Virus Evolution, 2024, **10(1)**, veae082
DOI: https://doi.org/10.1093/ve/veae082
Advance Access Publication 26 September 2024
Research Article

Structural models predict a significantly higher binding affinity between the NblA protein of cyanophage Ma-LMM01 and the phycocyanin of *Microcystis aeruginosa* NIES-298 compared to the host homolog

Isaac Meza-Padilla, Brendan J. McConkey, Jozef I. Nissimov

Department of Biology, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada
*Corresponding author. Isaac Meza-Padilla, Department of Biology, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada. E-mail: imezapad@uwaterloo.ca; Jozef I. Nissimov, E-mail: jnissimov@uwaterloo.ca

Abstract

Horizontal gene transfer events between viruses and hosts are widespread across the virosphere. In cyanophage-host systems, such events often involve the transfer of genes involved in photosynthetic processes. The genome of the lytic cyanophage Ma-LMM01 infecting the toxic, bloom-forming, freshwater *Microcystis aeruginosa* NIES-298 contains a homolog of the *non-bleaching A* (*nblA*) gene, which was probably transferred from a cyanobacterial host. The function of the NblA protein is to disassemble phycobilisomes, cyanobacterial light-harvesting complexes that can comprise up to half of the cellular soluble protein content. NblA thus plays an essential dual role in cyanobacteria: it protects the cell from high-light intensities and increases the intracellular nitrogen pool under nutrient limitation. NblA has previously been shown to interact with phycocyanin, one of the main components of phycobilisomes. Using structural modeling and protein–protein docking, we show that the NblA dimer of Ma-LMM01 is predicted to have a significantly higher binding affinity for *M. aeruginosa* NIES-298 phycocyanin $(\alpha\beta)_6$ hexamers, compared to the host homolog. Protein–protein docking suggests that the viral NblA structural model is able to bind deeper into the phycocyanin groove. The main structural difference between the virus and host NblA appears to be an additional α-helix near the N-terminus of the viral NblA, which interacts with the inside of the phycocyanin groove and could thus be considered partly responsible for this deeper binding. Interestingly, phylogenetic analyses indicate that this longer *nblA* was probably acquired from a different *Microcystis* host. Based on infection experiments and previous findings, we propose that a higher binding affinity of the viral NblA to the host phycocyanin may represent a selective advantage for the virus, whose infection cycle requires an increased phycobilisome degradation rate that is not fulfilled by the NblA of the host.

Keywords: Cyanophage; non-bleaching A (NblA) protein; structural modeling; AlphaFold2; protein–protein docking; virus evolution; virus ecology; horizontal gene transfer (HGT)

1. Introduction

Aquatic viruses influence global biogeochemical cycles, control the abundance and diversity of their hosts, shape ecological food webs, accelerate coevolutionary processes, and transfer genetic information in marine and freshwater environments (Fuhrman 1999; Rohwer and Thurber 2009, DeLong et al. 2023). Virus-host horizontal gene transfer (HGT) events, however, are not unique to aquatic ecosystems. In fact, they take place across the whole virosphere. Examples include the double jelly-roll major capsid protein of diverse archaeal viruses and bacteriophages within *Varidnaviria*, probably exapted from a family of bacterial enzymes involved in carbohydrate metabolism (Krupovic et al. 2022); the acquisition of cholera toxin genes from the temperate single-stranded DNA bacteriophage CTXΦ by *Vibrio cholerae* (Waldor and Mekalanos 1996, Davis and Waldor 2003); as well as several HGTs in eukaryotic viruses, including members of the *Nucleocytoviricota*, and their hosts (Irwin et al. 2022). Collectively, HGTs have played foundational roles in macroevolutionary processes of both viruses and cells throughout the history of life on Earth (Forterre and Prangishvili 2009, Koonin et al. 2022).

In marine cyanophage-host systems, HGT events often involve the transfer of genes involved in photosynthesis. At the beginning of the twenty-first century, it was found that S-PM2, a cyanophage infecting *Synechococcus* strains, encodes D1 (*psbA*) and D2 (*psbD*) proteins in its genome (Mann et al. 2003). D1 and D2 are photosystem II reaction center proteins (Barber 2013). Lindell et al. (2004) discovered three more cyanophages infecting *Prochlorococcus* that also contained auxiliary metabolic genes encoding for

proteins involved in photosynthesis, namely, *psbA, psbD, hli, petE*, and *petF. Hli, petE*, and *petF* encode high-light-inducible protein, plastocyanin, and ferredoxin, respectively, all components of the photosynthetic process (Gross 2013, Hanke and Mulo 2013, Konert et al. 2022). Further, phylogenetic analyses suggested that these genes were not only horizontally transferred from cyanobacterial hosts to viruses but also that they were probably transferred back to the hosts (Lindell et al. 2004).

Although research on freshwater viruses is relatively limited compared to their marine counterparts, various cases of the presence of photosynthesis-related genes in the genomes of freshwater cyanophages have already been reported (Yoshida et al. 2008, Gao et al. 2012, Ou et al. 2015, Meng et al. 2023b). The double-stranded DNA lytic Ma-LMM01 virus infecting the toxic, bloom-forming *Microcystis aeruginosa* NIES-298 is arguably the most comprehensively characterized freshwater cyanophage to date (e.g. Yoshida et al. 2006, 2008, Morimoto et al. 2018). Its genome contains a homolog of the *non-bleaching A* (*nblA*) gene, which was probably horizontally acquired from a *Microcystis* host (Yoshida et al. 2008, Ou et al. 2015). The function of the host NblA protein is to disassemble phycobilisomes, the light-harvesting complexes of cyanobacteria, which can comprise up to half of the cellular soluble protein content (Grossman et al. 1993, Baier et al. 2004, Bienert et al. 2006). NblA thus plays an essential dual role in cyanobacteria: it protects the cell from high-light intensities and increases the intracellular nitrogen pool under nutrient limitation (Grossman et al. 1993, Collier and Grossman 1994, Baier et al. 2004). NblA has previously been shown to interact with phycocyanin, one of the main components of phycobilisomes (Luque et al. 2003, Bienert et al. 2006, Dines et al. 2008, Karradt et al. 2008, Nguyen et al. 2017). The fact that the *nblA* gene has been fixed in the Ma-LMM01 population readily suggests that it provides a selective advantage for the virus. Indeed, it is highly transcribed during infection (Yoshida-Takashima et al. 2012, Honda et al. 2014, Morimoto et al. 2018). Here we employ structural modeling and protein–protein docking to investigate which of the two NblAs present in the Ma-LMM01/*M. aeruginosa* NIES-298 virocell (i.e. the virus- or the host-encoded one) is predicted to have a higher binding affinity to the phycocyanin of the host. We elaborate on the potential implications of such a difference in the eco-evolutionary context of this virus-host system.

