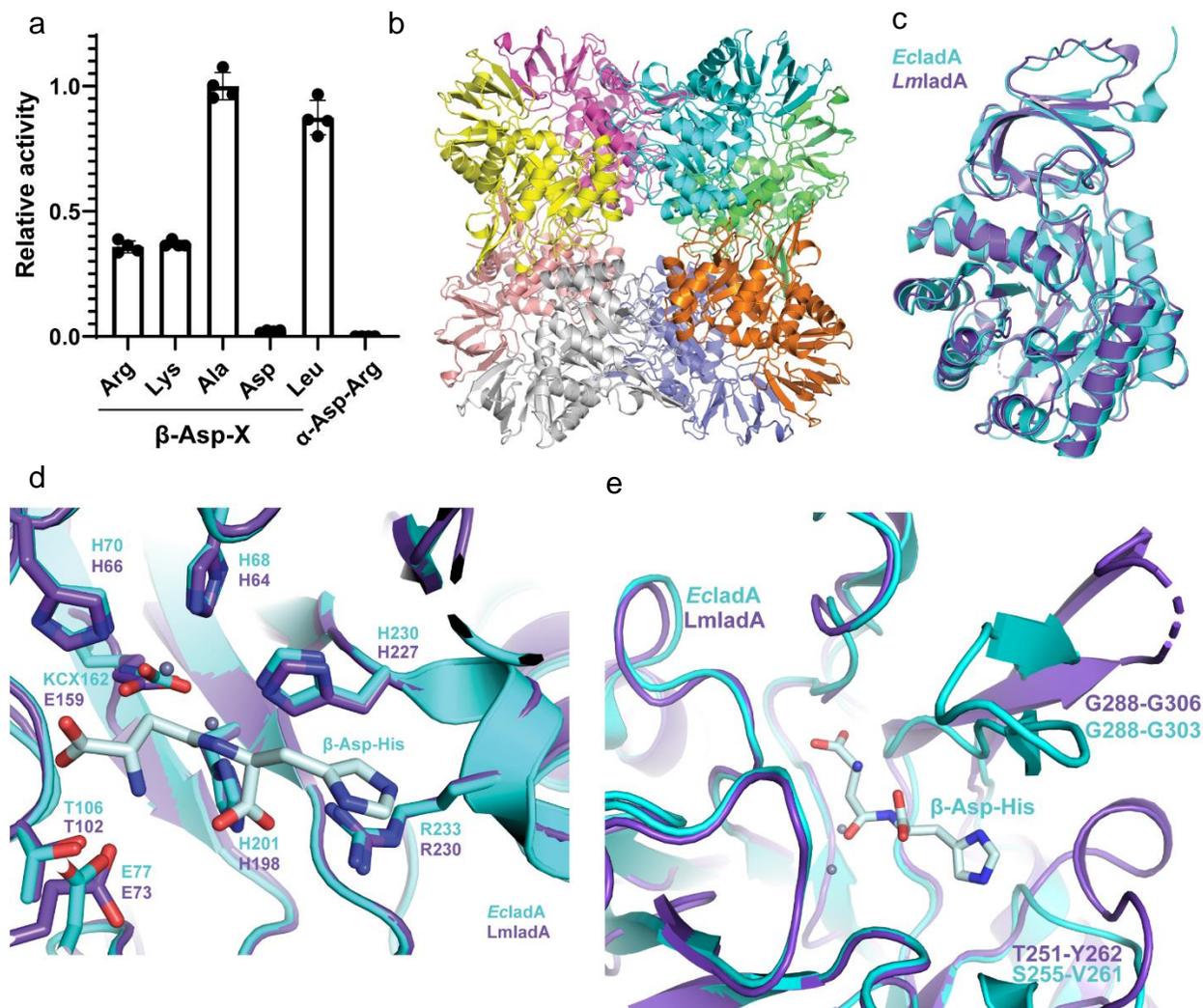
Previous studies which characterized the activity of canonical isoaspartyl dipeptidases found that both IadA<sup>4</sup> and IaaA<sup>129</sup> accept a wide range of  $\beta$ -aspartyl dipeptides as substrates. Subsequent structural results explained this lack of substrate specificity: while both enzymes make extensive interactions with the Asp portion of the substrate, the portion of the isoaspartyl dipeptidase surrounding the amino acid attached to the Asp side chain is large and able to accommodate the substrate rather than bind it specifically<sup>4,10</sup>.

We wondered whether the IaaA or IadA homologs present in cyanophycin metabolism clusters have evolved to specialize in cyanophycin degradation and display substrate preference for  $\beta$ -Asp-Arg (and  $\beta$ -Asp-Lys) over other  $\beta$ -aspartyl dipeptides. We therefore performed biochemical and structural characterization of a representative of IaaA and of IadA  $\beta$ -aspartyl dipeptidases whose genes are clustered with both *cphA1* and *cphB*: IadA from *Leucothrix mucor* DSM2157 (*LmIadA*) and IaaA from *Roseivivax halodurans* DSM15395 (*RhIaaA*).

*LmIadA* has 44% sequence identity to *E. coli* IadA (*EcIadA*<sup>4</sup>). Like *EcIadA*, the purified enzyme forms octamers in solution (Supplementary Fig. 6.2)<sup>4</sup>. We examined the activity of *LmIadA* towards several  $\beta$ -aspartyl dipeptides and found that it displayed no apparent preference towards  $\beta$ -Asp-Arg/Lys (Fig. 6.2a). To confirm the structural basis for this lack of specificity, we solved the structure of the wild type enzyme at 1.8 Å resolution and compared it to that of *EcIadA*<sup>4</sup> (Supplementary Table 6.1).

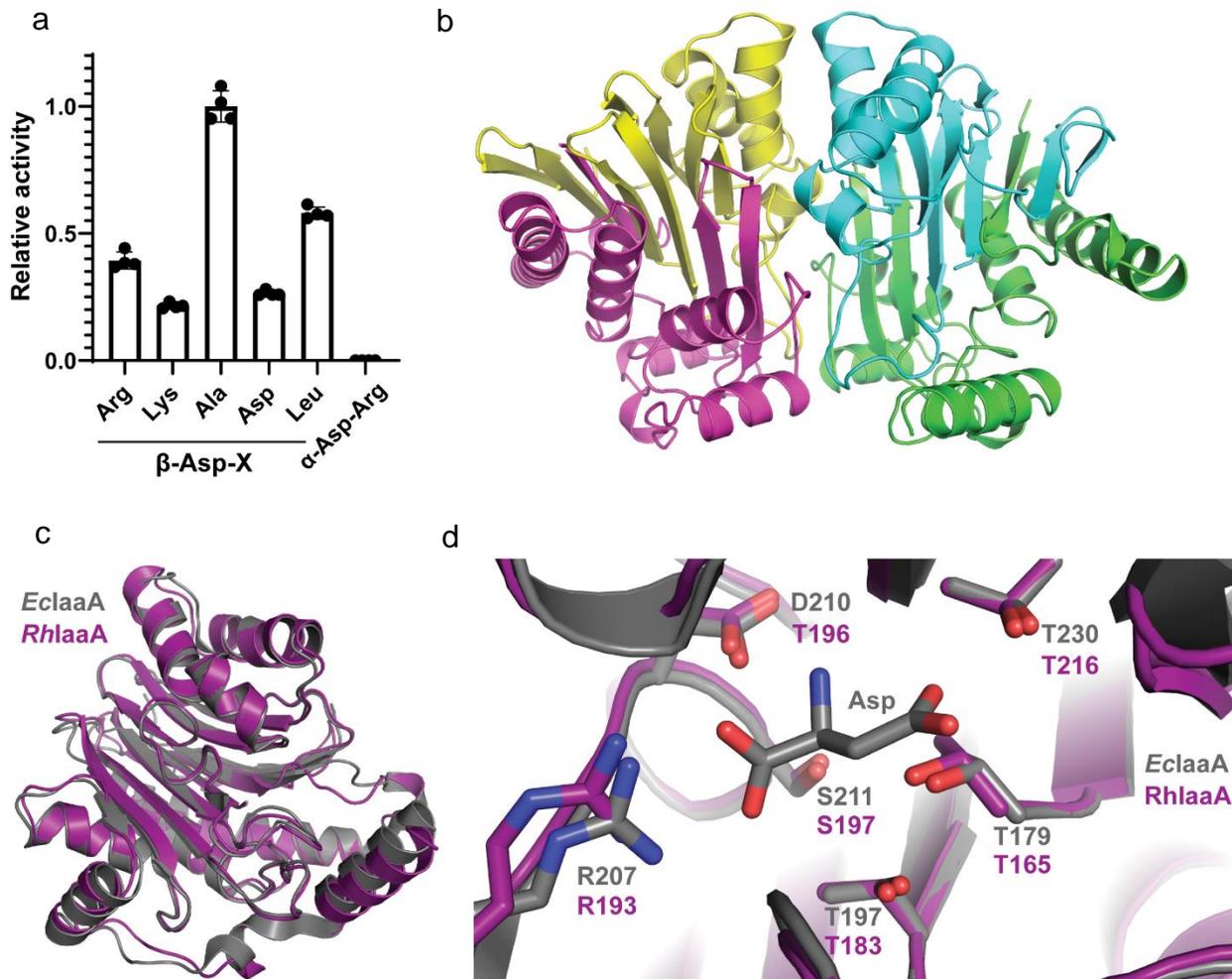
The crystal structure of *LmIadA* shows a homooctameric architecture as the asymmetric unit (Fig. 2b). It displays high similarity to that of *EcIadA*<sup>4</sup> (0.81 Å RMSD across 315 C $\alpha$  pairs, PDB code 1YBQ; Fig. 6.2c), with the active site residues almost identical in both sequence and structure (Fig. 6.2d). Two Zn<sup>2+</sup> ions are liganded by H64, H66, H198, H227 and E159, corresponding to *EcIadA* H68, H70, H201, H230 and carboxylated K162. Substrate binding residues in *EcIadA* such as E77, T106 and R233<sup>4</sup>, are also present at corresponding positions in *LmIadA* (E73, T102 and R230) and display similar conformations (Fig. 6.2d).

The published structure of *EcIadA* in complex with  $\beta$ -Asp-His<sup>4</sup> shows that the His side chain of the substrate forms minimal interactions with the enzyme. It faces an opening in the active



**Figure 6.2. Structure and activity of *LmIadA*.** (a) Asp release assay of *LmIadA* and different Asp-containing dipeptides. The enzyme is specific towards  $\beta$ -aspartyl dipeptides, but displays no specificity towards Arg or Lys as the  $\beta$ -linked amino acid. Error bars represent the standard deviation of the mean of  $n=4$  replicates. (b) The homooctameric crystal structure of *LmIadA*. (c) Overlay of *LmIadA* (purple) and *EcIadA*<sup>4</sup> (cyan, PDB code 1YBQ) monomers showing their high overall structural similarity. (d) Close-up view of the active sites of *LmIadA* and *EcIadA* in complex with the substrate  $\beta$ -Asp-His, showing they are similar in both sequence and structure. (e) Overlay of the regions around the active sites of *LmIadA* and *EcIadA*, showing both have large openings capable of accommodating a variety of  $\beta$ -aspartyl dipeptides as substrate.

site which, as expected, can accommodate a variety of substrates. *LmIadA* displays a somewhat different architecture in this region (Fig. 6.2e). The loop formed by *LmIadA* T251-Y262 is longer



**Figure 6.3. Structure and activity of *RhIaaA*.** (a) Asp release assay of *RhIaaA* with different Asp-containing dipeptides. The enzyme is specific towards  $\beta$ -aspartyl dipeptides, but displays no specificity towards Arg or Lys as the  $\beta$ -linked amino acid. Error bars represent the standard deviation of the mean of n=4 replicates. (b) The heterotetrameric crystal structure of *RhIaaA*. (c) Overlay of *RhIaaA* (purple) and *EcIaaA*<sup>10</sup> (gray, PDB code 2ZAL) heterodimers showing their high overall structural similarity. (d) Close view of the active sites of *RhIaaA* and *EcIaaA* in complex with the product Asp, showing they are similar in both sequence and structure.

and bulkier than the corresponding one of *EcIadA* (S255-V261), and as a result could restrict access to the active site. However, the partially flexible region between G288-G306 (*EcIadA* G288-G303) is oriented away from the binding pocket. This leads to a similarly sized opening in

the active site region surrounding the non-Asp portion of the substrate and explains the lack of specificity (Fig. 6.2e).

We performed analogous analyses with the IaaA enzyme encoded in the cyanophycin gene cluster of *Roseivivax halodurans*. *RhIaaA* has 51% sequence identity with *E. coli* IaaA (*EcIaaA*<sup>10</sup>). Like *EcIaaA* and other Ntn-family enzymes, the pro-enzyme is expressed as a single chain that undergoes autocatalytic cleavage into two subunits, a and b, which constitute the mature a<sub>2</sub>b<sub>2</sub> heterotetrameric enzyme (Supplementary Fig. 6.2). We assayed the activity of *RhIaaA* towards the same set of  $\beta$ -Asp dipeptides used to assess *LmIadA* and found that it could hydrolyze all of them with no apparent preference towards  $\beta$ -Asp-Arg/Lys (Fig. 6.3a). To confirm the structural basis for the lack of substrate specificity, we solved the structure of the wildtype enzyme at 2.7 Å resolution and compared it to that of *EcIaaA* (Supplementary Table 6.1).

The crystal structure of *RhIaaA* shows the expected heterotetrameric architecture (Fig. 6.3b). The enzyme displays high structural similarity to *EcIaaA* (0.58 Å RMSD across 230 C $\alpha$  pairs, PDB code 2ZAL<sup>10</sup>; Fig. 6.3c), with the active site residues being almost identical in both sequence and conformation. In *EcIaaA*, the Asp portion of the substrate is bound by T197, R207, D210, S211, T230 and G231, as well as the catalytic T179<sup>10</sup>. These residues are all present and in the same conformations in *RhIaaA* (T183, R193, D196, S197, T216 and G217, and the catalytic T165, Fig. 6.3d). As is the case with *IadA*, the substrate likely binds oriented in a way that positions the non-Asp portion of it facing a large opening in the active site (Fig. 6.3d). This presumably results in minimal interaction between IaaA and the substrate residue bound to Asp by the scissile isopeptide bond, which would enable the active site to accommodate a wide range of  $\beta$ -aspartyl dipeptides.

#### **6.4. Discussion**

Bacteria often use clustering to control expression of genes with related functions<sup>205</sup>. In the case of cyanophycin metabolism, clustering appears to be common for *cphA1* and cyanophycinase<sup>119</sup> (Table 6.1). Previous studies in cyanobacteria show that these two genes can also share some transcription regulation elements<sup>119</sup>. Clustering of genes for cyanophycinase and an isoaspartyl dipeptidase is very common in genomes that have those genes but not *cphA1* (Table 6.2). These are often accompanied by amino acid transporters and probably represent

cyanophycin-scavenging clusters, such as the ones described in the cyanobacteria-scavenger strain L21-Spi-D4<sup>62</sup> and in *Flammeovirga pacifica* strain WPAGA1<sup>206</sup>.

The clustering rate of isoaspartyl dipeptidases with *cphA1* and cyanophycinase in genomes that have all three is well above random distribution, but not as high as that of *cphA1*-cyanophycinase alone. There are several possible explanations why clustering is not strict. First, it is possible for these genes to be under control of the same transcription regulators even if they are not clustered. Second, since isoaspartyl dipeptidases are required outside of a cyanophycin context, there may be evolutionary pressure to keep those genes separate for regulatory purposes. Third, in some cases it is beneficial to have cyanophycin-metabolizing genes regulated independently of one another. An example for this can be seen in the heterocyst-forming cyanobacterium *Anabaena* sp. PCC7120. Heterocysts of this bacterium express cyanophycinase to degrade cyanophycin into dipeptides, which are shuttled to vegetative cells. These, in turn, express high levels of IaaA to convert the dipeptides into free amino acids<sup>40</sup>.

In general, the co-occurrence rates of genes involved in cyanophycin metabolism is lower than we expected. For example, detection of a recognizable cyanophycinase in only 41% of *cphA1*-containing genomes is unanticipated. Cyanophycin is only known to serve as a storage material, and the bacteria that store it must also possess the wherewithal to degrade it. It is possible that bacteria which possess *cphA1* but not *cphB/E/I* possess other, unknown cyanophycinase isozymes. The lack of an identifiable isoaspartyl dipeptidase gene in 27% of *cphA1*-containing genomes suggests that not all genes encoding enzymes with this dipeptidase activity were detected in our searches. Similarly, Füser *et al.* performed an analysis of 48 *cphA1* or cyanophycinase-containing genomes in 2007<sup>100</sup> and found that only 26 also had *iaaA* or *iada*. Isoaspartyl dipeptidase activity in these bacteria could be provided by distant homologues of *iaaA* or *iada* or by unrelated isozymes. Manual examination of genomes from the RefSeq database that have a CphA1-cyanophycinase cluster shows some of them to include adjacent genes which could potentially have isoaspartyl dipeptidase activity, such as those annotated as "S9 family peptidase" (in genome NZ\_CP029187.1), annotated as "M14 family metallopeptidase" or "succinylglutamate desuccinylase/aspartoacylase family protein" (in genome NZ\_VYQF01000002.1) and a gene weakly homologous (25-30% identity) to cocaine esterase<sup>207</sup> (in genome NZ\_SJEY01000003). The existence of cryptic isoaspartyl dipeptidase enzymes has been proposed before, for example in *Saccharomyces cerevisiae*<sup>208</sup>.

Both of the isoaspartyl dipeptidases from cyanophycin gene clusters that we cloned, expressed, purified and assayed display no substrate specificity towards  $\beta$ -Asp-Arg/Lys and accept a range of isoaspartyl dipeptides. The crystal structures of both enzymes were consistent with this promiscuity and show that the structural basis for this lack of specificity is shared with other IaaA and IadA enzymes. These results suggest that even when their genes are clustered with cyanophycin-related genes, IaaA and IadA function in both cyanophycin metabolism and the protein-degradation pathway, in line with the widely held belief that general isoaspartyl dipeptidases are responsible for the last step of cyanophycin degradation<sup>100,129</sup>.

## **6.5. Materials and methods**

### **6.5.1. Bioinformatics**

For the identification of gene clusters, we created a local database with all complete bacterial genomes in the NCBI (USA) Refseq<sup>204</sup> database (May 2022). We used cblaster<sup>209</sup> to search this database using several queries for *cphA1* (WP\_028947105.1, WP\_004925893.1, WP\_015942562.1), cyanophycinase (WP\_011058003.1, WP\_004925892.1, Q8KQN8.1), *iada* (WP\_188415469.1, WP\_138978951.1, WP\_037265155.1) and *iaaA* (MBS3792760.1, WP\_034545427.1, WP\_022952024.1). For the identification of putative isoaspartyl dipeptidases in *cphA1*-cyanophycinase clusters, MultiGeneBlast<sup>210</sup> was used to search *cphA1*-containing genomes for *cphA1*-cyanophycinase clusters, and the results were analyzed manually for putative isoaspartyl dipeptidases.

