bioRxiv preprint doi: https://doi.org/10.1101/2020.02.07.937771; this version posted March 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Prochlorococcus phage ferredoxin: structural characterization and interactions with cyanobacterial sulfite reductases

Ian J. Campbell[1,2], Jose L. Olmos, Jr. [1,2], Weijun Xu[2], Dimithree Kahanda[2], Joshua T. Atkinson[2], Othneil N. Sparks[2], Mitchell D. Miller[2], George N. Phillips[2,3], George N. Bennett[2,4], and Jonathan J. Silberg[2,4,5]*

From the [1]Biochemistry and Cell Biology Graduate Program, Rice University, 6100 Main Street, MS-140, Houston, TX 77005; [2]Department of Biosciences, Rice University, 6100 Main Street, MS-140, Houston, TX 77005; [3]Department of Chemistry, Rice University, 6100 Main Street, MS-140, Houston, TX 77005; [4]Department of Chemical and Biomolecular Engineering, Rice University, 6100 Main Street, MS-362, Houston, TX 77005; [5]Department of Bioengineering, Rice University, 6100 Main Street, MS-142, Houston, TX 77005

Running title: *Phage Fd characterization and host SIR interactions*

*To whom correspondence should be addressed: Jonathan J. Silberg: Biosciences Department, Rice University, Houston, TX 77005; joff@rice.edu; Tel. (713) 348-3849

Keywords: cyanobacteria, electron transfer, ferredoxin, marine, bacteriophage, sulfite reductase, reductase, hydrogen sulfide

ABSTRACT

Marine cyanobacteria are infected by phage whose genomes encode ferredoxin (Fd) electron carriers. While these Fds are thought to redirect the energy harvested from light to phage-encoded oxidoreductases that enhance viral fitness, it is not clear how the biochemical and biophysical properties of phage Fds relate to those in photosynthetic organisms. Bioinformatic analysis using a sequence similarity network revealed that phage Fds are most closely related to cyanobacterial Fds that transfer electrons from photosystems to oxidoreductases involved in nutrient assimilation. Structural analysis of myovirus P-SSM2 Fd (pssm2-Fd), which infects *Prochlorococcus marinus*, revealed high similarity to cyanobacterial Fds (≤0.5 Å RMSD). Additionally, pssm2-Fd exhibits a low midpoint reduction potential (-336 mV vs. SHE) similar to other photosynthetic Fds, albeit lower thermostability (T_m = 28°C) than many Fds. When expressed in an *Escherichia coli* strain with a sulfite assimilation defect, pssm2-Fd complemented growth when coexpressed with a *Prochlorococcus marinus* sulfite reductase, revealing that pssm2-Fd can transfer electrons to a host protein involved in nutrient assimilation. The high structural similarity with cyanobacterial Fds and reactivity with a host sulfite reductase suggests that phage Fds evolved to transfer electrons to both phage- and cyanobacterial-encoded oxidoreductases.

Introduction

Prochlorococcus marinus is thought to be the most prevalent photosynthetic organism on Earth, with a global abundance of 10^{27} cells (1), making it a key player in biogeochemical processes. This oligotrophic cyanobacterium inhabits the euphotic zone of oceans (40°N to 40°S) and is projected to expand both in density and range as global temperatures rise (1). Collectively fixing four gigatons of carbon annually (1), the diverse ecotypes of *P. marinus* thrive at many depths, where light, nutrient availability, and temperature vary (2–4). Up to 24% of the CO_2 fixed by *P. marinus* is released as dissolved organic carbon, providing critical feedstocks for heterotrophic organisms and supporting the larger oceanic ecosystem (5).

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.07.937771; this version posted March 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The core metabolic systems in *P. marinus* have diverged significantly from other cyanobacteria due in part to genomic streamlining (6–10). Additionally, *P. marinus* lacks phycobilisomes widely used by other cyanobacteria for light harvesting and instead uses a divinyl chlorophyll a_2/b_2 complex (11). Furthermore, this cyanobacterium contains only one set of core photosystem components while other cyanobacteria maintain several homologs (8, 12). Nitrogen assimilation pathways have been pruned, with species often lacking the ability to reduce either nitrate or nitrite (13, 14). Various ecotypes are incapable of heterotrophic growth, lacking several tricarboxylic acid cycle genes (8).

The life cycle of *P. marinus* is influenced by viruses. Cyanophages that infect *P. marinus* have evolved genomes with as many as 327 open reading frames (ORFs) (15). Viruses from three clades infect *P. marinus*, including T4-like myoviruses, T7-like podoviruses, and, less commonly, members of *Siphoviridae* (16). In some ecosystems, as many as 50% of cyanobacteria may be infected at any point in time (17). Some cyanophages exhibit infection cycles that are closely tied to day-night rhythms of cyanobacteria, although individual viruses vary in their responses (18). Once infection begins, many host genes are repressed in favor of phage homologs (19), which can decrease cyanobacterial carbon fixation, potentially keeping ~5 gigatons of carbon in the atmosphere (20). While recent estimates suggest that cyanophages are limited by energy rather than nutrients (21), our understanding of their controls over host electron transfer (ET) remain limited.

Following infection, cyanophages express metabolic genes to modulate host photosynthesis (22). Phage can express their own photosystem II reaction core (D1 protein) during infection (17, 23, 24), with viral transcripts equaling or even exceeding host transcripts in some settings. Certain cyanophages encode an entire photosystem I (PSI) complex that is thought to be a more promiscuous oxidizer of electron carriers than host PSI, which may push host cells toward cyclic photosynthesis (25). Viral plastocyanin homologs, which possess altered charged surfaces compared with host plastocyanin, can also be expressed and are thought to transfer electrons directly to cytochrome aa_3 oxidase rather than PSI as a mechanism for managing redox balance (26). Other phage genes implicated in oxygenic photosynthesis include *hli* (high-light inducible protein), *pebA* (15,16-dihydrobiliverdin:Fd reductase), *psbD* (PSII D2 protein), *pcyA* (phycocyanobilin:Fd reductase), and *petF* (Fd) (15). Although the sequence similarities of these genes to their cyanobacterial homologs give clues to their cellular roles, we do not know how these proteins interact with host proteins to tune the host electron fluxome.

The *Prochlorococcus* P-SSM2 phage Fd (pssm2-Fd) supports the reduction of phage phycoerythrobilin synthase (PebS), which catalyzes the reduction of biliverdin IX_α to *3E/3Z* phycoerythrobilin (27, 28). While prior studies showed that some Fds support ET to diverse partner oxidoreductases (29–32), it remains unclear how the biophysical properties of this and other phage Fds compare to host Fds, which can function as ET hubs that couple light harvesting to a range of reductive metabolic reactions (31–33). In addition, it is not known how the specific constraints of viral evolution might have modulated the primary structure and partner specificity of phage Fd.

In this study, we used a sequence similarity network (SSN) analysis to examine how cyanophage and cyanobacterial Fds relate to one another. We also determined the first structure of a cyanophage Fd, which has high structural similarity to cyanobacterial Fds. We show that this Fd has a low midpoint potential characteristic of photosynthetic Fds, establish that this protein has a low thermostability (T_m = 28°C), and show that it supports ET to a host oxidoreductase involved in sulfur assimilation. These results suggest that viral Fds support ET to both phage- and cyanobacterial-encoded oxidoreductases, and they extend our understanding about the ways viruses might alter marine cyanobacteria sulfur cycling.

Results

Cyanophage and host Fds: evolutionary links

To better understand the prevalence of cyanophage Fds, we searched all 8392 viral genomes available in the Integrated Microbial Genomes and Microbiomes database for ORFs matching the Interpro signature for [2Fe-2S] Fds (34). This genome mining yielded 26 Fds from phages that have been shown to infect *Prochlorococcus* and *Synechococcus* (including pssm2-Fd), a Fd from an uncultured marine virus, and a Fd from the giant *Pseudomonas* phage OBP

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.07.937771; this version posted March 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(35). All viral ORFs encoded Fds with 95 to 97 residues, with the exception of the *Pseudomonas* phage OBP which was 83 residues.

We compared the primary structures of the viral Fds with 807 cyanobacterial Fds obtained from a recent genome mining effort (36). *Zea mays* Fd (zm-Fd1) was included as a frame of reference because this protein has been intensively studied (37, 38). When the Fds were compiled into a SSN (39), the cyanobacterial Fds presented over a dozen distinct clusters, which arose because of variation in Fd length (Figure 1A). Across the twelve largest clusters, the average Fd length varies by ~100 amino acids. These clusters also vary in the number of cyanobacterial species represented and the average Fd size, with the top five most species diverse clusters being VI (100 ±8 residues) > IX (115 ±9 residues) > VII (106 ±13 residues) > I (79 ±2 residues) > XI (163 ±8 residues) (Dataset S1).

The phage Fds all localize to Cluster VI in close association with cyanobacterial Fds (Figure 1B), with the exception of *Pseudomonas* phage OBP Fd, which does not cluster with any Fds. Cluster VI includes numerous structurally-characterized Fds, including Fds from *Synechocystis sp.* PCC 6803 (s6803-Fd1), *Nostoc sp.* PCC 7119 (nsp-Fd1), *Thermosynechococcus elongatus* (te-Fd1), and *Mastigocladus laminosus* (ml-Fd1). Additionally, this cluster contains zm-Fd1, the Fd included as a frame of reference.

Among the Fds in Cluster VI, interactions with a range of oxidoreductases have been documented. s6803-Fd1 binds PSI, glutamate synthase (GS), and Fd-thioredoxin reductase (FTR) (40–43). nsp-Fd1 interacts with FNR and nitrate reductase (NIR) (44, 45). te-Fd1 interacts with FNR, PSI, heme oxygenase (HO), PcyA, and 4-Hydroxy-3-methylbut-2-enyl diphosphate synthase (GcpE) (46–48). This latter oxidoreductase is in the non-mevalonate pathway for isoprenoid biosynthesis (48). The tight association of cyanophage and cyanobacterial Fds having a wide range of protein-protein interactions suggested cyanophage Fds might be able to support ET to host oxidoreductases. In the case of pssm2-Fd, which is encoded by a myophage that infects *P. marinus*, it has the potential to interact with a total of eight host oxidoreductases (Figure 1C), in addition to the virally-encoded PebS (15, 27, 49).

pssm2-Fd contains sequence features of host Fds

To better understand how pssm2-Fd structure relates to cyanobacterial Fds, we analyzed the pairwise sequence identity of pssm2-Fd and cyanobacterial Fds in *P. marinus* hosts that are infected by cyanophage P-SSM2 (49). Fds from three host strains were collected by protein BLAST and designated pmN1-Fd1 (ecotype NATL1A), pmN2-Fd1 (NATL2A), and pm9211-Fd1 (MIT9211) (50). Like pssm2-Fd, these Fds reside within Cluster VI. We also evaluated the pairwise identity of pssm2-Fd with structurally-characterized Fds, including cyanobacterial (s6803-Fd1, nsp-Fd1, te-Fd1, and ml-Fd1) and plant (zm-Fd1) Fds. This analysis revealed most (5/7) cyanobacterial Fds exhibit ≥70% identity with each other (Figure 2A), while pssm2-Fd exhibits a range of sequence identities with these Fds (40-60%). Interestingly, pssm2-Fd presents slightly higher sequence identities with Fds from non-host cyanobacteria than from host cyanobacteria. The Fds from *P. marinus* ecotypes NATL1A and NATL2A (pmN1-Fd1 and pmN2-Fd1) exhibit the lowest identity with pssm2-Fd (43%) and appear to be the two most divergent Fds in the set. pm9211-Fd1 exhibits higher sequence identity with pssm2-Fd (52%) and appears more closely related to other cyanobacterial Fds (s6803-Fd1, te-Fd1, nspFd1, and ml-Fd1) than the other *P. marinus* ecotypes.

Fd binding specificities are influenced by charged surface residues and residues that make direct contacts with partner proteins (37, 38). To establish whether pssm2-Fd sequence is conserved at positions implicated in forming residue-residue contacts with partner proteins, a multiple sequence alignment was generated (Figure 2B). Residues implicated in contacting Fd partners from crystal structures of Fd-partner complexes are highlighted (37, 38, 46). In the case of the Fd-FNR complex (PDB = 1GAQ), all of the Fd residues that make contacts with FNR are completely conserved (38), suggesting a ubiquitous binding strategy to FNR across cyanobacteria. For the Fd-PSI complex (PDB = 5ZF0) and Fd-SIR (PDB = 5H92), greater variability is observed in the Fd residues that mediate partner contacts (37, 46). Only eight of the eleven Fd residues that contact PSI are either absolutely conserved or have residues with similar physicochemical properties. Similarly, four of the six Fd residues that contact SIR are either absolutely conserved or similar.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.07.937771; this version posted March 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

pssm2-Fd is a marginally stable metalloprotein

The tight association of pssm2-Fd with cyanobacterial Fds in Cluster VI suggested that these proteins may exhibit similar biophysical properties. To obtain recombinant pssm2-Fd for *in vitro* characterization, the gene encoding pssm2-Fd was overexpressed in *Escherichia coli* at 37°C and purified using a combination of anion exchange and size exclusion chromatography. This protocol yielded 0.6 mg/L of phage Fd, which was estimated to be >95% homogeneous by SDS-PAGE analysis. This yield is ~25-fold lower than that obtained with a thermophilic cyanobacterial Fd using an identical expression protocol (51).

