Structure of CphA2 from *Stanieria* sp.; insights into oligomeric behaviour and substrate binding

Linda M.D. Markus
Minor Research Internship Report

Supervision by
Prof. Dr. T. Martin Schmeing
Dr. Mike Strauss
Prof. Dr. Friedrich G. Förster
Itai Sharon, MSc

Contents

Laymen summary

Cyanobacteria, commonly known as blue-green algae, are bacteria that obtain energy through photosynthesis. A well-known issue in cyanobacteria is the separation of nitrogen fixation and photosynthesis. Cyanobacteria take up nitrogen from their environment to use in cellular processes, for example protein biosynthesis. However, the enzyme that fixates nitrogen, nitrogenase, is inactive in the presence of oxygen. Cyanobacterial species have developed several methods to separate nitrogen fixation and consumption, one of these methods involves the polymer cyanophycin. Cyanophycin is an efficient molecule for storage of nitrogen. By using cyanophycin, bacteria can fixate nitrogen when oxygen levels are low and store the fixed nitrogen for consumption when oxygen levels are high. Cyanophycin consists of a poly-aspartate backbone with arginine residues attached as sidechains and can be produced by two enzymes, cyanophycin synthetase 1 and 2 (CphA1 and CphA2). Besides its function as a nitrogen storage, cyanophycin has many commercial applications. For example, the polymer or derivatives can be used as biodegradable plastics, water softener or bandage material. Insights into the production of cyanophycin are of great importance for efficient large-scale production.

In this study, I aim to gain insights into the production of cyanophycin by CphA2. CphA2 uses three substrates, cyanophycin, β-aspartate-arginine (β-Asp-Arg) and ATP, to elongate cyanophycin. Two aspects of CphA2 are not completely understood and were studied in this research. The first aspect is the oligomeric behaviour of CphA2 from the cyanobacterial species *Stanieria* sp.. While CphA2 from a variety of species forms dimers, *Stanieria* sp. CphA2 forms higher oligomers. The exact oligomeric state and the structure of this oligomer remained unknown. The second aspect is the binding of the three substrates to CphA2. Previous studies have revealed the binding site of substrates in CphA1, however, it is not known what the similarities and differences are between CphA1 and CphA2 in substrate binding.

In this research project, I used cryo-electron microscopy (cryo-EM) to obtain the structure of CphA2 from the cyanobacterial species *Stanieria* sp.. Cryo-EM is a microscopy technique where the protein of interest is frozen at cryogenic temperatures. An electron beam is used to image the sample at a resolution that is not obtainable by light microscopy. Cryo-EM allows visualization of protein structure in a near-native state. Using cryo-EM, the structure of *Stanieria* sp. was obtained, showing a hexameric form. The hexamer was determined to be stable in solution. It was found that the hexamer formation is important for activity of the enzyme, because the hexamer was twice as active as a dimeric mutant. Furthermore, insights into the binding of substrates cyanophycin and ATP was obtained by the cryo-EM structure of *Stanieria* sp. CphA2 with substrates. The binding of cyanophycin and ATP has similarities and differences from their binding in CphA1. The approximate binding site of substrate β-Asp-Arg was determined by mutagenesis experiments.

Abstract

Cyanophycin is a natural polymer found in a wide range of bacteria, where it functions as a reservoir for fixed nitrogen. Additionally, cyanophycin has various commercial application in bio-industry and biomedicine. Cyanophycin consists of a poly-aspartate backbone with arginines attached to their sidechains. The polymer is synthesized by cyanophycin synthetase 1 and 2 (CphA1 and CphA2). While CphA1 uses aspartate and arginine as substrate, CphA2 uses the dipeptide β-Asp-Arg. Although the structure of dimeric CphA2 from *G. citriformis* has been resolved in a previous study, questions about oligomeric behaviour and substrate binding in CphA2 remained unanswered. In this study, I resolved the structure of the hexamer-forming CphA2 homolog from *Stanieria* sp. to 2.76 Å by cryo-EM. The hexamer form is stable and is important for enzyme activity. Structures of *Stanieria* sp. CphA2 with substrates obtained by cryo-EM at resolutions of 2.6 Å and 2.7 Å show the binding of cyanophycin and ATP or ATP analog, revealing similarities to and differences from substrate binding in CphA1. Moreover, mutagenesis experiments provided insights into the binding site of β-Asp-Arg. The structures and biochemical assay performed in this study assist in understanding of the reaction mechanism of cyanophycin formation by CphA2.

Introduction

Cyanophycin is a natural polymer discovered in the 19th century by Antonio Borzì in cyanobacteria.[1] It is composed of a poly-L-aspartate backbone with arginine side chains, where the amino group of the arginine is attached to the carboxyl group of the aspartate sidechain through an isopeptide bond (Fig. 1A).[2] One cyanophycin polymer typically ranges in length from 80 to 400 Asp-Arg polypeptides and in molecular mass from ~20 to ~100 kDa.[3]

The biological function of cyanophycin in cyanobacteria is as storage of fixed nitrogen.[4] With a nitrogen content of 24% by mass, cyanophycin can store nitrogen efficiently.[5] Cyanobacteria require nitrogen for many cellular processes, including protein biosynthesis. However, nitrogen fixation cannot occur simultaneously with photosynthesis because nitrogenase, the enzyme responsible for nitrogen fixation, is inactivated by oxygen.[6] Therefore, cyanobacteria need to separate nitrogen fixation and photosynthesis either spatially or temporally.[7] An example of this separation are single-cell cyanobacteria, where nitrogen taken up from the environment is fixated and stored in cyanophycin granules during periods of darkness, when oxygen levels are low. During periods of light, cyanophycin is consumed in protein biosynthesis and other cellular processes.[7] Furthermore, cyanophycin can also function as storage of carbon and energy.[8,9]

Besides its function in cyanobacteria, cyanophycin has interesting properties for commercial application as a green alternative to fossil-derived industrial products.[10] It is soluble in many conditions, has a high viscosity and is completely biodegradable.[11,12] Cyanophycin and cyanophycin derivatives have a wide range of potential applications, from biotechnology, for example as biodegradable plastics,[13] super-absorbant or antiscalant,[10] to biomedicines, as wound healing bandage material[14] or in nutrition.[15] Large-scale production of cyanophycin for industrial applications has been limited by the slow growth of cyanobacteria and low yield.[16] In recent years, many studies have aimed to improve production of cyanophycin,[17] for example by heterologous expression in different species.[18,19] Further insights into the production of cyanophycin can greatly benefit these studies to establish an efficient production method for wide-spread use.

In bacteria, cyanophycin synthesis is facilitated by cyanophycin synthetase 1 (CphA1) and 2 (CphA2).[20,21] The two cyanophycin synthetases have partially overlapping functions but use different substrates. CphA1 uses aspartate and arginine

as substrates in two sequential reactions, a first reaction extends the poly-L-aspartate backbone by one aspartate and a second reaction attaches an arginine to the recently added aspartate.[22] CphA2 produces cyanophycin by polymerization of the dipeptide β-aspartate-arginine (β-Asp-Arg), requiring only one reaction (Fig. 1B).[21,23] CphA1 and CphA2 are evolutionary related to each other. While CphA1 is present in a wide range of bacteria,[24] CphA2 evolved more recently in cyanobacteria.[21,23] In this study, I will focus on the structure and function of CphA2.

Cyanophycin synthesis by CphA2 is a two-step reaction.[21,25] In the first step, the terminal carboxyl group of an existing cyanophycin molecule is phosphorylated using ATP. In the second step, the resulting acylphosphate group undergoes nucleophilic attack by the α-amino group of the β-Asp-Arg dipeptide, resulting in the elongation of cyanophycin by one β-Asp-Arg dipeptide (Fig. 1B).[25]

Recently, the structure of CphA2 from the cyanobacterium *Gloeothece citriformis* PCC7424 was resolved using X-ray crystallography.[25] The structure shows that CphA2 consists an N-terminal domain with no structure similarity to other proteins besides CphA1 (N domain),[25] a middle domain homologous to gluthatione synthetase (G domain),[22,26] and a C-terminal domain with homology to MurE-like muramyl ligases (M domain).[27] The domains are arranged with the N domain between the G and M domains (Fig. 1C). The active site of CphA2 likely resides in the G domain, based on structural study of CphA1.[28] By obtaining the cryo-EM structure of CphA1 *Synechocystis* sp. UTEX2470 (*Su*CphA1) with substrates, Sharon *et al.* were able to visualize the binding of cyanophycin analog (β-Asp-Arg)$_8$-NH$_2$ and an ATP analog in the active sites in the G domain (Fig. 1D).[28] The G domain consists of a G_{core} region and two lobes, G_{lid} and G_{omega}. The binding pocket for cyanophycin is located in the G_{core}, ATP binds between the G_{core} and the G_{lid} lobe. Sharon *et al*. identified residues S166, E215 and R309 as important in the binding of cyanophycin (Fig. 1D).[28] Although Sharon *et al.* were not able to obtain a structure of *G. citriformis* CphA2 with substrates, they gained insight into the cyanophycin binding site by mutagenesis of residues corresponding to the aforementioned residues of CphA1.[25] The resulting reduction or absence of activity of the mutants supports the hypothesis that CphA2 binds cyanophycin similar to CphA1, however, this hypothesis had not been proven yet.

