RP TU Rheinland-Pfälzische Technische Universität Kaiserslautern Landau

Functional characterization of viral auxiliary metabolic genes involved in tetrapyrrole biosynthesis

vom Fachbereich Biologie

der Rheinland-Pfälzischen Technischen Universität Kaiserslautern-Landau

zur Verleihung des akademischen Grades Dr. rer. nat. genehmigte

Dissertation

von

Helen Wegner, M.Sc., geb. in Heidelberg

Mündliche Prüfung: 16.01.2025

Dekan:	Prof. Dr. Stefan Kins
Promotionskommissionsvorsitzender:	Prof. Dr. Michael Schroda
Berichterstattende:	Prof. Dr. Nicole Frankenberg-Dinkel
	Prof. Dr. Stefanie Müller-Schüssele

D 386

Table of contents

Abbreviations

AMG	auxiliary metabolic gene
APS	ammonium persulfate
BCIP	5-bromo-4-chloro-3-indolyl phosphate
BSA	bovine serum albumin
BV IXα	biliverdin IXα
Chl	chlorophyll
CV	column volume
DHBV	15,16-dihydrobiliverdin
DMSO	dimethyl sulfoxide
dNTP	deoxyribonucleotide triphosphate
DTT	dithiothreitol
Fd	ferredoxin
FDBR	ferredoxin-dependent bilin reductase
HO	heme oxygenase
IPTG	isopropyl-β-D-thiogalactopyranoside
LB	Lysogeny broth (Luria Bertani)
MW	molecular weight
NBT	nitro blue tetrazolium chloride
$OD_{600\ nm}$	optical density at 600 nm
PCB	phycocyanobilin
PcyA	phycocyanobilin:ferredoxin oxidoreductase
PcyX	ferredoxin-dependent bilin reductase yielding PEB
PCR	polymerase chain reaction
PEB	phycoerythrobilin
PebA	15, 16-dihydrobiliverdin:ferredoxin oxidoreductase
PebB	phycoerythrobilin:ferredoxin oxidoreductase
PebS	phycoerythrobilin synthase
PS	photosystem
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel
TEMED	N ,N ,N',N' – tetramethylethylene-1,2-diamine
T_m	melting temperature
UV-Vis	Ultra violet - Visible
v/v	volume per volume
w/v	weight per volume

1. Introduction

1.1 The significance of linear and cyclic tetrapyrroles in nature

Tetrapyrroles are present in all living beings, ranging from bacteria and plants to animals; therefore, they are called the pigments of life (Battersby 2000). Their structure is remarkably diverse, facilitating the biological processes they are included in, like photosynthesis, oxygen transport, and respiration. Tetrapyrroles generally consist of four pyrrole rings in either linear or cyclic structure. Their characteristic color is due to their conjugated ring system. The cyclic tetrapyrroles are the precursors of the linear tetrapyrroles. Their ability to chelate different central metal ions by the nitrogen atoms of the pyrrole rings gives them distinct chemical properties. The variation of tetrapyrroles is achieved through a series of enzyme-mediated transformations from the common template (uroporphyrinogen III) that alter the oxidation state, size, side chain composition, and the centrally chelated metal ion. The resulting tetrapyrroles include hemes, siroheme, chlorophylls, corrins (including vitamin B_{12}), cofactor F_{430}, and linear tetrapyrroles, which have diverse functions in the cell (Figure 1) (Bryant *et al.,* 2020).

Cobalamins, containing cobalt as a central metal ion, play a vital role in transcription, amino acid metabolism, and serves as a cofactor for multiple enzymes (Banerjee & Ragsdale 2003). The greenish siroheme is an essential cofactor in sulfur and nitrogen metabolism (Murphy & Siegel 1973, Murphy *et al.,* 1973, Murphy *et al.,* 1974). While, the yellow tetrapyrrolic cofactor F_{430} contains nickel. F_{430} can be found in methanogenic bacteria and archaea and is involved in the final step in methanogenesis as a cofactor for methyl-coenzyme-M reductase (Diekert *et al.,* 1980, Thauer & Bonacker 1994). The green cyclic tetrapyrrole chlorophyll has a chelated Mg^{2+}-ion, which is crucial for light-harvesting and energy transfer of the oxygenic photosystems in plants, algae, and cyanobacteria. Its related bacteriochlorophyll is present in anoxygenic photosynthesis of certain photosynthetic bacteria like green sulfur and purple bacteria, which absorb light in the infrared region, another wavelength plants or cyanobacteria use. Contributing to the red blood color of humans is heme, a component of hemoglobin consisting of a Fe-containing porphyrin (Dailey *et al.,* 2017). Its central iron exists in multiple oxidation states, which enables heme to act in diverse roles, e.g., as an electron

carrier in cytochromes, sensing of diatomic gases, or as a prosthetic group in enzymes for catalysis, e.g., cytochrome P450, catalase, peroxidase) (Hamza & Dailey 2012). As iron is often scarce, some pathogenic bacteria evolved mechanisms for heme uptake and cleavage to obtain iron for their growth benefit (Skaar *et al.,* 2004, Wilks & Schmitt 1998, Zhu *et al.,* 2000). This is not the only reason why cyclic tetrapyrroles get degraded. The chlorophyll in plants is broken down during leaf senescence to avoid the potential phototoxic effects of reactive oxygen species (ROS) formation, which is mediated by phototoxic tetrapyrrole (Foyer *et al.,* 1994, Hörtensteiner & Kräutler 2011). Also, free heme is cytotoxic, reacting with oxygen and forming ROS (Everse & Hsia 1997). Therefore, tetrapyrrole biosynthesis has to be tightly regulated, and excessive concentrations must be lowered; otherwise, cell death in bacteria can occur, while human diseases like X-linked protoporphyria or sideroblastic anemia are also characterized by unregulated or disturbed tetrapyrrole biosynthesis (Cartwright & Deiss 1975, Everse & Hsia 1997, Kumar & Bandyopadhyay 2005, Nakahigashi *et al.,* 1991). The catabolism of heme is mediated by heme oxygenases (HOs) that catalyze the oxidative cleavage of heme at the methine bridge to obtain the linear tetrapyrrole biliverdin IX (BV IX). BV IX has antioxidative properties and can further be utilized for the biosynthesis of chromophores, which some organisms use for light-sensing or -harvesting. Linear tetrapyrroles obtained from heme can act as light-harvesting chromophores of phycobiliproteins, present in cyanobacteria, cryptophytes, and red algae. Additionally, the chromophores can act as light sensors of phytochrome-like photoreceptors that can be found in fungi, bacteria, plants, and algae (Grombein *et al.,* 1975, Lagarias & Rapoport 1980, Sharrock 2008).

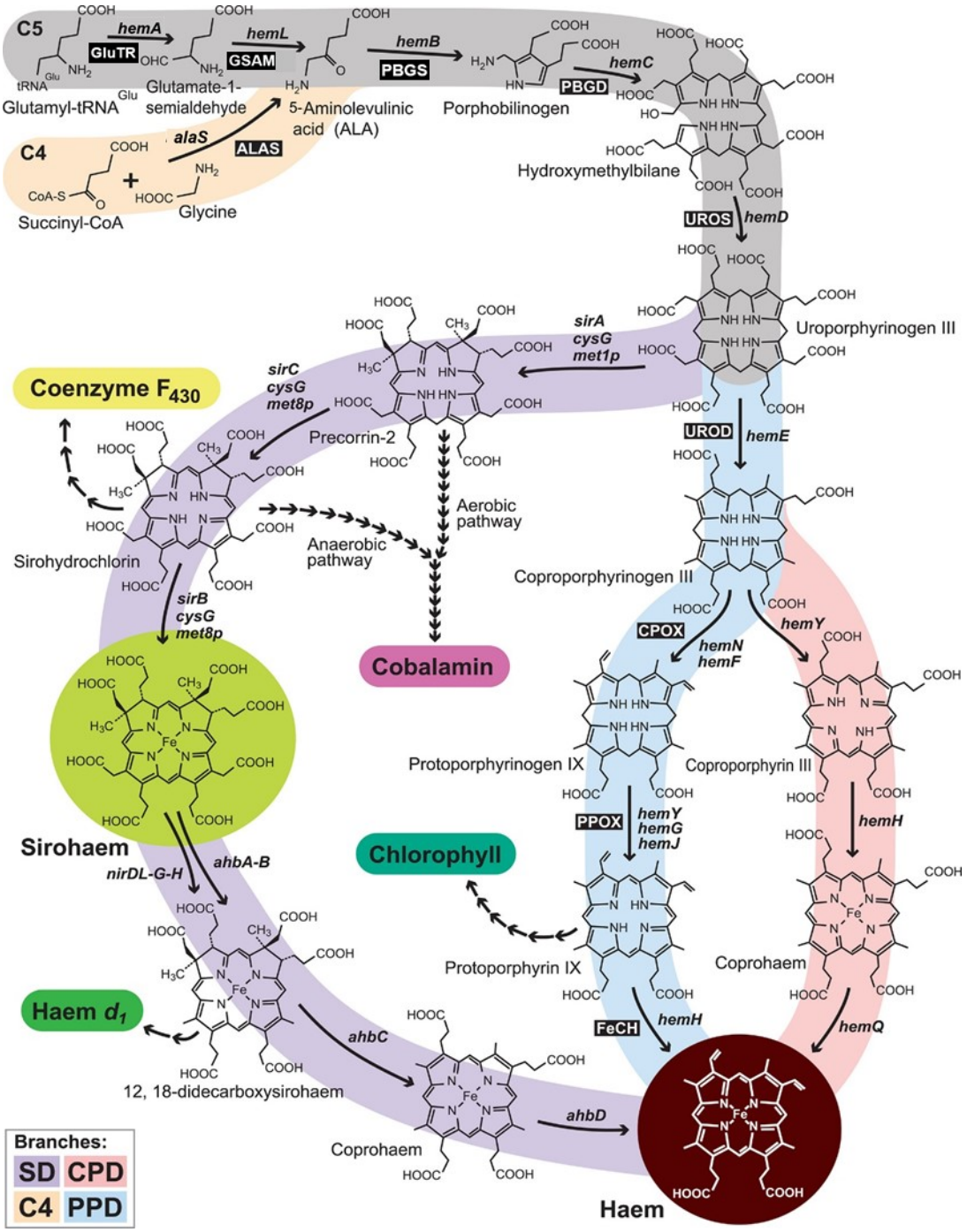


Figure 1: Tetrapyrrole biosynthesis pathways present in different organisms leading to heme synthesis. The two pathways, C4 and C5, lead to the synthesis of 5-aminolevulinic acid (ALA), the common precursor of tetrapyrroles. The C4 pathway is present in alphaproteobacteria, animals and other non-plant eukaryotes, while the C5 pathway is present in all remaining organisms. ALA is used to synthesize the first cyclic tetrapyrrole uroporphyrinogen III via different enzymes. The different pathways branch from here, while the archaeal siroheme-dependent (SD) pathway utilizes uroporphyrinogen III for synthesizing heme and other tetrapyrroles. The other two pathways first synthesize coproporphyrinogen III out of uroporphyrinogen III and then branch into the ancient coproporphyrin-dependent (CPD) pathway and the protoporphyrin-dependent (PPD) pathway. The pathway branches are depicted in the respective colors. The respective enzymes, which are associated with the PPD branch, are abbreviated in black rectangles. The gene names encoding individual enzymes are in italics, followed by the standard abbreviation. Please be aware that some genes encoding enzymes have a similar abbreviation while responding to different enzymes, e.g., *hemH,* which is protoporphyrin ferrochelatase in the PPD pathway, and coproporphyrin ferrochelatase in the CPD pathway. Dailey *et al*. 2017 suggested a more distinct abbreviation, which remains to be implemented. There are also alternative genes for some reactions of which some evolved independently. The synthesis of heme and other tetrapyrroles with the required number of enzymatic steps is indicated with arrows. Abbreviation: GluTR = Glutamyl-

tRNA reductase; GSAM= Glutamate-1-semialdehyde aminotransferase; ALAS = Aminolevulinic acid synthase; PBGS = Porphobilinogen synthase; PBGD = Porphobilinogen deaminase; UROS = Uroporphyrinogen synthase; UROD = Uroporphyrinogen decarboxylase; CPOX = Coproporphyrinogen oxidase; PPOX = Protoporphyrinogen oxidase; FeCH = Ferrochelatase. Figure adapted and modified from (Kořený *et al.,* 2022).

1.2 The biosynthesis of tetrapyrroles

The first steps in synthesizing tetrapyrroles are restricted to two pathways for synthesizing 5-aminolevulinic acid (ALA), the common precursor of tetrapyrroles in all organisms. Both pathways, the C4 and C5, occur exclusively in certain organisms and result in the synthesis of ALA. From there on, three highly conserved enzymatic steps are required to form the first cyclic tetrapyrrole uroporphyrinogen III. Contrary to the former belief that from the first cyclic tetrapyrrole, the synthesis of other tetrapyrroles like heme, (bacterio-) chlorophyll, corrinoid, and siroheme is highly conserved, it was discovered that the synthesis can vary substantially in different organisms.

1.2.1 The biosynthesis of the common tetrapyrrole precursor

The precursor of all tetrapyrroles is a molecule known as ALA, which can be produced via two unrelated and independent pathways, the C5 pathway (Beale & Castelfranco 1973, Beale 1990) and the C4 or Shemin pathway (Shemin & Rittenberg 1945) (Figure 1). In the C5 pathway, glutamate's five-carbon skeleton (Beale & Castelfranco 1973) is utilized, whereas the C4 pathway involves a condensation reaction between glycine and succinyl-Coenzyme A (succinyl-CoA) to generate ALA (Gibson *et al.,* 1958). The C4 reaction is catalyzed by a single homodimeric enzyme aminolevulinic acid synthase (ALAS) (EC 2.3.1.37), a pyridoxal 5'-phosphate (PLP) dependent enzyme (Scholnick *et al.,* 1972) which is present in alphaproteobacteria, animals, and yeast (Figure 2).

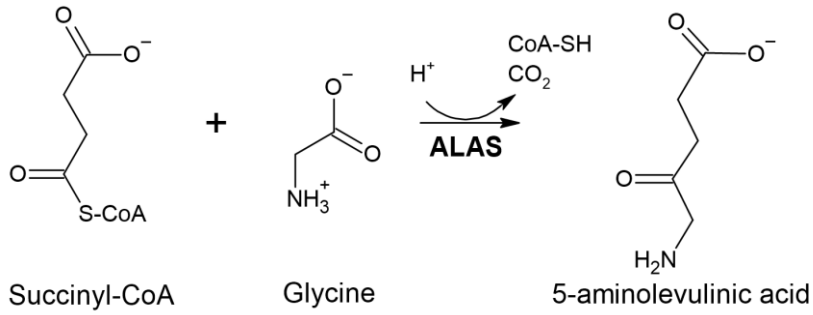


Figure 2: The reaction catalyzed by aminolevulinic acid synthase (ALAS) yielding 5-aminolevulinic acid (ALA). The C4 or Shemin pathway uses the decarboxylative condensation of the two substrates succinyl-Coenzyme A (Succinyl-CoA) and glycine to form ALA. Hereby the enzyme ALAS releases CO_2 and CoA-SH while using a proton to catalyze this reaction. ALAS is a pyridoxal 5-phosphate (PLP)-dependent enzyme with the unusual property of cleaving two bonds as it catalyzes both a decarboxylation and a deprotonation.

In comparison, the C5 pathway relies on a series of enzymatic reactions to convert glutamyl-tRNA to ALA and is predominant in most bacteria, archaea, and higher plants (Figure 1). The first step in the C5 pathway involves the NADPH-dependent enzyme glutamyl-tRNA reductase (GluTR) (EC 1.2.1.70) to catalyze the reduction of glutamyl-tRNA, which is also used for protein biosynthesis, to glutamate-1-semialdehyde. This is further transformed into ALA through the action of the PLP-dependent enzyme glutamate-1-semialdehyde aminotransferase (GSAM) (EC 5.4.3.8) (Ilag & Jahn 1992, Kannangara & Gough 1978, Schön *et al.,* 1986). Hereby, GSAM utilizes only a single substrate, thus functioning as an aminomutase.

Structurally similar to GSAM is the C4 pathway enzyme ALAS. It is a family member of α-oxoamine synthases, which are responsible for the formation of α-amino ketones (Schulze *et al.,* 2006). The PLP-mediated reaction of ALAS is catalyzed through the covalently linked lysine residue, forming an internal aldimine. This gets converted upon binding the first substrate glycine (produced in the serin biosynthesis) into an external aldimine (Kaufholz *et al.,* 2013). The binding of succinyl-CoA (produced in the TCA cycle) and release of CoA yield a quinonoid intermediate II, which, when protonated, releases ALA. Here, the reaction follows a *bi-bi*-kinetic mechanism with glycine binding prior to succinyl-CoA and the release of CO_2 and CoA prior to the product release of ALA via the untypical cleavage of two α-carbon bonds (Figure 3) (Astner *et al.*, 2005, Hunter *et al.,* 2007, Kaufholz *et al.*, 2013).