To determine if pssm2-Fd contains a [2Fe-2S] cluster, we acquired pssm2-Fd absorbance and circular dichroism spectra (Figures 3A, 3B). The absorption spectrum contains peaks (465, 420, and 330 nm) that are characteristic of holoFds (52, 53). Additionally, the circular dichroism spectrum presented ellipticity maxima (427 and 360 nm) and minima (550 and 505 nm) that are observed with holoFds (52, 53). The ratio of the ellipticity at 427 nm to the absorbance at 280 was comparable to a cyanobacterial Fd that was recently reported (51), suggesting that a large fraction of pssm2-Fd contains a [2Fe-2S] cluster.

Because the yield of pssm2-Fd was lower than a thermophilic Fd that we recently purified (51), we postulated that pssm2-Fd might have a lower melting temperature (T_m). To test this idea, we evaluated how the ellipticity of pssm2-Fd at 427 nm changes with temperature (Figure 3C). Assuming two-state unfolding (54), the fraction of folded pssm2-Fd at each temperature was analyzed in a Van't Hoff plot. A linear fit of this plot yielded $\Delta H_{unfolding}$ and $\Delta S_{unfolding}$ (Figure S1), which were used to calculate T_m of 28°C.

To determine how the midpoint potential of pssm2-Fd relates to photosynthetic Fds, we characterized the electrochemical properties of this Fd using thin film, square wave voltammetry (55). With this analysis, pssm2-Fd presented a low midpoint potential (E_m = -336 mV) relative to a standard hydrogen electrode (Figure 3D), a value that is similar to other photosynthetic Fds.

Cyanophage Fd structural characterization

We determined the pssm2-Fd crystal structure at 1.6 Å resolution. The resulting model (R_{cryst}/R_{free} are 0.1783/0.2134) contains all 96 residues for each of the two monomers in the asymmetric unit, excluding the initiator methionines. The two chains of the asymmetric crystal align well, with a low root-mean-square deviation of all atomic coordinates (0.17 Å). The main source of variation between the chains arises from subtle differences within the beta sheets. While chain A in the asymmetric unit is well ordered, chain B lacks density for residues around the iron-sulfur cluster with a poor fit to the electron density resulting in a large real-space R-value. Therefore, non-crystallographic symmetry (NCS) restraints refinements were carried out using chain A as a reference for chain B. The apo-Fd is unstable and evidence in the literature indicates that zinc can interfere with and disrupt iron-sulfur clusters (56–59). In chain B, the iron-sulfur cluster site was disrupted. The density surrounding the cluster in chain B is consistent with a mixture of zinc bound and an intact [2Fe-2S] cluster. To account for the poor density and the anomalous data, chain B was modeled as a grouped occupancy, where the model is partially comprised of an intact cysteine-coordinated iron-sulfur cluster and a zinc ion with a hydration shell. An interpretation of these data is that as the cluster is disrupted, the surrounding residues become disordered. The crystallographic data collection and refinement statistics are provided in Table S1.

The structure of the pssm2-Fd Fd is shown in Figure 4A. With five β-sheet strands and three α-helices, pssm2-Fd possesses a canonical β-grasp topology, which consists of a β-sheet that is four to five strands long with α-helices interspersed and packed against the sheet (60). This topology is observed in other Fds, such as ml-Fd1 and s6803-Fd1 (41, 61). To quantify how similar the pssm2-Fd is to cyanobacterial Fds in Cluster VI (1OFF, 5AUI, 1CZP, 1RFK, 3B2F), we calculated the pairwise RMSD between these structures (41, 61–64). The pssm2-Fd structure presents low RMSD (≤0.5 Å) in all atom positions with the cyanobacterial Fds (Table S2), including two thermophilic Fds (46, 61) and a plant Fd (zm-Fd1).

We next mapped the sequence variation of Fds within Cluster VI onto the pssm2-Fd structure (Figure 4B). To quantify variation, we calculated the conservation scores at each pssm2-Fd residue using ConSurf (65). This analysis revealed that the largest density of absolutely conserved residues occurs in the loop surrounding the [2Fe-2S] cluster,

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.07.937771; this version posted March 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

with the next most conserved regions being those adjacent to this loop. There are additional residues that are highly conserved across all of the Fds analyzed. However, these are dispersed across the structure and primarily localized to regions that are packed within the center of the pssm2-Fd structure or forming salt-bridges on the surface.

Analysis of the intramolecular interactions within pssm2-Fd and those observed within cyanobacterial Fds found in Cluster VI revealed additional differences (Table S2). Among the proteins analyzed, pssm2-Fd presents a lower number of salt bridges (n = 3) compared with all other Fds analyzed (n = 4 to 6). However, pssm2-Fd contains a similar number of total hydrogen bonds (n = 133) as the other Fds (n = 105 to 154). Only one salt bridge was conserved in all of the Fds, which arises from an interaction between an arginine at position 41 or 42 and either a glutamate or aspartate situated between positions 27 and 30. This interaction may be conserved because it plays a role in stabilizing [2Fe-2S] cluster coordination, as the arginine is part of the cluster binding loop (Figure 2B).

pssm2-Fd supports ET to a host oxidoreductase

In a previous study, we examined whether pssm2-Fd could support ET between *Zea mays* FNR (zm-FNR) and SIR (zm-SIR) using a synthetic ET pathway in *E. coli* that is required for cellular growth (51). In this assay, Fd-mediated ET from FNR to SIR is required for sulfur assimilation and growth on medium containing sulfate as the only sulfur source (Figure 5A) (66). This prior study found that pssm2-Fd could not support ET from zm-FNR to zm-SIR (51). However, these experiments were performed at 37°C where only a small fraction of pssm2-Fd is folded.

To increase the fraction of holo-pssm2-Fd in the cellular assay, we built an anhydrotetracycline (aTc)-regulated expression vector that initiates translation of pssm2-Fd using a synthetic ribosome binding site (RBS) with strong translation initiation. When the assay was performed using this vector, aTc-dependent complementation was observed at 37°C (Figure 5B). This finding indicates that pssm2-Fd supports ET from zm-FNR to zm-SIR but requires high protein expression to compensate for the low stability of pssm2-Fd.

To test whether pssm2-Fd can also support ET to host SIRs, we created vectors that constitutively express *P. marinus* NATL2A SIR (pmN2-SIR) and MIT9211 SIR (pm9211-SIR) in parallel with zm-FNR. SIR from the NATL1A host was omitted because it shares high sequence identity (98%) with pmN2-SIR (Figure S2). Analysis of the zm-Fd1/zm-SIR complex structure revealed that these *P. marinus* SIRs are identical with zm-SIR at five of the seven residues that bind zm-Fd1 (Figure S3) (37). We also built aTc-inducible vectors to express host Fds (pmN1-Fd1, pmN2-Fd1, and pm9211-Fd1). Using these vectors, we tested the ability of cyanophage and host Fds to support ET from zm-FNR to zm-SIR and from zm-FNR to two host SIRs (pmN2-SIR and zm-9211-SIR). Because ml-Fd1 supports ET from zm-FNR to zm-SIR (51), it was included in the analysis as a positive control. Assays were performed at 23, 30, and 37°C ±aTc, which regulates Fd expression.

At the highest temperature assayed (37°C), only cells expressing zm-SIR complemented growth when coexpressed with a Fd and FNR (Figure 6A). This growth was observed with ml-Fd1, pssm2-Fd, and pm9211-Fd1. Cells expressing pmN2-Fd did not support ET from zm-FNR to zm-SIR. At 30°C, complementation was again observed with the same three-component ET pathways (Figure 6B). Additionally, cells expressing pm9211-SIR complemented growth when coexpressing the zm-FNR and either the cognate host Fd, pssm2-Fd, or ml-Fd1. At 23°C, similar trends were observed with zm-SIR and pm9211-SIR as at 30°C (Figure 6C). Additionally, significant complementation, albeit small, was observed when pmN2-SIR was coexpressed with the pssm2-Fd and ml-Fd1.

To investigate if differences in Fd-dependent growth arise because each Fd is expressed at different levels, we created expression vectors that produce these proteins as fusions to red fluorescence protein (RFP). We performed whole cell fluorescence measurements using cells transformed with these vectors and compared the signal to cells expressing ml-Fd1 alone (Figure 7). This analysis revealed that pmN2-Fd is expressed at higher levels than the other three Fds under the expression conditions used for the cellular assay. This finding suggests that the weaker complementation observed with pmN2-Fd does not arise because it is expressed at lower levels than the Fds that complement growth.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.07.937771; this version posted March 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To evaluate whether variability in structural properties correlates with cellular complementation, we evaluated the surface charge distribution of Fd and SIR structures (zm-Fd1: 3B2F, pssm2-Fd: 6VJV, zm-SIR: 5H8V) as well as structural models for Fds and SIRs lacking structures (pmN2-Fd1, pm9211-Fd1, pmN2-SIR, and pm9211-SIR) (37, 64). With the Fds that complement corn SIR, we observed similar charge distribution proximal to the iron-sulfur cluster (Figure 8A). Notably, the host Fd that did not support cellular ET to corn SIR (pmN2-Fd1) is the least electronegative Fd. When we analyzed the charge distribution of the SIRs, we observed similar surface charges, with slightly more electronegative character in cyanobacterial SIRs compared to zm-SIR (Figure 8B).

Discussion

Our bioinformatic analysis revealed 12 major clusters of cyanobacterial Fds, with many clusters having a narrow Fd length distribution. This analysis also showed that some clusters represent Fds which are widespread across cyanobacteria, while others contain Fds from a small number of organisms, illustrating how Fd gene duplications have diversified Fds in specific cyanobacterial genera. The extent to which the SSN clusters represent Fds with distinct ET roles is not known. In the case of the cyanobacterial Fds found in Cluster VI, which includes all cyanophage Fds analyzed, several of these proteins represent Fds expressed under phototrophic conditions (31, 67, 68). In some cases, this expression is likely due to their critical role as the electron acceptor for PSI (42, 46). Since Cluster VI contains cyanobacterial Fds that have evolved to support ET to a wide range of oxidoreductases, it seems probable that cyanophage Fds arose from the duplication of a cyanobacterial Fd that mediates ET to one or more of these partner oxidoreductases.

Some clues about the cellular roles of the cyanobacterial Fds in each SSN cluster can be gathered from studies in the model cyanobacterium *Synechocystis* sp. PCC6803. This organism contains a Fd (s6803-Fd1), within Cluster VI and also has five more [2Fe-2S] Fd paralogs that map onto Clusters V, VI, IX, and X (31). The dispersion of *Synechocystis* Fds into these different clusters with a wide range of primary structure lengths suggests that evolutionary pressures have diversified these Fds into multiple roles. The variable regulation of these Fds supports this notion (31, 67, 69). Aside from s6803-Fd1, transcripts of the other five [2Fe-2S] Fds change markedly under various stresses, with differences observed in response to salt stress, high concentrations of metals, low light conditions, and blue light conditions. Despite this knowledge, we know little about the *in vivo* function of cyanobacterial [2Fe-2S] Fds outside of Cluster VI. Further studies will be needed to understand how cyanobacterial Fds control the proportion of electrons that flow between different partner oxidoreductases.

Some additional unique features are observed in the SSN clusters. Clusters I and II consist of [2Fe-2S] mini-Fds that have a tight size distribution ($\mu = 79 \pm 2$, $\mu = 81 \pm 5$) and have primary structures that all terminate in tryptophan. This terminal residue is an unusual feature in comparison to canonical Fds (*e.g.*, Cluster VI) which generally end with a tyrosine. These "tryptophan-terminating" Fds are found in >40 species, although Cluster II consists wholly of the related genera *Prochlorococcus* and *Synechococcus* (10). Structures of these Fds have not yet been reported. The only structurally characterized [2Fe-2S] Fd with a terminal tryptophan is a cytochrome P450-reducing homolog from *Pseudomonas putida* (1PDX) but it is 106 amino acids long and much larger than the Fds within Cluster I and II (70). Cluster III is solely populated by Fds from *P. marinus* ecotypes, which have not shown to be infected by P-SSM2 cyanophage. Cluster V contains only Fds from *Synechocystis* strains. Of the 87 species with Fds in Cluster IX, 96% have two Fds within this cluster. The smaller set of Fds exhibits a tight size distribution ($\mu = 106 \pm 5$) as does the larger set ($\mu = 122 \pm 3$). Finally, Cluster X is populated primarily by Fds from *P. marinus* and *Synechococcus sp.* strains, although it also contains two species of *Cyanobium*. Future studies will be required to better understand how sequence variation within and across clusters relates to the evolution of partner specificity profiles across different cyanobacteria.