Moreover, little is known about the binding site of β-Asp-Arg. No structure has been obtained of CphA2 with β-Asp-Arg or CphA1 with aspartate in the active site of the G domain. It has been suggested that two loops in the G_{omega} lobe, the large loop (residues 374-381 in *G. citriformis* CphA2) and an adjacent loop (328-340) are involved in the binding of β-Asp-Arg.[25,28,29] The loops are conserved among homologs of CphA2 and contribute to a shallow pocket where β-Asp-Arg could be orientated correctly for nucleophilic attack of cyanophycin acylphosphate intermediate. Structural information on the binding of β-Asp-Arg to CphA2 could provide insights into the reaction mechanism of cyanophycin synthesis.

An interesting aspect of cyanophycin synthetases is its oligomeric behavior. The crystal structure of CphA2 from *G. citriformis* showed that the protein adopts a dimeric state,[25] while CphA1 was found to exist as a tetramer.[28] Furthermore, Sharon *et al.* studied the oligomeric states of CphA2 from nine different cyanobacterial species by size exclusion chromatography (SEC) (Fig. 1E).[28] Seven homologs, including CphA2 from *G. citriformis*, eluted as a single species with masses between 156 and 186 kDa, suggesting a dimeric state for these homologs.[25] Two homologs showed a different elution pattern. CphA2 from *C. elsteri* showed a major peak at ~217 kDa and a minor peak at ~460 kDa suggested to correspond to a trimer and hexamer respectively, however the homolog was inactive.[25] More interestingly, CphA2 from *Stanieria* sp. was active and eluted as one peak

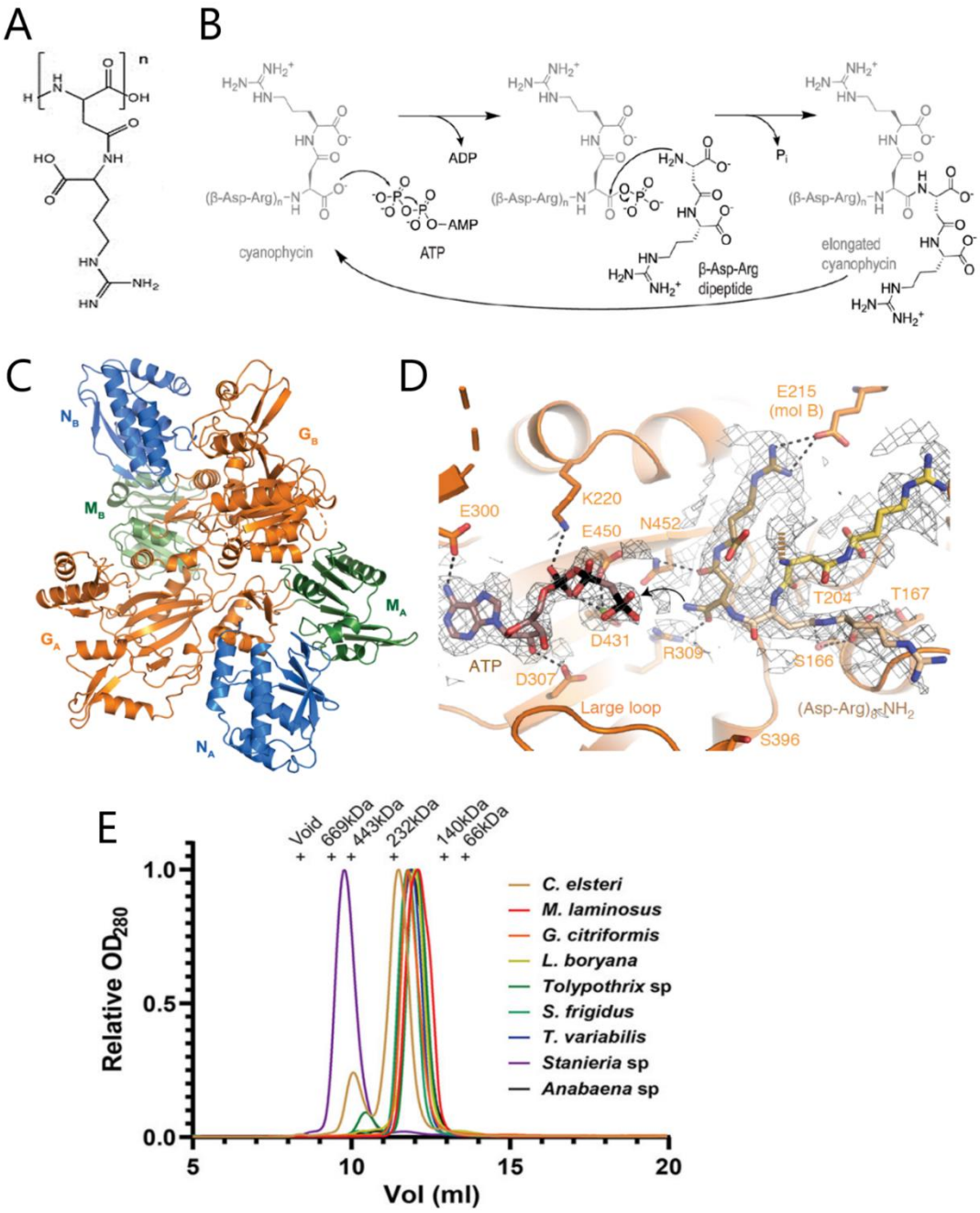

Figure 1: A: Chemical structure of cyanophycin. B: Reaction mechanism of cyanophycin synthesis by CphA2. The two-step reaction consists of phosphorylation of the C-terminal carboxyl of cyanophycin by ATP and subsequent nucleophilic attack by the α-amino of β-Asp-Arg. C: Structure of the *G. citriformis* CphA2 biological dimer.[25] D: Structure of the *Su*CphA1 G domain active site with $(Asp\text{-}Arg)_8\text{-}NH_2$ and ADPCP.[28] E: SEC chromatograms of nine CphA2 homologs. Peaks were normalized to the maximal peak height.[25]

with a mass of ~536 kDa, suggested to correspond to an heptameric or octameric state.[25] The exact oligomeric state of CphA2 from *Stanieria* sp., the structural arrangement, and the relevance of this oligomeric state remained unknown.

In this research project, I focused on the CphA2 homolog from *Stanieria* sp., with the aim to provide insights into the oligomeric state of this homolog. I have discovered that CphA2 from *Stanieria* sp. forms a stable hexameric species and have resolved the structure of the hexamer to 2.76 Å by cryo-electron microscopy. Furthermore, I successfully visualized ATP and cyanophycin bound in the expected region by resolving

cryo-EM structures of *Stanieria* sp. CphA2 to 2.6 and 2.7 Å, showing similarities to and differences from the binding of cyanophycin in CphA1.

Results

Oligomeric behaviour of *Stanieria* sp. CphA2

In order to establish that *Stanieria* sp. CphA2 forms the higher oligomeric species reported by Sharon *et al.* at concentrations suitable for cryo-EM, SEC was performed with a concentration range of 0.8 to 3.6 mg/ml. The results show that at all concentrations, the protein eluted as a higher oligomer similar to *Stanieria* sp. CphA2 reported by Sharon *et al.* (Fig. 2A).[25] This suggests that higher oligomer formation is stable over a range of concentrations.

To determine the oligomeric state of *Stanieria* sp. CphA2, analytical ultracentrifugation was performed. The c(S) distribution showed that 93.5% of protein particles migrated in one peak with a molecular weight corresponding to approximately 434 kDa, roughly the size of a hexamer (428 kDa) (Fig. 2B). 2.86% of particles migrated according to a molecular weight of approximately 878 kDa, corresponding to a dodecamer (876 kDa). Interestingly, 0.00% of particles migrated according a molecular weight smaller than a hexamer, indicating that the hexameric structure does not dissociate into smaller species under these conditions. From these results, we can conclude that the oligomeric form of *Stanieria* sp. CphA2 is a hexamer and the complex is very stable in solution.

