Cyanobacteria Phage S-RSM4 Optimization of Pentose Phosphate Pathway via Glucose 6-phosphate Dehydrogenase Homologous to Synechococcus Host

Zachary J. Babbitz

Follow this and additional works at: https://scholarworks.uvm.edu/hcoltheses

Recommended Citation
https://scholarworks.uvm.edu/hcoltheses/382

This Honors College Thesis is brought to you for free and open access by the Undergraduate Theses at UVM ScholarWorks. It has been accepted for inclusion in UVM Honors College Senior Theses by an authorized administrator of UVM ScholarWorks. For more information, please contact scholarworks@uvm.edu.
Cyanobacteria Phage S-RSM4 Optimization of Pentose Phosphate Pathway via Glucose 6-phosphate Dehydrogenase Homologous to Synechococcus Host

Cyanobacteria are essential to the biosphere for their primary production of oxygen and their fixation of carbon and nitrogen. The evolution of cyanobacteria in the ocean is controlled by infectious cyanophages. The cyanophage S-RSM4 carries a copy of glucose 6-phosphate dehydrogenase (G6PDH), a photosynthetic enzyme homologous to that of its host, *Synechococcus sp. W8103*. Both G6PDH homologs were cloned and expressed. The viral enzyme was purified, and its kinetic and mechanistic properties were investigated. It is likely that S-RSM4 encodes this enzyme to continue production of ribose 5-phosphate and NADPH during infection, independent of the host’s homeostatic constraints. A continued effort to understand this unique parasitism is pertinent to understanding the ocean’s microbiome as a whole.
Cyanobacteria are a taxon of photosynthetic bacteria which populate the oceans worldwide. Cyanophages are viruses which infect cyanobacteria. Infection by these viruses controls the dynamics and evolution of cyanobacteria in the oceans\(^1\). Being the only diazotrophs which release oxygen as a byproduct of their photosynthesis, cyanobacteria are a keystone species for nitrogen, carbon, and oxygen cycles of the ocean\(^2\). Cyanobacteria are responsible for >25% of the atmospheric oxygen the biosphere depends upon\(^3\). In fact, it is likely that the majority of Earth’s atmospheric oxygen was produced by cyanobacteria-like organisms during the late Proterozoic Era\(^4\). The two genera, *Prochlorococcus* and *Synechococcus*, contribute up to 90% of the primary nutrient production in certain nutrient deficient regions of the ocean\(^5\). The nine species\(^6\) of saltwater cyanobacteria from the genus *Trichodesmium* are estimated to be responsible for nearly half of all oceanic nitrogen fixation\(^7\).

There are 2,698 documented species of cyanobacteria and discovery curves predict there are actually 6,280 species\(^8\). The contribution of bioavailable nitrogen attributed to all cyanobacteria is not precisely known, but it is certainly staggering. Furthermore, cyanophage infection causes the lysis of cyanobacterial cells, accelerating the release of nutrients into the ocean.

Understanding the parasitosis of cyanophages on cyanobacterial hosts gives insight into the biological control of Earth’s oceans and atmosphere.

This study addressed how, specifically, T4-like cyanophages overtake the metabolism of cyanobacteria to promote their own replication. The T4-like cyanophage S-RSM4 and its host *Synechococcus sp. WH8103* were used as a model system to study how phage-encoded enzymes might direct the pentose-phosphate pathway to optimize ribose 5-phosphate and NADPH production, supporting phage DNA replication at the expense of host energy.
Since cyanobacteria photosynthesize, their metabolism is complex and highly regulated. Viruses have essentially no metabolism of their own; they depend upon the host for all biosynthesis. One would expect a phage to only carry genes necessary for infection and rely on the host’s metabolism for chemical energy and biosynthesis, but this is not true of S-RSM4.

The transfer of energy from photosynthesis overlaps with the nucleotide synthesis necessary for phage reproduction at the pentose-phosphate pathway (PPP). Three PPP enzymes: glucose 6-phosphate dehydrogenase (G6PDH), 6-phosphogluconate dehydrogenase, and trans-aldolase are found as homologs in both the virus and the host. Why would the virus carry redundant genes? The null hypothesis is that phages carry extra enzymes to increase the enzyme concentration in the cell during infection, boosting overall cellular metabolism and thus phage replication. Yet it would be even more interesting if the phage’s strategy is more “intelligent”. Investigation into these homologous genes starts with G6PDH, the rate-limiting enzyme of the oxidative phase of the pentose-phosphate pathway.

The host’s enzymes are regulated to maintain homeostasis. Since the phage does not gain fitness from the host maintaining homeostasis, we propose that the phage enzymes are regulated differently than those in the host’s system. Their regulatory mechanisms, substrate specificity, and turnover rates could all be optimized for phage replication at the expense of host homeostasis. Genomic and biochemical background about S-RSM4 and *Synechococcus* sp. *W8103* helped guide the hypotheses about these potential regulatory differences.

Two viruses which infect cyanobacteria of the genus *Synechococcus* are S-PM2 and S-RSM4. These viruses stand out among viruses like P-SMM2 and P-SMM4, which infect the genus *Prochlorococcus*, because they have up to 25 times as many tRNA sequences encoded in their
genome. The S-RSM4 virus encodes 25 tRNA molecules, vaguely suggesting that this virus has a more complex, self-sufficient system of survival and reproduction than similar strains. T4-like cyanophages lack certain proteins found in the true T4 genome, proteins which modify RNA polymerase specificity in favor of viral replication. It is predicted that most T4-like cyanophages utilize an unmodified host RNA polymerase, since most of them contain essential promoters for a translational sigma factor. This promoter is lacking in the S-RSM4 genome, as is a middle-transcription promoter commonly found in T4-like cyanophages. This lack of communication with the host replication system suggests replication during S-RSM4 infection is unique, and possibly more self-sufficient than other cyanophages.

