Structures and Function of the Amino Acid Polymerase Cyanophycin Synthetase

Itai Sharon[1], Asfarul S. Haque[1], Marcel Grogg[2], Indrajit Lahiri[3], Dieter Seebach[2], Andres E. Leschziner[3], Donald Hilvert[2] & T. Martin Schmeing[1]

[1]Department of Biochemistry and Centre de recherche en biologie structurale, McGill University, Montréal, QC, Canada, H3G 0B1.

[2] Laboratory of Organic Chemistry, ETH Zürich, CH-8093 Zürich, Switzerland.

[3]Department of Cellular and Molecular Medicine, and Section of Molecular Biology, Division of Biological Sciences, University of California San Diego, La Jolla, CA 92093.

Correspondence e-mail: martin.schmeing@mcgill.ca

Abstract

Cyanophycin is a natural biopolymer produced by a wide range of bacteria, consisting of a chain of poly-L-Asp residues with L-Arg residues attached to the □-carboxylate side chains by isopeptide bonds. 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. 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.

Introduction

Cyanophycin was described over 130 years ago as a light-scattering granule in cyanobacterial cells[1]. It is a biopolymer of a poly-L-Asp backbone with L-Arg residues attached via isopeptide bonds to the β-carboxylates of each Asp side chain, ranging in length from 80 to 400 dipeptides ($(\beta\text{-Asp-Arg})_{\sim 80\text{-}400}$)[2,3] (Fig. 1a). The high nitrogen content of cyanophycin makes it good for storage of fixed nitrogen[4]. With 24% nitrogen by mass, it does so more efficiently than proteins (~13-19%) and nucleic acids (~16%)[5], whereas glycogen and fat contain no nitrogen. Cyanophycin is useful for bacteria that keep nitrogen and carbon fixation separated, either spatially or temporally[6], because nitrogenase enzymes are inactivated in aerobic environments[7]. 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[6]. In multicellular cyanobacteria communities, cyanophycin metabolism can be performed in specialized heterocysts, dedicated cells for fixing nitrogen[8] or spore-like akinetes[9]. The heterocysts generate and store cyanophycin, which is degraded when needed and transferred to vegetative cells that cannot fix nitrogen[10]. Cyanophycin can also be used for carbon and energy storage[11,12], as a scavenged nutrient source[13], for other metabolic processes[11,14], in spore assembly[15], and in plant-symbiont relationships[16,17].

Commercial interest in cyanophycin has led to its heterologous production in hosts from bacteria to tobacco[18-20]. Potential commercial applications include processing to poly-Asp for use as a biodegradable antiscalant, water softener, and super swelling material[21]. Cyanophycin-type polymers are also interesting for biotechnological applications, like formation of heat-sensitive nanovesicles[22].

Cyanophycin is synthesized in bacteria by cyanophycin synthetase (CphA1) using ATP, aspartate and arginine[23] (Fig. 1b). CphA1 is a dimer[24] or tetramer[25] 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[26,27], and the C-terminal ~400 residues (“M domain”) have homology to MurE-like muramyl ligases[28](Fig. 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 polymerization (Fig. 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[23,26,28]. 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[23,26,28]. Cyanophycin synthesis is usually “primer dependent”[27], where CphA1 only extends an existing segment of cyanophycin or, less efficiently, another biopolymer[24,25].

Many studies sought to dissect, characterize and exploit CphA1, and its overall activity and substrate specificity is established[2,23,25,27,29,30]. 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.

Results

CphA1 is a common bacterial enzyme

Research on cyanophycin has largely focused on cyanobacteria, because of its discovery in that phylum. In 2007, Füser and Steinbüchel reported *cphA1* genes in 44 of 946 bacterial genomes analyzed, including in non-cyanobacterial species[31]. In the current NCBI non-redundant protein data sequence bank, we found >4000 *cphA1* sequences. Strikingly, only 18% of species encoding CphA1 are cyanobacterial (Extended Data Figure 1). CphA1 is found in most bacterial phyla, including groups like *Rhizobiales*, which form symbiotic relationships with legumes[32]; *Nitrosomonas* and *Nitrosospira*, which are important for the nitrogen cycle[33] and wastewater treatment[34]; 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[35,36]. There is evidence for both ancient and recent horizontal gene transfer and repeated loss of *cphA1* (Extended Data Figure 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) (Extended Data Figure 1). These CphA1s produced cyanophycin *in vitro* from Asp, Arg, ATP and cyanophycin primer with different kinetics (Fig. 2a,b), within the range of previously-reported rates[37].

Architectures of cyanophycin synthetase

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. 2c,d, Supplementary Tables 1-3). The structural and solution data indicate that all three enzymes are tetramers, assembled as dimers of dimers (Fig. 2c,d, Extended Data Figure 2). All CphA1 monomers and dimers are similar to each other, but they form two distinct tetrameric architectures (Fig. 2, Extended Data Figure 2a-d).

CphA1 monomers are tri-lobed, with each lobe corresponding to one of the three domains (Fig. 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 $Å^2$ of surface area (Fig. 2d – middle).