The cyanophage Fd characterized herein exhibits higher sequence identities with non-host cyanobacterial Fds than with host Fds. The underlying cause of this trend is not known. pssm2-Fd exhibits a similar overall structure to non-host cyanobacterial Fds (41, 61–63), presenting ≤0.5 Å

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.07.937771; this version posted March 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

RMSD with mesophilic (nsp-Fd1 and s6803-Fd1) and thermophilic (te-Fd1 and ml-Fd1) Fds. Additionally, pssm2-Fd exhibits a low midpoint potential that is similar to ml-Fd1 (E° = -336 mV) and higher than s6803-Fd1 (E° = -412 mV) and nsp-Fd1 (E° = -384 mV) (51, 71, 72). While structures and electrochemical properties of *P. marinus* Fds have not been established, it seems likely that they will be similar to characterized orthologs in other cyanobacteria that serve as PSI electron acceptors.

pssm2-Fd exhibits a low melting temperature (T_m = 28°C) compared with the other Fds in Cluster VI, which includes the thermophilic ml-Fd1 (T_m = 76°C) (73). This low thermostability may occur because *P. marinus* hosts grow in marine environments having temperatures ranging from 15 to 25°C (4). Additionally, this low T_m may have been selected to support more rapid turnover of cyanophage Fds or to regulate protein turnover in response to temperature changes. Currently, it is not known how Fds tune their stability without altering their midpoint potential. The native residues that control Fd stabilities (T_m = 28 to 76°C) could be established in future studies by creating chimeras between pssm2-Fd and more thermostable Fds with similar midpoint potentials (*e.g.*, ml-Fd1) and then evaluating how inheritance of different native residues relates to the thermostability of the resulting chimeras. This approach has previously been used to identify residues in other proteins that control thermostability (74, 75).

The growth selection used to monitor Fd ET illustrates how a cellular assay simplifies the analysis of Fd and partner oxidoreductase interactions (29). In previous studies, this cellular assay was used to evaluate if Fds from different organisms can mediate ET from zm-FNR to zm-SIR (51, 66). In addition, this approach has been used to select for engineered Fds that support cellular ET from FNR to SIR (51, 76, 77). The measurements described herein show how this assay can be used to rapidly assess whether phage-encoded Fds are capable of supporting ET to host-encoded oxidoreductases. The finding that pssm2-Fd supports ET to *P. marinus* SIRs indicates that this ET may occur during the cyanophage life cycle. However, the extent to which this reaction occurs in nature is not known, and it is unclear how phage Fds control the proportion of electrons transferred to host oxidoreductases like SIR versus phage-encoded oxidoreductases. One way to evaluate this competition in the future would be to examine how coexpression of viral PebS and host SIR in the cellular assay affects growth complementation.

How cyanophages influence both demand for and production of sulfur in their hosts will require further studies. Our finding that host SIRs can be reduced by cyanophage Fd supports the idea that P-SSM2 has evolved to supply reducing power to host sulfide production machinery during infection. However, it is unclear how the temporal expression of phage Fd changes the flux of nutrients during the cyclical lifecycle of cyanobacteria. Recent studies have revealed that cyanophage gene expression and replication is diurnally linked to host replication, with both processes occurring during the afternoon-evening period when light intensity is high (18, 78, 79). Additionally, phage transcripts drop at night and when photosystem is compromised (18). Despite these advances, we don't know what feedback loops govern expression of phage genes in infected cells, especially photosynthesis-related genes derived from host homologs.

The relationship of cyanophage infection cycles to the sulfur metabolism of marine photosynthetic cyanobacteria is not well established, although such interactions are likely critical to the health of oligotrophic oceanic gyres. Sulfate is highly abundant in these ecosystems, exhibiting concentrations as high as 28 mM (80), but reduced sulfur is more scarce, with dissolved dimethylsulfoniopropionate (DMSP) and methionine reaching concentrations of 0.5-10 nM and 10-40 pM, respectively, during the day (80–82). Reduced marine sulfur is in high demand, as DMSP is readily assimilated by gyre communities even when sulfate is $>10^6$-fold more abundant (80). Further, DMSP uptake is repressed by nM additions of more highly reduced compounds that are ostensibly easier to metabolize, such as methionine, cysteine, sulfide, and even dimethyl sulfoxide, indicating a hierarchy of preferred sulfur species among plankton (80, 82). Additionally, some heterotrophs from the Sar11 clade, which constitutes 30% of sea surface bacteria (83), are unable to metabolize sulfate and rely on dissolved reduced sulfur sources (84), potentially those produced by phototrophs, like *Prochlorococcus*. Conversely, recent work has shown that two ecotypes of *P. marinus* (MED4 and MIT9313) were unable to supply sufficient reduced sulfur to enable SAR11 growth in laboratory co-cultures (85).

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.07.937771; this version posted March 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

However, this study also demonstrated that only one of these strains was able to complement glycine auxotrophy in SAR11, suggesting that sulfide complementation may be observed in other *P. marinus* ecotypes or other species of cyanobacteria. The fact that environmental methionine levels peak during the day and drop below detection limits after sunset suggests that the diurnal cycle of cyanobacteria contributes to sulfur cycling in surface communities (82). Uncovering these relationships could be important for understanding nutrient availability across the Earth's oceans.

Experimental procedures

Materials

Tris was from Fisher Scientific, N-cyclohexyl-3-aminopropanesulfonic (CAPS) was from Acros Organics, 2-(N-Morpholino)ethanesulfonic acid (MES) and N-[Tris(hydroxymethyl)methyl]-3-aminopropanesulfonic acid (TAPS) were from Fluka Biochemika, and isopropyl β-D-1-thiogalactopyranoside (IPTG), dithiothreitol (DTT), kanamycin, chloramphenicol, and streptomycin were from Research Product International. 3-Morpholinopropane-1-sulfonic acid (MOPS), N-Cyclohexyl-2-aminoethanesulfonic acid (CHES), and all other chemicals were purchased from Sigma-Aldrich. *E. coli* EW11 was a gift from Pam Silver (Harvard University) (66), *E. coli* XL1-Blue was from Agilent, and Rosetta™(DE3) was from Novagen.

Vector Design

All plasmids are listed in Table S3. Genes were synthesized by Integrated DNA technologies or Twist Bioscience as G-blocks and Gene Fragments, respectively. Plasmids were constructed by ligating PCR products amplified with Q5 polymerase (New England Biolabs) using Golden Gate assembly (86). Translation initiation sites were designed using the RBS calculator (87). All plasmids were sequence verified.

Bioinformatics

To harvest viral Fd ORFs, all available viral genomes from Integrated Microbial Genomes and Microbiomes (n=8392) were scanned for genes encoding Fds with the Interpro signature IPR001041 (34). Viral genes with this IPR signature were compared with the 807 cyanobacterial [2Fe-2S] Fds listed in Dataset S1. To obtain pairwise sequence similarity scoring and generate a SSN, these Fds were analyzed using the Enzyme Function Initiative-Enzyme Similarity Tool (88). The resulting networks were analyzed and images were generated using Cytoscape with the prefuse force directed layout based on alignment score (89). Protein sequence alignments were generated using MUSCLE (90).

Protein purification

E. coli Rosetta transformed with pJTA022 were grown at 37˚C in lysogeny broth (LB) containing 50 µg/mL kanamycin to exponential phase, induced with 50 µM IPTG, and grown overnight at 37˚C while shaking at 250 rpm. Cells harvested by centrifugation (4k x *g*) were resuspended in 25 mL lysis buffer (per L of culture), which contained 10 mM Tris pH 8, 5 mM DTT, 10 mg/L DNase I, and 0.5 mg/mL lysozyme. After freezing at −80˚C, cells were thawed and mixed with a cOmplete Mini, EDTA-Free protease inhibitor tablet (Sigma-Aldrich) at a ratio of one tablet per 100 mL lysate. Samples were kept on ice or at 4˚C for all subsequent purification steps. Clarified lysate generated by centrifugation (47k x *g*) was diluted three-fold with TED buffer (25 mM Tris pH 8, 1 mM EDTA, and 1 mM DTT) to lower the salt concentration. This mixture was loaded onto a DE52 column (Whatman), the column was washed with TED containing 200 mM NaCl, and the Fd was eluted using sequential isocratic washes with TED containing 250 and 300 mM NaCl. Fractions appearing brown were mixed and diluted with TED to bring NaCl below 100 mM. These pooled fractions were loaded onto HiTrap Q XL column (GE Healthcare) using an AKTA Start FPLC system (GE Healthcare). The column was washed using 100 mL TED buffer. Protein was eluted using a linear gradient (0 to 375 mM NaCl in TED) followed by an isocratic wash (500 mM NaCl in TED). Brown fractions were pooled and then further purified using a HiLoad 16/600 Superdex 75 column containing TED. SDS-PAGE was performed to analyze purity at each step using NuPage 12% Bis-Tris Gels (Invitrogen) and the PageRuler Unstained Broad Range Protein Ladder (Thermo Scientific). Samples appearing homogeneous were pooled and concentrated using an Amicon Ultra 10 K MWCO spin column and then flash frozen with liquid nitrogen.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.07.937771; this version posted March 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Spectroscopy

Prior to all measurements, pssm2-Fd was dialyzed into TED. Absorbance spectra and ellipticity were acquired using a J-815 spectropolarimeter (Jasco, Inc.) using quartz cuvettes with a 1 cm path length. Scans used a 1 nm bandwidth, a 0.5 nm data pitch, and a 200 nm/min scan rate at 20°C. To assess stability, a cuvette with 50 µM pssm2-Fd was heated from 5 to 50°C at a rate of 1°C/min while monitoring ellipticity at 427 nm.

Electrochemistry

Electrochemical measurements were performed anaerobically using a three-electrode system. A Ag/AgCl/1M KCl electrode (CH Instruments) was used as the reference electrode, and a platinum wire was used as the counter electrode. An edge-plane pyrolytic graphite electrode was used as the working electrode to perform protein film electrochemistry. Prior to adding Fd, this electrode was treated with 100 mM neomycin trisulfate (Sigma-Aldrich) to improve the electrochemical signal (91). An 3 µL aliquot of pssm2-Fd Fd (620 µM) was then applied to the electrode surface, and the protein was allowed to adhere for 1 min at 23°C prior to placing in a glass vial containing a pH 7 buffer solution (5 mM acetate, MES, MOPS, TAPS, CHES, CAPS) and 100 mM NaCl at 23.5 °C. Square wave voltammograms were collected at 10 Hz frequency, and electrochemical signals were analyzed using QSoas open software. Similar results were obtained when experiments were performed using pssm2-Fd from different purifications. A CH Instruments potentiostat and CHI660E electrochemical analyzer were used for all measurements. All data is reported relative to Standard Hydrogen Electrode (SHE), taking into account the potential difference between SHE and Ag/AgCl/1M KCl, which is 0.222 V.

Crystallization

Purified pssm2-Fd (~20 mg/mL) was used for crystal screening with a mosquitoLCP pipetting robot (TTP Labtech, Boston, MA) and commercially available screens including Wizard Classic 1 and 2 and Wizard Classic 3 and 4 (Rigaku Reagents, Inc., Bainbridge Island, WA), MORPHEUS and MIDAS (Molecular Dimensions, Holland, OH), and PEGRx and IndexHT (Hampton Research, Alisa Viejo, CA). Protein crystals were grown via the sitting drop method of vapor diffusion at a 1:1 (v:v) ratio of protein:precipitant 200:200 nL, at 20 ºC. Crystals for diffraction and data collection were optimized and grown with a precipitant solution containing 10% (w/v) PEG 3000, 200 mM zinc acetate, and 100 mM sodium acetate/acetic acid pH 4.5. The crystals were cryo-protected with 10% glycerol, harvested using a MiTeGen micromesh loop (MiTeGen, Ithaca, NY), and flash-cooled in liquid nitrogen prior to data collection.

X-ray data collection, structure solution and refinement

Data were collected at the Advanced Photon Source GM/CA beamline 23ID-B using a 20 µm beam and 0.2 degree oscillations at a wavelength of 1.033 Å with an EIGER X 16M (DECTRIS Ltd, Philadelphia, PA) detector. The data were processed using the *autoPROC* toolbox (92), indexed and integrated using *XDS* (93), and scaled with *AIMLESS* (94). Initial phases were obtained by molecular replacement with *Phaser-MR* using the *Zea mays* Fd PDB ID: 5H57 (95) as a search model. Model building and refinement were performed with *Coot* (96) and *phenix.refine* (97). A custom graphics system was used for collaborative visualization (98). NCS restraints were used in initial refinements, with chain A as a reference for chain B, and removed in the final rounds of refinement. For TLS refinement, groups were automatically determined for chain A while chain B was treated as one group. Anomalous difference maps were calculated using *ANODE* and zinc from the precipitant solution was modeled in the structure (99). Data processing and refinement software were compiled and supported by the SBGrid Consortium (100). The structural model is available in the worldwide Protein Data Bank (PDB ID: 6VJV) (101). Data collection and refinement statistics are provided in Table S2.

