Structure and Function of the β-Asp-Arg Polymerase Cyanophycin Synthetase 2

Itai Sharon[1], Marcel Grogg[2], 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.

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

Abstract

Cyanophycin is a biopolymer composed of long chains of β-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 β-Asp-Arg dipeptides. CphA2 is evolutionarily related to CphA1, but questions about CphA2's altered structure and function remain unresolved. Cyanophycin and related biopolymers have drawn interest as potential "green" polymers. Because it only has a single active site, CphA2 could potentially be more useful than CphA1 for biotechnological applications seeking to produce modified cyanophycin or other polymers. In this study, we report biochemical assays on nine cyanobacterial CphA2 enzymes and report the crystal structure of CphA2 from *Gloeothece 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 accompanying biochemical assays explain how CphA2 binds and polymerizes β-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 *Gloeothece citriformis* CphA2 from a primer-dependent enzyme into a primer-independent enzyme.

Cyanophycin, first discovered in cyanobacteria almost 140 years ago[1], is a biopolymer produced by many bacterial species[2-3]. It has a poly-L-Asp backbone with an L-Arg attached to each sidechain through an isopeptide bond[4] (Fig. 1A). This composition gives it a high nitrogen content of 24% by mass, making it especially valuable for nitrogen storage[5], although it can also be useful for storing carbon and energy[6-7]. Cyanophycin's function as a nitrogen reservoir is especially beneficial in nitrogen-fixing cyanobacteria[8]: nitrogenase is inhibited by oxygen[9], so (aerobic) photosynthesis and nitrogen fixation must be separated, either temporally (in a day/night cycle)[10] or spatially (e.g. in heterocyst and vegetative cell types)[11]. 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[10-11].

In addition to its biological roles, cyanophycin also has attractive biotechnological potential, with. applications ranging from a source of poly-Asp (commonly used as a water softener and super swelling material as well as a biodegradable alternative to poly-acrylic acid) to a material for bandages[12]. Despite its attractive properties, 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 in order to maximize heterologous expression[13-15]. 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[16]. In contrast, CphA2 links β-Asp-Arg dipeptides together by a repetitive, ATP-dependent polymerization reaction at a single active site[17-18] (Fig. 1A). While primary metabolism provides the Asp and Arg building blocks for CphA1, the β-Asp-Arg dipeptide used by CphA2 is generated by degradation of cyanophycin by cyanophycinase[19]. CphA1 is found throughout bacteria[2-3], but CphA2 evolved from CphA1 fairly recently in cyanobacteria[17-18]. The two enzymes seem to have partially overlapping roles *in vivo*, as the functions of both have been shown to be important for maximal cyanophycin production in *Anabaena*[17-18]. 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[4]. In primer dependent synthesis, cyanophycin synthetase can extend an existing chain of cyanophycin polymer, but cannot perform *de novo* cyanophycin production from ATP, Asp and Arg (for CphA1), or from ATP and β-Asp-Arg (for CphA2). CphA1 enzymes are largely primer dependent, with one known exception[20]. Molecules other than cyanophycin can also serve as primers for CphA1, albeit with low efficiency[21], explaining how CphA1 can make cyanophycin in heterologous expression systems such as *E. coli* that do not possess pre-existing polymer[22]. CphA2 has been reported to be primer-independent *in vitro*, but the reactions were performed with a very high β-Asp-Arg concentration[18].

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[17-18] and the structures of the related CphA1 enzymes[2] 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[2, 18] (Sup. Fig. 1A). Moreover, the overall architecture of CphA2 is unknown and the oligomeric state uncertain[18]. 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 CphA2 enzymes from 9 different bacteria and solved the structure of one, *Gloeothece citriformis* PCC7424 CphA2, 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.

Results and Discussion

Cyanophycin synthesis by CphA2

The β-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 by other ATP-grasp enzymes. In the first step, the terminal carboxylate of the cyanophycin polymer is phosphorylated using ATP and, in a second step, the resulting acylphosphate intermediate undergoes nucleophilic attack by the alpha amine of β-Asp-Arg, resulting in extension of the polymer chain by one dipeptide (Fig. 1A). The published *in vitro* activity assay conditions for *A. variabilis* ATCC 29413 and *Cyanothece sp*. PCC 7425 CphA2 included 100 mM of the dipeptide substrate β-Asp-Arg[18], a very high concentration which is unlikely to be common in the cell. Although a cyanophycin primer, which is often required for CphA1 activity[16], was not necessary, the authors stated that the presence of purified soluble cyanophycin enhanced *in vitro* activity[18].

For more detailed biochemical characterization of how primers influence CphA2-catalyzed cyanophycin synthesis, we cloned, expressed and purified CphA2 from *Gloeothece citriformis* PCC7424 (Sup. Fig 1C). Cyanophycin formation was monitored by an increase in the OD_{600} of the reaction solution caused by scattering of light by insoluble cyanophycin[2]. Although no activity was observed for 2 mM β-Asp-Arg alone, addition of 50 μM $(\beta\text{-Asp-Arg})_3$ as a primer enabled robust polymerization of the dipeptide (Fig. 1B). The shorter $(\beta\text{-Asp-Arg})_2$ peptide could also prime the reaction but was less effective than $(\beta\text{-Asp-Arg})_3$, whereas increasing the length of the primer to $(\beta\text{-Asp-Arg})_4$ did not result in a further increase in rate (Fig. 1C). Thus, $(\beta\text{-Asp-Arg})_3$ 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. 1D) and at moderate potassium chloride concentrations (Fig. 1E). Potassium is required for CphA1 and CphA2 activity, and cannot be replaced by sodium[18]. However, high ionic strength was inhibitory, as activity decreased sharply with increasing KCl or NaCl concentrations (Fig. 1E,F).