2. Materials and methods

2.1. Structural bioinformatics pipeline

The amino acid sequences of Ma-LMM01 NblA (vNblA), *M. aeruginosa* NIES-298 NblA (hNblA), *M. aeruginosa* NIES-298 phycocyanin α-subunit, and *M. aeruginosa* NIES-298 phycocyanin β-subunit were downloaded from the NCBI Protein Database (GenBank accession numbers: BAF36096, GBD54109, GBD54899, and GBD54900, respectively; Sayers et al. 2022). Structural modeling for the vNblA dimer, hNblA dimer, and *M. aeruginosa* NIES-298 phycocyanin $(\alpha\beta)_6$ hexamer (PC) was performed using AlphaFold2 (AF2) ColabFold (Jumper et al. 2021, Mirdita et al. 2022) with an NVIDIA A100-SXM4-40GB graphics processor. Secondary structure predictions for vNblA and hNblA were further corroborated using PSIPRED 4.0 (Buchan and Jones 2019). The sequence alignment between vNblA and hNblA was generated using MUSCLE 3.8 through the EMBL-EBI (Edgar 2004, Madeira et al. 2022). Entropy-based conservation values were calculated and mapped onto the structural models using AL2CO and University of California, San Francisco (UCSF) ChimeraX (Pei and Grishin 2001, Meng et al. 2023a). The pairwise structure comparison tool of the DALI server (Holm et al. 2023) was used to generate single-chain alignments for vNblA and hNblA. Hits with a DALI Z-score > 2 (i.e. 2 SDs above expected) were considered significant (Holm and Sander 1995). Protein–protein docking was carried out using ClusPro 2.0 (Kozakov et al. 2017). To construct the vNblA–PC and hNblA–PC complexes, the PC and NblA structural models were used as the receptor and ligand, respectively. The 30 Balanced models ranked by cluster size were downloaded and subsequently uploaded to the PRODIGY server (Xue et al. 2016) for binding affinity prediction at 25°C.

2.2. Phylogenetic analyses

Viral and cyanobacterial NblA sequences were downloaded from the NCBI Protein Database (Sayers et al. 2022) to update the phylogeny constructed by Ou et al. (2015). Accession numbers are as follows: Ma-LMM01, BAF36096; MaMV-DC, AGR48570; MaMV-DH0, UWV19634; *Microcystis* sp. M19BS1, MCA2624452; *Microcystis* sp. M20BS1, WP_287636433; *Raphidiopsis raciborskii* Cr2010 (previously *Cylindrospermopsis raciborskii* Cr2010), UJL33151; *Nostoc* sp. PCC 7120 (chromosome), BAB76216; *Nostoc* sp. PCC 7120 (plasmid), BAB77423; *Planktothrix agardhii* NIVA-CYA 126/8, KEI67021; *M. aeruginosa* PCC 7806, CAO88905; *M. aeruginosa* NIES-298, GBD54109; *Synechococcus elongatus* PCC 7942, ABB58157; *Synechocystis* sp. PCC 6803 (NblA1), BAA17955; *Thermosynechococcus vulcanus* str. Copeland, 2QDO_A; PaV-LD, ADZ31529; *Synechocystis* sp. PCC 6803 (NblA2), BAA17954. The NblA sequences were aligned using MUSCLE (Edgar 2004) and visualized in SeaView (Gouy et al. 2010). Blocks of evolutionarily conserved sites were generated using Gblocks (Castresana 2000). A maximum likelihood phylogenetic tree was computed using PhyML (Guindon et al. 2003, 2010) with 100 bootstrap replicates.

2.3. Statistical analyses

Statistical analyses were conducted using R version 4.3.0 (R Core Team 2023). In order to investigate whether the vNblA–PC and hNblA–PC predicted complexes had significantly different binding affinities, Welch's *t*-tests were employed. Prior to analyses, the normality of data was tested using a series of Shapiro–Wilk tests. Gibbs free energy change (ΔG) data, as well as the number of charged-charged (CCs), and polar-nonpolar (PNs) intermolecular contacts at the interface within a threshold distance of 5.5 Å, followed normal distributions ($P > .062$ in all cases). However, dissociation constant (K_d) data, and the number of charged-polar, charged-nonpolar, polar-polar, and nonpolar-nonpolar intermolecular contacts (CPs, CNs, PPs, and NNs, respectively) departed from normality. Thus, a log transformation was applied to K_d, CPs, CNs, PPs, and NNs data. After transformation, K_d, CPs, CNs, PPs, and NNs data followed normal distributions ($P > .063$ in all cases). ΔG (kcal mol^{-1}), K_d (M), CCs, PNs, CPs, CNs, PPs, and NNs were used as the dependent variables, and complex (vNblA-PC, hNblA–PC) as the grouping variable. Welch's *t*-tests were also employed to test for differences in the phycocyanin to chlorophyll *a* (PC/Chl *a*) absorbance ratio and fold change in Ma-LMM01-infected and uninfected *M. aeruginosa* NIES-298 cultures. Significance was defined at $P < .05$.

3. Results

3.1. Structural models

The vNblA and hNblA AF2 structural models are displayed in Fig. 1a, b, d, and e. Space-filling views of the models can be found in Supplementary Fig. S1. An α-helix, absent in the host model, can be observed near the N-terminus of the viral NblA. The