S-RSM4 has the highest number of photosynthetic genes of any cyanophage sequenced to date. It is tied with Syn9 for the most carbon/phosphate metabolism genes. Only S-RSM4 and Syn9 have 6-phosphogluconate dehydrogenase and glucose 6-phosphate dehydrogenase annotated in their genomes. Syn9 and S-RSM4 also contain the gene trx, which encodes the regulatory protein thioredoxin, an inhibitor of G6PDH. Only the reduced form of thioredoxin regulates proteins. Thioredoxin is reduced by ferredoxin via the enzyme ferredoxin-thioredoxin reductase. S-RSM4 has putative genes for ferredoxin-thioredoxin reductase and ferredoxin.

Glucose-6 phosphate dehydrogenase catalyzes a rate limiting step of the pentose phosphate pathway: glucose 6-phosphate → 6-phosphogluconolactone. The oxidative phase of the pentose phosphate pathway produces two metabolites which are in high demand for cyanophage replication: ribose 5-phosphate, the precursor to nucleosides, and NADPH, the primary source of reducing power for most anabolic metabolism. The cyanophage would benefit from an increase in production of these metabolites. Additionally, cyanobacteria
produce the majority of their NADPH during the light reactions of photosynthesis\(^9\). Since replicating viruses require NADPH during the night as well, redundant G6PDH may be explained as an adaptation which allows for viral replication during the dark reactions of photosynthesis.

Synthetic genes for both the viral and the host G6PDH were optimized for expression in *E. coli*. The two enzymes were produced successfully, however the host G6PDH could not be solubilized. S-RSM4 G6PDH was purified with affinity chromatography. Kinetics and dynamics of the viral enzyme were investigated. These data were compared to other variants of G6PDH to explore the utility of this gene in the viral genome. Alternative strategies for the preparation of *Synechococcus* G6PDH were explored. This study opens doors to understanding the unique host/virus relationship of *Synechococcus sp. WH8103* and the S-RSM4 cyanophage.

**Methods**

**Cloning for Overexpression in *E. coli***

The open reading frame sequences for the gene zwf, encoding glucose-6-phosphate dehydrogenase (G6PDH), were extracted from the whole annotated genome sequences of cyanophage S-RSM4 (NCBI Reference Sequence: NC_013085.1) and cyanobacterium *Synechococcus sp. WH8103* (GenBank: CRY91594.1). Synthetic genes incorporating optimal codon usage for expression in *E. coli* were designed online using the Thermo Fisher/Invitrogen GeneArt\(^\text{TM}\) application. NdeI and BamHI restriction sites were added to the 5’ and 3’ ends, respectively, of the optimized S-RSM4 G6PDH gene. NdeI and HindIII restriction sites were added to the 5’ and 3’ ends, respectively, of the optimized WH8103 G6PDH gene. The synthetic
gene was then purchased from Thermo Fisher/Invitrogen, which supplied them in the vector pMA-T, containing a selectable Amp\(^R\) marker.

The synthetic S-RSM4 G6PDH gene was subcloned between the Ndel and BamHI sites of expression vector pET28a (Novagen), which fused the coding sequence for a thrombin-cleavable His\(_6\) affinity tag to the 5’ end of the coding sequence for the enzyme. Similarly, the synthetic WH8103 G6PDH gene was subcloned between the Ndel and HindIII sites of pET28a, which fused the same affinity tag to the 5’ end of this gene. See Figure 1 for a schematic of vector constructs. The sequences of the subcloned genes were verified by DNA Sanger sequencing (Vermont Integrative Genomics Resource).

The subclones in pET28a were transformed into competent *E. coli* Rosetta2 (DE3) cells (Novagen) and grown on LB agar with 50 µg/mL kanamycin (Kan) and 30 µg/mL chloramphenicol (Cam) plates at 37°C. The antibiotics selected for the presence of a Kan\(^R\) marker in the pET28a vector sequence, and a Cam\(^R\) marker in the lambda DE3 lysogen in *E. coli* Rosetta2 (DE3).

**Large-Scale Overexpression of Recombinant G6PDH in E. coli**

Single transformant colonies in Rosetta2 (DE3) were selected. These colonies were used to inoculate 50 mL of liquid LB containing Kan + Cam at concentrations equivalent to the drug conditions of the agar. These clones were amplified by overnight growth with shaking at 37°C. Each 50 mL overnight culture was used to inoculate 3 liters of LB + Kan + Cam. The 3-liter cultures were incubated with shaking at 37°C until cell density reached \(A_{600} = 0.6\), which marked mid-log phase growth. At this point, 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), an
artificial inducer of the *E. coli* lac operon, was added. Addition of IPTG induced the expression of T7 RNA polymerase from the DE3 lysogen in the Rosetta2 (DE3) cells, which in turn induced expression of N-terminally His\(_6\)-tagged G6PDH from the pET28a expression vector containing the synthetic, recombinant gene. Shaking was continued for 3 hours at 37°C. Cells were then harvested by centrifugation, resuspended in a small volume of media, and subjected to centrifugation again to form pellets in 50 mL conical tubes. The pellets were frozen in liquid N\(_2\) and stored at -80°C.

**Overexpression Assays**

Whole cell protein extractions were performed to monitor the success of overexpression. Following induction, but preceding the large-scale centrifugation, 1 mL of cell culture was aliquoted into a 1.5-mL microcentrifuge tube. The cells were subjected to centrifugation for 5 minutes at 13.2 thousand rpm. Supernatant was discarded, and the pellet was resuspended by Vortexing™ in 250 µL of 1X NuPAGE™ Sample Buffer (Thermo Fisher Scientific) with 15 mM dithiothreitol (DTT). Once resuspended, the cells were incubated at 95°C for 15 minutes. The sample was subjected to centrifugation for 1 minute at 13.2 thousand rpm, and then Vortexed™ until homogeneous. A final centrifugation for 1 minute at 13.2 thousand rpm created a pellet of solid cell debris.