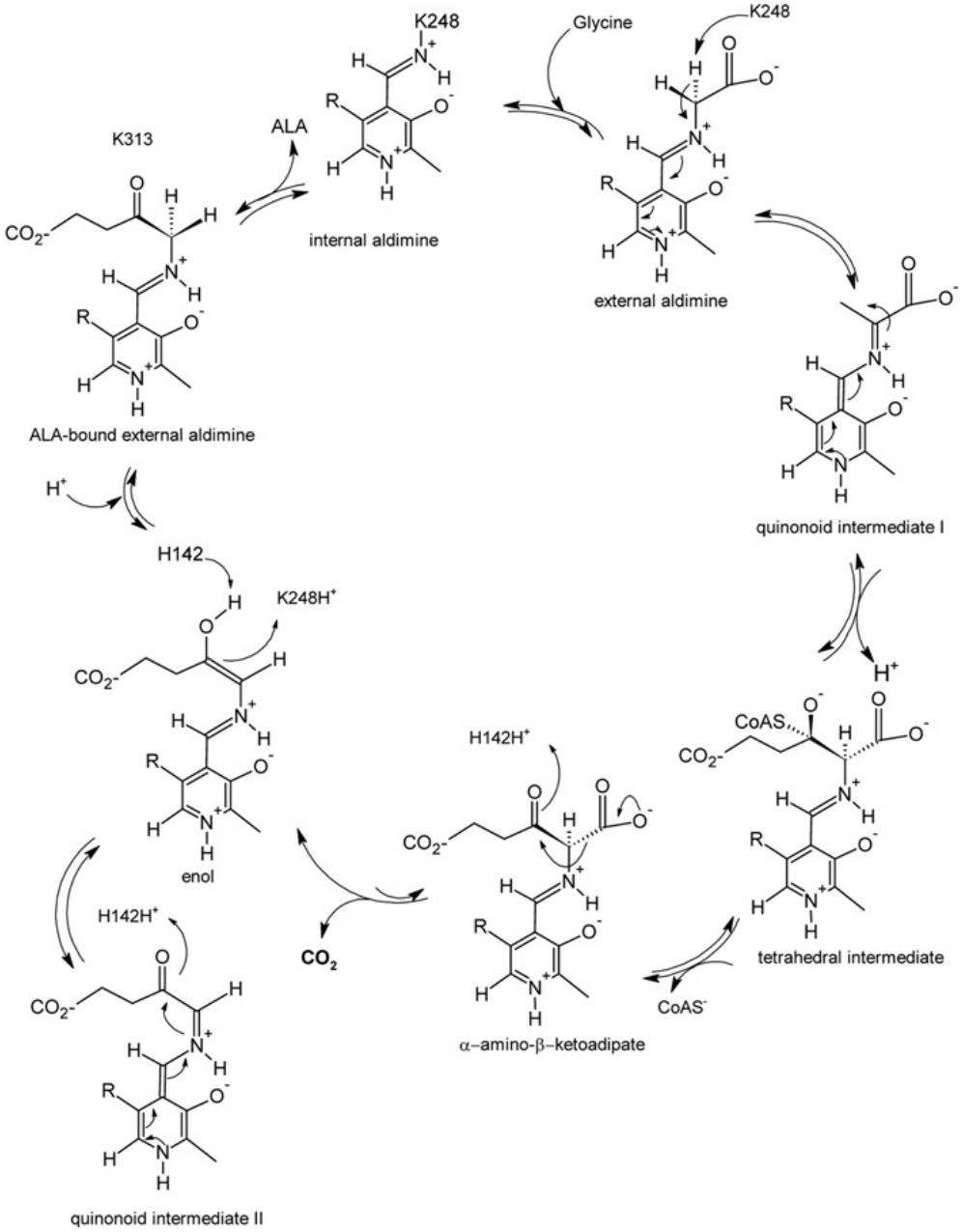


Figure 3: Proposed reaction mechanism of ALAS utilizing PLP as cofactor to synthesize ALA out of the substrates glycine and succinyl-CoA. Displayed are the proposed intermediates and important catalytic residues of the *R. capsulatus* enzyme ALAS required for the synthesis of ALA**.** The PLP cofactor is bound by a Schiff's base linkage on the active site lysine residue K248 of ALAS as internal aldimine. Following the binding of glycine the catalysis proceeds by quinonoid I intermediated, followed by quinonoid II intermediate after succinyl-CoA binding and release together with CO_2. Ultimately resulting in the release of ALA. Respective important amino acid residues are labelled regarding *R. capsulatus* numbering. Figure adapted from (Kaufholz *et al*., 2013).

Despite the fact that the catalyzed reaction is similar in alphaproteobacteria, animals, and yeasts, the amino acid sequence of ALAS enzymes varies. ALAS sequences from humans to alphaproteobacteria are quite similar, with sequence identities ranging from 40 % to 80 %. The three-dimensional structure of the homodimeric enzyme has a characteristic 'yin yang' structure formed by the two monomers, with the active site

positioned at the subunit interface as a hole. In the absence of the two substrates, the cofactor resides covalently attached via a Schiff's base linkage to the active site lysine in the center of the structure in a nearly non-aqueous environment to prevent side reactions. Additionally, a mobile loop (near the carboxy terminus) of amino acids closes over the active site if the substrates are bound and open up when the reaction is carried out and ALA dissociates out of the active site (Astner *et al*., 2005, Hunter & Ferreira 1999, Hunter *et al*., 2007). The ALAS enzyme can be present in an open and closed state. During the open state, the substrates can bind, and succinyl-CoA binding will induce the enzyme's close state (Zhang & Ferreira 2002).

Initially, the C4 and C5 pathways were thought not to co-occur in the same organism, but a few exceptions are known. For example, in the photosynthetic phytoflagellate *Euglena gracilis*, both pathways occur but are separated in different compartments (Weinstein & Beale 1983). In *Streptomyces nodosus* subsp. *asukaensis* also both pathways are present, but the produced ALA via the C4 pathway is exclusively used for the production of the antibiotic asukamycin and not incorporated into heme (Petříček *et al.,* 2006). Additionally, in some bacteria and vertebrates, two ALAS enzymes are present. In humans, there is a ubiquitously expressed housekeeping isoform (ALAS1) and an erythroid-specific isoform (ALAS2), which is upregulated during erythropoiesis to supply enough heme and it is exclusively present in erythroid cells (May *et al.,* 1995, Sutherland *et al.,* 1988). In bacteria, e.g., *Rhodobacter sphaeroides,* the two enzymes HemA and HemT are present, and only one isoform is sufficient for survival (Neidle & Kaplan 1993), which is not possible for the human isoforms. Mutations in the human erythroid ALAS are associated with severe illnesses like X-linked sideroblastic anemia when the mutation is near the PLP binding site or protoporphyria when genetic alterations cause overexpression (Bottomley *et al.,* 1995, Whatley *et al.,* 2008).

1.2.2 ALA-derived biosynthesis of cyclic tetrapyrroles

Subsequently, the formed ALA is utilized to create the cyclic tetrapyrroles via diverse porphyrin intermediates. The synthesis of these tetrapyrroles is organism- and pathway-specific and results in the formation of, e.g., corrins, cofactor F_{430}, (bacteria-) chlorophyll, and heme (Dailey *et al*., 2017) (Figure 1). Therefore, after synthesizing ALA via the C4 or C5 pathway, ALA is proceeded by three universal enzymatic steps common in all tetrapyrrole synthesizing organisms. First, two ALA molecules are asymmetrically condensed via porphobilinogen synthase (PBGS) (EC 4.2.1.24) to yield the pyrrole porphobilinogen (PBG) (Gibson *et al.,* 1955). Then, four PBG molecules are condensed by PBG deaminase (PBGD) (EC 2.5.1.61), forming a linear tetrapyrrole, which gets cyclized to uroporphyrinogen III in the last step by uroporphyrinogen-III synthase (UROS) (EC 4.2.1.75) (Frydman & Feinstein 1974, Levin & Coleman 1967, Tsai *et al.,* 1987) (Figure 1). This is the last universal intermediate, from here on, different heme pathways emerge in bacteria and archaea. Contrary to the belief that only one heme pathway exists in nature, which is only valid for eukaryotes, three pathways are known among prokaryotes (Dailey *et al*., 2017) (Figure 1). Namely, the classic or protoporphyrin-dependent pathway (PPD), the coproporphyrin-dependent pathway (CPD), and the siroheme-dependent pathway (SD), all named after their respective key intermediates. The SD pathway diverges after uroporphyrinogen III. In contrast, the CPD and PPD pathways split after the next synthesis step of uroporphyrinogen III, at coproporphyrin III. While the PPD pathway operates in many Gram-negative bacteria and eukaryotes, the CPD pathway is primarily present in Gram-positive bacteria. The SD pathway from sulfate-reducing bacteria and archaea is the most ancient one from an anaerobic world (Dailey *et al*., 2017). All leading ultimately to the synthesis of heme and branch off in between to synthesize other important tetrapyrroles, like (bacterio-) chlorophyll, vitamin B_{12}, siroheme, and cofactor F_{430}.

The presence or absence of oxygen is also reflected in the occurrence of specific pathways and their regulation. Adding to the complexity of the pathways is the presence of isozymes that catalyze the same reaction but are the oxygen-dependent/ -independent counterparts of specific enzymes in the pathway (Celis &

DuBois 2019, Dailey *et al*., 2017). The prominent risk of cellular damage by free iron, heme, and porphyrin intermediates is common for all three pathways. Therefore, tight regulation is crucial because free heme exhibits a cytotoxic effect (Everse & Hsia 1997). The control of the metabolic flow is also necessary because intermediate porphyrins, in the presence of light or certain metal ions, are highly photosensitive and can generate ROS such as singlet oxygen and hydroxyl radicals, leading to oxidative damage of the cells (Nakahigashi *et al*., 1991). Controversially, ROS formation can be detoxified by proteins like superoxide dismutases, catalases, and peroxidases, which depend on the availability of heme biosynthesis and iron. Catalase and peroxidase have a bound heme as a cofactor, thus bringing them under the regulation of heme biosynthesis. Meanwhile, superoxide dismutases have a bound iron, and their formation is controlled by iron availability and is also connected to heme synthesis (Kumar & Bandyopadhyay 2005). Tetrapyrroles contribute to ROS generation and are essential for detoxifying ROS. Therefore, the whole heme biosynthesis requires tight regulation and coordination of the iron and heme availability.

Proper heme biosynthesis is essential for maintaining cellular heme levels and functioning heme-dependent processes, which are strictly controlled by various endogenous and environmental factors (Donegan *et al.,* 2019, Lascelles 1960, May *et al*., 1995). The main regulation seems to occur at the start of the C4 and C5 pathways with the production of the first committed compound of the pathway, namely ALA (Chen *et al.,* 1996, Ikushiro *et al.,* 2018). Here heme was shown to inhibit GluTR activity in cell extract of bacterial lysates by allosteric feedback mechanisms or targeted proteolysis in the C5 pathway (Jones & Elliott 2010, Levicán *et al.,* 2007, Rieble & Beale 1988, Wang *et al.,* 1997). In the C4 pathway, ALAS activity was inhibited by negative feedback inhibition of heme at the synthesis of ALA (Burnham & Lascelles 1963, Ikushiro *et al*., 2018). This strong regulation at the entry point of heme biosynthesis regulates the product formation rate and thus, the speed of the pathway (Beator & Kloppstech 1993, Burnham & Lascelles 1963, Chen *et al*., 1996, McCormac *et al.,* 2001). The activity of the following pathway enzymes typically surpasses that of the first rate-limiting steps, resulting in

consistently high enzymatic activity and preventing the accumulation of porphyrin intermediates due to this excessive pressure (Kwon *et al.,* 2003).

Since ALA is a tightly regulated precursor for heme synthesis, it also affects the synthesis of branching tetrapyrroles such as (bacterio-) chlorophyll, vitamin B_{12}, siroheme, and cofactor F_{430}. These tetrapyrrole or -like structures participate in respiration, oxygen sensing, iron acquisition, detoxification, virulence, and biofilm formation. As a prosthetic group in hemoproteins, heme catalyzes various enzymatic reactions, enabling electron transfer, oxidative metabolism, and the breakdown of toxins and drugs.

1.2.3 The biosynthesis of bilins

Linear tetrapyrroles are capable of light-sensing and -harvesting derived from the cyclic heme via the ring-opening step of heme oxygenases (HO) (Cornejo & Beale 1988, Rhie & Beale 1992). The resulting BV IXα is further reduced by ferredoxin-dependent bilin reductases (FDBRs) to generate colorful chromophores like the blue phycocyanobilin (PCB), the pink phycoerythrobilin (PEB) or the orange phycourobilin (Beale & Cornejo 1991, Glazer & Hixson 1977).

Heme oxygenases

Heme is the starting point for the synthesis of bilins. The heme oxygenase family is currently divided into two groups, while further members are discussed: 1. Canonical HOs, 2. Non-canonical and IsdG-like (iron-regulated surface determinant system) HOs (Chim *et al.,* 2010, Lyles & Eichenbaum 2018, Nambu *et al.,* 2013, Skaar *et al*., 2004, Wilks & Heinzl 2014). The canonical HOs (EC 1.14.14.18) catalyze a regiospecific opening of the heme ring in an oxygen-dependent step to produce BV IX. The second group of HOs is characterized by its unique structure and novel cleavage of the protoporphyrin ring, yielding staphylobilin or mycobilin as the product. The most common product of canonical HOs is BV IXα, where the α-methine bridge is cleaved (Figure 5). The other isomers of BV IXβ, IXγ, and IXδ are also produced via enzymatically catalyzed reactions, whereas the presence of all isomers in equimolar amounts indicates a non-enzymatic coupled reaction (Bonnett & McDonagh 1973). The isomers BV IXβ and IXδ can be synthesized by the heme oxygenase HemO (formerly known as PigA) of *P. aeruginosa*

(Ratliff *et al.,* 2001), and the unusual dicysteinyl-BV IXγ occurs in the hematophagous insect *Rhodnius prolixus* (Paiva-Silva *et al.,* 2006). The synthesis of the other BV IX isomers is due to an unusual orientation of the heme in the active site.

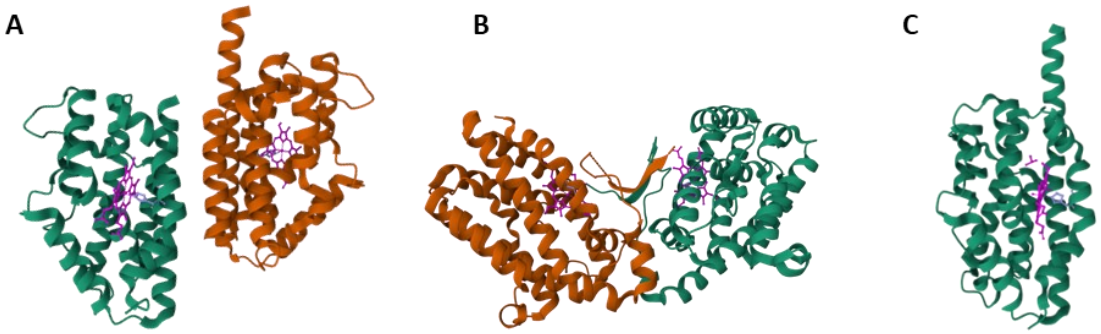


Figure 4: Crystal structure of different heme oxygenases in complex with heme. Displayed are heme oxygenases as monomer or dimer, the respectively bound heme is depicted in pink. **A** Human heme oxygenase 2 with bound heme, lacking three heme regulator motifs domains and the membrane binding region (RCSB PDB code: 2QPP). **B** Crystal structure of heme oxygenase 2 from *Synechocystis* sp. PCC6803 in complex with heme (RCSB PDB code: 1WOW). **C** Heme oxygenase 1 from cyanobacterium *Synechocystis* sp. PCC6803 in complex with heme (RCSB PDB code: 1WE1). Images were created using Mol* at RCSB.org (Berman *et al.,* 2000, Bianchetti *et al.,* 2007, Sehnal *et al.,* 2021, Sugishima *et al*., 2004, Sugishima *et al.,* 2005).

Bacterial heme oxygenases are generally similar in structure, consisting of primarily α-helical folds, while having low sequence similarity in common (Figure 4). The mammalian HOs are membrane-anchored C-terminally in the endoplasmic reticulum but otherwise have a similar structure. Interestingly, they are thought to promote self-association to form dimers and higher-order oligomers (Hwang *et al.,* 2009). Multimer complexes are not typical for bacterial HOs. They mainly occur monomeric or, with a few exceptions, homodimeric as the HO2 of *Synechocystis* sp. PCC 6803 (Sugishima *et al*., 2005). The heme is located between alpha helices, where the proximal helix has the histidine ligand with a water molecule that coordinates the heme in addition to the propionate groups present at the protein's surface (Figure 4). In the distal heme binding pocket, a network of hydrogen bonds of several water molecules is located and proposed to be necessary for HO activity while being interrupted by hydrophobic residues in *Synechocystis* sp. PCC 6803 HO2 and mammalian HO1 (Hirotsu *et al.,* 2004, Lightning *et al*., 2001, Sugishima *et al.,* 2000, Sugishima *et al.,* 2003).