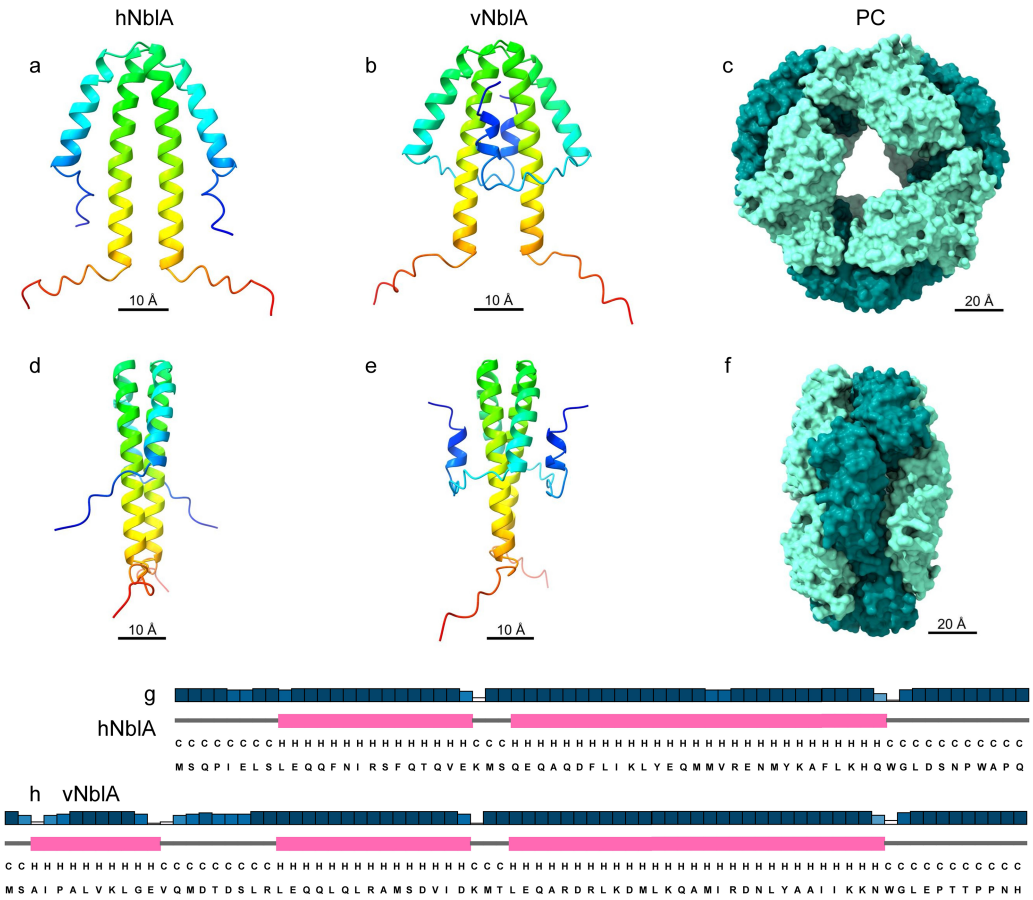

Figure 1. Structural models of the Ma-LMM01/*Microcystis aeruginosa* NIES-298 cyanophage-host system non-bleaching A (NblA) proteins. (a) Front and (d) side views of *M. aeruginosa* NIES-298 NblA (hNblA). (b) Front and (e) side views of Ma-LMM01 NblA (vNblA). NblA models are colored with a rainbow gradient, from the N-terminus in blue to the C-terminus in red. (c) Front and (f) side views of *M. aeruginosa* NIES-298 phycocyanin $(\alpha\beta)_6$ hexamer (PC). PC α- and β-subunits are colored in dark cyan and aquamarine, respectively. (g) hNblA and (h) vNblA secondary structure predictions. Helices are colored in pink and coils in gray. The confidence of prediction is depicted using a short-tall and white-blue gradient, with short white bars indicating a low prediction confidence, and tall blue bars a high prediction confidence. Note the additional α-helix near the N-terminus of the vNblA secondary and tertiary structural models. Space-filling views of vNblA and hNblA can be found in Supplementary Fig. S1.

structural model of PC, required for protein–protein docking and binding affinity prediction, is shown in Fig. 1c and f. Secondary structure prediction also supports the presence of the additional helix in vNblA, spanning from alanine 3 to glutamic acid 12 (Fig. 1g and h).

3.2. Virus-Host NblA sequence and structure conservation

The sequence conservation between vNblA and hNblA is mapped onto the structural models in Fig. 2a and b. The conserved residues may be directly involved in the interaction with the host's phycobilisomes. A significant DALI pairwise structure comparison between vNblA and hNblA (Z-score = 4.4) suggests that the structure of the vNblA model is more conserved compared to its amino acid sequence (Fig. 2c and d). Based on the virus-host NblA sequence alignment, the residues absent in hNblA span from leucine 7 to aspartic acid 18, a region containing the predicted α-helix unique to the vNblA model (Fig. 2e; Section 3.1).

3.3. Phylogenetic analysis of Ma-LMM01 NblA

To update the viral and cellular NblA phylogeny of Ou et al. (2015), we constructed a maximum likelihood phylogenetic tree including additional cyanophage and cyanobacterial sequences published since then (Fig. 3). The three *Microcystis* cyanophages, that is, Ma-LMM01, MaMV-DC, and MaMV-DH01, clustered together and further grouped with *Microcystis* sp. M19BS1 and *Microcystis* sp. M20BS1. *Microcystis aeruginosa* NIES-298, the known host of Ma-LMM01, was assigned to a different cluster shared by *Planktothrix*, *Raphidiopsis*, and *Nostoc* strains. Furthermore, a multiple sequence alignment shows that the NblA of *Microcystis* sp. M19BS1 (and, to a lesser extent, *Microcystis* sp. M20BS1) has a very similar sequence to that of Ma-LMM01, with an almost identical extension.

3.4. Protein–protein docking and binding affinity prediction

Statistical analyses indicate that the vNblA–PC and hNblA–PC predicted complexes have significantly different binding affinities, as determined by ΔG (Welch's *t*-test, $t_{49,667} = 3.017$, $P = .004$; Fig. 4a) and K_d ($t_{49,578} = 3.038$, $P = .004$; Fig. 4b). Likewise, the predicted complexes have significantly different CC, CP, CN, and PP intermolecular contacts at the interface within a threshold distance of 5.5 Å ($P < .041$ in all cases). No significant differences were detected regarding the number of PN and NN intermolecular contacts ($P > .05$ in both cases). This suggests that the lower ΔG and K_d values of the vNblA–PC complex are mainly due to an increased number of CC, CP, and CN intermolecular contacts.