### **6.5.2. Cloning, protein expression and purification**

The genes encoding *LmIadA* (WP\_022952024.1) and *RhIaaA* (WP\_037265155.1) were amplified from genomic DNA (DSMZ, Leibniz Institute, Germany). Both genes were cloned into a pJ411-derived plasmid encoding a C-terminal tobacco etch virus (TEV) protease recognition site and an octahistidine affinity tag. All cloning and mutagenesis were performed by transforming DH5- $\alpha$  *E. coli* cells with PCR fragments containing overlapping ends. Proteins were expressed in *E. coli* BL21(DE3) cells grown in TB media supplemented with 150  $\mu$ g/ml kanamycin. Cultures were grown at 37 °C until they reached an OD<sub>600</sub> of ~1. The growth temperature was then lowered to 18 °C and protein expression was induced with 0.2 mM isopropyl  $\beta$ -d-1-thiogalactopyranoside (IPTG) for ~20 hours. All subsequent protein purification steps were carried out at 4 °C. Following harvest by centrifugation, the cells were resuspended in buffer A (250 mM NaCl, 50 mM Tris pH

8.0, 10 mM imidazole, 2 mM  $\beta$ -mercaptoethanol) supplemented with a few crystals of lysozyme and DNase I, and lysed by sonication. The lysate was clarified by centrifugation at 40,000 g for 30 minutes and then applied onto a HisTrap HP column (Cytiva, USA). The column was washed extensively with buffer B (buffer A with 30 mM imidazole) and the protein was eluted with buffer C (buffer A with 250 mM imidazole). For structural studies, the protein was incubated with TEV protease for removal of the octahistidine tag while being dialyzed overnight against buffer D (250 mM NaCl, 20 mM Tris pH 8.0, 5 mM  $\beta$ -mercaptoethanol) prior to application to a HisTrap column and collection of the flow through. All protein preparations were then concentrated and applied to a Superdex200 16/60 column (Cytiva, USA) equilibrated in buffer E (100 mM NaCl, 20 mM Tris pH 8.0, 1 mM dithiothreitol). Fractions with the highest protein purity were concentrated, supplemented with glycerol to a final volume of 15% and flash frozen in liquid nitrogen for storage.

### **6.5.3. Protein crystallization, data collection, structure solution and refinement**

For crystallization trials, all proteins were buffer exchanged into buffer E and subjected to small-scale wide screen crystallization trials in 96-well plates using the sitting drop method. Optimization of crystallization conditions was performed using the sitting drop method by mixing 2  $\mu$ l of protein with 2  $\mu$ l of crystallization buffer and allowing this to equilibrate against 500  $\mu$ l of crystallization buffer. The crystallization buffer for *LmIadA* (20 mg/ml) contained 0.56 M  $\text{NaH}_2\text{PO}_4$  and 1.04 M  $\text{K}_2\text{HPO}_4$ . Crystals were grown at 22 °C and cryo-protected by briefly dipping them in crystallization solution supplemented with 20% glycerol before freezing in liquid nitrogen. Data were collected at the Advanced Light Source (ALS) beamline 5.0.1. The structure was solved by molecular replacement using *E. coli* IadA (PDB code 1YBQ) as a search model. The crystallization buffer for *RhIaaA* (10 mg/ml) contained 0.1 M bis-tris propane pH 8.5, 0.2 M disodium malonate and 25% PEG3350. Crystals were grown at 4 °C and cryo-protected by dipping them in crystallization solution supplemented with 10% PEG100 for 1 minute before freezing in liquid nitrogen. Data were collected at the Canadian Light Source (CLS) beamline CMCF-BM. The structure was solved by molecular replacement using *E. coli* IaaA (PDB code 2ZAL) as a search model. All datasets were processed in DIALS<sup>150</sup> and merged in AIMLESS<sup>196</sup> implemented in CCP4i2 suite<sup>184</sup>. The structures were refined in REFMAC5<sup>187</sup>, Rosetta<sup>152</sup>, Phenix<sup>188</sup> and Coot<sup>153</sup>. Figures were prepared in PyMOL (Schrödinger, USA).

#### **6.5.4. Enzyme activity assays**

Enzyme-catalyzed  $\beta$ -Asp-X dipeptide hydrolysis was measured with an Asp release assay<sup>4</sup>. The 100  $\mu$ l reactions contained 100 mM HEPES pH 8.2, 20 mM KCl, 5 mM  $\alpha$ -ketoglutarate, 1 mM NADH, 2.4 U aspartate aminotransferase, 0.3 U malate dehydrogenase, 1 mM dipeptide substrate and 500 nM purified enzyme. Data were collected by following 340 nm transmittance in 96-well plates using a SpectraMax Paradigm (Molecular Devices, USA) and analyzed using Prism (GraphPad, USA).  $\beta$ -Asp-Arg dipeptides were purified as previously described<sup>13</sup>.  $\beta$ -Asp-Ala and  $\alpha$ -Asp-Arg were purchased from Bachem (Switzerland).  $\beta$ -Asp-Lys and  $\beta$ -Asp-Leu were purchased from Toronto Research Chemicals (Canada).  $\beta$ -Asp-Asp was purchased from Advanced ChemBlocks (USA).

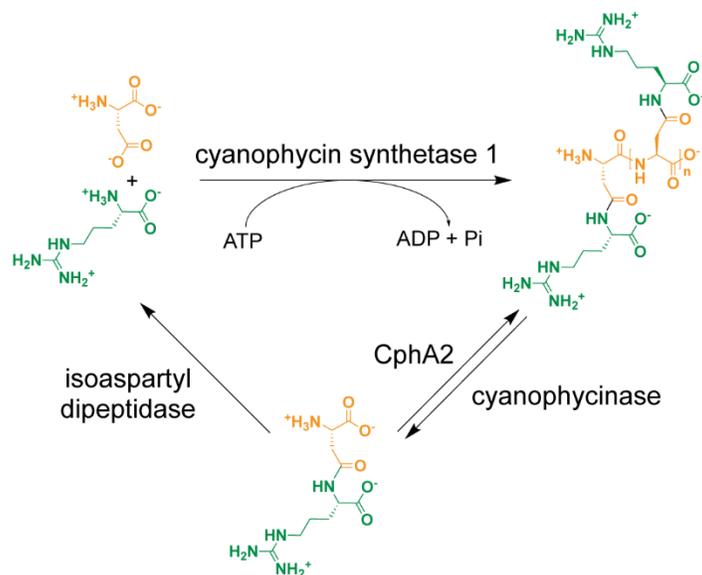
#### **6.6. Acknowledgements**

We thank all the members of the Schmeing lab for important advice and ongoing discussions on this project, Nancy Rogerson for proofreading and synchrotron staff J. Gorin (Canadian Light Source) and M. Allaire (Advanced Light Source) for facilitating remote collection of diffraction datasets. Part of the research described in this paper was performed using beamline CMCF-BM at the Canadian Light Source, a national research facility of the University of Saskatchewan, which is supported by the Canada Foundation for Innovation (CFI), the Natural Sciences and Engineering Research Council (NSERC), the National Research Council (NRC), the Canadian Institutes of Health Research (CIHR), the Government of Saskatchewan, and the University of Saskatchewan. Beamline 5.0.1 of the Advanced Light Source, a U.S. DOE Office of Science User Facility under Contract No. DE-AC02-05CH11231, is supported in part by the ALS-ENABLE program funded by the National Institutes of Health, National Institute of General Medical Sciences, grant P30 GM124169-01.

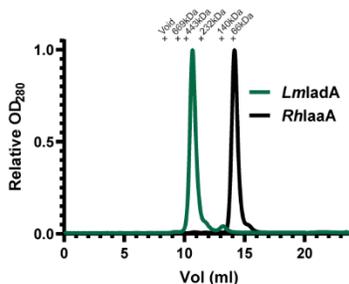
#### **6.7. Data Availability**

Diffraction data and structures determined in this study have been deposited to the Protein Data Bank: *LmIadA* (PDB 8DQN), *RhIaaA* (PDB 8DQM). All other relevant data are within the manuscript.

## 6.8. Supplementary information



**Supplementary Figure 6.1.** Schematic diagram of cyanophycin biosynthesis and degradation.



**Supplementary Figure 6.2.** Size exclusion chromatography traces of *LmiadA* and *RhlaaA* suggesting they migrate as octamer (expected Mw 335 kDa) and heterotetramer (expected Mw 66 kDa), respectively.

**Supplementary Table 6.1.** Statistics for crystallography data collection and structure refinement.

	DSM15395 (PDB: 8DQM)	DSM2157 (PDB: 8DQN)
<b>Data collection</b>		
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	62.2 154.6 197.9	153.5 163.7 170.4
$\alpha$ , $\beta$ , $\gamma$ (°)	90.0 90.0 90.0	90.0 90.0 90.0
Resolution (Å)	98.96-2.70 (2.78-2.70)	118.34-1.80 (1.86-1.80)
<i>R</i> <sub>merge</sub>	0.033 (0.134)	0.100 (0.868)

$R_{\text{pim}}$	0.033 (0.134)	0.028 (0.245)
$I / \sigma I$	8.60 (0.65)	9.68 (0.56)
$CC_{1/2}$	0.999 (0.972)	0.999 (0.895)
Completeness (%)	99.9 (99.9)	98.2 (97.5)
Redundancy	11.2 (7.9)	13.6 (13.2)
<b>Refinement</b>		
Resolution (Å)	98.96-2.70	85.23-1.80
No. reflections	53406 (5216)	386931 (38129)
$R_{\text{work}} / R_{\text{free}}$	0.244/0.268	0.171/0.189
No. atoms	8662	25936
Protein	8510	22779
Ligand/ion	4	104
Solvent	152	3053
$B$ -factors		
Protein	41.18	31.00
Ligands	25.38	56.08
Clashscore	3.09	2.58
Molprobity score	1.33	1.19
R.M.S. deviations		
Bond lengths (Å)	0.013	0.014
Bond angles (°)	1.80	1.83

## Bridge to chapter 7

The results presented in chapter 6 show that in the majority of cases, cyanophycin is degraded by either IaaA or IadA. These two well-studied enzymes have general isoaspartyl dipeptidase activity even when they are clustered with *cphA1* and cyanophycinase, and are believed to be required for processes other than cyanophycin degradation. However, our analysis also showed that a significant portion of bacteria that have *cphA1* or cyanophycinase lack a known isoaspartyl dipeptidase. This could be explained by the existence of uncharacterized enzymes capable of degrading  $\beta$ -Asp-Arg, and near the end of chapter 6 we list several examples of genes that are clustered with *cphA1* and cyanophycinase and may encode such enzymes. In chapter 7, I describe the characterization of the enzyme encoded by one of these genes.

7. Specific cyanophycin dipeptide hydrolase enzymes suggest widespread utility of cyanophycin

Described in: Sharon I, McKay G, Nguyen D, Schmeing TM. *Manuscript in preparation.*

## **7.1. Abstract**

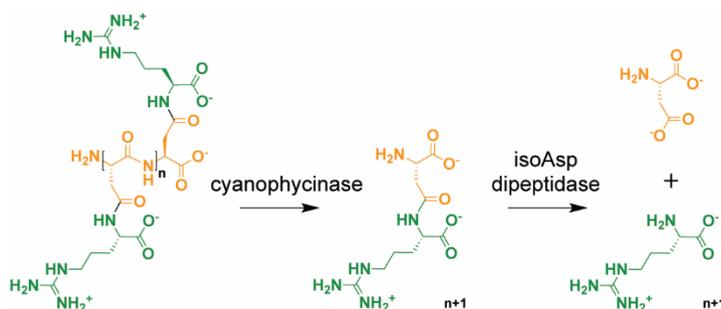
Cyanophycin is a bacterial polymer that is mainly used for nitrogen storage. It is composed of a peptide backbone of L-aspartate residues with L-arginines attached to their side chains through isopeptide bonds. Cyanophycin is degraded in a two-step process whereby cyanophycinase degrades the polymer into  $\beta$ -Asp-Arg dipeptides, and then enzymes with isoaspartyl dipeptidase activity hydrolyze these dipeptides into free Asp and Arg. Isoaspartyl dipeptidase (IadA) and isoaspartyl aminopeptidase (IaaA) have been shown to degrade  $\beta$ -Asp-Arg dipeptides, but it is not uncommon for bacteria which encode cyanophycin-metabolizing genes to lack *iaaA* and *iadA*. In this study, we interrogated a previously uncharacterized enzyme whose gene can cluster with cyanophycin-metabolizing genes. We used bioinformatic, structural and biochemical studies to show that, unlike IadA and IaaA, this common proteobacterial enzyme (CphZ) possesses highly specific cyanophycin-dipeptide hydrolase activity. Activity and sequence analysis demonstrate that AotO, a protein of unknown function in the arginine import and metabolism operon *aot* of *Pseudomonas aeruginosa*, is a CphZ-type cyanophycin dipeptide hydrolase. Using *in vivo* studies, we show that genes in the *aot* operon allow *P. aeruginosa* to utilize  $\beta$ -Asp-Arg as a nitrogen and carbon source. The results suggest that dipeptides resulting from cyanophycin degradation are commonly found in many environments and that CphZ is mainly found in bacteria that scavenge for them.

## 7.2. Introduction

Cyanophycin is a biopolymer produced by a wide range of bacteria<sup>2,31,40,85,99,100,119</sup>. It was first discovered around 140 years ago as large dark granules in cyanobacteria, visible under a simple light microscope<sup>14</sup>. Cyanophycin chains are composed of a backbone of L-aspartate residues with L-arginines attached to each aspartate side chains through isopeptide bonds, thus being a polymer of  $\beta$ -Asp-Arg dipeptide residues<sup>15</sup> (Fig. 7.1). Although most often described as a nitrogen storage polymer<sup>130</sup>, cyanophycin can also be used to store carbon and energy<sup>133,134</sup>. Use of cyanophycin as a dynamic nitrogen reservoir is beneficial for cells in a variety of conditions<sup>52,211,212</sup>. For example, nitrogen fixing cyanobacteria can use it to separate (aerobic) photosynthesis from the strictly anaerobic process of nitrogen fixation. By producing cyanophycin, cells can stockpile excess fixed nitrogen during anaerobic periods of a day/night cycle<sup>42</sup> or in anaerobic cell types<sup>40</sup>. This fixed nitrogen can then be mobilized and used in anaerobic conditions, because cyanophycin catabolism is insensitive to O<sub>2</sub>.

Cyanophycin is most commonly synthesized by cyanophycin synthetase 1 (CphA1)<sup>98</sup>. This enzyme has two synthetic active sites where Asp and Arg are alternately added to the nascent polymer in an ATP-dependent manner<sup>2,98</sup>, and (in many CphA1s) one hydrolytic site that generates cyanophycin primers for synthesis<sup>90</sup>. Some lysine can be incorporated into cyanophycin in place of Arg<sup>96</sup>, the amount of which depends on the CphA1 enzyme, the native or heterologous host and the growth conditions<sup>73</sup>, but is typically much lower than Arg. A second enzyme related to CphA1 called CphA2 is found only in cyanobacteria and can polymerize  $\beta$ -Asp-Arg dipeptides into cyanophycin using its single active site<sup>12</sup>.

Bacteria must degrade cyanophycin polymer into the constituent amino acids to access the carbon, nitrogen and energy stored in it<sup>40,42</sup>. This degradation happens in two steps (Fig. 7.1). First,



**Figure 7.1. The two catalytic steps of cyanophycin biodegradation.**

the polymer is degraded into  $\beta$ -Asp-Arg dipeptides by cyanophycinase<sup>121</sup>. These dipeptides are then cleaved by hydrolysis of the isopeptide bond, yielding Asp and Arg, which can feed into primary metabolism<sup>129</sup>. Unrelated enzymes capable of catalyzing this second reaction can be grouped under the label “isoaspartyl dipeptidases” (Fig. 7.1).

Isoaspartyl dipeptidases are common because they can participate in a fairly widespread protein damage pathway: Proteins can become spontaneously damaged by intramolecular rearrangement of Asp or Asn residues, which transfers the peptide backbone from the Asp/Asn main chain to its side chain. This can be repaired by L-isoaspartyl O-methyltransferases<sup>126</sup>, or the damaged protein can be degraded by proteases that hydrolyze the peptide backbone but leave isoaspartyl dipeptides, which they cannot digest<sup>125,198</sup>. These isodipeptides can accumulate to toxic levels if they are not or degraded<sup>127</sup>, by isoaspartyl dipeptidases<sup>128</sup>. Since various amino acids can become attached to the Asp side chain during protein damage, isoaspartyl dipeptidases accept a wide range of isoaspartyl dipeptides as substrates<sup>129,202</sup>. Isoaspartyl dipeptidases involved in cyanophycin degradation are thought to be moonlighting from their main role as part of the damaged protein repair pathway.

In bacteria, two unrelated isoaspartyl dipeptidase enzymes are known: “isoaspartyl dipeptidase” (IadA)<sup>4,128</sup> and “isoaspartyl aminopeptidase” (IaaA)<sup>129,199-201</sup>. Both can degrade  $\beta$ -aspartyl dipeptides derived from damaged protein, as well as  $\beta$ -Asp-Arg and  $\beta$ -Asp-Lys derived from cyanophycin<sup>213</sup>. However, over 25% of cyanophycin-producing bacteria do not possess genes that encode IadA and IaaA<sup>100,213</sup>. We have recently noted that a putative hydrolase that can cluster with cyanophycin metabolizing genes in some bacteria that lack IadA and IaaA<sup>100,213</sup>.

In this study, we perform structural, biochemical and bioinformatic studies to characterize the activity and role of this previously uncharacterized enzyme from *Acinetobacter baylyi*. We show that, unlike the unrelated, characterized isoaspartyl dipeptidases IaaA and IadA, this enzyme is specific for  $\beta$ -Asp-Arg/Lys, and thus name it cyanophycin dipeptide hydrolase (CphZ). We also show that *Pseudomonas aeruginosa* AotO, a common, previously uncharacterized proteobacterial enzyme commonly found in non-cyanophycin producing bacteria, is an orthologous cyanophycin dipeptide hydrolase. We find that genes in the aot operon, including *aotO*, allow *P. aeruginosa* to use  $\beta$ -Asp-Arg as a nitrogen and carbon source. These results imply that cyanophycin is a common material in many environments, and that bacteria specifically scavenge for its degradation products  $\beta$ -Asp-Arg/Lys.