A Novex NuPAGE™ 4-12% Bis-Tris protein gel with 1-mm wells was loaded with 5 µL of supernatant. The gel was submerged in 1X MES running buffer. An EC600 power supply was connected to the gel apparatus and was set to 200 Volts for 45 minutes. After three rinses in NANOpure™ water (Barnstead), the gel was submerged in GelCode Blue™ protein stain
For over an hour, the gel was kept in the dark, gently rocking in stain solution. A light table was then used to visualize the protein bands. BenchMark™ Protein Ladder Cat. no. 10747-012 (Thermo Fisher Scientific) was used to estimate the molecular mass of the proteins on the gel.

**Lysis and Purification**

Frozen cell pellets were thawed in an ice-water bath. They were resuspended in 4 volumes of Buffer A (Table 1) spiked with complete *Mini EDTA-free Protease Inhibitor Cocktail Tablets* (Roche) per the manufacturer’s specifications. The suspension was subjected to sonication in an ice-water bath. The sonicator was pulsed in order to avoid bubbles from forming. Sonication was complete when the initial $A_{600}$ reading of a diluted sample was reduced by ≥ 60%. The sonicate was then subjected to centrifugation for 30 minutes at 20,000 rpm in a Sorval RC6+ centrifuge with an F21-8x50y rotor, set to 4°C. The soluble fraction, or supernatant, was collected and stored on ice. The pellet was frozen in liquid N$_2$ and stored at -80°C.

Nickel affinity chromatography was performed at room temperature using a 5 mL HisTrap™ HP column (IMAC) attached to an ÄKTApurifier™ liquid chromatography station (GE Healthcare Life Sciences). Prior to loading the sample, the column was washed with 25 mL of Buffer B (Table 1), then equilibrated with 50 mL of Buffer A.

The sample supernatant (typically 30-40 mL from 7-10 grams of cell pellet) was applied to the column at a flow rate of 1 mL/min, then washed with 25 mL of Buffer A to remove unbound protein. Bound protein was then eluted with a 20-column volume (100 mL) gradient from 50 mM → 100 mM imidazole (Buffer A → Buffer B). His-tagged protein typically eluted in
a large peak from ~250 mM to ~400 mM imidazole, as shown by continuous UV absorbance monitoring of the eluant. The gradient was followed by 25 mL of Buffer B to clean the column. After, the column was rinsed with water and stored in 20% ethanol.

**Dialysis**

Peak fractions were pooled, aliquoted into three parts, and dialyzed into storage buffer. Precipitate developed after overnight dialysis at 4°C. However, after removal of precipitates by centrifugation, high concentrations of protein remained in the storage buffer. These were aliquoted separately into microfuge tubes, frozen in liquid N₂, and stored at -80°C.

**Purity Assay**

The eluted fractions were analyzed by denaturing SDS-PAGE using pre-cast NuPAGE™ Bis-Tris 10% polyacrylamide gels (Invitrogen). Molecular weights were determined by comparison to a BenchMark™ protein ladder (Invitrogen). After identification of peak fractions and verification of correct molecular weight, overloaded samples were analyzed by SDS-PAGE to reveal contaminant bands. Based on this analysis, the purified enzyme was judged to be over 95% pure, therefore no further purification was attempted before characterizing enzyme activity.

**S-RSM4 G6PDH Kinetics**

Michaelis-Menten kinetics, pH dependence, and product inhibition kinetics of the viral G6PDH were investigated. Assays were performed in a Varian Cary 50 UV-Visible
spectrophotometer, using 10-mm Quartz Spectrophotometer Cell Semi-micro cuvettes (Starna Cells). The wavelength was constant at 340 nm. Measurements were made as arbitrary absorbance units per minute and converted to reaction rates using the Beer-Lambert law\textsuperscript{13}. The product, NADPH, absorbs 340 nm light with an extinction coefficient ($\varepsilon_{340}$) of $6.3 \times 10^3$ M$^{-1}$ cm$^{-1}$\textsuperscript{14}. The spectrophotometric assays are based on this property.

Solid-state NADP$^+ \text{ reduced disodium salt, 97\%, (Thermo Fisher Scientific) and G6P sodium salt (Sigma) were dissolved in 100 mM Tris-HCl, pH 7.4 to stock concentrations of 9.84 mM and 31.07 mM, respectively. These concentrations were calibrated by repeated enzymatic assays, allowing the reactions to reach equilibrium. Since the equilibrium state of this reaction entails 100% conversion of NADP$^+$ to NADPH, $[\text{NADP}^+]_{\text{initial}} = [\text{NADPH}]_{\text{final}}$ when NADP$^+$ is limiting. When G6P is limiting, $[\text{G6P}]_{\text{initial}} = [\text{NADPH}]_{\text{final}}$ since both reactants and both products have a stochiometric coefficient of one. Stock solutions were diluted in 100 mM Tris-HCl, pH 7.4 when used in the assays.

The enzyme storage buffer was 20 mM Tris-HCl (pH 7.4), 2 mM $\beta$-mercaptoethanol, 5% glycerol, and 300 mM NaCl (Table 1). In order to determine the stock concentration of enzyme, the amino acid sequence was analyzed by the method of Gill and von Hippel\textsuperscript{2} via Protein Calculator version 3.4 (Source Forge). The algorithm calculated an extinction coefficient, $\varepsilon = 78,000$ M$^{-1}$ Cm$^{-1}$ at 280 nm. The absorbance reading, $A_{280} = 1.1161$, measured a stock concentration of 14.73 µM G6PDH. The stock enzyme solution was dissolved in buffer equivalent to the reaction conditions (Table 2) in a 1:10 dilution. Diluted enzyme was used for a final reaction concentration of 2 nM.
Identical reactions, under $V_{\text{max}}$ conditions, were conducted at varying pH levels. The pH was varied with either Tris-base or hydrochloric acid. Enzyme activity was maximal at pH 6.8. The reaction rate decreased by 25% at both pH 7 and pH 6.6. All following reactions were therefore conducted at pH 6.8.