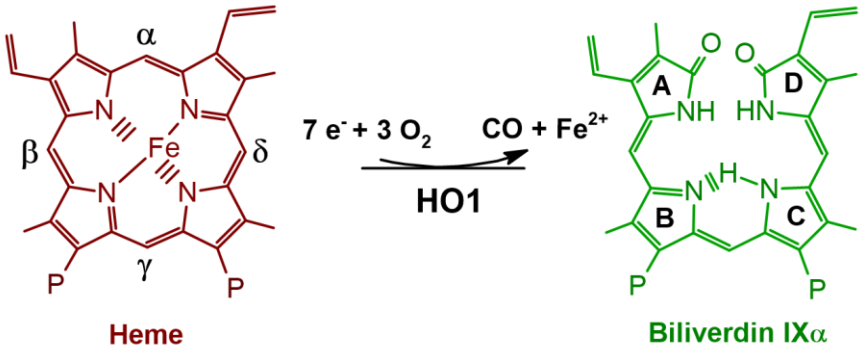


Figure 5: Simplified reaction mechanism of heme oxygenase. The ring-opening of the circular heme macrocycle is catalyzed by most HOs via the cleavage at the α-meso carbon bridge. The reaction requires three molecules of oxygen and seven electrons to yield the synthesis of biliverdin IXα. P= propionate side chain.

In general, the Ho1 reaction requires three molecules of O_2 and seven electrons to produce BV IX, CO, and Fe^{2+} (Figure 5). The required electron donors can vary depending on the organisms from which the HO originates. Mammalian and pathogenic bacterial HOs often use NADPH-cytochrome P450 reductase as the electron donor, contrary to bacterial, plant, and algae-derived HOs that take the electrons from reduced ferredoxins (Muramoto *et al.,* 1999, Rhie & Beale 1992, Schacter *et al.,* 1972, Wilks & Schmitt 1998). During the degradation reaction, heme acts as the substrate and the cofactor in it (Wilks & Heinzl 2014). The ferric-BV IX is the final product, while some HOs further reduce the iron to convert it to the ferrous state (Chu *et al.,* 1999, Wilks 2002). The rate-limiting step of this reaction is releasing the product from the active site in the absence of proteins able to bind BV IX. This is postulated to be driven forward by proteins able to bind BV IX like FDBRs, e.g., mammalian biliverdin reductase is capable of synthesizing bilirubin IXα out of BV IX as part of hemolysis (Liu & Ortiz de Montellano 2000). BV IX can serve in phytochromes as a light-sensing chromophore or as the precursor for synthesizing other light-harvesting and light-sensing proteins.

Ferredoxin-dependent bilin reductase (FDBR)

FDBRs found in cyanobacteria, higher plants, and red algae subsequently convert the precursor BV IXα to light-harvesting or chromophore precursors such as PCB, phytochromobilin (PΦB), or PEB. According to the number of electrons transferred during the site-specific reduction the FDBRs are grouped, resulting in two- and four-electron reduction groups (Dammeyer *et al.,* 2008a, Frankenberg *et al.,* 2001, Frankenberg & Lagarias 2003, Kohchi *et al.,* 2001). The electron donor is ferredoxin

(Beale & Cornejo 1991), which enables FDBRs to yield a variety of colorful tetrapyrroles, such as the pink-colored PEB and the blue-colored PCB (Figure 6). Different intermediates characterize the two routes of synthesizing PEB and PCB, while conserved characteristic amino acid residues catalyze the reaction.

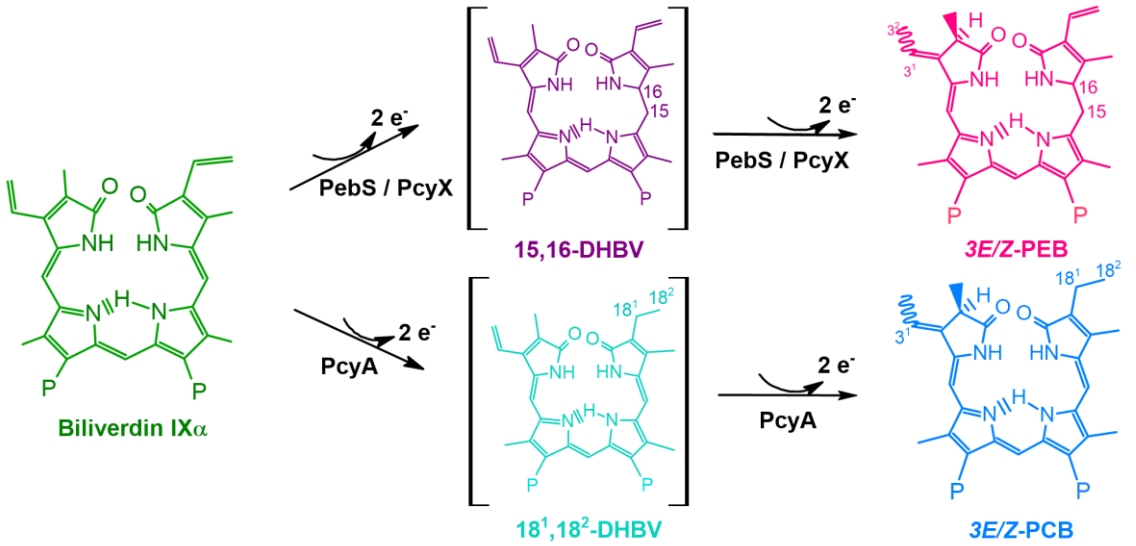


Figure 6: Tetrapyrrole biosynthesis of selected members of FDBR enzymes. All in common is the usage of biliverdin IXα as substrate. P = propionate side chains; 3*E/Z* = respective isomers. Dihydrobiliverdin (DHBV) is shown in brackets as this form is not released from the enzyme but an intermediate of the reaction. The blue pigment PCB is yielded by a PcyA-mediated four-electron reduction while the pink pigment PEB can be synthesized by the four-electron reduction by PebS or PcyX.

The catalyzing enzyme reactions of FDBRs are comparable, but there is only a minor sequence similarity between them. However, the occurrence of a homologous structure characterizes the entire FDBR enzyme family consisting of an α/β/α sandwich fold. Crystal structures revealed that all FDBRs are globular single-domain proteins with a central antiparallel β-sheet with two α-helices, one on each side (Dammeyer *et al.,* 2008b, Sommerkamp *et al.,* 2019, Sugishima *et al.,* 2020, Tu *et al.,* 2007). Thus, to determine the function of these enzymes, they often have to be analyzed *in vitro*.

The FDBR phycocyanobilin:ferredoxin oxidoreductase (PcyA; EC 1.3.7.5) catalyzes the four-electron reduction of BV IXα to red-absorbing PCB with the intermediate 18^1,18^2-dihydrobiliverdin (DHBV) (Figure 6) (Frankenberg *et al*., 2001, Frankenberg & Lagarias 2003). The substrate binds to PcyA in a cyclic configuration with A and D rings buried within the enzyme and the propionate side chains orientated toward the solvent (Frankenberg & Lagarias 2003, Tu *et al.,* 2004). Through a sterical hindrance, the exovinyl reduction of the D-ring can happen before the A-ring reduction can occur. Thus, the

enzyme consists of two separate activities: a 18^1, 18^2-DHBV:ferredoxin oxidoreductase and a PCB:ferredoxin oxidoreductase (Figure 6). These vinyl group reduction and diene system reduction proceed through bilin radical intermediates buried inside the enzyme so as not to react with molecular oxygen, which would generate reduced oxygen species (i.e., O_2^- and H_2O_2).

Meanwhile, to synthesize PEB, two enzymes are required for the BV IXα conversion into PEB. First, 15,16-dihydrobiliverdin:ferredoxin oxidoreductase (PebA, EC 1.3.7.2) reduces BV IXα to 15,16-DHBV, which is further reduced into PEB by phycoerythrobilin:ferredoxin oxidoreductase (PebB, EC 1.3.7.3). However, enzymes originating from phages showed the conversion of BV IXα to PEB catalyzed by a single enzyme, namely phycoerythrobilin synthase (PebS or PcyX, EC 1.3.7.6) (Figure 6) (Dammeyer *et al*., 2008b, Ledermann *et al.,* 2016). The viral enzymes PcyX and PebS catalyze the synthesis of the green-absorbing PEB via two independent mechanisms, having the same intermediate 15,16-DHBV and using four-electron reduction (Figure 6). Even though both enzymes produce PEB, they only share 9 % sequence identity while showing a globular α/β/α sandwich fold as described before (Dammeyer *et al*., 2008b, Ledermann *et al.,* 2018).

PebS combines the two enzymatic activities of PebA and PebB of cyanobacterial PEB synthesis in one protein. It uses a substrate radical mechanism where two conserved aspartate residues are located at the distal and proximal side of the substrate on the β-sheet, which are important for substrate protonation and conversion. PcyX is lacking these residues but possesses essential catalytic residues of PcyA. For PebS, the first Asp residue on the proximal side is essential for the first reduction (BV IXα to 15,16 DHBV), while the second aspartate residue on the distal side catalyzes the following reduction to PEB. This second Asp residue can also be found in PebB, which catalyzes a similar reduction of 15,16 DHBV to PEB (Busch *et al.,* 2011a). Interestingly, this second Asp residue is missing in PcyX and PcyA, where an Asn residue is located in the same position instead. The activity of PcyX is postulated to be similar to PcyA, where a proton channel is essential for the reprotonation of the first Asp residue, facilitated by a His residue. This catalytic important His residue is restricted to PcyX and PcyA sequences. Additionally, phylogenetic analysis showed that PcyX forms a distinct clade near PcyA

and far away from PebS, thus strengthening the evidence of two independent mechanisms of PebS and PcyX to produce PEB (Ledermann *et al*., 2018). Due to the spatial orientation of the C3 atom, *E* and *Z* isomers of the phycobilins are possible.

1.2.4 The physiological function of bilins and heme oxygenases

The different products resulting from the activity of heme oxygenases and FDBRs are essential for multiple physiological processes, e.g., light-harvesting and -sensing. Light-harvesting is crucial for photosynthesis and ATP generation in phototrophic organisms, while light-sensing can also be used by non-photosynthetic organisms.

Light-harvesting

Common for light-harvesting and sensing is the bound chromophore to a protein, which consists of a linear tetrapyrrole derived from heme. Therefore, a heme oxygenase must open the heme ring, forming BV IXα. This can now either be bound to a bacteriophytochrome for light-sensing or be further reduced by FDBRs to produce light-absorbing chromophores such as the green-light absorbing PEB or the red-light absorbing PCB. FDBRs are characterized in cyanobacteria, where their product acts as specialized light-harvesting chromophores bound by thioether linkages to proteins, thus forming phycobiliproteins. These are part of the significant light-harvesting antenna complexes, the phycobilisomes (PBSs). PBSs significantly increase the performance of oxygenic photosynthesis by covering the green gap of chlorophyll absorbance. They enable photosynthesis even in different water depths characterized by decreasing light intensity gradients and differentially filtered light colors (Glazer 1988, Kehoe 2010).

Interestingly, the bilin chromophores require the assistance of a bilin lyase to be attached to the cysteine of the apoprotein, thus forming the chromophorylated phycobiliprotein. The different chromophores bound to the α- and β-subunit of a phycobiliprotein enable light absorption at specific wavelengths based on their unique spectral properties. Multiple phycobiliproteins are then stacked together by a linker and form the rods and the core of the PBS. While the core consists of allophycocyanin (APC λmax 650-655 nm), the rods can consist of phycobiliproteins with different spectral

properties like phycocyanin (PC λmax 610-620 nm) and phycoerythrin (PE λmax 540-570 nm) or phycoerythrocyanin (PEC λmax 570-595 nm) (Bryant *et al.,* 1976, Moraes & Kalil 2009, Nies & Wehrmeyer 1980, Rippka *et al.,* 1979, Sonani *et al.,* 2014). The PBS is located on the stromal side of the thylakoid membrane above the reaction center of photosystem II and I (PSII, PSI). PBSs can harvest light by the absorption of light-energy. The energy is channeled from the rods to the PBS core and finally to the reaction center of the photosystem in the thylakoid membrane, ultimately reaching chlorophyll *a,* another tetrapyrrole.

Light-sensing

The second function of bilins is their occurrence as part of light-sensing phytochromes. Phytochromes were originally described as red and far-red light photochromic proteins (Butler *et al.,* 1959). Generally, they can be described as photochromic photoreceptors consisting of a bilin-bound protein capable of sensing light, which results in a signaling cascade. Signaling involves detecting a physical input (light) transducing into a biochemical output. Hereby, the phytochrome is switching between active and inactive forms because of a reversible photoisomerization. Due to irradiation, the bilin chromophore is subjected to photoisomerization at the 15,16 double bond between ring C and D, resulting in the flip of the D ring (Andel *et al.,* 1997). This leads to further sterical rearrangement of the chromophore. Consequently, the protein triggers the output signal (Figure 7).

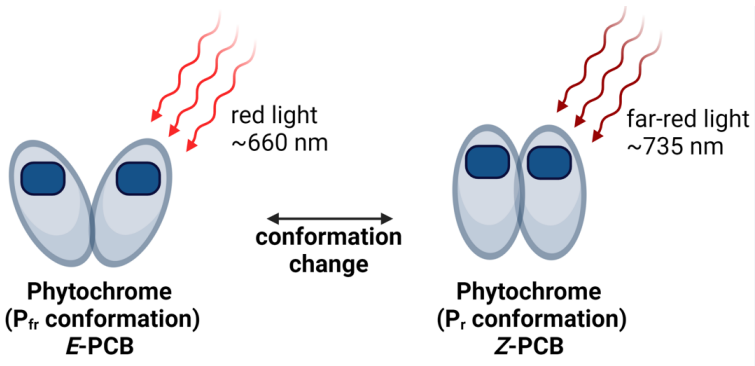


Figure 7: Illustration of PCB binding phytochrome conformation change upon illumination. The phytochrome dimer undergoes conformation change after certain light illumination, which results in sterical change of the bound chromophore (depicted as blue box) from *E*- to *Z*-isomer and *vice versa*. Figure created with BioRender.com.

Most phytochromes have a conserved domain architecture consisting of PAS-GAF-PHY (PAS = PerArntSim; GAF = cGMP-specific phosphodiesterases, cyanobacterial adenylate cyclases, and formate hydrogen lyase transcription activator FhlA; PHY = phytochrome-specific) (Mandalari *et al.,* 2013, Nagano 2016). The covalently attached chromophore can be located in the GAF domain and consists of a linear tetrapyrrole chromophore that is typically a BV IXα but can also be a PCB or PΦB, which is bound via a thioether bond to a cysteine residue in the protein. Typically, a histidine kinase domain is C-terminally located, responsible for the signal transduction pathways that lead to changes in different biological outputs after a photoisomerization event. This domain architecture enabled the identification of phytochrome-related sensors present in photosynthetic and nonphotosynthetic bacteria, fungi, plants, and unicellular green algae (Blumenstein *et al.,* 2005, Davis *et al.,* 1999, Froehlich *et al.,* 2005, Hughes & Lamparter 1997, Jiang *et al.,* 1999, Rockwell *et al.,* 2014). Phytochromes and phytochrome-related sensors from different organisms vary in this PAS/GAS/PHY domain architecture, and other output domains are also known to associate with it instead of histidine kinases (Ulijasz *et al.,* 2008).

Evolutionary related to phytochromes regarding their GAF domain are cyanobacterial phytochrome-related photosensors (CBCRs), but most of them lack conserved PAS or PHY domains (Ikeuchi & Ishizuka 2008). They extend the photosensory range to shorter wavelengths of visible light compared to regular phytochromes, as multiple blue/green CBCRs have already been described (Ikeuchi & Ishizuka 2008, Rockwell *et al.,* 2011, Ulijasz *et al.,* 2009, Yoshihara *et al.,* 2004). CBCR domains play key roles in complementary chromatic adaptation (green/red light adaptation) or other photobiological processes like phototaxis (Ikeuchi & Ishizuka 2008, Kehoe & Grossman 1996, Terauchi *et al.,* 2004). Moreover, the isomerization of PCB to the related phycoviolobilin (PVB) was demonstrated by a CBCR to enable sensing of green light (Ishizuka *et al.,* 2007). Light-sensing and -harvesting are essential for energy metabolism and adaption to environmental changes, while relying on the biosynthesis of specific linear tetrapyrroles derived from heme. These processes are often targeted by viruses that infect the host organism. Here, phages are known to harbor their own viral encoding FDBRs, e.g., PcyA, PebS, and PcyX, capable of synthesizing PCB and PEB.

