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

Itai Sharon, Sharon Pinus, Marcel Grogg, Nicolas Moitessier, Donald Hilvert & T. Martin Schmeing

Cyanophycin is a nitrogen reserve biopolymer in many bacteria that has promising industrial applications. Made by cyanophycin synthetase 1 (CphA1), it has a poly-L-Asp backbone with L-Arg residues attached to each aspartate sidechain. CphA1s are thought to typically require existing segments of cyanophycin to act as primers for cyanophycin polymerization. In this study, we show that most CphA1s will not require exogenous primers and discover the surprising cause of primer independence: CphA1 can make minute quantities of cyanophycin without primer, and an unexpected, cryptic metallopeptidase-like active site in the N-terminal domain of many CphA1s digests these into primers, solving the problem of primer availability. We present co-complex cryo-EM structures, make mutations that transition CphA1s between primer dependence and independence, and demonstrate that primer dependence can be a limiting factor for cyanophycin production in heterologous hosts. In CphA1, domains with opposite catalytic activities combine into a remarkable, self-sufficient, biosynthetic nanomachine.
Cyanophycin is a natural biopolymer discovered over 130 years ago as large, insoluble granules within cyanobacterial cells. Cyanophycin chains consist of a poly-L-Asp backbone with L-Arg residues attached to each Asp side chain through isopeptide bonds. Chains are typically ~80–400 β-Asp-Arg dipeptides in length \((β\text{-Asp-Arg})_{40–400}\). The high nitrogen content and inert nature of cyanophycin make it ideal for storing fixed nitrogen, as well as carbon and energy.

Cyanophycin is especially useful for nitrogen-fixing cyanobacteria that separate aerobic photosynthesis and anaerobic nitrogen fixation either spatially or temporally, has been shown to enhance the efficiency of nitrogen assimilation in non-diazotrophic strains, and is also produced by many other bacteria across the kingdom. Cyanophycin has promising potential commercial applications, from use as bandage material to providing a source of poly-Asp, a biodegradable antiscalant, water softener, and super-swelling material. Nevertheless, production yields of cyanophycin are currently too low for commercial viability and many studies have sought to increase them.

Cyanophycin synthetase 1 (CphA1) catalyzes polymerization of Asp and Arg into cyanophycin in two iterative, ATP-dependent reactions (Fig. 1a). In the first reaction, the C-terminus of a cyanophycin chain is activated by phosphorylation and then reacts with arginine to form an iso-Asp side chain. All CphA1 enzymes also contain an N-terminal domain active sites. Our study also visualized two separate tetrameric architectures for CphA1.

CphA1s have most often been described as possessing primer-dependent activity. It is widely accepted that primer-dependent CphA1s cannot synthesize cyanophycin de novo from only Asp, Arg, and ATP, but require existing chains to extend. Only CphA1 from *Thermosynechococcus elongatus* BP-1 has been shown to display robust primer-independent activity. Primer-dependent CphA1s are known to use long cyanophycin chains, trimer dipeptide segments \((β\text{-Asp-Arg})_n\), and (albeit with low efficiency) other biomaterials as primers, but characteristics of minimal and optimal primers are not established. It was also completely unknown what determines whether a CphA1 enzyme is primer dependent or primer independent. When heterologously expressed, CphA1 is catalytically active and produces cyanophycin within host cells, so understanding the nature of primers and primer independence could be important for bioproduction yields from these hosts.

Here, we report the discovery that the key to primer independence is a cryptic metallopeptidase-like active site in the N domain. We use a combination of cryo-EM, mass spectrometry, mutagenesis, and biochemical assays to characterize and identify the nature of the active site. Our studies provide a compelling mechanism for both primer-independent and primer-dependent CphA1 activity.

---

**Fig. 1** CphA1 structure and activity. **a** Schematic diagram of the biosynthetic reactions catalyzed by the G and M domains of CphA1. The overall structure of tetrameric CphA1 from *Synechocystis* sp. UTEX2470 (SuCphA1, PDB code 7LG5). ATP molecules mark the positions of the G (orange) and M (green) domain active sites. The N domain is colored in blue. **b** Cyanophycin biosynthesis rate comparison of synthesis by SuCphA1 and TmCphA1 with and without primer. TmCphA1 is completely inactive in the absence of primer. n = 4 independent experiments. Data are presented as individual measurements and mean value, error bars represent SD values. **d** Activity levels of TmCphA1 in the presence of various cyanophycin primers: 1mer \((β\text{-Asp-Arg})\), 1.5mer \((β\text{-Asp-Arg})\)-Asp, 2mer \((β\text{-Asp-Arg})_2\), 3mer \((β\text{-Asp-Arg})_3\), 4mer \((β\text{-Asp-Arg})_4\). n = 4 independent experiments. Data are presented as individual measurements and mean value, error bars represent SD values.
manipulate primer-independent CphA1 activity. The results show how the N domain enables biosynthesis by CphA1 without exogenous primers and demonstrate the implications of primer independence for in vivo cyanophycin production.