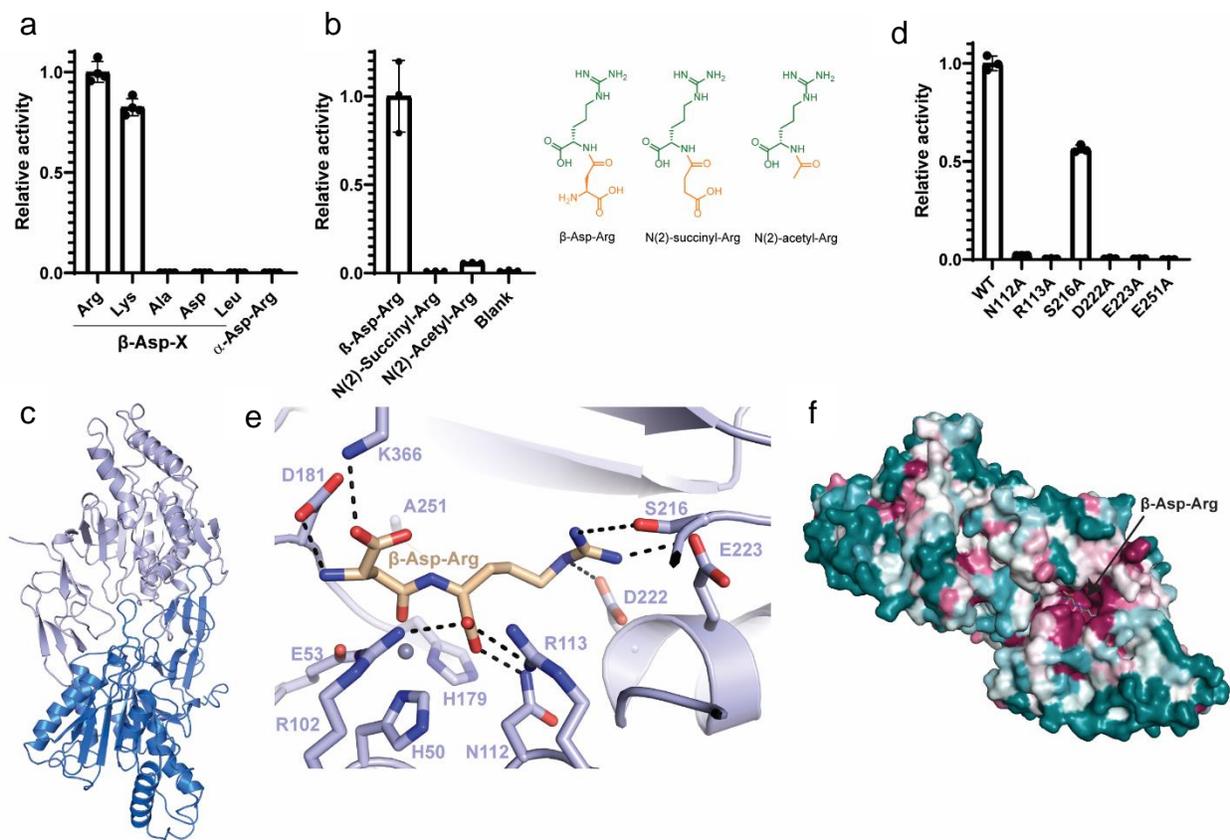
## **7.3. Results**

### **7.3.1. *AbCphZ* is a $\beta$ -Asp-Arg/Lys dipeptidase**

We recently performed a bioinformatic analyses of co-occurrence and clustering of cyanophycin metabolizing genes within bacterial genomes<sup>213</sup>. While cyanophycin synthetase and cyanophycinase often cluster together, *iadA* and *iaaA* usually do not cluster with cyanophycin metabolism genes, and 27% of the genomes that have cyanophycin synthetase do not have an identifiable *iadA* or *iaaA*. In some genomes, however, we identified cyanophycin synthetase-cyanophycinase clusters that also include a gene annotated as “M14 family metallopeptidase” or “succinylglutamate desuccinylase/aspartoacylase family protein”. Notably, among those genomes is that of *Acinetobacter baylyi* DSM587, one of the 5 genomes identified by Füsler et. al.<sup>100</sup> to have *cphA1* and cyanophycinase but not *iadA* or *iaaA*. We hypothesize that *Acinetobacter baylyi* DSM587 “M14 family metallopeptidase” (WP\_004925890.1) which shares ~22% identity with *E. coli* succinylglutamate desuccinylase (AstE<sup>214</sup>) may perform hydrolysis of  $\beta$ -Asp-Arg/Lys derived from cyanophycin, and tentatively named it *A. baylyi* cyanophycin dipeptide hydrolase CphZ (*AbCphZ*).

We cloned and expressed *AbCphZ* in *E. coli* for activity and structural studies. *AbCphZ* could be purified to homogeneity and migrates as a dimer in size exclusion chromatography (Supplementary Fig. 7.1). We tested the enzyme’s ability to hydrolyze various  $\beta$ -aspartyl dipeptides. It displayed a clear preference towards  $\beta$ -Asp-Arg/Lys dipeptides, and possessed almost no activity with other  $\beta$ -aspartyl dipeptides (Fig. 7.2a). *AbCphZ* also displayed specificity toward the  $\beta$ -linkage, as  $\alpha$ -Asp-Arg was not efficiently hydrolyzed (Fig. 7.2a). Activity assays showed that the enzyme is also specific towards the Asp portion of the substrate dipeptide, as it has low activity on N(2)-acetyl-Arg and no detectable activity on N(2)-succinyl-Arg, an intermediate in the arginine succinyltransferase (AST) arginine catabolism pathway which is very similar in structure to  $\beta$ -Asp-Arg<sup>214</sup> (Fig. 7.2b). This substrate specificity is remarkable, as we are unaware of any other enzyme that has specific activity for  $\beta$ -Asp-Arg/Lys dipeptides. Furthermore, cyanophycin degradation is the only abundant source of  $\beta$ -Asp-Arg/Lys dipeptides of which we are aware.

To determine the structural basis for this substrate specificity, we solved the crystal structure of WT *AbCphZ* at 2.5 Å resolution (Fig. 7.2c, Supplementary Table 7.1). *AbCphZ* is a dimer with an overall structure that shares modest similarity with those of *E. coli* AstE (PDB code



**Figure 7.2. Structural and biochemical characterization of *AbCphZ*.** (a) Asp release assay of *AbCphZ* with different Asp-dipeptides. The enzyme displays high specificity towards  $\beta$ -Asp-Arg/Lys dipeptides. Error bars represent the standard deviation of the mean of  $n=4$  replicates. (b) Arg release assay of *AbCphZ* with Arg-containing substrates. The enzyme preferentially cleaves  $\beta$ -Asp-Arg, suggesting it recognizes the Asp portion of the substrate. Error bars represent the standard deviation of the mean of  $n=3$  replicates. (c) The dimeric crystal structure of *AbCphZ*. (d) Asp release assay of *AbCphZ* mutants with  $\beta$ -Asp-Arg. The conserved N112, R113, D222, E223 and E251 are essential for activity, while the non-conserved S216 is not. Error bars represent the standard deviation of the mean of  $n=3$  replicates. (e) Close view of the active site of *AbCphZ* E251A with bound  $\beta$ -Asp-Arg. (f) Sequence conservation map of CphZ enzymes generated using ConSurf<sup>6</sup>. The enzyme displays very high sequence conservation in the active site.

1YW6, RMSD 3.8 Å across 109 C $\alpha$  pairs, Supplementary Fig. 7.2a) and bovine pancreatic carboxypeptidase A (PDB code 1HEE, RMSD 4.1 Å across 159 C $\alpha$  pairs). As with these two enzymes, the active site of *AbCphZ* contains a single metal ion liganded by a conserved H-H-E

triad (*AbCphZ* H50, E53, and H179, Supplementary Fig. 7.2b,c). The density map indicates the presence of another ion liganded by Y62 of one protomer and H315 and S330 of another protomer within the CphZ dimer. Inductively coupled plasma mass spectrometry (ICP-MS) showed our CphZ samples contain significant amounts of both zinc and manganese (Supplementary table 7.2). We verified that the ion in the active site is  $\text{Zn}^{2+}$  using single-wavelength anomalous diffraction (Supplementary Fig. 7.2d), and modelled  $\text{Mn}^{2+}$  in the second metal site. The conserved E251 is positioned close to the  $\text{Zn}^{2+}$  ion and putative substrate binding site (Supplementary Fig. 7.2c), and is essential for activity (Fig. 7.2d), like analogous glutamates in other zinc metallopeptidases<sup>169</sup>.

To understand how the enzyme specifically binds and cleaves cyanophycin-derived dipeptides, we solved the co-complex structure of the catalytically inactive mutant CphZ<sub>E251A</sub> bound to  $\beta$ -Asp-Arg at 2.4 Å resolution (Supplementary table 7.1). The maps show no substantial rearrangement from the apo structure and clear density for  $\beta$ -Asp-Arg in the active site (Supplementary Fig. 7.2e). Similarly to other zinc metallopeptidases<sup>215</sup>, the carbonyl of the scissile bond is positioned 2.0 Å away from the  $\text{Zn}^{2+}$  ion (Fig. 7.2e). The metal should serve as a Lewis acid and draw electrons from the bond and facilitate its cleavage, presumably by a water molecule activated by E251<sup>169</sup>. The structure explains how CphZ specifically recognizes and binds the Arg/Lys portion of the substrate. Conserved residues D222 and E223 form a negatively-charged pocket suitable for the binding of a positively charged guanidino or amino group (Fig. 7.2e). In addition, the conserved R102, N112 and R113 bind the carboxyl part of the substrate Arg residue. Mutation of any of N112, R113, D222 and E223 abolished activity (Fig. 7.2d). The portion of the active site surrounding the substrate's Asp residue shows lower sequence conservation in general (Fig. 7.2f), but D181 and K366 are well conserved and within hydrogen bonding distance from the Asp backbone (Fig. 7.2e). To determine the importance of the Asp and Arg portions of  $\beta$ -Asp-Arg for substrate recognition and binding, we performed isothermal titration calorimetry (ITC) binding studies with the E251A mutant and  $\beta$ -Asp-Arg, Asn and Arg. The results show that *AbCphZ* has the lowest  $K_D$  for  $\beta$ -Asp-Arg at  $7.6 \pm 0.1 \mu\text{M}$ , while its  $K_D$  for Arg is  $470.8 \pm 16.8 \mu\text{M}$  and experiments with Asn gave no measurable signal.

### **7.3.2. *P. aeruginosa* AotO is a CphZ**

Sequence alignment using BLAST<sup>63</sup> showed that residues 18-280 of *AbCphZ* have similarity to a domain annotated as “M14\_PaAOTO\_like” (accession cd06250) which is also annotated in the uncharacterized protein AotO<sup>216</sup> of *Pseudomonas aeruginosa* (WP\_128550578.1,

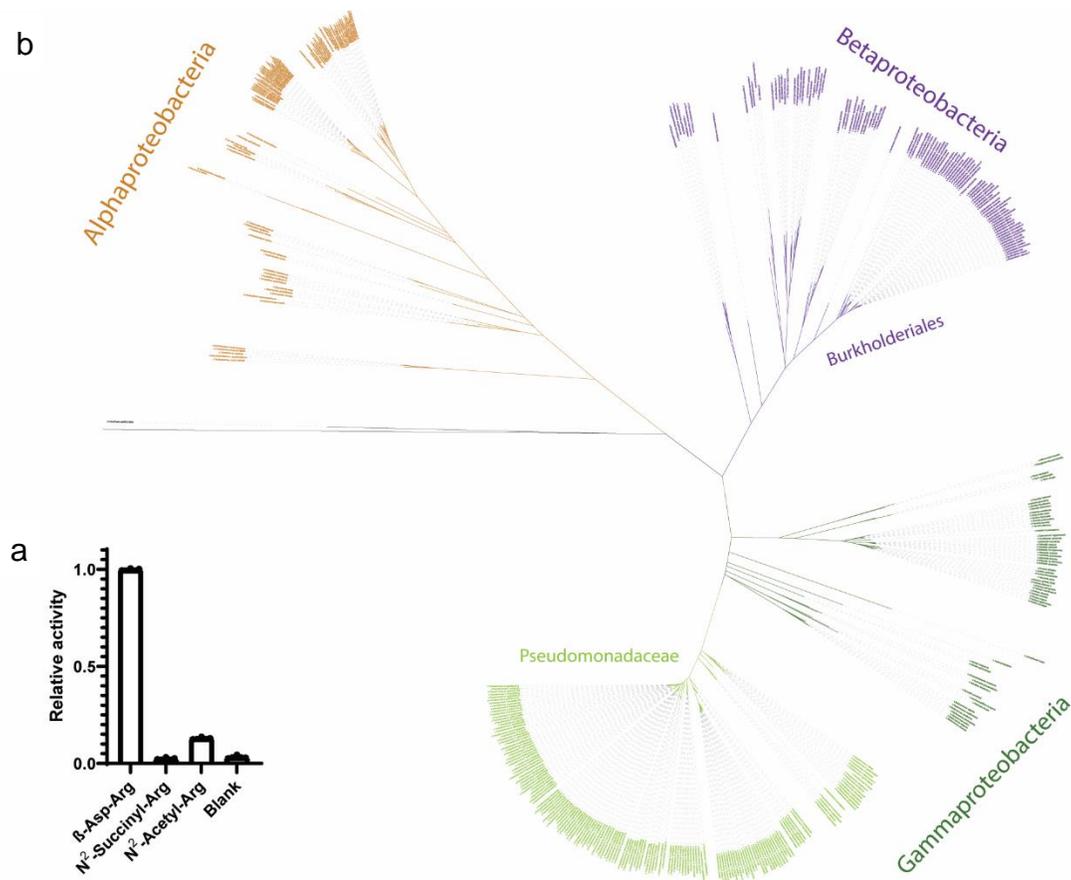
PA0891<sup>217</sup>, here *PaAotO*). Overall, *PaAotO* and *AbCphZ* share 38% sequence identity, with high similarity around the active site region (Supplementary Fig. 7.3a). Although there are no known CphA1 enzymes in *Pseudomonas* species, this similarity suggested that CphZ and AotO might have similar activity.

The gene *aotO* is part of the *aot* operon, which also includes a multi-component Arg transporter (*aotJQMP*) and an Arg-dependent transcription activator (*aotR*)<sup>216</sup>. This operon increases Arg uptake and is upregulated by it. A previous study found that knockout of *P. aeruginosa aotO* does not affect Arg uptake, and the authors could not determine this gene's role<sup>216</sup>. To verify whether CphZ and AotO have the same activity and should be considered homologs of the same enzyme, we purified *PaAotO* and tested its activity towards  $\beta$ -Asp-Arg, N(2)-succinyl arginine and N(2)-acetyl arginine (Fig. 7.3a). The enzyme displayed similar substrate specificity to that of *AbCphZ*, and hydrolyzed  $\beta$ -Asp-Arg at a rate comparable to that of *AbCphZ* and the previously characterized isoaspartyl dipeptidases *RhIaaA* and *LmIadA*<sup>213</sup> (Supplementary Fig. 7.3b). This shows *P. aeruginosa AotO* is a cyanophycin dipeptide hydrolase.

### 7.3.3. The genetic context of CphZ suggests a role in dipeptide scavenging

To better understand the role of CphZ, we performed a bioinformatic analysis of its distribution and genomic localization with respect to other cyanophycin biosynthesis genes. First, we searched the RefSeq database<sup>204</sup> for complete bacterial genomes that have CphZ and constructed a phylogenetic tree of these using phyloT<sup>158</sup> according to the genome taxonomy database classification<sup>218</sup> (Fig. 7.3b). The analysis shows that CphZ is a proteobacterial enzyme. It is particularly common among the family *Pseudomonadaceae* and order *Burkholderiales*. While many members of the order *Burkholderiales* have CphA1 homologs<sup>213</sup>, a BLAST search shows that cyanophycin producers are extremely rare in *Pseudomonadaceae*, with only one strain found to have a *cphA1*.

Next, we constructed a database of all the complete bacterial genomes in the RefSeq database and used Cblaster<sup>209</sup> to analyze the co-occurrence and clustering of CphA1, cyanophycinase (CphB/E/I) and cyanophycin dipeptide hydrolase (CphZ/AotO). Of the 27,349 genomes in that database, 1614 (~6%) have *cphA1*, 840 (~3%) have cyanophycinase and 1364 (~5%) have *cphZ*. In total, 3,095 (~11%) genomes have at least one of *cphA1*, cyanophycinase or *cphZ* (Table 7.1). However, the pattern of co-occurrence in these genomes is surprising. Very few have both *cphA1* and *cphZ* (56 genomes) or cyanophycinase and *cphZ* in (50 genomes). Forty one



**Figure 7.3. *PaAotO* activity and phylogeny.** (a) Arg release assay of *PaAotO* with Arg-containing substrates. The enzyme preferentially cleaves  $\beta$ -Asp-Arg, suggesting it recognizes the Asp portion of the substrate. Error bars represent the standard deviation of the mean of  $n=3$  replicates. (b) Unrooted phylogenetic tree of CphZ/AotO distribution.

of those genomes contain all three genes. We also examined the tendency of *cphZ* to cluster with *cphA1* and cyanophycinase. We defined clustering as having an intergenic region of up to 5 kilobases. Of the 56 genomes which have both *cphA1* and *cphZ*, they are found clustered in 13, whereas cyanophycinase genes and *cphZ* cluster in 32 of 52 genomes (Table 7.1). All three genes cluster in 12 genomes out of 41. These numbers are in line with those reported for *IaaA* and *IadA*<sup>213</sup>. Taken together, these results indicate that the cluster in *Acinetobacter baylyi* DSM587 is the exception rather than norm. The genomic context observed in *P. aeruginosa*, that has *aotO* without other cyanophycin metabolizing genes, is more common, partially thanks to this operon's ubiquity among *Pseudomonadaceae*.

	Total	CphA1	cyano- phycinase	CphZ	CphA1 + CphZ cyano- phycinase	CphA1 + cyano- phycinase	CphZ + cyano- phycinase	CphA1 + CphZ
In genome	3095	1614	840	1364	41	659	50	56
Clustered	563	543	562	33	12	542	32	13

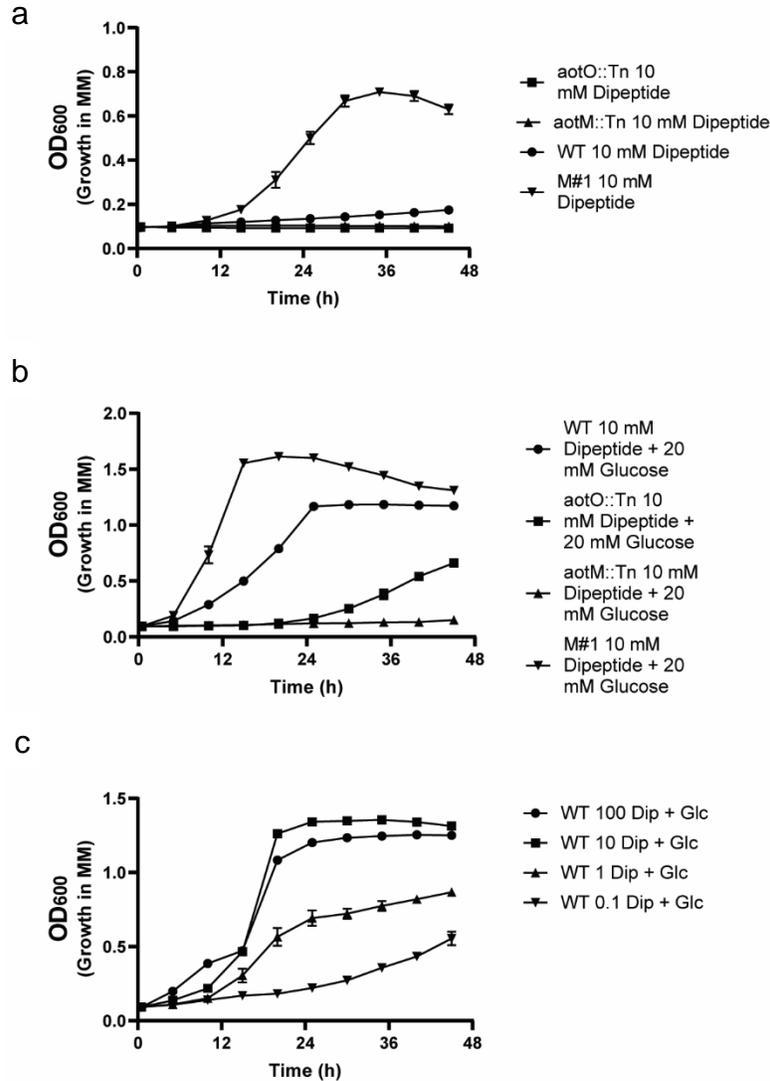
**Table 7.1. Analysis of *cphA1*, cyanophycinase and *cphZ* co-occurrence and clustering.**

### 7.3.4. The *aot* operon enables *P. aeruginosa* to utilize $\beta$ -Asp-Arg as a nitrogen source

Next, we set to examine the activity of AotO *in vivo*. We grew wild type (WT) *P. aeruginosa* PAO1, as well as transposon mutants with interrupted *aotO* (*aotO*::Tn) and *aotM* (*aotM*::Tn), in liquid minimal media containing  $\beta$ -Asp-Arg as a sole carbon and nitrogen source. Both mutants displayed no growth after a 2-day incubation, while the WT displayed low but measurable growth levels (Fig. 7.4a). However, when glucose was added as a carbon source the WT cells displayed robust growth rates, showing they can use  $\beta$ -Asp-Arg as a nitrogen source in a dose-dependent manner (Fig. 7.4b,c). The transposon mutant *aotO*::Tn could also use dipeptides for nitrogen, but not nearly as well as the WT, while *aotM*::Tn cells could not use dipeptides as a nitrogen source (Fig. 7.4b). All three strains displayed similar growth rates on media containing either Asp, Arg or NH<sub>4</sub>Cl and glucose as nitrogen and carbon sources (Supplementary Fig. 7.4a-e), suggesting the transposon mutations' effect is limited to the ability to metabolize  $\beta$ -Asp-Arg. None of the strains could use  $\beta$ -Asp-Ala instead of  $\beta$ -Asp-Arg, as this molecule did not support cell growth (Supplementary Fig. 7.4f).