The tetramer architectures differ between gammaproteobacterial and cyanobacterial CphA1 (Fig. 2d – top, Extended Data Figure 2). In *Su*CphA1, each tetramer interface buries only ~450 $Å^2$ of surface area, through M domain residue W672 inserting into a pocket near R470 of the G domain of the adjacent dimer (Extended Data Figure 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. 2d, Supplementary Movie 1, Extended Data Figure 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 $Å^2$ 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; Extended Data Figure 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. 2d). *Ab*CphA1, *Tm*CphA1 and *Su*CphA1 are all tetrameric in solution (Extended Data Figure 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. 2d, top right). The key *Su*CphA1 tetramerization residue, W672, is conserved in cyanobacteria and *Bacteroidetes*, but not elsewhere (Extended Data Figure 2). These groups represent ~30% of CphA1s, and they likely all have the morning star shape.

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.

Structure and mutational analysis of the G domain

The G domain catalyzes ATP-dependent addition of Asp to the C-terminus of cyanophycin polymer[23,26,28] (Fig. 1b). The active site of the G domain is located between the body of the G domain (G_{core}) and two subdomains, G_{lid} (*Su*CphA1 residues 235-305) and G_{omega} (residues 325-399) (Fig. 2d, 3a). G_{core} and G_{lid} are also present in bacterial glutathione synthetases[26] and *D*-alanine-*D*-alanine ligases[38] (Fig. 3b). The G_{omega} subdomain incorporates the "large loop"[39] of glutathione synthetase (Fig 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. 3b)[40]. The overall binding of ATP is similar to that of ATP-grasp enzymes (Extended Data Figure 3a), and has a partially ordered "P-loop" (residues 263-269)[26,41]

covering part of the active site (Fig. 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[42].

Cryo-EM datasets of *Su*CphA1 in the presence of the Asp, ADPCP and cyanophycin analog $(\beta\text{-Asp-Arg})_8\text{-NH}_2$ (Supplementary note section 3.5, Extended Data Figure 3f) gave a 2.6 Å resolution map with clear signal for ATP and cyanophycin analogs at the active site (Fig 3c). Three β-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_{core}/G_{alpha}, away from G_{lid} and G_{omega}, and toward the N domain. The polymer makes several interaction 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. 3d). The terminal amide group is ~6 Å from the γ-phosphate of ADPCP (Fig. 3c) in the average conformation represented by the EM map, just out of range for nucleophilic attack for phosphorylation.

G_{lid} and most of G_{omega} show clear evidence of mobility (Extended Data Figure 4a). The range of motion of G_{lid} is demonstrated by the *Tm*CphA1 structure, where crystal contacts hold G_{lid} of molecule B in a conformation rotated by 32°, which does not allow ATP binding (Extended Data Figure 4b). However, the average G_{lid} position does not change markedly upon polymer binding to *Su*CphA1, as there is a maximal ~2 Å shift in positions with and without polymer. Three-dimensional variability analysis with CryoSPARC[43] reveals distinct modes of movement for both G_{lid} and G_{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 γ-phosphate of ATP, bridging the ~6 Å gap we see, and allowing the G domain to catalyze its first reaction (Fig. 1b) to produce main-chain phosphorylated cyanophycin.

The large loop of G_{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[40]), 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* homoglutathione synthetase is important for selection of its amino acid substrate[44]. 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. 3d).

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[23,26,28] (Fig. 1b). The active site of the M domain in *Su*CphA1 is between the central body (M_{core}; *Su*CphA1 488-723) and a large subdomain (M_{lid}; 724-875) (Fig. 2d, 4a). M_{core} and M_{lid} are conserved with Mur-ligases, which additionally have an N-terminal lobe not present in CphA1 (Fig. 4b, Extended Data Figure 3c)[45]. M_{lid} displays a large range of motion, also observed in Mur ligases[46,47]. In the *Tm*CphA1 crystal structure, M_{lid} is disordered in one monomer and held far from an ATP-binding conformation by crystal contacts in the other (Extended Data Figure 4b). In the EM map of *Ab*CphA1, no sign of M_{lid} is present, even at low threshold, despite the presence of ATP in the sample. M_{lid} 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 (β-Asp-Arg)$_8$-Asn (Extended Data Figure 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. 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 (Extended Data Figure 3e). Thereafter, signal is visible only at low contour, and extends toward the N domain (Extended Data Figure 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. 4c,d). R561 orients the Asp (Asn in our analog), so its reactive side chain is 4.2 Å from the ATP γ-phosphate, in a good pre-attack conformation. Two Mg^{2+} coordinating the β and γ-phosphates of ATP are resolved (Fig. 4c). The structures and 3D variability analysis (Supplementary Movie 3) all indicate flexibility of M_{lid}, in which, as in Mur ligases[46], 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_{core} and M_{lid} for isopeptide bond formation, but we could not unambiguous identify it in the maps.

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 β sheet backed by two long, antiparallel helices, a long helix at 45° to those, and two or three shorter helices (Fig. 5a,b). Structure similarity searches shows good matches only for the four-stranded β sheet and two backing helices, e.g. with parts of *E. coli* RNA-polymerase α-subunit (Extended Data Figure 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 α-helices (Fig. 5a,b, Extended Data Figure 4d). *Su*CphA1 α_a (residues 119-138) includes four surface-exposed arginines, and α_b (144-161) includes six surface-exposed aspartates or glutamates (Fig. 5a). This trend is reversed in *Ab*CphA1 and *Tm*CphA1, where α_a is more negative, with five aspartates or glutamates, and α_b is more positive, with four arginines or lysines (Fig. 5b). In all three enzymes, α_a and α_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 (Extended Data Figure 4d). This pattern suggests that these charged patches, including helices α_a and α_b, could be involved in cyanophycin binding. Indeed, comparison of the unsharpened EM maps of *Su*CphA1 without a cyanophycin analog to those with (β-Asp-Arg)$_8$-Asn or (β-Asp-Arg)$_8$-NH_2 clearly shows additional features when a cyanophycin analog is present (Extended Data Figure 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_{omega} near the G domain active site (Supplementary Movie 3). Cryo-EM maps of *Su*CphA1 in the presence of (β-Asp-Arg)$_{16}$ are nearly identical to that of the (β-Asp-Arg)$_8$-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 α_a and α_b is repeatedly observed.