Cryo-EM structure of *Stanieria* sp. CphA2

To understand the nature of *Stanieria* sp. CphA2 hexamerization, we used cryo-EM to determine the structure of the hexameric form (Fig. S1A). During data processing, 2D classification indicated that different species with various sizes were present (Fig. S1B) and 3D classification yielded three unique classes corresponding to hexameric (42.0% of particles), tetrameric (22.2%) and dimeric (7.8%) species (Fig. S1C). Because AUC showed no dimeric or tetrameric particles, these species likely arise as an artefact from interaction on the grid or with the air-water interface and have not been considered for further processing. Further processing of the hexameric class resulted in a map with a global resolution of 2.76 Å (Fig. S1D) into which the

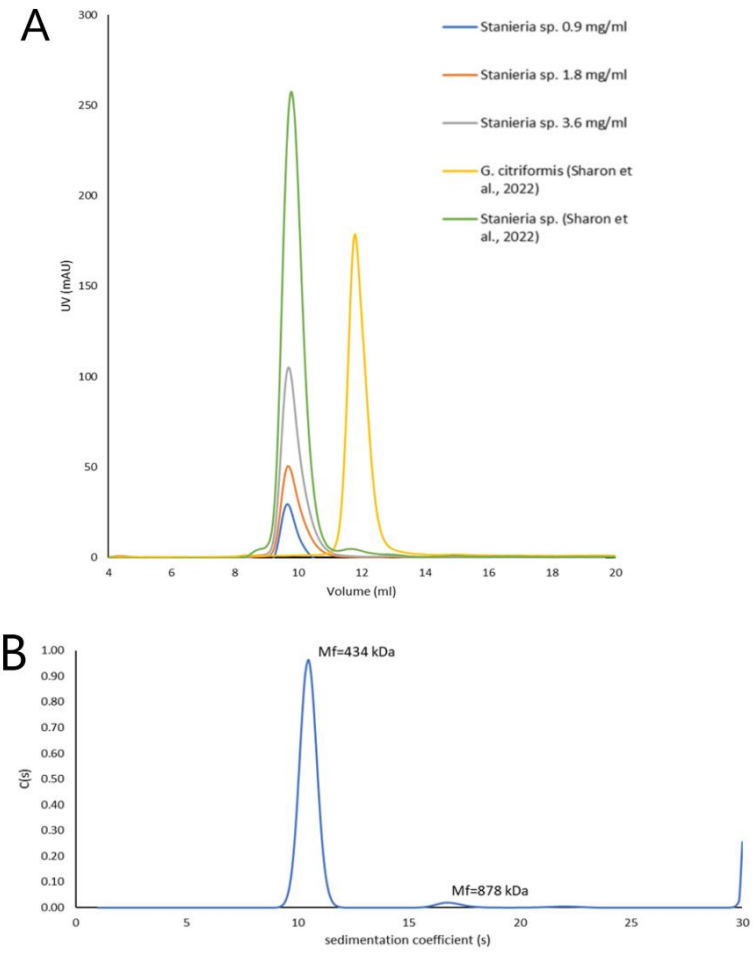

Figure 2: A: SEC chromatograms of *Stanieria* sp. ChpA2 at concentrations of 0.9 mg/ml (blue), 1.8 mg/ml (orange), 3.6 mg/ml (grey) compared to SEC chromatograms of *Stanieria* sp. CphA2 and *G. citriformis* CphA2 from previous study.[25] B: Sedimentation distribution measured by AUC of *Stanieria* sp. CphA2. Mf = frictional coefficient-dependent molecular weight.

structure of *Stanieria* sp. CphA2 was built (Fig. 3A and B). The structure shows that *Stanieria* sp. CphA2 forms a hexamer with the shape of an open disc, with C2 symmetry along the axis shown in figure 3A.

The hexamer can be divided into three dimers that are similar to the dimer structure of *G. citriformis* CphA2 (Fig. 3C and D).[25] As expected, each *Stanieria* sp. CphA2 molecule consists of the N domain (residues 1-158), G domain (159-486) and M domain (487-636) (Fig. 3C). The domains are arranged with the N domain residing between the G and M domains. The dimer interface is formed mainly by two G domain helices (204-228) and a loop (181-184) (Fig. 3D). The interface between two monomers in a dimer buries approximately 1220 Å^2. In the hexameric structure, the M domains and G_{core} are oriented towards each other, forming a rigid core, while the N domains and G_{lid} (234-300) and G_{omega} (324-398) lobes reach outwards (Fig. 3B). In this orientation, the active site residing in the G domain faces the solvent and is therefore accessible for binding of substrates.

The monomer and dimer structures of the *Stanieria* sp. CphA2 have similarities to and differences from *G. citriformis* CphA2 (Fig. 3C

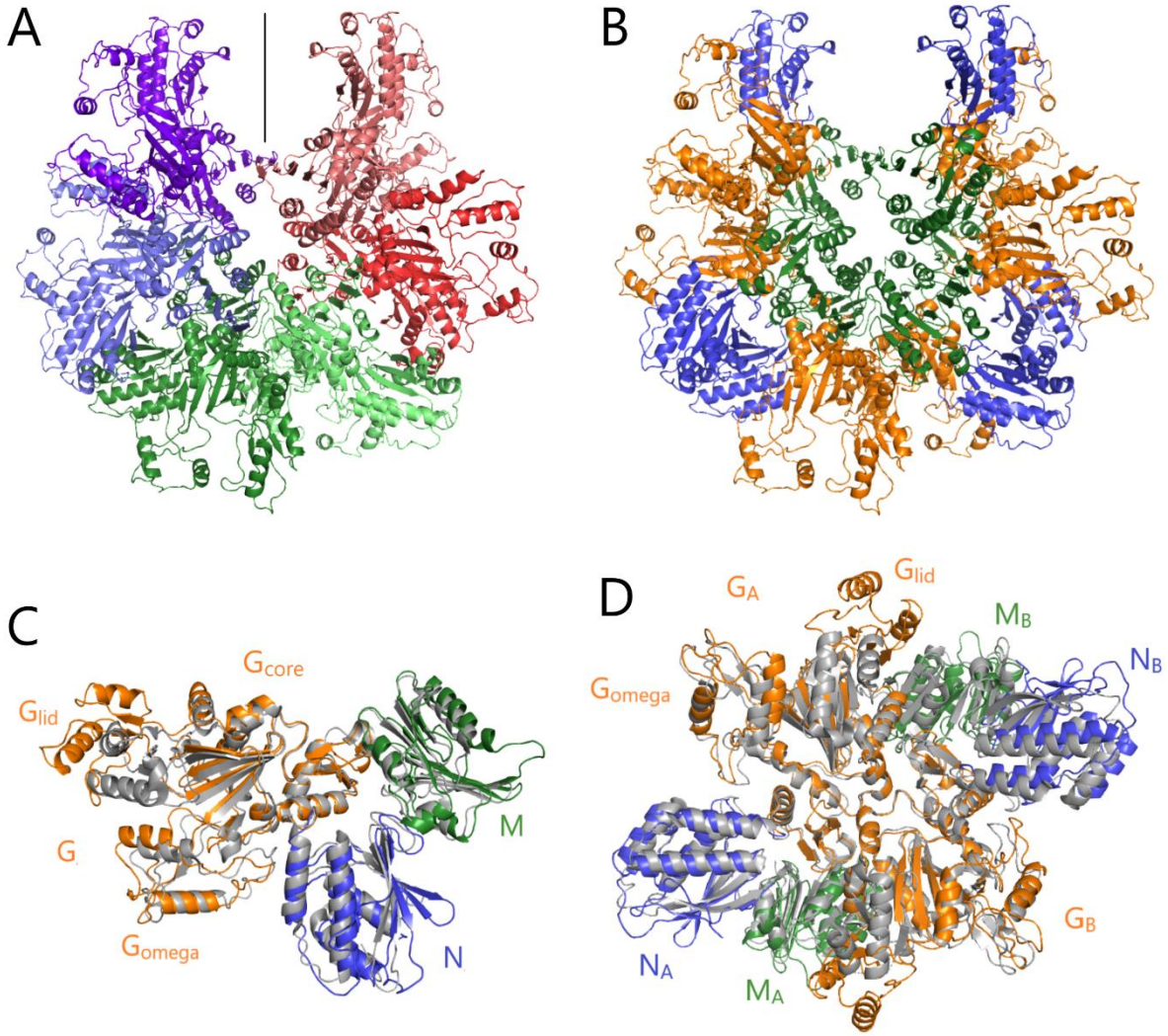

Figure 3: A: Hexameric structure of *Stanieria* sp. CphA2 at 2.76 Å, colored by molecule. Black line indicates C2 symmetry axis. B: Hexameric structure of *Stanieria* sp. CphA2 at 2.76 Å, colored by domain, blue: N domain, orange: G domain, green: M domain. C: molecule of *Stanieria* sp. CphA2 from the hexameric structure, colored as in B, compared to monomeric *G. citriformis* CphA2 (grey).[25] D: dimer of two *Stanieria* sp. CphA2 molecules from the hexameric structure, colored as in B, compared to monomeric *G. citriformis* CphA2 (grey).[25]

and D). The domains have the same arrangement in both homologs and the dimer interface is similar. The main difference between the obtained model of *Stanieria* sp. CphA2 and *G. citriformis* is the conformation of the G_{lid} lobe. While the lobe adopts a closed conformation for *G. citriformis* CphA2, the model of *Stanieria* sp. CphA2 shows an open conformation. However, 3D variability analysis of *Stanieria* sp. CphA2 showed that particles with both open and closed G_{lid} conformations were present in the sample (Movie 1; Fig. S3A and B).

