Received: 15 October 2019 | Accepted: 10 April 2020
DOI: 10.1002/prot.25896

RESEARCH ARTICLE

PROTEINS WILEY

Crystal structure of a novel fold protein Gp72 from the freshwater cyanophage Mic1

Ying Wang | Hua Jin | Feng Yang | Yong-Liang Jiang | Yan-Yan Zhao | Zhi-Peng Chen | Wei-Fang Li | Yuxing Chen | Cong-Zhao Zhou | Qiong Li

Hefei National Laboratory for Physical Sciences at the Microscale and School of Life Sciences, University of Science and Technology of China, Hefei, Anhui, China

Correspondence
Qiong Li and Cong-Zhao Zhou, Hefei National Laboratory for Physical Sciences at the Microscale and School of Life Sciences, University of Science and Technology of China, Hefei, Anhui 230026, China.
Email: liqiong@ustc.edu.cn (Q. L) and zcz@ustc.edu.cn (C.-Z. Z.)

Funding information
Ministry of Science and Technology of China, Grant/Award Numbers: 2016YFA0400900, 2018YFA0903100; National Natural Science Foundation of China, Grant/Award Numbers: 31621002, 31630001; The Chinese Academy of Sciences (the Innovative Academy for Seed Design)

Peer Review
The peer review history for this article is available at https://publons.com/publon/10.1002/prot.25896.

Abstract

Cyanophages, widespread in aquatic systems, are a class of viruses that specifically infect cyanobacteria. Though they play important roles in modulating the homeostasis of cyanobacterial populations, little is known about the freshwater cyanophages, especially those hypothetical proteins of unknown function. Mic1 is a freshwater siphocyanophage isolated from the Lake Chaohu. It encodes three hypothetical proteins Gp65, Gp66, and Gp72, which share an identity of 61.6% to 83%. However, we find these three homologous proteins differ from each other in oligomeric state. Moreover, we solve the crystal structure of Gp72 at 2.3 Å, which represents a novel fold in the $\alpha + \beta$ class. Structural analyses combined with redox assays enable us to propose a model of disulfide bond mediated oligomerization for Gp72. Altogether, these findings provide structural and biochemical basis for further investigations on the freshwater cyanophage Mic1.

KEYWORDS
crystal structure, freshwater cyanophage, hypothetical protein, oligomerization, redox

1 | INTRODUCTION

Viruses are the most abundant biological entities on the planet, most of which usually infect bacteria and are known as bacteriophages.[1] As a special class, bacteriophages that specifically infect cyanobacteria are termed cyanophages, which are widespread in aquatic systems.[2] Cyanobacteria and cyanophages adapt to and coevolve with each other during infection and anti-infection processes,[3] forming a mutualistic symbiosis relationship. In fact, cyanobacteria provide living place for cyanophages to replicate genomic DNA and release progeny phages, whereas cyanophages act as genetic information stores and carriers of cyanobacteria and regulate the abundancy, distribution, and diversity of cyanobacterial populations.[4,5] As we known, under the conditions of eutrophication and high temperature, cyanobacteria could proliferate quickly, resulting in the seasonal outbreak of cyanobacterial bloom in a wide range of fresh waterbodies all over the world.[6] The bloom not only causes the deterioration of water quality and imbalance of ecosystem, but also threats the health of surrounding humans and animals, thus becomes a serious social and economic issue.[7] Cyanophages play essential roles in modulating the homeostasis of cyanobacterial populations,[8] providing a putative environment-friendly solution for temporarily controlling the cyanobacterial bloom.[9] However, systematic investigations are needed to better understand the freshwater cyanophages.

Studies on cyanophages started from the 1960s.[10] Nearly all the isolated cyanophages contain a dsDNA genome,[11] and belong to the order *Caudovirales* with three families *Myoviridae*, *Podoviridae*, and *Siphoviridae* according to the morphology.[12] Thanks to the development of next generation sequencing, genome of about 108 cyanophages have been sequenced (http://www.genome.jp/virushostdb/view/), including only 14 freshwater cyanophages. The growing genomic sequence data indicate that the cyanophages

represent a huge pool of diverse novel genes, as more than half of the hypothetical open reading frames (ORFs) could not be annotated based on the previously identified proteins.[13,14] These genes occupy a large number of genome and are thought to provide a selective benefit to phages,[15,16] thus more functional analyses are necessary.

Mic1 is a freshwater siphocyanophage isolated from the Lake Chaohu that specifically infects the bloom-forming cyanobacteria *Microcystis aeruginosa*.[17] Its 92.6 kb genome harbors 98 putative ORFs (GenBank: MN013189), including 62 genes encoding hypothetical proteins of unknown function.[33] Notably, *gp65*, *gp66*, and *gp72* are three duplicated genes, which encode proteins possessing sequence identities of 61.6% to 83% with each other. We found that the three homologous proteins Gp65, Gp66, and Gp72 exist as different oligomeric states in solution. However, bioinformatic analyses showed that neither homologous protein nor similar motif could be found in the present databases. Therefore, we solved the crystal structure of Gp72 at 2.3 Å resolution, and revealed that it possesses a central barrel and displays in a novel α + β fold. Structural analyses in combination with redox assays indicated that Cys28 and Cys115 might regulate the oligomerization of Gp72. These findings provide structural and biochemical hints for further exploring the function of hypothetical proteins in the freshwater cyanophage Mic1.