Homology modeling

All-atom homology models were generated for pmN2-Fd1, pm9211-Fd1, pmN2-SIR, and pm9211-SIR using the fold and function assignment system (FFAS) server (102, 103). The Fe-S clusters for the Fd homology models were

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.07.937771; this version posted March 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

incorporated by aligning and copying them from our experimental model, PDB: 6VJV. The Fe-S clusters and sirohemes for the SIR homology models were incorporated by aligning the models to PDB: 5H8V and copying them.

Structural analysis

Root-mean-squared deviations were calculated, hydrogen bonds were detected (length cutoff: 3.5 Å), and structural images were generated using PyMOL (104). The illustration of conserved Fd residues mapped onto the pssm2-Fd structure was generated using the ConSurf Server with the WAG evolution model, seeded with all Fd sequences from Cluster VI (65). Electrostatic surface representations were calculated in PyMOL (version 2.3.2) using the APBS electrostatics plug-in with the prepwizard method default (Schrodinger) (104). All solvent molecules, ions, and ligands were removed, while the Fe-S clusters and sirohemes were used in the map calculation.

Cellular assay

To assess Fd ET from FNR to SIR, *E. coli* EW11 was transformed with two plasmids using electroporation as described (51, 66, 76, 77). One plasmid constitutively expresses FNR and SIR pairs, while the other expresses Fds under control of an anhydrotetracycline (aTc)-inducible promoter. To maintain both plasmids, all growth steps included chloramphenicol (34 µg/mL) and streptomycin (100 µg/mL). Starter cultures inoculated using single colonies were grown in deep-well 96-well plates for 18 h at 37 °C in 1 mL of a nonselective modified M9 medium (M9c), which contained sodium phosphate heptahydrate, dibasic (6.8 g/L), potassium phosphate, monobasic (3 g/L), sodium chloride (0.5 g/L), 2% glucose, ammonium chloride (1 g/L), calcium chloride (0.1 mM), magnesium sulfate (2 mM), ferric citrate (500 µM), p-aminobenzoic acid (2 mg/L), inositol (20 mg/L), adenine (5 mg/L), uracil (20 mg/L), tryptophan (40 mg/L), tyrosine (1.2 mg/L), and the remaining 18 amino acids (80 mg/L each). Starter cultures grown to stationary phase were then diluted 1:100 into a selective modified M9 medium (M9sa), which is identical to M9c but lacks cysteine and methionine. Cultures were inoculated in sterile Nunc™ Edge 2.0 96-well plates with side reservoirs filled with 1 mL of water. The edge between plate tops and bottoms were sealed with parafilm. Cells were grown in a Spark plate reader (Tecan) at 37°C while shaking (90 rpm) at an amplitude of 3 mm in double-orbital mode. Optical density (OD) was measured at 600 nm every 5 min for 48 h. For endpoint measurements, all conditions were the same except cells were grown in incubators at the indicated temperature while shaking at 250 rpm for 48 hours.

Fluorescence measurements

To evaluate the cellular expression of Fds, the genes encoding each Fd were expressed as fusions to a monomeric RFP with a twelve amino acid linker (105). Each vector was transformed into *E. coli* EW11 cells, alongside a plasmid (pSAC01) constitutively expressing *Zea mays* SIR (zm-SIR) and FNR (zm-FNR) (51). Transformed cells were grown under identical conditions to those described for the cellular assay, with the exception of being grown in M9c at all steps. After 24 hours, endpoint OD and RFP fluorescence ($\lambda_{excitation}$ = 560 nm; $\lambda_{emission}$ = 650 nm) were measured.

Statistics

Error bars represent standard deviation calculated from three or more biological replicates. Independent two-tailed t-tests were applied to compare differences between all relevant samples with α=0.05.

Data availability

The structure presented in this study is available in the PDB with the code 6VJV. All remaining data are contained within this article.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.07.937771; this version posted March 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Funding and additional information: This project was supported by DOE grant DE-SC0014462 (to J.J.S), NASA NAI grant 80NSSC18M0093 (to G.N.B and J.J.S), NSF grant 1231306 (to G.N.P), and Moore Foundation grant 7524 (to G.N.B and J.J.S.). GM/CA@APS has been funded in whole or in part with Federal funds from the National Cancer Institute (ACB-12002) and the National Institute of General Medical Sciences (AGM-12006). The Eiger 16M detector at GM/CA-XSD was funded by NIH grant S10 OD012289. This research used resources of the Advanced Photon Source, a DOE Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. I.J.C. and J.T.A. were supported by a Lodieska Stockbridge Vaughn Fellowship.

Conflict of interest: The authors declare that they have no conflicts of interest with the contents of this article.

Author contributions: I.J.C., J.T.A., G.N.B., and J.J.S. conceptualized the project. I.J.C. and J.L.O. performed bioinformatics. I.J.C., N.S., and J.T.A. purified recombinant protein. I.J.C. conducted spectroscopy. J.L.O., W.X., M.D.M., and G.N.P. crystallized pssm2-Fd and generated the structures. D.K. conducted electrochemistry. I.J.C. constructed vectors and performed cellular assays. I.J.C. and J.J.S. wrote the manuscript, and all other authors refined the text.

References

1. Flombaum, P., Gallegos, J. L., Gordillo, R. A., Rincon, J., Zabala, L. L., Jiao, N., Karl, D. M., Li, W. K. W., Lomas, M. W., Veneziano, D., Vera, C. S., Vrugt, J. A., and Martiny, A. C. (2013) Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc. Natl. Acad. Sci.* **110**, 9824–9829
2. Malmstrom, R. R., Coe, A., Kettler, G. C., Martiny, A. C., Frias-Lopez, J., Zinser, E. R., and Chisholm, S. W. (2010) Temporal dynamics of *Prochlorococcus* ecotypes in the Atlantic and Pacific oceans. *ISME J.* **4**, 1252–1264
3. Martiny, A. C., Tai, A. P. K., Veneziano, D., Primeau, F., and Chisholm, S. W. (2009) Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*. *Environ. Microbiol.* **11**, 823–832
4. Zinser, E. R., Johnson, Z. I., Coe, A., Karaca, E., Veneziano, D., and Chisholm, S. W. (2007) Influence of light and temperature on *Prochlorococcus* ecotype distributions in the Atlantic Ocean. *Limnol. Oceanogr.* **52**, 2205–2220
5. Bertilsson, S., Berglund, O., Pullin, M., and Chisholm, S. (2005) Release of dissolved organic matter by *Prochlorococcus*. *Vie Milieu*. **55**, 225–231
6. Dufresne, A., Garczarek, L., and Partensky, F. (2005) Accelerated evolution associated with genome reduction in a free-living prokaryote. *Genome Biol.* **6**, R14
7. Kettler, G. C., Martiny, A. C., Huang, K., Zucker, J., Coleman, M. L., Rodrigue, S., Chen, F., Lapidus, A., Ferriera, S., Johnson, J., Steglich, C., Church, G. M., Richardson, P., and Chisholm, S. W. (2007) Patterns and Implications of Gene Gain and Loss in the Evolution of *Prochlorococcus*. *PLoS Genet.* **3**, e231
8. Rocap, G., Larimer, F. W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N. A., Arellano, A., Coleman, M., Hauser, L., Hess, W. R., Johnson, Z. I., Land, M., Lindell, D., Post, A. F., Regala, W., Shah, M., Shaw, S. L., Steglich, C., Sullivan, M. B., Ting, C. S., Tolonen, A., Webb, E. A., Zinser, E. R., and Chisholm, S. W. (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature*. **424**, 1042–1047
9. Sun, Z., and Blanchard, J. L. (2014) Strong Genome-Wide Selection Early in the Evolution of *Prochlorococcus* Resulted in a Reduced Genome through the Loss of a Large Number of Small Effect Genes. *PLoS ONE*. **9**, e88837
10. Urbach, E., Scanlan, D. J., Distel, D. L., Waterbury, J. B., and Chisholm, S. W. (1998) Rapid Diversification of Marine Picophytoplankton with Dissimilar Light-Harvesting Structures Inferred from Sequences of *Prochlorococcus* and *Synechococcus* (Cyanobacteria). *J. Mol. Evol.* **46**, 188–201

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.07.937771; this version posted March 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

11. Chisholm, S. W., Frankel, S. L., Goericke, R., Olson, R. J., Palenik, B., Waterbury, J. B., West-Johnsrud, L., and Zettler, E. R. (1992) *Prochlorococcus marinus* nov. gen. nov. sp.: an oxyphototrophic marine prokaryote containing divinyl chlorophyll a and b. *Arch. Microbiol.* **157**, 297–300
12. Golden, S. S., Brusslan, J., and Haselkorn, R. (1986) Expression of a family of psbA genes encoding a photosystem II polypeptide in the cyanobacterium *Anacystis nidulans* R2. *EMBO J.* **5**, 2789–2798
13. Berube, P. M., Rasmussen, A., Braakman, R., Stepanauskas, R., and Chisholm, S. W. (2019) Emergence of trait variability through the lens of nitrogen assimilation in *Prochlorococcus*. *eLife*. **8**, e41043
14. Biller, S. J., Berube, P. M., Lindell, D., and Chisholm, S. W. (2015) *Prochlorococcus*: the structure and function of collective diversity. *Nat. Rev. Microbiol.* **13**, 13–27
15. Sullivan, M. B., Coleman, M. L., Weigele, P., Rohwer, F., and Chisholm, S. W. (2005) Three *Prochlorococcus* Cyanophage Genomes: Signature Features and Ecological Interpretations. *PLoS Biol.* **3**, e144
16. Labrie, S. J., Frois-Moniz, K., Osburne, M. S., Kelly, L., Roggensack, S. E., Sullivan, M. B., Gearin, G., Zeng, Q., Fitzgerald, M., Henn, M. R., and Chisholm, S. W. (2013) Genomes of marine cyanopodoviruses reveal multiple origins of diversity: Cyanopodovirus genomics. *Environ. Microbiol.* **15**, 1356–1376
17. Sieradzki, E. T., Ignacio-Espinoza, J. C., Needham, D. M., Fichot, E. B., and Fuhrman, J. A. (2019) Dynamic marine viral infections and major contribution to photosynthetic processes shown by spatiotemporal picoplankton metatranscriptomes. *Nat. Commun.* **10**, 1169
18. Liu, R., Liu, Y., Chen, Y., Zhan, Y., and Zeng, Q. (2019) Cyanobacterial viruses exhibit diurnal rhythms during infection. *Proc. Natl. Acad. Sci.* **116**, 14077–14082
19. Lindell, D., Jaffe, J. D., Coleman, M. L., Futschik, M. E., Axmann, I. M., Rector, T., Kettler, G., Sullivan, M. B., Steen, R., Hess, W. R., Church, G. M., and Chisholm, S. W. (2007) Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature*. **449**, 83–86
20. Puxty, R. J., Millard, A. D., Evans, D. J., and Scanlan, D. J. (2016) Viruses Inhibit CO_2 Fixation in the Most Abundant Phototrophs on Earth. *Curr. Biol.* **26**, 1585–1589
21. Puxty, R. J., Evans, D. J., Millard, A. D., and Scanlan, D. J. (2018) Energy limitation of cyanophage development: implications for marine carbon cycling. *ISME J.* **12**, 1273–1286
22. Lindell, D., Sullivan, M. B., Johnson, Z. I., Tolonen, A. C., Rohwer, F., and Chisholm, S. W. (2004) Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc. Natl. Acad. Sci.* **101**, 11013–11018
23. Clokie, M. R. J., Shan, J., Bailey, S., Jia, Y., Krisch, H. M., West, S., and Mann, N. H. (2006) Transcription of a "photosynthetic" T4-type phage during infection of a marine cyanobacterium. *Environ. Microbiol.* **8**, 827–835
24. Sharon, I., Tzahor, S., Williamson, S., Shmoish, M., Man-Aharonovich, D., Rusch, D. B., Yooseph, S., Zeidner, G., Golden, S. S., Mackey, S. R., Adir, N., Weingart, U., Horn, D., Venter, J. C., Mandel-Gutfreund, Y., and Béjà, O. (2007) Viral photosynthetic reaction center genes and transcripts in the marine environment. *ISME J.* **1**, 492–501
25. Sharon, I., Alperovitch, A., Rohwer, F., Haynes, M., Glaser, F., Atamna-Ismaeel, N., Pinter, R. Y., Partensky, F., Koonin, E. V., Wolf, Y. I., Nelson, N., and Béjà, O. (2009) Photosystem I gene cassettes are present in marine virus genomes. *Nature*. **461**, 258–262
26. Puxty, R. J., Millard, A. D., Evans, D. J., and Scanlan, D. J. (2015) Shedding new light on viral photosynthesis. *Photosynth. Res.* **126**, 71–97
27. Dammeyer, T., Bagby, S. C., Sullivan, M. B., Chisholm, S. W., and Frankenberg-Dinkel, N. (2008) Efficient Phage-Mediated Pigment Biosynthesis in Oceanic Cyanobacteria. *Curr. Biol.* **18**, 442–448
28. Ledermann, B., Schwan, M., Sommerkamp, J. A., Hofmann, E., Béjà, O., and Frankenberg-Dinkel, N. (2018) Evolution and molecular mechanism of four-electron reducing ferredoxin-dependent bilin reductases from oceanic phages. *FEBS J.* **285**, 339–356