The structures imply that cyanophycin binds CphA1 through loose anchoring to the N domain via salt bridges with α_a and α_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. 5d). Mutagenesis of charged residues on α_a and α_b also

support this binding mode: The triple α_a mutant R123A-R127A-R131A and quadruple α_b mutant D150A-E152A-D153A-D156A each display decreased activity, and the combined α_a-α_b septuple alanine mutations further reduced cyanophycin synthesis (Fig. 5e). However, an α_a-α_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 (Extended Data Figure 5c), and DSF of all mutants showed them to have a T_m similar to that of WT enzymes (Extended Data Figure 5d). The clear effect of mutating residues located so far from the active sites (~32 – 62 Å) strongly suggests that loose, N-domain anchoring is a key contributing feature of efficient cyanophycin synthesis.

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 α_a/α_b helix pairs important for cyanophycin binding. In a CphA1 monomer, the G and M active sites are ~60 Å apart, with the α_a/α_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 ~60 Å and ~80 Å (in *Su*CphA) or ~70 Å and ~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 suggest 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 (Extended Data Figure 2e, 6a), we mutated it to alanine, and indeed saw *Su*CphA1(W672A) to be dimeric (Fig. 6a). Interestingly, dimeric *Su*CphA1 displayed very similar cyanophycin synthesis activity in *in vitro* assays (Fig. 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. 6c, Extended Data Figure 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 comprised of one monomer encoded by each plasmid. We assessed combinations of dimeric *Su*CphA1(W672A): G^+M^+/G^+M^+; G^+M^+/G^-M^-; G^-M^+/G^+M^-, as well as G^-M^+/G^-M^+ and G^+M^-/G^+M^- negative controls (Fig. 6c,d). As expected, G^+M^+/G^-M^- and G^-M^+/G^+M^- have reduced activity compared to G^+M^+/G^+M^+, 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^+M^+/G^-M^-) compared to inactivating mutations of one active site in

each monomer ($G^{-}M^{+}/G^{+}M^{-}$) (Fig. 6d). Altering ionic strength did not change this result (Extended Data Figure 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.

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_{core}/G_{alpha}, M_{core}), with spiky projections formed by the N domain and mobile subdomains adjacent to the active sites (G_{lid}, G_{omega}, M_{lid}). 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. ellipsosporum* CphA1 increased thermal stability and activity *in vivo*[29,48]. *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. ellipsosporum* CphA1 by 31 residues did not inhibit activity, while truncation by 59 residues led to complete inactivation[29]: The former removes only the variable C-terminus while the latter also removes part of M_{lid}, including the central β-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[27]. 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 β-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 $(\beta\text{-Asp-Arg})_3$-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 α_a and α_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 β strand backbone angles and Asp side chain χ2 angles of ~150° positions the positive Arg guanidiniums in alignment with α_b and the negative Arg α-carboxylates in alignment with α_a (Extended Data Figure 5b). Switching the χ2 angles to ~-30° allows cyanophycin's positive charges to interact with α_a and its negative charges with α_b. This

would enable CphA1s with either helix-charge pattern (Fig. 5a,b), or a mixed pattern, to interact with cyanophycin. This plasticity also results in cryptic conservation of these patches. In cyanobacteria, α_a is largely positive and α_b largely negative (Fig. 5a,c), but this pattern is not conserved among other groups. In *Tm*CphA1 and *Ab*CphA1, α_a is largely negative and α_b positive (Fig. 5b), and other gammaproteobacterial CphA1s display other charge distribution patterns (Fig. 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. 6e, Extended Data Figure 8, Supplementary Movie 4). The sigmoidal shape of cyanophycin synthesis is consistent with distinct initiation and elongation phases of synthesis (Fig. 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 (Extended Data Figure 8). However, the position of the N domain in the tetramers (Fig. 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. 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.

