Structural bases for aspartate recognition and polymerization efficiency of cyanobacterial cyanophycin synthetase

Cyanophycin is a natural biopolymer consisting of equimolar amounts of aspartate and arginine as the backbone and branched sidechain, respectively. It is produced by a single enzyme, cyanophycin synthetase (CphA1), and accumulates as a nitrogen reservoir during N\textsubscript{2} fixation by most cyanobacteria. A recent structural study showed that three constituent domains of CphA1 function as two distinct catalytic sites and an oligomerization interface in cyanophycin synthesis. However, it remains unclear how the ATP-dependent addition of aspartate to cyanophycin is initiated at the catalytic site of the glutathione synthetase-like domain. Here, we report the cryogenic electron microscopy structures of CphA1, including a complex with aspartate, cyanophycin primer peptide, and ATP analog. These structures reveal the aspartate binding mode and phosphate-binding loop movement to the active site required for the reaction. Furthermore, structural and mutational data show a potential role of protein dynamics in the catalytic efficiency of the arginine condensation reaction.
are a significant challenge for microorganisms because the O₂ produced from CO₂ fixa- tion is inhibitory to nitrogenase, which catalyzes the conversion of N₂ to NH₃. To solve this problem, diazotrophs have developed new strategies to temporally and spatially separate nitrogenase from O₂. The heterocyst-forming cyanobacteria, such as *Anabaena* spp., perform CO₂ fixation and N₂ fixation in different cells: CO₂ fixation in the vegetative cells and N₂ fixation in heterocysts. Other cyanobacteria, such as *Trichodesmium* spp., temporally separate the processes by a dicro cycle with CO₂ fixation during the day and N₂ fixation at night⁸. The fixed nitrogen accumulates at night as cyanophycin granules isolated from the other cellular components but not as compounds that affect cellular redox dynamics. Cyanophycin is degraded by cyanophycinase and isopeptidase to release nitrogen to support cell growth during the day. Thus, cyanophycin granules serve as dynamic storage bodies in diatrophic cyanobacteria to uncouple N₂ fixation from overall growth dynamics. Cyanophycin accumulation enables the optimization of the nitrogen utilization of cyanobacteria in natural environments with fluctuating and limiting nitrogen supplies⁹.

The biosynthesis of cyanophycin is achieved by a single enzyme named cyanophycin synthetase (CphA1) that catalyzes peptide elongation coupled with ATP hydrolysis. CphA1 uses a low-molecular-weight cyanophycin (at least 3 dipeptides long) as a primer for cyanophycin polymerization⁶. Recently, the tertiary structure of CphA1 has become available and reveals that amino acid polymerization is driven by three distinct domains: the N-terminal domain (N domain), the middle glutathione synthetase-like domain (G domain) and the C-terminal domain (M domain). The discrete active sites of the G and M domains are used for extending the Asp backbone and attaching the Arg sidechain, respectively (Supplementary Fig. 1). The N domain has been proposed to anchor cyanophycin polymers through electrostatic interactions. However, it remains unclear how ATP-dependent addition of aspartate to cyanophycin is initiated at the catalytic site of the G domain.

In this work, given the importance of CphA1 in cyanophycin biosynthesis and the biotechnological value that structural information on enzyme-substrate interactions would provide, we investigate the CphA1 enzyme from the marine diazotroph *Trichodesmium erythraeum* IMS101 (TeCphA1), an important contributor to global nitrogen and carbon cycling. We solve the TeCphA1 structure bound to ATP analogs, cyanophycin primer peptides, and the substrate aspartate, in addition to apo and ATP analog-bound structures. Key elements, including residues and protein dynamics required for enzymatic activity, are identified based on experimental evidence of the G and M domains. In particular, we reveal the structural bases for aspartate recognition and condensation, which deepen our understanding of the reaction of cyanophycin biosynthesis in the G domain. Moreover, our data show a potential role of protein dynamics in the catalytic efficiency for the reaction of Arg sidechain condensation.

**Results**

**Cryo-EM structures of TeCphA1**

The TeCphA1 structures in three different states were determined by single-particle cryogenic electron microscopy (cryo-EM): apo, the adenosine-5′-(γ-thio)-triphosphate (ATPγS, an ATP analog)-bound state, and the substrate-bound state (TeCphA1-aspartate-ATPγS-cyanophycin primer peptide complex) at 2.91, 2.96, and 2.64 Å resolution, respectively (Fig. 1a, Supplementary Figs. 2–5, and Supplementary Table 1). The cryo-EM data of the substrate-bound state were collected under a high concentration of -aspartate (100 mM), cyanophycin primer peptide, (β-Asp-Arg)₄, and ATPγS. TeCphA1 consists of three domains: an N-terminal domain (N domain, residues 1–161) that adopts a partially similar structure to *E. coli* RNA-polymerase α-subunit, a glutathione synthetase-like domain (G domain, residues 162–487), and a MurE-like muramyl ligase domain (M domain, residues 488–876) (Fig. 1b). TeCphA1 basically adopts a tetrameric assembly with *D₄* symmetry (Fig. 1a and Supplementary Fig. 6), which is a common architecture with another cyanobacterial CphA1 from *Synecho-

The G domain, which is responsible for backbone elongation through linking aspartate to the C-terminal backbone carboxy group of cyanophycin, is based on Gcore residues 162–234, 306–324, and 400–487, with Glid (residues 235–305) and GM (residues 325–399) modules inserted into the sequence (Fig. 1b and Supplementary Fig. 1). Gcore is a central module that provides a large dimer interface with another Gcore in the TeCphA1 tetramer (Fig. 1c and Supplementary Fig. 8a). Gsid acts as a lid of the ATP-binding site of Gcore, and Gsid is positioned to surround the substrate-binding site together with Gsid. The relative map intensity of Gsid versus Gcore and Gsid is higher in the substrate-bound state than in the other two states (Fig. 1a), suggesting that Gsid is clearly visible as a result of reduced mobility, which is probably due to the binding of aspartate and (β-Asp-Arg)₄.

The M domain, which catalyzes the addition of the Arg sidechain to the C-terminal group of the cyanophycin primer (Fig. 1b), is divided into two modules, Mcore (residues 488–723) and Mlid (residues 724–876). Mcore contacts the intramolecular N domain and three Gcore to form the TeCphA1 tetramer (Fig. 1c and Supplementary Fig. 8c), suggesting the mobile nature of Mlid. However, in the substrate-bound state, the Mlid modules were partially resolved around ATPγS in chains A and B (Fig. 1a and Supplementary Fig. 6). The density of Mlid was relatively weak compared with that of Mcore. A large shift of Mlid was observed in the other CphA1 enzymes⁷.