**Results**

**Primer dependence of CphA1 enzymes.** Cyanophycin synthetases from *Synechocystis* sp. UTEX2470 (SuCphA1) and *Tatamella morbireosi* DSM23827 (TmCphA1) show high activity in the presence of a (β-Asp-Arg)3 primer11,18,19 (Fig. 1c). Notably, in the absence of a β-Asp-Arg, the observed rate increases progressively with addition of (β-Asp-Arg)2, and saw a similar activity with (β-Asp-Arg)4 (Fig. 1d). We then tested (β-Asp-Arg)-Asp, the CphA1 biosynthetic intermediate between β-Asp-Arg and (β-Asp-Arg)2, and observed a lag phase of ~15 min, followed by robust cyanophycin synthesis. This result is surprising, because only one other CphA1 enzyme has ever been reported to be primer independent22. In contrast to SuCphA1, no primer-independent activity is observed for TmCphA1, even at high protein concentrations and incubation with substrates over several days (Supplementary Fig. 1a). These results led us to ask what properties of the primer and enzyme control primer-dependent and -independent cyanophycin synthesis.

We first sought to define the minimal length of cyanophycin that can serve as a primer for CphA1 enzymes (Fig. 1d). TmCphA1 could not perform cyanophycin synthesis in the absence of β-Asp-Arg, but the synthesis was observed with (β-Asp-Arg)2, and the observed rate increases progressively in reactions with (β-Asp-Arg)3 and (β-Asp-Arg)4 (Fig. 1d). We then tested (β-Asp-Arg)-Asp, the CphA1 biosynthetic intermediate between β-Asp-Arg and (β-Asp-Arg)2, and saw a similar activity as with (β-Asp-Arg)2 (Fig. 1d). Assays with the primer-independent SuCphA1 showed similar results: Addition of (β-Asp-Arg)-Asp and longer fragments shortens the lag phase, and (Asp-Arg)3 and (β-Asp-Arg), give the highest rates of synthesis (Supplementary Fig. 1b). Together, these results define (β-Asp-Arg)-Asp as the minimal primer for CphA1.

**CphA1 primer independence does not depend on the G or M domain activity.** We next sought to identify the source of primer independence by comparing primer-independent SuCphA1 to primer-dependent TmCphA1. We hypothesized that affinity differences for short cyanophycin segments at the G and/or M domain active sites might dictate primer (in)dependence 27. SuCphA1 has several polar residues at the G and M domain active sites that could increase its affinity to nascent cyanophycin segments relative to TmCphA1, which has hydrophobic residues at the analogous positions11,19 (Supplementary Fig. 2a, b). However, experiments with 15 different mutants of TmCphA1 in which one or more of these SuCphA1 hydrophilic residues were introduced into TmCphA1’s G domain, M domain, or G and M domains did not lead to primer independent activity (Supplementary Fig. 2c).

We then reasoned that if the G or M domains of SuCphA1 were individually responsible for primer independence, providing the necessary domain in trans to a reaction including TmCphA1 would lead to primer-independent synthesis. Therefore, we prepared SuCphA1 with inactivating mutations31 in the G domain (H267A), the M domain (D585A H586A), or in both (H267A; D585A H586A) (Supplementary Fig. 2a, b). To our surprise, all three of these constructs enabled robust primer-independent activity when added to TmCphA1 (Fig. 2a). This result suggested that SuCphA1 has a third, previously unsuspected active site that is responsible for primer independence.

**A cryptic N domain active site responsible for primer independence.** Because our results indicated that the G and M domain active sites are not important for primer independence, we investigated whether the N domain is. We created a chimeric CphA1 (TmCphA1SuN) comprising the N domain of SuCphA1 and the G and M domains of TmCphA1. Intriguingly, the TmCphA1SuN chimera displayed robust primer-independent activity, suggesting that the N domain is responsible for primer independence (Fig. 2b). To verify this conclusion, we added the isolated SuCphA1 N domain in trans to a reaction including TmCphA1. Again, primer-independent activity was observed, proving that the N domain is vital for primer independence in CphA1 (Fig. 2b and Supplementary Fig. 2d).

A catalytic role for the N domain has never been suggested before, so this discovery led us to re-examine the sequence conservation of N domains to search for a cryptic active site. No putative catalytic residues are 100% conserved, which is expected since some proportion of CphA1 enzymes will be primer dependent like TmCphA1. However, an HxxEH motif28 can be seen by careful inspection of sequence alignments (Fig. 2c and Supplementary Fig. 3a). The N domain of our existing SuCphA1 structures11 shows the motif residues H79, E83, and H83 cluster together with C59, and are surrounded by several conserved arginine and histidine residues (Fig. 2d). All four residues of a Cx3-HxxEH motif are present in the primer-independent SuCphA1 and none of the four are present in the primer-dependent TmCphA1 (Supplementary Fig. 3b). Of CphA1s with a non-redundancy of 70% (nr70), the Cx3-HxxEH motif is fully present in 83% of sequences. Mutations in and near this putative N domain active site (H57A, C59A, R70A, H79A, E82Q29, and R100A) did not greatly affect primer-dependent activity (Supplementary Fig. 3c), but all except R100A reduced or abolished primer-independent activity (Fig. 2e), confirming that this N domain active site is responsible for primer independence. Notably, HxxEH is the active site motif for inverted zinc metallopeptidase28, suggesting the N domain may have peptidase activity important for primer generation from cyanophycin polymer.