Figures

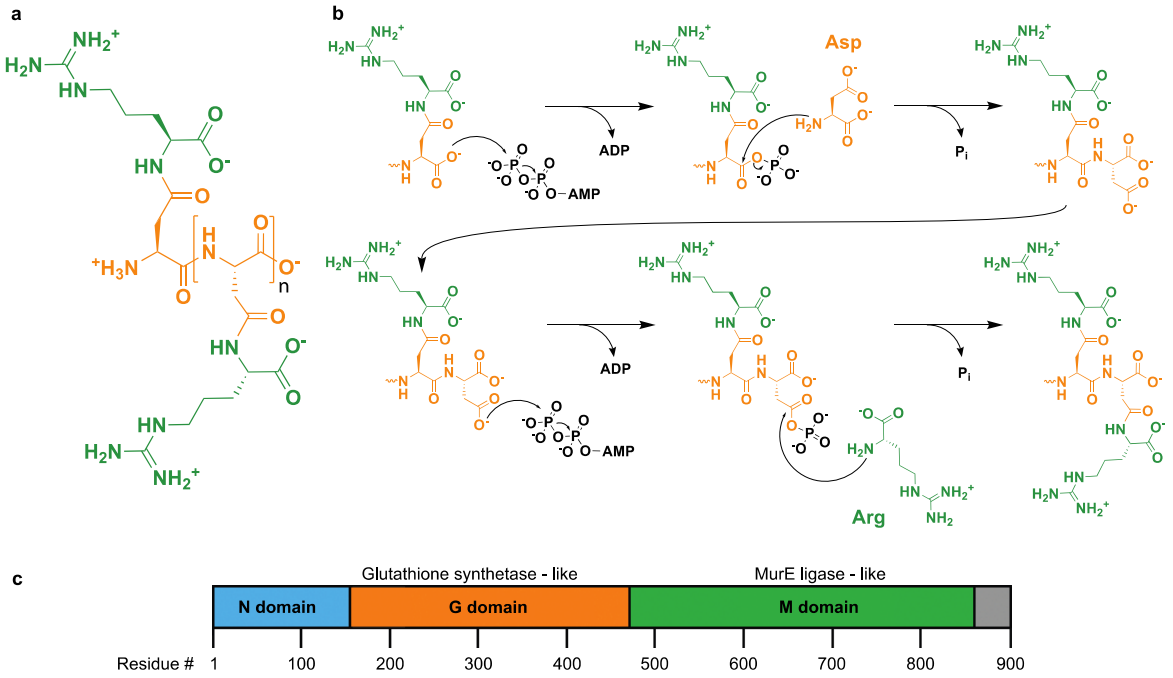

Figure 1. Cyanophycin and cyanophycin synthetase. (a) The chemical structure of cyanophycin. The backbone is made of L-Asp residues and each Asp side chain is linked through an isopeptide bond to an L-Arg residue. n=80-400. (b) The reactions catalyzed by CphA1. Top: First, the terminal carboxylate of 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.

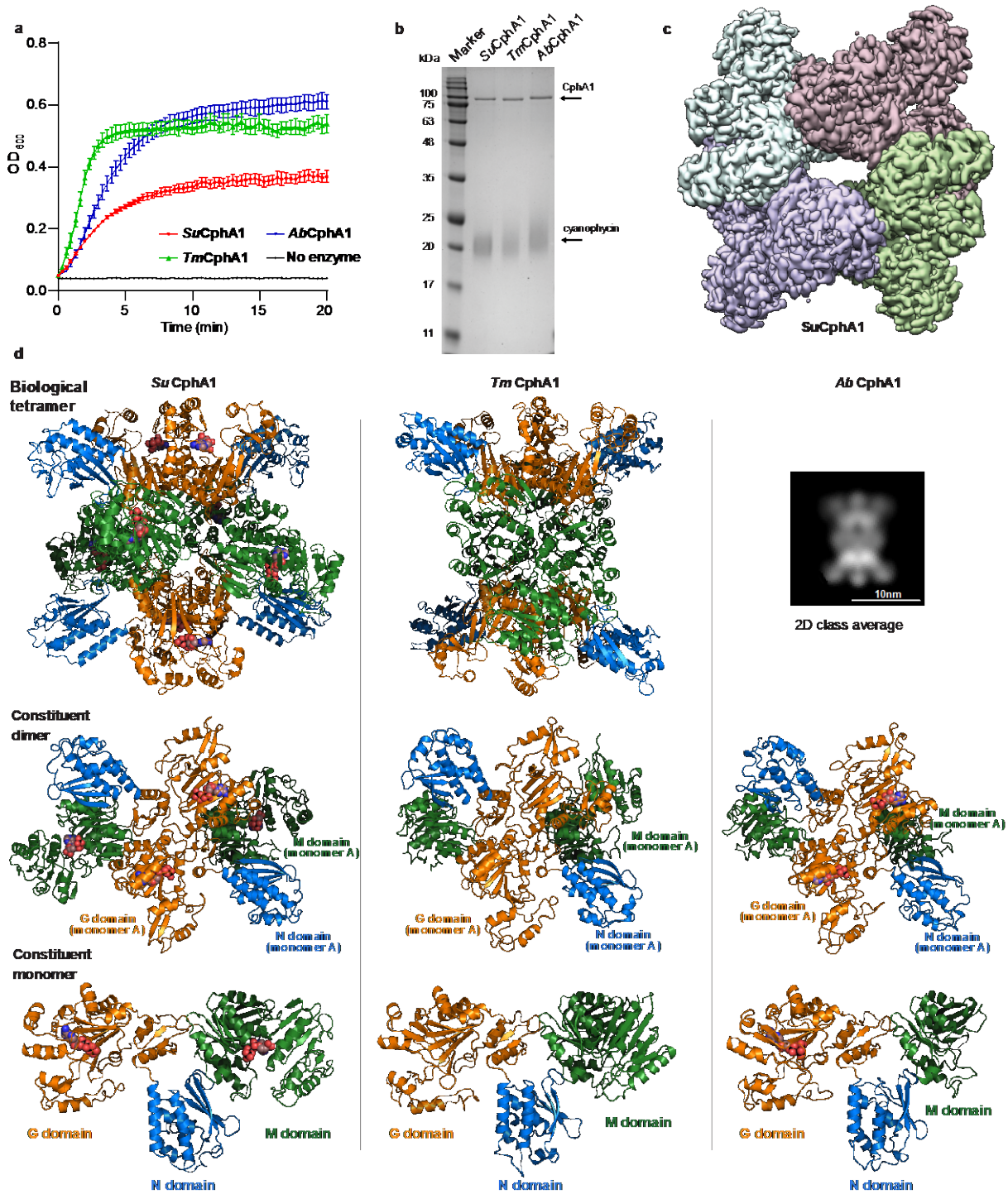

Figure 2. Overall structure and activity of CphA1. (a) Cyanophycin-synthesis activity of the three homologs used in this study. Cyanophycin polymer formed in the reaction scatters light, causing an increase in OD_{600}. The activity rates determined for each homolog are: *Su*CphA1 –

149U, *Tm*CphA1 – 460U, *Ab*CphA1 – 249U, where 1 U is defined as the incorporation of 1nmol (β-Asp-Arg)/min[37]. 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 $Å^2$) 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 (Extended Data Figure 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.