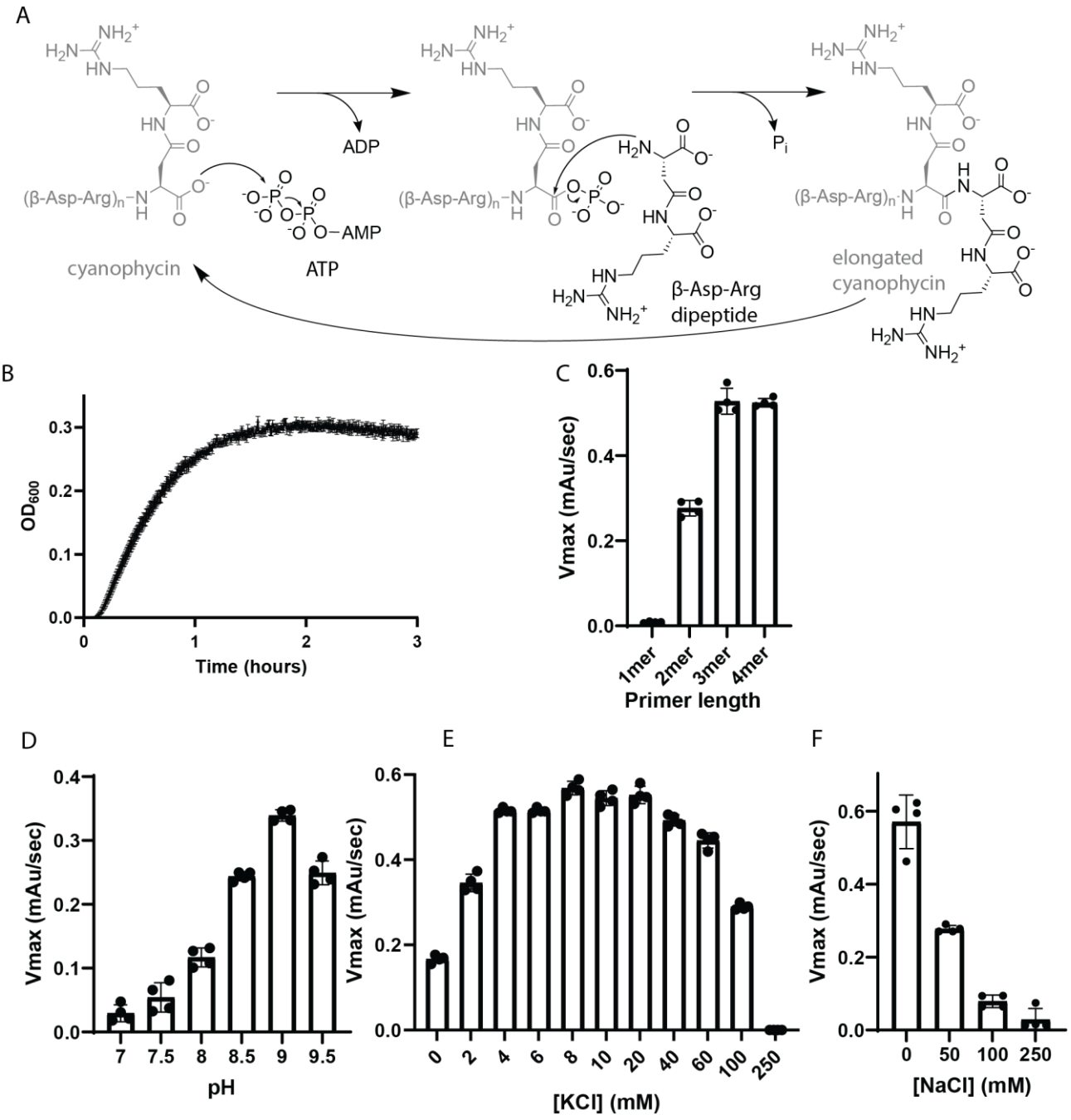

Figure 1. Cyanophycin synthesis by CphA2. (A) Schematic diagram of cyanophycin synthesis catalyzed by CphA2. First, the terminal backbone carboxylate of a cyanophycin chain is phosphorylated by ATP. This intermediate is then attacked by the alpha amino group of an incoming β-Asp-Arg dipeptide, thus extending the cyanophycin polymer. (B) Cyanophycin synthesis by *G. citriformis* CphA2 in the presence of 50 µM $(\beta\text{-Asp-Arg})_3$ primer, measured by increase in the OD_{600} of the reaction solution caused by scattering of light by insoluble cyanophycin. (C) *G. citriformis* CphA2 activity with different primer lengths shows that $(\beta\text{-Asp-Arg})_2$ is the minimal chain length that can serve as primer and at least $(\beta\text{-Asp-Arg})_3$ is required for maximal rate. (D) pH dependence of *G. citriformis* CphA2 activity. Bis-tris propane was used as buffer in these assays as it covers the optimal pH range better than HEPES, which was used in all other experiments. The enzyme displayed maximal activity around pH 9. (E) The dependence of *G. citriformis* CphA2 activity on KCl concentration. At least 8 mM are required for maximal activity rate. Substantially higher concentrations inhibit activity. (F) The dependence of *G. citriformis* CphA2 activity on NaCl concentration. Reactions contain 40 mM KCl in addition to the NaCl concentrations indicated. Increasing ionic strength decreases V_{max}, presumably because the high salt concentration interferes with the binding of cyanophycin, which is predominantly electrostatic. All experiments were carried out in quadruplicate. Bar height represents the mean value of the maximal activity, and error bars show the standard deviation.

Having established reaction conditions for one CphA2 enzyme, we expanded our study to additional examples, cloning and expressing CphA2 enzymes from an additional eight different cyanobacterial species: *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. These homologs have 51-97% identity to each other (Sup. Fig. 1B), and include members of cyanobacterial sections I-IV[23]. All enzymes expressed well in *E. coli* and could be purified in soluble form (Sup. Fig. 1C). Like *G. citriformis* CphA2, seven of the new enzymes displayed cyanophycin synthesis activity when provided with the (β-Asp-Arg)$_3$ primer (Fig. 2A, Sup. Fig. 1D). However, they exhibited substantial differences in V_{max} and in lag time before detectable cyanophycin synthesis; *C. elsteri* CphA2 did not display any activity under the conditions tested.

We next asked if any of the new CphA2 enzymes could synthesize cyanophycin in the absence of primer. Although *G. citriformis* CphA2 had shown no primer independent activity (Fig. 1C), primer-independent activity has been reported for one CphA1 enzyme[20] and, at very high substrate concentration, for *A. variabilis* and *Cyanothece sp*. CphA2[18]. We performed assays with the purified CphA2 enzymes using 2 mM β-Asp-Arg and no primer (Fig. 2B, Sup. Fig. 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 (Sup. Fig. 1E). Interestingly, 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 V_{max} yet ranked fifth for the primer-dependent V_{max} (Fig. 2A, B).