The structural basis for the catalytic activity of the N domain. To structurally characterize this cryptic N domain active site and its binding to cyanophycin, we turned to cryo-EM. We determined a structure of the inactivated29 SuCphA1(E82Q) in complex with (β-Asp-Arg)4 at 2.7 Å resolution by cryo-EM (Fig. 3a, b, Supplementary Fig. 4a, b, and Supplementary Table 1). The maps showed clear density for a chain of seven β-Asp-Arg dipeptide residues centered on the conserved region of the N domain. This region harbors a metallopeptidase-like active site29 (Fig. 3c): H79 and H83 from the active site helix (residues 77–92) ligate an ion. We have tentatively assigned this ion as Zn2+ because it is by far the most abundant metal detected in inducibly coupled plasma mass spectrometry (ICP-MS) analyses of SuCphA1 (and not of TmCphA1; Supplementary Table 2), and because of similarities to inverted zinc metallopeptidase28 and peptide deformylase (PDF)20 active sites, both of which can bind Zn2+ tightly. The two histidine side chains bind Zn2+ with their Nε atoms. H83, which forms π-stacking interactions with the conserved F67, also forms a hydrogen bond network with the conserved H57 and E87, stabilizing the tautomeric form in which its Nε lone-pair electrons face the Zn2+ site31. C59, present in a loop region, serves as the third metal ligand, a role typically played by a glutamate, aspartate, or histidine from the metallopeptidase “glutamate helix”29, Q82 (taking the place of the general base E82 in this SuCphA1(E82Q) construct) sits above the metal-binding histidines, as in metallopeptidases. The cyanophycin polymer makes an extensive hydrogen-bonding network with itself and SuCphA1 residues (Fig. 3b), including the backbones of Y14, C59, A96, G97, and T101, and the side chains of E90, R70, Y110, S60 and M domain residues S603 and E607. These interactions position four visualized dipeptides upstream of the zinc ion and three downstream. The main chain carboxyl
The N domain cleaves cyanophycin into primers. The structures suggest that the N domain possesses endo-cyanophycinase activity.

To directly observe this activity, we incubated purified cyanophycin with SuCphA1 and examined the reaction with SDS-PAGE (Supplementary Figs. 5a and 7). A slow but clear decrease in cyanophycinase activity was observed over several days, especially in the molecular weight range of ~15–20 kDa. We next performed mass spectrometry-based cyanophycin cleavage assays. Incubation of SuCphA1 with (β-Asp-Arg)_5-NH₂ over several hours led to the gradual formation of species with mass values corresponding to (β-Asp-Arg)_4-NH₂ and (β-Asp-Arg)_3, (Fig. 4a). Likewise, (β-Asp-Arg)_4-Asn is converted by SuCphA1 to species with masses corresponding to (β-Asp-Arg)_3-Asn and (β-Asp-Arg)_2 (Fig. 4b). Control reactions with primer-dependent TmCphA1 or the N domain SuCphA1(E82Q) variant with (β-Asp-Arg)_5-NH₂ or (β-Asp-Arg)_4-Asn did not result in the appearance of the product peaks (Supplementary Fig. 5b). SuCphA1 cleaved (β-Asp-Arg)_2 into major products (β-Asp-Arg)_4 and (β-Asp-Arg)_3 and minor products (β-Asp-Arg)_5 and (β-Asp-Arg)_6 (Supplementary Fig. 5c–e). Thus, the N domain of SuCphA1 is indeed a cryptic primer-generating endo-cyanophycinase that possesses a low catalytic rate and, at least with the cyanophycin segments used in these experiments, preferentially yields (β-Asp-Arg)_4 fragments as products.
The relative geometry of the CphA1 active sites is important for cyanophycin biosynthesis\textsuperscript{11}, so we interrogated whether it is also important for primer-generating cleavage. We combined the W672A mutation (which forces SuCphA1 to be dimeric instead of tetrameric\textsuperscript{11}) with mutations that abolish the activity of each active site (E82Q = N\textsuperscript{−}; H267A = G\textsuperscript{−}; D585A H586A = M\textsuperscript{−}) and used orthogonal affinity tags to purify different SuCphA1 heterodimers\textsuperscript{11}. SuCphA1 N\textsuperscript{−}G\textsuperscript{+}M\textsuperscript{+}/N\textsuperscript{+}G\textsuperscript{−}M\textsuperscript{−} (with all WT active sites on the same protomer) and SuCphA1 N\textsuperscript{+}G\textsuperscript{+}M\textsuperscript{+}/N\textsuperscript{−}G\textsuperscript{−}M\textsuperscript{−} (with the WT N domain on the opposite protomer as WT G and M domains) have similar activity in the presence of primer (Fig. 4c). However, N\textsuperscript{−}G\textsuperscript{+}M\textsuperscript{+}/N\textsuperscript{+}G\textsuperscript{−}M\textsuperscript{−} is somewhat less active in the absence of primer (Fig. 4d). This suggests that the proximity of the hydrolytic active site to the biosynthetic active sites is beneficial for primer-independent activity. Interestingly, SuCphA1 dimers\textsuperscript{11,19} exhibit the same primer-dependent activity rate as tetramers, but a lower primer-independent activity rate (Fig. 4e), indicating that the tetramer architecture of SuCphA1 is also beneficial for primer-independent activity.