2 | MATERIALS AND METHODS

2.1 | Oligomeric states analyses

Analytical gel filtration chromatography (AGFC) was used to determine the molecular weight of Gp72 in solution by a Superdex 200 Increase 10/300 GL column (GE Healthcare). The following standard molecular markers were used for calibration: ribonuclease A (13.7 kDa), ovalbumin (43.0 kDa), conalbumin (75.0 kDa), aldolase (158.0 kDa), ferritin (440.0 kDa), and thyroglobulin (669.0 kDa).

Size exclusion chromatography with multiangle light scattering (SEC-MALS) was used to determine the molecular weight of Gp66 in solution. The assay was performed using a Superdex 200 Increase 10/300 GL column connected to the DAWN HELEOS II light scattering detector (Wyatt Technology) and the Optilab T-rEx refractive index detector (Wyatt Technology). The protein sample (100 μL, 1.0 mg/mL) was injected into and then eluted from the column pre-equilibrated with the buffer of 40 mM Tris-HCl, pH 7.5, 100 mM NaCl, and 14 mM β-mercaptoethanol (β-ME). The results were recorded and processed by ASTRA 7.0.1 software (Wyatt Technology). The final figure was prepared using the OriginPro 2019 software.

2.2 | Cloning, expression, and purification

The coding sequences of Gp65, Gp66, and Gp72 were amplified from the genomic DNA of Mic1, and then cloned into a modified pET29b vector, respectively, with a C-terminal His-tag. The *Escherichia coli* BL21 (DE3) stain (Novagen) was used for the overexpression of the recombinant proteins. The transformed cells were grown in LB culture medium (10 g of Tryptone, 5 g of yeast extract, and 10 g of NaCl per liter) containing 30 μg/mL kanamycin at 37°C to an $A_{600\ nm}$ of 0.8, and then induced with 0.2 mM isopropyl β-D-1-thiogalactopyranoside at 16°C for 20 hours. The cells were harvested and resuspended in 30 mL lysis buffer of 40 mM Tris-HCl, pH 7.5, 100 mM NaCl, and 14 mM β-ME. After sonication and centrifugation, the supernatant was loaded onto a Ni-NTA column (GE healthcare) equilibrated with the binding buffer, which is the same as the lysis buffer. The target protein was eluted with 500 mM imidazole, and further applied to a Superdex 75 column (GE Healthcare) equilibrated with the binding buffer. The fractions containing the target protein were collected and stored at 4°C. The purity of protein was assessed by gel electrophoresis.

The selenomethionine (SeMet)-substituted Gp72 was overexpressed in *E. coli* B834 (DE3) strain. The transformed cells were cultured overnight in LB medium at 37°C, and then washed and resuspended with the SeMet medium (M9 medium supplied with 60 mg/L SeMet and other essential amino acids).[18] The following steps in protein expression and purification were the same as those for the native protein.

The version of Gp72 with an N-terminal His-tag, the site-directed mutagenesis of Gp72 and the truncation of Gp66 were performed using a standard PCR-based strategy with the plasmid encoding the wild-type Gp72 and Gp66 as the template, respectively. They were expressed and purified in the same way as the wild-type protein.

2.3 | Crystallization, data collection, and processing

The Gp72 protein was concentrated to 13 mg/mL for crystallization. Crystals were grown at 289 K using the sitting drop vapor diffusion method, with a drop of 1 μL protein solution mixed with an equal volume of the reservoir solution. However, after exhaustive optimization, the crystal diffraction was too low to determine the structure. Microseed matrix screening method was used for further crystal optimization.[19] Initial crystals are systematically transferred to all crystallographic kit and used as seeds to screen for other conditions that support ordered growth of crystals. The reservoir solution of SeMet-substituted Gp72 with good diffraction quality is 13% polyethylene glycol 6000, 0.1 M ADA pH 7.0, 0.01 M spermine tetrahydrochloride, and 0.1 M guanidine hydrochloride. Crystals were flash cooled with liquid nitrogen after addition of 10% glycerol to the reservoir solution as cryoprotectant. The X-ray diffraction data were collected at 100 K in a liquid nitrogen stream on beamline BL19U at the Shanghai Synchrotron Radiation Facility. The diffraction data were indexed, integrated, and scaled with the program *HKL2000*.[20]

2.4 | Structure determination and refinement

The crystal structure of Gp72 was determined by the single-wavelength anomalous dispersion method.[21] The *AutoSol* program in

PHENIX[22] was used to search the selenium atoms and to calculate the phase. Automatic model building was carried out by *AutoBuild* in *PHENIX*. Then, the initial model was refined by *REFMAC5* of *CCP4i* program suite, and rebuilt interactively with the program *Coot*.[23,24] The final model was evaluated with the web service *MolProbity*.[25] The crystallographic parameters were listed in Table 1. All structure figures were prepared with *PyMOL* (http://www.pymol.org).