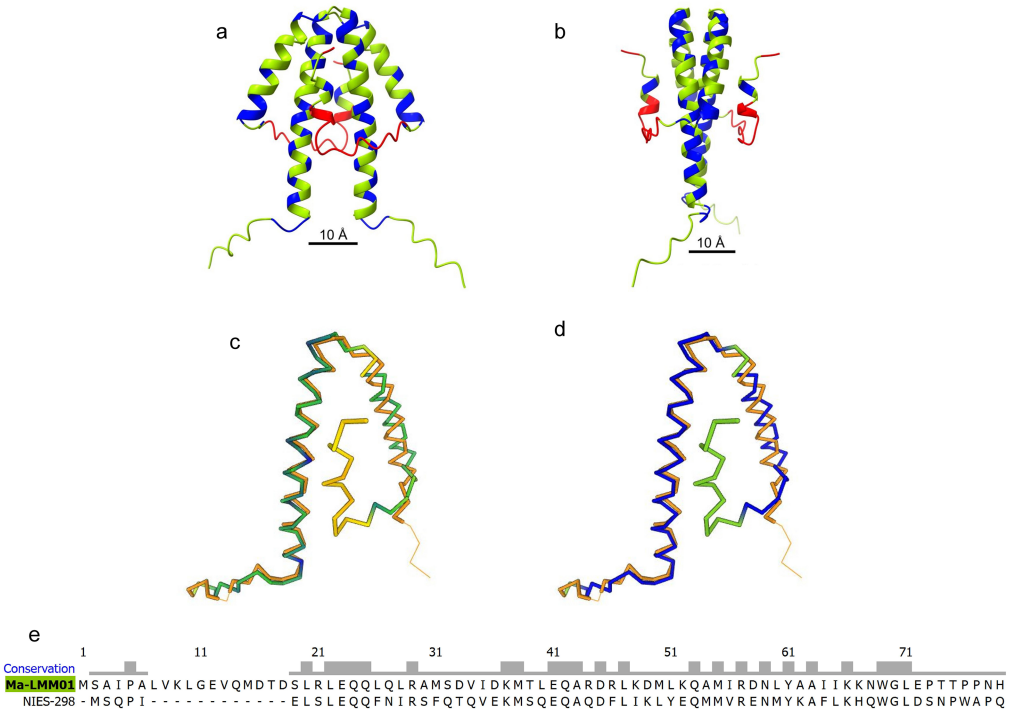

Figure 2. Structure and sequence conservation of the Ma-LMM01/*Microcystis aeruginosa* NIES-298 cyanophage-host system non-bleaching A (NblA) structural models. (a) Front and (b) side views of Ma-LMM01 NblA (vNblA) colored by sequence conservation according to the (g) Ma-LMM01/*M. aeruginosa* NIES-298 NblA sequence alignment. The AL2CO entropy-based amino acid sequence conservation measure between the two proteins is used. Since this is a pairwise comparison, light green and blue represent different and identical amino acids, respectively. Residues without conservation values, i.e. residues absent in *M. aeruginosa* NIES-298 NblA (hNblA) are shown in red. Single-chain superimpositions resulting from a DALI pairwise structure comparison with a significant *Z*-score = 4.4 are displayed in (c) and (d) using a yellow–green–blue scheme mapped onto the C-α trace of the vNblA model, yellow representing the lowest conservation values and blue the highest. In (c) and (d), the relative entropy (0–6.3 bits) was used as the conservation measure. In (c), vNblA is colored by amino acid sequence conservation; in (d), by structure conservation. hNblA is colored in orange. (e) Ma-LMM01/*M. aeruginosa* NIES-298 NblAs sequence alignment, where the short and tall bars above the residues indicate low and high sequence conservation values, respectively.

ΔG was found to be -15.89 ± 2.69 kcal mol^{-1} (mean ± SD; $n = 30$) for the hNblA–PC complex, and -17.66 ± 1.74 kcal mol^{-1} ($n = 30$) for the vNblA–PC complex; while K_d was -11.65 ± 1.97 M ($\log_{10}$ mean ± SD; $n = 30$) for the hNblA–PC complex, and -12.95 ± 1.27 M ($n = 30$) for the vNblA–PC complex. Both of these significantly lower ΔG and K_d values of the vNblA–PC predicted complex translate into a significantly higher binding affinity between the vNblA and PC structural models compared to the hNblA and PC models.

The best ranked hNblA–PC ClusPro 2.0 docking model (Fig. 4c and e) was predicted to have ΔG and K_d values of -14.9 kcal mol^{-1} and 1.20×10^{-11} M, respectively. In comparison, the best ranked vNblA–PC model (Fig. 4d and f) had a ΔG of -16.3 kcal mol^{-1} and a K_d of 1.10×10^{-12} M, almost an order of magnitude lower than the hNblA–PC model. It is worth noting that in most complexes the viral and cellular NblAs were in similar positions as they were in the best ranked models (Supplementary Fig. S2), showing consistency in the docking predictions. Most hNblA models (23/30) were bound superficially (compared to vNblA) to PC, with a smaller region inside the PC groove. In contrast, virtually all vNblA models were bound deep inside the PC groove (29/30). A closer view of the intermolecular interactions between vNblA and PC reveals that the two additional α-helices of the vNblA dimer interact with the inside of the PC groove (Fig. 5).

4. Discussion

The sequencing of the Ma-LMM01 genome (Yoshida et al. 2008) and more recently its *M. aeruginosa* NIES-298 host (Yamaguchi et al. 2018), along with the development of highly accurate structure prediction (Jumper et al. 2021) and other robust structural bioinformatics tools (Xue et al. 2016, Kozakov et al. 2017) provide a unique opportunity to model the NblA and PC structures of this cyanophage-host system and compare their predicted binding affinities. First, a qualitative analysis of the viral and host NblA models readily highlights an additional α-helix present in the viral structural model (Section 3.1; Fig. 1). There are a number of possible evolutionary scenarios for the acquisition of this longer version of *nblA* by Ma-LMM01. Phylogenetic evidence suggests that Ma-LMM01 *nblA* was horizontally acquired from a *Microcystis* host (Ou et al. 2015). The updated phylogeny presented here also supports that scenario. Structure conservation analyses, including a significant DALI pairwise structure comparison between vNblA and hNblA (Section 3.2; Fig. 2), support virus-host structural homology as well. Interestingly, however, Ma-LMM01 (or an ancestor of the virus) appears to have acquired the *nblA* gene from a *Microcystis* strain other than *M. aeruginosa* NIES-298, namely, a *Microcystis* sp. M19BS1- or *Microcystis* sp. M20BS1-like host (Section 3.3; Fig. 3).

An increase in genome size (the acquisition of this particular *nblA* version by the virus) must constitute a selective advantage