TeCphA1 has a high sequence identity (69.6%) and similarity (83.2%) to *Synechococcus* in the region consisting of the N, G, and M domains (residues 1–876). We compared the structures of TeCphA1 in the substrate-bound state and *Synechococcus* that bound a cyanophycin analog (β-Asp-Arg)₄-NH₂ and an ATP analog 5′-adenylylmethylphosphonate (AMPPCP) to the G domain (PDB 7LGJ)⁹. These structures are similar to each other. The overall structures of the N domain and each module of the G and M domains (Gcore, Glid, GMcore, and Mlid) were quite similar between TeCphA1 and *Synechococcus* (root-mean-square deviation (RMSD) = 0.521–1.601 Å). The spatial arrangement of the N domain and the Gcore and Mcore modules was also well conserved between TeCphA1 and *Synechococcus* (Supplementary Fig. 7c, d). In particular, the distance between the Mlid and N domain of TeCphA1 is shorter than that of *Synechococcus* (Supplementary Fig. 7d).

**Functional roles of Gcore and Gsid in adding aspartate to cyanophycin primer peptide**

The TeCphA1 structure in the substrate-bound state visualizes the initial state of the catalytic reaction for the ATP-dependent addition of aspartate to the C-terminal backbone carboxy group of cyanophycin (Fig. 1b and Supplementary Fig. 1). In the cryo-EM map processed with *C₁* symmetry, all the TeCphA1 protomers in the tetramer bound a cyanophycin primer peptide, (β-Asp-Arg)₄, on the surface of Gcore (Fig. 1a). The cryo-EM map showed a clear density for the C-terminal β-Asp-Arg dipeptide unit (4th unit) and the other Asp backbones with a hook-like shape (Fig. 2a). The Arg sidechain of the 3rd unit partially showed weak density. Since the Arg sidechains of the 1st and 2nd units

**Article**

https://doi.org/10.1038/s41467-022-32834-8
were invisible in the cryo-EM map, these models could not be built in the substrate-bound TeCphA1 structure. In contrast, the Arg sidechain of the 2nd unit was visible in the cryo-EM map of SuCphA1₁, and its guanidium group was located near Ala188. This residue is substituted with Phe on many cyanobacterial CphA1, including TeCphA1 (Fig. 2a and Supplementary Fig. 9). The residue type at position 188 may affect the Arg-sidechain orientation of the 2nd unit having Gcore.

In the TeCphA1 structure, the C-terminal backbone carboxy group and Arg sidechain of the 4th unit interact with Arg309 (the nearest atomic distance, 2.9 Å) and Glu213, which is derived from another protomer forming the Gcore-Gcore dimer (Glu215) (2.8 Å), respectively (Fig. 2a and Supplementary Fig. 8a). Since the catalytic activity of TeCphA1 was impaired by the Ala substitutions of Glu215 and Arg309 (Fig. 2b and Supplementary Fig. 9), the interaction between the Gcore-Gcore dimer and the 4th unit seems to be required for the activity of TeCphA1.

The mutations of Glu215 and Arg309 (E215A and R309A) of the G domain were commonly analyzed in the two studies on TeCphA1 and SuCphA1. Of the two, R309A (and its corresponding mutation in SuCphA1) showed almost no activities. Both residues seem to be essential for the catalytic activities of TeCphA1 and SuCphA1. On the other hand, while the assay conditions were not the same between the two, the E215A mutation was different between TeCphA1 and SuCphA1.
The E215A of TeCphA1 lost most of the catalytic activity (Fig. 2b), but the corresponding mutation of SuCphA1 retained nearly half the activity. Since the interaction between E215 and the C-terminal Arg sidechain of cyanophycin contributes more to the cyanophycin binding in TeCphA1 than in SuCphA1, the results of this mutational analysis seem to be explained by the difference in the structures. Although the importance of the Arg sidechains for the lack of poly-Asp polymerase activity has been explained by the structural observation of SuCphA1, our results showed that the Arg sidechain of the 4th unit is a major contributor to the (β-Asp-Arg)₄ recognition and determines the position of the 4th Asp backbone at the active site of the G domain to avoid polymerization progression with the aspartate backbone alone. Since (β-Asp-Arg)₄ is well accommodated at the binding site of G₄, a cyanophycin molecule with at least 3–4 dipeptides long seems to be suitable as a primer peptide for TeCphA1 activity (Fig. 2c). This observation is consistent with the previous results of the other CphA1. According to the common structural features between TeCphA1 and SuCphA1, the cyanophycin primer peptide seems to adopt a hook-like shape at the active site of CphA1 (Fig. 2a).

While (β-Asp-Arg)₄ was observed in all protomers, the substrate aspartate was observed only in one protomer (chain D) (Fig. 2a). The map intensity of G₄ was significantly different among the four protomers, and G₄ of chain D showed the highest map intensity of the four G₄ (Fig. 2d). The substrate aspartate was visualized in the most stable G₄. The other G₄ modules seemed to be more mobile than those in chain D. TeCphA1 required four substrates, L-aspartate, L-arginine, (β-Asp-Arg)₄, and ATP, for the catalytic reaction (Supplementary Fig. 10a). Steady-state kinetic analysis showed a typical Michaelis-Menten kinetics for L-arginine, (β-Asp-Arg)₄, and ATP (Supplementary Fig. 10b). While we have determined the apparent Km value for ATP, TeCphA1 has two active sites in the G and M domains for the different catalytic reactions with ATP. As observed for the other CphA1 enzyme, the sequential condensation reaction of TeCphA1 may be rate-limited primarily by the site with lower affinity for ATP when the Km values for ATP differ between the G and M domains. Unlike these substrates, aspartate acts as a positive effector of the catalytic reaction with the Hill coefficient (h) 1.92 ± 0.07. In addition, K₅₀ and K₈₅ (K₅ₐ₅) were estimated as 18.0 ± 1.1 mM and 254 ± 21 mM, respectively. As shown in Supplementary Fig. 10b, the catalytic activity of TeCphA1 increases with the increasing aspartate concentration; the cellular aspartate concentration may be a key regulator for the catalytic activity of TeCphA1. The results of the kinetic analysis for aspartate (Supplementary Fig. 10b) suggest that 100 mM aspartate is appropriate for the cryo-EM analysis. While it is unclear why the aspartate molecule was visible in only one subunit, the cryo-EM structure of TeCphA1 with ATPγS, (β-Asp-Arg)₄, and aspartate may be a
impaired Te aspartate was close to the sidechain of Arg458 (2.5 Å) in the consistent with the notion that Glg mutation. The substituted Trp residue prevented the reaction of bone elongation.

The numbers under the logo correspond to the residue numbers of the GshF ATP-grasp-like module. The P-loop sequences are shown in red boxes.