2.5 | Redox assays

Gp72 and its mutants were purified using the buffer of 40 mM Tris-HCl, pH 7.5, 100 mM NaCl, respectively. Then, they were incubated with 0.5 mM $CuCl_2$ on ice for 30 minutes to introduce disulfide bonds. Each sample was divided into two parts with or without 100 mM β-ME, and applied to sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) or native polyacrylamide gel electrophoresis (native PAGE) to detect the formation of disulfide bonds.

TABLE 1 Crystal parameters, data collection, and structure refinement

	SeMet-Gp72
Data collection	
Wavelength (Å)	0.97918
Space group	*C2*
Unit cell parameters	
a, *b*, *c* (Å)	85.702, 45.534, 67.368
α, β, γ (°)	90.00, 94.29, 90.00
Resolution range (Å)	50.00-2.30 (2.38-2.30)[a]
Unique reflections	11 607 (1125)
Completeness (%)	98.8 (98.2)
$<I/\sigma(I)>$	25.23 (4.90)
R_{merge}[b] (%)	8.3 (30.1)
Average redundancy	5.9 (5.3)
Structure refinement	
Resolution range (Å)	42.77-2.29
R_{factor}[c]/R_{free}[d] (%)	20.1/23.8
Number of protein atoms	1,976
Number of water molecules	57
RMSD[e] bond lengths (Å)	0.001
RMSD bond angles (°)	1.239
Mean B factors (Å^2)	53.24
Ramachandran plot[f]	
Most favored (%)	96.1
Additional allowed (%)	3.9
Protein Data Bank entry	6L2W

Abbreviation: RMSD, root mean square deviation.

[a]The values in parentheses refer to statistics in the highest bin.

[b]$R_{merge} = \sum_{hkl}\sum_i|I_i(hkl) - <I(hkl)>|/\sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of an observation, and $<I(hkl)>$ is the mean value for its unique reflection. Summations are over all reflections.

[c]$R_{factor} = \sum_h|Fo(h) - Fc(h)|/\sum_h Fo(h)$, where Fo and Fc are the observed and calculated structure-factor amplitudes, respectively.

[d]R_{free} was calculated with 5% of the data excluded from the refinement.

[e]RMSD from ideal values.

[f]Categories were defined by MolProbity.

3 | RESULTS

3.1 | Homologous protein Gp65, Gp66, and Gp72 possess different oligomeric states

The genome of freshwater siphocyanophage Mic1 harbors three duplicated genes *gp65*, *gp66*, and *gp72*, which encode three small proteins of similar molecular weight and primary sequence identity at 61.6% to 83% (Figure 1A). Notably, Gp65, Gp66, and Gp72 have been all detected in the mass spectrometric analysis, indicating that they might be structural components of the mature Mic1. Though *gp65*, *gp66*, and *gp72* locate closely to genes involved in DNA replication and nucleotide metabolism in the genome, their function cannot be annotated based on searching homologous proteins or similar motifs in the present databases. Interestingly, the recombinant Gp65, Gp66, and Gp72 exhibit different elution volumes in the gel filtration column (Figure 1B). Further analysis with AGFC and SEC-MALS showed that the apparent molecular weights of Gp72 and Gp66 are about 25.7 and 63 kDa, whereas the theoretical molecular weights of the monomer are 14.6 and 15.5 kDa, suggesting that Gp72 and Gp66 exist as a monodispersing dimer and tetramer in solution, respectively (Figure 1C,D). Moreover, secondary structure prediction combined with multi-sequence alignment revealed that Gp66 has a longer helix α2 compared to that of Gp72 (Figure 1A). Gel filtration assays showed that deletion of the 4 or 8 C-terminal residues truncations, termed $Gp66_{C\Delta4}$ and $Gp66_{C\Delta8}$, respectively, could significantly alter the oligomeric state (Figure 1E), and deletion the last 12 C-terminal residues ($Gp66_{C\Delta12}$) resulted in the formation of a dimer (Figure 1E). It indicated that the helix α2 is involved in regulating the tetramerization of Gp66. These results suggested that Gp72 and Gp66 might possess different functions in the assembly or infection of Mic1.

3.2 | Overall structure of Gp72

To explore the functions based on their structures, Gp65, Gp66, and Gp72 were purified and applied to crystal screening, but only high-quality crystals of Gp72 were obtained after exhaustive screening and optimization. Eventually, we solved the structure of Gp72 at 2.3 Å resolution in the space group *C2*. The Gp72 structure is refined to final *R* and R_{free} values of 20.1% and 23.8%, respectively. The crystallographic parameters are listed in Table 1.