The reaction conditions for the following Michaelis-Menten experiments are summarized in Table 2. The Michaelis-Menten constant for NADP$^+$ was determined by maintaining saturating G6P, and varying NADP$^+$ concentrations. Initial experiments were used to estimate $K_{M,\text{NADP}^+}$. Once an estimation was established, NADP$^+$ concentrations were varied from under $1/40 \times K_M$ to over $5X K_M$. Reaction rates for NADP$^+$ were corrected by a factor of 1.94 to compensate for a loss in enzyme activity over repeated freeze-thaw cycles.

Similarly, the Michaelis-Menten constant for G6P was determined by maintaining saturating NADP$^+$ and varying $[\text{G6P}]$. Initial experiments allowed for an estimation of $K_{M,\text{G6P}}$. From this estimation, experiments were designed and executed to vary G6P concentrations from under $1/6 \times K_M$ to over $7X K_M$.

Linear reaction rates occurred within the first 10 minutes of the assays. The spectrophotometric data was exported to Microsoft Excel, which calculated a least-squares regression line. Any regression with a coefficient of determination less than 0.99 was rejected. Although cleaning was strongly emphasized, dust and fibers would sometimes enter the cuvette. When these particulates crossed the beam, small peaks would appear in the data. These data were removed only if their removal improved the regression, and only if the interference peak occurred over an insignificant amount of time compared to the 10-minute reaction time.
Michaelis-Menten kinetics were analyzed with *Prism 8* (GraphPad Software, LLC). Two non-linear regressions were utilized: Michaelis-Menten and allosteric sigmoidal. No data correlated with an allosteric sigmoidal curve, whereas the Michaelis-Menten function explained the data precisely.

The software had options to either float or constrain $V_{\text{max}}$. Floating $V_{\text{max}}$ was more useful for $K_M$ estimations, whereas constraining $V_{\text{max}}$ increased the precision of the $K_M$ reported. Therefore, $V_{\text{max}}$ was estimated computationally and then determined experimentally. $V_{\text{max}}$ experiments were conducted with both substrates at saturating concentrations. The data were then retrofitted to the experimental $V_{\text{max}}$. This constraint made insignificant changes in the computed $K_M$ while increasing precision substantially.

Inhibition studies examined the product-inhibition by NADPH. Solid state NADPH sodium salt (Cayman Chemical Company) was dissolved in 100 mM Tris-HCl, pH 7.4 to create a 1.19 mM stock solution. With $[\text{NADP}^+] = K_M$ and $[\text{G6P}] = \text{saturating}$, NADPH concentrations were varied from 0 to 5 µM. These reaction rates were plotted against inhibitor concentrations. Using the generalized inhibition regression provided by *Prism 8*, a curve was fitted to the inhibition data. After 5 minutes of a 10-minute reaction with 5 µM NADPH, the NADP$^+$ concentration was increased by 1X $K_M$ equivalent. A comparison of the reaction rate before and after doubling $[\text{NADP}^+]$ was used to categorize the product inhibition.

**G6PDH/NADPH Binding Studies**

The binding of NADPH to free G6PDH enzyme was quantified by the quenching of protein fluorescence by NADPH. Fluorescence measurements were made with a
QuantaMaster™ QM-6 spectrofluorometer (Photon Technology International, South Brunswick, NJ). The excitation wavelength was 280 nm with a slit width of 1 nm, and the emission wavelength was 322 nm with a slit width of 5 nm. Titrations of G6PDH with NADPH were performed at room temperature (~21°C) in reaction buffer (25 mM Bis-Tris-HCl, pH 6.8, and 2 mM MgCl₂). In a quartz fluorescence cuvette, a 3-mL initial volume of 1 µM G6PDH in reaction buffer was titrated by adding small aliquots of a concentrated NADPH solution. After each addition, the solution was mixed, then allowed to equilibrate for 30 seconds before recording the fluorescence intensity signal in machine units. Each data point was corrected for the background signal of a buffer blank, as well as for dilution. The cumulative dilution factor in each titration was less than 10% of the starting volume. Binding curves were generated by converting raw data into fractional saturation of protein with ligand (Y) at each NADPH concentration, according to the relationship Y = (F₀ – F)/F₀, where F₀ is the initial (corrected) fluorescence signal of free protein and F is the observed (corrected) fluorescence signal at each NADPH concentration. Binding data thus generated from two different titration experiments were fit globally to the following equation using GraphPad Prism:

\[ Y = \frac{B_{\text{max}}X^h}{(K_d^h + X^h)} \]

where B_{\text{max}} is the maximum saturation of protein with ligand (assumed to be 1), X is the ligand (NADPH) concentration, K_d is the dissociation constant or ligand concentration at half-saturation (Y = 0.5), and h is the Hill coefficient, an indicator of binding cooperativity. Error bars shown in the data represent the SEM of the results of two experiments.
Results

S-RSM4 G6PDH Preparation

The whole-cell protein electrophoresis strongly supported a successful induction of S-RSM4 G6PDH (Figure 2A). The expected molecular mass of the His-tagged protein is 54.5 kDa, and very dark band was visible between the 60 kDa and 50 kDa ladder markers. Bands at this molecular mass in the uninduced sample were insignificant. When the scale of the induction was increased, these results remained consistent.