1.3 Phages with AMGs are key players in aquatic ecosystems

Viruses are one of the most abundant biological entities on the planet (Suttle 2007) and almost always exceed the number of bacteria by a factor of 10 in the studied environments of air, soil, and water (Rastrojo & Alcamí 2017). This large genetic reservoir remains mostly unexplored, as well as its impact and importance in different environments. Through the development of sequencing technology and bioinformatic analysis in the last couple of years, a glimpse of the complex microbial communities and the role of viruses in different environments could be gained.

The ocean comprises the largest ecosystem on the planet, and the microorganisms present constitute 90 % of the ocean biomass, making them the main players in nutrient and energy cycles (Suttle 2007). Regarding this, the role of viruses infecting them can be divided into three aspects: 1. Viruses keep microbial levels in check and regulate ecosystems; they are essential for shaping microorganism communities by being the main predator of the planktonic community. 2. The lysis caused by viruses increases nutrient availability for other microorganisms, influencing the carbon shunt cycle. 3. Viruses can modulate the host metabolism and DNA composition through horizontal gene transfer and lysogenic conversion (Rodriguez-Brito *et al.,* 2010, Suttle 2007). Interestingly, host genes could be detected in bacteriophages that modulate host metabolic pathways. These so-called auxiliary metabolic genes (AMGs) are thought to redirect host metabolic pathways and resources to enhance phage survival and reproduction by actively manipulating the host metabolism (Howard-Varona *et al.,* 2020, Jacobson *et al.,* 2021). Among others, AMGs were reported to alter various metabolic pathways like photosynthesis (Fridman *et al.,* 2017, Lindell *et al.,* 2005, Mann 2003, Sharon *et al.,* 2009), carbon metabolism (Millard *et al.,* 2009, Thompson *et al.,* 2011), energy production (Rihtman *et al.,* 2019), and pigment biosynthesis (Dammeyer *et al*., 2008a, Ledermann *et al*., 2016, Ledermann *et al*., 2018).

The possession of AMGs in viruses is of current interest in research. The AMG composition was found to vary with respect to viral lifestyle, habitats, and hosts. Even within each lifestyle, the AMG composition was shown to be habitat-dependent and specific to the host infected (Luo *et al.,* 2022). While lysogenic viruses often encode AMGs related to physiology regulation to enhance virus-host symbiosis, lytic viruses

harbor more genes associated with boosting nutrient metabolism to benefit viral replication (Luo *et al*., 2022). In general, AMGs in phages are thought to exploit host resources, thus enhancing phage progeny release. This concept challenges the traditional view that phage infection leads only to phage replication, followed by the rapid death of the host and phage progeny release. The presence of AMGs demonstrates the complex interplay between viruses and their hosts in microbial ecosystems, which is facilitated through horizontal gene transfer (HGT) events.

Interestingly, the genes encoding for enzymes responsible for creating this wide variety of chromophores, the FDBRs, are also present in the genome of viruses, especially cyanophages (phages infecting cyanobacteria), and are thought to derive from their cyanobacterial host counterpart or ancestral bacterial host. Thus far, three distinct phage-encoded FDBR enzymes are known, namely PebS (Dammeyer *et al*., 2008b), PcyX (Ledermann *et al*., 2016), and PcyA (Dammeyer *et al*., 2008a). While cyanobacteria also encode PcyA, PebS is found only in cyanophages, and PcyX has been found so far solely in uncultured marine phages and the recently isolated pelagibacter phage Mosig (Buchholz *et al.,* 2021). Also, the presence of a *ho1* encoding for an active heme oxygenase on some phage genomes harboring FDBRs was verified, which were located next to each other (Buchholz *et al*., 2021, Dammeyer *et al*., 2008a). However, the occurrence of *pcyX* is still enigmatic as *pcyX*-carrying phages are likely to have an alphaproteobacterial host, like Mosig. Alphaproteobacteria are not known to harvest light using phycobilins (Imhoff 2006). Therefore, PcyX has been proposed to play a role in maintaining metabolic flux through the tetrapyrrole pathway, which in turn will also steadily release iron from the HO reaction, serving as an internal source of iron for reproduction (Ledermann *et al*., 2016, Ledermann *et al*., 2018). In contrast to its iron-regulating features, PcyX could provide the chromophore for phytochrome assembly, thereby supporting the host in the adaptation to environmental light conditions.

AMGs are primarily found in metagenomic data in addition to cultured and sequenced host-phage cultures. Therefore, it is an essential resource for discovering new AMGs to characterize their impact, environment, and distribution in phages.

1.4 Viral metagenomes as a source for new AMGs

Analysis of viral metagenomic data is crucial for future research on novel viral enzymes that could harbor environmental or medical therapeutic significance, among others. As the acquisition of metagenomic data does not require the cultivation and isolation of the phage and its host, it is also cost-efficient and generates large data repositories. Bypassing this limitation of cultivation is crucial as only around 1 % of bacterial species have been cultured (Pham & Kim 2012), which explains why the number of sequenced genomes of viruses is 10-fold lower than bacterial genomes (Rastrojo & Alcamí 2017). With exponentially growing numbers of metagenomic samples primarily focused on bacteria and not viruses, the aquatic environment is the most characterized habitat. This is also true for total viral reads, especially ocean samples, which comprised ~40 % of the total viral reads in Metavir (Rastrojo & Alcamí 2017, Roux *et al.,* 2016). However, the viral reads of other aquatic environments like freshwater, thermal spring water, sewage water, river estuary, and reprocessed water remain elusive and are currently gaining researchers' interest. Terrestrial habitats are beginning to be uncovered in terms of their rich environmental metagenomic samples. Recently, a functional viral chitin-degrading enzyme was characterized in this habitat, and its crystal structure was solved (Wu *et al.,* 2022). Other habitats, including air, remain to be investigated.

The main form of viruses detected metagenomically are dsDNA viruses that infect bacteria due to technical biases during sample preparation, thus neglecting single-stranded viruses, RNA viruses, and viruses infecting other domains of life. RNA viruses were recently proven to be just as abundant as DNA viruses. However, other viruses, such as ssDNA phages, are also overlooked because of their difference from dsDNA phages. This leads to experimental bias during the sample preparation, filtering them out. Thus, their ecological distribution and role are mostly unexplored (Steward *et al.,* 2013, Székely & Breitbart 2016). Further challenges in the identification of new viral genomes are the absence of the universal marker genes for viruses, which is established for prokaryotes and eukaryotes using ribosomal genes (16S and 18S) or for fungi by an internal transcribed spacer. There are rather diverse metagenomic marker genes for viruses to study viral diversity, as they are not always present in all viral families. Belonging to these marker genes are, e.g., RNA-dependent RNA polymerase gene (Culley

et al., 2003), major capsid protein (*gp23*) gene from T4-like phage (Filée *et al.,* 2005) or neck protein (*gp13*) gene that often clusters with photosynthetic genes and is used for phylogenetic analysis (Béjà *et al.,* 2012).

In order to unravel viral diversity and its impact on environmental communities and biological and biogeochemical cycles, further metagenomic studies have to be accomplished. This will be facilitated by reduced costs, new emerging sequencing techniques, and new data analysis tools, eliminating the previously mentioned biases. Also, the genetic material of the complete microbial community should be examined to obtain valuable insights into environmental communities.

1.5 Aims

Earlier investigations of marine metagenomic datasets uncovered a conserved viral gene cassette featuring *ho1* and FDBR genes known for encoding functional enzymes utilized in cyanobacteria for light-harvesting in oxygenic photosynthesis. Notably, the phage is predicted to infect alphaproteobacterial hosts, which do not utilize such light-harvesting systems (Ledermann *et al.*, 2016). Using new metagenomic datasets that emerged, a novel conserved AMG cluster of viruses present in freshwater and marine samples was identified (Figure 8). This cluster encodes for the putative enzymes Ho1, an FDBR (most likely PcyA or PcyX, named PcyA/X), and ALAS, suggesting potential involvement in pigment and heme biosynthesis of this virus when infecting its host. As ALAS is exclusively present in the bacterial domain of alphaproteobacteria, the focus lies in investigating the biochemical function of the novel putative ALAS originating from uncultured phages. Further, the Ho1 and PcyA/X characterization was carried out biochemically as these are the first freshwater-derived viral enzymes of their kind.

This is the first report of a phage-encoded *alaS* and its occurrence in a cluster with cyanophage-associated AMGs from marine and freshwater environments, suggesting a potential role in pigment biosynthesis outside of oxygenic phototrophs.

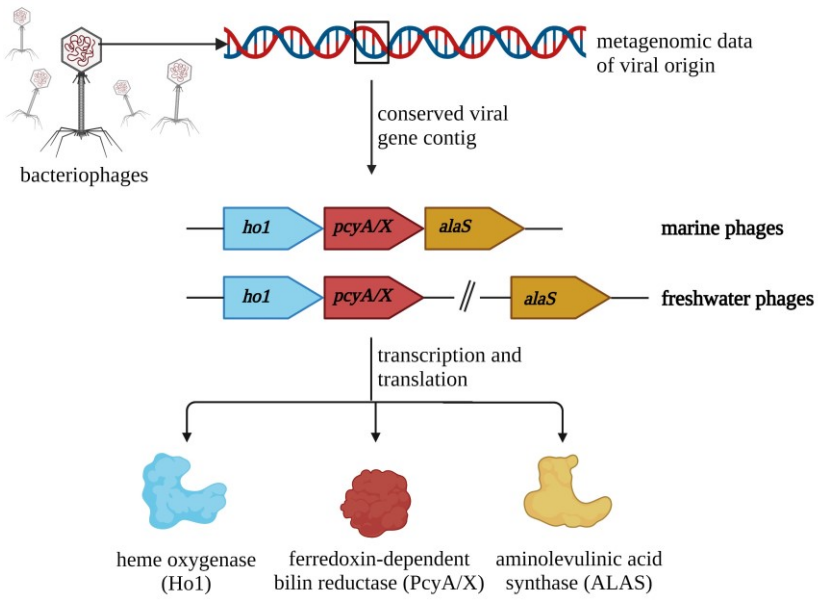


Figure 8: Viral metagenomic cluster of AMGs encoding for enzymes responsible in tetrapyrrole biosynthesis. The common tetrapyrrole precursor ALA is synthesized by the enzyme aminolevulinic acid synthase (ALAS). Heme can then be converted via heme oxygenase to linear biliverdin IXα which can be used by FDBRs like PcyA or PcyX to form phycocyanobilin (PCB) or phycoerythrobilin (PEB) respectively. Marine and freshwater phages have different genomic organizations of the three gene cassette, in freshwater phages *alaS* is always located further away from the cassette. Figure created with BioRender.com.

2. Material and methods

2.1 Materials

2.1.1 Devices

Table 1: Used Equipment

Type of instrument	Name	Manufacturer
Agarose gel electrophoresis	Com Phor L Mini Com Phor L Midi	Biozym GmbH
Autoclave	VX 150	Systec GmbH
Blotting Equipment	Semidry Blot Trans-blot® SD	Bio-Rad Laboratories GmbH
Centrifuges	Centrifuge 5415D Centrifuge 5810R Z32HK Sorvall Lynx 6000	Eppendorf AG Hermle Labortechnik GmbH Thermo Fischer Thermo-Fischer
Cell disruption	LM10 Microfluidizer	Microfluidics
Incubator Shaker	New Brunswick Innova® 44 Universal Shaker SM 30-control	Eppendorf AG Edmund Bühler GmbH
Fluorescence spectrometer	FP-8300	Jasco Deutschland GmbH
Gel documentation	GeliX20 Imager	Intas Science Imaging Instruments GmbH
HPLC	1100 Series	Agilent
Lyophyll/Freeze-dryer	Alpha 2-4 LSC plus	Martin Christ
Microtiter plate reader	Tecan Infinite® F200 Pro	Tecan Group AG
pH meter	Basic pH Meter p-11	Sartorius AG
Photometers	8453 UV visible system Nanodrop™ Lite Novaspec III	Agilent Technologies Thermo Scientific Bio-Rad Laboratories GmbH
Power Supply	PowerPac 300	Bio-Rad Laboratories GmbH
Scales	AccuLab Research	Sartorius AG
SDS-PAGE	Mini-Protein® Tetra cell system	Bio-Rad Laboratories GmbH
Shaker	MR Hei-Standard	Heidolph Instruments GmbH & Co. KG
Thermocycler	T1 Thermocycler	Biometra GmbH
Ultrasonic homogenizer	HD 2200 with tip KE 76	Bandelin GmbH & Co. KG
Ultra-pure water system	MilliQ® Integral Water Purification System	Merck KGaA

UV-Vis spectrophotometer	8453 UV-Visible System	Agilent
Vacuum Pump	KNF LABOPORT® Series Laboratory Vacuum Pumps	KNF Neuberger, Inc.
Vortexer	Vortex Shaker	VWR International GmbH

2.1.2 Chemicals and reagents

Unless otherwise indicated, all of the chemicals and reagents utilized in this study were of ACS grade or higher. They were acquired from New England Biolabs (Schwalbach), AppliChem (Darmstadt), Carl Roth (Karlsruhe), Merck (Darmstadt), Sigma Aldrich (Munich), Fluka (USA) and Serva (Heidelberg).

2.1.3 Enzymes, kits, antibodies, and special chemicals

Table 2: Special chemicals and materials

Product	Name	Manufacturer
Column material for affinity chromatography	Protino™ Glutathione-Agarose 4B	Macherey - Nagel GmbH & Co. KG
	Ni-NTA Agarose	Qiagen
	Talon®Superflow™	Cytiva
Centrifugal concentrator	Amicons Ultra, Cellulose 10000 NMWL	Diagonal/Merck
Dialysis tube	Dialysis visking, cellulose MWCO 10000	Roth
DNA loading dye	DNA Gel Loading Dye (6x)	Thermo Scientific
DNA ladder	GeneRuler™ DNA ladder 1kb plus	Thermo Scientific
Protein standards	Unstained Protein Standard, Broad Range	NEB
	Color Prestained Protein Standard, Broad Range	NEB
PVDF membrane	Roti®-PVDF-Membrane	Carl Roth
Sep-Pak® light C18		Waters
Sterile filter	PES, 0.45	Starlab
	PVDF, 0.2 µm	Diagonal
	PVDF, 0.45 µm	

Table 3: Enzymes and kits

Product	Name	Manufacturer
DNase I	DNase I	AppliChem
DNA ligase	T4-DNA Ligase	Thermo Scientific
DNA polymerase	Phusion High-Fidelity DNA polymerase	Produced in this laboratory
Lysozyme	Lysozyme	Carl Roth
PCR clean-up kit	NucleoSpin® Gel and PCR Clean-up	Macherey - Nagel GmbH & Co. KG
	DNA Clean & Concentrator™ -5	Zymo Research Europe GmbH
Plasmid miniprep kit	NucleoSpin® Plasmid EasyPure	Macherey - Nagel GmbH & Co. KG
	E.Z.N.A.® Plasmid DNA Mini Kit I	Omega Bio-Tek Inc.
Protease	PreScission protease	Produced in this laboratory
Restriction endonucleases	FastDigest restriction endonucleases	Thermo Scientific
	High Fidelity restriction endonucleases	New England Biolabs Inc.

Table 4: Antibodies

Antibody	Antigen	Manufacturer
Goat anti-GST antibody (1:10 000)	GST-tag	Pharmacia Biotech
Rabbit anti-goat IgG alkaline phosphatase (1:30 000)	Goat IgG	Immuno research
Mouse 6x-His-Tag Monoclonal Antibody (HIS.H8) (1:3 000)	His_6-tag	Invitrogen
Goat anti-mouse IgG-alkaline phosphatase (1:10 000)	Mouse IgG	Sigma-Aldrich

Table 5: Antibiotics

Antibiotics	Solvent	Stock conc.	Working conc.
Ampicillin	50 % EtOH	100 mg/ml	100 µg/ml
Chloramphenicol	70 % EtOH	34 mg/ml	34 µg/ml
Gentamycin	dH_2O	20 mg/ml	20 µg/ml
Kanamycin	dH_2O	50 mg/ml	50 µg/ml
Streptomycin	dH_2O	50 mg/ml	50 µg/ml
Tetracycline	EtOH	25 mg/ml	25 µg/ml

2.1.4 Viral sequences

Sequences were derived from different metagenomic datasets and publications.