We also grew WT *P. aeruginosa* cultures on solid minimal media containing  $\beta$ -Asp-Arg as a sole nitrogen and carbon source. While very little growth was observed initially, following a week-long incubation several colonies appeared which showed rapid growth rates. We isolated several of these and tested their ability to utilize  $\beta$ -Asp-Arg in liquid media cultures. To our surprise, these isolates displayed high growth rates using only dipeptides as a source of nitrogen and carbon (M#1 in Fig. 7. 4a,b, Supplementary Fig. 7.4). To detect what mutations may have led to this increased ability to utilize dipeptides as a nutrient source, we performed whole-genome sequencing of three of the isolates. All three had mutations that are likely to inactivate AruF, the first enzyme in the AST pathway – the major Arg and Orn utilization pathway in *P. aeruginosa*<sup>214,219</sup>. As with the other strains, the isolates were also unable to utilize  $\beta$ -Asp-Ala

instead of  $\beta$ -Asp-Arg (Supplementary Fig. 7.4f). However, a transposon mutant *aruF*::Tn mutant did not show identical growth characteristics to M#1.



**Figure 7.4. *In vivo* growth of *P. aeruginosa* using  $\beta$ -Asp-Arg.** The experiments compare the growth of WT and three strains: a transposon mutant disrupting *aotO* (*aotO*::Tn), a transposon mutant disrupting *aotM* (*aotM*::Tn) and a strain isolated after growing for 1 week on minimal media supplemented only with  $\beta$ -Asp-Arg. Each panel shows growth on minimal media supplemented with different nitrogen and carbon sources: (a) 10 mM  $\beta$ -Asp-Arg. (b) 10 mM  $\beta$ -Asp-Arg and 20 mM glucose. (c) 20 mM glucose and different  $\beta$ -Asp-Arg concentrations. Each data point is the average of n=3 independent replicates.

#### **7.4. Discussion**

The biochemical and structural results show that CphZ specifically recognizes both the Arg/Lys and Asp portions of its B-Asp-Arg/Lys substrates, making it the first known enzyme that is seemingly dedicated to the degradation of cyanophycin dipeptides. This is unlike IaaA and IadA, which have broad substrate specificity and are believed to only degrade cyanophycin metabolites as a secondary function. Because *AbCphZ* and *PaAotO* have high sequence identity around their active site and display the same substrate specificity, CphZ and AotO should be classified as homologs of the same enzyme, cyanophycin dipeptide hydrolase. Additionally, CphZ is highly similar to two uncharacterized proteins with structures deposited in the PDB: *Shewanella amazonensis* succinylglutamate desuccinylase aspartoacylase (PDB code 3FMC, 27% identity, 1.8 Å RMSD across 220 C $\alpha$  pairs) and *Shewanella frigidimarina* putative succinylglutamate desuccinylase/aspartoacylase (3LWU, 29% identity, 1.7 Å RMSD across 220 C $\alpha$  pairs). A BLAST<sup>63</sup> search showed that both of these gammaproteobacterial species have cyanophycinase homologs, so these proteins may also be CphZ enzymes.

Our bioinformatic analysis shows that CphZ/AotO is common among proteobacteria. Surprisingly, we mostly found it in strains that cannot produce cyanophycin. Another study aiming to identify core genes among *Pseudomonas* groups found the *aot* operon, and *aotO* specifically, to be highly conserved among several common species, and partially conserved in others<sup>220</sup>. These results, coupled with the substrate specificity and isolation of cyanophycin-degrading bacteria from a variety of environments and microbiomes<sup>57-61</sup>, suggest that cyanophycin is a common material in many environments, making it advantageous for bacteria to have mechanisms for its scavenging. The low co-occurrence of CphZ with cyanophycinase in the same genomes suggests that most of the cyanophycin found in the environment is in the form of already degraded dipeptides. Alternatively, some proteobacteria may have uncharacterized enzymes with cyanophycinase activity that complement CphZ's dipeptide hydrolase activity.

Our *in vivo* results show that *P. aeruginosa* can utilize  $\beta$ -Asp-Arg as a nitrogen and carbon source. Moreover, this ability depends on genes from the *aot* operon, confirming their roles in the uptake (*aotM*) and degradation (*aotO*) of these dipeptides. A previous study<sup>216</sup> found that AotJQMP can function as an Arg transporter, but is not essential for its uptake by *P. aeruginosa*. Thus, it seems like the unique advantage conferred by this transporter is its ability to import  $\beta$ -Asp-Arg/Lys dipeptides. In agreement with our *in vitro* results, AotO seems to be important for  $\beta$ -

Asp-Arg hydrolysis *in vivo*. The *aotO::Tn* mutant had reduced ability to use the dipeptides as a nitrogen source, suggesting other enzymes could not fully compensate for its loss of AotO activity.

The dipeptide-utilizing mutants we isolated all had frameshift or truncation mutations in the *aruF* gene that are likely to inactivate AruF. This enzyme allows Arg to enter the AST pathway, the main arginine catabolism pathway under aerobic conditions in *P. aeruginosa*<sup>219</sup>. It is not yet clear how, or indeed if, these mutations allow the cells to utilize  $\beta$ -Asp-Arg more efficiently. One possibility is that the inactivation of AruF may lead to an increase in intracellular Arg levels. As the *aot* operon is upregulated by Arg, this likely leads to higher expression of the AotJQMP transporter and AotO, and so enhance the cells' ability to utilize exogenous  $\beta$ -Asp-Arg.

As CphZ is specific for  $\beta$ -Asp-Arg/Lys dipeptides, the question about an apparent lack of a general isoaspartyl dipeptidase in bacteria like *A. baylyi* and *P. aeruginosa* remains. These species do normally have isoaspartyl-O-methyltransferases, which are capable of repairing isoaspartyl dipeptides embedded in proteins, and it is possible that these enzymes are sufficiently efficient to prevent the harmful accumulation of  $\beta$ -Asp dipeptides in the cells<sup>127</sup>. Alternatively, some bacteria may have yet uncharacterized isoaspartyl dipeptidases with broad substrate specificity.

## **7.5. Methods**

### **7.5.1. Cloning, protein expression and purification**

The gene encoding *AbCphZ* (WP\_004925890.1) was amplified from genomic DNA (DSMZ, Germany). The gene encoding *PaAotO* (WP\_128550578.1) was codon optimized for expression in *E. coli* and synthesized (Biobasic, Canada). The genes encoding *AbCphZ* and *paAotO* were cloned into a pJ411-derived plasmid with a C-terminal tobacco etch virus (TEV) protease recognition site and an 8xHis affinity tag. All cloning and mutagenesis were performed by transforming DH5- $\alpha$  *E. coli* cells with PCR fragments containing overlapping ends. Protein expression was carried out in *E. coli* BL21(DE3) grown in TB media supplemented with 150  $\mu$ g/ml. Cultures were grown at 37 °C until an OD<sub>600</sub> of ~1 was reached, and the growth temperature was then lowered to 18 °C. Protein expression was induced by the addition of 0.2 mM isopropyl  $\beta$ -d-1-thiogalactopyranoside (IPTG) for ~20 hours, and the cells were then harvested by centrifugation. Following this, all the purification steps were carried out at 4 °C. The cell pellets were resuspended in buffer A (250 mM NaCl, 50 mM Tris pH 8.0, 10 mM imidazole, 2 mM  $\beta$ -

mercaptoethanol) supplemented with a few crystals of lysozyme and DNaseI and lysed on ice by sonication. The lysate was clarified by centrifugation at 40,000 g for 30 minutes and then applied onto a HisTrap HP column (Cytiva, USA). Following loading, the protein was washed with at least 20 column volumes of buffer B (buffer A with 30 mM imidazole) and eluted with buffer C (buffer A with 250 mM imidazole). For structural studies, the protein was dialyzed overnight against buffer D (250 mM NaCl, 20 mM Tris pH 8.0, 5 mM  $\beta$ -mercaptoethanol) in the presence of TEV protease for removal of the 8xHis tag. Following tag cleavage, the protein was again applied to a HisTrap column and the flow through was collected. For all protein preparations, the next purification step was concentration using an Amicon Ultra centrifugal filter and applied to a Superdex200 16/60 column (Cytiva, USA) equilibrated in buffer E (100 mM NaCl, 20 mM Tris pH 8.0, 1 mM dithiothreitol). Fractions containing the highest protein purity were pooled, and following concentration were supplemented with glycerol to a final volume of 15% and flash frozen in liquid nitrogen for storage.

### **7.5.2. Protein crystallization, data collection, structure solution and refinement**

Protein crystals were grown using the sitting drop method using a reservoir volume of 500  $\mu$ l and a drop containing 2  $\mu$ l of protein sample in buffer E and 2  $\mu$ l of crystallization buffer. The crystallization buffer for WT *AbCphZ* (5 mg/ml) contained 0.1 M bis-tris propane pH 7.5, 24% PEG3350, 0.2 M NaBr, 10 mM spermine and 10 mM LiCl. Crystals were grown at 4 °C, dehydrated by allowing the drop to equilibrate through vapor diffusion against 0.1 M bis-tris propane pH 7.5, 30% PEG3350, 0.2 M NaBr and 10 mM spermine overnight, and then cryo-protected by dipping them in 0.1 M bis-tris propane pH 7.5, 20% PEG3350, 20% ethylene glycol, 0.2 M NaBr and 10 mM spermine. Data were collected at the Advanced Photon Source (APS) beamline 24-ID-C. The same crystals were also used for EDS experiments performed at the APS. The structure was solved by molecular replacement using PDB codes 3FMC and 3LWU as search models. The E251A *AbCphZ* with  $\beta$ -Asp-Arg crystals were grown in similar conditions to the WT *AbCphZ* ones. The crystals were soaked in a cryo-protection solution containing 10 mM  $\beta$ -Asp-Arg for 30 minutes prior to freezing. Data were collected at the ALS beamline 5.0.1. All datasets were processed in DIALLS<sup>150</sup> and merged in AIMLESS<sup>196</sup> implemented in CCP4i2 suite<sup>184</sup>. The structures were refined in REFMAC<sup>187</sup>, Rosetta<sup>152</sup>, Phenix<sup>188</sup> and Coot<sup>153</sup>. Figures were prepared in PyMol. For anomalous difference map calculation, data from E251A *AbCphZ* crystals with  $\beta$ -

Asp-Arg were collected at the Canadian Light Source (CLS) beamline CMCF-BM at a wavelength of 1.2828 Å and the map was calculated using “calculate unusual map coefficients” in CCP4i2.

### 7.5.3. Enzyme activity assays

For the detection of Asp release from  $\beta$ -Asp-X dipeptides, an Asp release assay similar to the one previously described<sup>4</sup> was used. Each 100  $\mu$ l reaction contained 100 mM HEPES pH 8.2, 20 mM KCl, 5 mM  $\alpha$ -ketoglutarate, 500 nM purified enzyme, 0.3 U malate dehydrogenase, 1 mM NADH, 2.4 U aspartate aminotransferase and 1 mM dipeptide substrate. For other Arg containing substrates, Arg release was monitored by using a free Arg detection kit (K-LARGE, NEOGEN, USA). The 135  $\mu$ l reactions contained 15  $\mu$ l buffer solution, 10  $\mu$ l NADPH solution, 1  $\mu$ l GIDH suspension, 2.5  $\mu$ l urease solution, 1  $\mu$ l arginase suspension, 1 mM substrate and 500 nM enzyme. For both assays, reaction progression was monitored by following 340 nm transmittance in 96-well plates. Data were collected using a SpectraMax Paradigm (Molecular Devices, USA) and analyzed using GraphPad Prism (GraphPad, USA). Other than  $\beta$ -Asp-Arg, all molecules used as substrates were purchased from several vendors:  $\beta$ -Asp-Ala and  $\alpha$ -Asp-Arg from Bachem (Switzerland); N<sup>2</sup>-acetyl arginine,  $\beta$ -Asp-Lys and  $\beta$ -Asp-Leu from Toronto Research Chemicals (Canada);  $\beta$ -Asp-Asp from AchemBlock (USA); and N<sup>2</sup>-succinyl arginine from BLD Pharmatech (USA).  $\beta$ -Asp-Arg dipeptides were purified as previously described<sup>13</sup>.

### 7.5.4. Isothermal titration calorimetry

ITC experiments were carried out using a MicroCal iTC200 (GE Healthcare, USA) at 25 °C. For  $\beta$ -Asp-Arg, the cell contained 200  $\mu$ M purified E251A *AbCphZ* in buffer F (100 mM NaCl, 20 mM Tris pH 8.0, 5 mM  $\beta$ -mercaptoethanol) and the syringe contained 1 mM  $\beta$ -Asp-Arg in buffer F. For L-Arg and L-Asn, the cell contained 400  $\mu$ M purified E251A *AbCphZ* and the syringe contained 10 mM substrate, both in buffer F. A total of nineteen 2  $\mu$ l injections were interspaced by 180 seconds each. Data were analyzed using Microcal Origin 7.0 (OriginLab, USA) with a binding model stoichiometry of 1:1. Measurements were performed in triplicates.

### 7.5.5. Inductively coupled plasma mass spectrometry (ICP-MS)

A sample of WT *AbCphZ* was buffer-exchanged into 100 mM (NH<sub>4</sub>)<sub>2</sub>CO<sub>3</sub> by performing gel filtration with a Superdex S200 10/300 column equilibrated with that buffer. Protein containing fractions were concentrated to 100  $\mu$ M and analyzed by ICP-MS at the Center for Applied Isotope Studies, University of Georgia. A sample of the buffer eluted from the column was used as a control.

### **7.5.6. Bioinformatic analysis of gene co-occurrences**

All the complete bacterial genomes in the Refseq database were used to create a local database (May 2022). Cblaster<sup>209</sup> was used in local mode to search this database using several templates for CphA1 (accession codes WP\_028947105.1, WP\_004925893.1 and WP\_015942562.1), cyanophycinase (WP\_011058003.1, WP\_004925892.1 and Q8KQN8.1) and *PaAotO* and *AbCphZ*. The resulting binary table was analyzed using Excel.

### **7.5.7. Synthesis and purification of $\beta$ -Asp-Arg**

$\beta$ -Asp-Arg dipeptides were made from purified insoluble cyanophycin produced in *E. coli* by *SuCphA1*<sup>197</sup>. The purified polymer was washed by resuspension in ddH<sub>2</sub>O and centrifugation at 3500g for 10 minutes. The washed polymer was resuspended in ddH<sub>2</sub>O and digested with purified cyanophycinase from *Synechocystis sp.* PCC6803<sup>124</sup> until the suspension became clear, filtered using a 3 kDa molecular weight cut-off Amicon centrifugation concentrator (EMD Millipore) and lyophilized. The isolated compound was analyzed using an Arg/NH<sub>4</sub>/Urea detection kit (K-LARGE, NEOGEN, USA) to verify it contained no other potential nitrogen sources.  $\beta$ -Asp-Ala dipeptides were purchased from Bachem (Switzerland).

### **7.5.8. *P. aeruginosa* growth assays**

*P. aeruginosa* cultures were grown in minimal media<sup>221</sup> supplemented with different nitrogen and carbon sources, as indicated. Liquid media cultures were grown at 37 °C under shaking. Growth was measured by monitoring OD<sub>600</sub>. For isolation of strains from solid media, cultures were grown at 37 °C on agar-based minimal media supplemented with 10 mM  $\beta$ -Asp-Arg as the sole nitrogen and carbon source.

## **7.6. Acknowledgements**

We thank all the members of the Schmeing lab for important advice and ongoing discussions on this project, Kim Munro for help with ITC experiments and synchrotron staff D. Neau (Advanced Photon Source) and M. Allaire (Advanced Light Source) for facilitating remote collection of diffraction datasets. This study includes work based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the National Institutes of Health (P30 GM124165). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory

under Contract No. DE-AC02-06CH11357. Beamline 5.0.1 of the Advanced Light Source, a U.S. DOE Office of Science User Facility under Contract No. DE-AC02-05CH11231, is supported in part by the ALS-ENABLE program funded by the National Institutes of Health, National Institute of General Medical Sciences, grant P30 GM124169-01.