The ÄKTApurifier™ monitored the affinity chromatography with UV spectrophotometry. There was a large peak from the load and wash. After about 10 column volumes, the UV absorbance was close to zero. At this point, imidazole was increased from 40 mM to 500 mM, while being assayed by conductivity. About halfway along this gradient, the majority of the protein was eluted. The majority of the protein peak occurred over 4 column volumes.

Cellular G6PDH Preparation

WH8103 G6PDH was also successfully induced (Figure 2B). The expected molecular mass was 59.3 kDa, and a dark band appeared slightly ahead of the 60 kDa marker. The uninduced samples did not display this band. These results are consistent between the small-scale and large-scale inductions.

The cellular enzyme was insoluble in aqueous buffer. There was no protein peak off of the affinity column. A NuPAGE™ gel confirmed that the protein was exclusively located in the insoluble pellet from the final centrifugation step. There was suspicion that the WH8103 G6PDH was bound by inclusion bodies, formed by the E. coli cells. Samples from the insoluble pellet
were subjected to an inclusion body denaturing procedure using B-PER™ (Thermo Fisher Scientific). A NuPAGE™ gel demonstrated that the protein was still insoluble after the B-PER™ procedure. Next, milder induction conditions were attempted. Inductions were performed as described above, except 0.25 mM IPTG was used and cells were harvested after only 15 minutes. These mild induction conditions produced significant quantities of WH8103 G6PDH. The issue of insolubility persisted, however, since neither sonication nor the B-PER™ procedure yielded soluble protein.

**S-RSM4 G6PDH Kinetics**

\[ V_{\text{max}} \] was 0.774 nM/min at an enzyme concentration of 2 nM. This corresponds to a turnover rate \( (k_{\text{cat}}) \) of 23.2 s\(^{-1}\). The Michaelis-Menten kinetics of NADP\(^+\) were analyzed with 26 different experiments, ranging from \([\text{NADP}^+] = 0.25\) to 20 µM (Figure 3A). \( K_{M, \text{NADP}^+} \) was calculated to be 9.14 µM. The 95% confidence interval for \( K_{M, \text{NADP}^+} \) spanned from 8.44 to 9.92 µM. The coefficient of determination was 0.974. The floated \( V_{\text{max}} \) value was 0.773 nM/min with a 95% confidence interval of 0.679 to 0.896 nM/min. The catalytic efficiency \( (k_{\text{cat}}/K_{M}) \) of NADP\(^+\) was 2.54 µM\(^{-1}\) s\(^{-1}\) (Table 3).

The Michaelis-Menten kinetics of G6P utilized data from 17 different experiments, ranging from \([\text{G6P}] = 10\) to 350 µM (Figure 3B). \( K_{M, \text{G6P}} \) was calculated to be 61.8 µM. The 95% confidence interval for \( K_{M, \text{G6P}} \) spanned from 55.7 to 68.4 µM. The coefficient of determination was 0.981. The floated \( V_{\text{max}} \) value was 0.778 nM/min with a 95% confidence interval of 0.725 to 0.839 nM/min. Catalytic efficiency for G6P was 0.376 µM\(^{-1}\) s\(^{-1}\). The constraints on both the NADP\(^+\) and the G6P analyses were \( K_{M} > 0 \) and \( V_{\text{max}} = 0.774 \) (Table 3).
The SRSM4 variant of G6PDH is strongly inhibited by its product, NADPH (Figure 4). In a reaction containing saturating G6P and \([\text{NADP}^+] = K_{\text{M, NADP}^+}\), NADPH concentrations were varied from 0 to 5 µM. As compared to the control, 5 µM NADPH reduced the reaction rate by more than 50% (Table 4). These data were correlated to an inhibition kinetics regression with a coefficient of determination of 0.987. After 5 minutes of a 10-minute reaction where \([\text{NADPH}]_0 = 5 \text{ µM, } 1 \times K_{\text{M, NADP}^+}\) equivalent of NADP⁺ was added to the cuvette. The doubling of \([\text{NADP}^+]\) nearly doubled the reaction rate (Figure 5).

**G6PDH/NADPH Binding Studies**

The fluorescence data evinced very high-affinity, non-cooperative binding of NADPH to S-RSM4 G6PDH (Figure 6). \(B_{\text{max}}\) was calculated to be 1.15 with a 95% confidence interval from 1.10 to 1.19. The dissociation constant, \(K_d\) was 2.23 µM, with a 95% confidence interval from 1.93 to 2.57 µM. The Hill coefficient was 1.10 (Table 4). The coefficient of determination for this regression was 0.979.

**Discussion**

The coevolution of cyanobacteria and cyanophages follows a very unique narrative. Homologous photosynthetic genes, likely sourced from random recombination events⁵, became fixed in certain cyanophage lineages. Limited by their own molecular machinery, these viruses must depend on the host to express these genes during the infection cycle. Since the virus experiences an energy cost associated with bearing and replicating these genes, there ought to be a discernable fitness advantage to this coevolutionary development.
The cyanophage S-RSM4 carries more photosynthetic genes than any catalogued cyanophage, and thus it is an ideal model system. Certain genes encoding enzymes of the oxidative phase of the PPP; glucose-6-phosphate dehydrogenase (G6PDH), 6-phosphogluconate dehydrogenase, and trans-aldolase are found in the genomes of S-RSM4 and its host, *Synechococcus sp. WH8103*. G6PDH is of special interest, as it is rate-limiting in this pathway.1

The pentose phosphate pathway produces ribose 5-phosphate and NADPH, both of which are in high demand during viral replication. It is possible that G6PDH is carried by S-RSM4 simply to increase concentration of this enzyme and thus increase concentrations of the aforementioned metabolites. Rate-limiting enzymes tend to be highly regulated. This study suggests the S-RSM4 G6PDH is regulated differently from its host’s homologue, tailoring the PPP to its own benefit, at the expense of host energy and homeostasis. It is also possible that viral G6PDH is designed to maintain NADPH concentration during the dark reactions of photosynthesis.