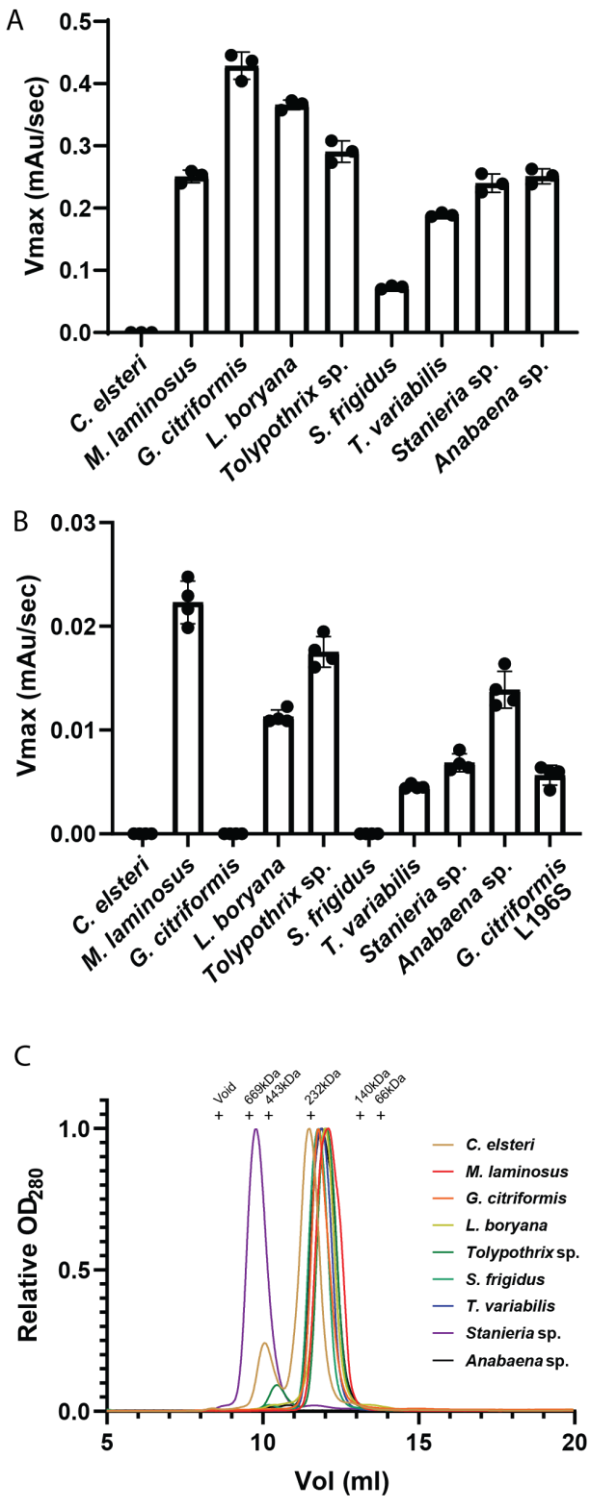

Figure 2. Activity and size exclusion profile of nine CphA2 enzymes. (A) V_{max} values for cyanophycin synthesis by nine different CphA2 enzymes in the presence of 50 µM (β-Asp-Arg)$_3$ primer. (B) V_{max} values for cyanophycin synthesis by the nine CphA2 enzymes in the absence of primer. (C) SEC chromatograms of the 9 homologs. While most displayed a major peak corresponding to ~156-186kDa, *C. elsteri* had two peaks corresponding to ~217kDa and ~460kDa, and *Stanieria* sp. eluted as one peak with a molecular mass corresponding to ~536kDa, showing that some homologs form oligomers other than dimers in solution. All peaks were normalized to the maximal peak height.

Oligomeric state of CphA2 enzymes

Oligomerization of CphA2 would be expected, as CphA1 typically exists as a tetramer[2], while many ATP-grasp domain enzymes, to which the G domain of CphA2 is evolutionarily related, form dimers (e.g. D-Ala-D-Ala ligase[24]) or tetramers (e.g. glutathione synthetase[25]). 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[18]. To determine whether CphA2 has a conserved oligomerization state, we performed SEC experiments with the 9 enzymes we purified in this study (Fig. 2C). Although the molecular mass of each CphA2 protomer is very similar (71-74 kDa), there were differences in the SEC elution profiles. Seven of the enzymes eluted as species with a molecular mass in the range 156-186 kDa. Although these masses are somewhat higher than the expected ~146 kDa for a dimer, higher oligomeric states appear most unlikely for these variants. In contrast, *C. elsteri* CphA2, which was 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. 2C).

The crystal structure of CphA2

To understand how CphA2 recognizes and extends cyanophycin chains, and to help explain the variable primer-dependence and oligomerization states we observed, we sought to determine crystal structures 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. 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 with good statistics at 3.0 Å resolution (Supplementary Table 1).

CphA2 is, as expected, a three-domain protein, consisting of an N domain (residues 1-143), a G domain (residues 144-470), and an M domain (residues 471-616). The N domain is nestled between the G and M domains (Fig. 3A). The protein is dimeric, consistent with the SEC results for *G. citriformis* CphA2 and most other CphA2s (Fig. 2C, Fig. 3B). The asymmetric unit contains one CphA2 protomer (Fig. 3A), with a symmetry mate completing the dimer (Fig. 3B). The dimer interface, which is mainly formed by two G domain helices (residues 189-213) and a loop (166-170), buries 1526 $Å^2$ of surface area.

The structure of CphA2 shows clear similarities to, and differences from, CphA1[2]. Superimposition of CphA1 and CphA2 protomers shows the domains to be arranged in a similar way, with the N domain shifted by ~24 degrees (Fig. 3C). CphA2 lacks a portion of CphA1's M_{core} (residues 644-723 in *Synechocystis sp.* UTEX2470 CphA1 (*Su*CphA1)[2]) and the entire M_{lid} lobe (*Su*CphA1 residues 724 to the C terminus). As a consequence of these deletions, the ATP binding site of the M domain is completely missing. In addition,