## **7.7. Supplementary information**

### **7.7.1 Supplementary tables**

	<i>AbCphZ</i>	<i>AbCphZ</i> E251A + $\beta$ -Asp-Arg
<b>Data collection</b>		
Space group	C121	C121
Cell dimensions		
$a, b, c$ (Å)	152.5 127.7 106.9	151.9 127.0 109.4
$\alpha, \beta, \gamma$ (°)	90 129.5 90	90.0 130.3 90.0
Resolution (Å)	58.85-2.70 (2.80-2.7)	54.46-2.40 (2.49-2.40)
$R_{\text{merge}}$	0.037 (0.273)	0.07169 (0.201)
$R_{\text{pim}}$	(0.037 (0.273))	0.049 (0.148)
$I / \sigma I$	15.26 (0.21)	6.6800 (0.30)
$CC_{1/2}$	0.998 (0.847)	0.999 (0.947)
Completeness (%)	83.81 (26.47)	99.3 (98.3)
Redundancy	5.3 (5,3)	2.8 (2.5)
<b>Refinement</b>		
Resolution (Å)	58.85-2.70	85.6-2.40
No. reflections	36397	61419 (6064)
$R_{\text{work}} / R_{\text{free}}$	0.227/0.247	0.209/0.246
No. atoms	11493	11994
Protein	11485	11469
Ligand/ion	8	88
Solvent	0	437
<b>B-factors</b>		
Protein	94.85	45.49
Ligands	111.58	52.35
Clashscore	2.55	3.06
Molprobability score	1.17	1.10
R.M.S. deviations		

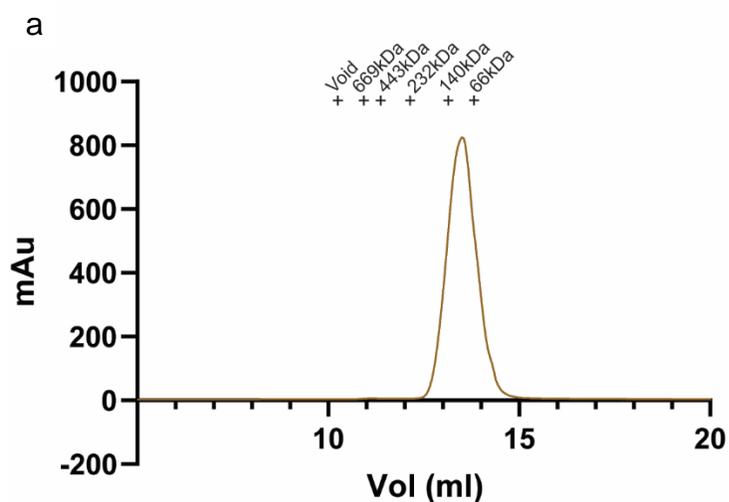
Bond lengths (Å)	0.014	0.012
Bond angles (°)	1.77	1.58

**Supplementary table 7.1. Statistics for X-ray crystallography data collection and structure refinement.**

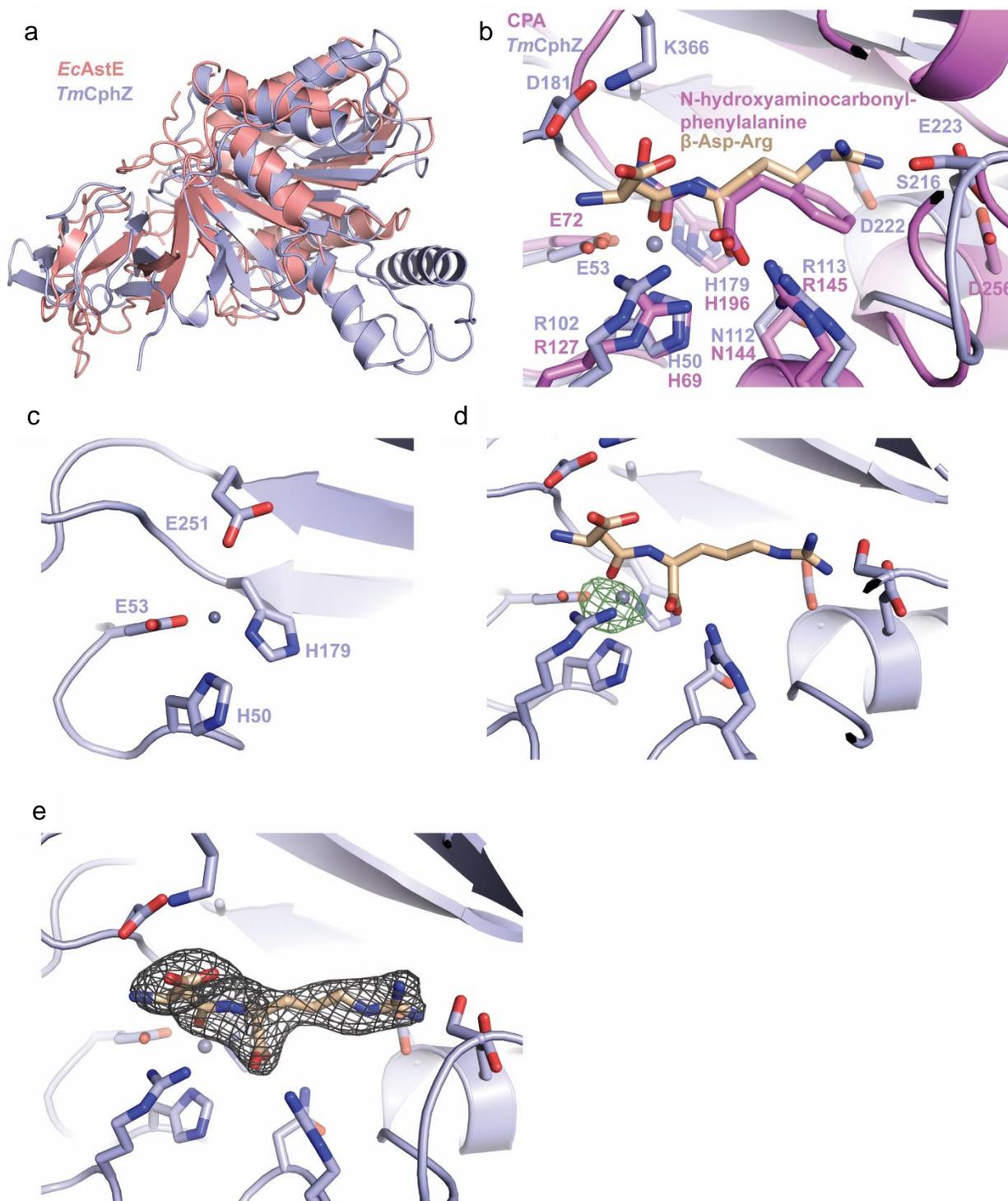
Sample	<sup>55</sup> Mn	<sup>56</sup> Fe	<sup>59</sup> Co	<sup>60</sup> Ni	<sup>65</sup> Cu	<sup>66</sup> Zn
Blank	BQ < 0.992	BQ < 29.1	BQ < 0.677	BQ < 3.7	BQ < 1.54	2.28
CphZ	2485	103	1.38	27.0	15.0	1659

**Supplementary table 7.2. CphZ metal analysis.** ICP-MS analysis of *AbCphZ* and a buffer control sample showing high amounts of manganese and zinc are present only in the protein sample. All values are in µg/kg units.

### 7.7.2. Supplementary figures

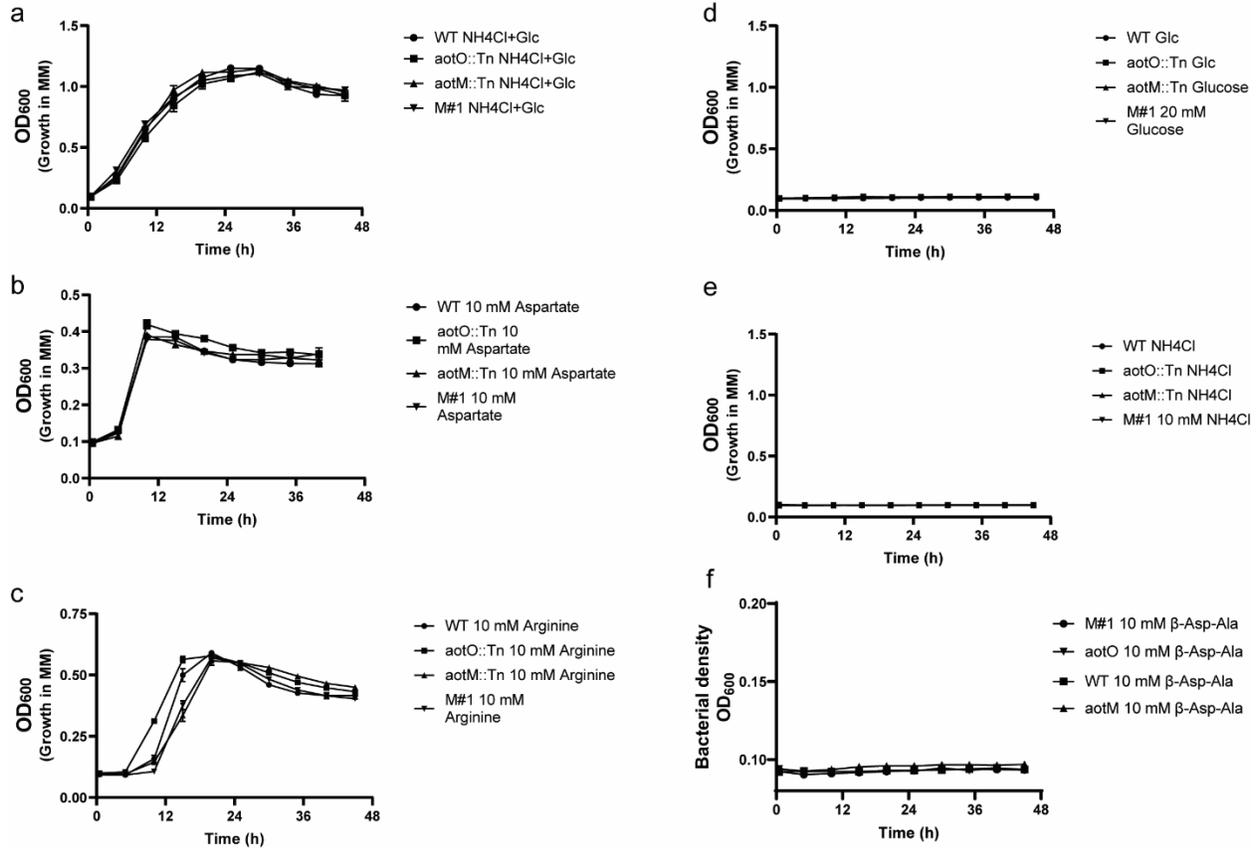


**Supplementary figure 7.1. SEC chromatogram of *AbCphZ*.** The protein migrates as a dimer (expected size 86 kDa).



**Supplementary figure 7.2. Structural characterization of *AbCphZ*.** (a) Overlay of WT *AbCphZ* and *EcAstE* (PDB code 1YW6) monomers showing modest structural similarity. (b) Overlay of the active sites of *AbCphZ* E251A in complex with  $\beta$ -Asp-Arg and bovine pancreatic carboxypeptidase A in complex with the inhibitor L-N-hydroxyaminocarbonyl phenylalanine





**Supplementary figure 7.4. *In vivo* growth of *P. aeruginosa* using  $\beta$ -Asp-Arg.** The experiments compare the growth of WT and three strains: a transposon mutant disrupting *aotO* (*aotO*::Tn), a transposon mutant disrupting *aotM* (*aotM*::Tn) and a strain isolated after growing for 1 week on minimal media supplemented only with  $\beta$ -Asp-Arg. Each panel shows growth on minimal media supplemented with different nitrogen and carbon sources: (a) 10 mM  $\text{NH}_4\text{Cl}$  and 20 mM glucose. (b) 10 mM Asp. (c) 10 mM Arg. (d) 20 mM glucose. (e) 10 mM  $\text{NH}_4\text{Cl}$ . (f)  $\beta$ -Asp-Ala. Each data point is the average of  $n=3$  independent replicates.

## 8. Outlook and general discussion

This thesis presented structural, biochemical and bioinformatic characterization of the enzymes that are involved in cyanophycin metabolism. Chapters 2 and 3 characterized cyanophycin synthetase 1 and showed how it is able to create cyanophycin primers and extend them using Asp and Arg. Chapter 4 discussed cyanophycin synthetase 2, and answered some questions about its activity and structure. Chapter 5 showed how cyanophycinase can hydrolyze cyanophycin into dipeptides, the essential first step in this polymers' degradation. Chapters 6 and 7 investigated isoaspartyl dipeptidases in the context of cyanophycin degradation, and showed that some bacteria have specialized cyanophycin-dipeptide degradation pathways. Together, these results answer many questions about the synthesis and biodegradation of this polymer, and strengthen the claim that it is a common material in many environments. The following discussion will contemplate remaining challenges and open questions in the field of cyanophycin metabolism.

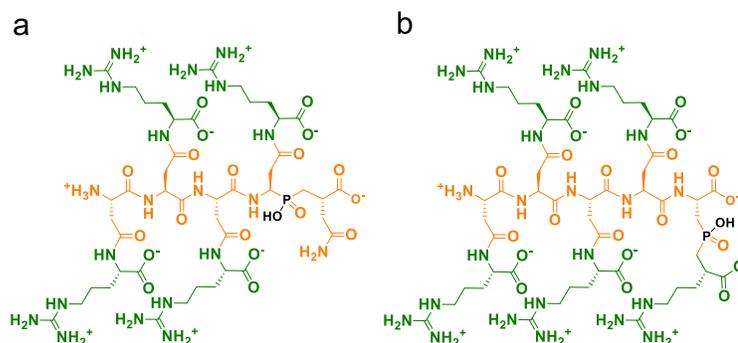
### **8.1 Cyanophycin synthetases**

#### **8.1.1. CphA1 co-complexes with Asp and Arg**

In chapter 2, co-complex structures of CphA1 with ATP and cyanophycin analogs are presented. Despite multiple attempts, we were unable to obtain maps that show unambiguous density for Asp and Arg, even when they were present in high concentrations (up to 20 mM). We encountered similar difficulties when attempting to obtain the CphA2- $\beta$ -Asp-Arg co-complex. This is not surprising, as structures of ATP-grasp and Mur-ligase enzymes with the free form of the equivalent substrates are very rare. These consistent difficulties suggest that they result from inherent features of these enzymes and their catalytic mechanism. Given the high flexibility of both the G and M domain active sites, it is likely that the binding of Asp and Arg involves unstable conformations that may only be favorable during intermediate steps of catalysis. For example, the formation of an acyl-phosphate intermediate in the G domain may be required to promote a G<sub>omega</sub> conformation that favors Asp binding. The resulting ternary complex may be too reactive to be captured in normal crystallography or cryo-EM experiments.

Transition-state substrate analogs may be able to solve this problem by stabilizing transient interactions and conformations (Fig. 8.1). For example, the substitution of an electrophilic carboxylate with a phosphate group can mimic a tetrahedral intermediate. Such analogs were

successfully used to gain insight into the activity of some ATP-grasp enzyme<sup>22,222,223</sup>. An example can be seen in the published structures of tubulin tyrosine ligase-like 6 (TTLL6) in complex with various substrate analogs<sup>22</sup>. These structures show the enzyme in complex with ADP and a phosphorylated intermediate analog. The phosphorylation occurred *in situ* during crystallization, and leads to the stabilization of the P-loop through multiple direct and indirect interactions with the intermediate.



**Figure 8.1.** Proposed substrate-analog intermediates for the reactions catalyzed by the G (a) and M (b) domains of CphA1. Residues in black show phosphinate moieties that mimic tetrahedral intermediates formed during catalysis.

However, the synthesis of such analogs for cyanophycin synthetases is not trivial and is complicated by factors such as stability and the need to use multiple orthogonal protecting groups. A second option, commonly used in the case of Mur-ligases<sup>224</sup>, is to identify inhibitors that bind in the enzyme's active sites. While these inhibitors do not necessarily reflect all the interactions formed by the native substrates, they can give insights into what an enzyme-substrate intermediate complex might look like<sup>225</sup>. As the development of potent and specific inhibitors for any enzyme is a lengthy and difficult process, synthesis or rationally-designed substrate analogs remains the preferred option.

### 8.1.2. The prospects of bioengineering CphA1

Multiple studies attempted to bioengineer CphA1. These attempts had two goals: to increase the enzyme's stability and thus activity<sup>88</sup>, and to change its substrate specificity<sup>76</sup>. The structures of CphA1 offer a good tool for the design of variants with increased stability. By comparing all three CphA1s described in chapter 2 it is easier to identify structurally conserved regions and parts of the enzyme that may be altered without directly affecting its catalytic activity. For example, the beneficial truncation described by Hai et al.<sup>88,108</sup> removed a small, non-conserved

portion of the enzyme's C-terminus. This part of the enzyme is probably flexible, as no density for it was observed in our structures. It is thus not surprising that this truncation led to increased thermal stability, and consequently increased activity *in vivo*.

On the other hand, bioengineering CphA1 to change its substrate specificity is much more challenging. The structures highlight the difficulties that any researcher attempting to do this will face. Both of the enzyme's active sites bind cyanophycin in distinct ways, and use multiple interactions to orient it for catalysis. In the G domain, ordered density is visible for 3 dipeptides, and another 3 are visible in the M domain. Thus, an attempt to mutate CphA1's substrate specificity from Arg to another residue will require changes in 6 peptidyl residue-binding positions, in addition to the Arg binding pocket. This may be feasible in the case of substituting Arg with another basic amino acid such as Lys or Orn. Some CphA1s can use these amino acids both *in vivo* and *in vitro*<sup>73</sup>. The fact that enzymes from different bacteria display different incorporation rates of non-Arg residues into cyanophycin suggests that the mutations required for such changes in substrate specificity are not unrealistically extensive.

On the other hand, mutation of CphA1 to use non-basic amino acids instead of Arg will probably require more extensive modifications to the enzyme. In addition to active site interactions, the favorable, non-specific interactions of cyanophycin with the charged patches in the N domain will need to be adjusted as well, as they considerably increase the enzyme's activity rate. Even with the recent advances in enzyme design and the available tools for protein bioengineering, this remains a non-trivial challenge.

### **8.1.3. CphA1's N domain activity is highly conserved**

The ability of CphA1 to generate its own primers is conserved in over 80% of these enzymes. Among the ~20% that do not have it, a large proportion belong to the order *Burkholderiales*. These bacteria often have two copies of CphA1, sometimes called CphA3 and CphA3'<sup>100</sup> or CphA and CphA'<sup>226</sup>. Interestingly, while *cphA3* encodes an active cyanophycin synthetase<sup>226</sup> and does not have an N domain active site, sequence alignment suggests that *cphA3'* has inactivating mutations in the G and M active sites but does have an active N domain. By having both CphA3 and CphA3', members of *Burkholderiales* are able to split the synthetic and hydrolytic activity of CphA1, potentially allowing them to better control the ratio of cyanophycin synthesis to primer generation. This also means that out of the ~20% of CphA1s that do not have active N

domains, a large number are coupled to enzymes that do. In effect, this further increases the percentage of CphA1s that are found with primer-generating N domains.

Why active N domains are so prevalent is not clear, as cyanophycin can be synthesized in their absence. Some bacteria that do not have known primer-generating enzymes, like *Acinetobacter baylyi* DSM587, are nevertheless capable of producing large amounts of cyanophycin<sup>26</sup>. In addition, as shown in chapter 2, considerable amounts of cyanophycin are produced in heterologous hosts even when using CphA1s with inactive N domains. It is possible that in certain native hosts, CphA1 activity in the absence of cyanophycin primers is high enough to reach the required synthesis levels. As previously noted<sup>104</sup>, some biomaterials can serve as primers in the absence of cyanophycin and may be sufficiently abundant in those bacteria. A possible explanation for the high conservation of active N domains is that while non-specific primers can be used by CphA1, this process is not efficient enough for most bacteria. The use of cyanophycin as primer likely leads to faster accumulation of polymer upon CphA1 expression. Unicellular cyanobacteria, for example, often transition between cyanophycin production and consumption in cycles of several hours, depending on light availability. Thus, a delay of several hours in cyanophycin production due to the lack of ideal primers may render this system irrelevant.

#### **8.1.4. CphA2-substrate co-complexes**

In chapter 4, I described our unsuccessful attempts to solve the structure of CphA2 in complex with cyanophycin or  $\beta$ -Asp-Arg. The difficulty in forming a stable CphA2 –  $\beta$ -Asp-Arg complex is analogous to that of CphA1, and may be addressed in similar ways. On the other hand, the CphA2 – cyanophycin co-complex is expected to be similar to the one formed by CphA1's G domain. If that is true, it is likely relatively stable and so should not pose as big of a problem as the CphA2 –  $\beta$ -Asp-Arg co-complex. As pointed out in chapter 4, the immediate problem we faced when attempting to visualize cyanophycin bound to CphA2 was the non-specific binding of tartrate molecules from the crystallization conditions, which probably prevented the binding of cyanophycin in the active site. As very specific crystallization conditions were required to achieve well-diffracting crystals, the tartrate concentration could not be lowered.

Identification of other crystallization conditions would not have necessarily solved this problem, as other factors likely contributed to the low binding of cyanophycin to CphA2 in the crystal. Both CphA1 and CphA2 are much less active in buffers that contain high salt concentrations. As shown in chapters 2 and 4, even 50 mM NaCl lead to reduced cyanophycin

synthetase activity. This is likely because electrostatic interactions are a major factor in the binding of cyanophycin to CphA1 and CphA2, and buffers with a high ionic strength interfere with these interactions. As many crystallization conditions have high concentrations of salts, they are not ideal for binding of cyanophycin to CphA1 or CphA2 and thus not suitable for co-complex formation. Furthermore, the cyanophycin segments that are needed for maximal binding to the enzyme are relatively large molecules, and may interfere with the packing in some crystal forms. For example, some of the structures of *Su*CphA1 in chapters 2 and 3 are with the large polymer segment ( $\beta$ -Asp-Arg)<sub>16</sub>. Examination of the crystal structure of *Tm*CphA1 shows that many of the solvent channels in this crystal form are not large enough to accommodate such a large cyanophycin segment.