Although *Synechococcus* G6PDH has been prepared before, there are no thorough kinetic data with which to compare the S-RSM4 G6PDH. Our study of S-RSM4 G6PDH is the first Michaelis-Menten kinetics investigation of a viral zwf gene. The zwf gene of filamentous cyanobacteria species *Nostoc punctiforme* was co-expressed and prepared with allosteric effector OpcA (Table 5). The $K_M$ of OpcA + G6PDH$_{N.~punctiforme}$ was 1,900 µM for G6P, and 24 µM for NADP$^+$. G6PDH$_{S-RSM4}$ displayed substrate affinities of $K_M, G6P = 61.8$ µM and $K_M, NADP^+ = 9.14$ µM. As compared to the *Nostoc punctiforme* enzyme, the S-RSM4 enzyme had over 30 times higher affinity for G6P and over 2.5 times higher affinity for NADP$^+$. Furthermore, when
G6PDH<sub>N. punctiforme</sub> was independent of OpcA, the S-RSM4 enzyme had over 1,300 times higher affinity for G6P and over 5 times higher affinity for NADP<sup>+</sup>.

If the G6PDH of *Synechococcus sp. WH8103* is at all similar to the G6PDH of *Nostoc punctiforme*, S-RSM4’s zwf gene has undergone divergent evolution to increase production of ribose 5-phosphate and NADPH even when their precursors are scarce. Whereas the cyanobacteria benefit from maintaining homeostasis through sensitivity to nutrient (G6P) and cofactor (NADP<sup>+</sup>) availability, the virus benefits from constitutive production of metabolites needed for viral replication, even if it drains the host of G6P and NADP<sup>+</sup>. The host G6PDH may be dependent on OpcA, but the virus apparently bypasses this regulatory constraint with its specialized zwf gene. The k<sub>cat</sub> of S-RSM4 G6PDH was 23.2 s<sup>-1</sup>, which is ~7 times slower than that of the *E. coli* enzyme<sup>20</sup>. Depending on how the *Synechococcus* enzyme’s turnover rate compares to that of *E. coli*, the effects of viral G6PDH may only be significant to overall cell metabolism at very low substrate concentrations. (Alternatively, the virus may overproduce G6PDH to compensate for its relatively low k<sub>cat</sub>-.) With a K<sub>M</sub> of 9.14 µM, it is evident that the viral enzyme is designed to operate even when substrate concentrations are very low.

S-RSM4 G6PDH is product inhibited by NADPH. The inhibition constant is estimated to be ~5 µM. This value indicates ~3X more potent NADPH inhibition than what is observed in *E. coli*<sup>20</sup>. Depending on how this K<sub>i</sub> compares with that of G6PDH<sub>WH8103</sub>, the viral enzyme may be more sensitive to NADPH than the host enzyme. This would be consistent with the hypothesis that S-RSM4 G6PDH is specialized to maintain NADPH concentrations during the dark reactions of photosynthesis. It may be very strongly inhibited in the daylight, while the light reactions are
producing NADPH, but become active at night. It is evident in the binding studies that NADPH binds very strongly to the free enzyme ($K_d = 2.23 \mu M$), again supporting this hypothesis.

The binding experiments and inhibition studies serve as initial experiments to discover the mechanism of action of the enzyme. Currently, the possibilities to explore are as follows:

**Ordered Sequential:** Most NADP+-dependent dehydrogenases exhibit Ordered Sequential Kinetics, in which the binding of substrates and release of products occurs in an obligatory, defined order, and all substrates must bind before catalysis may occur. Typically, NADP+ is the first substrate to bind and NADPH is the last product to leave in the forward reaction, and vice-versa in the reverse reaction. If cyanophage G6PDH has this kinetic mechanism, we would call it “Ordered Sequential Bi Bi”. (“Bi” stands for bimolecular reaction, and it would be “Bi Bi” because it would have to be bimolecular in both the forward and reverse directions.) We could represent this mechanism with the following schematic:

![Ordered Sequential Kinetics Diagram](image)

where $A = \text{NADP}^+$, $B = \text{G6P}$, $P = \text{6PG}$, $Q = \text{NADPH}$, $E = \text{free enzyme}$, $EA = \text{enzyme-NADP}^+ \text{ complex}$, $EQ = \text{enzyme-NADPH complex}$, and $[EAB/EPQ] = \text{ternary complex intermediate that undergoes catalysis}$.
If the mechanism is truly *Sequential* (not Ping-Pong), then in kinetics experiments done by varying \([\text{NADP}^+]\) at different fixed concentrations of G6P, or vice-versa, the Lineweaver-Burk \((1/v_0 \text{ vs. } 1/[S])\) plots of the data sets will be a series of intersecting lines.

If the mechanism is truly *Ordered Sequential* (not Random), then the product NADPH should exhibit competitive inhibition when \(\text{NADP}^+\) is the variable substrate, since both \(\text{NADP}^+\) and NADPH can only bind to free enzyme and compete for the same binding site. Other product inhibition patterns (e.g., NADPH as inhibitor with G6P as variable substrate, and 6PG as inhibitor with either \(\text{NADP}^+\) or G6P as variable substrate) will be non-competitive. Table 6 summarizes the expected results of product inhibition studies for the three types of Bi Bi kinetic mechanisms described above.

**Random Sequential:** There are examples of \(\text{NADP}^+\)-dependent dehydrogenases, including some G6PDH enzymes, that exhibit Random Sequential Kinetics. Therefore, we must consider it. In this mechanism, substrates bind and products release in random order, but the reaction still proceeds through the obligatory \([\text{EAB/EPQ}]\) ternary intermediate. The schematic to represent a “Random Sequential Bi Bi” mechanism looks like this:

![Diagram of Random Sequential Bi Bi mechanism](image)

where there are two different ways to form the \([\text{EAB/EPQ}]\) ternary intermediate in either direction.
Again, because the mechanism is Sequential (not Ping-Pong), kinetics experiments done by varying [NADP+] at different fixed concentrations of G6P, or vice-versa, will still give intersecting Lineweaver-Burk plots of the data sets.