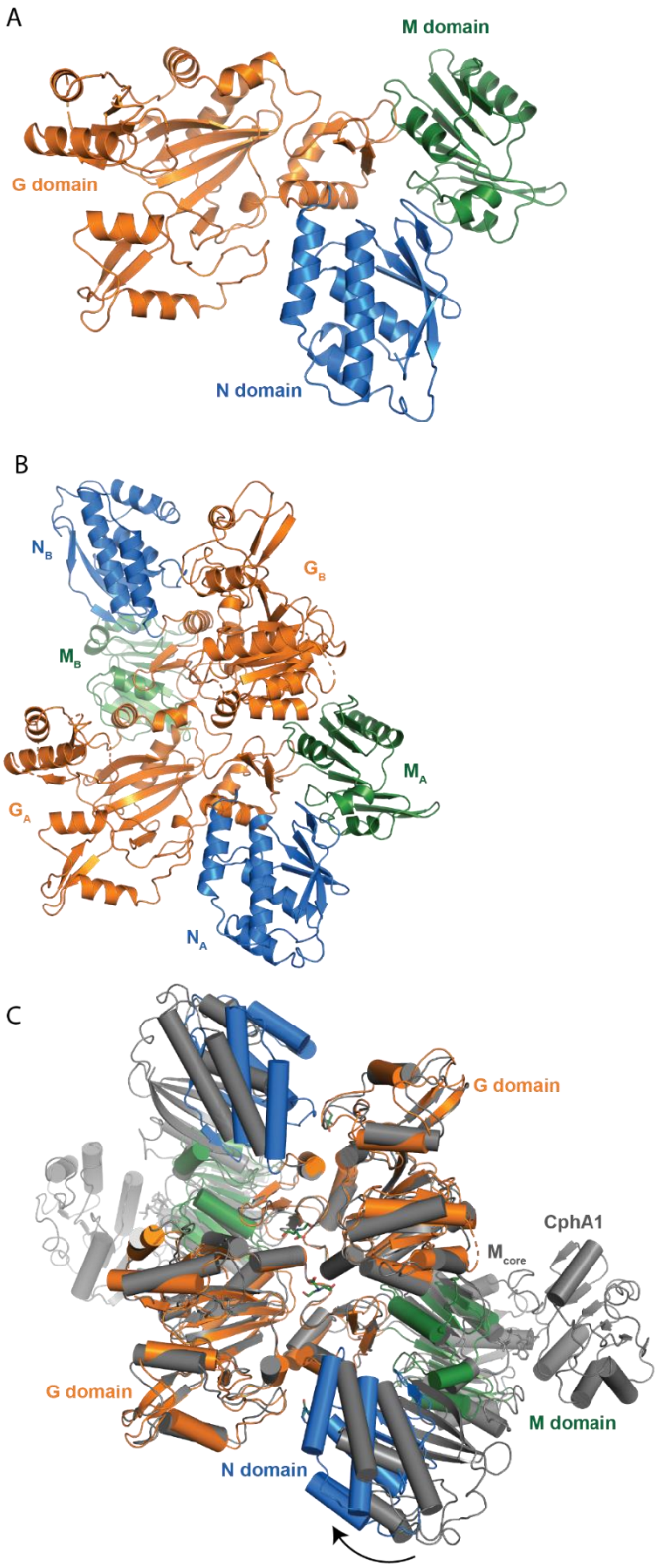

Figure 3. The crystal structure of *G. citriformis* CphA2. (A) View of the CphA2 protomer, composed of 3 domains: the N domain (blue), G domain (orange) and M domain (green). (B) View of the biological dimer of CphA2. (C) CphA2 is similar in architecture to the constituent dimer of a tetrameric CphA1.

many residues that are important for cyanophycin binding to the M domain of CphA1 (for example T538, E533, R561 and S542 in *Su*CphA1) are not conserved in CphA2. This is consistent with CphA2 requiring only one active site, in the G domain, for dipeptide polymerization, whereas CphA1 requires a second active site in the M domain to ligate Arg to the main-chain Asp residue. The dimer interface we observe in CphA2 is similar to that in CphA1, which is also largely formed by the respective G domains. However, both possible tetramerization interfaces observed in CphA1 enzymes[2] are missing in CphA2: CphA1 enzymes have been observed to adopt two different tetramer arrangements, with both tetramer interfaces involving residues in the M domain. Because its M domain is truncated, CphA2 lacks these residues, and is unable to form these contacts.

Structure and mutation of the N domain

Since the CphA2 structure confirms that the M domain was drastically altered and inactivated over the course of evolution, we focused further analysis on the contributions of the other domains to polymerase activity. The N domain of CphA2 has the same overall fold to that of

CphA1 N domains[2]. 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. 4A, B). In CphA2, the two long anti-parallel helices are α_a (residues 103-122) and α_b (residues 128-141) (Fig. 4A). Unlike the corresponding helices in CphA1 N domains, the surface of CphA2 α_a and α_b does not contain many conserved charged residues (Fig. 4C). The conserved charged patches in CphA1 were shown to be important for binding the growing cyanophycin polymer and their mutation drastically decreased CphA1 activity[2], so we 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 α_a and R131, K134 and E137 on α_b (Fig. 4B). Of these, only D115 seems to be conserved (Fig. 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 wild-type (WT) CphA2, while CphA2$\alpha_a$mut displayed substantially reduced activity (Fig. 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. 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 α_a to obtain CphA2 α_a variants that oligomerized normally (Fig. 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. 4D). Differential scanning fluorimetry (DSF) showed that all enzymes have similar melting temperatures, suggesting no significant difference in stability (Sup. Fig. 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 (CphA2ΔN) did not produce protein under the

conditions used for WT expression. Conversely, constructs including only the N domains from these enzymes yielded well behaved, soluble proteins (Sup. Fig. 2B).