In the structure, each asymmetric unit contains two molecules of Gp72 (Figure 2A), which form a homodimer with a buried interface area of 1000 Å^2, as calculated by QTPISA.[26] The interface is mainly stabilized by hydrophobic interactions (Figure 2B), in addition to

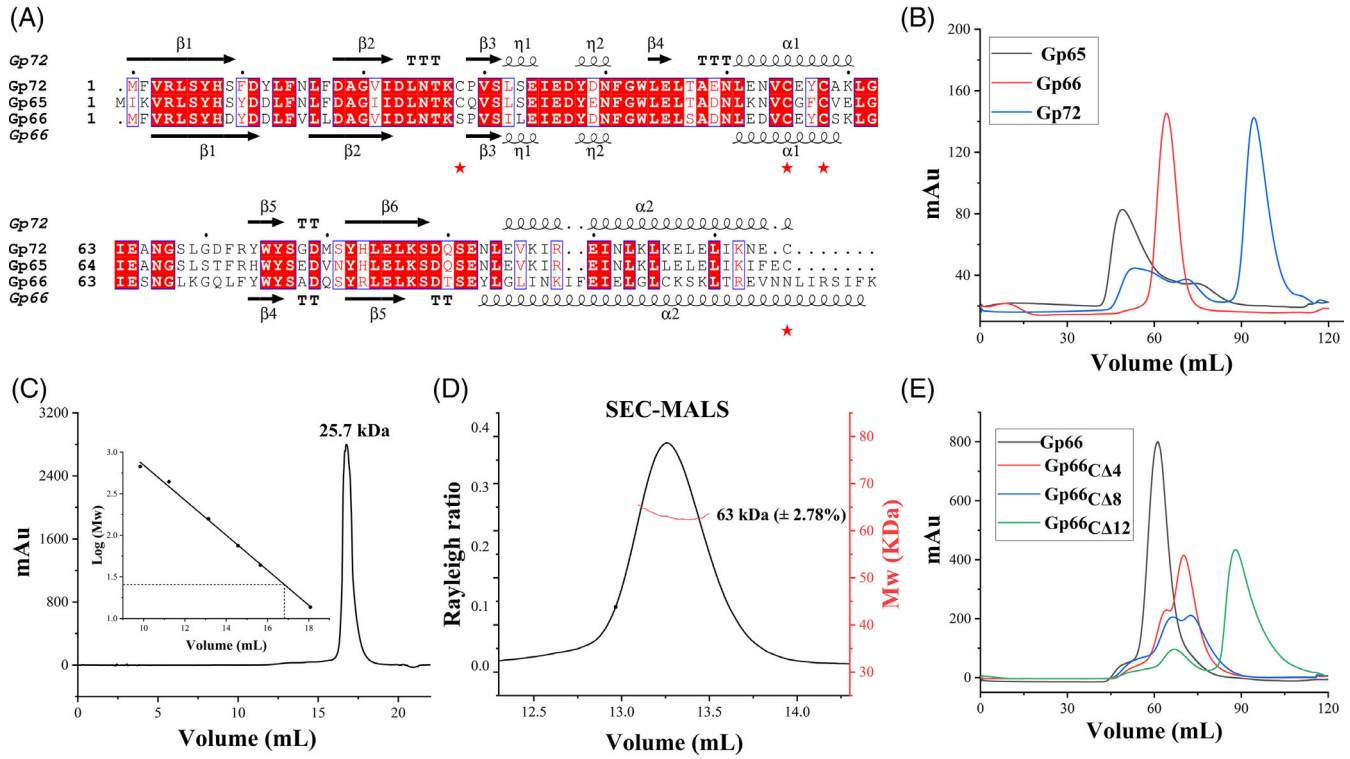

FIGURE 1 Different oligomeric states of Gp65, Gp66, and Gp72. A, Multiple-sequence alignment of Gp65, Gp66, and Gp72. The alignment was performed with the programs Clustal Omega and Espript. The secondary structural elements of Gp72 and Gp66 are shown above and below the sequences, respectively. The structure of Gp66 is gained by homology modeling using Robetta[31] based on Gp72. Four cysteine residues are marked with red asterisks. B, Gel filtration profiles of Gp65, Gp66, and Gp72. mAU, milliabsorbance units. C, Analysis of analytical gel filtration chromatography (AGFC) showed that Gp72 exists as a dimer in solution. (D) Analysis of size exclusion chromatography with multiangle light scattering (SEC-MALS) assays showed that Gp66 exists as a tetramer in solution. The eluted peak is in correspondence with the X axis and the Rayleigh ratio of the Y axis on the left. The jagged short line represents the molecular weight of the Y axis on the right. E, Gel filtration profiles of full-length and truncated Gp66

hydrophilic interactions including hydrogen bonds and salt bridges. The dimerization of Gp72 in the crystal structure is consistent with the result of AGFC (Figure 1C), proving that Gp72 also exists as a stable dimer in solution. Further analysis showed that the two subunits of Gp72 pack against each other in a symmetric manner, forming a central barrel, which is closed at the bottom and open at the top (Figure 2A). The bottom is blocked by the residues Tyr76, Met80, and His8, whereas the top is surrounded by the residues Asp39 and Arg73 (Figure 2C).

Each subunit consists of two α-helices, two η-helices and six antiparallel β-strands, which are arranged with a topological architecture of β1-β2-β3-η1-η2-β4-α1-β5-β6-α2 (Figure 2D). The six β-strands consist of half of the central barrel, whereas the two α-helices pack against the barrel on each side (Figure 2A). The two subunits in the asymmetric unit share an overall structure similar to each other, with a root mean square deviation (RMSD) of 0.5677 Å, except that a loop between strands β1and β2 (Phe14-Leu16) of subunit A is missing in the electron density map. A homology search against the DALI database[27] yielded an output of several functionally unrelated proteins with a Z-score of ≤5. Structural superposition gave an RMSD of 3.4 Å over 82 Cα atoms and a sequence identity of 5% against the top hit, the so-called PHCCEx domain of guanine nucleotide-exchange factors Tiam2 from *Mus musculus*.[28] Notably, Gp72 resembles a computationally designed novel-fold protein termed Top7,[29] with an RMSD of 3.5 Å over 69 Cα atoms. Moreover, we did not find a protein that harbors a similar fold with Gp72 in the SCOP database (http://scop.mrc-lmb.cam.ac.uk/scop/). Altogether, Gp72 represents a novel fold in the α + β class.