As it was with CphA1, cryo-EM can be a good solution to this problem. This technique both enables better control over the buffer conditions that are used and keeps the enzyme molecules in solution so that they can bind large substrates. The main challenge faced in this regard is the size of CphA2. Most of the CphA2 enzymes that we characterized form dimers and, at <150 kDa, are somewhat small for cryo-EM data processing. Higher-order oligomers of this enzyme should be ideal candidates for cryo-EM studies, as they would combine size with high symmetry. There are signs that CphA2 can sometimes form such oligomers. Based on our SEC results, out of the nine homologs characterized in chapter 4, two form stable oligomers larger than dimers. Other CphA2s show a tendency to oligomerize as well: the triple mutant of *G. citriformis* that we described seems to form octamers, and the twinned crystal-structures that we were able to solve but not refine all show a similar hexameric architecture. By inducing oligomerization of these enzymes, for example through mutagenesis or optimization of buffer conditions, they may become suitable targets as well. CphA2 may thus be more suitable for cryo-EM studies than it appears, and this technique may be the right tool to visualize it in complex with its substrates.

## **8.2. Cyanophycinase**

### **8.2.1. Is cyanophycinase unique?**

With the research described in this thesis, two kinds of enzymes are now known to perform cyanophycin hydrolysis: CphB/E/I are exocyanophycinases that degrade cyanophycin to dipeptides, and active N domains of CphA1 are endocyanophycinase which cleave cyanophycin into short segments. Only exocyanophycinase activity has been linked to utilization of

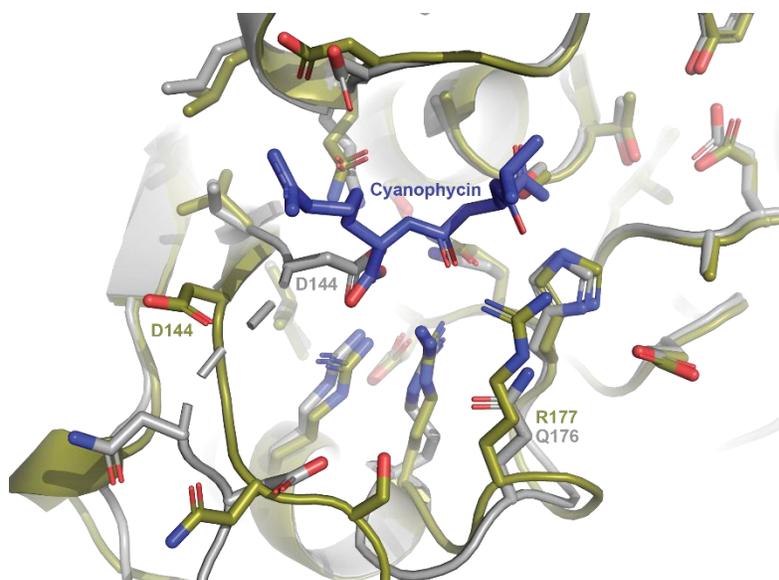
cyanophycin, making its presence a prerequisite for its catabolism. However, as the results in chapter 6 show, many bacteria have *cphA1* copies but no detectable cyanophycinase gene. Several possible explanations to this observation exist, two of which are more likely.

The first likely explanation is that there is high sequence variation among cyanophycinases, and some homologous enzymes were not detected. To avoid false-positive results, the bioinformatic searches in chapter 6 were limited to 30% identity and 70% coverage of several known enzymes. These parameters may be too conservative and miss some hits. However, the use of three different queries (*cphB*, *cphE* and *cphI*) should limit the extent of this problem. The second explanation is that genes with cyanophycinase-like activity exist that are unrelated to cyanophycinase. This would be analogous to the function of CphZ as an alternative to IadA and IaaA. One possible way to detect such unknown enzymes is by screening for cyanophycin degraders among bacteria that do not have known cyanophycinase genes. Multiple studies showed that detection of cyanophycin-degrading bacteria can be done fairly easily by growing isolates on cyanophycin-containing solid media<sup>57-59</sup>. Since cyanophycin is insoluble and scatters light at neutral pH, its degradation can be easily detected by the formation of transparent halos around colonies.

Theoretically, two other potential explanations for the mismatch in *cphA1* and cyanophycin distributions are possible. The first is the degradation of cyanophycin through metabolic pathways that do not require its cleavage to dipeptides first. For example, hydrolysis of only the Arg residues will result in biodegradable poly-Asp. Such a pathway, in addition to its biological significance, would be very desirable for biotechnological applications. However, no evidence exists for the formation of poly-Asp by bacteria. Another possible explanation is that some bacteria use cyanophycin in ways that do not require its degradation to amino acids. However, no such uses are currently known.

### **8.2.2. *PcCphB* as a test case for protein structure prediction**

The lack of measurable cyanophycinase activity by *PcCphB* is very surprising and provides a good example of how subtle and seemingly minor variations in sequence can lead to significant differences in protein folding and activity. As discussed in chapter 5, virtually all the residues that were identified as important for cyanophycinase substrate binding and catalytic activity are conserved in *PcCphB*. The main differences in sequence between it and *SyCphB* are in non-conserved residues that do not form part of the active site. Although we were unable to find



**Figure 8.2.** Overlay of the crystal structure (7UQV, gray) and AlphaFold model (A0A0L6JSP1, olive) of *PcCphB*. The position of cyanophycin in the active site of *SyCphB* (7UQW, blue) shows the clash with D144 and distance from Q176 in the crystal structure of *PcCphB*. In contrast, the AlphaFold model shows D144 and R177 in conformations that are similar to those seen in *SyCphB*.

mutations that restored *PcCphAB* activity, it seems likely that the conformational differences we identified between it and *SyCphB* are the cause for its loss of function. These conformational differences are probably the result of the observed sequence variations in the aforementioned non-conserved residues.

This protein also serves as an example for the importance of experimental structural biology in this new era of computational protein structure prediction. During and following the work on *PcCphB*, I modeled it using various methods to verify this protein's predicted identity as a cyanophycinase (before solving its structure) and for comparison with the structure once it has been solved. Surprisingly, despite using state of the art structure prediction programs such as AlphaFold<sup>227</sup> and RoseTTAFold<sup>228</sup>, none of the models I generated predicted the conformations we observed in the crystal structures. Based on the models, *PcCphB* should be an active protein with a fold similar to that of *SyCphB*. In fact, the current model of *PcCphB* available in the AlphaFold database (entry A0A0L6JSP1) differs from the crystal structure in an analogous way (Fig. 8.2). Without determining the protein's structure, the causes for its inactivity would have remained a mystery.

### **8.3. Isoaspartyl dipeptidases**

#### **8.3.1. Unknown isoaspartyl dipeptidases likely exist**

As already noted, the only known pathway of cyanophycin degradation ends with the hydrolysis of  $\beta$ -Asp-Arg dipeptides into Asp and Arg. Thus, it is expected that all cyanophycin-producing bacteria should have an enzyme that can catalyze this reaction. However, the results presented in chapters 6 and 7 show that only 77% of the genomes that have *cphA1* also have at least one of *iadA*, *iaaA* and *cphZ*. This number is even lower when considering cyanophycinase-containing genomes – only ~65% of these have a known isoaspartyl dipeptidase. The simplest explanation to this observation is that there are other, yet uncharacterized enzymes that serve as isoaspartyl dipeptidases in bacteria that do not have *IadA*, *IaaA* or *CphZ*. This would not be surprising, as in the analysis described in chapter 6 we identified several other putative isoaspartyl dipeptidases clustered with *cphA1* and cyanophycinase. Moreover, relying on clustering to detect these genes is very likely to miss potential candidates, as our results show that isoaspartyl dipeptidases normally do not cluster with other cyanophycin metabolizing genes.

Another possible explanation is the existence of other pathways for  $\beta$ -Asp-Arg degradation. For example, the AST pathway for arginine catabolism<sup>214</sup> seems like it could be adjusted for  $\beta$ -Asp-Arg degradation fairly easily. This pathway primes Arg for degradation by coupling it with a succinyl molecule through the Arg backbone nitrogen. As pointed out in chapter 7, the resulting molecule – N(2)-succinyl-arginine – is very similar to  $\beta$ -Asp-Arg. The only difference between the two molecules is the presence of the backbone nitrogen of the Asp residue in  $\beta$ -Asp-Arg, which also leads to the formation of another chiral center in this molecule. However, this relatively modest difference in substrate may not be so hard to overcome. It is conceivable that some variants of the AST pathway can accept dipeptides as substrate and degrade them in a similar way to N(2)-succinyl-arginine, resulting in L-glutamate and L-aspartate as its final products. This is also a possible explanation for the results observed in the *in vivo* studies of cyanophycin utilization by *P. aeruginosa* presented in chapter 7. The bacteria can use the dipeptides as a nitrogen source even when *aotO* is disrupted, albeit with lower efficiency compared to the WT strain. This suggests that they have other, less efficient mechanisms for the utilization of the nitrogen stored in these dipeptides. As with cyanophycinase, other theoretical explanations

are the degradation of cyanophycin that does not form  $\beta$ -Asp-Arg, or unknown uses of this polymer.

#### **8.4. Concluding remarks**

Cyanophycin is a very common biomaterial. Perhaps because of its name or for historical reasons, much of the research conducted on it has focused on cyanobacteria. However, in recent years there is an increasing realization that cyanophycin and the genes that metabolize it are widespread throughout the bacterial kingdom. In chapter 2 I present a phylogenetic analysis of CphA1 sequences, and show that <20% of them are found in cyanobacteria. This means that the majority of the cyanophycin world exists outside of this phylum. I think it is very likely that as we go deeper into this largely-unexplored world, we will discover more and more unexpected pathways and cellular functions that involve cyanophycin.

## References

1. Berg, H. et al. Biosynthesis of the cyanobacterial reserve polymer multi-L-arginyl-poly-L-aspartic acid (cyanophycin): mechanism of the cyanophycin synthetase reaction studied with synthetic primers. *Eur J Biochem* **267**, 5561–70 (2000).
2. Sharon, I. et al. Structures and function of the amino acid polymerase cyanophycin synthetase. *Nat Chem Biol* **17**, 1101–1110 (2021).
3. Krehenbrink, M. & Steinbuchel, A. Partial purification and characterization of a non-cyanobacterial cyanophycin synthetase from *Acinetobacter calcoaceticus* strain ADP1 with regard to substrate specificity, substrate affinity and binding to cyanophycin. *Microbiology* **150**, 2599–608 (2004).
4. Marti-Arbona, R. et al. Mechanism of the reaction catalyzed by isoaspartyl dipeptidase from *Escherichia coli*. *Biochemistry* **44**, 7115–24 (2005).
5. Bartlett, G.J. & Woolfson, D.N. On the satisfaction of backbone-carbonyl lone pairs of electrons in protein structures. *Protein Sci* **25**, 887-97 (2016).
6. Ashkenazy, H. et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* **44**, W344-50 (2016).
7. Stout, J., De Vos, D., Vergauwen, B. & Savvides, S.N. Glutathione biosynthesis in bacteria by bifunctional GshF is driven by a modular structure featuring a novel hybrid ATP-grasp fold. *Journal of molecular biology* **416**, 486–494 (2012).
8. Crooks, G.E., Hon, G., Chandonia, J.M. & Brenner, S.E. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188–90 (2004).
9. Basavannacharya, C., Robertson, G., Munshi, T., Keep, N.H. & Bhakta, S. ATP-dependent MurE ligase in *Mycobacterium tuberculosis*: biochemical and structural characterisation. *Tuberculosis (Edinb)* **90**, 16–24 (2010).
10. Michalska, K., Brzezinski, K. & Jaskolski, M. Crystal structure of isoaspartyl aminopeptidase in complex with L-aspartate. *J Biol Chem* **280**, 28484-91 (2005).
11. Thompson, J.D., Higgins, D.G. & Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–80 (1994).
12. Klemke, F. et al. CphA2 is a novel type of cyanophycin synthetase in N<sub>2</sub>-fixing cyanobacteria. *Microbiology* **162**, 526–36 (2016).
13. Sharon, I., Grogg, M., Hilvert, D. & Schmeing, T.M. Structure and function of the  $\beta$ -Asp-Arg polymerase cyanophycin synthetase 2. *ACS Chem Biol* **17**(2022).
14. Borzi, A. *Le comunicazioni intracellulari delle nostochinee*, (Messina, 1886).
15. Simon, R.D. Cyanophycin granules from the blue-green alga *Anabaena cylindrica*: a reserve material consisting of copolymers of aspartic acid and arginine. *Proc Natl Acad Sci U S A* **68**, 265–7 (1971).
16. Simon, R.D. & Weathers, P. Determination of the structure of the novel polypeptide containing aspartic acid and arginine which is found in cyanobacteria. *Biochim Biophys Acta* **420**, 165–76 (1976).
17. Simon, R.D. The effect of chloramphenicol on the production of cyanophycin granule polypeptide in the blue green alga *Anabaena cylindrica*. *Arch Mikrobiol* **92**, 115-22 (1973).
18. Takehara, M., Saimura, M., Inaba, H. & Hirohara, H. Poly( $\gamma$ -l-diaminobutanoic acid), a novel poly(amino acid), coproduced with poly( $\epsilon$ -l-lysine) by two strains of *Streptomyces celluloflavus*. *FEMS Microbiology Letters* **286**, 110-117 (2008).

19. Xu, Z. et al. Systematic unravelling of the biosynthesis of poly (L-diaminopropionic acid) in *Streptomyces albulus* PD-1. *Scientific Reports* **5**, 17400 (2015).
20. Kunioka, M. Biosynthesis and chemical reactions of poly(amino acid)s from microorganisms. *Applied Microbiology and Biotechnology* **47**, 469-475 (1997).
21. Bajaj, I. & Singhal, R. Poly (glutamic acid) – An emerging biopolymer of commercial interest. *Bioresource Technology* **102**, 5551-5561 (2011).
22. Mahalingan, K.K. et al. Structural basis for polyglutamate chain initiation and elongation by TLL family enzymes. *Nat Struct Mol Biol* **27**, 802-813 (2020).
23. Yamanaka, K., Maruyama, C., Takagi, H. & Hamano, Y.  $\epsilon$ -Poly-L-lysine dispersity is controlled by a highly unusual nonribosomal peptide synthetase. *Nature Chemical Biology* **4**, 766-772 (2008).
24. Mariotti, F., Tome, D. & Mirand, P.P. Converting nitrogen into protein--beyond 6.25 and Jones' factors. *Crit Rev Food Sci Nutr* **48**, 177–84 (2008).
25. Simon, R.D., Lawry, N.H. & McLendon, G.L. Structural characterization of the cyanophycin granule polypeptide of *Anabaena cylindrica* by circular dichroism and Raman spectroscopy. *Biochim Biophys Acta* **626**, 277–81 (1980).
26. Elbahloul, Y., Krehenbrink, M., Reichelt, R. & Steinbuchel, A. Physiological conditions conducive to high cyanophycin content in biomass of *Acinetobacter calcoaceticus* strain ADP1. *Appl Environ Microbiol* **71**, 858–66 (2005).
27. Ingram, L.O., Thurston, E.L. & Van Baalen, C. Effects of selected inhibitors on growth and cell division in *Agmenellum*. *Arch Mikrobiol* **81**, 1-12 (1972).
28. Kolodny, N.H. et al. Effect of nitrogen source on cyanophycin synthesis in *Synechocystis* sp. strain PCC 6308. *J Bacteriol* **188**, 934–40 (2006).
29. Allen, M.M., Hutchison, F. & Weathers, P.J. Cyanophycin granule polypeptide formation and degradation in the cyanobacterium *Aphanocapsa* 6308. *J Bacteriol* **141**, 687–93 (1980).
30. Stevens, S.E. & Paone, D.A. Accumulation of Cyanophycin Granules as a Result of Phosphate Limitation in *Agmenellum quadruplicatum*. *Plant Physiol* **67**, 716–9 (1981).
31. Mackerras, A.H., de Chazal, N.M. & Smith, G.D. Transient accumulations of cyanophycin in *Anabaena cylindrica* and *Synechocystis* 6308. *Microbiology* **136**, 2057–2065 (1990).
32. Page-Sharp, M., Behm, C.A. & Smith, G.D. Cyanophycin and glycogen synthesis in a cyanobacterial *Scytonema* species in response to salt stress. *FEMS Microbiology Letters* **160**, 11-15 (1998).
33. Yalcin, Y.S., Aydin, B.N., Sayadujjehara, M. & Sittler, V. Antibiotic-Induced Changes in Pigment Accumulation, Photosystem II, and Membrane Permeability in a Model Cyanobacterium. *Front Microbiol* **13**, 930357 (2022).
34. Carr, N.G. Nitrogen reserves and dynamic reservoirs in cyanobacteria. *Biochemistry of the Algae and Cyanobacteria Proceedings of the Phytochemical Society of Europe, No* **28**, 14-21 (1988).
35. Whitton, B.A. & Potts, M. Introduction to the Cyanobacteria. in *Ecology of Cyanobacteria II: Their Diversity in Space and Time* (ed. Whitton, B.A.) 1-13 (Springer Netherlands, Dordrecht, 2012).
36. Einsle, O. & Rees, D.C. Structural Enzymology of Nitrogenase Enzymes. *Chem Rev* **120**, 4969-5004 (2020).

37. Gallon, J.R. Reconciling the incompatible: N<sub>2</sub> fixation And O<sub>2</sub>. *New Phytologist* **122**, 571-609 (1992).
38. Popa, R. et al. Carbon and nitrogen fixation and metabolite exchange in and between individual cells of *Anabaena oscillarioides*. *ISME J* **1**, 354–60 (2007).
39. Sherman, D.M., Tucker, D. & Sherman, L.A. Heterocysts development and localization of cyanophycin in N<sub>2</sub>-fixing cultures of *Anabaena* sp. PCC7120 (cyanobacteria). *Journal of Phycology* **36**, 932-941 (2000).
40. Burnat, M., Herrero, A. & Flores, E. Compartmentalized cyanophycin metabolism in the diazotrophic filaments of a heterocyst-forming cyanobacterium. *Proc Natl Acad Sci U S A* **111**, 3823–8 (2014).
41. Zhang, H. & Yang, C. Arginine and nitrogen mobilization in cyanobacteria. *Molecular Microbiology* **111**, 863-867 (2019).
42. Li, H., Sherman, D.M., Bao, S. & Sherman, L.A. Pattern of cyanophycin accumulation in nitrogen-fixing and non-nitrogen-fixing cyanobacteria. *Arch Microbiol* **176**, 9–18 (2001).
43. Finzi-Hart, J.A. et al. Fixation and fate of C and N in the cyanobacterium *Trichodesmium* using nanometer-scale secondary ion mass spectrometry. *Proceedings of the National Academy of Sciences* **106**, 6345-6350 (2009).
44. Mackerras, A.H., Youens, B.N., Weir, R.C. & Smith, G.D. Is cyanophycin involved in the integration of nitrogen and carbon metabolism in the cyanobacteria *Anabaena cylindrica* and *Gleotheca* grown on light/dark cycles? *Microbiology* **136**, 2049-2056 (1990).
45. Harke, M.J. et al. A review of the global ecology, genomics, and biogeography of the toxic cyanobacterium, *Microcystis* spp. *Harmful Algae* **54**, 4-20 (2016).
46. Paerl, H.W. & Otten, T.G. Harmful Cyanobacterial Blooms: Causes, Consequences, and Controls. *Microbial Ecology* **65**, 995-1010 (2013).
47. Bláha, L., Babica, P. & Maršálek, B. Toxins produced in cyanobacterial water blooms - toxicity and risks. *Interdisciplinary Toxicology* **2**, 36-41 (2009).
48. Coffey, M.M. et al. Assessing cyanobacterial frequency and abundance at surface waters near drinking water intakes across the United States. *Water Research* **201**, 117377 (2021).
49. Hampel, J.J. et al. Ammonium recycling supports toxic *Planktothrix* blooms in Sandusky Bay, Lake Erie: Evidence from stable isotope and metatranscriptome data. *Harmful Algae* **81**, 42–52 (2019).
50. Lu, Z. et al. Cyanophycin accumulated under nitrogen-fluctuating and high-nitrogen conditions facilitates the persistent dominance and blooms of *Raphidiopsis raciborskii* in tropical waters. *Water Res* **214**, 118215 (2022).
51. Adams, D.G. & Duggan, P.S. Tansley Review No. 107. Heterocyst and akinete differentiation in cyanobacteria. *New Phytologist* **144**, 3-33 (1999).
52. Sukenik, A. et al. Carbon assimilation and accumulation of cyanophycin during the development of dormant cells (akinetes) in the cyanobacterium *Aphanizomenon ovalisporum*. *Frontiers in Microbiology* **6**(2015).
53. Garg, R. et al. Changes in Envelope Structure and Cell–Cell Communication during Akinete Differentiation and Germination in Filamentous Cyanobacterium *Trichormus variabilis* ATCC 29413. *Life* **12**, 429 (2022).