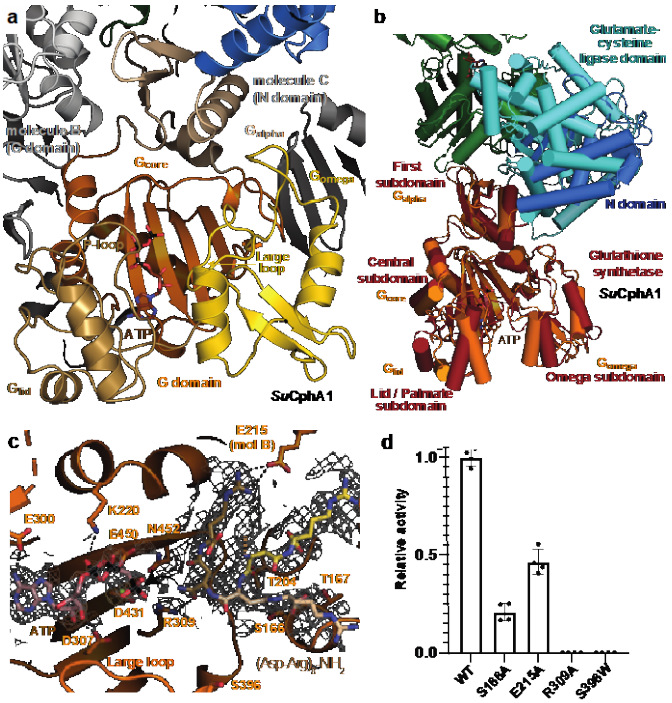

Figure 3. Structure and mutagenesis of the G domain. (a) The overall structure of *Su*CphA1 G domain, colored by subdomain. (b) Overlay of the *Su*CphA1 G domain and glutathione synthetase-cysteine ligase from *S. agalactiae*[40] showing the overall structure, including G_{core}, G_{lid} and G_{omega}. (c) The structure of the *Su*CphA1 G domain complexed with $(Asp\text{-}Arg)_8\text{-}NH_2$ and ADPCP. The Cryo-EM map was carved 2Å 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.

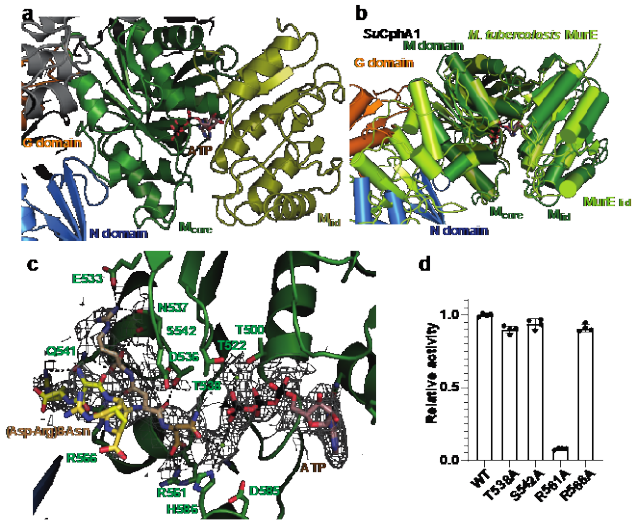

Figure 4. Structure and mutagenesis of the M domain. (a) The overall structure of the *Su*CphA1 M domain, colored by subdomains. (b) Overlay of the *Su*CphA1 M domain and MurE ligase from *M. tuberculosis*[47] showing the similar overall structure, including M_{lid} in the closed conformation. (c) The structure of the *Su*CphA1 M domain with (Asp-Arg)$_8$-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.