Mutagenesis of the dimer-dimer interface

The dimer-dimer interface buries a surface area of approximately 1684 $Å^2$, consisting of the most C-terminal β-strand (612-617) and α-helix (624-636) of a molecule of one dimer interacting with both molecules from the neighboring dimer (Fig. 4A). Identifying molecules A-F according to figure 4A, molecule B of the dimer formed by A and B interacts with molecules C and D. In the binding interface between molecules B and C, R315 and Q416 of molecule C form hydrogen bonds with the backbone of residues K610 and L613, respectively, of molecule B (Fig. 4B). In the binding interface between molecules B and D, R528 of molecule D and Y616 of molecule B form a stacking interaction (Fig. 4C). Additionally, R528 forms electrostatic interactions with E619 and E622 of molecule B.

In order to determine the effect of hexamerization on the activity of *Stanieria* sp. CphA2, I aimed to use the obtained knowledge on the dimer-dimer interface to produce a dimeric mutant of *Stanieria* sp. CphA2 and compare the activity to the hexamer. I made a G611' mutant truncating the C-terminal β-strand and α-helix (Fig. S2A) and mutated the aforementioned residues at the dimer-dimer interface to R315A, Q416A, R528G and Y616R (Fig. 4B and C).

The ability of the mutants to form hexamers was assessed by size exclusion chromatography. The G611' mutant appeared unstable on the SEC chromatogram (Fig. S2B) and was not pursued further. For the R315A mutation, the chromatogram was identical to wildtype *Stanieria* sp. CphA2 concluding that this mutation has no effect on hexamer formation (Fig. 4D, yellow). The Y616R mutant appeared as a predominantly hexameric species, with a main peak at the same volume as the wildtype (Fig. 4D, light blue). A small peak was present with a molecular weight of approximately 140 kDa, corresponding to a dimer (146 kDa). Moreover, the main peak had a shoulder suggesting intermediates between hexamer and dimer. However, the Y616R mutation did not cause significant disruption of the hexamer into dimers. Single mutants Q416A and R528G appeared as two distinct peaks after SEC, with elution volumes corresponding to molecular weights of approximately 300 and 140 kDa respectively (Fig. 4D, orange and green). The peak corresponding to 140 kDa represents a dimeric species. The percentage of dimeric species is approximately 80% for Q416A and 50% for R528G based on peak area. These results show that mutation of Q416 or R528 partially disrupts hexamer formation, however, there is a significant amount of higher oligomers present. In order to obtain a predominantly dimeric sample, a Q416A R528G double mutant was made. The double mutant appeared as 90% dimer by SEC (Fig. 4D, dark blue). To determine the activity of the dimer compared to hexamer, an activity assay was performed with the Q416A R528G double mutant and wildtype (WT). Relative to the hexameric wildtype, the dimeric Q416A R528G double mutant showed ~50% reduction in activity (Fig. 4E). These result show that formation of the hexamer has a significant effect on the synthesis of cyanophycin.

Visualization of CphA2 active site with substrates

To visualize the active site of CphA2 with substrates bound, a cryo-EM dataset was obtained of *Stanieria* sp. CphA2 in the presence of β-Asp-Arg, cyanophycin $(β\text{-Asp-Arg})_4$, the β-Asp-Arg dipeptide, and the ATP

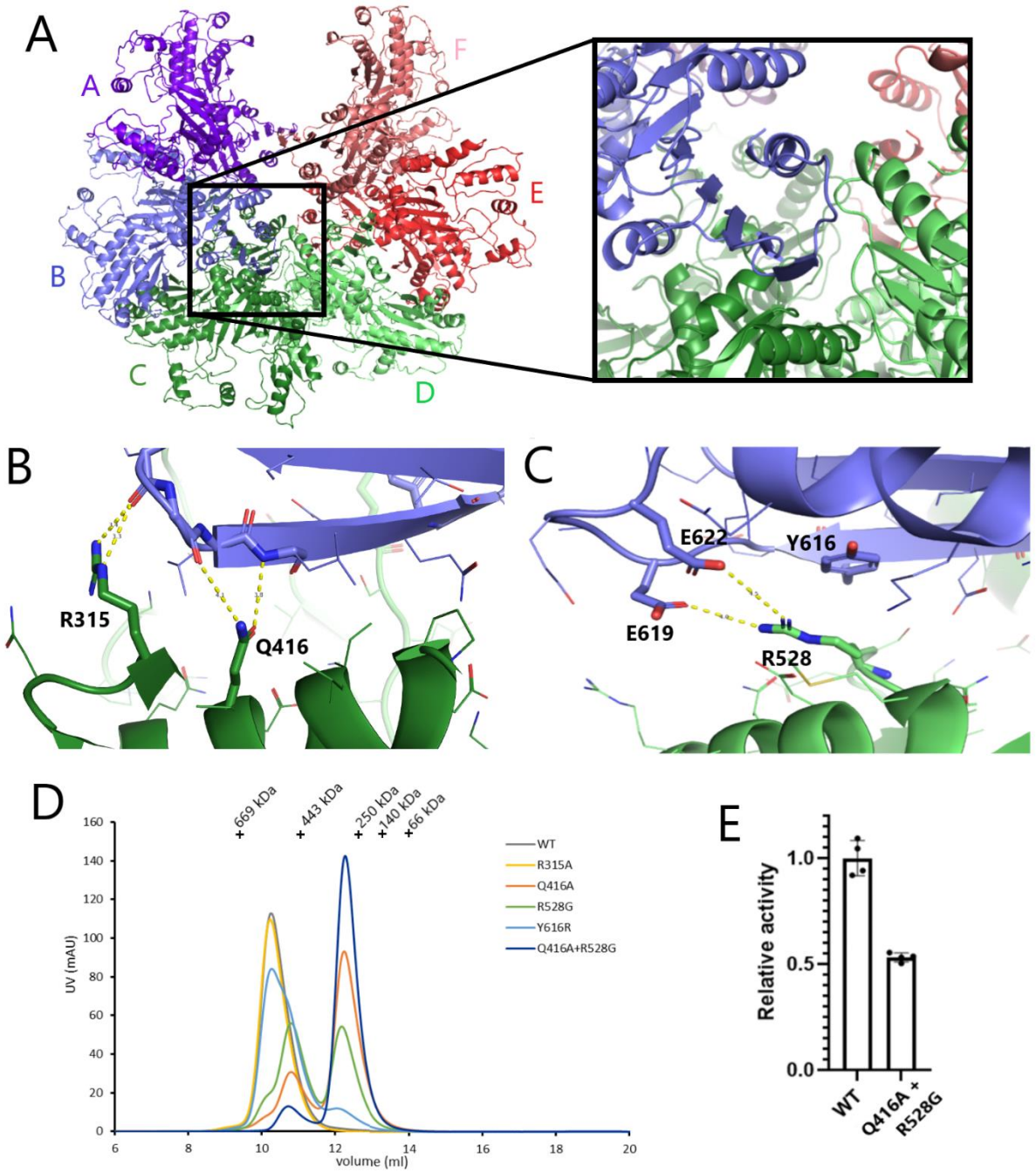

Figure 4: A: Dimer-dimer interface between molecule B from A-B dimer of and molecule C and D from the C-D dimer. B: Interactions between R315 and Q416 of molecule C and backbone of molecule B. C: Interactions between R528 from molecule D and Y616, E619 and E622 of molecule B. D: SEC chromatograms of single mutants R315A (yellow), Q416A (orange), R528G (green), Y616R (light blue) and double mutant Q416A+R528G (dark blue). E: Cyanophycin synthesis activity of wildtype (WT) *Stanieria* sp. CphA2 and double mutant Q416A+R528G. Activity was normalized to the activity of the WT.

analog 5'-adenosyl-methylene-triphosphate (ADPCP). ADPCP was used instead of ATP to catch the enzyme in a state where substrates are bound without allowing the reaction to proceed. The dataset resulted in a map at 2.60 Å resolution and hexameric structure similar to the structure of *Stanieria* sp. CphA2 without substrates (S4A and B).

Binding of cyanophycin and ATP

In the structure, strong density is present for the ATP analog and cyanophycin. ADPCP and

cyanophycin bind in the binding pocket in the G domain (Fig. 5A), as expected from the structure of *Su*CphA1. For cyanophycin, the three C-terminal β-Asp-Arg dipeptides are visible. The carboxyl group on the most C-terminal aspartate of cyanophycin is orientated towards ATP and interacts with R308 (Fig. 5A). From C-terminus to N-terminus, the first dipeptide extends towards the G_{core} of the other monomer in the dimeric unit (molecule B) where the guanidinium group interacts with D212, and H208 from molecule B. The second dipeptide extends towards the N domain and forms hydrogen bonds with N156 and T163 (Fig. 5A). The third dipeptide resides between the first and second dipeptides and interacts with both D212 and H208 from molecule B (Fig. 5A). The most N-terminal dipeptide reaches into the solvent and is has little observable density due to flexibility (Fig. 5A). The observed interactions of cyanophycin with R308, T163 and D212 are in agreement with the loss of activity after mutagenesis of the corresponding residues R292 and S148 in *G. citriformis* CphA2.[25]