The substrate aspartate was accommodated in a small space beside Ser396 (Fig. 2a). This binding mode seemed to be conserved in SuCphA1 because SuCphA1 activity was lost by the S396W mutation. The substituted Trp residue prevented the reaction of Asp backbone elongation by occupying the space for substrate aspartate binding (Fig. 2a). Residues contributing to substrate aspartate recognition are conserved in the cyanobacterial CphA1 enzymes (Fig. 2a and Supplementary Fig. 9). Asn394, a residue on the large loop of Gcore (residues 389–399), and Arg323 interacted with the β-carboxy group of the substrate aspartate at distances of 2.5 Å and 3.4 Å, respectively (Fig. 2a). The α-carboxy group of the substrate aspartate was close to the sidechain of Arg458 (2.5 Å) in the TeCphA1 structure. Each Ala mutation of Asn394, Arg323 or Arg458 impaired TeCphA1 activity (Fig. 2b). These observations are consistent with the notion that Glg functions as a module for substrate aspartate recognition.

Catalytically important residues in Gcore and Glg for Asp backbone elongation

The G domain adopts an ATP-grasp fold that requires nucleophilic attack of the C-terminal backbone carboxy group of cyanophycin on ATP for Asp backbone elongation. However, since these two functional groups are negatively charged, it is necessary to avoid electrostatic repulsion between them. The ATP bound to the ATP-grasp fold typically interacts with one to three Mg ions. In the G domain of TeCphA1, the γ-phosphate chelates two Mg ions together with two acidic residues, Asp431 and Glu450 (Figs. 2a and 3a). This structural feature is similar to SuCphA1 and other ATP-grasp enzymes, such as bacterial glutathione synthetases and N-carboxyaminimidazole ribonucleotide synthetase (PurK). In addition, the α- and β-phosphates of ATPS interact with the positively charged sidechains of Lys261 and Lys220, respectively (Fig. 3a). While the triphosphate of ATP is neutralized overall by these interactions, there are still two conserved charged residues, Arg309 and Arg323, around the γ-phosphate. The substrate-bound TeCphA1 structure shows that these residues electrostatically interact with the C-terminal backbone carboxy group of (β-Asp-Arg)₄, and β-carboxy group of the bound aspartate (Fig. 3a and Supplementary Fig. 9).

The substrate-bound-state structure suggests that the phosphate-binding loop of Glg (hereafter P-loop, residues 263–269) (Fig. 3b) plays a critical role in the catalytic reaction. In the substrate-bound state, His267 on the P-loop is located between γ-phosphate and the C-terminal backbone carboxy group of (β-Asp-Arg)₄ (Fig. 3a). This His267 position seems to be suitable to stabilize an acylphosphate intermediate together with Arg309, following the nucleophilic attack of the substrate aspartate. The importance of these residues was confirmed by the conservation of cyanobacterial CphA1 enzymes, including SuCphA1 and decreased activity of a mutant substituted with...
Ala both in TeCphA1 and SuCphA1 (Figs. 2b and 3b). This loop corresponds to the B loop of PurT and PurK, which are ATP-grasp enzymes of *E. coli*. These enzymes function in the de novo purine biosynthetic pathway, and the B loop is required to efficiently generate an acylphosphate intermediate in these enzymes. However, His267, which seems to be important to stabilize the acylphosphate intermediate in TeCphA1, is not conserved in PurT and PurK.

C-terminus motion in the catalytic mechanism of Asp backbone elongation

Due to the charge neutralization of the γ-phosphate of ATP and the C-terminal backbone carboxy group of cyanophycin, these two groups can approach each other until reaching the distance of nucleophilic attack. A comparison of the three states of the TeCphA1 structure (Fig. 3c) revealed conformational changes of Cβ induced by ATP, the substrate aspartate and (β-Asp-Arg)4 binding. These conformational changes seem to contribute to optimizing the relative dispositions of ATP, aspartate, and (β-Asp-Arg)4 for the catalytic reaction. The binding of ATPγS causes Cβ to move toward Cγ by 4.6 Å, which is the distance between the Cα atoms of Lys293 (Fig. 3c, asterisk), and the substrate aspartate and (β-Asp-Arg)4 binding induces a further shift of Cβ by 4.3 Å. This Cβ movement seems to change the binding position of ATPγS (β-Asp-Arg)4. Interestingly, Cγ and ATPγS in the substrate-bound TeCphA1 structure are located at a position closer to Cβ than those in the SuCphA1 structure bound with the cyanophycin analog (β-Asp-Arg)8-NH2 (Fig. 3d). Since the residues contributing to the reaction of Asp backbone elongation in the G domain are conserved between TeCphA1 and SuCphA1 (Supplementary Fig. 9), the binding of the substrate aspartate seems to modulate the orientation of ATPγS together with Cβ.

While the binding of ATP, the substrate aspartate, and (β-Asp-Arg)4 seems to cause the rearrangement of their relative orientations for the catalytic reaction, the γ-phosphate of ATP is less susceptible to nucleophilic attack by the C-terminal backbone carboxy group of (β-Asp-Arg)4, because they are located at a distance of ~6 Å (Fig. 3c). Although Cβ moves toward the active site to change the relative positions of ATP and (β-Asp-Arg)4, a further mechanism is required to bridge this distance gap for the catalytic reaction to proceed. One of the possible mechanisms is the thermal motion of CαS, which is mentioned in relation to SuCphA1. Since the cryo-EM structures of TeCphA1 only represent the average position in the equilibrium of the CαS motion, it may be possible to consider that the protein dynamics allow these two molecules to occasionally access the distance of the catalytic reaction.

While the relative orientation of ATPγS and (β-Asp-Arg)4 in the substrate-bound state needs a greater change for the catalytic reaction, the relative orientation of (β-Asp-Arg)4 and the substrate aspartate in the G domain seems to be suitable for ATP-dependent backbone elongation (Fig. 1b and Supplementary Fig. 1). The amino group of substrate aspartate is located 4.1 Å from the carboxy group at the C-terminus of (β-Asp-Arg)4, and the position of the carboxy group may be adjusted by phosphorylation (Figs. 2a and 3c).

Structural and functional roles of Mlid in Arg sidechain condensation

In the substrate-bound state, Mlid was clearly visible in the four protomers of the TeCphA1 tetramer. ATPγS interacts with the conserved P-loop of Mlid (hereafter P-loopβ, residues 495–501) (Fig. 4a and Supplementary Fig. 9). Lys499 and an Mg2+ ion are located between the β-phosphate and γ-thiophosphate groups of ATPγS, and the Mg2+ ion is chelated by Thr500, Thr522, and Glu538. These interactions with ATP are highly conserved in the Mur ligases, which synthesize the peptide stem of bacterial peptidoglycan. Since the Ala mutation of Lys499 impaired TeCphA1 activity (Fig. 4b), the interaction between Lys499 and ATP is also critical to the catalytic activity of CphA1 enzymes.

While ATPγS is visible in the Mcore modules of all protomers, the cryo-EM densities for Mlid were observed weakly only in protomers A and B (Supplementary Fig. 11), suggesting a relatively mobile nature of Mlid even in the substrate-bound state. In protomers A and B, Mlid and Mcore sandwich ATPγS (Figs. 1a and 4a), and Arg731 and His748 in Mlid interact with the ϕ-phosphate and γ-thiophosphate groups of ATPγS. Since Mlid was not observed in the ATPγS-bound state, Mlid is unlikely essential to the stable binding of ATPγS.