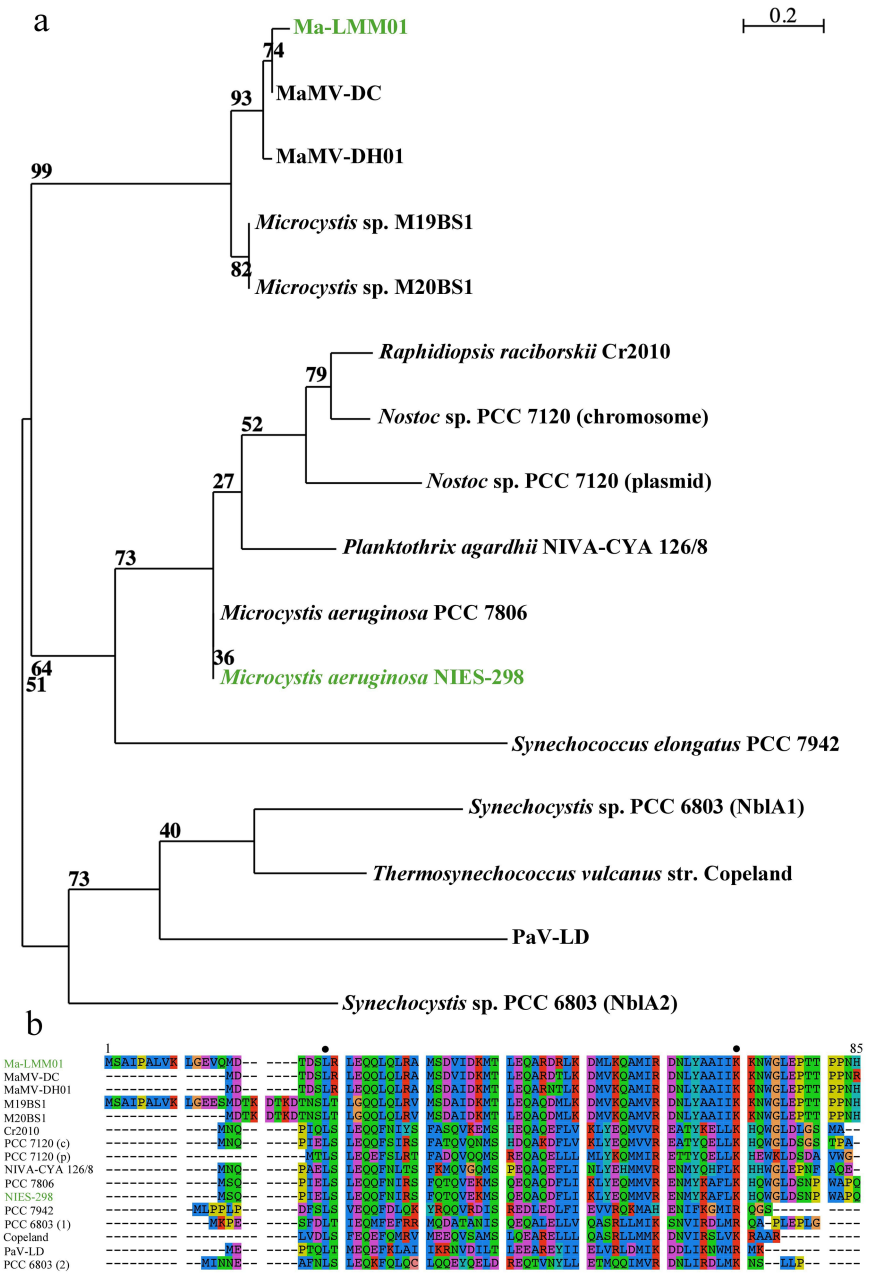

Figure 3. Maximum likelihood phylogenetic tree and multiple sequence alignment of viral and cyanobacterial NblA proteins. (a) PhyML tree constructed with 100 bootstrap replicates. Branch support values are shown above the branches. The scale bar indicates 0.2 amino acid substitutions per site. Note how the three *Microcystis* cyanophages, including Ma-LMM01, cluster together and further group with *Microcystis* sp. M19BS1 and *Mycrocystis* sp. M20BS1. (b) Multiple sequence alignment using the Clustal 2 color scheme. Closed circles denote the interval of evolutionarily conserved sites generated via Gblocks. Ma-LMM01 and *M. aeruginosa* NIES-298 are highlighted in green. Note the highly similar NblA proteins of Ma-LMM01 and *Microcystis* sp. M19BS1, both in sequence and length. The genome of *Synechocystis* sp. PCC 6803 contains two *nblA* genes, while *Nostoc* sp. PCC 7120 has one *nblA* gene in its chromosome and another in a plasmid. See Section 2.2 for GenBank accession numbers.

strong enough to counter the possible trade-off associated with the greater amount of time and resources required to replicate the viral genome (Mills et al. 1967, DiMaio 2012, Edwards et al. 2021). The quantitative analyses conducted on the vNblA–PC and hNblA–PC predicted complexes indicate a significantly lower ΔG and K_d for the vNblA–PC complex, which translate into a higher binding affinity (Section 3.4; Fig. 4). Not surprisingly, the number of intermolecular contacts was also predicted to be greater in vNblA–PC compared to the hNblA–PC complex in most contact property categories. The exact binding mechanism of NblA in phycobilisomes has not been resolved experimentally to date. However, various studies have provided insights into the mode of action of NblA (Luque et al. 2003, Baier et al. 2004, Bienert et al. 2006, Dines et al. 2008, Karradt et al. 2008, Sendersky et al. 2014, 2015, Nguyen et al. 2017, Levi et al. 2018, Hu et al. 2020). For example, Bienert et al. (2006) determined the crystal structure of NblA from *Nostoc* sp. PCC 7120, and using pull-down assays, found that NblA bound the α-subunits of phycocyanin and phycoerythrocyanin; while Nguyen et al. (2017) detected a protein interaction between *S. elongatus* UTEX 2973 NblA and β-phycocyanin. Dines et al. (2008) determined the NblA structures of *S. elongatus* PCC 7942 and *T. vulcanus*, and using random mutagenesis, found that