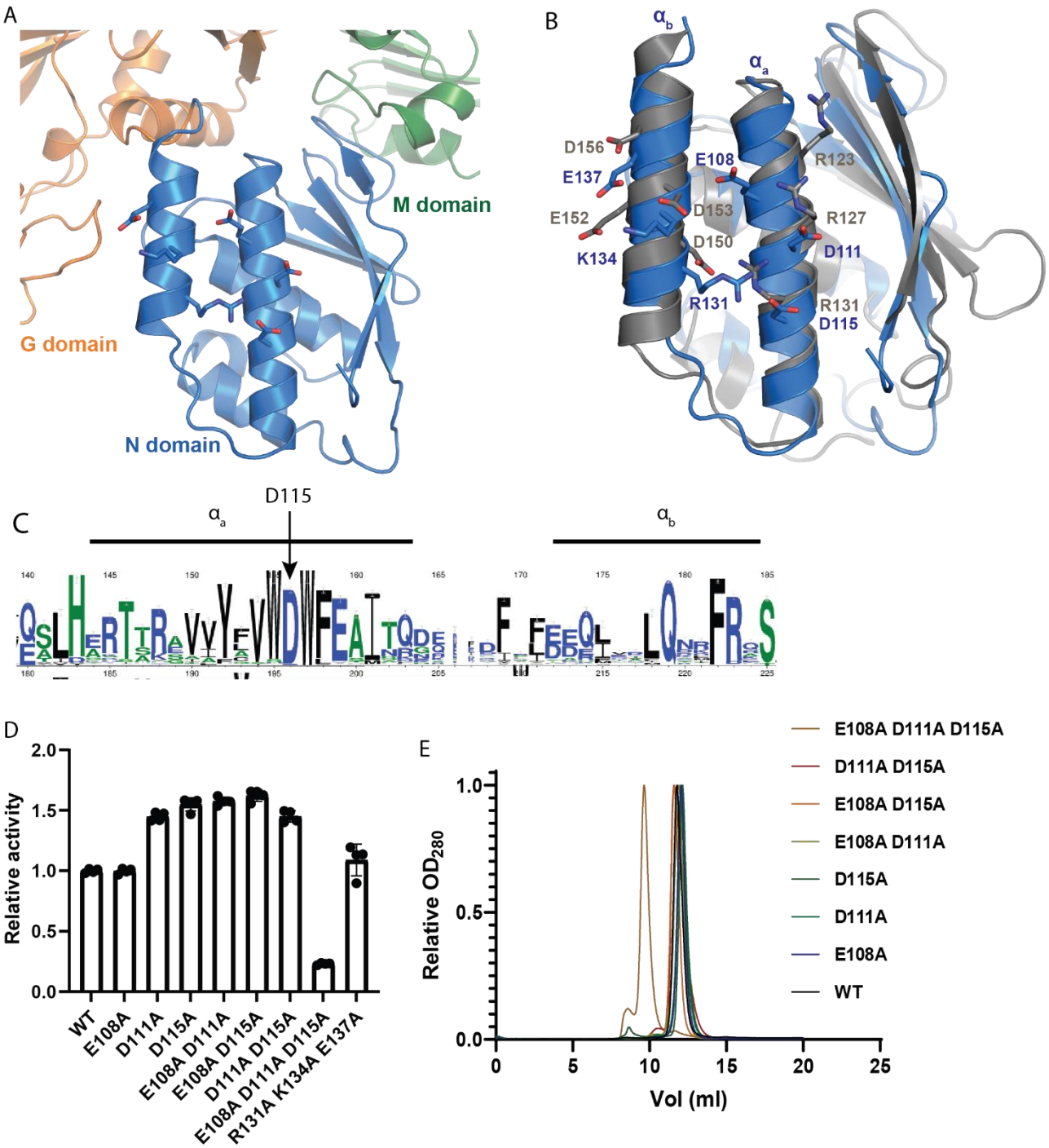

Figure 4. The N domain of CphA2. (A) The CphA2 N domain has two long antiparallel helices, α_a and α_b, supported by a central β-sheet and smaller helices. Several non-conserved charged residues on the surface of these two helices are highlighted. (B) Overlay of CphA2 (blue) and *Su*CphA1 (gray) N domains. The two domains have the same β-β-α-α-α-β-β-α-α fold. The conserved charged residues of CphA1 and non-conserved charged residues of α_a and α_b CphA2 are highlighted. (C) Weblogo[40] showing conservation of CphA2 α_a and α_b. Among the charged residues on the surface of these helices in *G. citridormis* CphA2, only D115 is clearly conserved. (D) Activity assays of *G. citriformis* N domain mutants. The triple-mutant R131A K134A E137A displayed similar activity to that of WT. The triple mutant E108A D111A D115A displayed much lower activity rate. Additionally, every mutant containing either D111A or D115A displayed a higher activity rate than WT. All 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 mutants migrated similarly to WT, except for the triple-mutant E108A D111A D115A, which had a peak corresponding to ~573kDa, close to the size of an octamer (584kDa). All peaks were normalized to the maximal peak height.

Structure and mutation of the G domain

The G domain of CphA2 is composed of a main body (G_{core}, residues 144-218, 284-307, 383-470) and two flexible lobes: G_{lid} (residuse 219-283) and G_{omega} (residues 308-382) (Fig. 5A). G_{lid} does not make crystal contacts and has poorer electron density and higher B-factors than other parts of the enzyme. This G_{lid} flexibility is conserved among ATP-grasp enzymes and is believed to be important for activity[26]. G_{lid} contains the flexible P-loop present in all ATP-grasp enzymes[25, 27] (Fig. 5A), and the conserved H247, which is also present in most CphA1s, is centrally positioned in the CphA2 P-loop. G_{omega} has only been previously observed in the structures of CphA1 and a bifunctional glutathione synthetase[28], and contains the "large loop"[29] typical for ATP-grasp enzymes (Fig. 5A). The density and B-factors of G_{omega} also suggest the presence of some flexibility, although not as extensive as in G_{lid}.

G_{omega} and its large loop, are thought to be important for binding and recognition of the incoming substrate[30]. 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. 5B), in accordance with differing substrate identity (Asp vs β-Asp-Arg; Fig 1A, Sup. Fig. 4A). The large loop forms a putative substrate binding pocket near the G domain active site that is capped by the adjacent loop (Fig. 5B). Mutation to alanine of the highly conserved T335, T337 or S338 all lead to large decreases in activity (Fig. 5C), consistent with a role in β-Asp-Arg recognition for this region of G_{omega}. We then asked if transplanting G_{omega} from CphA2 onto a CphA1 would confer β-Asp-Arg recognition capabilities to the CphA1 enzyme. We created a chimeric protein containing G_{omega} from *G. citriformis* CphA2 and all other (sub)domains from *Tatumella morbirosei* DSM23827 CphA1. However, while the resulting chimeric protein, *Tm*$CphA1_{Gcit\text{-}omega}$, could be produced and purified in soluble form, 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 either $(\beta\text{-Asp-Arg})_3$ or $(\beta\text{-Asp-Arg})_4$ did not show extra density near the active site that we could confidently attribute to cyanophycin. This is likely because

tartrate was present at very high concentration (200 mM) 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 (Sup. Fig. 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 (β-Asp-Arg)$_8$-NH$_2$ (Fig. 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_{core}. Notably, 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* CphA2 should thus bind the reactive C-terminal carboxylate of cyanophycin through interaction with R292 (*Su*CphA1 R309), the first β-Asp-Arg dipeptide with T200 and N435 (*Su*CphA1 C218 and N452), the second β-Asp-Arg dipeptide with S148 and T186 (*Su*CphA1 S166 and T204), and the third with S149 (*Su*CphA1 T167)(Fig. 5D). We tested these predictions by mutating R140, S148, T186, D197 or R292 to alanine, and found that the mutants displayed reduced (R140A, T186A) or no (S148A, D197A, R292A) activity (Fig. 5E), supporting the hypothesis that CphA2 binds cyanophycin much like CphA1 does. The deleterious effects of increasing ionic strength (Fig. 1E, F) underscore the importance of electrostatic interactions for substrate binding to the enzyme.