The freshwater contigs were derived from German lakes located in Berlin and from the publication of (Chen *et al.,* 2020). The latter harbors sequences from freshwater reservoirs like lakes or bogs in the United States, Canada, and Switzerland. The annotation of the metagenomic data from the lakes in Berlin was carried out by our collaborator Anne Kupczok (Wageningen University & Research, Netherlands), and the sequences were kindly provided by Hans-Peter Grossart and Jason Woodhouse (Leibniz Institute of Freshwater Ecology and Inland Fisheries, Stechlin and Potsdam University; University Hamburg, Germany). Four lakes near Berlin were sampled between 2003 and 2020 for the time series data from German lakes (LIMNOS lake dataset PRJEB47226), yielding 1096 metagenome samples. Individual samples have been assembled and sequenced (manuscript in preparation; unpublished data). First, contigs measuring at least 500 bp in length were examined for virus signals using VIBRANT (Kieft *et al.,* 2020). After that, CheckV was used to evaluate the quality of each hit; only virus sequences rated as medium or high quality were kept (Nayfach *et al.,* 2021). The resulting 50,913 viral contigs were annotated using geNomad (Camargo *et al.,* 2023) and scanned against a protein database of previously identified ALAS sequences using Diamond v2 (Buchfink *et al.,* 2021). In total, this procedure yielded twelve contigs with a respective hit. The respective sequences used in this dissertation can be found in the Appendix.

The collaborators Sheila Roitman (Max Planck Institute for Biology Tübingen, Germany) and Oded Béjà (Technion-Israel Institute of Technology Haifa, Israel) found and annotated the marine contigs. They were derived from the Tara Oceans and re-assembled as described elsewhere (Philosof *et al.,* 2017, Roitman *et al.,* 2018). Using TBLASTX (Altschul *et al.,* 1990, Camacho *et al.,* 2009) with default parameters, viral fatty acid desaturase sequences from a prior study (Roitman *et al*., 2018) were used as queries to find scaffolds with similar genes in the reassembled Tara Ocean database and publicly accessible datasets from the JGI GOLD (Mukherjee *et al.,* 2023). Based on gp13 phylogeny, one scaffold most likely derived from a pelagiphage encoded the ALAS, Ho1, an FDBR, and a fatty acid desaturase. The virally encoded ALAS was deployed as bait in

the same databases using the same process to obtain more relevant contigs. GeneMark (Besemer & Borodovsky 1999, Zhu *et al.,* 2010) was used to identify ORFs in the contigs, and BLASTX was used with default parameters to manually annotate the putative genes in the contigs against the NCBI non-redundant (nr) protein database.

The viral sequences were either used as original versions without codon adaption or with codon adaption using diverse adaption tools, e.g., Backtranseq (EMBL-EBI), JCat (Java Codon adaption tool), JGI BOOST Juggler Balanced (Joint Genome Institute). The marine and freshwater *alaS* sequences were used as original and codon-adapted versions. From the contig '3300009785_Ga0115001_100000', the following sequences were derived: ALASsyn, ALASorg, ALASopt, and ALASart. The DNA sequence was at the beginning not known for this ALAS; therefore, ALASsyn was created from the protein sequence, which was backtranslated using EMBOSS Backtranseq (EMBL-EBI) and then adapted for the codon usage of the *E. coli* K12 strain (Madeira *et al.,* 2022). The original DNA sequence of ALASorg was used to derive the coding sequence of ALASsyn (not codon-adapted). In contrast, ALASopt was generated by adapting the codon-usage of the original ALASorg DNA sequence for the *E. coli* K12 strain using the codon adaptation tool of JCat (Grote *et al*., 2005). ALASart was created by hand using the codon usage and occurrence of *Rhodobacter capsulatus* (*R. capsulatus*) ALAS, and the occurrence of the codons of the viral ALASsyn was adopted. The codon usage of the viral ALASsyn was compared to the occurrence and codon usage of *R. capsulatus* ALAS, and identical amino acids with deviating codon triplets were adapted to the *R. capsulatus* ALAS. From the contig 'SI39nov09', the ALAS sequences SI and SIopt were derived. Where SI is the original sequence of ALAS, SI opt is the codon adapted one using the Online tool JGI BOOST Juggler with the setting 'balanced', which ensures a harmonized occurrence of codon usage for *E. coli* (Oberortner *et al*., 2017). The contig '3300009790_Ga0115012_10000076 761838' harbored an ALAS, referred to as '33'. The sequence was not codon adapted for this work's expression- and enzyme activity studies. The respective sequences used in this dissertation can be found in the Appendix.

The freshwater metagenomic viral sequence of ALAS was derived from the contig crystal-bog-phage-scaffold_2_498, which is present on the assembled phage genome

named Crystal_bog_pmoC-phage_2 (Chen *et al*., 2020). This phage (metagenomic sequence) was found in samples derived from a peat bog named Crystal Bog (Madison, WI, United States) (Chen *et al*., 2020). This sequence was used as an original and codon-adapted version. The codon adaption was carried out using the JGI BOOST Juggler Balanced. The v*ho1* and v*pcyA* were derived from the same phage and only used as codon-adapted versions using the JGI BOOST Juggler Balanced.

All viral sequences were chemically synthesized (Twist Bioscience, San Francisco, CA). Lyophilized synthetic genes were resuspended in autoclaved TE buffer (pH 8.0) and stored at -20 °C.

TE buffer:

10 mM Tris pH 8.0 (HCl)

1 mM EDTA pH 8.0 (NaOH)

2.2 Microbial strains

2.2.1 *E. coli*

Cloning procedures were conducted in *E. coli* DH5α. The heterologous protein production was performed in *E. coli* strains described below (Table 6). Liquid cultures of *E. coli* were grown in Luria Bertani (LB) Lennox with the appropriate antibiotics and supplements. Test expressions were carried out in standard media, 2YT, Minimal media M9 (MM), or Autoinduction media (AI) supplemented with the appropriate antibiotics. In order to prepare solid media for the cultivation of *E. coli,* 1.5 % (w/v) agar-agar was added to the media before sterilization.

LB-medium		**2YT-medium**	
NaCl	5 g/l	NaCl	5 g/l
Tryptone	10 g/l	Tryptone	16 g/l
Yeast extract	5 g/l	Yeast extract	10 g/l

Minimal medium M9 (Maniatis *et al.,* 1987)

For the preparation of Minimal media (MM), different stock solutions were prepared separately and sterilized by either autoclaving or filter sterilization. The final stocks were mixed as described below and adjusted to one liter with sterile water. $MgSO_4$ stock solution (1 M) was prepared separately and sterilized by autoclaving. Glucose stock solution (1 M) was sterilized by passing it through a 0.22 µm filter, and the thiamin stock solution (1.5 M) which was stored at -20 °C. Yeast extract stock solution (10 % w/v) was autoclaved for sterilization.

5x M9 salts		200 ml
	Na_2HPO_4 x 7 H_2O	64 g/l
	KH_2PO_4	15 g/l
	$NaCl_2$	2.5 g/l
	NH_4Cl	5 g/l
1 mM $MgSO_4$ x 7 H_2O		1 ml
40 mM Glucose		40 ml
0.1 % Yeast extract		10 ml
1.5 mM Thiaminchlorid hydrochloride		1 ml
Fill up with H_2O to 1 l		

Autoinduction medium (Studier 2005, Studier 2014) adapted by Imperial College London

The Autoinduction medium was prepared by setting up the phosphate buffer and adding tryptone, yeast extract, and NaCl. The resulting medium was then autoclaved. Further, a sugar mix is required; therefore, stock solutions of glycerol, glucose, and lactose were prepared. The sugar mix is mixed as described below and filtered-sterilized before adding it to the previously prepared medium under sterile conditions. When using this media, it is crucial to adjust antibiotic concentrations for ampicillin 50 µg/ml, kanamycin 100 µg/ml, and chloramphenicol 35 µg/ml.

Phosphate buffer (pH 7.2)

Na_2HPO_4 / KH_2PO_4	6 g / 3 g per l
Tryptone	20 g/l
Yeast extract	5 g/l
NaCl	5 g/l

Filter-sterilize the following sugar mix and add sterile to the before-prepared media. The recipe corresponds to one-liter medium (above).

60 % (v/v) Glycerol	10 ml
10 % (w/v) Glucose	5 ml
8 % (w/v) Lactose	25 ml

Further media supplements, e.g., betaine and sorbitol, chloramphenicol, and ethanol, were added to the media before the inoculation with *E. coli* strains or before induction when the cells reached the desired optical density (OD) at 600 nm.

Table 6: Microbial strains of *E. coli.*

***E. coli* strain**	**Genotype**	**Reference**
Arctic Express (DE3) (B strain)	F⁻ *ompT hsdS*(r_B^- m_B^-) *dcm*⁺ Tet^R gal λ(DE3) *endA* Hte [cpn10 cpn60 Gen^R]	Agilent
BL21 (DE3) (B strain)	F⁻ *ompT gal dcm lon hsdS*$_B$($r_B^-m_B^-$) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*⁺]$_{K-12}$(λ^S)	(Studier & Moffatt 1986)
BL21 (DE3) CodonPlus RIL (B strain)	F⁻ *ompT hsdS* (r_B^- m_B^-) *dcm*⁺ Tet^R *gal* λ(DE3) *endA* Hte [*argU ileY leuW* Cam^R]	Stratagene
BL21 (DE3) pGro7 (B strain)	F⁻ *ompT gal dcm lon hsdS*$_B$(r_B-m_B-) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*⁺]$_{K-12}$(λ^S) pGro7 (Cam^R)	TaKaRa
Rosetta-gami (DE3) (K strain)	Δ*ara-leu*7697 Δ*lacX74* Δ*phoA PvuII phoR araD139 ahpC galE galK rpsL* (DE3) F'[*lac*⁺ *lacI*^{q} *pro*] *gor522::*Tn*10 trxB* pRARE (Cam^R, Kan^R, Str^R, Tet^R)	Novagen
Rosetta (DE3) pLysS (B strain)	F⁻ *ompT gal dcm lon? hsdS*$_B$(r_B^- m_B^-) λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*⁺]$_{K-12}$(λ^S) pLysSRARE[*T7p20 ileX argU thrU*	Novagen

	tyrU glyT thrT argW metT leuW proL ori$_{p15A}$](Cam^R)	
DH5α (K strain)	F$^-$1$^-$*supE44*D(*argF-lac)U169 j80dlac*ZΔM*15 hsdR17 recA1 endA1 gyrA96 thi-1 relA1*	(Grant *et al.,* 1990)
Shuffle T7 Express (B strain)	*fhuA2 lacZ::T7 gene1 [lon] ompT ahpC gal λatt::p*NEB3-r1-*cDsbC* (Spec^R*, lacI^q) ΔtrxB sulA11 R(mcr-73::miniTn10*--Tet^S *)2* [dcm] *R(zgb-210::Tn10* --Tet^S *) endA1 Δgor Δ(mcrC-mrr)114::IS10*	NEB
AD494 (DE3) (K strain)	*Δara-leu7697 ΔlacX74 Δ phoA PvuII phoR ΔmalF3* F'[*lac$^+$* (*lacI^q*)*pro*] *trxB::kan(DE3)*	Novagen
Origami (DE3) (K strain)	*Δ(ara-leu)7697 ΔlacX74 ΔphoA Pvu*II *phoR araD139 ahpC galE galK rpsL* F'[*lac$^+$ lacI^q pro*] (DE3) *gor522*::Tn*10 trxB* (Kan^R, Str^R, Tet^R)	Novagen
Tuner (DE3) (B strain)	F$^-$ *ompT hsdS$_B$*(r$_B^-$ m$_B^-$) *gal dcm lacY1* (DE3)	Novagen
S17.1λ*pir*	*pro, res$^-$ hsdR17* (r$_K^-$ m$_K^+$) *recA$^-$* with an integrated *RP4-2-Tc::Mu-Km::Tn7, Tp^r, λpir* lysogen	(De Lorenzo *et al.,* 1993, Simon *et al.,* 1983)
ST18	S17*.1λpir*Δ*hemA*	(Thoma & Schobert 2009)

pRARE and pLysSRARE contain the tRNA genes *argU, araW, ileX, glyT, leuW, proL, metT, thrT, tyrU* and *thrU*. These genes support the translation of rare codons, including AGG, AGA, AUA, CUA, CCC, and GGA (Novy 2001). Strains with the Str^R marker possess a mutation in ribosomal protein (*rpsL*) that confers resistance to streptomycin. However, streptomycin is not required to maintain the genotype of these strains.

2.2.2 Plasmids

The plasmids generated in this study (Table 7) were constructed by Gibson Assembly or T4-based ligation. The cloning was confirmed by DNA sequencing of the constructed plasmids using Sanger sequencing (Eurofins Genomics, Luxembourg) using the respective primer listed below (Table 8). Samples for sequencing were prepared according to the company's instructions.

Table 7: Plasmids

Plasmids	Description	Tag	Reference
pGEX-6P-1/3	Expression vector; *tac* promotor; IPTG inducible; PreScission Protease cleavage site between tag and gene of interest (goi); Amp^R	GST-tag (N-terminal)	GE Healthcare Life Sciences
pET28b	Expression vector; T7 promotor; IPTG inducible; thrombin cleavage site between tag and goi; Kan^R	His_6-tag (N-terminal)	Novagen
pCold-TF	Expression vector; *cspA* promotor; induction via cold shock and IPTG addition; PreScission Protease cleavage site between tag and goi; Amp^R	His_6-tag (N-terminal)	TaKaRa
pCYB1	Expression vector; *tac* promotor; IPTG inducible; Factor Xa cleavage site between intein and Chitin binding domain (CBD); Amp^R	Intein and CBD (C-terminal)	New England Biolabs
pACYC-Duet1	Duet expression vector; T7 promotor; IPTG inducible; Cm^R	1.MCS: His_6-tag (N-terminal) 2. MCS: S-tag (C-terminal)	Novagen
pRcA	derivate of pGEX-6P-1; *alaS* from *Rhodobacter capsulatus* SB1003 encoding ALAS	GST-tag (N-terminal)	(Hännig 2011)
pMkA	derivate of pGEX-6P-1; *gtrR* encoding a glutamyl-tRNA reductase from *Methanopyrus kandleri*	GST-tag (N-terminal)	(Moser *et al.,* 1999)
pGro7	Chaperone (*gro*ES-*gro*EL) plasmid for optimizing protein production; size of chaperones GroEL 60 kDa, GroES 10 kDa; *araB* promotor; induction with 0.05 mg/ml arabinose; Cm^R	No tag	TaKaRa

pET his-tag *cph1* crystal	pET-derivate; Phytochrome gene *cph1* from *Synechocystis* sp. PCC 6803; T7 promotor; IPTG inducible; Amp^R	His_6-tag	J. Clark Lagarias
pTD*ho1*	pACYC-Duet1 derivate; 1. MCS empty, 2. MCS: *ho1* from cyanophage P-SSM2	2.MCS: *ho1* S-tag (C-terminal)	Thorben Dammeyer
pGEX-*bphO*	Derivative of pGEX-6P-1; *bphO* encoding for a heme oxygenase from *Pseudomonas aeruginosa*	GST-tag (N-terminal)	(Wegele *et al*., 2004)
pGEX-*pcyA*-Anab	Derivative of pGEX-6P-1; *pcyA* from *Anabaena* sp. PCC 7120 (known as *Nostoc* sp. PCC7120)	GST-tag (N-terminal)	(Frankenberg *et al*., 2001)
Marine metagenomic viral sequences			
pGEX-6P-ALASsyn	derivate of pGEX-6P-3; *alaS* from marine metagenomic data from the contig 3300009785_Ga0115001_100000 encoding for ALAS; amino acid sequence was backtranslated using Backtranseq (EMBL-EBI) with codon usage of *E. coli* K12	GST-tag (N-terminal)	This work
pGEX-6P-ALASorg	derivate of pGEX-6P-3; *alaS* from marine metagenomic data encoding for ALAS; original sequence, is not codon adapted	GST-tag (N-terminal)	This work
pGEX-6P-ALASopt	derivate of pGEX-6P-3; *alaS* from marine metagenomic data encoding for ALAS; original sequence was codon adapted using JCat with codon usage of *E. coli* K12	GST-tag (N-terminal)	This work
pET28-ALASorg	derivate of pET28b; *alaS* from marine metagenomic data encoding for ALAS; original sequence, is not codon adapted	His_6-tag (N-terminal)	This work
pET28-ALASopt	derivate of pET28b; *alaS* from marine metagenomic data encoding for ALAS; original sequence was codon adapted using JCat with codon usage of *E. coli* K12	His_6-tag (N-terminal)	This work
pGEX-ALASart	derivate of pGEX-6P-3; *alaS* artificially created sequence from ALASsyn derived AS sequence, codon adapted to *R. capsulatus* ALAS codon usage	GST-tag (N-terminal)	This work