3.3 | Cys28 and Cys115 regulate the oligomerization of Gp72

As shown in Figure 3A, when C-terminal His-tagged Gp72 was applied to SDS-PAGE, we found that it displays as a mix of monomer, dimer and oligomer in the presence of oxidant $CuCl_2$, whereas it becomes a homogeneous monomer in the presence of reductant β-ME. In contrast, the N-terminal His-tagged Gp72 displays as a much regular ladder of oligomers in the presence of $CuCl_2$ (Figure 3A), suggesting that Gp72 exists as different oligomeric states in solution upon oxidation. Considering that the hydrophobic dimerized interface seen in the structure could be destroyed by SDS; thus, it indicated

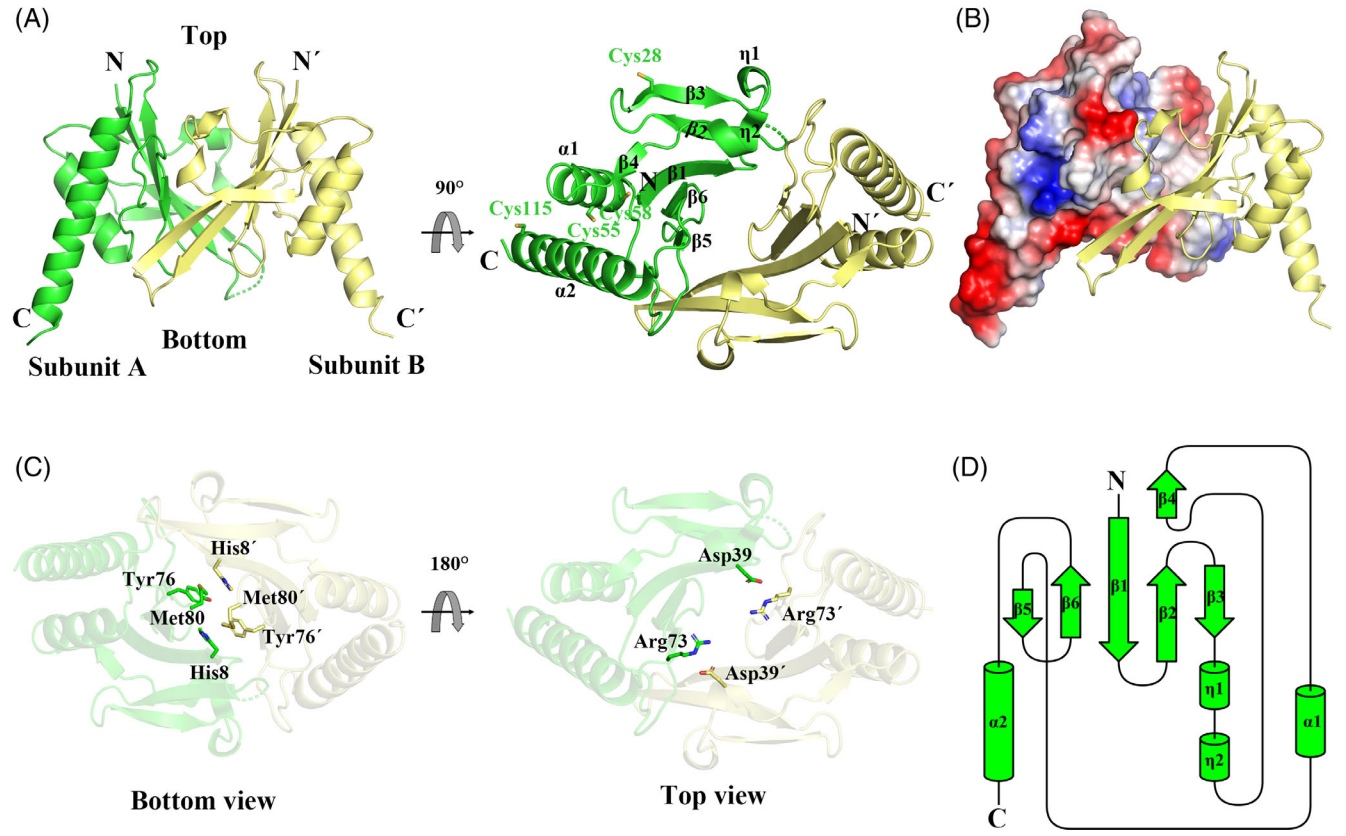

FIGURE 2 Overall structure of Gp72. A, Cartoon representation of dimeric Gp72 in the side view (left) and top view (right). Two subunits are colored in green and yellow, respectively. The secondary elements are labeled in subunit A. The disordered regions are shown as dotted lines. The four cysteine residues at subunit A are shown as sticks and labeled. B, The hydrophobic interface of dimeric Gp72. Subunits A and B are shown as electrostatic potential and cartoon, respectively. C, The central barrel of Gp72 in the bottom view (left) and top view (right). The involved residues are labeled and shown as sticks. D, Topology diagram of Gp72, which is drew by Topdraw.[32]