The product inhibition patterns will be different in Random Sequential, however. NADPH will be competitive when G6P is the variable substrate, and either competitive or non-competitive when NADP+ is the variable substrate. Meanwhile 6PG will be non-competitive when G6P is the variable substrate, and competitive when NADPH is the variable substrate.

We do not really need to consider a “Ping-Pong” or double displacement type of mechanism because NADP+-dependent dehydrogenases simply do not work that way. Ping-Pong mechanisms involve transfers of chemical functional groups via covalent intermediates with enzyme. They skip the ternary complex intermediate that has to occur in electron transfer reactions involving NADP+/NADPH coenzymes. A Ping-Pong mechanism would be very easy to distinguish from either Ordered or Random Sequential because in Ping-Pong, the kinetics experiments with one substrate variable and the other fixed would give parallel Lineweaver-Burk plots.

Future experiments necessitate the purification of G6PDH from Synechococcus sp. WH8103. This will likely be accomplished through co-expression of G6PDH and OpcA. Cloning and expression studies are currently underway to test co-expression of recombinant WH8103 G6PDH and OpcA proteins from a pET28a expression vector in E. coli. Figure 1C shows the design of a co-expression cassette in pET28a. In this construct, the synthetic OpcA gene is cloned downstream of the cloned, synthetic gene that expresses His-tagged WH8103 G6PDH. The OpcA gene has its own T7 promoter, Lac O operator, and ribosome binding site (RBS).
These control elements were added during gene synthesis. This construct has been made and is currently undergoing DNA sequence analysis. As soon as the sequence is verified, the plasmid will be transformed into *E. coli*, grown and induced in liquid media as described in Methods. It is expected that whole cell electrophoresis will demonstrate co-expression of both WH8103 G6PDH and OpcA. Follow-up experiments will test whether soluble G6PDH can be recovered from these cells. If WH8103 G6PDH is soluble in aqueous buffers, kinetic and binding studies will be used to compare the host and viral enzymes. This comparison is crucial to understanding why S-RSM4 carries a unique version of G6PDH.
References


### Tables

<table>
<thead>
<tr>
<th>Buffer Type</th>
<th>A, Binding</th>
<th>B, Elution</th>
<th>Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Tris-HCl] mM</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>[NaCl] mM</td>
<td>300</td>
<td>300</td>
<td>300</td>
</tr>
<tr>
<td>[β-mercaptoethanol] mM</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>[Imidazole] mM</td>
<td>40</td>
<td>500</td>
<td>0</td>
</tr>
<tr>
<td>% Glycerol (w/v)</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>pH</td>
<td>7.4</td>
<td>7.4</td>
<td>7.4</td>
</tr>
</tbody>
</table>

*Table 1: Buffers for Protein Preparation.*

<table>
<thead>
<tr>
<th>Reaction Volume</th>
<th>900 µL</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Bis-Tris]</td>
<td>25 mM</td>
</tr>
<tr>
<td>[Tris-HCl]</td>
<td>3.89 mM</td>
</tr>
<tr>
<td>[MgCl₂]</td>
<td>2 mM</td>
</tr>
<tr>
<td>[NaCl]</td>
<td>407 µM</td>
</tr>
<tr>
<td>[NADP⁺] = Saturating</td>
<td>50 µM</td>
</tr>
<tr>
<td>[G6P] = Saturating</td>
<td>500 µM</td>
</tr>
<tr>
<td>[G6PDH] (Enzyme)</td>
<td>2 nM</td>
</tr>
<tr>
<td>pH = Optimal</td>
<td>6.8</td>
</tr>
<tr>
<td>Temperature</td>
<td>24-24.5 °C</td>
</tr>
</tbody>
</table>

*Table 2: Cuvette Reaction Conditions for Kinetics.*

| $V_{\text{max}}$ | 0.774 nM min$^{-1}$, 95% C.I. (0.679, 0.896) |
| $k_{\text{cat}}$ | 23.2 s$^{-1}$                                    |
| $K_M, \text{ NADP}^+$ | 9.14 µM, 95% C.I. (8.44, 9.92) |
| $k_{\text{cat}} / K_M, \text{ NADP}^+$ | 2.54 µM$^{-1}$ s$^{-1}$ |
| $K_M, \text{ G6P}$ | 61.8 µM, 95% C.I. (55.7, 68.4) |
| $k_{\text{cat}} / K_M, \text{ G6P}$ | 0.376 µM$^{-1}$ s$^{-1}$ |

*Table 3: S-RSM4 G6PDH Kinetics Results.*
Tables Cont.

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B_{\text{max}}$, NADPH</td>
<td>1.15, 95% C.I. (1.10, 1.19)</td>
</tr>
<tr>
<td>$K_d$, NADPH</td>
<td>2.23 μM, 95% C.I. (1.93, 2.57)</td>
</tr>
<tr>
<td>Hill coefficient</td>
<td>1.105</td>
</tr>
<tr>
<td>$K_i$, NADPH (Estimate)</td>
<td>~5 μM</td>
</tr>
</tbody>
</table>

Table 4: S-RSM4 Binding/Inhibition Studies Results.