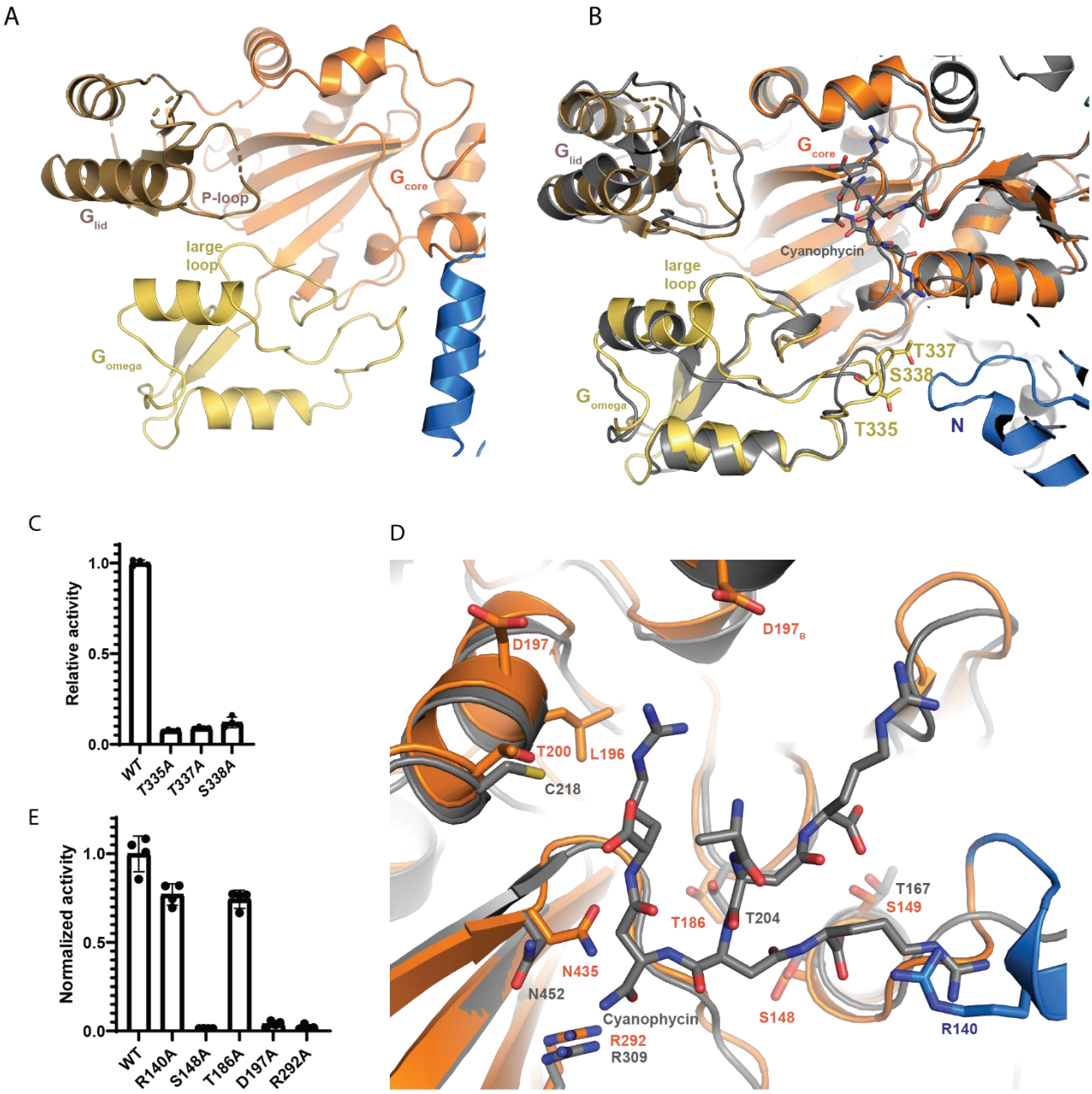

Figure 5. The G domain of CphA2. (A) The structure of the G domain, colored by subdomains: G_{core} (orange), G_{lid} (brown) and G_{omega} (yellow). (B) Overlay of CphA2 (colored) and cyanophycin-bound *Su*CphA1 (gray) G domains. The overall structures are very similar, with the main differences arising from small conformational variation between the flexible lobes G_{lid} and G_{omega}, which are likely influenced by crystal packing. Conserved loop residues which could are shown as yellow sticks. (C) Mutation of those conserved loop residues decreases 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 of both enzymes are highlighted. These residues are very similar in CphA1 and CphA2, suggesting that both enzymes bind cyanophycin polymer / primer in a similar way. (E) Activity assays of *G. citriformis* CphA2 G domain mutants. Mutations of several conserved residues that potentially bind cyanophycin resulted in decreased or abolished activity.

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. 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 would likely contact a cyanophycin 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 the binding site of what would be the first dipeptide of a cyanophycin primer (Fig. 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 β-Asp-Arg dipeptide, which could stabilize binding of that β-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. 2B), demonstrating the importance of this interaction for the binding of β-Asp-Arg as the carboxylate donor in the absence of longer primers.

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[31]. The elongated polymer likely then dissociates from this product position and re-associates, placing the newly added β-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 orientations and 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 of interactions 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 other homologs that could be phased but not refined to an R_{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 β-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_{omega} and G_{lid} could influence V_{max} in CphA2 enzymes. Nonetheless, that a single mutation in a primer-binding residue was sufficient to allow *G. citriformis* CphA2 to synthesize CphA2 in the absence of primer, confirms that affinity for β-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.