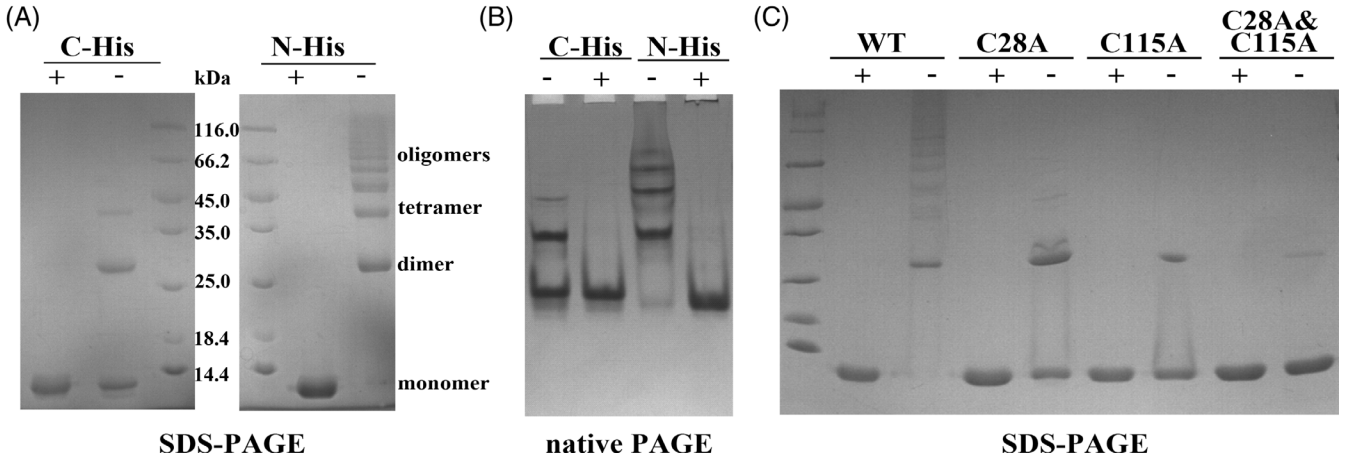

FIGURE 3 Cys28 and Cys115 are involved in regulating the oligomerization of Gp72. A, Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) profiles of Gp72 with the C-terminal or N-terminal His-tag. Gp72 in the presence of β-mercaptoethanol and $CuCl_2$ are shown as "+" and "–," respectively. B, Native PAGE profiles of Gp72 with the C-terminal or N-terminal His-tag. C, SDS-PAGE profiles of wild-type and mutated Gp72

that the oligomerization of Gp72 we observed might be correlated with redox, and the C-terminal His-tag could partly disturb the process. In addition, the results of SDS-PAGE were further confirmed by native PAGE (Figure 3B). Notably, though it presents as a dimer in the presence of β-ME due to lack of SDS, Gp72 forms oligomers of higher molecular weight (tetramer and so on) in the oxidized condition.