The binding of the ATP and cyanophycin analogs in *Stanieria* sp. CphA2 is mostly similar to the binding in *Su*CphA1 with minor differences in the orientation of the first and third dipeptide of cyanophycin (Fig. 5B). The interactions of cyanophycin with R308 and T163 are similar to the interactions with corresponding residues R309 and S166 in *Su*CphA1 (Fig. 5B).[28] Residue D212 in *Stanieria* sp. CphA2 and the corresponding residue D197 in *Su*CphA1 both interact with cyanophycin, but in different ways. In *Stanieria* sp. CphA2, D212 from molecule A interacts with the C-terminal dipeptide and D212 from molecule B interacts with the third dipeptide, while in *Su*CphA1, D212 from molecule B interacts with the first dipeptide (Fig. 5B). Furthermore, an interesting difference is the presence of I211 in *Su*CphA1 instead of H208 in *Stanieria* sp. CphA2. H208 interacts with the guanidinium groups of both the first and third dipeptide (Fig. 5A). The absence of the hydrophilic group at this position in *Su*CphA1 likely plays a role in the difference in orientation of the first and third dipeptide between *Su*CphA1 and *Stanieria* sp. CphA2. Another difference between the structures that likely affects the orientation of the third cyanophycin dipeptide is the orientation of the G domain loop at

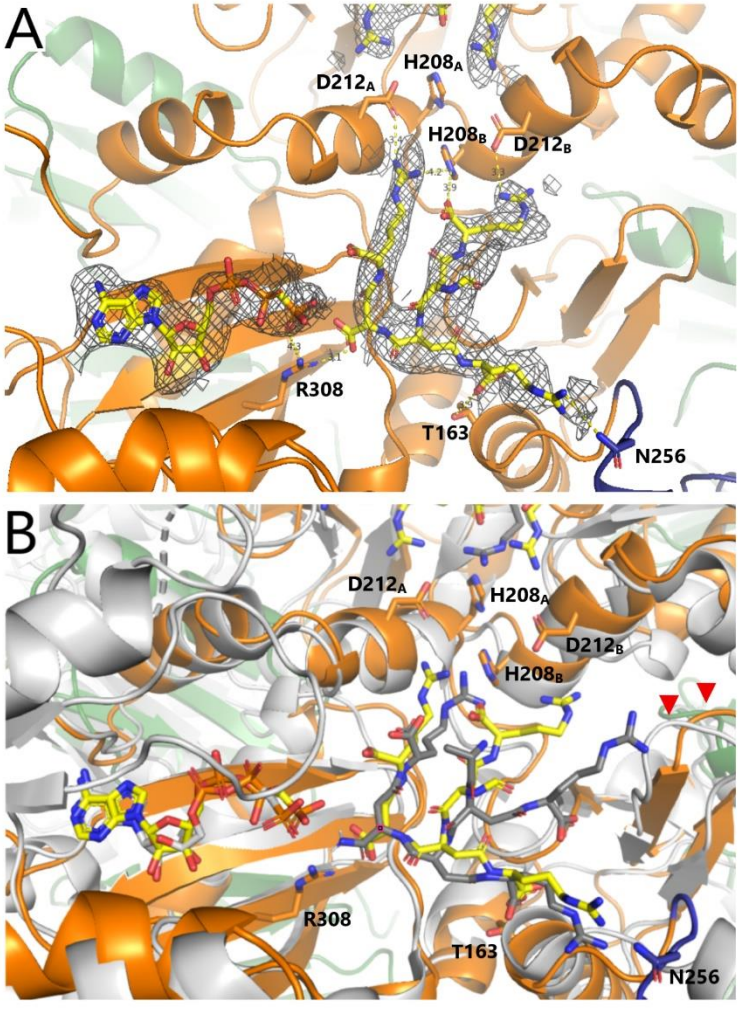

Figure 5: A: Structure of the *Stanieria* sp. CphA1 G domain complexed with cyanophycin (β-Asp-Arg)$_4$ and ADPCP at 2.60 Å. The cryo-EM map was carved 2 Å around the substrates at level 2.0. B: Structure of the *Stanieria* sp. CphA1 G domain complexed with cyanophycin (β-Asp-Arg)$_4$ and ADPCP overlaid with the structure of *Su*CphA1 with (Asp-Arg)$_8$-NH$_2$ and ADPCP (grey). Red arrows indicate loop at residues 181-185 in *Su*CphA1 and 183-190 in *Stanieria* sp. CphA2.

residues 181-185 in *Su*CphA1 and 183-190 in *Stanieria* sp. CphA2 (Fig. 5B, red arrows). In *Su*CphA1, this loop is orientated towards the cyanophycin binding pocket and the backbone of R187 can interact with the guanidinium group of the third dipeptide. In contrast, the loop is shifted away from the binding pocket in *Stanieria* sp. CphA2 and is not in proximity to interact with cyanophycin.

Binding of β-Asp-Arg

Unfortunately, no density could be observed for the β-Asp-Arg dipeptide near the large and adjacent loop, where it has been suggested to bind. In a second attempt to visualize the dipeptide, a cryo-EM dataset was collected with ATP instead of ADPCP and (N^2)-succinyl-L-arginine (succinyl-Arg) instead of β-Asp-Arg. The hypothesis behind this approach is that the binding of the β-Asp-Arg could be affected by the phosphorylation of the carboxyl group of cyanophycin. Succinyl-Arg was used instead of β-Asp-Arg to catch the enzyme in act, because this analog lacks the α-amino group required for nucleophilic attack. Additionally, KCl was added to the sample buffer because potassium is essential for enzymic activity and might play a role in binding of substrates. A map was obtained at 2.70 Å (Fig. S4C) and the previously obtained *Stanieria* sp. CphA2 model was refined into the map (Fig. S4D).

Upon inspection of the active site, it can be seen that ATP and the cyanophycin analog bind in the same position and orientation as in the dataset with β-Asp-Arg (Fig. S3E). The same residues are involved in interactions with cyanophycin. However, there was no density present for succinyl-Arg at the expected binding site near the large and adjacent loop.

Alternatively, insights into the potential binding site were gained by site-directed mutagenesis of residues in the large and adjacent loop. Four residues were targeted for mutagenesis to alanines, D362, R389, N393 and S395 (Fig. 6A). These residues are all conserved among cyanophycin synthetases (Fig. 6B), contain charged or hydrophilic sidechains, and are orientated towards the binding pocket (Fig. 6A). The mutants were expressed and purified and an activity assay was performed to determine the effect of the residues on the formation of cyanophycin. It was found that the D362A, N393A and S393A

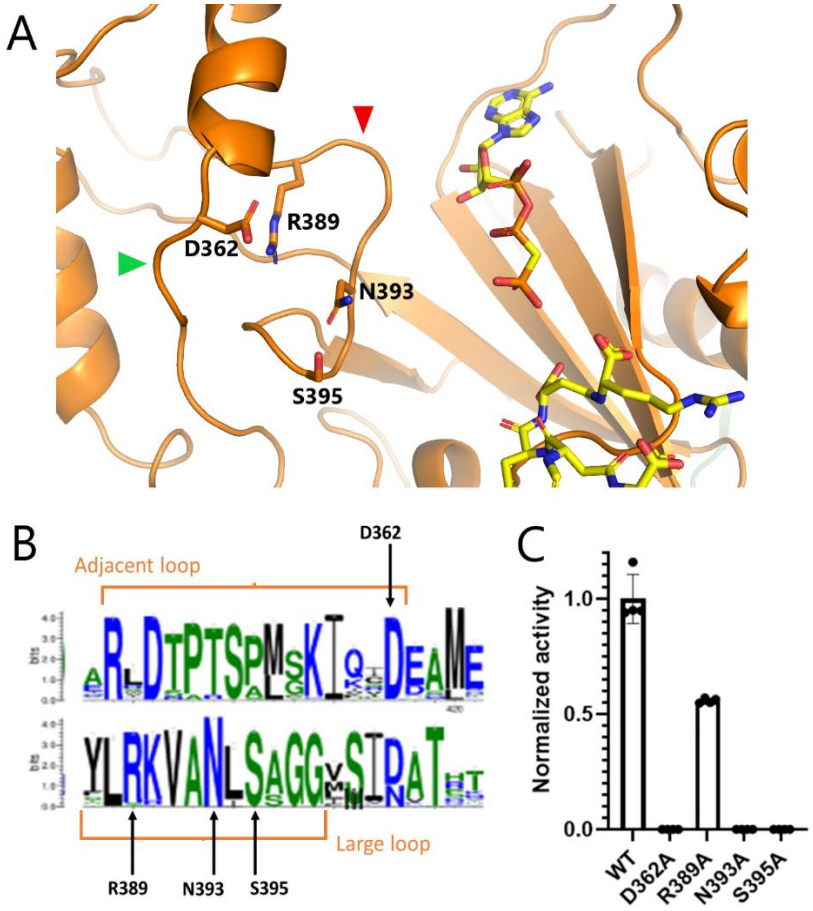

Figure 6: A: Residues in large loop (red arrow) and adjacent loop (green arrow) targeted for mutagenesis. B: Multiple sequence alignment of large and adjacent loop residues of CphA2 homologs by WebLogo.[30] C: Cyanophycin synthesis activity of WT *Stanieria* sp. CphA2 and mutants D362A, R389A, N393A, S395A. Activity was normalized to the activity of WT *Stanieria* sp. CphA2.

were completely inactive, while the activity of the R389A mutant was reduced by ~50% (Fig. 6C). These results show that the large and adjacent loop are important for cyanophycin synthesis activity and support the hypothesis that they are involved in the binding of β-Asp-Arg.