pCold-ALASart	derivate of pCold-TF; *alaS* artificially created sequence from ALASsyn derived AS sequence, codon adapted to *R. capsulatus* ALAS codon usage	His_6-tag + Trigger factor (N-terminal)	This work
pGEX-SI	derivate of pGEX-6P-3; *alaS* from marine metagenomic data from the contig SI39nov09 encoding for ALAS; original sequence, is not codon adapted	GST-tag (N-terminal)	This work
pGEX_SI opt	derivate of pGEX-6P-3; *alaS* from marine metagenomic data from the contig SI39nov09 encoding for ALAS; codon adapted using JGI BOOST Juggler Balanced	GST-tag (N-terminal)	This work
pGEX_33	derivate of pGEX-6P-3; *alaS* from marine metagenomic data from the contig 3300009790_Ga0115012_10000076761838 encoding for ALAS; original sequence, is not codon adapted	GST-tag (N-terminal)	This work
Freshwater metagenomic viral sequences			
pCold-ALAScry	derivate of pCold-TF; *alaS* derived from phage CB_2 (Chen 2020); original sequence, is not codon adapted	His_6-tag + Trigger factor (N-terminal)	This work
pCold-ALAScryopt	derivate of pCold-TF; *alaS* derived from phage CB_2 (Chen 2020) codon adapted using JGI BOOST Juggler Balanced	His_6-tagm + Trigger factor (N-terminal)	This work
pGEX-ALAScry	derivate of pGEX-6P-3; *alaS* derived from phage CB_2 (Chen 2020); original sequence, is not codon adapted	GST-tag (N-terminal)	This work
pGEX-ALAScryopt	derivate of pGEX-6P-3; *alaS* derived from phage CB_2 (Chen 2020) codon adapted using JGI BOOST Juggler Balanced	GST-tag (N-terminal)	This work
pCYB1-ALAScryopt	derivate of pCold-TF; *alaS* derived from phage CB_2 (Chen 2020) codon adapted using JGI BOOST Juggler Balanced; Stop codon present thus no intein fusion protein with ALAS will be translated	No tag	This work

pGEX-*ho1*-CB2	derivate of pGEX-6P-3; *ho1* derived from phage CB_2 (Chen 2020) codon adapted using JGI BOOST Juggler Balanced	GST-tag (N-terminal)	Vanessa Braun/Helen Wegner
pGEX-*pcyA*-CB2	derivate of pGEX-6P-3; *pcyA* derived from phage CB_2 (Chen 2020) codon adapted using JGI BOOST Juggler Balanced	GST-tag (N-terminal)	Vanessa Braun/Helen Wegner
pTD*ho1*_v*pcyA*	derivate of pTDho1; 1. MCS: *pcyA* from phage CB_2 (Chen 2020) codon adapted using JGI BOOST Juggler Balanced; 2. MCS: *ho1* from P-SSM2	1.MCS: *pcyA* (CB_2) His_6-tag (N-terminal) 2. MCS: *ho1* (P-SSM2) S-tag (C-terminal)	Vanessa Braun/Helen Wegner
pTDv*ho1*_v*pcyA*	derivate of pTDho1_v*pcyA*; 1. MCS: *pcyA* from phage CB_2 (Chen 2020) codon adapted using JGI BOOST Juggler Balanced; 2.MCS: *ho1* from phage CB-2 (Chen 2020) codon adapted using JGI BOOST Juggler Balanced	1.MCS: *pcyA* (CB_2) His_6-tag (N-terminal) 2. MCS: *ho1* (CB_2) S-tag (C-terminal)	This work
pCYB1_ALAScryopt_K250A	Quikchange mutagenesis for PLP binding site K250 to A	No tag	This work
pRcA_K248A	Quikchange mutagenesis for PLP binding site K248 to A	GST-tag	This work

The viral sequences of v*ho1* and v*pcyA* were derived from the same phage as ALAScryopt, namely Crystal_bog_pmoC-phage_2. The sequences were also codon-adapted and chemically synthesized, as stated before. The *E. coli* strains carrying the plasmid pCold-TF or its derivate should not be stored at 4 °C and should either be transformed freshly or streaked out of glycerol cryocultures from -80 °C.

2.2.3 Oligonucleotides

Table 8: Oligonucleotides

Primer	Sequence 5‘ → 3‘ orientation
Sequencing or amplification primer	
pGEX-fw	CCAGCAAGTATATAGCATGG
pGEX-rev	GCTTACAGACAAGCTGTGAC
T7_fwd	TAATACGACTCACTATAGGG
T7term/DuetDown2	GCTAGTTATTGCTCAGCGG
Twist Primer Set1_fwd	CCCCGTCACCTTTGGCTTATCAGT
Twist Primer Set1_rev	CGGTCTTAGCGTCCAGATGTCACT
Twist2_for	CAATCCGCCCTCACTACAACCG
Twist2_rev	CTACTCTGGCGTCGATGAGGGA
pCOLDrev	GCAGGGATCTTAGATTCTGTG
pCOLDfwd	GTAGTGGTGGTATCGAAGGTA
p*tac*_seq	GAGCGGATAACAATTTCACACAG
pACYCDuettUP1	GGATCTCGACGCTCTCCCT
Duetup2	TTGTACACGGCCGCATAATC
DuettDOWN1	GATTATGCGGCCGTGTACAA
ALAS_syn-seq_rev	ATGGAACACCTGGACAAATTC
ALAS_syn-seq_fwd	TTATTTGGTTTTACCGCAACGAC
ALASopt_fw	ATGGAACACCTGGACAAATTCA
ALASopt_rev	TTATTTGGTTTTACCGCAACGAC
ALAScontig_*hemA*_ fw (ALASorg)	ATGGAACATTTAGATAAATTTAAAAAAATAATAACAG
ALAScontig_*hemA*_ rev (ALASorg)	TCATTTAGTTTTACCACATCTTCG
ALAScry-fwd	ATGTTAAAATCAAAAACACAAGATGCT
ALAScry-rev	TTAAATATATTTGTCAATAGAATTTCTTAAAGAT
ALAScryopt-fwd	ATGCTTAAATCTAAAACCCAGGA
ALAScryopt-rev	TTATATATATTTATCAATGGAATTGCGA
Gibson Assembly or T4-based ligation	
pGEX6P3_ALAS_syn_ fwd	AGGGGCCCCTGGATCCCCGATGGAACACCTGGACAAATTC
pGEX6P3_ALAS_syn_ rev	TCACGATGCGGCCGCTCGAGTTATTTGGTTTTACCGCAAC

pET_ALASorg _fwd	CAGCAAATGGGTCGGATGGAACATTTAGATAAATTTAAAAAAATAATAAC
pET_ALASorg _rev	CGGAGCTCGAATTCGTCATTTAGTTTTACCACATCTTCG
pET_ALASopt _fwd	CAGCAAATGGGTCGGATGGAACACCTGGACAAATTC
pET_ALASopt _rev	CGGAGCTCGAATTCGTTATTTGGTTTTACCGCAACG
pGEX6P_*Not*IALASopt_ fwd	CCCCTGGGATCCCCGATGGAACACCTGGACAAATTC
pGEX6P_*Eco*RIALASopt _rev	CAGTCAGTCACGATGCTTATTTGGTTTTACCGCAACG
pGEX6P_*Not*IALASorg_ fwd	CCCCTGGGATCCCCGATGGAACATTTAGATAAATTTAAAAAAATAATAAC
pGEX6P_*Eco*RIALASorg_ rev	CAGTCAGTCACGATGCTCATTTAGTTTTACCACATCTTCG
SI39nov09_*Eco*RI_fwd	GGTTGAATTCTATGGAACATTTAGACAAGTTTAAACAAGTAATAACAG
SI39nov09_*Sal*I_rev	GGTGTCGACTCATTTAGTTCTACCACATCTTCGCA
SI39nov09opt_*Eco*RI_ fwd	GGTTGAATTCTATGGAACATCTGGATAAATTTAAACAAGTGATTACC
SI39nov09opt_*Sal*I_rev	GGTGTCGACTTATTTGGTGCGGCCGC
3300009790Ga_*Eco*RI_ fwd	GGTTGAATTCTATGGAACATTTAGATAAATTTAAGGCTGTAATAAAAGATTT
3300009790_Ga_*Sal*I_ rev	GGTGTCGACTCATTTCTTATTACCACATCTTCTCATAGCAGA
pColdTK-ALAScry-fwd	GGTAGGCATATGGAGCTCATGTTAAAATCAAAAACACAAGATG
pColdTK-ALAScry-rev	CAGGTCGACAAGCTTGTTAAATATATTTGTCAATAGAATTTCTTAAAGATTTTC
pColdTK-ALAScryopt-fwd	GGTAGGCATATGGAGCTATGCTTAAATCTAAAACCCAGG
pColdTK-ALAScryopt-rev	CAGGTCGACAAGCTTGTTATATATATTTATCAATGGAATTGCGAAG
pCold_ALASart_fw	GGTAGGCATATGGAGCTCATGGAACACCTCGACAAATTC
pCold_ALASart_rev	CAGGTCGACAAGCTTGAATTCTCACTTGGTCTTCCCACAG

pGEX_ALASart-fw	CCCCTGGGATCCCCGATGGAACACCTCGACAAATTC
pGEX_ALASart-rev	CAGTCAGTCACGATGCTCACTTGGTCTTCCCACAG
pCYB1-ALAScry-fwd	CAACAAGGACCATAGCATATGTTAAAATCAAAAACACAAGATG
pCYB1-ALAScry-rev	GAAGAGCCCTCGAGGTTAAAATATTTGTCAATAGAATTTCTTAAAGATTTTC
pCYB1-ALAScryopt-fwd	CAACAAGGACCATAGCATATGCTTAAATCTAAAACCCAGG
pCYB1-ALAScryopt-rev	GAAGAGCCCTCGAGGTTATATATATTTATCAATGGAATTGCGAAG
pGEX-ALAScry-fwd	CCCCTGGGATCCCCGATGTTAAAATCAAAAACACAAGATG
pGEX-ALAScry-rev	CAGTCAGTCACGATGCTTAAATATATTTGTCAATAGAATTTCTTAAAGATTTTC
pGEX-ALAScryopt-fwd	CCCCTGGGATCCCCGATGCTTAAATCTAAAACCCAGG
pGEX-ALAScryopt-rev	CAGTCAGTCACGATGCTTATATATATTTATCAATGGAATTGCGAAG
pGEX_CB2_*ho1*_fwd	CCCCTGGGATCCCCGATGGCTACACTGAAAGAGG
pGEX_CB2_*ho1*_rev	CAGTCAGTCACGATGCTCACTTTCCATCTTTATAGTGATTC
pGEX-CB2_*pcyA*_fwd	CCCCTGGGATCCCCGGTGACCGATTACATAGAACTAC
pGEX-CB2_*pcyA*_rev	CAGTCAGTCACGATGCTTATACGGTTGGAAACAGAATC
pACDuet_v*pcyA*_fw	CACAGCCAGGATCCGGTGACCGATTACATAGAACTA
pACDuet_v*pcyA*_rev	GCGGCCGCAAGCTTGTTATACGGTTGGAAACAGAATC
pACDuet_v*ho1*_fw	TATAAGAAGGAGATATACAATGGCTACACTGAAAGAGG
pACDuet_v*ho1*_rev	GCGGTTTCTTTACCAGACTTGCACTTTCCATCTTTATAGTG
Site-directed mutagenesis	
ALAScryopt_PLP K250A_fw	CTTAGGCGCGGCGTTCGGTATCCAGGGAGG
ALAScryopt_PLP K250A_rev	GAACGCCGCGCCTAAGGTGCCCTGTATAATGTCC
RcA_PLP_K248A_fw	CTGGCGGCAGCCTACGGCGTCTTCGGCGGC
RcA_PLP_K248A_rev	GGCTGCCGCCAGCGTGCCGTTGAAGATGTC

Primers used for site-directed mutagenesis have the changed nucleotides marked in red.

2.3 Microbial methods

2.3.1 Sterilization

Buffer and medium were sterilized before usage for at least 15 min at 121 °C and 1 bar by autoclaving. Heat-sensitive supplements were sterilized by filtration through a Polyether sulfone (PES) filter with a pore size of 0.22 µm and added under sterile conditions to sterilized medium.

2.3.2 Cultivation and storage of *E. coli* cells

A starter culture of *E. coli* cells was set up with an appropriate volume of LB media supplemented with the respective antibiotic and inoculated with a single colony from a freshly transformed *E. coli* plate or with a sterile loop from a glycerol stock. The starter culture was incubated overnight (O/N) under constant shaking at 37 °C with 1.8 * g. For the main culture preparation, LB media containing the appropriate supplement was inoculated with a 1:100 dilution of a starter culture. The main culture was then incubated at 37 °C with 0.5 * g until an $OD_{600\,nm}$ ~ 0.5 – 0.8 was reached. The cell density was determined by a photometrical measurement of the OD at 600 nm of 1 ml cell culture. The cultivation media without cells was used as the reference sample. The induction of gene expression was conducted either after the incubator reached 15 °C, 17 °C or 30 °C at different time points under constant shaking at 0.5 * g. For long-term storage of *E. coli* strains, cryocultures were prepared by using starter cultures grown O/N of the respective strain and adding sterile glycerol to obtain a 20 % (v/v) glycerol concentration. The cryocultures are flash-frozen in liquid nitrogen and stored at -80 °C.

2.3.3 Preparation of chemically competent *E. coli* cells

A starter culture of *E. coli* was grown overnight and used for inoculation, as stated before, of 100 ml LB media. Cultures were incubated at 37°C under shaking conditions until an OD_{600} of 0.5 was reached. After centrifugation with 1480 * g for 15 min at 4 °C, the cells were re-suspended in 50 ml $CaCl_2$-buffer and incubated for 1 h on ice. Following another centrifugation step with 1480 * g for 15 min at 4 °C, the sedimented cells were then re-suspended in 5 ml 50 mM $CaICl_2$ + 15 %-glycerol. The obtained chemically competent cells were distributed in reaction tubes in 200 µl aliquots, flash-frozen in liquid nitrogen, and stored at -80 °C for several months.

2.3.4 Transformation of chemically competent *E. coli* cells

For a transformation approach, the respective chemically competent *E. coli* cells were thawed on ice for 20 minutes. After the addition of 1 µl of plasmid DNA [≥ 50 ng/µl] or 10 µl – 20 µl of a ligation mixture or Gibson assembly mixture, the cells were incubated for 30 min on ice. Afterward, a heat-shock for 90 s was done, and the cells recovered on ice for 2 min. Following the addition of 700 µl LB media, the cells were incubated by shaking at 1.8 * g and 37 °C for 45 min and plated on agar plates containing the appropriate antibiotic. The plates were incubated O/N at 37 °C and stored at 4 °C for no longer than two weeks. Except for the *E. coli* strains carrying the plasmid pCold-TF or its derivate, these were not stored at 4 °C and were either transformed freshly or streaked out of cryocultures from -80 °C.

2.3.5 Functional complementation of auxotrophic *E. coli* Δ*hemA*

For examination of *in vivo* activity of viral ALAS, complementation of the synthesis of ALA should be carried out in the respective *E. coli* ST18 Δ*hemA*. This strain is a deletion mutant of the gene *hemA,* which encodes for the C5 pathway enzyme glutamyl-tRNA reductase. The strain is auxotrophic in growth when not supplemented with ALA [50 µg/ml] or complemented with an enzyme forming ALA, e.g., ALAS or GluTR. As positive controls the vector pRcA consisting of pGEX-6P with the *alaS* gene of *R. capsulatus* (Hännig 2011) encoding an aminolevulinic acid synthase and the vector pMkA of pGEX-6P with the *hemA* gene of *Methanopyrus kandleri* (Moser *et al*., 1999) encoding a glutamyl-tRNA reductase (GluTR) were selected and transformed into the auxotrophic *E. coli* ST18. The respective empty vectors were used as control.

2.3.6 Chemical competent cells for functional complementation

Freshly prepared chemically competent *E. coli* ST18 Δ*hemA* cells were crucial for functional complementation. Therefore, an LB-agar plate with 50 µg/ml ALA was used to streak out *E. coli* ST18 Δ*hemA* and incubated overnight at 37 °C. ALA is not stable when dissolved in water and should be prepared freshly. The following O/N culture was inoculated with a single colony of the streak out of *E. coli* ST18 Δ*hemA* in LB medium with 50 µg/ml ALA and incubated shaking at 37 °C. With this culture, chemically competent cells were prepared by inoculating 5 ml LB medium + 50 µg/ml ALA with 1:100 of the O/N culture. The cells were grown at 37 °C shaking until $OD_{600\,nm}$ reached

0.5 – 0.7. The biomass of the 5 ml culture was harvested at 4 °C for 10 min at 5,000 * g, and the supernatant was discarded. The cell pellet was resuspended in 1 ml 0.1 M $CaCl_2$ following another centrifugation for 5 min at 4 °C with 5,000 * g. Washing was repeated with 0.5 ml 0.1 M $CaCl_2$ following the same centrifugation as stated before and a final resuspension of the cells in 0.2 ml of 0.1 M $CaCl_2$. The resuspended cells are incubated on ice for 30 min. For each transformation approach, 0.1 ml of these cells can be used. The addition of plasmid DNA was followed by incubation at 37 °C for 30 sec in a heating block. The transformation approach is then incubated on ice for 40 min, followed by a heat shock of 1 min at 42 °C. Further, 1 ml of LB medium + 50 µg/ml ALA was added and incubated for 30 min at 37 °C shaking. Plating of the transformation approach was carried out on respective selective agar containing the appropriate antibiotic, supplemented with 50 µg/ml ALA, and incubated O/N at 37 °C.