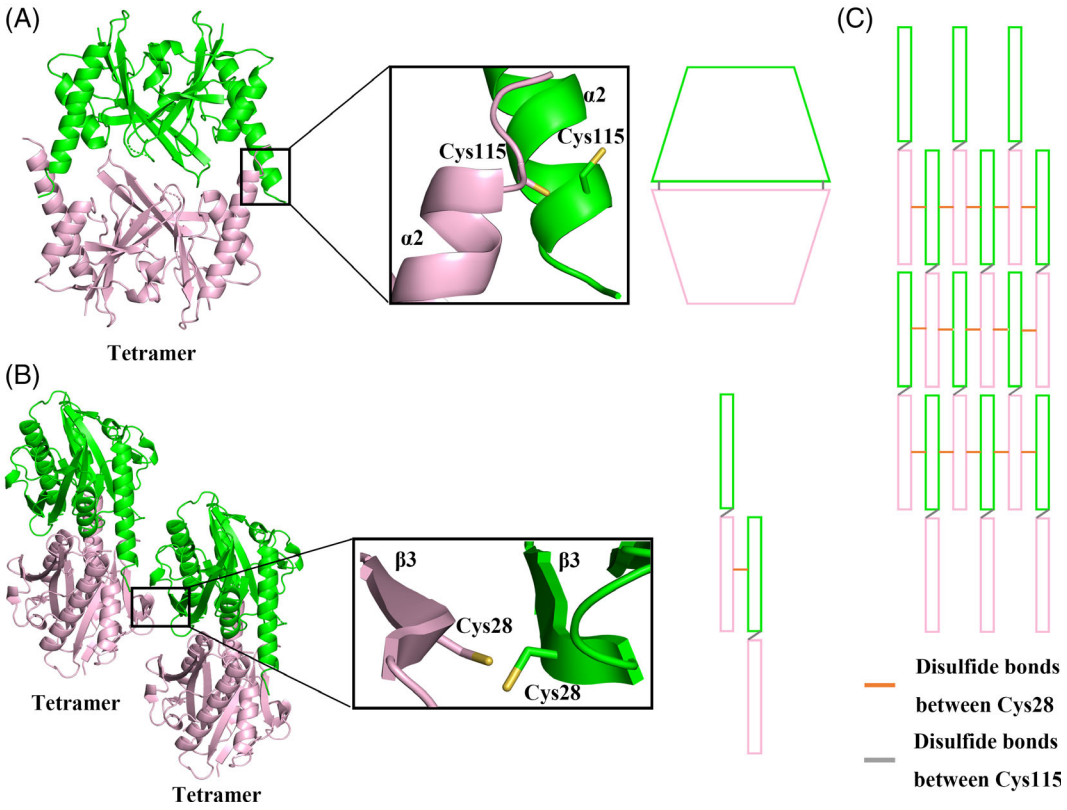

FIGURE 4 A model for the oligomerization of Gp72. A, Two Gp72 dimers form a tetramer via the disulfide bonds between two pairs of Cys115 residues. A schematic representation is shown in the right. B, Two Gp72 tetramers form an octamer via the disulfide bonds of Cys28 residues. C, Gp72 octamers pack against each other to form super oligomers. The intermolecular disulfide bonds between pairs of Cys115 and Cys28 residues are shown as gray line and orange line, respectively

Structural analysis showed that two Gp72 homodimers could pack against each other at the bottom, and form a tetramer by symmetry operation (Figure 4A). Gp72 has four cysteine residues (Figure 1A), which probably form redox-regulated disulfide bonds. As shown in the crystal structure, Cys55 and Cys58 are buried at the interface between the helix α1 and the central barrel, whereas Cys28 at the strand β3 and Cys115 at the helix α2 are exposed to the solvent (Figure 2A). It indicated that Cys28 and Cys115 are involved in the formation of disulfide bonds that regulate the oligomerization of Gp72. Accordingly, we mutated these two cysteine residues to alanine. As shown in Figure 3C, compared with the wild-type N-terminal-His-tagged Gp72, single mutation of C28A or C115A led to the loss of oligomerized ladders except for the dimer at the oxidized condition; and moreover, double mutation of C28A and C115A resulted in almost complete disappearance of the oxidized dimer. It confirmed that both Cys28 and Cys115 play an essential role in regulating the oligomerization of Gp72.

4 | DISCUSSION

Recently, we isolated a siphocyanophage Mic1 infecting *Microcystis aeruginosa* from the Lake Chaohu in China, which suffers from seasonal algal blooms every year. The capsid structure of Mic1 has been solved using cryoelectron microscopy, providing hints for the assembly of freshwater cyanophages.[17] However, although its genome was sequenced, there are still many unknown aspects about the coding proteins of Mic1, especially the hypothetical proteins of unknown function.

The three duplicated ORFs *Gp65*, *Gp66*, *and Gp72* of Mic1 encode homologous proteins of different oligomeric states (Figure 1B). The duplicated genes not only directly increase gene dosage, but also contribute to the adaptation of cyanophage to specific environmental conditions.[30] Further structural investigations together with biochemical assays enable us to propose a model to elucidate the oligomerization of Gp72. First, two monomeric Gp72 form a stable dimer via hydrophobic interactions, as shown in our crystal structure (Figure 2B). Afterwards, two Gp72 dimers further form a tetramer in a bottom-to-bottom manner via the disulfide bonds between two pairs of Cys115 residues (Figure 4A). Moreover, Cys28 of one tetramer forms a disulfide bond with that of the neighboring tetramer, resulting in the formation of Gp72 octamer in a lateral manner (Figure 4B). Eventually, Gp72 octamers could pack against each other to form super oligomers with the help of these two pairs of disulfide bonds (Figure 4C). Notably, the tetramerization of Gp66 might just be contributed by the longer helix α2, but not the cysteines, which are missing in Gp66 (Figure 1A,E).

Although mass spectrometric analysis indicated that Gp65, Gp66, and Gp72 might be structural components of Mic1, nothing is known

for their molecular functions. Crystal structure of Gp72, in combination with the various oligomeric states regulated by redox of the homologs, implied the distinct roles of these three homologs in the lifecycle of Mic1. These findings provide structural and biochemical hints for further elucidation of hypothetical proteins in the freshwater cyanophage Mic1.

ACKNOWLEDGMENTS

The authors thank Dr Haiteng Deng at Tsinghua University for the mass spectrometric analysis and the staff at the Shanghai Synchrotron Radiation Facility for technical assistance. This work is supported by the Ministry of Science and Technology of China (Grant Nos. 2018YFA0903100 and 2016YFA0400900), the National Natural Science Foundation of China (Grant Nos. 31630001 and 31621002), and the Chinese Academy of Sciences (the Innovative Academy for Seed Design).

CONFLICT OF INTEREST

The authors declare no potential conflicts of interest.