Variability of the G_{lid} lobe

3D variability analysis in CryoSPARC showed that both structures of *Stanieria* sp. CphA2 with substrates show extensive variability in the G_{lid} lobe (Movies 2 and 3; Fig. S3C and D), representing the open and closed conformations. Because both structures show density for both conformations, the decision was made to build the model of *Stanieria* sp. CphA2 with cyanophycin, β-Asp-Arg and ADPCP with G_{lid} in the open conformation, while the model of *Stanieria* sp. CphA2 with cyanophycin, succinyl-Arg and ATP was built in the closed conformation (Fig. S5). The angle between the open and closed conformations was calculated to be 68 degrees.

Discussion

The hexamer form of *Stanieria* sp. found in this research was not suggested based on SEC results in previous research, where the approximate molecular weight of the elution peak suggested a heptamer or octamer.[25] However, SEC is not a precise method for determination of the molecular weight, thus, the hexameric species is still roughly in agreement with the SEC results reported by Sharon *et al.*.[25]

From the activity assay with hexameric *Stanieria* sp. CphA2 and the dimeric mutant, it can be concluded that hexamer formation is important for cyanophycin synthesis activity of *Stanieria* sp. CphA2. The positive effect of hexamerization on the activity of the enzyme can be contributed to two aspects. Firstly, it can be hypothesized that the stability of the hexamer benefits enzyme activity. However, to prove this, one would have to determine the stability of the dimer in solution and compare it to the stability of the hexamer. For future research, it would be interesting to perform a thermal stability assay on the wildtype and the Q416A R528G double mutant to obtain more details on the relative stability between the hexamer and dimer. Secondly, the higher activity for the hexamer can be explained by the presence of six active sites in close proximity to each other, opposed to two in the dimer. Because cyanophycin is a substrate and a product in the reaction, it can benefit greatly from multiple active sites concentrated in one area. After dissociation from the active site in one CphA2 molecule, cyanophycin can bind in the active site of a nearby molecule. In order to test this hypothesis, the dependency of cyanophycin synthesis on the concentration of CphA2 could be studied.

By obtaining structures of *Stanieria* sp. CphA2 with substrates, we were able to visualize the binding of cyanophycin and ATP or ATP analog in the G domain active site. The structures show clear similarities to substrate binding in *Su*CphA1 in the interactions between cyanophycin and R308 (R309 in *Su*CphA1) and T163 (S166 in *Su*CphA1).[28] Differences between *Stanieria* sp. CphA2 and *Su*CphA1 are found in the interaction between cyanophycin and D212 (D197 in *Su*CphA1), H208 (I211 in *Su*CphA1) and the loop at residues 183-190 (181-185 in *Su*CphA1).[28] The interactions between cyanophycin and residues R308, T163 and D212 are in agreement with the observed loss of activity when the residues were mutated to alanines in *G. citriformis* CphA2.[25]

The mutagenesis experiments on residues in the large and adjacent loops in the G_{omega} lobe support the hypothesis that the loops are important for binding of β-Asp-Arg. This is consistent with previous studies that found the G_{omega} lobe to be important for substrate recognition and enzymatic activity in CphA1 and other ATP-grasp enzymes.[28,29]

In all thee datasets, variability in conformation of the G_{lid} lobe was found. Flexibility of the G_{lid} lobe was anticipated, as it is conserved in ATP grasp enzymes and suggested to play a role in their activity.[28,31]

Further analysis on the variability in particles of the datasets is necessary to deduce definite conclusions on a potential relationship with substrate binding.

Altogether, the structures and mutagenesis experiments performed in this study are in agreement with the process of cyanophycin synthesis by CphA2 as suggested in Sharon *et al.*, 2022.[25] Our findings confirm that an existing cyanophycin molecule binds at the G domain active site by interactions of its three C-terminal β-Asp-Arg dipeptides. The reaction method by which the C-terminal carboxylate is phosphorylated and extended is similar to other ATP-grasp enzymes.[32] The insights on the oligomeric behaviour of *Stanieria* sp. CphA2 and the binding of substrates in the active site together with previously obtained knowledge provide a broad understanding of cyanophycin synthesis. This knowledge can contribute to the improvement of large-scale bio-industrial synthesis of cyanophycin for commercial purposes.

Methods

Protein expression and purification

Plasmid containing the gene encoding CphA2 from *Stanieria* sp. and *G. citriformis* were constructed in a previous study.[25] For protein expression, BL21(DE3) *E. coli* cells were transformed with a plasmid containing the CphA2 gene and grown in TB media supplemented with 150 µg/ml kanamycin at 37 °C. Upon reaching an OD_{600} of ~1, culture temperature was lowered to 18 °C and protein expression was induced by addition of 0.2 mM isopropyl β-d-1-thiogalactopyranoside (IPTG) and grown for ~20 hours before harvesting. All purification steps were performed at 4 °C. Cells were harvested by centrifugation at 4000 rpm using a JLA8.1000 rotor for 10 minutes and resuspended for 30 minutes in buffer A (50 mM Tris pH 7.5, 250 mM NaCl, 3 mM BME, 10 mM imidazole) with 100 µM phenylmethylsulfonyl fluoride (PMSF) and a few crystals of DNAseI (Roche) and lysozyme (Bio Basic Inc.). Cells were lysed by sonication on ice water and the lysate was clarified by centrifugation at 40,000 g for 30 minutes. The clarified lysate was loaded onto a 5 ml HisTrap HP column (Cytiva) and washed with at least 75 ml buffer A with 30 mM imidazole pH 8. The protein was eluted with ~20 ml buffer B (50 mM Tris pH 7.5, 250 mM NaCl, 250 mM imidazole pH 8, 3 mM β-mercaptoethanol (BME)) and concentrated to <4 ml using a 100,000 kDa Amicon® Ultra spin column (Millipore). The protein was loaded on a Superdex 200 pg 16/60 column (Cytiva) equilibrated in buffer C (20 mM Tris pH 7.5, 100 mM NaCl, 1 mM dithiothreitol (DTT)). Fractions with the highest purity were pooled and concentrated. Aliquots of purified protein were stored in 20% glycerol at -80 °C. Samples were taken at various purification steps and visualized on an 11% SDS-PAGE gel (Fig. S6).

Site-directed mutagenesis

Site-directed mutagenesis of CphA2 *Stanieria* sp. was performed by the quickchange method. PCR reactions contained 10 ng template in Phusion® HF buffer (NEB), 0.2 mM of each DNTP, 0.5 mM each of forward and reverse primers (table M1) and 0.5 U Phusion® HF DNA polymerase (NEB) in a reaction volume of 25 µl. Following the reaction, template DNA was digested by the addition of 1 µl DpnI (NEB) and incubation at 37 °C for 1 hour. *E. coli* DH-5α cells were transformed with 2 µl of the reaction product and plated on LB agar plates containing 50 µg/ml kanamycin and incubated overnight at 37 °C. For plasmid preparation, single colonies were cultured in LB medium containing 50 µg/ml kanamycin overnight. Plasmid DNA was isolated using EZ-10 Spin Column Plasmid DNA Miniprep Kit (BioBasic) and sequenced by Sanger sequencing at Centre d'expertise et de services Génome Québec.

Size exclusion assays (SEC)

SEC was performed with a S200 increase 10/300 column (Cytiva) in buffer C at 4 °C. A flow rate of 0.5 ml/min and injection volume of 150 µl were used. The column was calibrated

with Amersham™ HMW Calibration Kit (GE Healthcare).

Activity assays

Reactions contained 2 µM purified protein, 2 mM ATP, 2 mM β-Asp-Arg, 50 µM synthetic cyanophycin (β-Asp-Arg)$_3$ as primer, 100 mM HEPES pH 8.2, 5 mM $MgCl_2$ and 20 mM KCl with a total reaction volume of 100 µl. Reactions were executed at 23 °C in quadruplicate. Formation of cyanophycin was measured by the increase in the OD_{600} as a result of light scattering by insoluble cyanophycin,[28] using a SpectraMax Paradigm spectrophotometer (Molecular Devices), with 5 second linear shaking between reads. Reactions proceeded for approximately 2 hours. Data were analyzed with GraphPad Prism. For the calculation of activity rates, the maximum of the first derivative of each OD_{600} curve was used and smoothed with a 2nd order polynomial to reduce noise.