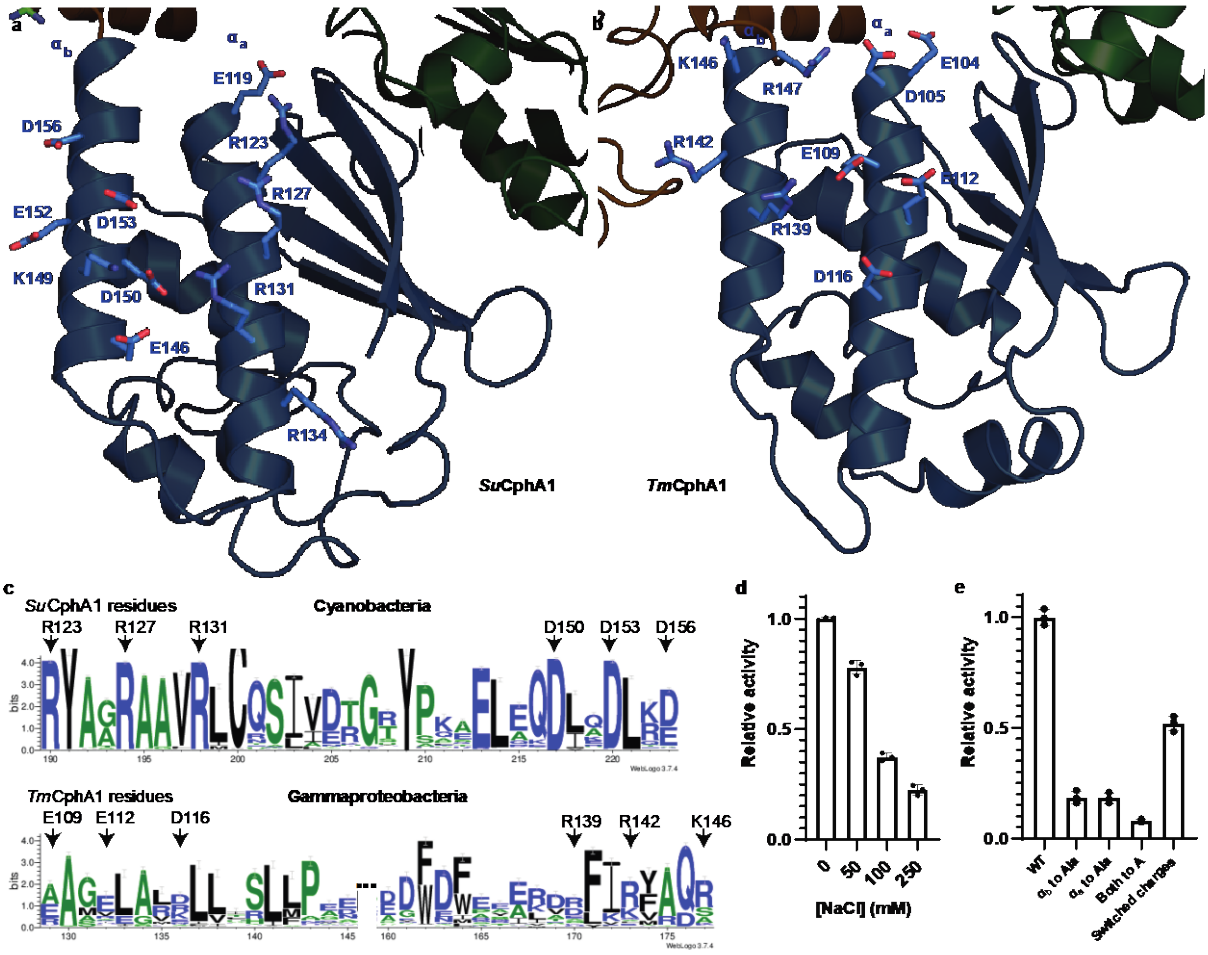

Figure 5. Structure, conservation and mutagenesis of the N domain. (a,b) The charged residues on α_a and α_b of *Su*CphA1 (a) and *Tm*CphA1 (b) form patches of positive and negative charges on their surface. (c) Weblogo[49] analysis of the region covering α_a and α_b of *Su*CphA1 (top) and *Tm*CphA1 (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 α_a (R123A-R127A-R131A) and on α_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.