ORCID

Cong-Zhao Zhou https://orcid.org/0000-0002-6881-7151
Qiong Li https://orcid.org/0000-0003-2838-9380

REFERENCES

1. Clokie MR, Millard AD, Letarov AV, Heaphy S. Phages in nature. *Bacteriophage*. 2011;1:31-45.
2. Xia H, Li T, Deng F, Hu Z. Freshwater cyanophages. *Virol Sin*. 2013; 28:253-259.
3. Shestakov SV, Karbysheva EA. The role of viruses in the evolution of cyanobacteria. *Biol Bull Rev*. 2015;5:527-537.
4. Hambly E, Suttle CA. The viriosphere, diversity, and genetic exchange within phage communities. *Curr Opin Microbiol*. 2005;8:444-450.
5. Marston MF, Pierciey FJJr, Shepard A, et al. Rapid diversification of coevolving marine *Synechococcus* and a virus. *Proc Natl Acad Sci U S A*. 2012;109:4544-4549.
6. Huisman J, Codd GA, Paerl HW, Ibelings BW, Verspagen JMH, Visser PM. Cyanobacterial blooms. *Nat Rev Microbiol*. 2018;16: 471-483.
7. Lee J, Lee S, Jiang X. Cyanobacterial toxins in freshwater and food: important sources of exposure to humans. *Annu Rev Food Sci Technol*. 2017;8:281-304.
8. Deng LI, Hayes PK. Evidence for cyanophages active against bloom-forming freshwater cyanobacteria. *Freshw Biol*. 2008;53:1240-1252.
9. Thuduhenage DACT, Thuduhena AC. Harmful cyanobacterial blooms and developed cyanophages as a biological solution. *Bact Empire*. 2019;2:6-9.
10. Safferman RS, Morris ME. Algal virus: isolation. *Science*. 1963;140: 679-680.
11. Zhang QY, Gui JF. Diversity, evolutionary contribution and ecological roles of aquatic viruses. *Sci China Life Sci*. 2018;61:1486-1502.
12. Safferman RS, Cannon RE, Desjardins PR, et al. Classification and nomenclature of viruses of cyanobacteria. *Intervirology*. 1983;19: 61-66.
13. Edwards RA, Rohwer F. Viral metagenomics. *Nat Rev Microbiol*. 2005; 2:504-510.
14. You S, Wang M, Jiang Y, et al. The genome sequence of a novel cyanophage S-B64 from the Yellow Sea, China. *Curr Microbiol*. 2019; 76:681-686.
15. Hendrix RW. Evolution of dsDNA tailed phages. *Origin and Evolution of Viruses*. Florida, FL: Academic Press; 2008:219-227.
16. Brüssow H, Hendrix RW. Phage genomics: small is beautiful. *Cell*. 2002;108:13-16.
17. Jin H, Jiang Y-L, Yang F, et al. Capsid structure of a freshwater cyanophage siphoviridae Mic1. *Structure*. 2019;27:1-9.
18. Walden H. Selenium incorporation using recombinant techniques. *Acta Cryst D*. 2010;D66:352-357.
19. D'Arcy A, Bergfors T, Cowan-Jacob SW, Marsh M. Microseed matrix screening for optimization in protein crystallization: what have we learned?*Acta Cryst F*. 2014;F70:1117-1126.
20. Zbyszek Otwinowski WM. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol*. 1997;276:307-326.
21. Brodersen DE, Vonrhein C, Bricogne G, Nyborg J, Kjeldgaard M. Applications of single-wavelength anomalous dispersion at high and atomic resolution. *Acta Cryst D*. 2000;D56:431-441.
22. Adams PD, Afonine PV, Bunko´czi G et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst D*. 2010;D66:213-221.
23. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Cryst D*. 2010;D66:486-501.
24. Winn MD, Ballard CC, Cowtan KD et al. Overview of the CCP4 suite and current developments. *Acta Cryst D*. 2011;D67:235-242.
25. Chen VB, Arendall WB3rd, Headd JJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst D*. 2010;66:12-21.
26. Krissinel E. Stock-based detection of protein oligomeric states in jsPISA. *Nucleic Acids Res*. 2015;43:W314-W319.
27. Liisa Holm CS. Protein structure comparison by alignment of distance matrices. *J Mol Biol*. 1993;233:123-138.
28. Joshi M, Gakhar L, Fuentes EJ. High-resolution structure of the Tiam1 PHn-CC-Ex domain. *Acta Cryst F*. 2013;F69:744-752.
29. Brian Kuhlman GD, Ireton GC, Varani G, Stoddard BL, Baker D. Design of a novel globular protein fold with atomic-level accuracy. *Science*. 2003;302:1364-1368.
30. Wolfe GCCAKH. Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet*. 2008;9:938-950.
31. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res*. 2004;32:W526-W531.
32. Bond CS. TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics*. 2003;19:311-312.
33. Yang Feng, Jin Hua, Wang Xiao-Qian, et al. Genomic Analysis of Mic1 Reveals a Novel Freshwater Long-Tailed Cyanophage. *Frontiers in Microbiology*. 2020;11. http://dx.doi.org/10.3389/fmicb.2020.00484.

How to cite this article: Wang Y, Jin H, Yang F, et al. Crystal structure of a novel fold protein Gp72 from the freshwater cyanophage Mic1. *Proteins*. 2020;1–7. https://doi.org/10.1002/prot.25896