Analytical ultracentrifugation (AUC)

Sedimentation velocity analysis by AUC was performed using a Beckman Coulter Optima analytical Ultracentrifuge. Protein was dialyzed extensively against a buffer containing 20 mM Tris pH 7.5, 100 mM NaCl, 1 mM Tris-(2-carboxyethyl)phosphine (TCEP) and diluted to a concentration of 0.8 mg/ml. The sample and reference buffer were housed in a 12mm 2-sector Epon-charcoal cell. AUC was performed at 10°C, using an An-60 Ti rotor spun at 18,000 rpm, and sedimentation behavior was observed by recording 175 absorbance scans (A_{280nm}) at intervals of 180 s. The oligomeric distribution was obtained by fitting the resultant data set to the Lamm equation model using the SEDFIT software package[33] and temperature-corrected buffer and protein parameters generated by SEDNTERP.[34]

Cryo-EM grid preparation

For testing of conditions on the Tecnai F20, the following conditions were used: (1) C-flat 2Cu T-50 grids and 0.3 mg/ml *Stanieria* sp. CphA2, (2) Quantifoil™ 200 Cu mesh grids and 0.3 mg/ml *Stanieria* sp. CphA2, (3) Quantifoil™ 200 Cu mesh grids and 1.8 mg/ml *Stanieria* sp. CphA2, (4) Quantifoil™ 200 Cu mesh grids and 2.2 mg/ml *Stanieria* sp. CphA2. All samples were in a 20 mM Tris, 100 mM NaCl pH 8, 1 mM DTT buffer (Fig. S7).

For the datasets collected on the Titan Krios, the following sample conditions were used: (1) no substrates sample: 2.2 mg/ml *Stanieria* sp. CphA2 in 20 mM Tris pH 8, 100 mM NaCl, 1 mM DTT, (2) cyanophycin, β-Asp-Arg, ADPCP sample: 2.4 mg/ml *Stanieria* sp. CphA2, 1 mM cyanophycin (β-Asp-Arg)$_4$, 20 mM β-Asp-Arg, 5 mM ADPCP in 20 mM Tris pH 8, 100 mM NaCl, 1 mM DTT, 10 mM $MgCl_2$, (3) cyanophycin succinyl-Arg, ATP sample: 2.4 mg/ml *Stanieria* sp. CphA2, 1 mM cyanophycin (β-Asp-Arg)$_4$, 20 mM succinyl-Arg, 5 mM ATP in 20 mM Tris pH 8, 100 mM NaCl, 1 mM DTT, 10 mM $MgCl_2$, 10 mM KCl with a 30 minute incubation period at 20 °C before vitrification. Protein was concentrated to desired concentration using 100,000 kDa Amicon® Ultra spin column (Millipore).

For vitrification, 3 µl sample was applied to glow-discharged grids, blotted for 1.5 seconds at force 10, and plunge-frozen in liquid ethane using a Vitrobot Mark IV (Thermo Fisher Scientific).

Cryo-EM grid quality assessment

Quality assessment of grid and sample was performed on a 200 keV Tecnai G2 F20 Transmission Electron Microscope (Thermo Fisher Scientific) equipped with TemCam XF416(ES) 16MP CMOS camera (TVIPS). Micrographs were collected using SerialEM software[35] at -3 µm defocus (Fig. S7).

Cryo-EM data acquisition

The sample was imaged on a 300 keV Titan Krios G3 Transmission Electron Microscope (Thermo Fisher Scientific) equipped with GIF BioQuantum LS imaging filter (Gatan) and K3 24MP direct electron detector (Gatan). For all datasets, SerialEM was used to collect micrographs of 40 frames at a dose of 2.0 $e^-/Å^2$, pixel size of 0.675 Å/pix and defocus range of -1.0 to -2.5 µm. For dataset (1), the dose rate

was 17.57 e^-/pix/s and 5663 micrographs were collected. For dataset (2), the dose rate was 16.02 e^-/pix/s and 7368 micrographs were collected. For dataset (3), the dose rate was 7.51 e^-/pix/s and 8639 micrographs were collected (Table M2).

Cryo-EM data processing

Number of micrographs and particles per dataset are specified in table M2. Data processing up to and including Bayesian polishing were performed in RELION4.0.[36,37] Beam induced drift was corrected by MotionCor2[38] and CTF estimation was performed by CTFfind4.[39] Particle picking was performed using Topaz.[40] For dataset (1), 769492 particles were picked, for dataset (2), 1902283 particles, and for dataset (3), 1204400. Particles were extracted from the micrographs and classified by 2D classification. Good 2D classes were selected (for dataset (1), 597148 particles, for dataset (2), 1416742 particles, for dataset (3) 707062 particles) and used for 3D classification. For the dataset without substrate, an initial model was generated in RELION to use as reference for 3D classification. For the datasets with substrates, the obtained model without substrates was low-pass filtered to 60 Å and used as reference. 3D classes consisting of hexameric particles were selected (for dataset (1), 251001 particles, for dataset (2), 336737 particles, for dataset (3) 549663 particles) and refined first without, and then with mask. Higher order aberration refinement, anisotropic magnification and per-particle defocus were performed. Two rounds of Bayesian polishing in RELION followed by 3D refinement with CTF refinement in CryoSPARC2[41] were performed. Local resolution and local filtering were performed in CryoSPARC2. 3D variability was assessed in CryoSPARC2.[42] Maps obtained by local filtering were subjected to local anisotropic sharpening in Phenix.[43]

Model building

An initial monomeric model was generated from protein sequence using Alphafold2[44] with the published structure of CphA2 from *G. citriformis* as reference.[25] The monomer was copied six times and roughly fitted in the density manually in Coot.[45] The model was subjected to real-space and rigid-body refinement in Phenix. Manual adjustments to the model were made in Coot on one symmetric unit and superimposed on the other symmetric unit using SSM superimpose.[46] The model was validated in Phenix and manual adjustments were made in Coot until satisfactory validation scores were obtained (Table M3).

Table M1: Site-directed mutagenesis primers

Mutation	Forward primer	Reverse primer
R315A	CTGTGTGTTAACGGTGCCTTCGTGGCG	CGCCACGAAGGCACCGTTAACACACAG
D362A	CTGGAGGAACAGGGTCTGGATCTG	GTACAGGTGCATAGCTTCAGCAGTACGG
R389A	CTGTCTAGCGGTGGCTTCAGCATC	GTTAGCAACTTTAGCCAGGTAGATGGTACGG
N393A	GTAAAGTTGCTGCCCTGTCTAGCGGTG	GCAGGTAGATGGTACGGTCACGATC
S395A	CTAACCTGGCTAGCGGTGGCTTCAG	CAACTTTACGCAGGTAGATGGTACGGTC
Q416A	GATAACATTATCCTGGCGGCAGACATCGCGC	GCGCGATGTCTGCCGCCAGGATAATGTTATC
R528G	GGTATCCTGATTAACGGTTCTGAGAAAATTCTG	CAGAATTTTCTCAGAACCGTTAATCAGGATACC
G611’	CCATTAAACGTAAAGGTGAGAATTTGTACTTCC	GGAAGTACAAATTCTCACCTTTACGTTTAATGG
Y616R	GCTGGAACAGCGTGAACTGGAAGC	GCTTCCAGTTCACGCTGTTCCAGC

Table M2: Data collection and processing

	Stanieria sp. CphA2	*Stanieria* sp. CphA2 + cyanophycin + β-Asp-Arg + ADPCP	*Stanieria* sp. CphA2 + cyanophycin + succinyl-Arg + ATP
Dose rate.(e-/pix/s)	17.57	16.02	7.51
Pixel size (A/pix)	0.675	0.675	0.675
Total dose (e-/A2/frame)	80	80	80
Defocus range (μm)	-1.0 to -2.5	-1.0 to -2.5	-1.0 to -2.5
Micrographs	5663	7368	8639
Particles extracted	769492	1902283	1204400
Particles selected from 2D classifications	597148	1416742	707062
Hexameric particles selected from 3D classification	251001	336737	549663

Table M3: Structure refinement scores

	Stanieria sp. CphA2	*Stanieria* sp. CphA2 + cyanophycin + β-Asp-Arg + ADPCP	*Stanieria* sp. CphA2 + cyanophycin + succinyl-Arg + ATP
Clash score	6.63	5.85	5.63
Ramachandran			
Favored (%)	91.84	93.84	93.52
Allowed (%)	7.79	6.47	6.32
Outliers (%)	0.37	0.05	0.16
Rotamer outliers (%)	1.16	0.85	2.19

References

[1] Borzi, A. (1886). *Le communicazioni intracellulari delle Nostochinee*. G. Capra e c. 0.

[2] Simon, R. D. (1976). The biosynthesis of multi-L-arginyl-poly (L-aspartic acid) in the filamentous cyanobacterium Anabaena cylindrica. *Biochimica et Biophysica Acta (BBA)-Enzymology*, *422*(2), 407-418.

[3] Simon, R. D. (1971). Cyanophycin granules from the blue-green alga Anabaena cylindrica: a reserve material consisting of copolymers of aspartic acid and arginine. *Proceedings of the national academy of sciences*, *68*(2), 265-267.

[4] Liotenberg, S., Campbell, D., Rippka, R., Houmard, J., & de Marsac, N. T. (1996). Effect of the nitrogen source on phycobiliprotein synthesis and cell reserves in a chromatically adapting filamentous cyanobacterium. *Microbiology*, *142*(3), 611-622.

[5] Mariotti, F., Tomé, D., & Mirand, P. P. (2008). Converting nitrogen into protein—beyond 6.25 and Jones' factors. *Critical reviews in food science and nutrition*, *48*(2), 177-184.