**Effect of primer independence on heterologous cyanophycin synthesis.** Understanding the basis of primer dependence in cyanophycin biosynthesis and obtaining primer-independent and primer-dependent variants of the same CphA1 enzymes allowed us to examine the importance of primer availability for cyanophycin accumulation in a heterologous host. To that end, we separately expressed primer-independent SuCphA1 and primer-dependent SuCphA1(E82Q) in Escherichia coli BL21(DE3) and quantified the amount of polymer produced in vivo by each variant. *E. coli* harboring WT SuCphA1 produced on average 2.3-fold more cyanophycin than the E82Q mutant, measured as milligrams of cyanophycin per liter of growth culture (Table 1). The total wet cell mass was 12% lower for cells expressing the WT enzyme, suggesting they divert more resources to cyanophycin synthesis from cell growth compared to those harboring the primer-dependent mutant. Similarly, in experiments with primer-dependent *Tm*CphA1 and the primer-independent chimera *Tm*CphA1\textsubscript{SuN}, the chimera produced 2-fold more cyanophycin than the primer-dependent WT enzyme, in a lower total wet cell mass (Table 1).
The sequence, structure, and activity data all indicate that the N domain of CphA1 has cryptic metallopeptidase activity. The endo-cyanophycinase activity of the N domain presents an explanation for the primer dependence and independence in CphA1 enzymes: We propose that all CphA1 enzymes possess explanation for the primer dependence and independence in the first steps of cyanophycin synthesis, e.g., ligating Asp and Arg to β-Asp-Arg, and ligating β-Asp-Arg and Asp to (β-Asp-Arg)-Asp. In the absence of primer, the two enzymes display similar primer-dependent activity, but the dimer has lower primer-independent activity.

**Table 1 Heterologous cyanophycin production in E. coli.**

| Enzyme          | Cell pellet (mg) | Polymer (mg) | % of cell mass (w/w) |
|-----------------|-----------------|--------------|----------------------|
| SuCphA1 WT      | 4730 ± 160      | 413 ± 35     | 8.7                  |
| SuCphA1 E82Q    | 5390 ± 140      | 177 ± 28     | 3.2                  |
| TmCphA1 WT      | 5570 ± 380      | 247 ± 18     | 4.4                  |
| TmCphA1Fin      | 4880 ± 170      | 486 ± 52     | 9.9                  |

**Discussion**

The sequence, structure, and activity data all indicate that the N domain of CphA1 has cryptic metallopeptidase activity. The endo-cyanophycinase activity of the N domain presents an explanation for the primer dependence and independence in CphA1 enzymes: We propose that all CphA1 enzymes possess very low levels of true primer-independent activity for the first steps of cyanophycin synthesis, e.g., ligating Asp and Arg to β-Asp-Arg, and ligating β-Asp-Arg and Asp to (β-Asp-Arg)-Asp. In the absence of primer, the two enzymes display similar primer-dependent activity, but the dimer has lower primer-independent activity. n = 4 independent experiments. Data are presented as individual measurements and mean value, error bars represent SD values. e Activity assays and activity rate comparison of WT SuCphA1 (tetramer) and the W672A mutant (dimer) with (+) and without (−) primer. The two enzymes display similar primer-dependent activity, but the dimer has lower primer-independent activity.

**Fig. 4 Catalytic activity of the N domain and cyanophycin synthesis by dimeric CphA1.** a Mass spectra of (β-Asp-Arg)8-NH2, an 8mer cyanophycin segment in which the terminal carboxylate is replaced by an amide, before (left) and after (right) incubation with WT SuCphA1. After incubation with enzyme, the peak corresponding to (β-Asp-Arg)8 (expected at 2187.0 Da) is reduced, and two peaks with sizes matching to (β-Asp-Arg)4 (expected at 1102.5 Da) and (β-Asp-Arg)4-NH2 (expected at 1101.5 Da) appear. Peaks corresponding to Na+ and K+ adducts are also labeled. Peaks at 1093.4 and 1112.2 Da are deconvolution artifacts (Supplementary Fig. 5). b Mass spectra of (β-Asp-Arg)8-Asn before (left) and after (right) incubation with WT SuCphA1. After incubation with enzyme, the peak corresponding to (β-Asp-Arg)8-Asn (expected at 2301.1 Da) disappears and two peaks with sizes matching to (β-Asp-Arg)4 (expected at 1102.5 Da) and (β-Asp-Arg)4-Asn (expected at 1216.6 Da) appear. c, d Activity rate and lag time of SuCphA1 dimer complementation assays with c and without d primer. In presence of exogenous primer, the two mutant combinations G-M-N/-/G-M-N+ and G-M+N +/G-M-N display a similar activity rate, although the G-M-N+/-G-M-N combination displays a somewhat longer lag time. In the absence of primer, the G-M-N/-G-M-N+ combination displays a somewhat higher activity rate and lower lag time. n = 4 independent experiments. Data are presented as individual measurements and mean value, error bars represent SD values. e Activity assays and activity rate comparison of WT SuCphA1 (tetramer) and the W672A mutant (dimer) with (+) and without (−) primer. The two enzymes display similar primer-dependent activity, but the dimer has lower primer-independent activity. n = 4 independent experiments. Data are presented as individual measurements and mean value, error bars represent SD values.
the next steps of elongation of these intermediates, the rate of polymerization increases, and a long chain of cyanophycin is made. In CphA1 enzymes with N domain metallopeptidase activity, the chain is cleaved to generate cyanophycin segments such as (β-Asp-Arg)₄ that act efficiently as primers. This leads to more long chains and more primers, and thus rapid accumulation of cyanophycin after the initial lag phase we observe. CphA1 enzymes that lack active N domains make long cyanophycin chains as well, but because they are limited by the very slow initial rates in absence of primers, they make so few as to be undetectable in light-scattering or ATP hydrolysis assays3,19–21 (Supplementary Fig. 1a).