Methods

Cloning, protein expression and purification

The gene encoding CphA2 from *Anabaena sp.* UTEX2576 (from genome NC_003272.1, encoding protein WP_010994749.1) was cloned from genomic DNA (purchased from the University of Texas (UTEX) collection). Other *cphA2* genes were codon optimized for expression in *E. coli* and synthesized by BioBasic Canada (*Stenomitos frigidus* ULC18, *Stanieria* sp. NIES3757 and *Tolypothrix* sp. NIES4075) or by the US

Department of Energy Joint Genome Institute (*Anabaena variabilis* PCC7120, *Calothrix elsteri* CCALA953, *Gloeothece citriformis* PCC7424, *Leptolyngbya boryana* NIES2135 and *Mastigocladus laminosus* UU774). All genes were cloned into pJ411-derived plasmids encoding C-terminal tobacco etch virus (TEV) protease recognition sites and an 8xHis affinity tag by transforming DH5-α *E. coli* cells with PCR fragments containing overlapping ends. All proteins were expressed in BL21(DE3) *E. coli*. Cells were grown in TB media supplemented with 150 µg/ml kanamycin at 37 °C until they reached an OD_{600} of ~1. At this point, the culture temperature was lowered to 18 °C, protein expression was induced with 0.2mM isopropyl β-d-1-thiogalactopyranoside (IPTG) and the cultures were then incubated for ~20 hours before harvesting. All protein purification steps were carried out at 4 °C. Following centrifugation, the cells were resuspended in buffer A (250 mM NaCl, 50 mM Tris pH 8, 10 mM imidazole, 2 mM β-mercaptoethanol) supplemented with a few crystals of lysozyme and DNAseI and lysed by sonication. The lysate was clarified by centrifugation at 40,000 g, and then loaded onto a 5 ml HisTrap HP column (Cytiva), washed with at least 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 ratio by mass) for removal of the 8xHis tag while being dialyzed overnight against buffer D (250 mM NaCl, 20 mM Tris-HCl pH 8, 3 mM β-mercaptoethanol) prior to being applied to a HisTrap column again. The flow through 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). Fractions with the highest protein purity were pooled and concentrated to 20 mg/ml, and following the addition of glycerol to 10% v/v the protein was flash frozen in liquid nitrogen and stored at -80 °C until use.

Protein crystallization, data collection and structure solution

Protein crystals were grown using the sitting drop method at 22 °C. *G. citriformis* CphA2 in buffer E at a concentration of 12.5 mg/ml was mixed with 1:5 v/v of 15 mM FOS-choline-12, resulting in a protein concentration of 10 mg/ml. Samples (2 µl) were mixed with 1.6 µl of well solution (0.1 M Tris-bicine pH 8.5, 8.8% w/v PEG8000, 17.6% v/v ethylene glycol, 60 mM $MgCl_2$, 200 mM Na/K tartrate, 5 mM betaine and 3%

v/v 2,2,2-trifluoroethanol) and 0.4 µl 100 mM $NiCl_2$, and allowed to equilibrate against well solution. Once the crystals reached their final size, they 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 $MgCl_2$, 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[32]. The structure was solved using an ensemble of the 3 known CphA1 structures[2] as a search model using PHASER implemented in CCP4i2[33], and refined using Rosetta[34], LORESTR[35], Phenix refine[36] and Coot[37]. Figures were created using PyMOL.

Size exclusion chromatography

SEC experiments were performed using a Superdex 200 increase 10/300 gl column (Cytiva) equilibrated in buffer E, running at 0.5 ml/min and an injection volume of 150 µl. The column's elution volumes were calibrated using an HMW standard (Cytiva).

CphA2 activity assays

Reactions contained 2 µM purified CphA2, 100 mM HEPES pH 8.2, 40 mM KCl, 10 mM $MgCl_2$, 2 mM each of β-Asp-Arg and ATP, and 50 µM synthetic cyanophycin primer. NaCl was also added in some experiments as indicated. The reactions were carried out in quadruplicate at 23 °C, in 96-well plates with a total reaction volume of 100 µl. OD_{600} was monitored using a SpectraMax Paradigm spectrophotometer (Molecular Devices), with 5 second linear shaking between reads. Typical reaction times were about 1 hour. 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 2nd order polynomial to reduce noise in measurements.

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[2, 38-39].

Differential scanning fluorimetry (DSF)

DSF assays were performed with 0.5mg/ml protein in buffer E and 5x Sypro™ Orange in a total reaction volume of 20 µl. The temperature was ramped from 5 °C to 95 °C over 2 hours and data were collected using a One Step Plus RT-PCR (Applied Biosystems).

Acknowledgements

We thank all the members of the Schmeing lab for important advice and ongoing discussions on this project and synchrotron staff Shaun Labiuk, Kathryn Janzen and Kiran 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.