[6] Fay, P. (1992). Oxygen relations of nitrogen fixation in cyanobacteria. *Microbiological reviews*, *56*(2), 340-373.

[7] Li, H., Sherman, D. M., Bao, S., & Sherman, L. A. (2001). Pattern of cyanophycin accumulation in nitrogen-fixing and non-nitrogen-fixing cyanobacteria. *Archives of microbiology*, *176*(1), 9-18.

[8] Liang, B., Wu, T. D., Sun, H. J., Vali, H., Guerquin-Kern, J. L., Wang, C. H., & Bosak, T. (2014). Cyanophycin mediates the accumulation and storage of fixed carbon in non-heterocystous filamentous cyanobacteria from coniform mats. *PLoS One*, *9*(2), e88142.

[9] Wingard, L. L., Miller, S. R., Sellker, J. M., Stenn, E., Allen, M. M., & Wood, A. M. (2002). Cyanophycin production in a phycoerythrin-containing marine Synechococcus strain of unusual phylogenetic affinity. *Applied and Environmental Microbiology*, *68*(4), 1772-1777.

[10] Gross, R. A., & Kalra, B. (2002). Biodegradable polymers for the environment. *Science*, *297*(5582), 803-807.

[11] Lang, N. J., Simon, R. D., & Wolk, C. P. (1972). Correspondence of cyanophycin granules with structured granules in Anabaena cylindrica. *Archiv für Mikrobiologie*, *83*(4), 313-320

[12] Simon, R. D., & Weathers, P. (1976). Determination of the structure of the novel polypeptide containing aspartic acid and arginine which is found in cyanobacteria. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, *420*(1), 165-176.

[13] Conrad, U. (2005). Polymers from plants to develop biodegradable plastics. *Trends in plant science*, *10*(11), 511-512.

[14] Uddin, Z., Fang, T. Y., Siao, J. Y., & Tseng, W. C. (2020). Wound Healing Attributes of Polyelectrolyte Multilayers Prepared with Multi-L-arginyl-poly-L-aspartate Pairing with Hyaluronic Acid and γ-Polyglutamic Acid. *Macromolecular Bioscience*, *20*(8), 2000132.

[15] Sallam, A., & Steinbüchel, A. (2010). Dipeptides in nutrition and therapy: cyanophycin-derived dipeptides as natural alternatives and their biotechnological production. *Applied microbiology and biotechnology*, *87*(3), 815-828.

[16] Aboulmagd, E., Voss, I., Oppermann-Sanio, F. B., & Steinbüchel, A. (2001). Heterologous expression of cyanophycin synthetase and cyanophycin synthesis in the industrial relevant bacteria Corynebacterium glutamicum and Ralstonia eutropha and in Pseudomonas putida. *Biomacromolecules*, *2*(4), 1338-1342.

[17] Mooibroek, H., Oosterhuis, N., Giuseppin, M., Toonen, M., Franssen, H., Scott, E., Sanders, J. & Steinbüchel, A. (2007). Assessment of technological options and economical feasibility for

cyanophycin biopolymer and high-value amino acid production. *Applied Microbiology and Biotechnology*, *77*(2), 257-267.

[18] Nausch, H., Hausmann, T., Ponndorf, D., Hühns, M., Hoedtke, S., Wolf, P., Zeyner, A. & Broer, I. (2016). Tobacco as platform for a commercial production of cyanophycin. *New biotechnology*, *33*(6), 842-851.

[19] Steinle, A., Witthoff, S., Krause, J. P., & Steinbüchel, A. (2010). Establishment of cyanophycin biosynthesis in Pichia pastoris and optimization by use of engineered cyanophycin synthetases. *Applied and environmental microbiology*, *76*(4), 1062-1070.

[20] Ziegler, K., Diener, A., Herpin, C., Richter, R., Deutzmann, R., & Lockau, W. (1998). Molecular characterization of cyanophycin synthetase, the enzyme catalyzing the biosynthesis of the cyanobacterial reserve material multi-L-arginyl-poly-L-aspartate (cyanophycin). *European journal of biochemistry*, *254*(1), 154-159.

[21] Klemke, F., Nuernberg, D. J., Ziegler, K., Beyer, G., Kahmann, U., Lockau, W., & Volkmer, T. (2016). CphA2 is a novel type of cyanophycin synthetase in N2-fixing cyanobacteria. *Microbiology*, *162*(3), 526-536.

[22] Berg, H., Ziegler, K., Piotukh, K., Baier, K., Lockau, W., & Volkmer-Engert, R. (2000). Biosynthesis of the cyanobacterial reserve polymer multi-L-arginyl-poly-L-aspartic acid (cyanophycin) Mechanism of the cyanophycin synthetase reaction studied with synthetic primers. *European Journal of Biochemistry*, *267*(17), 5561-5570.

[23] Picossi, S., Valladares, A., Flores, E., & Herrero, A. (2004). Nitrogen-regulated genes for the metabolism of cyanophycin, a bacterial nitrogen reserve polymer: expression and mutational analysis of two cyanophycin synthetase and cyanophycinase gene clusters in the heterocyst-forming cyanobacterium Anabaena sp. PCC 7120. *Journal of Biological Chemistry*, *279*(12), 11582-11592.

[24] Füser, G., & Steinbüchel, A. (2007). Analysis of genome sequences for genes of cyanophycin metabolism: identifying putative cyanophycin metabolizing prokaryotes. *Macromolecular bioscience*, *7*(3), 278-296.

[25] Sharon, I., Grogg, M., Hilvert, D., & Schmeing, T. M. (2022). Structure and Function of the β-Asp-Arg Polymerase Cyanophycin Synthetase 2. *ACS Chemical Biology*, *17*(3), 670-679.

[26] Hara, T., Kato, H., Katsube, Y., & Oda, J. I. (1996). A pseudo-Michaelis quaternary complex in the reverse reaction of a ligase: structure of Escherichia coli B glutathione synthetase complexed with ADP, glutathione, and sulfate at 2.0 Å resolution. *Biochemistry*, *35*(37), 11967-11974.

[27] van Heijenoort, J. (2001). Recent advances in the formation of the bacterial peptidoglycan monomer unit. *Natural product reports*, *18*(5), 503-519.

[28] Sharon, I., Haque, A. S., Grogg, M., Lahiri, I., Seebach, D., Leschziner, A. E., Hilvert, D. & Schmeing, T. M. (2021). Structures and function of the amino acid polymerase cyanophycin synthetase. *Nature Chemical Biology*, *17*(10), 1101-1110.

[29] Galant, A., Arkus, K. A., Zubieta, C., Cahoon, R. E., & Jez, J. M. (2009). Structural basis for evolution of product diversity in soybean glutathione biosynthesis. *The Plant Cell*, *21*(11), 3450-3458.

[30] Crooks, G. E., Hon, G., Chandonia, J. M., & Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome research*, *14*(6), 1188-1190.

[31] Li, H., Fast, W., & Benkovic, S. J. (2009). Structural and functional modularity of proteins in the de novo purine biosynthetic pathway. *Protein Science*, *18*(5), 881-892.

[32] Fawaz, M. V., Topper, M. E., & Firestine, S. M. (2011). The ATP-grasp enzymes. *Bioorganic chemistry*, *39*(5-6), 185-191.

[33] Schuck, P., Perugini, M. A., Gonzales, N. R., Howlett, G. J., & Schubert, D. (2002). Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems. *Biophysical journal, 82*(2), 1096-1111.

[34] Laue, T. M., Shah, B. D., Ridgeway, T. M., Pelletier, S. L., Harding, S. E., Rowe, A. J., & Horton, J. C. (1992). Analytical Ultracentrifugation in Biochemistry and Polymer Science. *Royal Society of Chemistry*.

[35] Mastronarde, D. N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. *Journal of structural biology, 152*(1), 36-51.

[36] Kimanius, D., Dong, L., Sharov, G., Nakane, T., & Scheres, S. H. (2021). New tools for automated cryo-EM single-particle analysis in RELION-4.0. *Biochemical Journal, 478*(24), 4169-4185.

[37] Scheres, S. H. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of structural biology, 180*(3), 519-530.

[38] Zheng, S. Q., Palovcak, E., Armache, J. P., Verba, K. A., Cheng, Y., & Agard, D. A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature methods, 14*(4), 331-332.

[39] Rohou, A., & Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of structural biology, 192*(2), 216-221.

[40] Bepler, T., Morin, A., Rapp, M., Brasch, J., Shapiro, L., Noble, A. J., Berger, B. (2019). Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature Methods 16*, 1153–1160.

[41] Punjani, A., Rubinstein, J. L., Fleet, D. J., & Brubaker, M. A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature methods, 14*(3), 290-296.

[42] Punjani, A., & Fleet, D. J. (2021). 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *Journal of structural biology, 213*(2), 107702.

[43] Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallographica Section D: Structural Biology*, 75(10), 861-877.

[44] Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nature Methods*, 1-4.

[45] Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D60*, 2126-2132.

[46] Krissinel, E. & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions *Acta Crystallogr. D60*, 2256-2268.