In vivo, these CphA1 enzymes that do not have N domain active sites likely use remnant strands of cyanophycin left over from the last round of catabolism, or other cellular small molecules25 as a primer. The maximum rates we observe in vitro indicate that polymerization of cyanophycin is several-fold faster than hydrolytic cleavage, but it is difficult to relate these rates to the situation in vivo, where cellular conditions are not constant and the availability of cyanophycin chains will change as molecules aggregate into granules. However, the accumulation of large amounts of cyanophycin in native bacteria and heterologous hosts clearly indicates the relative in vivo rates of polymerization and hydrolytic primer production are well tuned for cyanophycin biosynthesis.

Intriguingly, our experiments with dimeric CphA1 mutants (Fig. 4) suggest that a nascent chain could be polymerized at one end while being cleaved near the other end. In the absence of exogenous primer, mutant heterodimeric SuCphA1 displays higher synthesis rates when all three intact active sites are in the same protomer, hinting the increased rate is from cleavage in cis. A cyanophycin chain being elongated at its C-terminus by the G and M domains with soft anchoring on the N domain helices as the exogenous primer, mutant heterodimeric SuCphA1 displays higher synthesis rates when all three intact active sites are in the same protomer, hinting the increased rate is from cleavage in cis. A cyanophycin chain being elongated at its C-terminus by the G and M domains with soft anchoring on the N domain helices could intermittently wrap around to be cleaved at that same N domain’s active site (Supplementary Fig. 6a). Similarly, geometrical considerations can rationalize why dimeric SuCphA1 is less active than tetrameric SuCphA1 in the absence of exogenous primer: In the tetramer, two N domains active sites face each other and are 55 Å apart (protomers A and C; Supplementary Fig. 6a). It is possible that after cleavage, the new N-terminus of a cyanophycin chain experiences increased local concentration of N domain active sites, facilitating binding and increasing the rate of hydrolytic primer production. We note that the primer-dependent TmCphA1 has a different tetramer architecture11, in which the equivalent N domain positions are ~80 Å apart (Supplementary Fig. 6b).

The N domain appears distantly related to the M16 peptidase family32,33, which includes endopeptidases such as pitrilysin29 and insulin-degrading enzyme34. The family is also known as inverzincins35 because the active site motif HxxEH is inverted from the HExxH of the canonical mononuclear metallopeptidase motif. Active CphA1 N domains share the inverzincin HxxEH motif, as well as three structural elements: an active site helix, an adjacent β sheet, and the “backing helix”35 (Supplementary Fig. 6c). The CphA1 backing helix doubles as the α helix, which binds nascent cyanophycin chains through its surface-exposed side during biosynthesis11. Substrate binding in CphA1 and pitrilysin29 is similar, with the scissile peptide bond in analogous positions (Supplementary Fig. 6d). However, CphA1 N domains are clearly distinct from known inverzincins: CphA1 N domains are much smaller (~160 residues vs. up to 1000 residues), the structural similarity is modest and confined to the region around the active site, and crucially, the third metal-binding residue in CphA1 is a Cys upstream in sequence from the histidines in a C-H-M-metal-binding triad, rather than a downstream Asp, Glu or His in an M16 peptidase H-D/E/H triad31,35,36.

Two or more Cys ligands in structural Zn²⁺-binding motifs, but Cys as a Zn²⁺ ligand in an active metallopeptidase is rare37–39. The best-known example of a metallopeptidase with C-H-H metal coordinating residues is PDF, a ubiquitous enzyme responsible for deformylation of N-terminal fMet residues40. CphA1 N domains and PDFs share a very little structural similarity, and their active site helices are in opposite orientations (Supplementary Fig. 6e), but the geometry of the metal-binding residues of the two enzymes are remarkably similar (Supplementary Fig. 6f)40. PDF can bind Zn²⁺ tightly, but has a lower catalytic rate when bound with zinc than when bound with cobalt, nickel41, or iron42. The peptidase activity of CphA1 needs to be properly tuned so hydrolysis can generate primers but not efficiently compete with polymerization in this biosynthetic enzyme tasked with making long cyanophycin chains for storage. Because the biosynthetic and hydrolytic activities are both encoded into the same enzyme, the balance of these activities cannot be regulated by protein expression levels. Other features of the N domain active site that may temper the rate of hydrolysis and the lack of a residue for transition-rate stabilization (inverted zinc metallopeptidase28 has a Tyr; PDF13 has a Glu), or the lack of an active site residue accepting a hydrogen bond from N₉₉ of H79. The latter would promote the N₁ lone pair electrons facing the metal-binding site that can be important for activity31,44. H83 has such an interaction, but it is common for it to be seen for both active site histidines31,44.

Long strands of cyanophycin precipitate into granules for storage45. This precipitation may also serve to sequester cyanophycin from CphA1’s hydrolytic activity, since these chains of cyanophycin are largely stable both in vivo and in vitro in the presence of CphA1s which have active N domains19,22,45 (Supplementary Fig. 5a). Similarly, sequestration from the polymerizing G and M domains by precipitation into multistrand granules may be involved in determining cyanophycin chain length, which varies with CphA1, host, and other factors19,20,24,45,46. The exocyanophycinase CphB has a high Vₘₐₓ and an active site that is shallow and accessible47,48, allowing rapid degradation of strands in granules when needed.