However, since the deletion mutant of Mlid (ΔMlid) showed substantially reduced activity of TeCphA1, the Mlid module is necessary for the efficient catalytic reaction of TeCphA1 (Fig. 4b) and seems to play a significant role in ATP-dependent Arg sidechain condensation. Superposition of TeCphA1 in apo- and ATPγS-bound structures reveals a conformational change of Phe692 in Mcore upon ATPγS binding, and Phe692 and Asn497 sandwich the adenine ring of ATPγS (Fig. 4c). This sidechain orientation of Phe692 causes steric hindrance with the methyl group of Ala755 in Mlid (Fig. 4c and Supplementary Fig. 12). On the other hand, the sidechain position of Phe692 in the apo form inhibits ATP binding but not the interaction between Mcore and Mlid (Fig. 4e). The flexible nature of Mlid and the conformational change of Phe692 seem to cause mutually exclusive binding conformations of ATP and Mlid, which may facilitate the exchange of ATP and ADP in the reaction cycle. The invisible Mlid in protomers C and D can be explained by mutually exclusive binding conformations, and protomers C and D are considered ATPγS-binding states. Intriguingly, protomers A and B of the substrate-bound state showed simultaneous binding of ATPγS and Mlid to Mcore. We consider that Mlid and ATPγS in protomers A and B are in a metastable state that can be realized due to a high concentration of ATPγS. The interaction between Ala755 and Phe692 in protomers A and B may be energetically unfavored and thus easily changed to the Mlid binding or ATPγS-binding state under lower ATPγS concentrations.

While ATPγS was visible on all M domains, (β-Asp-Arg)4 was not observed on the M domain of the substrate-bound TeCphA1 structure. This observation is consistent with the previous result that the Arg-modified Asp residue at the C-terminus of cyanophycin is required to bind to the M domain in the SuCphA1 structure. When the TeCphA1 structure is superposed with the SuCphA1 structure in complex with the cyanophycin analog (β-Asp-Arg)8-Asn3, the cyanophycin analog is located between the N domain and Mcore of TeCphA1. Interestingly, we found a densely charged region with multiple Arg and Asp/Glu residues on the surface of Mlid (Fig. 4d). These residues seem to interact with the cyanophycin analog, as observed in SuCphA1. Mlid may contribute to guiding cyanophycin into the active site of Mcore. In the Mur ligases, the interaction with the substrate induces a rotation of the C-terminal domain, which corresponds to Mlid of CphA1 enzymes, resulting in a closed conformation. However, the orientation of Mlid is almost unchanged by the substrate (cyanophycin) interaction in SuCphA1 (RMSD = 0.158 Å, 2,583 atoms of Mcore and Mlid).

Discussion

Cyanophycin is a unique biopolymer that consists of the β-Asp-Arg dipeptide as a repeating unit (Supplementary Fig. 1) and is produced by CphA1-dependent peptide synthesis. Our understanding of this polymerization process has been greatly improved by a recent report that provided structural snapshots of CphA1. However, the catalytic and regulatory mechanisms of this reaction have not been fully elucidated. In the present study, we determined the CphA1 structure with a complete set of substrates for the reaction of the G domain, aspartate and (β-Asp-Arg)4, which visualizes the initial state for ATP-dependent backbone elongation. This structure proposes a feasible catalytic mechanism for the Asp backbone elongation reaction that occurs in the G domain (Fig. 3). In the apo state, Mlid adopts an open conformation and is the furthest away from Mcore. ATP binding induces the movement of Mlid toward Mcore to sandwich an ATP molecule, whereas

Article

https://doi.org/10.1038/s41467-022-32834-8

Nature Communications | (2022) 13:5097

6
Gin remains flexible. Cyanophycin and aspartate binding further
changes Gin to a closed conformation, and the orientation of Gin is
fixed to localize aspartate in a suitable position for the reaction by
the interaction of Asn394 on the large loop with the substrate. Thus, ATP,
the substrate aspartate, cyanophycin, and residues, including His267
on the P-loop25, are arranged to initiate the reaction of Asp backbone
elongation. The C-terminal backbone carboxy group of cyanophycin is
activated by ATP to form a high-energy acylphosphate intermediate
required for the ATP-grasp enzymes22, which is stabilized by His267 on
the P-loop25 together with Arg309. The amino group of the substrate
aspartate nucleophilically targets the carbonyl group of the acylphosphate,
thus generating a tetrahedral intermediate in which an oxygen
is stabilized by the interaction with Arg309 and Gly456. A new
peptide bond forms between cyanophycin and the substrate aspartate
with the release of phosphate.

The G domain of CphA1 enzymes adopts a fold similar to the ATP-
grasp-like module of bacterial hybrid-type glutathione synthetase,
GshF, which catalyzes peptide bond formation between the amino
group of glycine and the C-terminal backbone carboxy group of γ-
glutamyl cysteine23,24. The amino acid sequence of the P-loop25 in CphA1
is different from that in GshF. In particular, the P-loop25 has a His residue
(His267) in CphA1 but a Tyr or Phe residue in GshF (Fig. 3b). In
addition, Glu215 and Arg458 of TeCphA1 are substituted with Leu and
Met/Tyr in GshF, respectively. These different residues of CphA1 from
GshF seem to be optimized for the catalytic reaction for the Asp
backbone elongation of cyanophycin. Mutational analysis suggested
that His267 plays a critical role in the catalytic reaction for TeCphA1
(Figs. 2b and 5). Glu215 and Arg458 are required for the recognition of
(β-Asp-Arg)7 and the substrate aspartate at the active site of Gcore,
respectively (Fig. 2a, b).

Although Trz2120 from Thermosynechococcus elongatus BP-1 has
been found to be the only CphA1 enzyme that synthesizes cyanophycin
in a primer-independent manner26, the other known CphA1 enzymes
show primer-dependent activity. Efficient CphA1 activity requires a
cyanophycin primer at least 3–4 dipeptides long (Fig. 2c)27. Hence, to
rapidly initiate cyanophycin synthesis upon the transition to the N2
fixation phase, cyanophycin primer peptides must be available in
cyanobacterial cells. Another critical factor for CphA1 activity is the
cellular level of aspartate. In the process of N2 fixation, aspartate is
produced by the incorporation of nitrogen into oxaloacetate and is
consumed directly for cyanophycin synthesis2. Aspartate is also
required as a nitrogen donor for arginine biosynthesis in
cyanobacteria27. Therefore, the intracellular aspartate concentration is
one of the factors most likely to affect the progression of cyanophycin
synthesis. The steady-state kinetics of TeCphA1 suggests an allosteric
effect with positive cooperativity toward aspartate, and the reaction is,
in fact, accelerated with an increasing aspartate concentration up to
quite a high concentration range (Supplementary Fig. 10b). The
structural comparison between the substrate-bound TeCphA1 and the
SuCphA1 bound with the cyanophycin analog (β-Asp-Arg)7-NH2 showed
that the binding of aspartate modulates the orientation of ATP
(Fig. 3d). This aspartate-mediated regulation of Gin seems reasonable to control the CphA1 activity associated with storage of
fixed nitrogen in cyanobacteria to avoid wasteful phosphorylation

![Figure 4](https://doi.org/10.1038/s41467-022-32834-8)
of primer peptides under conditions in which aspartate is not available to CphA1 as a substrate.