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.07.937771; this version posted March 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

29. Atkinson, J. T., Campbell, I., Bennett, G. N., and Silberg, J. J. (2016) Cellular Assays for Ferredoxins: A Strategy for Understanding Electron Flow through Protein Carriers That Link Metabolic Pathways. *Biochemistry*. **55**, 7047–7064
30. Burkhart, B. W., Febvre, H. P., and Santangelo, T. J. (2019) Distinct Physiological Roles of the Three Ferredoxins Encoded in the Hyperthermophilic Archaeon *Thermococcus kodakarensis*. *mBio*. **10**, e02807-18
31. Cassier-Chauvat, C., and Chauvat, F. (2014) Function and Regulation of Ferredoxins in the Cyanobacterium, *Synechocystis* PCC6803: Recent Advances. *Life*. **4**, 666–680
32. Terauchi, A. M., Lu, S., Zaffagnini, M., Tappa, S., Hirasawa, M., Tripathy, J. N., Knaff, D. B., Farmer, P. J., Lemaire, D., Hase, T., and Merchant, S. S. (2009) Pattern of Expression and Substrate Specificity of Chloroplast Ferredoxins from *Chlamydomonas reinhardtii*. **284**, 25867–25878
33. Onda, Y., Matsumura, T., Kimata-Ariga, Y., Sakakibara, H., Sugiyama, T., and Hase, T. (2000) Differential interaction of maize root ferredoxin:NADP(+) oxidoreductase with photosynthetic and non-photosynthetic ferredoxin isoproteins. *Plant Physiol.* **123**, 1037–1045
34. Finn, R. D., Attwood, T. K., Babbitt, P. C., Bateman, A., Bork, P., Bridge, A. J., Chang, H.-Y., Dosztányi, Z., El-Gebali, S., Fraser, M., Gough, J., Haft, D., Holliday, G. L., Huang, H., Huang, X., Letunic, I., Lopez, R., Lu, S., Marchler-Bauer, A., Mi, H., Mistry, J., Natale, D. A., Necci, M., Nuka, G., Orengo, C. A., Park, Y., Pesseat, S., Piovesan, D., Potter, S. C., Rawlings, N. D., Redaschi, N., Richardson, L., Rivoire, C., Sangrador-Vegas, A., Sigrist, C., Sillitoe, I., Smithers, B., Squizzato, S., Sutton, G., Thanki, N., Thomas, P. D., Tosatto, S. C. E., Wu, C. H., Xenarios, I., Yeh, L.-S., Young, S.-Y., and Mitchell, A. L. (2017) InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Res.* **45**, D190–D199
35. Cornelissen, A., Hardies, S. C., Shaburova, O. V., Krylov, V. N., Mattheus, W., Kropinski, A. M., and Lavigne, R. (2012) Complete Genome Sequence of the Giant Virus OBP and Comparative Genome Analysis of the Diverse KZ-Related Phages. *J. Virol.* **86**, 1844–1852
36. Campbell, I. J., Bennett, G. N., and Silberg, J. J. (2019) Evolutionary Relationships Between Low Potential Ferredoxin and Flavodoxin Electron Carriers. *Front. Energy Res.* **7**, 79
37. Kim, J. Y., Nakayama, M., Toyota, H., Kurisu, G., and Hase, T. (2016) Structural and mutational studies of an electron transfer complex of maize sulfite reductase and ferredoxin. *J. Biochem. (Tokyo)*. **160**, 101–109
38. Kurisu, G., Kusunoki, M., Katoh, E., Yamazaki, T., Teshima, K., Onda, Y., Kimata-Ariga, Y., and Hase, T. (2001) Structure of the electron transfer complex between ferredoxin and ferredoxin-NADP(+) reductase. *Nat. Struct. Biol.* **8**, 117–21
39. Atkinson, H. J., Morris, J. H., Ferrin, T. E., and Babbitt, P. C. (2009) Using Sequence Similarity Networks for Visualization of Relationships Across Diverse Protein Superfamilies. *PLoS ONE*. **4**, e4345
40. Dai, S., Friemann, R., Glauser, D. A., Bourquin, F., Manieri, W., Schürmann, P., and Eklund, H. (2007) Structural snapshots along the reaction pathway of ferredoxin–thioredoxin reductase. *Nature*. **448**, 92–96
41. van den Heuvel, R. H. H., Svergun, D. I., Petoukhov, M. V., Coda, A., Curti, B., Ravasio, S., Vanoni, M. A., and Mattevi, A. (2003) The active conformation of glutamate synthase and its binding to ferredoxin. *J. Mol. Biol.* **330**, 113–128
42. Lelong, C., Boekema, E. J., Kruip, J., Bottin, H., Rögner, M., and Sétif, P. (1996) Characterization of a redox active cross-linked complex between cyanobacterial photosystem I and soluble ferredoxin. *EMBO J.* **15**, 2160–2168
43. Palma, P. N., Lagoutte, B., Krippahl, L., Moura, J. J. G., and Guerlesquin, F. (2005) *Synechocystis* ferredoxin/ferredoxin-NADP $^{+}$ -reductase/NADP $^{+}$ complex: Structural model obtained by NMR-restrained docking. *FEBS Lett.* **579**, 4585–4590
44. Hirasawa, M., Rubio, L. M., Griffin, J. L., Flores, E., Herrero, A., Li, J., Kim, S.-K., Hurley, J. K., Tollin, G., and Knaff, D. B. (2004) Complex formation between ferredoxin and *Synechococcus* ferredoxin:nitrate oxidoreductase. *Biochim. Biophys. Acta BBA - Bioenerg.* **1608**, 155–162

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.07.937771; this version posted March 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

45. Morales, R., Kachalova, G., Vellieux, F., Charon, M.-H., and Frey, M. (2000) Crystallographic studies of the interaction between the ferredoxin-NADP $^{+}$ reductase and ferredoxin from the cyanobacterium *Anabaena*: looking for the elusive ferredoxin molecule. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 1408–1412
46. Kubota-Kawai, H., Mutoh, R., Shinmura, K., Sétif, P., Nowaczyk, M. M., Rögner, M., Ikegami, T., Tanaka, H., and Kurisu, G. (2018) X-ray structure of an asymmetrical trimeric ferredoxin–photosystem I complex. *Nat. Plants*. **4**, 218–224
47. Okada, K. (2009) HO1 and PcyA proteins involved in phycobilin biosynthesis form a 1:2 complex with ferredoxin-1 required for photosynthesis. *FEBS Lett.* **583**, 1251–1256
48. Okada, K., and Hase, T. (2005) Cyanobacterial Non-mevalonate Pathway: (*E*)-4-hydroxy-3-methylbut-2-enyl diphosphate synthase interacts with ferredoxin in *Thermosynechococcus elongatus* BP-1. *J. Biol. Chem.* **280**, 20672–20679
49. Sullivan, M. B., Waterbury, J. B., and Chisholm, S. W. (2003) Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature*. **424**, 1047–1051
50. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410
51. Atkinson, J. T., Campbell, I. J., Thomas, E. E., Bonitatibus, S. C., Elliott, S. J., Bennett, G. N., and Silberg, J. J. (2019) Metalloprotein switches that display chemical-dependent electron transfer in cells. *Nat. Chem. Biol.* **15**, 189–195
52. Stephens, P. J., Thomson, A. J., Dunn, J. B. R., Keiderling, T. A., Rawlings, J., Rao, K. K., and Hall, D. O. (1978) Circular dichroism and magnetic circular dichroism of iron-sulfur proteins. *Biochemistry*. **17**, 4770–4778
53. Ta, D. T., and Vickery, L. E. (1992) Cloning, sequencing, and overexpression of a [2Fe-2S] ferredoxin gene from *Escherichia coli*. *J. Biol. Chem.* **267**, 11120–11125
54. Moczygemba, C., Guidry, J., Jones, K. L., Gomes, C. M., Teixeira, M., and Wittung-Stafshede, P. (2001) High stability of a ferredoxin from the hyperthermophilic archaeon *A. ambivalens*: Involvement of electrostatic interactions and cofactors. *Protein Sci.* **10**, 1539–1548
55. Osteryoung, J. G., and Osteryoung, R. A. (1985) Square wave voltammetry. *Anal. Chem.* **57**, 101–110
56. Jin, S., Kurtz, D. M., Liu, Z.-J., Rose, J., and Wang, B.-C. (2004) Displacement of iron by zinc at the diiron site of *Desulfovibrio vulgaris* rubrerythrin: X-ray crystal structure and anomalous scattering analysis. *J. Inorg. Biochem.* **98**, 786–796
57. Iametti, S., Uhlmann, H., Sala, N., Bernhardt, R., Ragg, E., and Bonomi, F. (1996) Reversible, Non-Denaturing Metal Substitution in Bovine Adrenodoxin and Spinach Ferredoxin and the Different Reactivities of [2Fe-2S]-Cluster-Containing Proteins. *Eur. J. Biochem.* **239**, 818–826
58. Kojoh, K., Matsuzawa, H., and Wakagi, T. (1999) Zinc and an N-terminal extra stretch of the ferredoxin from a thermoacidophilic archaeon stabilize the molecule at high temperature. *Eur. J. Biochem.* **264**, 85–91
59. Sazinsky, M. H., LeMoine, B., Orofino, M., Davydov, R., Bencze, K. Z., Stemmler, T. L., Hoffman, B. M., Argüello, J. M., and Rosenzweig, A. C. (2007) Characterization and Structure of a Zn^{2+} and [2Fe-2S]-containing Copper Chaperone from *Archaeoglobus fulgidus*. *J. Biol. Chem.* **282**, 25950–25959
60. Lo Conte, L. (2000) SCOP: a Structural Classification of Proteins database. *Nucleic Acids Res.* **28**, 257–259
61. Fish, A., Danieli, T., Ohad, I., Nechushtai, R., and Livnah, O. (2005) Structural Basis for the Thermostability of Ferredoxin from the Cyanobacterium *Mastigocladus laminosus*. *J. Mol. Biol.* **350**, 599–608
62. Morales, R., Charon, M. H., Hudry-Clergeon, G., Pétillot, Y., Norager, S., Medina, M., and Frey, M. (1999) Refined X-ray structures of the oxidized, at 1.3 A, and reduced, at 1.17 A, [2Fe-2S] ferredoxin from the cyanobacterium *Anabaena* PCC7119 show redox-linked conformational changes. *Biochemistry*. **38**, 15764–15773

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.07.937771; this version posted March 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