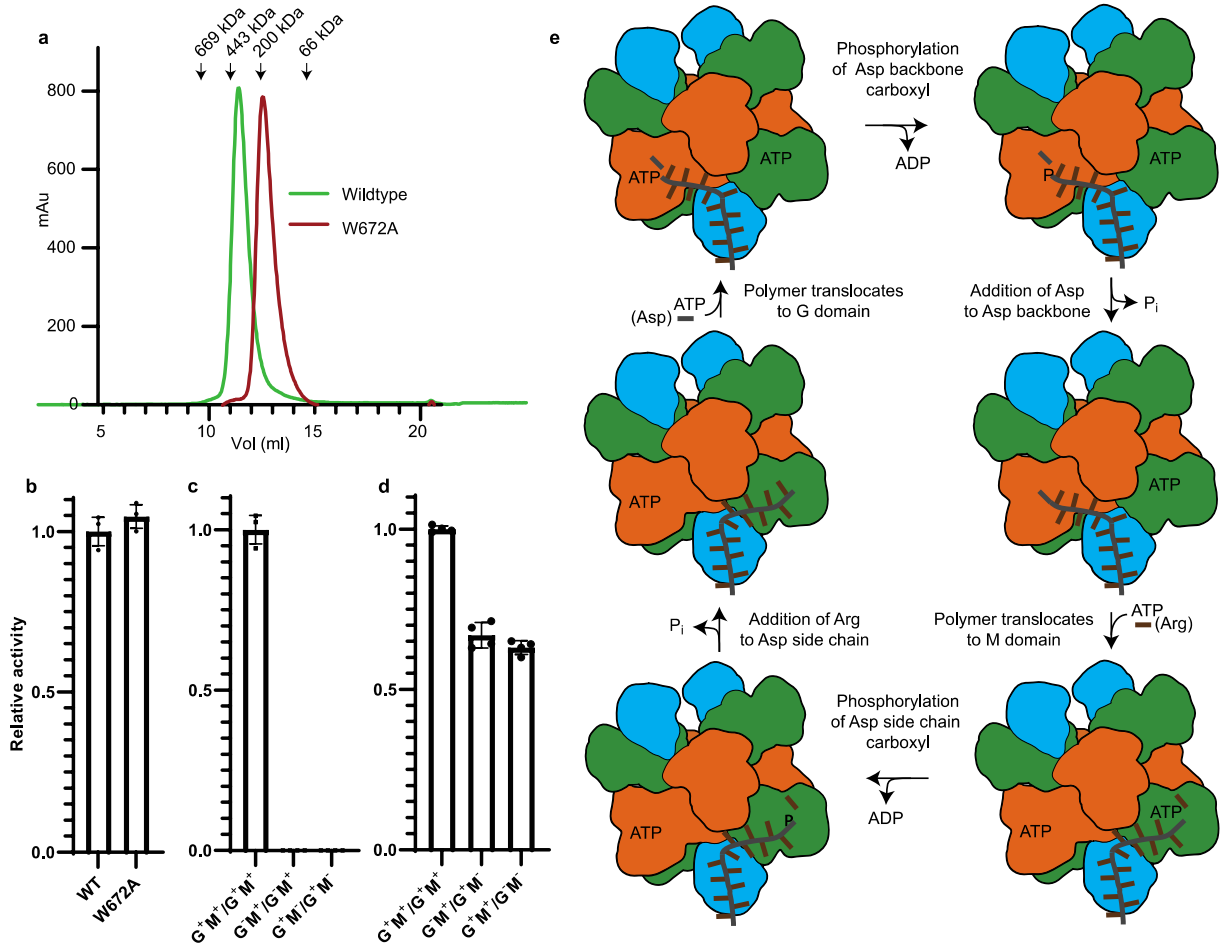

Figure 6. Dimeric *Su*CphA1 mutants and model of cyanophycin synthesis.

(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 Extended Data Figure 8 for the schematic of cyanophycin synthesis by G^-M^+/G^+M^-.

Methods

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 harboruing these plasmids were grown in LB or TB media supplemented with 100ug/ml kanamycin at 37 °C until they reached an OD_{600} 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-α *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_{600} 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 β-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 β-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 β-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 *Tm*CphA1 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). 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_{600} 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 and additional 40 hours prior to harvesting. Cells were lysed and nickel affinity chromatography was performed as described above. Pooled fractions were mixed with $CaCl_2$ 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 $CaCl_2$, 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 cetnrifucation concentrators (EMD Millipore). Glycerol was added to a final of 10% and sample was flash frozen and stored at -80 °C until use.

Crystallography of TmCphA1

Selenomethionine-labelled *Tm*CphA1 was crystallized using the sitting drop vapour diffusion method. Drop solution of 2µl of *Tm*CphA1 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[50] and then data from 6 crystals were analyzed using BLEND[51] 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 *Ab*CphA1 and models of residues 1-150 and 715-850 generated by Rosetta[52] as search models. The model was manually re-built and completed in Coot[53] and refined using REFMAC[54], LORESTR and Rosetta. Crystallography data statistics are listed in Supplementary Table 1.

Cryo-EM sample preparation and data collection

*Su*CphA1 in buffer F was mixed with substrate to give the following 4 samples: (1) No polymer sample - 2mg/ml *Su*CphA1, 10mM $MgCl_2$, 2mM ATP, 20mM Asp and 20mM Arg (dataset,); (2) G domain substrate analog sample - 2mg/ml *Su*CphA1, 10mM $MgCl_2$, 2mM AMPPCP, 20mM Asp, and 5mM (β-Asp-Arg)$_8$-NH_2; (3) Long G domain substrate analog sample - 3.5 mg/ml *Su*CphA, 10mM $MgCl_2$, 2mM ATP, and 1mM (β-Asp-Arg)$_{16}$-OH; and (4) M

domain substrate analog sample - 3.5 mg/ml *Su*CphA, 10mM $MgCl_2$, 2mM ATP, 20mM Arg, and 5mM $(\beta\text{-Asp-Arg})_8$-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/$Å^2$ and defocus range of -0.75 to -2.5µm at a nominal magnification of 105,000, resulting in a pixel size of 0.855$Å^2$. For *Ab*CphA1, protein in buffer G was mixed with 10mM $MgCl_2$, 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 *Su*CphA1. 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/$Å^2$ at a nominal magnification of 30 thousand resulting in an unbinned pixel size of 0.58$Å^2$. Data collection details are listed in Supplementary Table 2.

Cryo-EM data processing

*Su*CphA1 micrographs were motion corrected using Relion3.1[55]. The micrographs were imported to CryoSPARC2[43] 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. *Ab*CphA1data micrographs were processed in CryoSPARC2 unless otherwise stated. Patch motion corrected micrographs were CTF estimated using GCTF[56]. 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.

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

CphA1 activity assays

Reactions contained 20µg purified CphA1, 100mM HEPES pH8.2, 20mM KCl, 10mM $MgCl_2$, 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_{600} curve was taken. The derivatives curves were smoothed with a 2^{nd} 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_{600} on cyanophycin concentration, allowing us to determine specific activity in comparable units to those used by previous studies[37].

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[57] and displayed and manually annotated using iTOL[58].

Synthesis of cyanophycin segments

Solid phase synthesis was used for the synthesis of all molecules using Fmoc-(β-Asp-Arg)(O*t*Bu)-OH as building blocks in a manner similar to that previously described[59,60]. Full synthesis procedures are detailed in Supplementary note.

Differential Scanning Fluorimetry

DSF assays were performed with 0.5mg/ml protein in a buffer containing 50mm HEPES pH 8.2, 100mM NaClm 1mM DTT and 5x Sypro™ 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).