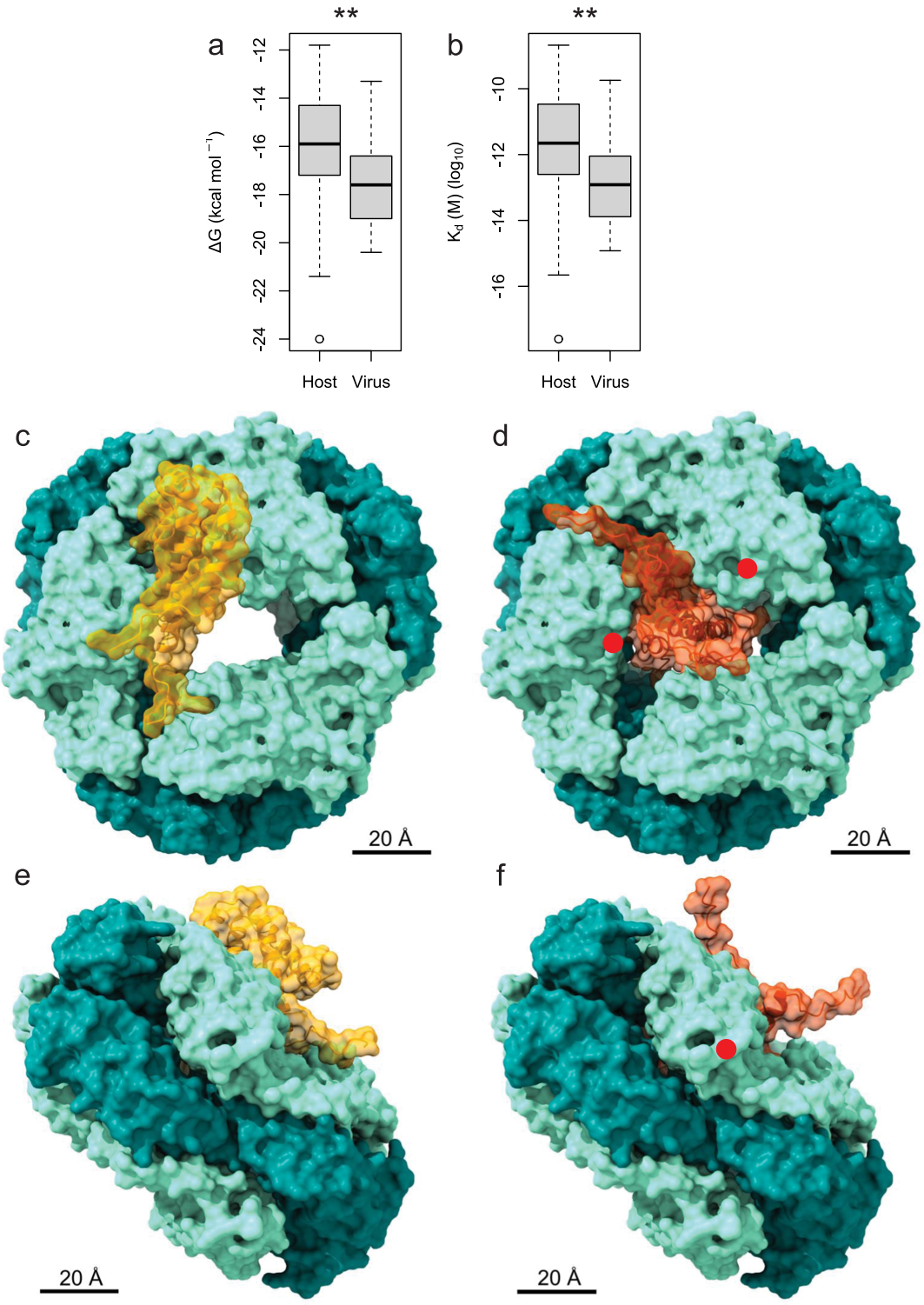

Figure 4. Protein-protein docking and binding affinity prediction between Ma-LMM01 non-bleaching A dimer (vNblA) and *Microcystis aeruginosa* NIES-298 phycocyanin $(\alpha\beta)_6$ hexamer (PC) structural models compared to *M. aeruginosa* NIES-298 NblA dimer (hNblA) and PC models. (a) ΔG (kcal mol^{-1}) and (b) K_d (M) of hNblA–PC (host) and vNblA–PC (virus) predicted complexes. See text for statistical values. (c) Front and (e) side views of the best ranked ClusPro 2.0 hNblA–PC docking model. (d) Front and (f) side views of the best ranked vNblA–PC model. hNblA is colored in orange, while vNblA in orange red; PC α- and β-subunits are colored in dark cyan and aquamarine, respectively. In (d) and (f), closed red circles indicate the sites where the additional vNblA α-helices interact with the inside of the PC groove. Note how vNblA is predicted to bind deeper into the PC groove compared to hNblA. See text for binding affinities of the displayed docking models.

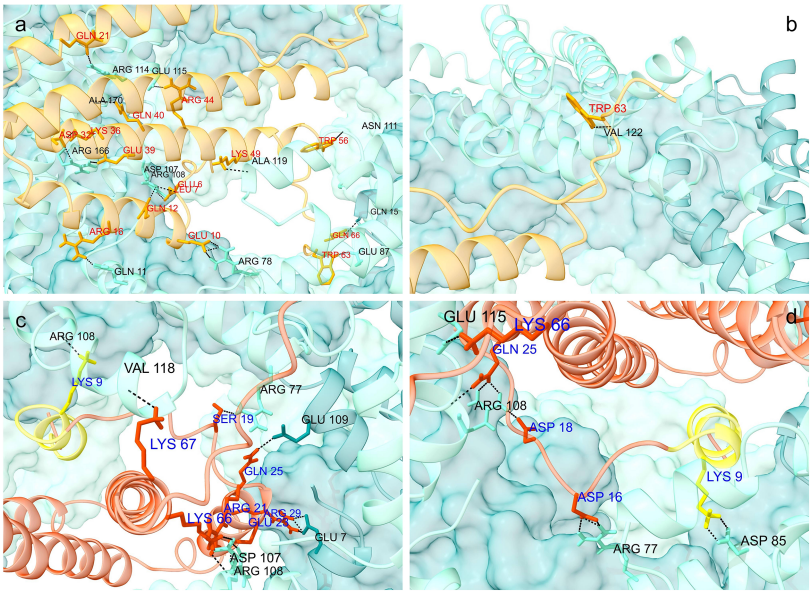

Figure 5. Close-up views of the intermolecular interactions between Ma-LMM01 non-bleaching A dimer (vNblA), *Microcystis aeruginosa* NIES-298 NblA dimer (hNblA), and *M. aeruginosa* NIES-298 phycocyanin $(\alpha\beta)_6$ hexamer (PC) structural models. (a and b) Interactions between hNblA and PC. hNblA is colored in orange; PC α- and β-subunits are colored in dark cyan and aquamarine, respectively. Hydrogen bonds are represented by dotted black lines. hNblA and PC interacting residues are labeled in red and black, respectively. Atoms of interacting residues are displayed in stick style. For clarity purposes, the non-interacting PC chains are represented using space-filling instead of ribbons. (c and d) Interactions between vNblA and PC. vNblA is colored in orange red, while its interacting residues are labeled in blue. The additional vNblA α-helix is highlighted in yellow. Note how lysine 9 in the additional helix of both NblA chains interacts with the inside of the PC groove.

amino acids essential for the interaction with phycobilisome proteins were located across the two NblA helices. The hNblA–PC docking models from the present study are consistent with those reported by Nguyen et al. (2017), where *S. elongatus* PCC 7942 NblA and PC X-ray structures docked similarly to *M. aeruginosa* NIES-298 NblA and PC structural models. In contrast, the vNblA–PC docking models differ drastically from both hNblA–PC and those reported previously for *S. elongatus* (Nguyen et al. 2017). In particular, the vNblA model was predicted to bind deeper into the PC groove. The additional α-helix, which appears to be the main structural difference between vNblA and these two cellular NblAs, interacts with the inside of the PC groove and may then be considered partly responsible for this deeper binding. Therefore, we suggest that the selective advantage referred to above could possibly be the increased binding affinity that the viral NblA is predicted to have for PC, as discussed below.