2.3.7 Growth curve of complemented *E. coli* strains

Cultivation was conducted in a sterile 100 ml glass flask filled with 20 ml LB medium and the respective antibiotic. The inoculum was O/N cultures from the transformation plates supplemented with respective antibiotics and ALA. Before inoculation, the cells were washed two times in LB medium without ALA and then used as an inoculum adjusted to start $OD_{600\ nm}$ of 0.03. During incubation at 37 °C with shaking at 1.8 * g, the flasks were fully covered in aluminum foil to prevent light entry. Samples were measured using a spectrophotometer and diluted accordingly when reaching $OD_{600\ nm}$< 0.9. During growth, the gene expression was induced by adding 0.5 mM IPTG after the cultures reached an $OD_{600\ nm}$ of 0.1 - 0.15 or were not induced to determine the effect of the known leaky *tac* promotor. After incubation for 24 h, the cultures were plated out after serial dilution on agar plates without ALA supplementation but containing antibiotics and with or without IPTG to determine functional complementation by incubating at 37 °C O/N. The formation of colonies indicated successful complementation; empty vectors were also plated and showed no colony formation. All experiments were conducted as independent biological triplicates with technical duplicates.

2.4 Molecular biological methods

2.4.1 Preparation of plasmid DNA

Plasmid DNA preparation was conducted from single colonies of freshly transformed *E. coli* DH5α cells. One colony was inoculated in 5 ml LB media containing the appropriate antibiotic and incubated overnight at 37 °C and 1.8 * g. The plasmid DNA was isolated using the NucleoSpin® EasyPure Kit (Macherey-Nagel) or the E.Z.N.A.® Plasmid DNA Mini Kit I (Omega Bio-tek) according to the manufacturer's instructions.

2.4.2 Determination of the concentration of nucleic acids

Determination occurred after the isolation of nucleic acids, by measuring the absorbance of the DNA solution at 260 nm (NanoDrop™ Lite). The purity of the isolated nucleic acids was determined by the calculation of the ratio of the absorbance at 260 nm and 280 nm. For a sufficiently pure double-stranded DNA solution, the absorption ratio of A_{260}/A_{280} is ~ 1.8 – 2.0.

2.4.3 Analysis of nucleic acids by agarose gel electrophoresis

For separating and identifying DNA molecules according to their size, agarose gel electrophoresis of nucleic acids was performed. Analytical and gel extraction approaches were set up by dissolving 0.8 % (w/v) of agarose in boiling 1x TAE buffer. Once the solution was cooled down, it was filled into the prepared gel chambers. During the polymerization of the gel, the DNA samples were prepared by adding specific amounts of 6x DNA loading dye (Thermo Fisher) and water to reach 1x. The prepared samples were then loaded into the gel wells, and the DNA fragments were separated in 1x TAE buffer at a constant current of 120 V.

TAE buffer (1x)

Tris/acetate	40 mM
EDTA	1 mM
	pH 8.0

In order to determine the size, the reference GeneRuler™ DNA 1 kb plus ladder (Thermo Fisher) was used for comparison. After separating the DNA, the gel was incubated in an ethidium bromide bath (consisting of 0.05 % (v/v) ethidium bromide in Aqua dest.) for 20 min, rinsed with water and visualized by a gel documentation station. If gel extraction

was required, the desired fragment was cut out from the gel and prepared using the NucleoSpin® Gel and PCR Clean-up Kit (Macherey-Nagel) according to the manufacturer's instructions.

2.4.4 The polymerase chain reaction (PCR)

The specific amplification of DNA fragments was conducted by the polymerase chain reaction (Table 9 +Table 10). The amplification process employed the utilization of the Phusion® High-Fidelity DNA Polymerase (Thermo Fisher or home-made). To verify the size of the generated PCR products, they were analyzed via agarose gel electrophoresis (see chapter 2.4.3). Amplified DNA fragments for cloning into a plasmid were purified using the NucleoSpin® Gel and PCR Clean-up Kit (Macherey-Nagel) or DNA Clean & Concentrator™-5 (Zymo Research) according to the manufacturer's instructions.

Table 9: PCR reaction with Phusion® High-Fidelity DNA polymerase (50 µl).

Component	Final concentration
Phusion® High-Fidelity DNA Polymerase	0.02 U/µl
5x HF Phusion® Buffer	1x
10 mM dNTPs	200 µM each
Primer forward	100 pmol
Primer reverse	100 pmol
Template DNA	5 - 10 ng
Sterile Aqua dest.	ad 50 µl

Table 10: Reaction step using Phusion® High-Fidelity DNA polymerase

Reaction step	Temperature	Time	
Initial denaturation	98 °C	30 s	
Denaturation	98 °C	10 s	x 35 cycles
Annealing	T_m-5 °C	15 – 20 s	
Elongation	72 °C	15 s / kb	
Final elongation	72 °C	5 min	

2.4.5 Colony PCR

Colony PCR was frequently used to screen *E. coli* colonies for the presence of desired plasmid DNA after transformation events from cloning approaches. Therefore, primers were chosen that would allow amplification of a DNA fragment from the vector backbone to either vector backbone with the insert in-between or to the 3'end of the insert. With this, a verification of the cloning success could be indicated, which was

confirmed by sequencing. For the colony PCR, single colonies of *E. coli* were picked with a sterile toothpick from transformation plates and preserved by streaking on a fresh LB-agar plate with the appropriate antibiotic. Then, the picked colony was resuspended multiple times into the PCR tube containing already all PCR components except of the polymerase, thus used directly as PCR template. The PCR reaction was set up as described in chapter 2.4.4. except for extending the first denaturation step to 5 min to achieve complete cell lysis. After analysis of the amplified DNA using agarose gel electrophoresis, the selected primers could amplify a DNA fragment according to the size of the insert regarding the chosen primer pair. The identified correct clone was picked from the master plate. From the suspected correct clone, the plasmid DNA was extracted and sent for sequencing (see chapter 2.2.2) to confirm correct cloning.

2.4.6 Restriction endonuclease digestion of nucleic acids

The plasmid DNA or PCR products were digested using FastDigest restriction endonucleases (Thermo Fisher) or High-Fidelity (HF®) restriction endonucleases (NEB) in a total volume of 50 µl. The digestion was performed following the manufacturer's instructions regarding buffer and enzyme usage and concentration. Following digestion, the enzymes were heat-inactivated at either 60 °C or 80 °C for 20 min. For plasmid DNA, the restriction process was confirmed by agarose gel electrophoresis (see chapter 2.4.3).

2.4.7 Ligation of DNA molecules

In the conventional cloning method, the vector DNA and insert DNA, previously digested with respective restriction endonucleases, are ligated together. The ligation was carried out using T4 DNA ligase in a volume of 20 µl according to the manufacturer's instructions with a ratio of 3:1 (insert:vector) for at least 30 min at room temperature or overnight. The ligation process was inactivated by thermal inactivation at 65 °C for 10 min. Then, 10 µl – 20 µl of the ligation sample was transformed into chemically competent *E. coli* DH5α cells and plated on LB agar plates containing the appropriate antibiotic.

2.4.8 Gibson Assembly® of DNA molecules

Gibson Assembly® was used as the alternative to the conventional cloning technique. This method utilizes a 5'- exonuclease, a 3'-extension, and a DNA-ligase activity in a single tube reaction (Gibson *et al.,* 2009). For the primer design with the respective overhangs, the NEBuilder® Assembly Tool (http://nebuilder.neb.com/) was used. Target

genes were then amplified using the generated primers, thus achieving vector-specific overhangs. The designated vector was digested with the appropriate restriction enzyme(s). Both linearized vector DNA and PCR product were prepared as described in chapter 2.4.3, and 10 - 100 ng of each fragment was added in a total volume of 5 µl autoclaved MilliQ water to achieve an insert:vector ratio of 3:1. Following the addition of 15 µl of the Gibson assembly master mix and incubation in a heating block at 50 °C for 1 h. Then, the transformation of the assembled vector into *E. coli* DH5α was carried out as described in chapter 2.3.4 while using the total volume of the Gibson reaction.

5x isothermal Gibson assembly reaction buffer

PEG 8000	25 % (w/v)
Tris-HCl pH 7.5	500 mM
$MgCl_2$	50 mM
DTT	50 mM
dNTPs	1 mM each

Gibson assembly master mix

5x isothermal Gibson assembly reaction buffer	320 µl
T5-exonuclease (10 U/µl)	0.64 µl
Phusion® High-Fidelity DNA polymerase (2 U/µl)	20 µl
Taq DNA ligase (40 U/µl)	160 µl
H_2O	fill up to 1200 µl

2.4.9 Site-directed mutagenesis

Variants of RcA and ALAS were generated using site-directed mutagenesis (Wang & Malcolm 1999). The exchange of certain bases in the protein's respective coding sequence results in protein variants, which should be analyzed in complementation approaches or enzyme assays. For the generation of these variants, specific primers for site-directed mutagenesis were created (Table 8; the site of mutation is marked in red) that had the to be exchanged bases at the 5'-site of each primer, following 16-20 bases without base changes. For the generation of the protein variants, the target plasmids were used in a dual-step PCR procedure using the Phusion High-Fidelity DNA Polymerase with the corresponding primer pair. The two-step approach prevents dimerization of the

oligonucleotides and increases the number of mutated single-stranded plasmid copies. In the first step, two separate PCR reactions (A and B) were performed (Table 11), each utilizing only one primer (Table 12). In the second step (Table 14), the products from the two reactions were combined (Table 13), generating double-stranded plasmids with introduced mutation sites. The PCR product was purified (DNA Clean & Concentrator Kit), and to remove the methylated parental plasmid DNA, *Dpn*I digestion (New England Biolabs) was conducted following heat-inactivation according to the manufacturer's instructions. Subsequently, the resulting plasmid mixture was transformed into chemically competent *E. coli* DH5α cells (see chapter 2.3.4).

Table 11: PCR reaction setups of the first step

Component	A	B
5x HF Phusion® buffer	10 µl	10 µl
Plasmid DNA-template (30 ng/µl)	1 µl	1 µl
Forward primer (10 nM)	2.5 µl	–
Reverse primer (10 nM)	–	2.5 µl
dNTPs (200 µM each)	1 µl	1 µl
Phusion® High-Fidelity DNA polymerase (2 U/µl)	0.5 µl	0.5 µl
H_2O	ad 50 µl	ad 50 µl

Table 12: PCR thermocycler program of the first step

Reaction step	Temperature	Time	
Initial denaturation	98 °C	30 s	
Denaturation	98 °C	15 s	x 8 cycles
Annealing	55 °C	30 s	
Elongation	72 °C	6 min	

Table 13: PCR reaction setup of the second step

Component	Volume
PCR mix A	25 µl
PCR mix B	25 µl
Phusion® High-Fidelity DNA polymerase (2 U/µl)	0.5 µl

Table 14: PCR thermocycler program of the second step

Reaction step	Temperature	Time	
Initial denaturation	98 °C	30 s	
Denaturation	98 °C	15 s	x 18 cycles
Annealing	55 °C	30 s	
Elongation	72 °C	4 min	
Final Elongation	72 °C	10 min	

2.5 Protein biochemical and analytical methods

2.5.1 Production of recombinant proteins in *E. coli* - Analytic scale

Recombinant protein production should be optimized regarding the solubility of the target protein. Therefore, different *E. coli* strains, culturing conditions, gene induction concentrations, and lysis buffer (+/- additives) were tested (Table 15, Table 16). To determine the solubility of the tested approaches, test expressions were conducted in small scale (5 - 20 ml medium) to medium scale (20 - 500 ml medium) following either a cell lysis procedure known as freeze-thaw (Johnson & Hecht 1994), microfluidizer or sonification. Freeze-thaw was carried out for analytic scale where 2 ml of cell culture was harvested by centrifugation (5 min, 13 000 * g), the supernatant was removed and frozen in liquid nitrogen for 2 min, and the sample was then thawed for 8 min in an ice-water bath. This freeze-thaw cycle was repeated two times, and the pellet was resuspended after adding lysis buffer (50 µl – 75 µl). To achieve resuspension of the pellet and obtain soluble protein, the reaction tube containing it was repeatedly scraped against the tube rack for 30 sec. The mixture was then incubated in an ice water bath for 30 min following a centrifugation step at 4 °C, for 10 min at 10 000 * g. The supernatant containing the soluble proteins was pipetted off from the pellet, and both supernatant and pellet were analyzed using SDS-PAGE to determine solubility. Cell lysis using a microfluidizer or sonification was carried out in medium scale setups (20 – 500 ml) as described in chapter 2.5.2.

Table 15: Overview of the strategies pursued in the protein production studies to obtain soluble protein testing different vectors and *E. coli* strains.

Strategy	**Conditions**
Titration of inducer concentration (only IPTG)	1 µM, 10 µM, 50 µM, 100 µM, 500 µM, 1 mM
Temperature for protein production	15 °C, 17 °C, 30 °C
Duration of protein production	1 h, 3 h, 6 h, O/N
Different *E. coli* strains	Arctic Express (DE3), BL21(DE3), BL21(DE3) CodonPlus RIL, BL21 (DE3) pGro7, Rosetta gami (DE3), Rosetta (DE3) pLysS, Shuffle T7 Express, AD494(DE3), Origami(DE3), Tuner
Co-expression with chaperon plasmid	pGro7 induced with 0.05 mg/ml arabinose
Media	LB, 2YT, MM, AI
Media supplements	100 mM sorbitol with 2.5 mM betaine, 3 % (v/v) EtOH, 1 % glucose, LB pH 6.5
Cell disruption	Sonification, freeze thaw, microfluidizer
Lysis buffer	Different buffer systems (Tris, Phosphate, HEPES), pHs (5-9), detergents (Triton X-100, NP40)

Table 16: Overview of tested expression conditions to improve solubilization for viral *alaS* genes encoding for marine ALASsyn, opt and org.

	ALASsyn – GST N-term.		ALASopt N-term His		ALASopt N-term. GST		ALASorg N-term His		ALASorg N-term. GST	
***E. coli* strains**	Codon Plus RIL	Rosetta gami, Shuffle T7, AD494, Origami, pGro7, Tuner	Codon Plus RIL	Origami, AD494 Rosetta pLysS, Rosetta, pGro7, Shuffle T7, Tuner	Codon Plus RIL	pGro7, AD494, Shuffle T7 Express, Tuner	Codon Plus RIL	Origami, AD494, Rosetta pLysS, pGro7, Shuffle T7, BL21 Tuner	Codon Plus RIL	pGro7, AD494, Shuffle T7 Express, Tuner
Expression temperature										
15 °C	X		X				X			
17 °C	X	X	X	X	X	X	X	X	X	X
30 °C	X		X				X			
Medium										
LB	X	X	X	X	X	X	X	X	X	X
M9	X				X		X		X	
2YT	X				X		X		X	
LB-BS	X				X		X		X	
LB + 3% (V/V)										
EtOH	X				X		X		X	
LB+ 1% glucose	X				X		X		X	
LB pH 6.5	X				X		X		X	
Inductor titration										
1 µM	X									
10 µM	X									
50 µM	X		X				X			
100 µM	X	X	X	X	X	X	X	X	X	X
500 µM	X									
1 mM										
Expression duration										
1 h	X		X				X			
3 h	X	X	X	X	X		X		X	
6 h	X	X		X						
O/N	X		X		X	X	X	X	X	X
Cell disruption										
Sonification	X									
Freeze thaw	X	X	X	X	X	X	X	X	X	X
Microfluidizer	X				X		X		X	
Lysis buffer										
Buffer system	X				X		X		X	
pH	X				X		X		X	
detergents	X				X		X		X	

Production of recombinant proteins in *E. coli* – Preparative scale

All recombinant proteins used during this study were produced heterologously in different *E. coli* strains. GST-tagged proteins were GST-RcA, GST-PcyA (*Anabaena* sp. PCC 7120, following named *Anabaena*), GST-BphO (Ho1 from *Pseudomonas aeruginosa*, following named *Pseudomonas*), GST-ALASsyn, GST-Ho1 (CB_2) and GST-PcyA (CB_2). The trigger factor fused proteins were His_6-tagged, namely TF-ALAScry, TF-ALAScryopt. For the production, they were freshly transformed in respective *E. coli* strains or streaked out from glycerol cultures. The colonies were grown O/N in LB media

containing the respective antibiotic. The main culture was set up as described (see chapter 2.3.2) and incubated at 37 °C with 0.5 * g until the $OD_{600\ nm}$ reached ~ 0.5. For GST-ALASsyn, the LB medium was supplemented with 2.5 mM betaine and 100 mM sorbitol.