63. Mutoh, R., Muraki, N., Shinmura, K., Kubota-Kawai, H., Lee, Y.-H., Nowaczyk, M. M., Rögner, M., Hase, T., Ikegami, T., and Kurisu, G. (2015) X-ray Structure and Nuclear Magnetic Resonance Analysis of the Interaction Sites of the Ga-Substituted Cyanobacterial Ferredoxin. *Biochemistry*. **54**, 6052–6061
64. Sakakibara, Y., Kimura, H., Iwamura, A., Saitoh, T., Ikegami, T., Kurisu, G., and Hase, T. (2012) A new structural insight into differential interaction of cyanobacterial and plant ferredoxins with nitrite reductase as revealed by NMR and X-ray crystallographic studies. *J. Biochem. (Tokyo)*. **151**, 483–492
65. Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344-350
66. Barstow, B., Agapakis, C. M., Boyle, P. M., Grandl, G., Silver, P. A., and Wintermute, E. H. (2011) A synthetic system links FeFe-hydrogenases to essential *E. coli* sulfur metabolism. *J. Biol. Eng.* **5**, 7
67. Houot, L., Floutier, M., Marteyn, B., Michaut, M., Picciocchi, A., Legrain, P., Aude, J.-C., Cassier-Chauvat, C., and Chauvat, F. (2007) Cadmium triggers an integrated reprogramming of the metabolism of *Synechocystis* PCC6803, under the control of the Slr1738 regulator. *BMC Genomics*. **8**, 350
68. Kimata, Y., and Hase, T. (1989) Localization of Ferredoxin Isoproteins in Mesophyll and Bundle Sheath Cells in Maize Leaf. *Plant Physiol.* **89**, 1193–1197
69. Singh, A. K., Elvitigala, T., Cameron, J. C., Ghosh, B. K., Bhattacharyya-Pakrasi, M., and Pakrasi, H. B. (2010) Integrative analysis of large scale expression profiles reveals core transcriptional response and coordination between multiple cellular processes in a cyanobacterium. *BMC Syst. Biol.* **4**, 105
70. Pochapsky, T. C., Jain, N. U., Kuti, M., Lyons, T. A., and Heymont, J. (1999) A Refined Model for the Solution Structure of Oxidized Putidaredoxin[†,‡]. *Biochemistry*. **38**, 4681–4690
71. Bottin, H., and Lagoutte, B. (1992) Ferrodoxin and flavodoxin from the cyanobacterium *Synechocystis* sp. PCC 6803. *Biochim. Biophys. Acta BBA - Bioenerg.* **1101**, 48–56
72. Hurley, J. K., Weber-Main, A. M., Hodges, A. E., Stankovich, M. T., Benning, M. M., Holden, H. M., Cheng, H., Xia, B., Markley, J. L., Genzor, C., Gomez-Moreno, C., Hafezi, R., and Tollin, G. (1997) Iron-sulfur cluster cysteine-to-serine mutants of *Anabaena* -2Fe-2S- ferredoxin exhibit unexpected redox properties and are competent in electron transfer to ferredoxin:NADP+ reductase. *Biochemistry*. **36**, 15109–15117
73. Nechushtai, R., Lammert, H., Michaeli, D., Eisenberg-Domovich, Y., Zuris, J. A., Luca, M. A., Capraro, D. T., Fish, A., Shimshon, O., Roy, M., Schug, A., Whitford, P. C., Livnah, O., Onuchic, J. N., and Jennings, P. A. (2011) Allostery in the ferredoxin protein motif does not involve a conformational switch. *Proc. Natl. Acad. Sci.* **108**, 2240–2245
74. Chang, C.-J., Lee, C.-C., Chan, Y.-T., Trudeau, D. L., Wu, M.-H., Tsai, C.-H., Yu, S.-M., Ho, T.-H. D., Wang, A. H.-J., Hsiao, C.-D., Arnold, F. H., and Chao, Y.-C. (2016) Exploring the Mechanism Responsible for Cellulase Thermostability by Structure-Guided Recombination. *PLOS ONE*. **11**, e0147485
75. Otey, C. R., Landwehr, M., Endelman, J. B., Hiraga, K., Bloom, J. D., and Arnold, F. H. (2006) Structure-Guided Recombination Creates an Artificial Family of Cytochromes P450. *PLoS Biol.* **4**, e112
76. Mutter, A. C., Tyryshkin, A. M., Campbell, I. J., Poudel, S., Bennett, G. N., Silberg, J. J., Nanda, V., and Falkowski, P. G. (2019) De novo design of symmetric ferredoxins that shuttle electrons *in vivo*. *Proc. Natl. Acad. Sci.* **116**, 14557–14562
77. Wu, B., Atkinson, J. T., Kahanda, D., Bennett, G. N., and Silberg, J. J. (2020) Combinatorial design of chemical-dependent protein switches for controlling intracellular electron transfer. *AIChE J.* **66**, e16796

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.07.937771; this version posted March 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

78. Aylward, F. O., Boeuf, D., Mende, D. R., Wood-Charlson, E. M., Vislova, A., Eppley, J. M., Romano, A. E., and DeLong, E. F. (2017) Diel cycling and long-term persistence of viruses in the ocean's euphotic zone. *Proc. Natl. Acad. Sci.* **114**, 11446–11451
79. Yoshida, T., Nishimura, Y., Watai, H., Haruki, N., Morimoto, D., Kaneko, H., Honda, T., Yamamoto, K., Hingamp, P., Sako, Y., Goto, S., and Ogata, H. (2018) Locality and diel cycling of viral production revealed by a 24 h time course cross-omics analysis in a coastal region of Japan. *ISME J.* **12**, 1287–1295
80. Kiene, R. P., Linn, L. J., González, J., Moran, M. A., and Bruton, J. A. (1999) Dimethylsulfoniopropionate and methanethiol are important precursors of methionine and protein-sulfur in marine bacterioplankton. *Appl. Environ. Microbiol.* **65**, 4549–4558
81. Andreae, M. O. (1990) Ocean-atmosphere interactions in the global biogeochemical sulfur cycle. *Mar. Chem.* **30**, 1–29
82. del Valle, D., Martínez-García, S., Sañudo-Wilhelmy, S., Kiene, R., and Karl, D. (2015) Methionine and dimethylsulfoniopropionate as sources of sulfur to the microbial community of the North Pacific Subtropical Gyre. *Aquat. Microb. Ecol.* **75**, 103–116
83. Morris, R. M., Rappé, M. S., Connon, S. A., Vergin, K. L., Siebold, W. A., Carlson, C. A., and Giovannoni, S. J. (2002) SAR11 clade dominates ocean surface bacterioplankton communities. *Nature*. **420**, 806–810
84. Tripp, H. J., Kitner, J. B., Schwalbach, M. S., Dacey, J. W. H., Wilhelm, L. J., and Giovannoni, S. J. (2008) SAR11 marine bacteria require exogenous reduced sulphur for growth. *Nature*. **452**, 741–744
85. Becker, J. W., Hogle, S. L., Rosendo, K., and Chisholm, S. W. (2019) Co-culture and biogeography of *Prochlorococcus* and SAR11. *ISME J.* **13**, 1506–1519
86. Engler, C., Kandzia, R., and Marillonnet, S. (2008) A one pot, one step, precision cloning method with high throughput capability. *PloS One*. **3**, e3647
87. Salis, H. M. (2011) The ribosome binding site calculator. *Methods Enzymol.* **498**, 19–42
88. Gerlt, J. A. (2017) Genomic Enzymology: Web Tools for Leveraging Protein Family Sequence–Function Space and Genome Context to Discover Novel Functions. *Biochemistry*. **56**, 4293–4308
89. Shannon, P. (2003) Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **13**, 2498–2504
90. Edgar, R. C., Drive, R. M., and Valley, M. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797
91. Li, B., and Elliott, S. J. (2016) The Catalytic Bias of 2-Oxoacid:ferredoxin Oxidoreductase in CO_2: evolution and reduction through a ferredoxin-mediated electrocatalytic assay. *Electrochimica Acta*. **199**, 349–356
92. Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T., and Bricogne, G. (2011) Data processing and analysis with the *autoPROC* toolbox. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 293–302
93. Kabsch, W. (2010) *XDS*. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132
94. Evans, P. R., and Murshudov, G. N. (2013) How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214
95. Shinohara, F., Kurisu, G., Hanke, G., Bowsher, C., Hase, T., and Kimata-Ariga, Y. (2017) Structural basis for the isotype-specific interactions of ferredoxin and ferredoxin: NADP+ oxidoreductase: an evolutionary switch between photosynthetic and heterotrophic assimilation. *Photosynth. Res.* **134**, 281–289
96. Emsley, P., and Cowtan, K. (2004) *Coot*: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
97. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with *phenix.refine*. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.07.937771; this version posted March 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

98. Yennamalli, R., Arangarasan, R., Bryden, A., Gleicher, M., and Phillips, G. N. (2014) Using a commodity high-definition television for collaborative structural biology. *J. Appl. Crystallogr.* **47**, 1153–1157
99. Thorn, A., and Sheldrick, G. M. (2011) *ANODE*: anomalous and heavy-atom density calculation. *J. Appl. Crystallogr.* **44**, 1285–1287
100. Morin, A., Eisenbraun, B., Key, J., Sanschagrin, P. C., Timony, M. A., Ottaviano, M., and Sliz, P. (2013) Collaboration gets the most out of software. *eLife*. **2**, e01456
101. Berman, H., Henrick, K., and Nakamura, H. (2003) Announcing the worldwide Protein Data Bank. *Nat. Struct. Mol. Biol.* **10**, 980–980
102. Xu, D., Jaroszewski, L., Li, Z., and Godzik, A. (2014) FFAS-3D: improving fold recognition by including optimized structural features and template re-ranking. *Bioinformatics*. **30**, 660–667
103. Canutescu, A. A., Shelenkov, A. A., and Dunbrack, R. L. (2003) A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci.* **12**, 2001–2014
104. Schrödinger, LLC (2015) The PyMOL Molecular Graphics System, Version 1.8
105. Campbell, R. E., Tour, O., Palmer, A. E., Steinbach, P. A., Baird, G. S., Zacharias, D. A., and Tsien, R. Y. (2002) A monomeric red fluorescent protein. *Proc. Natl. Acad. Sci.* **99**, 7877–7882

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.07.937771; this version posted March 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ABBREVIATONS

aTc	anhydrotetracycline
DMSP	dimethylsulfoniopropionate
DTT	dithiothreitol
ET	electron transfer
Fd	ferredoxin
FNR	Fd-NADP reductase
FTR	Fd-thioredoxin reductase
GcpE	4-Hydroxy-3-methylbut-2-enyl diphosphate synthase
GS	glutamate synthase
HO	heme oxygenase
IPTG	isopropyl β-D-1-thiogalactopyranoside
LB	lysogeny broth
ml-Fd1	*Mastigocladus laminosus* Fd 1
NIR	nitrate reductase
nsp-Fd1	*Nostoc sp.* PCC 7119 Fd 1
OD	optical density
ORFs	open reading frames
PcyA	phycocyanobilin:ferredoxin oxidoreductase
PebS	phycoerythrobilin synthase
pm9211-Fd1	*Prochlorococcus marinus* MIT9211 Fd 1
pm9211-SIR	*Prochlorococcus marinus* MIT9211 SIR

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.07.937771; this version posted March 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

pmN1-Fd1	*Prochlorococcus marinus* NATL1A Fd 1
pmN2-Fd1	*Prochlorococcus marinus* NATL2A Fd 1
pmN2-SIR	*Prochlorococcus marinus* NATL2A SIR
PSI	photosystem I
pssm2-Fd	myovirus P-SSM2 Fd
RFP	red fluorescent protein
s6803-Fd1	*Synechocystis sp.* PCC 6803 Fd 1
SHE	standard hydrogen electrode
SIR	sulfite reductase
SSN	sequence similarity network
te-Fd1	*Thermosynechococcus elongatus* Fd 1
zm-Fd1	*Zea mays* Fd 1
zm-FNR	*Zea mays* FNR
zm-SIR	*Zea mays* SIR

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.07.937771; this version posted March 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

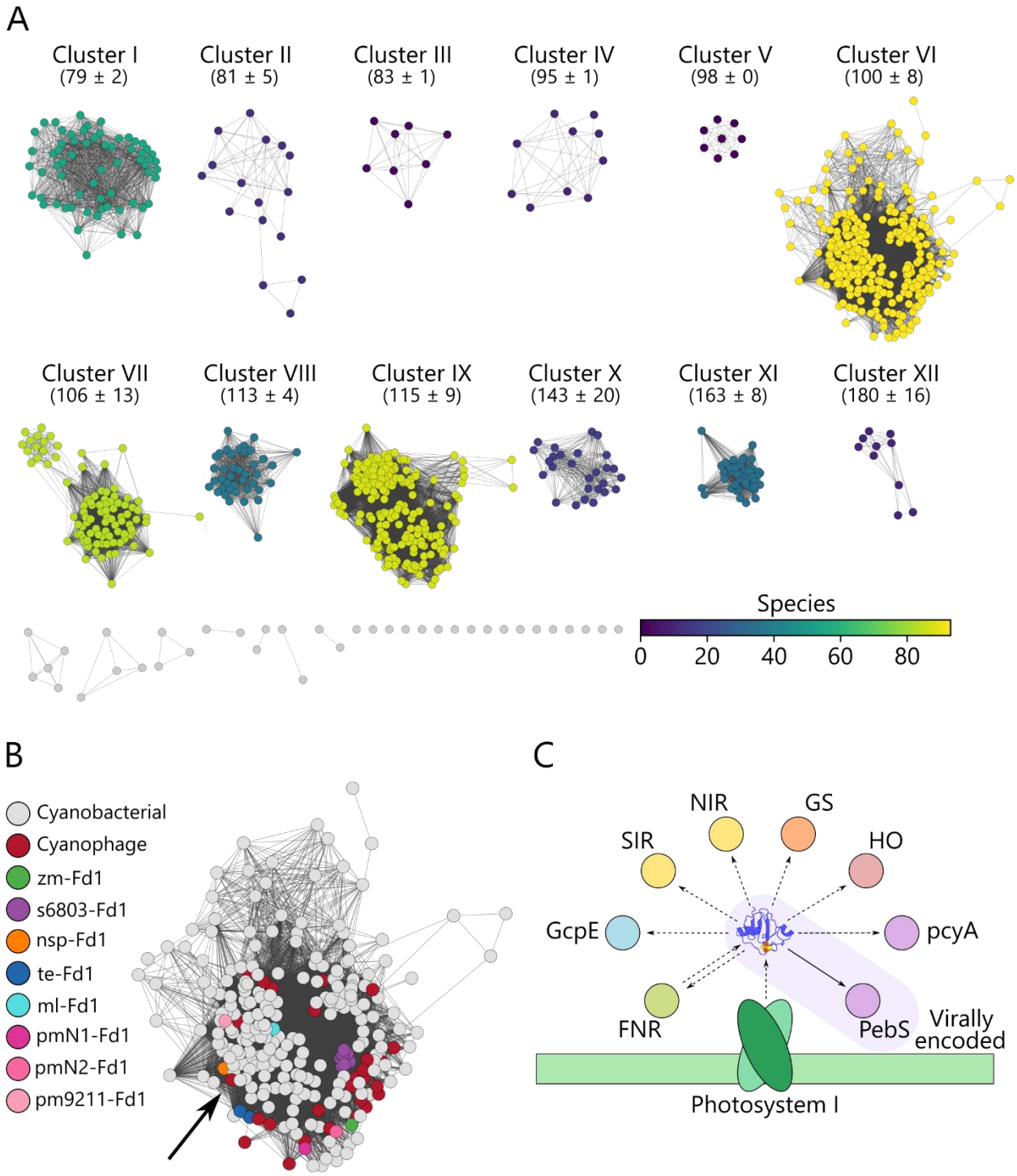

Figure 1. Network analysis of cyanophage and cyanobacterial Fds. (**A**) A network illustrating the relationships of 807 cyanobacterial and 28 cyanophage Fds. Edges are drawn between Fd pairs that exhibit a sequence alignment score ≥30. For the largest clusters (I to XII), the average protein lengths encoded by the ORFs in the cluster are provided ±1σ. Clusters are colored using a viridis gradient by the number of species represented in each cluster. (**B**) Cyanophage and cyanobacterial Fds are highlighted within Cluster VI with zm-Fd1 as a frame of reference. The arrow indicates pssm2-Fd. Fds from multiple species and strains share the same sequence as s6803-Fd1 and te-Fd1, so they are all colored identically. (**C**) *P. marinus* partner oxidoreductases that have the potential to interact with pssm2-Fd during an infection.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.07.937771; this version posted March 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