The in vivo experiments we performed show that primer dependence can be a limiting factor for cyanophycin production in heterologous hosts. This understanding can help guide future efforts for more efficient polymer production in vivo, for example by prioritizing primer-independent enzymes. With ~80% of CphA1s having the Cx₉₉HxxEH motif, primer-independent CphA1s are more common than previously realized, and primer independence can be conferred by using N domain chimeras like TmCphA1SuN. Of the four constructs we assayed here, the chimera that introduces an active N domain into a primer-dependent CphA1 produced the highest cyanophycin yields.

Although cryptic active sites are not unheard of49, it is unusual to discover a cryptic active site in an enzyme that has been studied for decades. However, it was completely unexpected that cyanophycin synthetase would generate its own primers or have hydrolytic activity for any reason, given its biosynthetic role. Furthermore, the active site motif was obscured in sequence alignments (Fig. 2c and Supplementary Fig. 3a) by the ~20% of primer-dependent CphA1 enzymes that do not have the active site, and the structural similarity to metalloproteases is so modest that they do not appear in the top 100 results of DALI50 searches. Only the observation that the N domain confers primer independence led to the discovery of the N domain cyanophycinase site, thus showing that evolution combined three different enzymes into one elegant macromolecular machine. We are not aware of any other polymerase that has a dedicated active site to create primers needed for its biosynthetic cycle, making CphA1 a truly remarkable, multifunctional enzyme.
Methods
Cloning, protein expression, and purification. The genes encodingSucCphA1 (protein WP_028947105.1) and TomCphA1 (protein WP_004925883.1) were cloned into pJ411-derived plasmids in a previous study.11 Point mutants and (sub)domain chimeras used in this study were generated by transforming DH5α E. coli cells with PCR fragments containing overlapping ends. Phusion DNA polymerase (New England BioLabs) was used for all PCR reactions. Sequences of DNA primers used for PCR reactions are found in Supplementary Table 1. All purification steps were similar to those already described up to the elution from the HisTrap HP column. Following elution, the protein was mixed with CaCl2 to a final concentration of 2 mM and loaded onto a column of calmodulin-sepharose (Agilent) equilibrated in buffer A (250 mM NaCl, 20 mM Tris pH 8, 5 mM β-mercaptoethanol) and eluted with buffer B (250 mM NaCl, 20 mM Tris pH 8, 0.1 mM diethiothreitol). Following gel filtration, fractions with the highest purity were pooled and concentrated to ~20 mg/ml.

MS analysis of cyanophycin degradation. Synthetic cyanophycin segments were digested in 100 μl reactions containing 1 μM purified CphA1, 100 mM (NH4)2CO3, 20 mM KCl and 5 mM MgCl2, and 2 mM cyanophycin segments. Samples of 10 μl were taken at specific time points and diluted into 90 μl of 100 mM (NH4)2CO3, then directly injected for 2 min at 40 μl/min into a Bruker amaZon speed ETD ion trap mass spectrometer operating at positive ionization mode. The resulting spectra were deconvoluted using the max entropy method.

SDS-PAGE analysis of cyanophycin degradation. Reactions contained 20 μM purified CphA1, 50 mM (NH4)2CO3, 20 mM KCl and 5 mM MgCl2, 5 mg/ml cyanophycin purified from E. coli, 0.01% NaN3, and 200 μM phenylmethylsulfonyl fluoride. The reactions were incubated at room temperature. Samples of 20 μl were mixed with 10 μl of 5× loading buffer, boiled for 1 min, and analyzed on a 17% polyacrylamide gel.

Cyanophycin purification. E. coli BL21(DE3) cells were transformed with the same plasmids used for protein expression and plated on LB plates supplemented with 50 μg/ml kanamycin. The next day, single colonies were picked and used to inoculate 10 ml of LB supplemented with 100 μg/ml kanamycin. The starter culture was grown overnight with shaking at 37 °C and then used to inoculate 1 l of LB supplemented with 100 μg/ml kanamycin. One-liter cultures were grown with shaking at 37 °C until OD600 reached 0.5, and then the temperature was reduced to 25 °C. After 1 h, protein expression was induced with 0.25 mM IPTG, and the cultures grown for another 20 h. The next day, cells were harvested by centrifugation, resuspended in 1 ml dH2O for every 0.2 g of cell pellet, and lysed by sonication at room temperature. The lysates were acidified to pH 0.9 using concentrated HCl and clarified by centrifugation at 3500 g for 20 min. The pH of the purified lysate was then neutralized using 2 M NaOH. Following centrifugation at 3500 g for 10 min, the pellets contained insoluble cyanophycin and the lysate contained soluble cyanophycin. The pellets were resuspended in 0.1 M HCl, centrifuged at 3500 g for 10 min and the resulting pellets were discarded. The pH of the liquid phase was neutralized with 2 M NaOH and centrifuged for 10 min at 3500 g. The resulting pellets, consisting of purified insoluble cyanophycin, were lyophilized and weighed. The lysate containing soluble cyanophycin was mixed with 1 volume of 95% EtOH and centrifuged for 10 min at 3500 g. The resulting pellet was resuspended in dH2O, mixed with 1 volume of 95% EtOH, and centrifuged for 10 min at 3500 g. The resulting pellet, consisting of purified soluble cyanophycin, was lyophilized and weighed. The reported amounts of purified cyanophycin are the sum of the soluble and insoluble polymer from each culture.