Funding sources

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. Gene syntheses were 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. Sharon, I.; Haque, A. S.; Grogg, M.; Lahiri, I.; Seebach, D.; Leschziner, A. E.; Hilvert, D.; Schmeing, T. M., Structures and function of the amino acid polymerase cyanophycin synthetase. *Nat Chem Biol* **2021,** 17, 1101-1110.
3. Fuser, G.; Steinbuchel, A., Analysis of genome sequences for genes of cyanophycin metabolism: identifying putative cyanophycin metabolizing prokaryotes. *Macromol Biosci* **2007,** 7, 278-296.
4. Simon, R. D., The biosynthesis of multi-L-arginyl-poly(L-aspartic acid) in the filamentous cyanobacterium Anabaena cylindrica. *Biochim Biophys Acta* **1976,** 422, 407-418.
5. 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* **1996,** 142, 611-622.
6. Liang, B.; Wu, T. D.; Sun, H. J.; Vali, H.; Guerquin-Kern, J. L.; Wang, C. H.; Bosak, T., Cyanophycin mediates the accumulation and storage of fixed carbon in non-heterocystous filamentous cyanobacteria from coniform mats. *PLoS One* **2014,** 9, e88142.
7. Wingard, L. L.; Miller, S. R.; Sellker, J. M.; Stenn, E.; Allen, M. M.; Wood, A. M., Cyanophycin production in a phycoerythrin-containing marine synechococcus strain of unusual phylogenetic affinity. *Appl Environ Microbiol* **2002,** 68, 1772-1777.
8. Watzer, B.; Forchhammer, K., Cyanophycin synthesis optimizes nitrogen utilization in the unicellular cyanobacterium Synechocystis sp. PCC 6803. *Appl Environ Microbiol* **2018**.
9. Fay, P., Oxygen relations of nitrogen fixation in cyanobacteria. *Microbiol Rev* **1992,** 56, 340-373.
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* **2014,** 111, 3823-3828.
11. Li, H.; Sherman, D. M.; Bao, S.; Sherman, L. A., Pattern of cyanophycin accumulation in nitrogen-fixing and non-nitrogen-fixing cyanobacteria. *Arch Microbiol* **2001,** 176, 9-18.
12. 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* **2020,** 20, e2000132.
13. Mooibroek, H.; Oosterhuis, N.; Giuseppin, M.; Toonen, M.; Franssen, H.; Scott, E.; Sanders, J.; Steinbuchel, A., Assessment of technological options and economical feasibility for cyanophycin biopolymer and high-value amino acid production. *Appl Microbiol Biotechnol* **2007,** 77, 257-267.
14. Nausch, H.; Hausmann, T.; Ponndorf, D.; Huhns, M.; Hoedtke, S.; Wolf, P.; Zeyner, A.; Broer, I., Tobacco as platform for a commercial production of cyanophycin. *N Biotechnol* **2016,** 33, 842-851.
15. Steinle, A.; Witthoff, S.; Krause, J. P.; Steinbuchel, A., Establishment of cyanophycin biosynthesis in Pichia pastoris and optimization by use of engineered cyanophycin synthetases. *Appl Environ Microbiol* **2010,** 76, 1062-1070.
16. Berg, H.; Ziegler, K.; Piotukh, K.; Baier, K.; Lockau, W.; Volkmer-Engert, R., 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* **2000,** 267, 5561-5570.
17. 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. *J Biol Chem* **2004,** 279, 11582-11592.
18. Klemke, F.; Nurnberg, D. J.; Ziegler, K.; Beyer, G.; Kahmann, U.; Lockau, W.; Volkmer, T., CphA2 is a novel type of cyanophycin synthetase in N2-fixing cyanobacteria. *Microbiology* **2016,** 162, 526-536.
19. 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* **2002,** 277, 25096-25105.
20. Arai, T.; Kino, K., A cyanophycin synthetase from Thermosynechococcus elongatus BP-1 catalyzes primer-independent cyanophycin synthesis. *Appl Microbiol Biotechnol* **2008,** 81, 69-78.

21. 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. *Appl Environ Microbiol* **2002,** 68, 93-101.
22. Krehenbrink, M.; Oppermann-Sanio, F.-B.; Steinbuchel, 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* **2002,** 177, 371-380.
23. 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* **2006,** 103, 5442-5447.
24. Kitamura, Y.; Ebihara, A.; Agari, Y.; Shinkai, A.; Hirotsu, K.; Kuramitsu, S., Structure of D-alanine-D-alanine ligase from Thermus thermophilus HB8: cumulative conformational change and enzyme-ligand interactions. *Acta Crystallogr D Biol Crystallogr* **2009,** 65, 1098-1106.
25. Yamaguchi, H.; Kato, H.; Hata, Y.; Nishioka, T.; Kimura, A.; Oda, J.; Katsube, Y., Three-dimensional structure of the glutathione synthetase from Escherichia coli B at 2.0 A resolution. *J Mol Biol* **1993,** 229, 1083-1100.
26. Li, H.; Fast, W.; Benkovic, S. J., Structural and functional modularity of proteins in the de novo purine biosynthetic pathway. *Protein Sci* **2009,** 18, 881-892.
27. 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* **1996,** 35, 11967-11974.
28. 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* **2012,** 416, 486-494.
29. Hibi, T.; Nishioka, T.; Kato, H.; Tanizawa, K.; Fukui, T.; Katsube, Y.; Oda, J., Structure of the multifunctional loops in the nonclassical ATP-binding fold of glutathione synthetase. *Nat Struct Biol* **1996,** 3, 16-18.
30. 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* **2009,** 21, 3450-3458.
31. Fawaz, M. V.; Topper, M. E.; Firestine, S. M., The ATP-grasp enzymes. *Bioorg Chem* **2011,** 39, 185-191.
32. Beilsten-Edmands, J.; Winter, G.; Gildea, R.; Parkhurst, J.; Waterman, D.; Evans, G., Scaling diffraction data in the DIALS software package: algorithms and new approaches for multi-crystal scaling. *Acta Crystallogr D Struct Biol* **2020,** 76, 385-399.
33. Potterton, L.; Agirre, J.; Ballard, C.; Cowtan, K.; Dodson, E.; Evans, P. R.; Jenkins, H. T.; Keegan, R.; Krissinel, E.; Stevenson, K.; Lebedev, A.; McNicholas, S. J.; Nicholls, R. A.; Noble, M.; Pannu, N. S.; Roth, C.; Sheldrick, G.; Skubak, P.; Turkenburg, J.; Uski, V.; von Delft, F.; Waterman, D.; Wilson, K.; Winn, M.; Wojdyr, M., CCP4i2: the new graphical user interface to the CCP4 program suite. *Acta Crystallogr D Struct Biol* **2018,** 74, 68-84.
34. Song, Y.; DiMaio, F.; Wang, R. Y.; Kim, D.; Miles, C.; Brunette, T.; Thompson, J.; Baker, D., High-resolution comparative modeling with RosettaCM. *Structure* **2013,** 21, 1735-1742.
35. Kovalevskiy, O.; Nicholls, R. A.; Murshudov, G. N., Automated refinement of macromolecular structures at low resolution using prior information. *Acta Crystallogr D Struct Biol* **2016,** 72, 1149-1161.
36. Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Echols, N.; Headd, J. J.; Hung, L. W.; Jain, S.; Kapral, G. J.; Grosse Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R. D.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., The Phenix software for automated determination of macromolecular structures. *Methods* **2011,** 55, 94-106.
37. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **2004,** 60, 2126-2132.
38. Grogg, M.; Hilvert, D.; Ebert, M. O.; Beck, A. K.; Seebach, D.; Kurth, F.; Dittrich, P. S.; Sparr, C.; Wittlin, S.; Rottmann, M.; Maser, P., 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* **2018,** 101.

39. Grogg, M.; Hilvert, D.; Beck, A. K.; Seebach, D., Syntheses of Cyanophycin Segments for Investigations of Cell-Penetration. *Synthesis* **2019,** 51, 31-39.
40. Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E., WebLogo: a sequence logo generator. *Genome Res* **2004,** 14, 1188-1190.