<table>
<thead>
<tr>
<th>N. punctiforme enzyme +/ effector</th>
<th>$K_M$, NADP+</th>
<th>$K_M$, G6P</th>
</tr>
</thead>
<tbody>
<tr>
<td>G6PDH</td>
<td>49 ± 4.2 μM</td>
<td>65 ± 10 mM</td>
</tr>
<tr>
<td>G6PDH + OpcA</td>
<td>24 μM</td>
<td>1.9 ± 0.2 mM</td>
</tr>
</tbody>
</table>

Table 5: Nostoc punctiforme G6PDH Kinetics (Hagan & Meeks, 2001).

**COMPARISON: Product Inhibition Patterns for BiBi Mechanisms**

<table>
<thead>
<tr>
<th>Variable Substrate</th>
<th>Product Inhibitor</th>
<th>Ordered Sequential</th>
<th>Random Sequential</th>
<th>Ping-Pong</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>P</td>
<td>Non-competitive</td>
<td>Competitive</td>
<td>Non-competitive</td>
</tr>
<tr>
<td>A</td>
<td>Q</td>
<td>Competitive</td>
<td>Competitive or Non-competitive</td>
<td>Competitive</td>
</tr>
<tr>
<td>B</td>
<td>P</td>
<td>Non-competitive</td>
<td>Non-competitive</td>
<td>Competitive</td>
</tr>
<tr>
<td>B</td>
<td>Q</td>
<td>Non-competitive /None*</td>
<td>Competitive</td>
<td>Non-competitive</td>
</tr>
</tbody>
</table>

* At very high concentrations of fixed substrate.

Table 6: Expected Results for Different BiBi Mechanisms
Figure 1: Plasmid Constructs for Expression of Recombinant Cyanophage and Cyanobacterial Genes in E. coli. A) Commercial expression vector pET28a. The multi-cloning site (MCS) contains a variety of clustered restriction sites that we take advantage of for cloning cyanophage or cyanobacterial synthetic genes. B) For solo expression of His$_6$-tagged S-RSM4 G6PDH or WH8103 OpcA, synthetic genes are cloned between the NdeI (5’ end) and BamHI (3’ end) sites. For solo expression of WH8103 G6PDH, the synthetic gene was cloned between the NdeI site (5’ end) and HindIII site (3’ end). All three constructs will express n-terminally His-tagged protein upon induction in Rosetta2(DE3) E. coli cells. C) For simultaneous expression of WH8103 G6PDH and OpcA proteins, a synthetic gene cassette containing WH8103 OpcA under the control of an inducible T7 promoter was cloned between the HindIII (5’ end) and NotI (3’ end) sites, downstream of the synthetic gene for WH8103 G6PDH described in panel B, above.
**Figure 2: Inductions of His$_6$-tagged G6PDH from S-RSM4 and *Synechococcus sp. WH8103 zwf* gene in *E. coli.***

His-tagged, optimized genes were transfected into Rosetta2(DE3) *E. coli.* Cells were harvested after 3 hours under 1 mM IPTG at 37°C. Whole cell extracts prepared for Novex NuPAGE™ 4-12% Bis-Tris gel as described in methods. BenchMark Protein Ladder Cat. no. 10747-012 and GelCode Blue™ protein stain (Thermo Scientific) were used. “U” = uninduced, “I” = induced sample. A) S-RSM4 His$_6$-G6PDH visible at the expected molecular mass of 54.5 kDa, only in induced samples. B) *Synechococcus sp. WH8103* His$_6$-G6PDH visible at the expected molecular mass of 59.3 kDa, only in induced samples.
Figure 3: Michaelis Menten Kinetics of S-RSM4 G6PDH. $V_{\text{max}}$ is 0.774 nM min$^{-1}$ for 2 nM G6PDH, confirmed by both computational and experimental determination. A) the $K_M$ for NADP$^+$ is 9.14 µM. Reaction rates are corrected for differences in enzyme concentration. B) the $K_M$ for G6P is 61.8 µM. All rates are converted from $A_{340}$ to rate of NADPH formation. Reaction conditions are defined in Table 2. Data points are means. Error bars are standard deviations. Kinetics parameters reported in Table 3. Plots and calculations generated by Prism 8, GraphPad.

Figure 4: S-RSM4 G6PDH is Inhibited by Product, NADPH. Under the conditions of saturating G6P and [NADP$^+]=K_M$, initial [NADPH] was varied. With 5 µM NADPH, the reaction rate was halved. All rates are converted from $A_{340}$ to rate of NADPH formation. Reaction conditions are defined in Table 2. Plots and calculations generated by Prism 8, GraphPad.
Figure 5: NADPH may be a Competitive Inhibitor of NADP⁺. Blue, solid trendline) the linear reaction rate of S-RSM4 G6PDH with saturating G6P, [NADP⁺] = K_M, and [NADPH]₀ = 5 µM. Orange, dashed trendline) the same reaction as (Blue) after addition of one K_M, NADP⁺ equivalent. Slopes displayed are reaction rates, Δ[NADPH] min⁻¹. Addition of NADP⁺ nearly doubled the reaction rate, suggesting NADPH blocks NADP⁺ at the active site. Reaction conditions are defined in Table 2. Plots generated in Microsoft Excel.

Figure 6: Binding of NADPH to Free S-RSM4 G6PDH. The equilibrium binding affinity of G6PDH for NADPH was measured by titrating free enzyme with NADPH while monitoring quenching of the intrinsic fluorescence of G6PDH by the coenzyme. Fluorescence readings were taken using an excitation wavelength of 280 nm and emission wavelength of 322 nm. The starting concentration of free enzyme was 1 µM. Other details of the experimental conditions are listed under Methods. Raw data were corrected for dilution and for the background fluorescence of buffer. Corrected raw data were converted into fractional saturation of protein with ligand \( Y = (F₀ - F)/F₀ \) and plotted vs. NADPH concentration. Data were fit globally to the isotherm \( Y = \frac{B_{max}X}{K_d + X} \) as described under Methods. Binding parameters are listed in Table 4. Plots and calculations generated by Prism 8, GraphPad.