In the apo and ATPγS-bound states, a large density was observed between two Glid modules and over the binding site of (β-Asp-Arg)₄ (Fig. 1a, arrowhead). CphA1 is known to synthesize cyanophycin in E. coli cells; hence, cyanophycin is the most likely candidate for the molecule with increased density. In addition, the polymerization reaction of CphA1 often terminates up to a polymer length of 25–30 kDa. The mobility of Glid is likely to be affected by the association of a high-molecular-weight molecule into the space between two Glid modules, and the regulation of Glid motion may also be related to the length determination and polymerization termination of cyanophycin, although this mechanism is a question that needs to be addressed in further study. On the other hand, the windshield wiper model using the N domain has been proposed as a promising model for predicting the efficiency of reactions using two distant active sites alternately. CphA2, which synthesizes cyanophycin by directly linking β-Asp-Arg dipeptides, lacks not only the P-loop but also the Mlid module to abolish the activity for Arg sidechain condensation, which is supported by our mutational data of TeCphA1, which does not exert the CphA2 activity (Fig. 4b and Supplementary Fig. 10a). This molecular evolution may imply that the presence of Mlid is unfavorable for the reaction to proceed efficiently only in the G domain, rather than alternately in the two domains. The mobility of Mlid and its densely charged region (Fig. 4c) may play a role in the reactivity of both active sites together with the N domain.

Fig. 5 | Schematic diagram of the catalytic reaction for aspartate addition to the C-terminal backbone carboxy group of cyanophycin. (i) G domain from a protomer in the apo state. Three modules, Gcore, Glid, and Gω, are shown in blue, pink, and purple, respectively. (ii) ATP-bound state. The chemical structure of ATP is depicted as AMP diphosphate. (iii) Substrate-bound state. The orientation of Glid is further changes compared to the ATP-bound state. (iv) Feasible catalytic mechanism in the G domain. In this mechanism, a tetrahedral intermediate is formed by nucleophilic attack to an acylphosphate intermediate of cyanophycin by the amino group of aspartate. Arg309 is assumed to stabilize the acylphosphate intermediate and the tetrahedral intermediate together with His267 and Gly456, respectively.
Methods
Materials
Cyanophycin primer peptides β-Asp-Arg, (β-Asp-Arg)2, (β-Asp-Arg)3, (β-Asp-Arg)4, and (β-Asp-Arg)5 were synthesized via the solid phase method by Sangon Biotech (Shanghai, China). The synthesized peptides were confirmed using a TripleTOF 5600+ system (Sciex) with a TurboScout™ (TIS) probe (Turbo V ion source, Sciex). Time-of-flight-mass spectroscopy (TOF-MS) spectra (m/z range, 100–2000 Da) were recorded in positive-ion mode with the following parameters: curtain gas, 15 psi; ion source gas 1, 15 psi; ion source gas 2, 0 psi; ion spray floating voltage, 5500 V; and interface heater temperature, 400 °C. The mass spectrometer was recalibrated using an APCI positive calibration solution (Sciex). The obtained spectra are listed in Supplementary Fig. 13. All the other chemicals were purchased from Sigma–Aldrich or FUJIFILM Wako Pure Chemical Corporation.

Expression and purification of 7eCphA1
The gene encoding 7eCphA1 (ENA code: ABG51217) was amplified from the genomic DNA of T. erythraeum IMS101 with forward and reverse primers (CphA-for: TATCATATGAAATCTCAAACTCCGA CAT-TACG; CphA-rev: GGTGCTCGAGAATTGAACTTTTTAAAACTT) containing 5′-10 bp overhangs complementary to the vector sequence for seamless cloning (TransGen Biotech). The target gene was cloned into the pET-22b(+) vector (Novagen) between the restriction sites Ndel and XhoI along with a C-terminal hexa-histidine tag. All mutants were prepared by inverse PCR for site-specific mutagenesis using PrimeSTAR Max DNA polymerase (Takara Bio) and primers listed in Supplementary Table 2.

Recombinant proteins were expressed in the E. coli strain BL21(DE3) (Novagen). Cells were inoculated in LB media supplemented with 50 mg ml−1 ampicillin and cultured overnight. Upon reaching an optical density of ~0.6 to 0.660 nm, the temperature was decreased from 37°C to 16°C, and the culture was supplemented with 0.5 mM isopropyl β-D-thiogalactopyranoside. Protein expression was performed for 16 h, after which cells were harvested by centrifugation at 5,180 × g for 10 min, followed by resuspension in lysis buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 1 mM dithiothreitol (DTT), and 5 mM imidazole). Cells were lysed by sonication, and cell debris was then removed by centrifugation at 40,000 × g for 30 min. The supernatant fraction was loaded onto an equilibrated Ni-NTA resin (Qiagen). After extensive washing and debris was then removed by centrifugation at 40,000 × g for 30 min. The supernatant fraction was loaded onto an equilibrated Ni-NTA resin (Qiagen). After extensive washing and

Grid preparation and cryo-EM data collection
Cryo-EM samples were prepared by mixing 24 μM 7eCphA1 with each solution. The final sample compositions were as follows: (1) apo state, 12 μM 7eCphA1 in a solution containing 20 mM Tris-HCI (pH 8.0), 400 mM NaCl, and 1 mM DTT; (2) ATPS-bound state, 2 μM 7eCphA1 in a solution containing 50 mM Tris-HCl (pH 8.2), 200 mM NaCl, 20 mM KCl, 20 mM MgCl2, 1 mM DTT, 4 mM ATPS, 5 mM l-aspartate, 0.5 mM l-arginine, and 10 mM (β-Asp-Arg)5 and (3) substrate-bound state, 4 μM 7eCphA1 in a solution containing 50 mM Tris-HCl (pH 8.2), 67 mM NaCl, 20 mM KCl, 20 mM MgCl2, 1 mM DTT, 10 mM ATPS, 100 mM l-aspartate, and 20 mM (β-Asp-Arg)5. Three microliters of sample was applied to a holey carbon grid (Quantifoil, Cu, RL2/1.3, 300 mesh) rendered hydrophilic by a 30 s glow-discharge in air (11 mA current with PIB-10). The grid was blotted for 15 s with a blot force of 0 (sample 1), for 15 s with a blot force of 10 (sample 2), and for 10 s with a blot force of 10 (sample 3) and flash-frozen in liquid ethane using Vitrobot Mark IV (Thermo Fisher Scientific) at 18°C and 100% humidity.