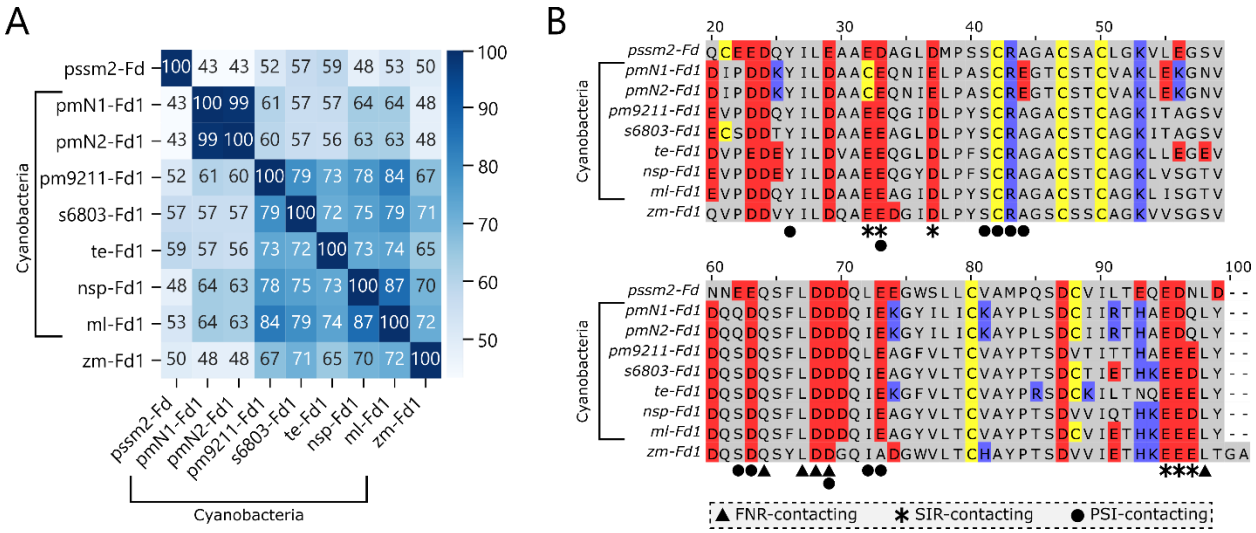

Figure 2. Cyanophage and host cyanobacterial Fd sequence comparisons. (**A**) A matrix showing the pairwise sequence identity of pssm2-Fd and Fds from host cyanobacteria, cyanobacteria with known structures, and one plant family member (zm-Fd1). The blue shading indicates the percentage sequence identity. (**B**) A multiple sequence alignment illustrates the conservation of cysteines (yellow), positively charged residues (blue), negatively charged residues (red), and residues implicated in binding FNR, SIR, and PSI (37, 38, 46). Residues are indexed with respect to ml-Fd1.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.07.937771; this version posted March 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

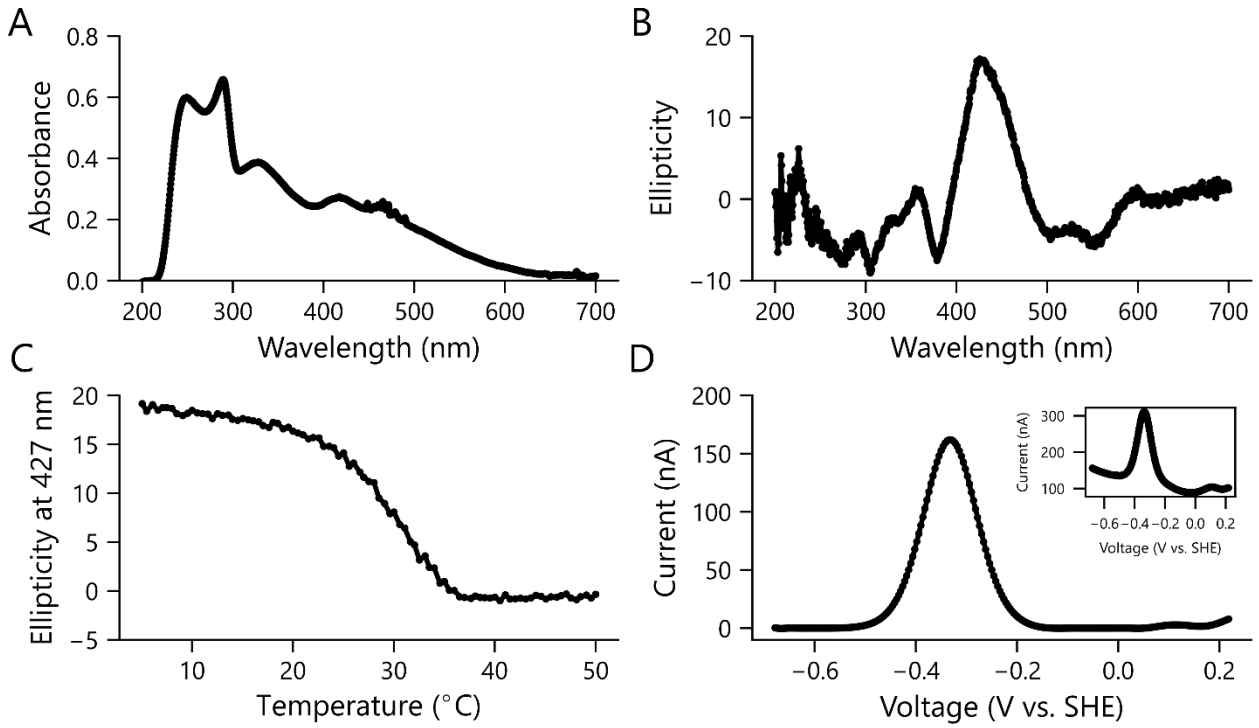

Figure 3. Biophysical properties of purified pssm2-Fd. (**A**) The absorbance and (**B**) circular dichroism spectra of recombinant purified pssm2-Fd (50 µM) reveal features characteristic of [2Fe-2S] Fds that support photosynthesis, such as zm-Fd1. (**C**). Effect of temperature on the ellipticity of pssm2-Fd (50 µM) at 427 nm. The midpoint of the transition occurs at 29°C. (**D**) Square wave voltammetry of pssm2-Fd (620 µM) reveals a midpoint potential of -334 mV. Voltammetry prior to background subtraction is shown as an inset.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.07.937771; this version posted March 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

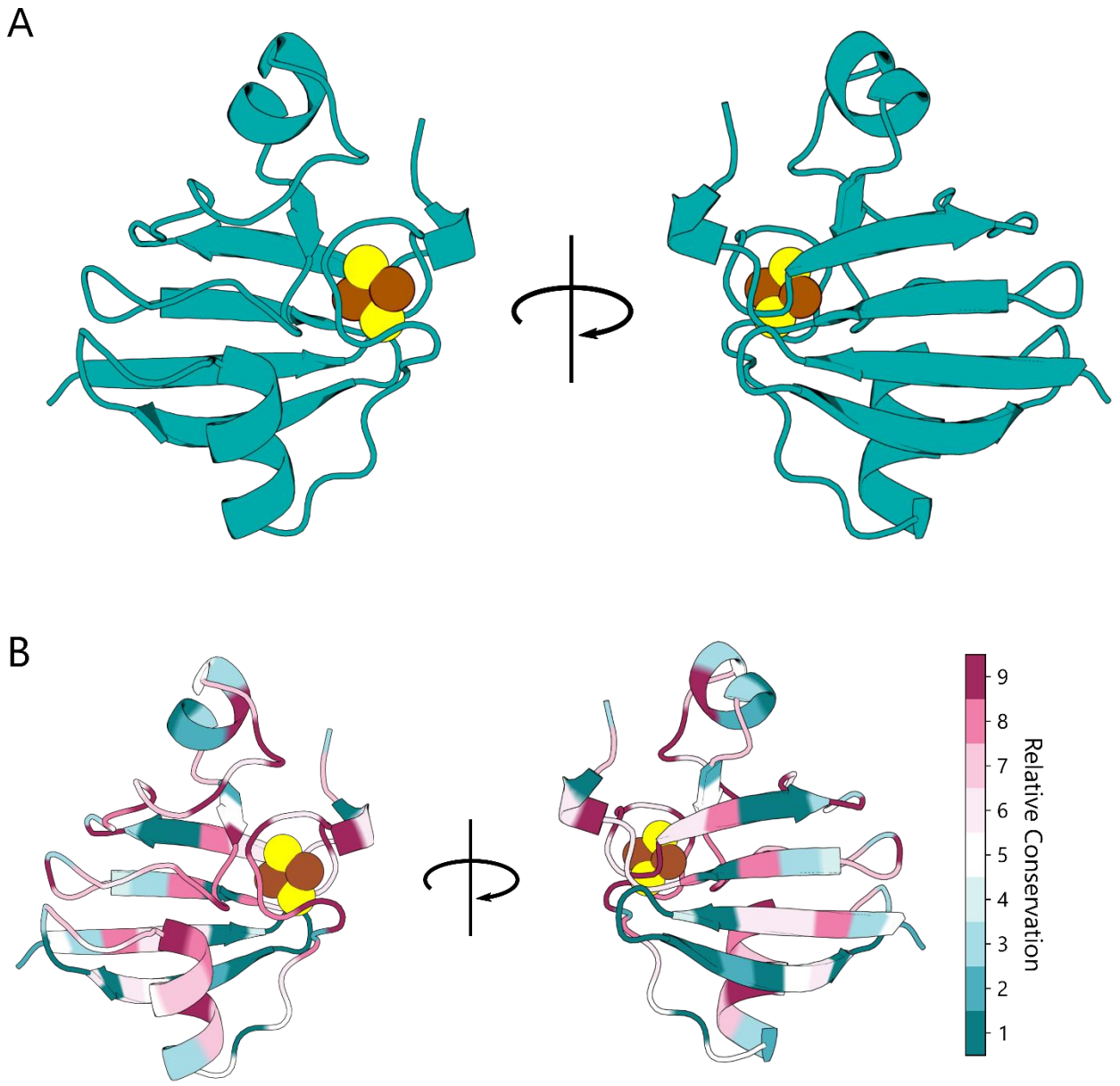

Figure 4. The phage Fd structure. (**A**) Front- and back-facing ribbon structures of the pssm2-Fd structure. Iron and sulfur atoms are shown as brown and yellow spheres, respectively. (**B**) Ribbon structures of pssm2-Fd colored with ConSurf conservation scores generated using the Fds in Cluster VI. The most conserved residues are shown in magenta, while the least conserved residues are in cyan.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.07.937771; this version posted March 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