What may a higher binding affinity between vNblA and PC imply physiologically? In cyanophages MaMV-DC and PaV-LD, *nblA* expression is accompanied by a significant decline in the phycocyanin and phycocyanobilin content of their hosts (Gao et al. 2012, Ou et al. 2015). Furthermore, when Ma-MVDC and PaV-LD *nblA* genes were introduced and expressed in the model cyanobacterium *Synechocystis* sp. PCC 6803, there was a significant decrease in phycocyanin and phycocyanobilin peaks, as well as in the PC/Chl *a* absorbance ratio (Gao et al. 2012, Ou et al. 2015). We have observed a very similar pattern of significant phycocyanin absorbance decline in *M. aeruginosa* NIES-298 during Ma-LMM01 infection (Supplementary Fig. S3). Moreover, transcription analyses have shown that infected and uninfected *M. aeruginosa* NIES-298 cells maintain the same level of expression of the cyanobacterial *nblA* (Yoshida-Takashima et al. 2012). In contrast, the transcription of the viral *nblA* gene has been found to be 20-fold higher compared to the host's, suggesting that vNblA acts on the phycobilisome degradation process in a more specific and effective manner than hNblA (Yoshida-Takashima et al. 2012). Thus, although a tighter binding does not necessarily imply an increased catalytic rate and may even hinder enzymatic function, based on our results and previous findings, we propose that a higher binding affinity of vNblA to the host phycocyanin would be expected to contribute to an increase in the disassembly rate of the host phycobilisomes. This could be essential for the virus if the performance of the host protein is not sufficient during infection. As previously described, phycobilisomes are the light-harvesting complexes of cyanobacteria, and can comprise up to half of the cellular soluble protein content (Grossman et al. 1993, Baier et al. 2004, Bienert et al. 2006). A higher phycobilisome degradation rate in the Ma-LMM01/*M. aeruginosa* NIES-298 virocell could lead to a larger burst size after a steeper increase in the intracellular nitrogen pool (Yoshida et al. 2008, Morimoto et al. 2018, 2020). A faster degradation of phycobilisomes would also imply a more efficient protection for the virocell from high-light intensities to which *M. aeruginosa* NIES-298 is highly susceptible (Yoshida et al. 2008, Morimoto et al. 2018, 2020).

Based on the significantly lower ΔG and K_d of the vNblA–PC docking models, we conclude that vNblA is predicted to have a higher binding affinity for PC compared to hNblA. The potential implications of this appear to be in line with the biology of the Ma-LMM01/*M. aeruginosa* NIES-298 virocell. However, rather than aiming to provide a definitive answer regarding the mode of action of the vNblA protein, these structural models invite further investigation through experimental approaches, including structure determination efforts. They also highlight the statistics-based hypothesis generating potential of the structural bioinformatics pipeline implemented here, particularly for comparative interactomics in the fields of evolutionary virology and viral ecology.

Acknowledgements

This study would not have been possible without the continuous support of Dr Medardo Meza Olea and Dr Esther Padilla Calderón. Thanks are also due to two anonymous reviewers, whose insightful comments greatly helped to enhance the quality of this work.

Author contributions

I.M.-P. conceived the idea, carried out the bioinformatics pipeline, conducted the infection experiment and statistical analyses, analyzed the data, and wrote the manuscript. B.J.M. provided additional insights on structural modeling and molecular docking. J.I.N. administered the project, provided resources and funding, and supervised the work. I.M.-P. and J.I.N. designed the infection experiment. All authors contributed to the refinement (review and editing) of the final manuscript prior to its submission and publication.

Supplementary data

Supplementary data is available at *VEVOLU Journal* online.

Conflict of interest: None declared.

Funding

This work was funded by Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants (2022-03350 and 2022-00329), a Phycological Society of America Norma J. Lang Early Career Researcher Fellowship awarded to J.I.N.; a Mitacs Globalink Graduate Fellowship, a University of Waterloo International Master's Award of Excellence, and a Consejo Nacional de Humanidades, Ciencias y Tecnologías (Conahcyt) Becas de Posgrado para Maestrías y Doctorados en Ciencias y Humanidades en el Extranjero scholarship awarded to I.M.-P.

Data availability

The binding affinity prediction data is available in Supplementary Material (Supplementary Tables S1, S2). The structural models can be generated following the structural bioinformatics pipeline described in Section 2.1. Alternatively, the models can be provided upon reasonable request to I.M.-P.

References

Baier K, Lehmann H, Stephan DP *et al.* NblA is essential for phycobilisome degradation in *Anabaena* sp. strain PCC 7120 but not for development of functional heterocysts. *Microbiology* 2004;**150**:2739–49.

Barber J. Photosystem II: Redox and protein components. In: Lennarz WJ, Lane MD (eds.), *Encyclopedia of Biological Chemistry II*. USA: Elsevier, 2013, 512–9.

Bienert R, Baier K, Volkmer R *et al.* Crystal structure of NblA from *Anabaena* sp. PCC 7120, a small protein playing a key role in phycobilisome degradation. *J Biol Chem* 2006;**281**: 5216–23.

Buchan DW, Jones DT. The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res* 2019;**47**:W402–W407.

Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000;**17**:540–52.

Collier JL, Grossman A. A small polypeptide triggers complete degradation of light-harvesting phycobiliproteins in nutrient-deprived cyanobacteria. *EMBO J* 1994;**13**:1039–47.

Davis BM, Waldor MK. Filamentous phages linked to virulence of *Vibrio cholerae*. *Curr Opin Microbiol* 2003;**6**:35–42.

DeLong JP, Van Etten JL, Dunigan DD. Lessons from chloroviruses: the complex and diverse roles of viruses in food webs. *J Virol* 2023;**97**:e00275–23.

DiMaio D. Viruses, masters at downsizing. *Cell Host Microbe* 2012;**11**:560–61.

Dines M, Sendersky E, David L *et al.* Structural, functional, and mutational analysis of the NblA protein provides insight into possible modes of interaction with the phycobilisome. *J Biol Chem* 2008;**283**:30330–40.

Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinf* 2004;**5**:1–19.

Edwards KF, Steward GF, Schvarcz CR. Making sense of virus size and the tradeoffs shaping viral fitness. *Ecol Lett* 2021;**24**:363–73.

Forterre P, Prangishvili D. The great billion-year war between ribosome-and capsid-encoding organisms (cells and viruses) as the major source of evolutionary novelties. *Ann NY Acad Sci* 2009;**1178**:65–77.

Fuhrman JA. Marine viruses and their biogeochemical and ecological effects. *Nature* 1999;**399**:541–48.

Gao EB, Gui JF, Zhang QY. A novel cyanophage with a cyanobacterial nonbleaching protein A gene in the genome. *J Virol* 2012;**86**:236–45.

Gouy M, Guindon S, Gascuel O. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 2010;**27**:221–24.

Gross EL. Plastocyanin. In: Lennarz WJ, Lane MD (eds.), *Encyclopedia of Biological Chemistry II*. USA: Elsevier, 2013, 530–2.

Grossman AR, Schaefer MR, Chiang GG *et al.* The phycobilisome, a light-harvesting complex responsive to environmental conditions. *Microbiol Rev* 1993;**57**:725–49.

Guindon S, Dufayard JF, Lefort V *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010;**59**:307–21.

Guindon S, Gascuel O, Rannala B. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 2003;**52**:696–704.