For apo and ATPS-bound states, cryo-EM datasets were acquired with a Talos Arctica (Thermo Fisher Scientific) transmission electron microscope (TEM) operating at 200 kV in nanoprobe mode using EPU software for automated data collection (Supplementary Fig. 5a). The movie frames were collected by a 4 k × 4 k Falcon 3 direct electron detector (DED) in electron counting mode at a nominal magnification of 120,000, which yielded a pixel size of 0.88 Å pixel−1. Forty-nine and fifty movie frames were recorded at an exposure of 1.02 and 1.00 electrons per Å2 per frame, corresponding to a total exposure of 50 e− 2 Å−2, respectively. The defocus range was −1.0 to −2.5 μm for the apo and ATPS-bound states. The dataset for the substrate-bound state was obtained with a 300 kV Titan Krios G3i TEM (Thermo Fisher Scientific) equipped with a Gatan Imaging Filter (GIF), Quantum energy filter (Gatan), and a K3 Summit DED (Gatan) (Supplementary Fig. 5a). The dataset was acquired in counting mode with a nominal magnification of 105,000 (yielding a calibrated pixel size of 0.83 Å pixel−1) using SerialEM software25. Forty-nine movie frames were recorded at an exposure of 1.00 electrons per Å2 per frame, corresponding to a total exposure of 49 e− 2 Å−2 and −0.84 to −1.8 μm of defocus range. Data collection details are listed in Supplementary Table 1.

Cryo-EM data processing
For the apo state dataset, 2,109 acquired movies were subjected to beam-induced motion correction by MotionCor231. The contrast transfer function (CTF) parameters were calculated using Gctf32. A total of 340,000 particles were fully automatically selected using SUPIRE crYOLO33,34 with a generalized model using a selection threshold of 0.1. The subsequent processes were performed using RELION31. The particles were extracted by downsampling to a pixel size of 2.64 Å pixel−1 and subjected to 2D classification. A total of 180,018 particles (17 classes) that displayed secondary-structural elements were selected for ab initio map reconstruction. D20 symmetry was imposed on the generated ab initio map and used as an initial reference map for the 3D classification assuming asymmetry (C1). A total of 103,041 particles associated with the best 3D class with the highest resolution were recentered, re-extracted with a pixel size of 0.88 Å pixel−1 and subjected to 3D refinement with D20 symmetry. Three rounds of CTF refinement and Bayesian polishing followed by 3D refinement with a soft-edged 3D mask yielded a D20 map with 3.11 Å resolution. To further improve the resolution, additional particles were selected using SUPIRE crYOLO33,34 with a relaxed selection threshold of 0.02. A total of 356,730 particles were extracted by downsampling to a pixel size of 2.64 Å pixel−1 and subjected to 2D classification. A total of 429,149 particles (14 classes) that displayed secondary-structural elements were selected. The D20 map with a resolution of 3.11 Å obtained above was used as a reference map for the 3D classification assuming asymmetry (C1). A total of 426,572 particles associated with the best three 3D classes were recentered and re-extracted with a pixel size of 0.88 Å pixel−1 and subjected to 3D refinement with D20 symmetry. Three rounds of CTF refinement and Bayesian polishing followed by 3D refinement with a soft-edged 3D mask yielded a 3.03 Å resolution map. Then, no-alignment 3D classification was conducted (D20 symmetry, two expected classes, T = 8) with a soft-edged 3D mask, and 161,823 particles were selected by

Nature Communications | (2022) 13:5097 https://doi.org/10.1038/s41467-022-32834-8
choosing the best 3D class. The last 3D refinement (D2 symmetry, with a mask diameter of 260 Å) with a soft-edged 3D mask and subsequent postprocessing yielded a map with a global resolution of 2.91 Å according to the Fourier shell correlation (FSC) = 0.143 criteria. The details of the cryo-EM data processing are shown in Supplementary Fig. 2, and the associated parameters and statistics are summarized in Supplementary Table 1. Directional FSC plots were calculated using the 3DFSC server to evaluate the evenness of the distribution of particle orientations (Supplementary Fig. 5b, c).

For the dataset of the ATPγS-bound state, 1,730 acquired movies were subjected to beam-induced motion correction by MotionCor230. The CTF parameters were calculated using Gctf32, and the particles were subjected to beam-induced motion correction by MotionCor230. The CTF parameters were calculated using CTFFIND439, and 3,730 micrographs were then selected with the estimated maximum resolutions. Then, no-alignment 3D classification was conducted using RELION 3.135, and 3,735 micrographs were subjected to beam-induced motion correction by MotionCor230, and then subjected to 3D refinement. After two cycles of CTF refinement38 and Bayesian polishing36, the resulting particles were subjected to two rounds of 3D refinement imposing C2 symmetry (first no 3D mask and second with a soft-edged 3D mask). Additionally, to investigate the possible asymmetrical nature of substrate binding as well as the Gx and Mx modules of our research interests, the last two rounds of 3D refinement were repeated assuming asymmetry. Postprocessing yielded a map with a global resolution of 2.52 Å for the C2 process and 2.64 Å for the C1 process according to the FSC = 0.143 criteria. The details of the cryo-EM data processing are shown in Supplementary Fig. 4, and the associated parameters and statistics are summarized in Supplementary Table 1. Directional FSC plots were calculated using the 3DFSC server to evaluate the evenness of the distribution of particle orientations (Supplementary Fig. 5b, c).

**Model building and refinement**

Each of the four models of Gscore and Mscore in the TeCphA1 tetramer was manually built based on the post-processed D2 cryo-EM map of the apo state using Coot40. Model refinement was performed using phenix.real_space_refine with the secondary-structure restraints generated using phenix.secondary_structure_restraints in the PHENIX program suite41,42. The built models were fit to the C2 cryo-EM map of the substrate-bound state with UCSF Chimera. The fitted models were readjusted into the map and automatically refined using Coot and PHENIX. The structural models of the other modules, the N domain, Gxux, Gux, and Mxux, were generated using the protein structure modeling tools SWISS-MODEL44 and AlphaFold245 to utilize them as initial models for the further model building to generate relatively weak density maps. SWISS-MODEL was run using a part of the sequence of TeCphA1 (residues 218–481) to build the initial models of Gxux and Mxux. The model with the highest GMQE (0.75) and QMEAN (-0.84) scores were obtained from the crystal structure of γ-glutamate-cysteine ligase/glutathione synthetase (GshF, PDB 3LN7), sequence identity of 35.5% as a template. AlphaFold2 predicted the initial models of the N domain (residues 1–161) and Mxux (724–879), in which almost the entire region showed high predicted local distance test (pLDDT) scores (>90), except for the terminal residues (residues 160 161, 724–733, and 877–879) and loop regions (residues 58–65 and 826–833). The initial models were fit to the C2 cryo-EM map of the substrate-bound state with UCSF Chimera. Manual model rebuilding and refinement were performed using Coot and PHENIX, respectively. The refined models were applied to the C1 cryo-EM maps of the substrate-bound state and D2 maps of the apo and ATPγS-bound states. After fitting each module to the maps with UCSF Chimera, the models were readjusted to the map and refined using Coot and PHENIX. Finally, ligands were manually fit in the map of the ATPγS-bound and substrate-bound states. The PDB coordinates AGS and 7ID were used for the models of ATPγS and a β-Asp-Arg dipeptide. All the final models were separately validated using MolProbity in PHENIX. The statistics of the model refinement are summarized in Supplementary Table 1. Figures were generated using UCSF Chimera and PyMol (Schrödinger).