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 $(\beta\text{-Asp-Arg})_8\text{-NH}_2$ (EMD-23325), *Su*CphA1 bound with ATP and $(\beta\text{-Asp-Arg})_8$-Asn (EMD-23328), *Su*CphA1 with ATP and $(\beta\text{-Asp-Arg})_{16}$ (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 $(\beta\text{-Asp-Arg})_8\text{-NH}_2$ (7LGJ), *Su*CphA1 with ATP and (β-

Asp-Arg)$_8$-Asn (7LGQ), *Su*CphA1 with ATP, *Ab*CphA1 with ATP (7LGM), and *Tm*CphA1 (7LGN).

Author contributions

TMS and IS designed the study. IS performed all molecular biology and biochemical experiments. IS, SAH, IL and TMS performed cryo-EM data collection and initial structure determination. IS performed crystallography and structure determination, as well as modelling and refinement of all structures. MG performed chemical synthesis under the direction of DS and DH. TMS and IS wrote the manuscript and TMS, IS, AEL, IL and DH edited the manuscript.

Competing Interests

The authors declare no competing interests.

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.

References

1. Borzi, A. *Le communicazioni intracellulari delle nostochinee*, (Messina, 1886).
2. 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).
3. 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).
4. 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).
5. 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).
6. 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).
7. Fay, P. Oxygen relations of nitrogen fixation in cyanobacteria. *Microbiol Rev* **56**, 340-73 (1992).
8. Kumar, K., Mella-Herrera, R.A. & Golden, J.W. Cyanobacterial heterocysts. *Cold Spring Harb Perspect Biol* **2**, a000315 (2010).
9. 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).
10. 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).
11. 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).

12. 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).
13. 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).
14. 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).
15. 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).
16. Gorelova, O.A. & Kleimenov, S. The accumulation and degradation dynamics of cyanophycin in cyanobacteria grown in symbiotic associations with plant tissues and cells. *Mikrobiologiia* **72**, 361-9 (2003).
17. 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. *Mikrobiologiia* **68**, 191-197 (1999).
18. Aboulmagd, E., Voss, I., Oppermann-Sanio, F.B. & Steinbuchel, 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).
19. Nausch, H. et al. Tobacco as platform for a commercial production of cyanophycin. *N Biotechnol* **33**, 842-851 (2016).
20. 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).
21. Gross, R.A. & Kalra, B. Biodegradable polymers for the environment. *Science* **297**, 803-7 (2002).
22. 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).
23. 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). *European journal of biochemistry* **254**, 154-9 (1998).
24. Hai, T., Oppermann-Sanio, F.B. & Steinbuchel, 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. *Applied and environmental microbiology* **68**, 93-101 (2002).
25. Arai, T. & Kino, K. A cyanophycin synthetase from Thermosynechococcus elongatus BP-1 catalyzes primer-independent cyanophycin synthesis. *Applied microbiology and biotechnology* **81**, 69-78 (2008).
26. 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 A resolution. *Biochemistry* **35**, 11967-74 (1996).
27. 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. *European journal of biochemistry* **267**, 5561-70 (2000).
28. van Heijenoort, J. Recent advances in the formation of the bacterial peptidoglycan monomer unit. *Nat Prod Rep* **18**, 503-19 (2001).
29. Hai, T., Frey, K.M. & Steinbuchel, 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).
30. Neubauer, K. et al. Isolation of cyanophycin from tobacco and potato plants with constitutive plastidic cphATe gene expression. *J Biotechnol* **158**, 50-8 (2012).
31. Fuser, G. & Steinbuchel, A. Analysis of genome sequences for genes of cyanophycin metabolism: identifying putative cyanophycin metabolizing prokaryotes. *Macromol Biosci* **7**, 278-96 (2007).
32. Laranjo, M., Alexandre, A. & Oliveira, S. Legume growth-promoting rhizobia: an overview on the Mesorhizobium genus. *Microbiol Res* **169**, 2-17 (2014).
33. Zehr, J.P. & Ward, B.B. Nitrogen cycling in the ocean: new perspectives on processes and paradigms. *Appl Environ Microbiol* **68**, 1015-24 (2002).
34. 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).
35. Watzer, B. & Forchhammer, K. Cyanophycin synthesis optimizes nitrogen utilization in the unicellular cyanobacterium Synechocystis sp. PCC 6803. *Appl Environ Microbiol* (2018).
36. 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).
37. 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).
38. 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).
39. 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).
40. 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).
41. Yamaguchi, H. et al. Three-dimensional structure of the glutathione synthetase from Escherichia coli B at 2.0 A resolution. *J Mol Biol* **229**, 1083-100 (1993).
42. 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).
43. 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).

44. 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).
45. 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).
46. Smith, C.A. Structure, function and dynamics in the mur family of bacterial cell wall ligases. *J Mol Biol* **362**, 640-55 (2006).
47. 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).
48. 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).
49. Crooks, G.E., Hon, G., Chandonia, J.M. & Brenner, S.E. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-90 (2004).
50. 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).
51. 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).
52. Song, Y. et al. High-resolution comparative modeling with RosettaCM. *Structure* **21**, 1735-42 (2013).
53. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32 (2004).
54. 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).
55. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**(2018).
56. Zhang, K. Gctf: Real-time CTF determination and correction. *J Struct Biol* **193**, 1-12 (2016).
57. 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).
58. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* **47**, W256-W259 (2019).
59. 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**(2018).
60. Grogg, M., Hilvert, D., Beck, A.K. & Seebach, D. Syntheses of Cyanophycin Segments for Investigations of Cell-Penetration. *Synthesis* **51**, 31-39 (2019).