54. Liu, H., Ray, W.K., Helm, R.F., Popham, D.L. & Melville, S.B. Analysis of the Spore Membrane Proteome in *Clostridium perfringens* Implicates Cyanophycin in Spore Assembly. *J Bacteriol* **198**, 1773–1782 (2016).
55. Elbahloul, Y. & Steinbuchel, A. Engineering the genotype of *Acinetobacter* sp. strain ADP1 to enhance biosynthesis of cyanophycin. *Appl Environ Microbiol* **72**, 1410–9 (2006).
56. Radzicka, A. & Wolfenden, R. Rates of Uncatalyzed Peptide Bond Hydrolysis in Neutral Solution and the Transition State Affinities of Proteases. *Journal of the American Chemical Society* **118**, 6105-6109 (1996).
57. Obst, M., Sallam, A., Luftmann, H. & Steinbuchel, A. Isolation and characterization of gram-positive cyanophycin-degrading bacteria-kinetic studies on cyanophycin depolymerase activity in aerobic bacteria. *Biomacromolecules* **5**, 153–61 (2004).
58. Obst, M., Krug, A., Luftmann, H. & Steinbuchel, A. Degradation of cyanophycin by *Sedimentibacter hongkongensis* strain KI and *Citrobacter amalonaticus* strain G Isolated from an anaerobic bacterial consortium. *Appl Environ Microbiol* **71**, 3642–52 (2005).
59. Sallam, A. & Steinbuchel, A. Anaerobic and aerobic degradation of cyanophycin by the denitrifying bacterium *Pseudomonas alcaligenes* strain DIP1 and role of three other coisolates in a mixed bacterial consortium. *Appl Environ Microbiol* **74**, 3434–43 (2008).
60. Sallam, A. & Steinbuchel, A. Cyanophycin-degrading bacteria in digestive tracts of mammals, birds and fish and consequences for possible applications of cyanophycin and its dipeptides in nutrition and therapy. *J Appl Microbiol* **107**, 474–84 (2009).
61. Obst, M., Oppermann-Sanio, F.B., Luftmann, H. & Steinbuchel, A. Isolation of cyanophycin-degrading bacteria, cloning and characterization of an extracellular cyanophycinase gene (cphE) from *Pseudomonas anguilliseptica* strain BI. The cphE gene from *P. anguilliseptica* BI encodes a cyanophycinhydrolyzing enzyme. *J Biol Chem* **277**, 25096–105 (2002).
62. Ben Hania, W. et al. Characterization of the first cultured representative of a Bacteroidetes clade specialized on the scavenging of cyanobacteria. *Environmental Microbiology* **19**, 1134-1148 (2017).
63. Johnson, M. et al. NCBI BLAST: a better web interface. *Nucleic Acids Res* **36**, W5–9 (2008).
64. Tseng, W.C., Fang, T.Y., Lin, Y.C., Huang, S.J. & Huang, Y.H. Reversible Self-Assembly Nanovesicle of UCST Response Prepared with Multi-L-arginyl-poly-L-aspartate Conjugated with Polyethylene Glycol. *Biomacromolecules* **19**, 4585–4592 (2018).
65. Grogg, M. et al. Cell penetration, herbicidal activity, and *in-vivo*-toxicity of oligo-arginine derivatives and of novel guanidinium-rich compounds derived from the biopolymer cyanophycin. *Helv Chim Acta* **101**, e1800112 (2018).
66. Uddin, Z., Fang, T.Y., Siao, J.Y. & Tseng, W.C. Wound healing attributes of polyelectrolyte multilayers prepared with multi-L-arginyl-poly-L-aspartate pairing with hyaluronic acid and gamma-polyglutamic acid. *Macromol Biosci* **20**, e2000132 (2020).
67. Zou, K. et al. Cyanophycin Granule Polypeptide: a Neglected High Value-Added Biopolymer, Synthesized in Activated Sludge on a Large Scale. *Appl Environ Microbiol*, e0074222 (2022).
68. Sallam, A. & Steinbuchel, A. Dipeptides in nutrition and therapy: cyanophycin-derived dipeptides as natural alternatives and their biotechnological production. *Appl Microbiol Biotechnol* **87**, 815–28 (2010).

69. Krzemińska, A., Kwiatos, N., Arenhart Soares, F. & Steinbüchel, A. Theoretical Studies of Cyanophycin Dipeptides as Inhibitors of Tyrosinases. *International Journal of Molecular Sciences* **23**, 3335 (2022).
70. Adelnia, H., Tran, H.D.N., Little, P.J., Blakey, I. & Ta, H.T. Poly(aspartic acid) in Biomedical Applications: From Polymerization, Modification, Properties, Degradation, and Biocompatibility to Applications. *ACS Biomater Sci Eng* **7**, 2083-2105 (2021).
71. Adelnia, H., Blakey, I., Little, P.J. & Ta, H.T. Hydrogels Based on Poly(aspartic acid): Synthesis and Applications. *Front Chem* **7**, 755 (2019).
72. Hasson, D., Shemer, H. & Sher, A. State of the Art of Friendly “Green” Scale Control Inhibitors: A Review Article. *Industrial & Engineering Chemistry Research* **50**, 7601–7607 (2011).
73. Frommeyer, M., Wiefel, L. & Steinbüchel, A. Features of the biotechnologically relevant polyamide family "cyanophycins" and their biosynthesis in prokaryotes and eukaryotes. *Crit Rev Biotechnol* **36**, 153–64 (2016).
74. Krehenbrink, M., Oppermann-Sanio, F.-B. & Steinbüchel, A. Evaluation of non-cyanobacterial genome sequences for occurrence of genes encoding proteins homologous to cyanophycin synthetase and cloning of an active cyanophycin synthetase from *Acinetobacter sp.* strain DSM 587. *Arch Microbiol* **177**, 371–80 (2002).
75. Aboulmagd, E., Voss, I., Oppermann-Sanio, F.B. & Steinbüchel, A. Heterologous expression of cyanophycin synthetase and cyanophycin synthesis in the industrial relevant bacteria *Corynebacterium glutamicum* and *Ralstonia eutropha* and in *Pseudomonas putida*. *Biomacromolecules* **2**, 1338–42 (2001).
76. Wördemann, R., Wiefel, L., Wendisch, V.F. & Steinbüchel, A. Incorporation of alternative amino acids into cyanophycin by different cyanophycin synthetases heterologously expressed in *Corynebacterium glutamicum*. *AMB Express* **11**, 55 (2021).
77. Voss, I., Diniz, S.C., Aboulmagd, E. & Steinbüchel, A. Identification of the *Anabaena sp.* strain PCC7120 cyanophycin synthetase as suitable enzyme for production of cyanophycin in gram-negative bacteria like *Pseudomonas putida* and *Ralstonia eutropha*. *Biomacromolecules* **5**, 1588–95 (2004).
78. Abd-El-Karem, Y., Elbers, T., Reichelt, R. & Steinbüchel, A. Heterologous expression of *Anabaena sp.* PCC7120 cyanophycin metabolism genes *cphA1* and *cphB1* in *Sinorhizobium (Ensifer) meliloti* 1021. *Appl Microbiol Biotechnol* **89**, 1177–92 (2011).
79. Steinle, A., Bergander, K. & Steinbüchel, A. Metabolic engineering of *Saccharomyces cerevisiae* for production of novel cyanophycins with an extended range of constituent amino acids. *Appl Environ Microbiol* **75**, 3437–46 (2009).
80. Steinle, A., Witthoff, S., Krause, J.P. & Steinbüchel, A. Establishment of cyanophycin biosynthesis in *Pichia pastoris* and optimization by use of engineered cyanophycin synthetases. *Appl Environ Microbiol* **76**, 1062–70 (2010).
81. Meussen, B.J., Weusthuis, R.A., Sanders, J.P.M. & Graaff, L.H.d. Production of cyanophycin in *Rhizopus oryzae* through the expression of a cyanophycin synthetase encoding gene. *Applied microbiology and biotechnology* **93**, 1167–74 (2012).
82. Nausch, H. et al. Tobacco as platform for a commercial production of cyanophycin. *N Biotechnol* **33**, 842–851 (2016).
83. Huckauf, J. et al. Sustainable Production of the Cyanophycin Biopolymer in Tobacco in the Greenhouse and Field. *Front Bioeng Biotechnol* **10**, 896863 (2022).

84. Huhns, M. et al. Tuber-specific *cphA* expression to enhance cyanophycin production in potatoes. *Plant Biotechnol J* **7**, 883–98 (2009).
85. Allen, M.M. & Weathers, P.J. Structure and composition of cyanophycin granules in the cyanobacterium *Aphanocapsa* 6308. *J Bacteriol* **141**, 959–62 (1980).
86. Hai, T., Oppermann-Sanio, F.B. & Steinbuechel, A. Purification and characterization of cyanophycin and cyanophycin synthetase from the thermophilic *Synechococcus* sp. MA19. *FEMS Microbiol Lett* **181**, 229–36 (1999).
87. Aboulmagd, E., Oppermann-Sanio, F.B. & Steinbuechel, A. Purification of *Synechocystis* sp. strain PCC6308 cyanophycin synthetase and its characterization with respect to substrate and primer specificity. *Applied and environmental microbiology* **67**, 2176–82 (2001).
88. Hai, T., Frey, K.M. & Steinbuechel, A. Engineered cyanophycin synthetase (CphA) from *Nostoc ellipsosporum* confers enhanced CphA activity and cyanophycin accumulation to *Escherichia coli*. *Applied and environmental microbiology* **72**, 7652–60 (2006).
89. Elbahloul, Y., Frey, K., Sanders, J. & Steinbüchel, A. Protamylase, a Residual Compound of Industrial Starch Production, Provides a Suitable Medium for Large-Scale Cyanophycin Production. *Applied and Environmental Microbiology* **71**, 7759-7767 (2005).
90. Sharon, I. et al. A cryptic third active site in cyanophycin synthetase creates primers it uses for polymerization. *Nat Commun (submitted)* (2022).
91. Hai, T., Ahlers, H., Gorenflo, V. & Steinbüchel, A. Axenic cultivation of anoxygenic phototrophic bacteria, cyanobacteria, and microalgae in a new closed tubular glass photobioreactor. *Applied Microbiology and Biotechnology* **53**, 383-389 (2000).
92. Zhang, Y., Kumar, A., Vadlani, P.V. & Narayanan, S. Production of nitrogen-based platform chemical: cyanophycin biosynthesis using recombinant *Escherichia coli* and renewable media substitutes. *Journal of Chemical Technology & Biotechnology* **88**, 1321-1327 (2013).
93. Huhns, M. et al. Plastid targeting strategies for cyanophycin synthetase to achieve high-level polymer accumulation in *Nicotiana tabacum*. *Plant biotechnology journal* **6**, 321–36 (2008).
94. Du, J. et al. Isolation and characterization of a novel cyanophycin synthetase from a deep-sea sediment metagenomic library. *Appl Microbiol Biotechnol* **97**, 8619–28 (2013).
95. Wiefel, L., Broker, A. & Steinbuechel, A. Synthesis of a citrulline-rich cyanophycin by use of *Pseudomonas putida* ATCC 4359. *Appl Microbiol Biotechnol* **90**, 1755–62 (2011).
96. Frommeyer, M. & Steinbuechel, A. Increased lysine content is the main characteristic of the soluble form of the polyamide cyanophycin synthesized by recombinant *Escherichia coli*. *Appl Environ Microbiol* **79**, 4474–83 (2013).
97. Wiefel, L. & Steinbuechel, A. Solubility behavior of cyanophycin depending on lysine content. *Appl Environ Microbiol* **80**, 1091–6 (2014).
98. Ziegler, K. et al. Molecular characterization of cyanophycin synthetase, the enzyme catalyzing the biosynthesis of the cyanobacterial reserve material multi-L-arginyl-poly-L-aspartate (cyanophycin). *Eur J Biochem* **254**, 154–9 (1998).
99. Simon, R.D. The biosynthesis of multi-L-arginyl-poly(L-aspartic acid) in the filamentous cyanobacterium *Anabaena cylindrica*. *Biochim Biophys Acta* **422**, 407–18 (1976).

100. Fuser, G. & Steinbüchel, A. Analysis of genome sequences for genes of cyanophycin metabolism: identifying putative cyanophycin metabolizing prokaryotes. *Macromol Biosci* **7**, 278–96 (2007).
101. Page, M.J. & Cera, E.D. Role of Na<sup>+</sup> and K<sup>+</sup> in Enzyme Function. *Physiological Reviews* **86**, 1049-1092 (2006).
102. Watzer, B. & Forchhammer, K. Cyanophycin synthesis optimizes nitrogen utilization in the unicellular cyanobacterium *Synechocystis sp.* PCC 6803. *Appl Environ Microbiol* (2018).
103. Liebergesell, M., Sonomoto, K., Madkour, M., Mayer, F. & Steinbüchel, A. Purification and Characterization of the Poly(Hydroxyalkanoic Acid) Synthase from *Chromatium vinosum* and Localization of the Enzyme at the Surface of Poly(Hydroxyalkanoic Acid) Granules. *European Journal of Biochemistry* **226**, 71-80 (1994).
104. Hai, T., Oppermann-Sanio, F.B. & Steinbüchel, A. Molecular characterization of a thermostable cyanophycin synthetase from the thermophilic cyanobacterium *Synechococcus sp.* strain MA19 and in vitro synthesis of cyanophycin and related polyamides. *Appl Environ Microbiol* **68**, 93–101 (2002).
105. Arai, T. & Kino, K. A cyanophycin synthetase from *Thermosynechococcus elongatus* BP-1 catalyzes primer-independent cyanophycin synthesis. *Appl Microbiol Biotechnol* **81**, 69–78 (2008).
106. Aboulmagd, E., Oppermann-Sanio, F.B. & Steinbüchel, A. Molecular characterization of the cyanophycin synthetase from *Synechocystis sp.* strain PCC6308. *Arch Microbiol* **174**, 297–306 (2000).
107. Ziegler, K., Deutzmann, R. & Lockau, W. Cyanophycin synthetase-like enzymes of non-cyanobacterial eubacteria: characterization of the polymer produced by a recombinant synthetase of *Desulfitobacterium hafniense*. *Zeitschrift fur Naturforschung C, Journal of biosciences* **57**, 522–9 (2002).
108. Hai, T., Lee, J.-S., Kim, T.-J. & Suh, J.-W. The role of the C-terminal region of cyanophycin synthetase from *Nostoc ellipsosporum* NE1 in its enzymatic activity and thermostability: a key function of Glu(856). *Biochimica et biophysica acta* **1794**, 42–9 (2009).
109. Chou, C.Y. & Tong, L. Structural and biochemical studies on the regulation of biotin carboxylase by substrate inhibition and dimerization. *J Biol Chem* **286**, 24417-25 (2011).
110. Smith, C.A. Structure, function and dynamics in the mur family of bacterial cell wall ligases. *J Mol Biol* **362**, 640–55 (2006).
111. Glieder, A. Protein Engineering Handbook. Edited by Stefan Lutz and Uwe T. Bornscheuer. *ChemBioChem* **10**, 2111-2112 (2009).
112. Fawaz, M.V., Topper, M.E. & Firestone, S.M. The ATP-grasp enzymes. *Bioorg Chem* **39**, 185–91 (2011).
113. Ye, D. et al. Investigation of the Catalytic Site within the ATP-Grasp Domain of *Clostridium symbiosum* Pyruvate Phosphate Dikinase\*. *Journal of Biological Chemistry* **276**, 37630-37639 (2001).
114. Miller, G.J., Wilson, M.P., Majerus, P.W. & Hurley, J.H. Specificity Determinants in Inositol Polyphosphate Synthesis: Crystal Structure of Inositol 1,3,4-Trisphosphate 5/6-Kinase. *Molecular Cell* **18**, 201-212 (2005).

115. Wang, W., Kappock, T.J., Stubbe, J. & Ealick, S.E. X-ray Crystal Structure of Glycinamide Ribonucleotide Synthetase from *Escherichia coli*. *Biochemistry* **37**, 15647-15662 (1998).
116. Hibi, T. et al. Structure of the multifunctional loops in the nonclassical ATP-binding fold of glutathione synthetase. *Nat Struct Biol* **3**, 16–8 (1996).
117. Galant, A., Arkus, K.A., Zubieta, C., Cahoon, R.E. & Jez, J.M. Structural basis for evolution of product diversity in soybean glutathione biosynthesis. *Plant Cell* **21**, 3450–8 (2009).
118. Pederick, J.L., Thompson, A.P., Bell, S.G. & Bruning, J.B. d-alanine-d-alanine ligase as a model for the activation of ATP-grasp enzymes by monovalent cations. *Journal of Biological Chemistry* **295**, 7894-7904 (2020).
119. Picossi, S., Valladares, A., Flores, E. & Herrero, A. Nitrogen-regulated genes for the metabolism of cyanophycin, a bacterial nitrogen reserve polymer: expression and mutational analysis of two cyanophycin synthetase and cyanophycinase gene clusters in heterocyst-forming cyanobacterium *Anabaena* sp. PCC 7120. *The Journal of biological chemistry* **279**, 11582–92 (2004).
120. Forchhammer, K. & Watzler, B. Microbiology Comment. *Microbiology* **162**, 727-729 (2016).
121. Richter, R., Hejazi, M., Kraft, R., Ziegler, K. & Lockau, W. Cyanophycinase, a peptidase degrading the cyanobacterial reserve material multi-L-arginyl-poly-L-aspartic acid (cyanophycin): molecular cloning of the gene of *Synechocystis* sp. PCC 6803, expression in *Escherichia coli*, and biochemical characterization of the purified enzyme. *Eur J Biochem* **263**, 163–9 (1999).
122. Lawry, N.H. & Simon, R.D. The normal and induced occurrence of cyanophycin inclusion bodies in several blue-green algae<sup>1</sup>. *Journal of Phycology* **18**, 391–399 (1982).
123. Huguenin-Dezot, N. et al. Trapping biosynthetic acyl-enzyme intermediates with encoded 2,3-diaminopropionic acid. *Nature* **565**, 112–117 (2019).
124. Law, A.M., Lai, S.W., Tavares, J. & Kimber, M.S. The structural basis of beta-peptide-specific cleavage by the serine protease cyanophycinase. *J Mol Biol* **392**, 393–404 (2009).
125. Aswad, D.W., Paranandi, M.V. & Schurter, B.T. Isoaspartate in peptides and proteins: formation, significance, and analysis. *J Pharm Biomed Anal* **21**, 1129–36 (2000).
126. Lowenson, J.D., Kim, E., Young, S.G. & Clarke, S. Limited accumulation of damaged proteins in l-isoaspartyl (D-aspartyl) O-methyltransferase-deficient mice. *J Biol Chem* **276**, 20695–702 (2001).
127. Kim, E., Lowenson, J.D., MacLaren, D.C., Clarke, S. & Young, S.G. Deficiency of a protein-repair enzyme results in the accumulation of altered proteins, retardation of growth, and fatal seizures in mice. *Proc Natl Acad Sci U S A* **94**, 6132–7 (1997).
128. Gary, J.D. & Clarke, S. Purification and characterization of an isoaspartyl dipeptidase from *Escherichia coli*. *J Biol Chem* **270**, 4076–87 (1995).
129. Hejazi, M. et al. Isoaspartyl dipeptidase activity of plant-type asparaginases. *Biochem J* **364**, 129–36 (2002).
130. Liotenberg, S., Campbell, D., Rippka, R., Houmard, J. & de Marsac, N.T. Effect of the nitrogen source on phycobiliprotein synthesis and cell reserves in a chromatically adapting filamentous cyanobacterium. *Microbiology* **142**, 611–622 (1996).