**CphA1 activity assays**

The reaction mixture contained 2.5–10 μg of purified TeCphA1, 50 mM Tris-HCl (pH 8.2), 20 mM KCl, 20 mM MgCl2, 1 mM DTT, 4 mM ATP, 5 mM L-asparate, 0.5 mM L-arginine, and 2 mM synthetic cyanophycin primer peptide of different lengths (1- to 5-mer). These are essentially the same conditions as in a previous report47 with a small modification (replacement of 10 mM 2-mercaptoethanol with 1 mM DTT). For enzyme kinetics, the activity of TeCphA1 (1 μM as a tetramer) was measured by varying the concentration of one of the four substrates,
ATP, L-aspartate, L-arginine, and (β-Asp-Arg)_4. Kinetic analysis revealed that our enzyme-assay conditions for arginine, (β-Asp-Arg)_4, and ATP were appropriate. When the concentration of aspartate is rather low for a stable enzyme-activity measurement, we determined the aspartate concentration at 5 mM by referring to an earlier experiment. In the analysis of TcrPhAI mutants, a 4-mer peptide, (β-Asp-Arg)_4, was used as a primer. The reaction was performed at 30°C for 30 min, with a total reaction volume of 25 µL. To confirm the TcrPhAI activity, the amount of released phosphate was measured by the colorimetric molybdenum blue method. The postreaction solution was mixed with 250 µL of molybdate solution (0.5% ammonium molybdate and 0.5 M H_2SO_4) and 50 µL of SnCl_2 solution (2 mg mL^-1 SnCl_2 dissolved in HCl and 0.5 M H_2SO_4). After incubation for 6 min at 25°C, the absorbance at 660 nm was measured using a Shimadzu UV-1800 spectrophotometer (Shimadzu Corporation). Data were analyzed using GraphPad Prism (GraphPad Software). In the steady-state kinetics, the data for ATP, L-arginine, and (β-Asp-Arg)_4 were fitted with a Michaelis-Menten curve, and the curve fitting for L-aspartate was performed using an allosteric sigmoidal equation as follows:

\[
\text{Reaction rate} = \frac{V_{\text{max}} [S]^h}{K_{\text{half}}^h + [S]^h} = \frac{V_{\text{max}} [S]^h}{K_{\text{prime}}^h + [S]^h}
\]

where \(V_{\text{max}}\) is the maximum enzyme velocity; \([S]\) is the substrate concentration; \(K_{\text{half}}\) is the concentration of substrate that produces a half-maximal enzyme velocity, and \(h\) is the Hill coefficient.

Sequence alignments
A homologous sequence search was performed using the amino acid sequences of TcrPhAI (Accession No. MBS9770029.1) and GshF from Pasteurella multocida (Accession No. WP_001960990.1) in Protein BLAST. CLUSTAL W was used for multiple sequence alignments using default parameters, and the results were displayed by ESPript 3.0. Sequence logos were created using multiple sequence alignment data in the WebLogo server.

Reporting summary
Further information on research reporting is available in the Nature Research Reporting Summary linked to this article.

Data availability
Cryo-EM maps generated in this study have been deposited in the Electron Microscopy Data Bank (EMDB) under accession codes EMD-32381 for TcrPhAI in the apo state, EMD-32382 for TcrPhAI bound with ATP,S, and EMD-32383 (C) and EMD-32384 (C) for TcrPhAI bound with ATP,S, aspartate, and (β-Asp-Arg)_4. The structural coordinates are available in the PDB under accession codes 7WAC for TcrPhAI in the apo state, 7WAD for TcrPhAI bound with ATP,S, and 7WAE (C) and 7WAF (C) for TcrPhAI bound with ATP,S, aspartate, and (β-Asp-Arg)_4. Enzymatic activity data generated in this study are provided in the Source Data file. The PDB coordinates used in this study are as follows: 7LGJ and 7LGQ. Source data are provided with this paper.