Synthesis of cyanophycin segments. (β-Asp-Arg)n dipeptides were made from purified cyanophycin made in vitro in a primer-independent reaction by SucCphA1. The produced polymer was isolated by centrifugation at 3500 g for 10 min, washed with dH2O, and resuspended in 50 mM (NH4)2CO3. The polymer suspension was digested with purified cyanophycosynase from Synechocystis sp. PCC680351 until the suspension became clear, then filtered using a 3 kDa molecular weight cut-off Amicon centrifugation concentrator (EMD Millipore) and lyophilized.

Other cyanophycin segments were prepared by manual Fmoc solid-phase peptide synthesis (SPPS) as previously described.11 Briefly, (β-Asp-Arg)n, where n = 2, 3, 4, 8, or 16, were synthesized on an HMPB-ChemMatrix resin (Bachem). Couplings and peptide release from the resin were otherwise the same as for the other derivatives. All products were purified by reverse phase preparative HPLC and analyzed by high-resolution mass spectrometry.11

Cyanophycin reductase assay. E. coli cells were grown in LB media supplemented with 100 μg/ml kanamycin at 37 °C until OD600 reached ~0.5. The growth temperature was then reduced to 22 °C and protein expression was induced with 0.25 mM isopropyl β-d-1-thiogalactopyranoside (IPTG) for ~20 h. Following harvesting by centrifugation, the cultures were resuspended in buffer A (250 mM NaCl, 50 mM Tris pH 8.0, 10 mM imidazole, 2 mM β-mercaptoethanol) supplemented with a few crystals of lysozyme and DNAse I, and lysed by sonication at 0 °C. The lysate was clarified by centrifugation at 35000 g until use.

Synthesis of cyanophycin segments. β-Asp-Arg)n, where n = 4 and 8, were purified already, but using Fmoc-(OrBu)-OH (OrBuOH)-OH for the first coupling to the resin; the allyl protecting group was removed with Pd(PPh3)4 and PhSH in the final deprotection step. The C-terminal amides (β-Asp-Arg)n-NH2 (n = 4 and 8) were synthesized on manual Fmoc-SPPS on an N-alkylated PAL resin (Bachem); couplings and peptide release from the resin were otherwise the same as for the other derivatives. All products were purified by reverse phase preparative HPLC and analyzed by high-resolution mass spectrometry.11

Metal analysis. For metal analysis, purified protein samples of SucCphA1, its extruded N domain, and TomCphA1 were buffer exchanged into 100 mM (NH4)2CO3 by performing gel filtration with a Superdex 200 10/300 column equilibrated with that buffer. Protein-containing fractions were concentrated to 10 mg/ml and analyzed by ICP-OES at the Center for Isotope Studies, University of Georgia. A sample of the buffer eluted from the column was used as a control.
Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability. The structural models and maps (Supplementary Fig. 4 and Supplementary Table 1) generated in this study are available in the Protein Data Bank database under accession codes 7TXU and 7TXV and the Electron Microscopy Data Bank under accession code EMD-26161. The biochemical data (Figs. 1c, d, 2a, b, e, and 4, Table 1, Supplementary Figs. 1, 2c, 3c, and 5a–e, and Supplementary Table 4) generated in this study are provided in the source data file. Source data are provided with this paper.