Hanke GUY, Mulo P. Plant type ferredoxins and ferredoxin-dependent metabolism. *Plant Cell Environ* 2013;**36**:1071–84.

Holm L, Laiho A, Törönen P *et al.* DALI shines a light on remote homologs: one hundred discoveries. *Protein Sci* 2023;**32**:e4519.

Holm L, Sander C. Dali: a network tool for protein structure comparison. *Trends Biochem Sci* 1995;**20**:478–80.

Honda T, Takahashi H, Sako Y *et al.* Gene expression of *Microcystis aeruginosa* during infection of cyanomyovirus Ma-LMM01. *Fish Sci* 2014;**80**:83–91.

Hu PP, Hou JY, Xu YL *et al.* The role of lyases, NblA and NblB proteins and bilin chromophore transfer in restructuring the cyanobacterial light-harvesting complex. *Plant J* 2020;**102**:529–40.

Irwin NA, Pittis AA, Richards TA *et al.* Systematic evaluation of horizontal gene transfer between eukaryotes and viruses. *Nat Microbiol* 2022;**7**:327–36.

Jumper J, Evans R, Pritzel A *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;**596**:583–89.

Karradt A, Sobanski J, Mattow J *et al.* NblA, a key protein of phycobilisome degradation, interacts with ClpC, a HSP100 chaperone partner of a cyanobacterial Clp protease. *J Biol Chem* 2008;**283**:32394–403.

Konert MM, Wysocka A, Koník P *et al.* High-light-inducible proteins HliA and HliB: pigment binding and protein–protein interactions. *Photosynth Res* 2022;**152**:317–32.

Koonin EV, Dolja VV, Krupovic M. The logic of virus evolution. *Cell Host Microbe* 2022;**30**:917–29.

Kozakov D, Hall DR, Xia B *et al*. The ClusPro web server for protein–protein docking. *Nat Protoc* 2017;**12**:255–78.

Krupovic M, Makarova KS, Koonin EV. Cellular homologs of the double jelly-roll major capsid proteins clarify the origins of an ancient virus kingdom. *Proc Natl Acad Sci* 2022;**119**:e2120620119.

Levi M, Sendersky E, Schwarz R. Decomposition of cyanobacterial light harvesting complexes: NblA-dependent role of the bilin lyase homolog NblB. *Plant J* 2018;**94**:813–21.

Lindell D, Sullivan MB, Johnson ZI *et al*. Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc Natl Acad Sci* 2004;**101**:11013–18.

Luque I, Ochoa de Alda JA, Richaud C *et al*. The NblAI protein from the filamentous cyanobacterium *Tolypothrix* PCC 7601: regulation of its expression and interactions with phycobilisome components. *Mol Microbiol* 2003;**50**:1043–54.

Madeira F, Pearce M, Tivey AR *et al*. Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res* 2022;**50**:W276–W279.

Mann NH, Cook A, Millard A *et al*. Bacterial photosynthesis genes in a virus. *Nature* 2003;**424**:741.

Meng EC, Goddard TD, Pettersen EF *et al*. UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci* 2023a;**32**:e4792.

Meng LH, Ke F, Zhang QY *et al*. Biological and genomic characteristics of MaMV-DH01, a novel freshwater *Myoviridae* cyanophage strain. *Microbiol Spectr* 2023b;**11**:e02888–22.

Mills DR, Peterson RL, Spiegelman S. An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc Natl Acad Sci* 1967;**58**:217–24.

Mirdita M, Schütze K, Moriwaki Y *et al*. ColabFold: making protein folding accessible to all. *Nat Methods* 2022;**19**:679–82.

Morimoto D, Kimura S, Sako Y *et al*. Transcriptome analysis of a bloom-forming cyanobacterium *Microcystis aeruginosa* during Ma-LMM01 phage infection. *Front Microbiol* 2018;**9**:317806.

Morimoto D, Šulčius S, Yoshida T. Viruses of freshwater bloom-forming cyanobacteria: genomic features, infection strategies and coexistence with the host. *Environ Microbiol Rep* 2020;**12**:486–502.

Nguyen AY, Bricker WP, Zhang H *et al*. The proteolysis adaptor, NblA, binds to the N-terminus of β-phycocyanin: implications for the mechanism of phycobilisome degradation. *Photosynth Res* 2017;**132**:95–106.

Ou T, Gao XC, Li SH *et al*. Genome analysis and gene *nblA* identification of *Microcystis aeruginosa* myovirus (MaMV-DC) reveal the evidence for horizontal gene transfer events between cyanomyovirus and host. *J Gen Virol* 2015;**96**:3681–97.

Pei J, Grishin NV. AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics* 2001;**17**:700–12.

R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 2023. https://www.R-project.org/ (April 2024, date last accessed).

Rohwer F, Thurber RV. Viruses manipulate the marine environment. *Nature* 2009;**459**:207–12.

Sayers EW, Beck J, Bolton EE *et al*. Database resources of the national center for biotechnology information. *Nucleic Acids Res* 2022;**50**:D20–D26.

Sendersky E, Kozer N, Levi M *et al*. The proteolysis adaptor, NblA, initiates protein pigment degradation by interacting with the cyanobacterial light-harvesting complexes. *Plant J* 2014;**79**:118–26.

Sendersky E, Kozer N, Levi M *et al*. The proteolysis adaptor, NblA, is essential for degradation of the core pigment of the cyanobacterial light-harvesting complex. *Plant J* 2015;**83**:845–52.

Waldor MK, Mekalanos JJ. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* 1996;**272**:1910–14.

Xue LC, Rodrigues JP, Kastritis PL *et al*. PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics* 2016;**32**:3676–78.

Yamaguchi H, Suzuki S, Kawachi M. Improved draft genome sequence of *Microcystis aeruginosa* NIES-298, a microcystin-producing cyanobacterium from Lake Kasumigaura, Japan. *Genome Announc* 2018;**6**:10–1128.

Yoshida T, Nagasaki K, Takashima Y *et al*. Ma-LMM01 infecting toxic *Microcystis aeruginosa* illuminates diverse cyanophage genome strategies. *J Bacteriol* 2008;**190**:1762–72.

Yoshida T, Takashima Y, Tomaru Y *et al*. Isolation and characterization of a cyanophage infecting the toxic cyanobacterium *Microcystis aeruginosa*. *Appl Environ Microbiol* 2006;**72**:1239–47.

Yoshida-Takashima Y, Yoshida M, Ogata H *et al*. Cyanophage infection in the bloom-forming cyanobacteria *Microcystis aeruginosa* in surface freshwater. *Microbes Environ* 2012;**27**:350–55.

Virus Evolution, 2024, 10(1), veae082 , DOI: https://doi.org/10.1093/ve/veae082, Advance Access Publication 26 September 2024, Research Article