131. Fay, P. Oxygen relations of nitrogen fixation in cyanobacteria. *Microbiol Rev* **56**, 340–73 (1992).
132. Kumar, K., Mella-Herrera, R.A. & Golden, J.W. Cyanobacterial heterocysts. *Cold Spring Harb Perspect Biol* **2**, a000315 (2010).
133. Liang, B. et al. Cyanophycin mediates the accumulation and storage of fixed carbon in non-heterocystous filamentous cyanobacteria from coniform mats. *PLoS One* **9**, e88142 (2014).
134. Wingard, L.L. et al. Cyanophycin production in a phycoerythrin-containing marine *synechococcus* strain of unusual phylogenetic affinity. *Appl Environ Microbiol* **68**, 1772–7 (2002).
135. Gorelova, O.A. & Kleimenov, S. The accumulation and degradation dynamics of cyanophycin in cyanobacteria grown in symbiotic associations with plant tissues and cells. *Mikrobiologiya* **72**, 361–9 (2003).
136. Korzhenevskaya, T.G., Gorelova, O.A., Baulina, O.I. & Gusev, M.V. Accumulation of Reserve Polymers by Nostoc muscorum CALU 304 Cells Grown in Mixed Culture with Plant Tissue. *Mikrobiologiya* **68**, 191–197 (1999).
137. Gross, R.A. & Kalra, B. Biodegradable polymers for the environment. *Science* **297**, 803–7 (2002).
138. Hara, T., Kato, H., Katsube, Y. & Oda, J. A pseudo-michaelis quaternary complex in the reverse reaction of a ligase: structure of Escherichia coli B glutathione synthetase complexed with ADP, glutathione, and sulfate at 2.0 Å resolution. *Biochemistry* **35**, 11967–74 (1996).
139. van Heijenoort, J. Recent advances in the formation of the bacterial peptidoglycan monomer unit. *Nat Prod Rep* **18**, 503–19 (2001).
140. Neubauer, K. et al. Isolation of cyanophycin from tobacco and potato plants with constitutive plastidic cphATe gene expression. *J Biotechnol* **158**, 50–8 (2012).
141. Laranjo, M., Alexandre, A. & Oliveira, S. Legume growth-promoting rhizobia: an overview on the Mesorhizobium genus. *Microbiol Res* **169**, 2–17 (2014).
142. Zehr, J.P. & Ward, B.B. Nitrogen cycling in the ocean: new perspectives on processes and paradigms. *Appl Environ Microbiol* **68**, 1015–24 (2002).
143. Mota, C., Head, M.A., Ridenoure, J.A., Cheng, J.J. & de Los Reyes, F.L., 3rd. Effects of aeration cycles on nitrifying bacterial populations and nitrogen removal in intermittently aerated reactors. *Appl Environ Microbiol* **71**, 8565–72 (2005).
144. Bolotin, E. & Hershberg, R. Bacterial intra-species gene loss occurs in a largely clocklike manner mostly within a pool of less conserved and constrained genes. *Sci Rep* **6**, 35168 (2016).
145. Diaz-Saez, L., Torrie, L.S., McElroy, S.P., Gray, D. & Hunter, W.N. Burkholderia pseudomallei d-alanine-d-alanine ligase; detailed characterisation and assessment of a potential antibiotic drug target. *FEBS J* **286**, 4509–4524 (2019).
146. Yamaguchi, H. et al. Three-dimensional structure of the glutathione synthetase from Escherichia coli B at 2.0 Å resolution. *J Mol Biol* **229**, 1083–100 (1993).
147. Li, H., Fast, W. & Benkovic, S.J. Structural and functional modularity of proteins in the de novo purine biosynthetic pathway. *Protein Sci* **18**, 881–92 (2009).
148. Punjani, A., Rubinstein, J.L., Fleet, D.J. & Brubaker, M.A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290–296 (2017).

149. Gordon, E. et al. Crystal structure of UDP-N-acetylmuramoyl-L-alanyl-D-glutamate: meso-diaminopimelate ligase from *Escherichia coli*. *J Biol Chem* **276**, 10999–1006 (2001).
150. Beilsten-Edmands, J. et al. Scaling diffraction data in the DIALS software package: algorithms and new approaches for multi-crystal scaling. *Acta Crystallogr D Struct Biol* **76**, 385–399 (2020).
151. Foadi, J. et al. Clustering procedures for the optimal selection of data sets from multiple crystals in macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **69**, 1617–32 (2013).
152. Song, Y. et al. High-resolution comparative modeling with RosettaCM. *Structure* **21**, 1735–42 (2013).
153. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–32 (2004).
154. Skubak, P., Murshudov, G.N. & Pannu, N.S. Direct incorporation of experimental phase information in model refinement. *Acta Crystallogr D Biol Crystallogr* **60**, 2196–201 (2004).
155. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**(2018).
156. Zhang, K. Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**, 1–12 (2016).
157. Minh, B.Q. et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **37**, 1530–1534 (2020).
158. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* **47**, W256–W259 (2019).
159. Grogg, M., Hilvert, D., Beck, A.K. & Seebach, D. Syntheses of cyanophycin segments for investigations of cell-penetration. *Synthesis* **51**, 31–39 (2019).
160. Mooibroek, H. et al. Assessment of technological options and economical feasibility for cyanophycin biopolymer and high-value amino acid production. *Appl Microbiol Biotechnol* **77**, 257–67 (2007).
161. Watzer, B. et al. Metabolic pathway engineering using the central signal processor PII. *Microb Cell Fact* **14**, 192 (2015).
162. Becker, A.B. & Roth, R.A. An unusual active site identified in a family of zinc metalloendopeptidases. *Proc Natl Acad Sci U S A* **89**, 3835–9 (1992).
163. King, J.V. et al. Molecular basis of substrate recognition and degradation by human presequence protease. *Structure* **22**, 996–1007 (2014).
164. Rajagopalan, P.T., Datta, A. & Pei, D. Purification, characterization, and inhibition of peptide deformylase from *Escherichia coli*. *Biochemistry* **36**, 13910-8 (1997).
165. Fukasawa, K.M., Hata, T., Ono, Y. & Hirose, J. Metal preferences of zinc-binding motif on metalloproteases. *J Amino Acids* **2011**, 574816 (2011).
166. Perlman, R.K., Gehm, B.D., Kuo, W.L. & Rosner, M.R. Functional analysis of conserved residues in the active site of insulin-degrading enzyme. *J Biol Chem* **268**, 21538–44 (1993).
167. Rawlings, N.D. et al. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res* **46**, D624–D632 (2018).

168. Perlman, R.K. & Rosner, M.R. Identification of zinc ligands of the insulin-degrading enzyme. *J Biol Chem* **269**, 33140–5 (1994).
169. Cerda-Costa, N. & Gomis-Ruth, F.X. Architecture and function of metallopeptidase catalytic domains. *Protein Sci* **23**, 123–44 (2014).
170. Tallant, C., Marrero, A. & Gomis-Ruth, F.X. Matrix metalloproteinases: fold and function of their catalytic domains. *Biochim Biophys Acta* **1803**, 20–8 (2010).
171. Springman, E.B., Angleton, E.L., Birkedal-Hansen, H. & Van Wart, H.E. Multiple modes of activation of latent human fibroblast collagenase: evidence for the role of a Cys73 active-site zinc complex in latency and a "cysteine switch" mechanism for activation. *Proc Natl Acad Sci U S A* **87**, 364–8 (1990).
172. Lee, Y.M. & Lim, C. Physical basis of structural and catalytic Zn-binding sites in proteins. *J Mol Biol* **379**, 545–53 (2008).
173. Rawlings, N.D. & Salvesen, G. *Handbook of proteolytic enzymes*, 3 volumes (liv, 3932 pages) (Elsevier/AP, Amsterdam, 2013).
174. Chan, M.K. et al. Crystal structure of the *Escherichia coli* peptide deformylase. *Biochemistry* **36**, 13904–9 (1997).
175. Jain, R., Hao, B., Liu, R.P. & Chan, M.K. Structures of *E. coli* peptide deformylase bound to formate: insight into the preference for Fe<sup>2+</sup> over Zn<sup>2+</sup> as the active site metal. *J Am Chem Soc* **127**, 4558–9 (2005).
176. Rajagopalan, P.T.R., Yu, X.C. & Pei, D. Peptide deformylase: A new type of mononuclear iron protein. *Journal of the American Chemical Society* **119**, 12418–12419 (1997).
177. Hao, B. et al. Structural basis for the design of antibiotics targeting peptide deformylase. *Biochemistry* **38**, 4712–9 (1999).
178. Fukasawa, K.M., Hirose, J., Hata, T. & Ono, Y. In rat dipeptidyl peptidase III, His<sup>568</sup> is essential for catalysis, and Glu<sup>507</sup> or Glu<sup>512</sup> stabilizes the coordination bond between His<sup>455</sup> or His<sup>450</sup> and zinc ion. *Biochim Biophys Acta* **1804**, 2063–9 (2010).
179. Neumann, K. et al. Production of cyanophycin, a suitable source for the biodegradable polymer polyaspartate, in transgenic plants. *Plant Biotechnol J* **3**, 249–58 (2005).
180. Zhao, B. et al. Crystal structure of albaflavenone monooxygenase containing a moonlighting terpene synthase active site. *J Biol Chem* **284**, 36711–36719 (2009).
181. Holm, L. DALI and the persistence of protein shape. *Protein Sci* **29**, 128–140 (2020).
182. Pettersen, E.F. et al. UCSF Chimera – a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–12 (2004).
183. Burnley, T., Palmer, C.M. & Winn, M. Recent developments in the CCP-EM software suite. *Acta Crystallogr D Struct Biol* **73**, 469–477 (2017).
184. Potterton, L. et al. CCP4i2: the new graphical user interface to the CCP4 program suite. *Acta Crystallogr D Struct Biol* **74**, 68–84 (2018).
185. Tomitani, A., Knoll, A.H., Cavanaugh, C.M. & Ohno, T. The evolutionary diversification of cyanobacteria: molecular-phylogenetic and paleontological perspectives. *Proc Natl Acad Sci U S A* **103**, 5442–7 (2006).
186. Kitamura, Y. et al. Structure of D-alanine-D-alanine ligase from *Thermus thermophilus* HB8: cumulative conformational change and enzyme-ligand interactions. *Acta Crystallogr D Biol Crystallogr* **65**, 1098–106 (2009).

187. Kovalevskiy, O., Nicholls, R.A. & Murshudov, G.N. Automated refinement of macromolecular structures at low resolution using prior information. *Acta Crystallogr D Struct Biol* **72**, 1149–1161 (2016).
188. Adams, P.D. et al. The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94–106 (2011).
189. Du, J., Li, L. & Zhou, S. Microbial production of cyanophycin: From enzymes to biopolymers. *Biotechnol Adv* **37**, 107400 (2019).
190. Nausch, H. et al. Direct delivery of health promoting beta-Asp-Arg dipeptides via stable co-expression of cyanophycin and the cyanophycinase CphE241 in tobacco plants. *Front Plant Sci* **11**, 842 (2020).
191. Ponndorf, D. et al. Stable production of cyanophycinase in *Nicotiana benthamiana* and its functionality to hydrolyse cyanophycin in the murine intestine. *Plant Biotechnol J* **15**, 605–613 (2017).
192. Ponndorf, D., Broer, I. & Nausch, H. Expression of CphB- and CphE-type cyanophycinases in cyanophycin-producing tobacco and comparison of their ability to degrade cyanophycin in plant and plant extracts. *Transgenic Research* **26**, 491–499 (2017).
193. Nausch, H. & Broer, I. Cyanophycinase CphE from *P. alcaligenes* produced in different compartments of *N. benthamiana* degrades high amounts of cyanophycin in plant extracts. *Appl Microbiol Biotechnol* **101**, 2397–2413 (2017).
194. Radisky, E.S., Lee, J.M., Lu, C.J. & Koshland, D.E., Jr. Insights into the serine protease mechanism from atomic resolution structures of trypsin reaction intermediates. *Proc Natl Acad Sci U S A* **103**, 6835–40 (2006).
195. Kabsch, W. XDS. *Acta Crystallographica Section D* **66**, 125–132 (2010).
196. Evans, P.R. & Murshudov, G.N. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204–14 (2013).
197. Sharon, I. et al. A cryptic third active site in cyanophycin synthetase creates primers for polymerization. *Nat Commun* **13**, 3923 (2022).
198. Radkiewicz, J.L., Zipse, H., Clarke, S. & Houk, K.N. Accelerated Racemization of Aspartic Acid and Asparagine Residues via Succinimide Intermediates: An ab Initio Theoretical Exploration of Mechanism. *Journal of the American Chemical Society* **118**, 9148–9155 (1996).
199. Borek, D. et al. Expression, purification and catalytic activity of *Lupinus luteus* asparagine beta-amidohydrolase and its *Escherichia coli* homolog. *Eur J Biochem* **271**, 3215–26 (2004).
200. Prah, A., Pazgier, M., Hejazi, M., Lockau, W. & Lubkowski, J. Structure of the isoaspartyl peptidase with L-asparaginase activity from *Escherichia coli*. *Acta Crystallogr D Biol Crystallogr* **60**, 1173–6 (2004).
201. Flores, E., Arévalo, S. & Burnat, M. Cyanophycin and arginine metabolism in cyanobacteria. *Algal Research* **42**, 101577 (2019).
202. Haley, E.E. Purification and Properties of a  $\beta$ -Aspartyl Peptidase from *Escherichia coli*. *Journal of Biological Chemistry* **243**, 5748–5752 (1968).
203. Noronkoski, T., Stoineva, I.B., Ivanov, I.P., Petkov, D.D. & Mononen, I. Glycosylasparaginase-catalyzed synthesis and hydrolysis of beta-aspartyl peptides. *J Biol Chem* **273**, 26295–7 (1998).

204. O'Leary, N.A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research* **44**, D733-D745 (2016).
205. Osbourn, A.E. & Field, B. Operons. *Cell Mol Life Sci* **66**, 3755-75 (2009).
206. Chan, Z. et al. Draft genome sequence of an agar-degrading marine bacterium *Flammeovirga pacifica* WPAGA1. *Marine Genomics* **20**, 23-24 (2015).
207. Bresler, M.M., Rosser, S.J., Basran, A. & Bruce, N.C. Gene cloning and nucleotide sequencing and properties of a cocaine esterase from *Rhodococcus* sp. strain MB1. *Appl Environ Microbiol* **66**, 904-8 (2000).
208. Patananan, A.N., Capri, J., Whitelegge, J.P. & Clarke, S.G. Non-repair pathways for minimizing protein isoaspartyl damage in the yeast *Saccharomyces cerevisiae*. *J Biol Chem* **289**, 16936-53 (2014).
209. Gilchrist, C.L.M. et al. cblaster: a remote search tool for rapid identification and visualization of homologous gene clusters. *Bioinformatics Advances* **1**, vbab016 (2021).
210. Medema, M.H., Takano, E. & Breitling, R. Detecting sequence homology at the gene cluster level with MultiGeneBlast. *Mol Biol Evol* **30**, 1218-23 (2013).
211. Maheswaran, M., Ziegler, K., Lockau, W., Hagemann, M. & Forchhammer, K. PII-regulated arginine synthesis controls accumulation of cyanophycin in *Synechocystis* sp. strain PCC 6803. *J Bacteriol* **188**, 2730-4 (2006).
212. Ziegler, K., Stephan, D.P., Pistorius, E.K., Ruppel, H.G. & Lockau, W. A mutant of the cyanobacterium *Anabaena variabilis* ATCC 29413 lacking cyanophycin synthetase: growth properties and ultrastructural aspects. *FEMS microbiology letters* **196**, 13-8 (2001).
213. Sharon, I. & Schmeing, T.M. Characterization of promiscuous isoaspartyl dipeptidases that catalyze the final step in cyanophycin degradation. *Manuscript in preparation* (2022).
214. Schneider, B.L., Kiupakis, A.K. & Reitzer, L.J. Arginine catabolism and the arginine succinyltransferase pathway in *Escherichia coli*. *J Bacteriol* **180**, 4278-86 (1998).
215. Alberts, I.L., Nadassy, K. & Wodak, S.J. Analysis of zinc binding sites in protein crystal structures. *Protein Sci* **7**, 1700-16 (1998).
216. Nishijyo, T., Park, S.M., Lu, C.D., Itoh, Y. & Abdelal, A.T. Molecular characterization and regulation of an operon encoding a system for transport of arginine and ornithine and the ArgR regulatory protein in *Pseudomonas aeruginosa*. *J Bacteriol* **180**, 5559-66 (1998).
217. Lu, C.D., Yang, Z. & Li, W. Transcriptome analysis of the ArgR regulon in *Pseudomonas aeruginosa*. *J Bacteriol* **186**, 3855-61 (2004).
218. Parks, D.H. et al. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Research* **50**, D785-D794 (2021).
219. Itoh, Y. Cloning and characterization of the *aru* genes encoding enzymes of the catabolic arginine succinyltransferase pathway in *Pseudomonas aeruginosa*. *J Bacteriol* **179**, 7280-90 (1997).
220. Nikolaidis, M., Mossialos, D., Oliver, S.G. & Amoutzias, G.D. Comparative Analysis of the Core Proteomes among the *Pseudomonas* Major Evolutionary Groups Reveals Species-Specific Adaptations for *Pseudomonas aeruginosa* and *Pseudomonas chlororaphis*. *Diversity* **12**(2020).

221. LaBauve, A.E. & Wargo, M.J. Growth and laboratory maintenance of *Pseudomonas aeruginosa*. *Curr Protoc Microbiol* **Chapter 6**, Unit 6E 1 (2012).
222. Panvert, M. et al. Cdc123, a Cell Cycle Regulator Needed for eIF2 Assembly, Is an ATP-Grasp Protein with Unique Features. *Structure* **23**, 1596-1608 (2015).
223. Ellsworth, B.A., Tom, N.J. & Bartlett, P.A. Synthesis and evaluation of inhibitors of bacterial D-alanine:D-alanine ligases. *Chem Biol* **3**, 37-44 (1996).
224. Hrast, M. et al. Mur ligases inhibitors with azastilbene scaffold: Expanding the structure-activity relationship. *Bioorg Med Chem Lett* **40**, 127966 (2021).
225. Roper, D.I., Huyton, T., Vagin, A. & Dodson, G. The molecular basis of vancomycin resistance in clinically relevant Enterococci: crystal structure of D-alanyl-D-lactate ligase (VanA). *Proc Natl Acad Sci U S A* **97**, 8921-5 (2000).
226. Adames, K., Euting, K., Broker, A. & Steinbuechel, A. Investigations on three genes in *Ralstonia eutropha* H16 encoding putative cyanophycin metabolizing enzymes. *Appl Microbiol Biotechnol* **97**, 3579-91 (2013).
227. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
228. Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871-876 (2021).