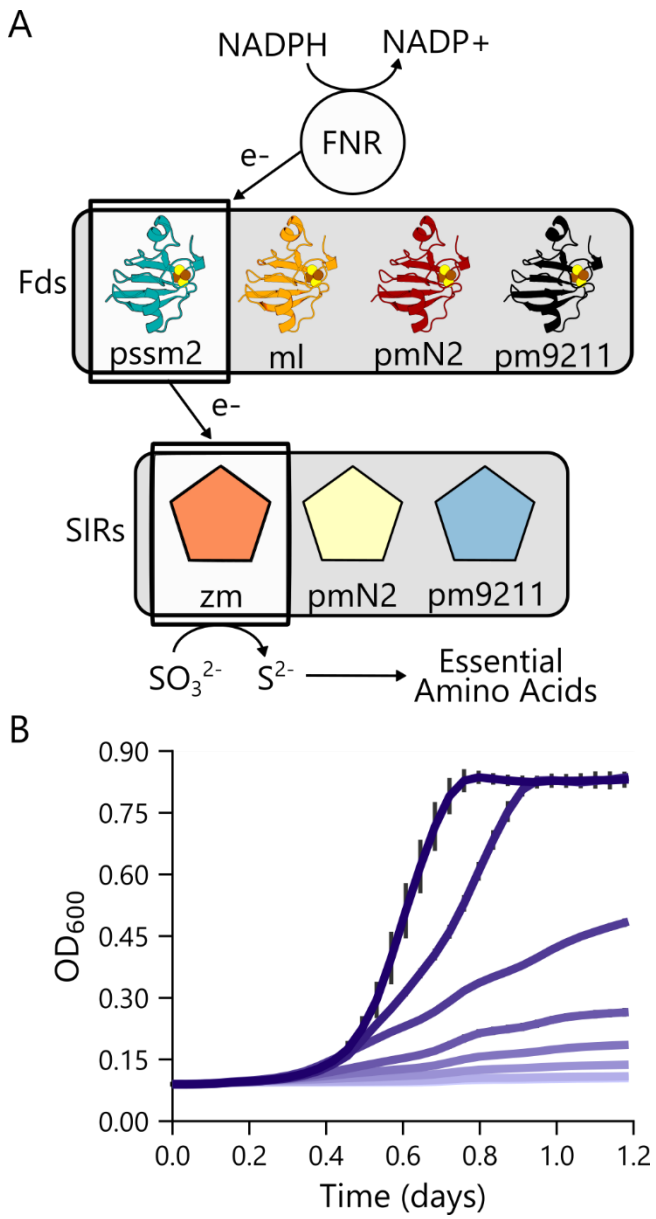

Figure 5. Monitoring pssm2-Fd ET using a cellular assay. (**A**) An *E. coli* auxotroph unable to produce sulfide can grow on sulfite as a sulfur source when cells express a three-component pathway consisting of FNR, Fd, and SIR. In this study, four different Fds and three different SIRs were examined. (**B**) The effect of aTc concentration on the growth complementation of *E. coli* EW11 transformed with vectors that express pssm2-Fd using an aTc-inducible promoter and zm-FNR/zm-SIR using constitutive promoters. The aTc concentrations (0, 1.6, 3.1, 6.3, 12.5, 25, 50, 100 ng/mL) are color coded as a gradient from light purple (no aTc) to dark purple (highest aTc). Average values from three biological replicates are shown with error bars representing ±1σ.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.07.937771; this version posted March 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

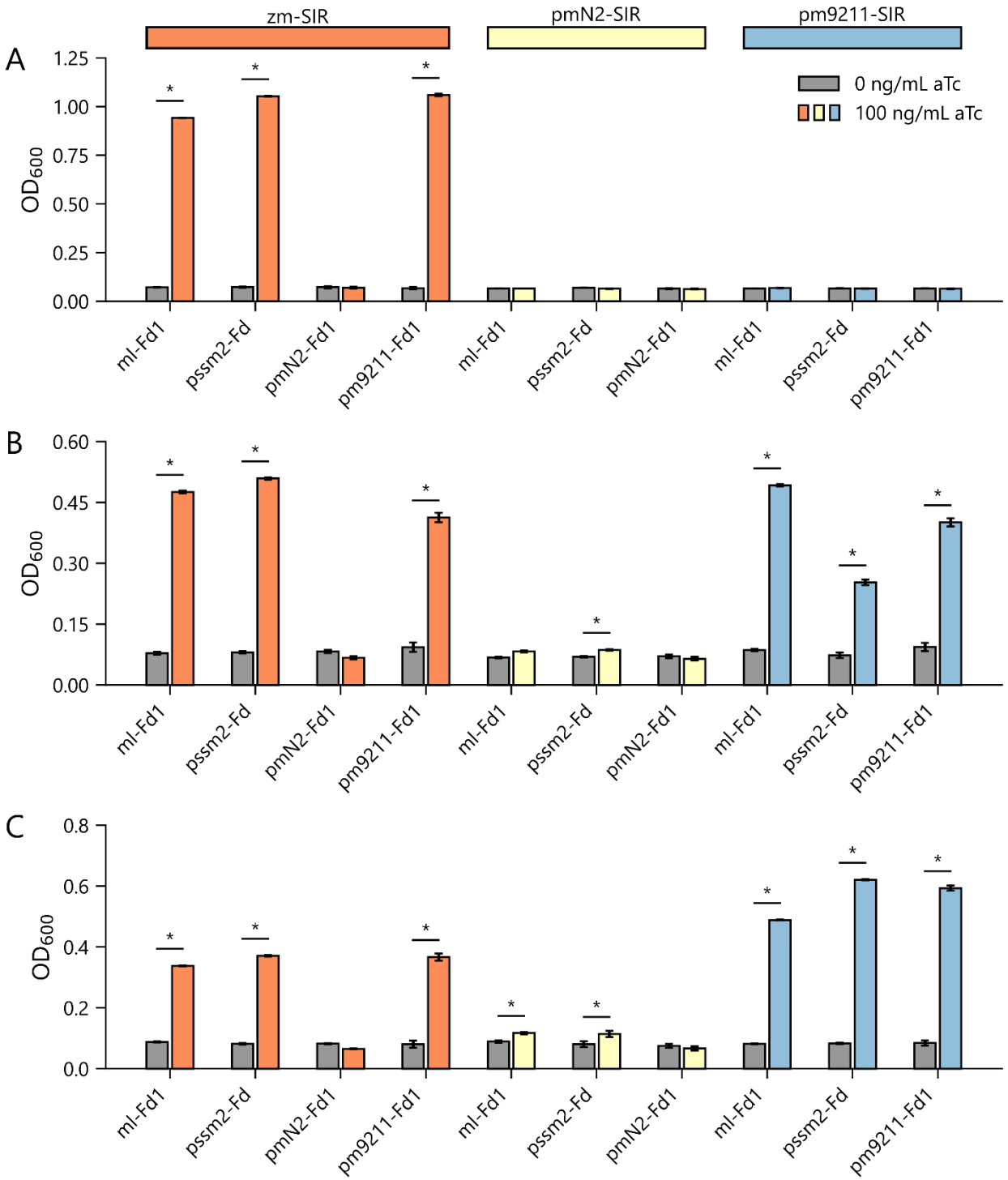

Figure 6. pssm2-Fd supports ET to host SIRs. Growth complementation observed at (**A**) 37°C, (**B**) 30°C, and (**C**) 23°C with *E. coli* EW11 transformed with vectors that constitutively expresses zm-FNR with three different SIRs (zm-SIR, pmN2-SIR, pm9211-SIR) and a vector that uses an aTc-inducible promoter to express different Fds (ml-Fd1, pssm2-Fd, pmN2-Fd1, pm9211-Fd1). Experiments were performed in the absence (grey bars) and presence (colored bars) of 100 ng/mL aTc to evaluate the dependence of growth upon Fd expression. Error bars represent ±1σ from three biological replicates. For each strain, an independent two-tailed t-test was used to compare the OD in the absence and presence of aTc (α=0.05). Significant differences are noted with asterisks.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.07.937771; this version posted March 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

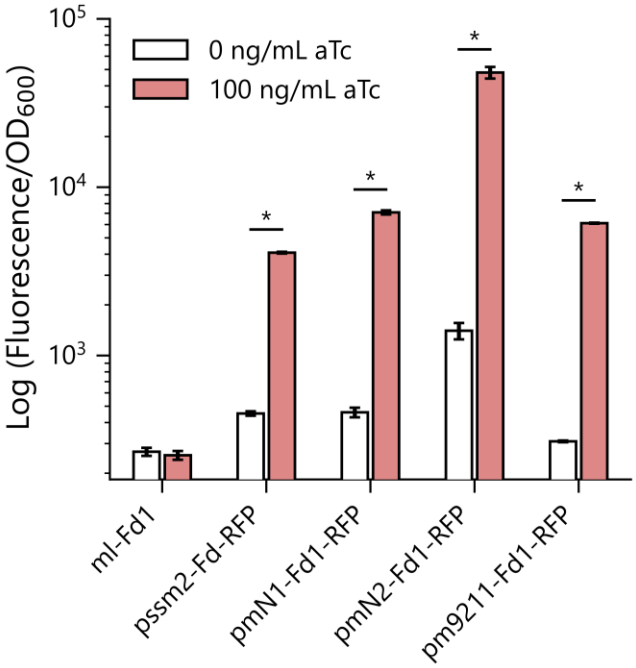

Figure 7. Expression of Fd-RFP fusion proteins. Fluorescence was measured in *E. coli* EW11 transformed with vectors that express Fds using an aTc-inducible promoter. Cells were grown in non-selective conditions. For each strain, an independent two-tailed t-test was used to compare fluorescence normalized to optical density (OD_{600}) ±aTc (α=0.05). Error bars represent ±1σ from three biological replicates. Significant differences are noted with asterisks.

bioRxiv preprint doi: https://doi.org/10.1101/2020.02.07.937771; this version posted March 19, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

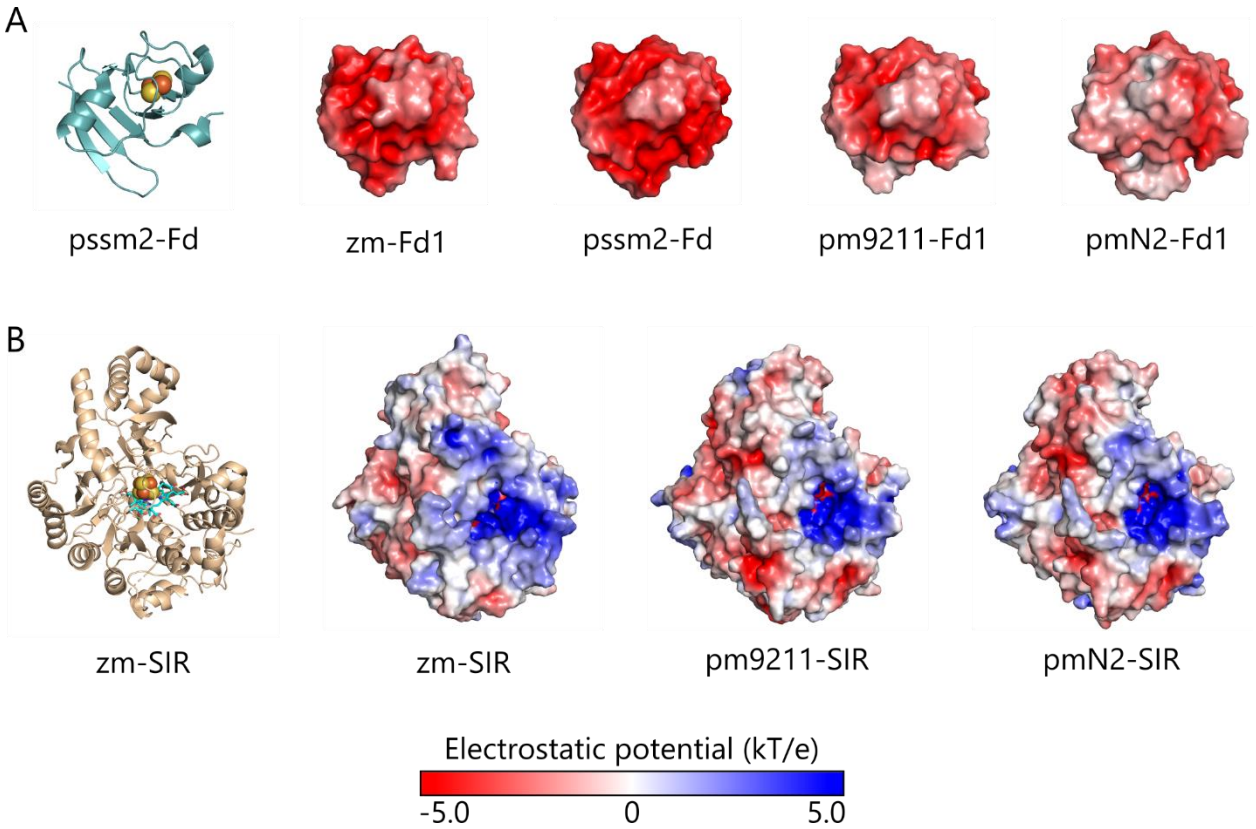

Figure 8. A comparison of the electrostatic potential surfaces of Fds and SIRs. The surfaces of (**A**) Fd and (**B**) SIR homologs are colored according to electrostatic potential, ranging from blue to red (kT/e), where kT is thermal energy, and e is the elementary charge. zm-Fd1, pssm2-Fd, and zm-SIR were generated using crystal structure data, while all other structures represent homology models (37, 64). The front faces of the Fds are negatively charged to varying degrees, while the SIRs display positive binding surfaces with small variations. These oppositely charged patches interact during binding, bringing the metallocofactors from each protein into close proximity to facilitate ET (37). Iron-sulfur clusters are displayed as spheres and sirohemes as sticks.