Received: 1 March 2022; Accepted: 14 June 2022; Published online: 07 July 2022

References

1. Borzi, A. Le comunicazioni intracellulari delle nostochinee (Messina, 1886).
2. Simon, R. D. The biosynthesis of multi-L-arginyl-poly-L-aspartate (cyanophycin) in the filamentous cyanobacterium Anabaena cylindrica. Biochim. Biophys. Acta 422, 467–481 (1976).
3. Simon, R. D. The biosynthesis of multi-L-arginyl-poly-L-aspartic acid in the filamentous cyanobacterium Anabaena cylindrica. Biochim. Biophys. Acta 422, 467–481 (1976).
4. Simon, R. D. The biosynthesis of multi-L-arginyl-poly-L-aspartate (cyanophycin). Eur. J. Biochem. 254, 154–159 (1998).
5. Aboulmagd, E., Oppermann-Sanio, F. B. & Steinbuchel, A. Molecular characterization of the cyanophycin synthetase from Synechocystis sp. strain MA19. Appl. Microbiol. Biotechnol. 297–306 (2000).
6. 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, 2599–2608 (2004).
7. Nakamura, T. & Kino, K. A cyanophycin synthetase from Thermosynechococcus elongatus BP-1 catalyzes primer-independent cyanophycin synthesis. Appl. Microbiol. Biotechnol. 81, 69–78 (2008).
8. Hai, T., Oppermann-Sanio, F. B. & Steinbuchel, A. Purification and characterization of cyanophycin and cyanophycin synthetase from the thermophilic cyanobacterium Synechococcus sp. strain MA19 and in vitro synthesis of cyanophycin and related polypeptides. Appl. Environ. Microbiol. 68, 93–101 (2002).
9. Uddin, Z., Fang, T. Y., Siao, J. Y. & Tseng, W. C. Wound healing attributes of Leucomelanospora strain of unusual phylogenetic affinity. Environ. Microbiol. 17, 1110–1111 (2022).
10. Nausch, H. et al. Tobacco as platform for a commercial production of cyanophycin synthetase from a deep-sea sediment metagenomic library. Appl. Microbiol. Biotechnol. 97, 8619–8628 (2013).
11. Sharon, I., Grogg, M., Hülvert, D. & Schmeing, T.M. Structure and function of the β-Asp-Arg polymerase cyanophycin synthetase 2. ACS Chem. Biol. 17, 418–807 (2022).
12. Beckler, A. B. & Roth, R. A. An unusual active site identified in a family of zinc metalloendopeptidases. Proc. Natl Acad. Sci. USA 89, 3835–3839 (1992).
13. King, J. V. et al. Molecular basis of substrate recognition and degradation by human prion protease. Structure 22, 996–1007 (2014).
14. Rajagopalan, P. T., Datta, A. & Pei, D. Purification, characterization, and inhibition of peptide deformylase from Escherichia coli. Biochemistry 36, 13910–13918 (1997).
15. Fukasawa, K. M., Hata, T., Ono, Y. & Hirose, J. Metal preferences of zinc-binding motif on metalloproteases. J. Amino Acids 2011, 574816 (2011).
16. Perlman, R. K., Gehm, B. D., Kuo, W. L. & Rosner, M. R. Functional analysis of conserved residues in the active site of insulin-degrading enzyme. J. Biol. Chem. 268, 21538–21544 (1993).
17. Rawlings, N. D. et al. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptides in the PANTHER database. Nucleic Acids Res. 46, D624–D632 (2018).
18. Perlman, R. K. & Rosner, M. R. Identification of zinc ligands of the insulin-degrading enzyme. J. Biol. Chem. 269, 33140–33145 (1994).
19. Kondo-Costa, N. & Gomis-Ruth, F. X. Architecture and function of metalloendopeptidase catalytic domains. Protein Sci. 23, 123–144 (2014).
20. Tallant, C., Marrero, A. & Gomis-Ruth, F. X. Matrix metalloproteinases: fold and function of their catalytic domains. Biochem. Biophys. Acta 1803, 20–28 (2010).
21. Springman, E. B., Angleton, E. L., Birkedal-Hansen, H. & Van Wart, H. E. Multiple modes of activation of latent human fibroblast collagenase: evidence for the role of a Cys73 active-site zinc complex in latency and a “cysteine switch” mechanism for activation. Proc. Natl Acad. Sci. USA 87, 364–368 (1990).
22. Lee, Y. M. & Lim, C. Physical basis of structural and catalytic Zn-binding sites in proteins. J. Mol. Biol. 379, 545–553 (2008).
23. Rawlings, N.D. & Salvesen, G. Handbook of Proteolytic Enzymes, 3 volumes, (liv, 3932 pages) (Elsevier/AP, Amsterdam, 2013).
24. Chan, M. K. et al. Crystal structure of the Escherichia coli peptide deformylase. Biochemistry 36, 13904–13909 (1997).
25. Jain, R., Hao, B., Liu, R. P. & Chan, M. K. Structures of E. coli peptide deformylase bound to formate: insight into the preference for Fe2+ over Zn2+ as the active site metal. J. Am. Chem. Soc. 127, 4558–4559 (2005).
26. Rajagopalan, P. T. R., Yu, X. C. & Pei, D. Peptide deformylase: a new type of mononuclear iron protein. J. Am. Chem. Soc. 119, 12418–12419 (1997).
27. Hao, B. et al. Structural basis for the design of antibiotics targeting peptide deformylase. Biochemistry 38, 4712–4719 (1999).
28. Fukasawa, K. M., Hirose, J., Hata, T. & Ono, Y. In rat depeptidylpeptidase III, His368 is essential for catalysis, and Glu707 or Glu722 stabilizes the coordination bond between His455 or His450 and zinc ion. Biochim. Biophys. Acta 1804, 2063–2069 (2010).
29. Frommeyer, M. & Steinbuchel, A. Increased lysine content is the main characteristic of the soluble form of the polypeptide cyanophycin synthesized by recombinant Escherichia coli. Appl. Environ. Microbiol. 79, 4474–4483 (2013).
We thank all the members of the Schmeing lab for advice and ongoing discussions on this project, J.P. Trempe for advice with mass spectrometry, Christopher Thibodeaux and Kenneth Johnson for advice with data interpretation, Nancy Rogerson for proofreading, staff at McGill Facility of EM Research (Kaustov Basu and Kelly Sears) for support during data collection and the Plasma Chemistry Laboratory at the Center for Applied Isotope Studies, University of Georgia for ICP-MS. The work (10.46936/10.25585/60001153) conducted by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02- 05CH11231. This work was funded by CIHR Project Grant 178084 and a Canada Research Chair to T.M.S., and the Schweizerischer Nationalfonds and ETH Zurich to D.H.

**Author contributions**

M.G. and S.P. performed the chemical synthesis of cyanophycin segments. I.S. performed all other biochemical and structural experiments and data processing. I.S. and T.M.S. wrote the manuscript with editing from D.H. N.S. supervised S.P.

**Competing interests**

The authors declare no competing interests.

**Additional information**

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-31542-7.

**Correspondence** and requests for materials should be addressed to T. Martin Schmeing.

**Peer review information** Nature Communications thanks Karl Forchhammer, Satish Nair, Max Cryle, and the other anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

| Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022