References
1. Moradali, M. F. & Rehm, B. H. A. Bacterial biopolymers: from pathogenesis to advanced materials. Nat. Rev. Microbiol. 18, 195–210 (2020).
2. Rehm, B. H. A. Bacterial polymers: biosynthesis, modifications and applications. Nat. Rev. Microbiol. 8, 578–592 (2010).
3. Cai, S. F. et al. A novel DNA-binding protein, PhaR, plays a central role in the regulation of polyhydroxylalkanoate accumulation and granule formation in the haliarchaeon Haloferax mediterranei. Appl. Environ. Microbiol. 81, 373–385 (2015).
4. Frommeyer, M., Wielf, L. & Steinbuechel, A. Features of the biotechnologically relevant polyamide family “cyanophycins” and their biosynthesis in prokaryotes and eukaryotes. Crit. Rev. Biotechnol. 36, 153–164 (2016).
5. Du, J. K., Li, L. & Zhou, S. N. Microbial production of cyanophycin: From enzymes to biopolymers. Biotechnol. Adv. 37, 107400 (2019).
6. Numata, K. Poly(aminocids)/polypeptides: Advantages of potential functional and structural materials. Polym. J. 47, 537–545 (2015).
7. Simon, R. & Weathers, P. Determination of the structure of the novel polypeptide containing aspartic acid and arginine which is found in cyanobacteria. Biochim. Biophys. Acta. Protein Struct. 420, 165–176 (1976).
8. Hoffman, B. M. et al. Mechanism of nitrogen fixation by nitrogenase: the next stage. Chem. Rev. 114, 4041–4062 (2014).
9. Burnat, M., Herrero, A. & Flores, E. Compartmentalized cyanophycin metabolism in the diazotrophic filaments of a heterocyst-forming cyanobacterium. Proc. Natl Acad. Sci. USA. 111, 3823–3828 (2014).
10. Finzi-Hart, J. A. et al. Fixation and fate of C and N in the cyanobacterium Trichodesmium using nanometer-scale secondary ion mass spectrometry. Proc. Natl Acad. Sci. USA. 106, 6345–6350 (2009).
11. Watzer, B. & Forchhammer, K. Cyanophycin synthesis optimizes nitrogen utilization in the unicellular cyanobacterium Synechocystis sp. strain PCC 6803. Appl. Environ. Microb. 84, e01298–18 (2018).
12. Aboulmagd, E., Sanio, F. B. O. & Steinbuchel, A. Purification of Synechocystis sp. strain PCC6308 cyanophycin synthetase and its characterization with respect to substrate and primer specificity. Appl. Environ. Microb. 67, 2176–2182 (2001).
13. Sharon, I. et al. Structures and function of the amino acid polymerase cyanophycin synthetase. Nat. Chem. Biol. 17, 1101–1110 (2021).
14. Capone, D. G. et al. Trichodesmium, a globally significant marine cyanobacterium. Science 276, 1221–1229 (1997).
15. Smith, C. A. Structure, function and dynamics in the mur family of bacterial cell wall ligases. J. Mol. Biol. 362, 640–655 (2006).
16. Deva, T., Baker, E. N., Squire, C. J. & Smith, C. A. Structure of Escherichia coli UDP-N-acetylmuramoyl-L-alanine ligase (MurC). Acta Crystallogr. D. Biol. Crystallogr. 62, 1466–1474 (2006).
17. Favini-Stabile, S., Contreras-Martel, C., Thielen, N. & Dessen, A. MreB and MurG as scaffolds for the cytoplasmic steps of peptidoglycan biosynthesis. Environ. Microbiol. 15, 3218–3228 (2013).
18. Berg, H. et al. Biosynthesis of the cyanobacterial reserve polymer multi-L-arginyl-poly-L-aspartic acid (cyanophycin): mechanism of the cyanophycin synthetase reaction studied with synthetic primers. Eur. J. Biochem 267, 5561–5570 (2000).
19. Krehenbrink, M. & Steinbüchel, 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. Microbiol. (Read.) 150, 2599–2608 (2004).
20. Fawaz, M. V., Topper, M. E. & Firestine, S. M. The ATP-grasp enzymes. Bioorg. Chem. 39, 185–191 (2011).
21. Hara, T., Kato, H., Katsube, Y. & Oda, J. A pseudo-michaelis quaternary complex in the reverse reaction of a ligase: structure of Escherichia coli β glutathione synthetase complexed with ADP, glutathione, and sulfate at 2.0 Å resolution. Biochemistry 35, 11967–11974 (1996).
22. Li, H., Fast, W. & Benkovic, S. J. Structural and functional modularity of proteins in the de novo purine biosynthetic pathway. Protein Sci. 18, 881–892 (2009).
23. Mol, C. D. et al. Crystal structures of active fully assembled substrate- and product-bound complexes of UDP-N-acetylmuramic acid-L-alanine ligase (MurC) from Haemophilus influenzae. J. Bacteriol. 185, 4152–4162 (2003).
24. Šink, R. et al. MurD enzymes: some recent developments. *Biomol. Concepts* **4**, 539–556 (2013).

25. Stout, J. S., De Vos, D., Vergauwen, B. & Savvides, S. N. Glutathione biosynthesis in bacteria by bifunctional glutaredoxin is driven by a modular structure featuring a novel hybrid ATP-grasp fold. *J. Mol. Biol.* **416**, 486–494 (2012).

26. Arai, T. & Kino, K. A cyanophycin synthetase from *Thermosynechococcus elongatus* BP-1 catalyzes primer-independent cyanophycin synthesis. *Appl. Microbiol. Biotechnol.* **81**, 69–78 (2008).

27. Zhang, H. et al. The cyanobacterial ornithine–ammonia cycle involves an arginine dihydrolase. *Nat. Chem. Biol.* **14**, 575–581 (2018).

28. Ziegler, K. et al. Molecular characterization of cyanophycin synthetase, the enzyme catalyzing the biosynthesis of the cyanobacterial reserve material multi-L-arginyl-poly-L-aspartate (cyanophycin). *Eur. J. Biochem.* **254**, 154–159 (1998).

29. Klemke, F. et al. CphA2 is a novel type of cyanophycin synthetase in *N*-fixing cyanobacteria. *Microbiology* **162**, 526–536 (2016).

30. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).

31. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

32. Zhang, K. Cctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).

33. Moriya, T. et al. High-resolution Single Particle analysis from Electron Cryo-microscopy Images Using SPHIRE. *J. Vis. Exp.* **796** (2017).

34. Wagner, T. et al. SPHIRE-cryOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, 218 (2019).

35. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, e42166 (2018).

36. Zivanov, J., Nakane, T. & Scheres, S. H. W. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6**, 5–17 (2019).

37. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryocryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).

38. Tan, Y. Z. et al. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* **14**, 793–796 (2017).

39. Rohou, A. & Grigorieff, N. CTFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Bio.* **192**, 216–221 (2015).

40. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. In *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501 (2010).

41. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. In *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213–221 (2010).

42. Wang, L. et al. Real-space quantum-based refinement for cryo-EM: QRR3. In *Acta Crystallogr. D. Biol. Crystallogr.* **76**, 1184–1191 (2020).

43. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

44. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).

45. Tunyasuvunakool, K. et al. Highly accurate protein structure prediction for the human proteome. *Nature* **596**, S90–S96 (2021).

46. Chen, V. B. et al. Molprobity: all-atom structure validation for macromolecular crystallography. In *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 12–21 (2010).

47. Nagui, E. A., McKelvie, I. D., Worsfold, P. & Kolev, S. D. The molybdenum blue reaction for the determination of orthophosphate revisited: opening the black box. *Anal. Chim. Acta* **890**, 60–82 (2015).

48. Copeland, R. A. *Enzymes: A Practical Introduction to Structure, Mechanism, and Data Analysis*, 2nd Edition (Wiley, Hoboken, 2000).

49. Als chol, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

50. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).

51. Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308 (1999).

52. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).

**Acknowledgements**

We thank Y. Sakamaki and M. Kikkawa (Cryo-EM facility at the University of Tokyo) for their help in cryo-EM data collection and A. Nakamura (Gakusyuin University) for the technical support to build the structural model of (β-Asp-Arg)₄. This research was supported by the Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)) from AMED under Grant Number JP21am0101077/JP22ama12101 (T. Miyakawa) and JP21am0101071 (T. Muramatsu) (support number 1678) (T.S.), and the Finance Science and Technology Project of Hainan Province under Grant Number ZDKJ202018.

**Author contributions**

M.T. and T.S. supervised the project. T.Miyakawa, J.Y., and M.T. designed the experiments. J.Y., A.F., and Y.M. performed recombinant proteins. A.F. and Y.M. performed biochemical studies with assistance from T.Miyakawa. T.Miyakawa, N.A., and M.K. collected cryo-EM data. T.Miyakawa and M.K. processed the data with the support of N.A., T.Moriya, and T.Muramatsu. T.Miyakawa and A.F. performed model building and refinement. T.Miyakawa, J.Y., M.K., N.A., A.F., and T.Moriya wrote the manuscript with assistance from all the authors. M.T. and T.S. edited the manuscript.

**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/s41467-022-32834-8.

**Correspondence** and requests for materials should be addressed to Toshiya Senda or Masaru Tanokura.

**Peer review information** Nature Communications thanks the 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.