***E. coli* strain for protein production**	**(Fusion-) protein**	**IPTG concentration**
BL21(DE3)	GST-PcyA (*Anabaena*),	0.5 mM
	GST-vHo1 (CB_2),	0.1 mM
	TF-ALAScry, TF-ALAScryopt	0.5 mM
	Cph1, Ho1, vPcyA (CB_2)	1 mM
	Cph1, vHo1 (CB_2), vPcyA (CB_2)	0.1 mM
BL21 RIL codon plus	GST-RcA,	0.2 mM
	GST-BphO (*Pseudomonas*),	1 mM
	GST-ALASsyn	0.5 mM
Rosetta pLysS	GST-vPcyA (CB_2)	0.1 mM

The cultures with constructs for GST-tagged proteins were shifted in temperature to 17 °C, whereas the ones with His_6-tagged trigger factor proteins were quickly cooled down in an ice water bath for 30 min. Following this, gene expression was induced with IPTG addition regarding the respective construct. The cells were incubated overnight at 17 °C (GST) or 15 °C (TF) with constant shaking at 0.5 * g.

The *in vivo* assembly of a chromophore with a phytochrome was carried out using a coexpression system. Therefore, *E. coli* BL21(DE3) with the plasmid pET_His-tag Cph1 crystal and one of the following plasmids pTD_v*ho1*_v*pcyA* or pTD*ho1*_v*pcyA* were used. The inoculation and growth conditions were as stated before. The culture was induced with 0.1 or 1 mM IPTG (see above) and incubated at 17 °C O/N in the absence of illumination.

2.5.2 Cell harvest and lysis

The cells were harvested by centrifugation for 15 min at 4 °C with 17 200 * g (Sorvall LYNX 6000 Centrifuge, rotor T9), and the pelleted cells were transferred into 50 ml centrifuge tubes (Falcon™) while the supernatant was discarded. The cells were either directly lysed for protein purification or stored. For storage, the pelleted cells were flash-frozen in liquid nitrogen and then stored at -20 °C. Regarding cell disruption, the pellet was thawed on ice, resuspended in the appropriate lysis buffer (1x PBS, Tris buffer, HEPES buffer) (5 - 10 ml buffer/g wet weight of the pellet), and incubated on ice for 30 min after the addition of one spatula tip of DNaseI and lysozyme. The cells were then mechanically disrupted by using Microfluidizer^{TM} three times at a constant pressure of 15 000 PSI or sonification (Ultrasonic homogenizer, Bandelin Sonoplus HD 2200 with tip KE 76) in an ice-water bath for a total of 1 min 30 sec sonification time with intervals of 30 sec sonification and 1 min cool down. Sonification pulse was 40 % intensity, 5 x cycle. The soluble fraction was separated from insoluble components by centrifugation at 48,000 *g (1 h, 4°C; Sorvall LYNX 6000 Centrifuge, rotor T29).

10x Phosphate-buffered saline (GST)

NaCl	1.4 M
KCl	27 mM
Na_2HPO_4	100 mM
KH_2PO_4	18 mM
pH 7.4	
for usage diluted to 1x buffer	

Tris buffer (GST-BphO)

Tris	50 mM
NaCl	100 mM
Triton X-100	0.05 % (v/v)
DTT	1 mM
pH 8.0 (HCl)	

HEPES buffer (GST)

HEPES	50 mM
NaCl	300 mM
pH 7.5 (NaOH)	

HEPES buffer (His-Cobalt)

HEPES	50 mM
NaCl	1 M
Imidazol	45 mM
pH 7.5 (NaOH)	

HEPES buffer (His-Nickel)

HEPES	50 mM
NaCl	300 mM
pH 7.5 (NaOH)	

HEPES buffer (GST-ALASsyn)

HEPES	50 mM
NaCl	200 mM
DTT	10 mM
PLP	20 µM
pH 7.0 (NaOH)	

2.5.3 Purification of recombinant proteins by affinity chromatography

Affinity chromatography was used to isolate the proteins from the total cell lysate. According to the fused tags, affinity chromatography columns with the column material Protino™ Glutathione-Agarose 4B, TALON® Superflow™ (Cobalt), or Ni-NTA Agarose were used. For GST-RcA purification, all buffers contained additionally 20 µM PLP. BphO, PcyA (*Anabaena*), Ho1 (CB_2), and PcyA (CB_2) were purified using 1x PBS buffer (GST) and elution buffer (GST). All ALASs and RcA were purified using HEPES wash and elution buffer (either GST, His-Nickel, or His-Cobalt) regarding their tags and used column material. As for the coexpression system, TALON® Superflow™ (Cobalt) Agarose was used only to enrich the Cph1, which contains a His_6-tag, and the PcyA that also has a His_6-tag here, the buffers were used according to the manufacturer's instructions. Shortly, the wash buffer was 1xPBS (see below), and the elution buffer was 1xPBS with 250 mM imidazole, both pH 7.4.

The gravity flow columns containing the respective amount of column material were equilibrated according to the manufacturer's instructions, and the lysate was added to the column. The column was then washed with the appropriate wash buffer (1xPBS or HEPES wash buffer), and the bound protein was then eluted from the column with an appropriate elution buffer (elution buffer, HEPES elution buffer) in 1 ml fractions and collected. The column was then regenerated according to the manufacturer's instructions and stored at 4 °C on 20 % (v/v) ethanol until further use. In order to control the purification procedure, samples were taken from some elution fractions as well as from lysate, flow through, and wash fractions and analyzed by using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE, see chapter 2.5.5). Suitable elution fractions were combined after purification. If necessary, the GST-tag or Trigger Factor of the fusion proteins was subsequently cleaved off. For this purpose, the elution fractions were mixed with 20 µl laboratory proprietary PreScission® protease (2 U/100 µg) per ml of elution fraction and dialyzed overnight at 4 °C against PreScission®-Protease buffer (see chapter 2.5.4). For removing the GST-tag, the fraction was then added again to a second affinity chromatography column with Protino™ Glutathione Agarose 4B column material to remove the cleaved GST-tag and the added

PreScission® protease from the protein solution. Removing the Trigger factor required the addition of the collected fractions to a second affinity chromatography column consisting of either cobalt or nickel column material to remove the cleaved Trigger factor. If required, the PreScission® protease could be removed by a followed application onto a Protino™ Glutathione-Agarose 4B column, the respective flow through should then contain the desired protein.

10x Phosphate-buffered saline (PBS) wash buffer (GST)

NaCl	1.4 M
KCl	27 mM
Na_2HPO_4	100 mM
KH_2PO_4	18 mM
pH 7.4	

for usage diluted to 1x buffer

Elution buffer (GST)

Tris	50 mM
Glutathione red.	10 mM
pH 8.0 (HCl)	

HEPES wash buffer (GST)

HEPES	20 mM
NaCl	200 mM
pH 7.5 (NaOH)	

HEPES elution buffer (GST)

HEPES	
Glutathione red.	10 mM

HEPES wash buffer (His-Cobalt)

HEPES	50 mM
NaCl	1 M
Imidazol	45 mM
pH 7.5 (NaOH)	

HEPES elution buffer (His-Cobalt)

HEPES	50 mM
NaCl	1 M
Imidazol	45 mM
pH 7.5 (NaOH)	

HEPES wash buffer (His-Nickel)

HEPES	50 mM
NaCl	300 mM
Imidazol	5 mM
pH 7.5 (NaOH)	

HEPES elution buffer (His-Nickel)

HEPES	50 mM
NaCl	300 mM
Imidazol	150 mM
pH 7.5 (NaOH)	

PreScission®-Protease-Puffer

HEPES	50 mM
NaCl	200 mM
DTT	1 mM
EDTA	1 mM
pH 7.5 (NaOH)	

2.5.4 Dialysis and concentration of isolated proteins

For dialysis, the protein solutions were incubated at 4 °C overnight with continuous stirring in the desired buffer for further experiments. For this purpose, the protein solution was transferred to a dialysis tube with an exclusion limit of 10 kDa. The buffer volume was approximately 250 times the volume of the protein solution. The dialyzed protein solution was then transferred to Vivaspin® 6 concentrators (MWCO 10 kDa) to reduce it to the desired volume (0.5 - 1 ml) by centrifugation at 4 °C with 1480 * g (Centrifuge 5415D, Rotor F-45-24-11). Previously, the concentrators were washed with Aqua dest., and then equilibrated with buffer.

2.5.5 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins were analyzed using sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (SDS-PAGE) by separation according to their molecular weight. The proteins were separated in a discontinuous system consisting of a stacking gel (pH 6.8) with an acrylamide concentration of 5.25 % and a separation gel (pH 8.8) with an acrylamide concentration of 12.5 %.) SDS is a denaturing detergent that masks the proteins' charge and replaces them with a negative charge to separate the proteins in an electric field. The acrylamide concentration determines the resolution of the separation. Larger proteins migrate slower through the acrylamide matrix than smaller proteins. Protein samples were prepared by adding 4x sample buffer, boiled for 10 min, centrifuged for 5 min, and loaded on the gel for separation. The electrophoresis was performed by applying 200 V for 1 h or until the loading dye reached the bottom of the gel. As a reference, the Color Prestained Protein Standard, Broad Range (NEB), was used to identify the proteins of interest for a following western blot, and the Unstained Protein Standard, Broad Range (NEB), was used only for SDS-PAGE. For visualization of the proteins, the gel was incubated in Coomassie staining solution for 10 min. Afterward, the gel was destained using the same solution without Coomassie until the excess dye was removed. The proteins were then detectable as blue signals on the gel.

Stop gel

Aqua dest.	330 µl
Separation buffer	250 µl
Acrylamide 30 %	420 µl
10 % APS	10 µl
TEMED	2 µl

Separation gel

Aqua dest.	5.3 ml
Separation buffer	4 ml
Acrylamide 30 %	6.7 ml
10 % APS	80 µl
TEMED	8 µl

Stacking gel

Aqua dest.	4.6 ml
Stacking gel buffer	2 ml
Acrylamide 30 %	1.4 ml
10 % APS	30 µl
TEMED	20 µl

4x Separation gel buffer

Tris-HCl pH 8	1.5 M
SDS	0.4 % (w/v)

4x Stacking gel buffer

Tris-HCl pH 6,8	1.5 M
SDS	4 % (w/v)

5x SDS-sample buffer

4x Stacking gel buffer	3.125 ml
SDS	0.75 g
87 % Glycerol	3 ml
β-Mercaptoethanol	1.25 ml
Bromophenol blue	2 ml

10x SDS-running buffer

Tris-HCl pH 8,8	0.25 M
Glycine	1.92 M
SDS	1 % (w/v)

Staining solution

Acetic acid	10 % (v/v)
Coomassie Brilliant Blue	0.25 % (w/v)
Ethanol	30 % (v/v)

Destaining solution

Acetic acid	10 % (v/v)
Ethanol	30 % (v/v)

2.5.6 Immuno-detection of immobilized proteins (western blot)

Western blot analyses were used for the immunological detection of recombinant fusion proteins with a GST- or His_6-tag using specific polyclonal antibodies. After an SDS-PAGE to separate the proteins (see chapter 2.5.5), they were transferred to a polyvinylidene difluoride (PVDF) membrane by the semi-dry blot method. For this purpose, the SDS-polyacrylamide gel, Whatman paper (3 mm), and a methanol-activated PVDF membrane were first equilibrated in Towbin buffer for 10 min. The protein transfer was carried out for 20 min at 15 V through a semidry blot system. For the transfer of the proteins, the

gel and the membrane were stacked together, surrounded by the blotting paper in a sandwich formation, and placed between the anode and cathode. Subsequently, non-specific binding sites on the membrane were saturated with a blocking solution at 4 °C or slightly shaken for 1 h at RT. Incubation with the first antibody (Table 4) in 10 ml blocking solution followed, gently shaking for 1 h at RT. After three five-minute washing steps with TBS-T buffer, the membrane was incubated with the second antibody (Table 4) in 10 ml blocking solution and shaken gently for 1 h at RT. Again, two times five-minute washing steps with TBS-T buffer and two with TBS buffer were performed. Optical detection was performed by the chromogenic reaction of alkaline phosphatase by the addition of 33 µl nitro blue tetrazolium (NBT, 100 mg/ml in 70 % (v/v) DMF) and 66 µl 5-bromo-4-chloro-3-indolyl phosphate (BCIP, 50 mg/ml in DMF) in 10 ml AP buffer. The color reaction was stopped by washing the membrane with water.

TBS buffer		**AP-buffer**	
Tris-HCl pH 7.4	50 mM	Tris-HCl pH 9.5	100 mM
NaCl	140 mM	NaCl	100 mM
TBS-T buffer		$MgCl_2$	5 mM
1x TBS			
Tween-20	0.1 % (v/v)		
Towbin transfer buffer		**Blocking solution**	
Tris-HCl pH 8.3	25 mM	1x TBS	
Glycine	192 mM	Albumin fraction V	3 % (w/v)
Detection solution			
AP buffer	10 ml		
NBT solution	33 µl		
BCIP solution	66 µl		

2.5.7 Determination of protein concentration

Protein concentration of purified proteins

The protein concentration was determined photometrically by measuring the absorbance at 280 nm. Considering the chosen dilution, the concentration could be calculated according to the Lambert-Beer law. For this purpose, the extinction

coefficient ε_{280nm} was determined with the software 'Protein Calculator v3.4' according to the method of Gill and von Hippel (Table 17) (Gill & von Hippel 1989).

$$c = \frac{A_{280nm} \times F}{\varepsilon_{280nm} \times d}$$

A_{280nm} = Absorption at 280 nm

c = concentration in g/l

d = pathlength of the cuvette in cm

ε_{280nm} = molar extinction coefficient at 280 nm in $M^{-1}cm^{-1}$

F = dilution factor

The molecular weight of the respective proteins, which was also determined by the 'Protein Calculator v3.4 software', was used to calculate the concentration in g/l.

$$c\left[\frac{g}{l}\right] = M_r\left[\frac{g}{mol}\right] \times c\left[\frac{mol}{l}\right]$$

c = concentration in g/l or mol/l

M_r = molecular weight in g/mol

Table 17: Molar extinction coefficient for determination of protein concentration. The molar extinction coefficients of the proteins are given with the possible disulfide bridges.

Protein	**Calculated molecular weight [Da]**	**Extinction coefficient ε_{280nm} [$M^{-1}cm^{-1}$]**
ALASsyn	44379	34900
RcA	43647	35110
ALAScry+TF/ ALAScryopt+TF	96670	52560
ALAScry/ALAScryopt	46118	36630
PcyA (*Anabaena*)	28726	26390
vPcyA (CB_2)	24552	22430
BphO (*Pseudomonas*)	21858	40540
vHo1 (CB_2)	20690	41370

Protein concentration of cell lysate

Bradford reagent was used to determine cell lysate protein concentration. Therefore, cell debris-free sample was incubated with Bradford reagent (Roti®Quant) in a transparent 96-well plate (Sarstedt, PS, flat based) as recommended by the manufacturer. Using a calibration curve with BSA solubilized in Tris buffer (50 mM Tris-HCl buffer pH 7.5) and measured in a plate reader (Tecan Infinite® F200 Pro) at 595 nm.

2.5.8 Size exclusion chromatography (SEC)

The oligomerization state of a protein can be determined using size exclusion chromatography (SEC). In a column filled with a porous gel matrix, proteins can be separated according to their hydrodynamic radius. Here, small proteins can enter the hydrophilic pores of the gel matrix and cannot pass the matrix as fast as big proteins with no access to the pores. The size of the molecule correlates with its migration speed in the column and increases the time of its elution (elution volume, Ve). Characterized proteins and their corresponding molecular mass can be used as standards to determine the relative molecular mass of unknown proteins.

The SEC was carried out using an FPLC system (ÄktaPure, GE Healthcare) in a cold room (4 °C) with a Superdex® 200 Increase 10/300 GL column (GE Healthcare) and a 500 µl sample loop. All buffers were filter sterilized (0.22 µm, PES) and degassed before usage (30 min, sonification bath). The system was washed with H_2O and equilibrated with the respective buffer. The sample was filtered and centrifuged before applying it onto the system. The flow rate was 0.2 ml/min, and the sample elution was monitored spectroscopically at 280 nm. Elution fractions were collected continuously, and appropriate elution fractions containing proteins were precipitated using TCA DOC and analyzed via SDS PAGE.

Protein standards used to calculate the oligomerization state of the protein were alcohol dehydrogenase (150 000 Da), albumin (66 500 Da), carbonic anhydrase (29 000 Da), and lysozyme (14 000 Da). The void volume (Vo) was determined using Blue dextran (2 000 000 Da). The calibration curve was blotted using the gel-phase distribution coefficient (Kav) against the logarithm of the molecular weight. Kav = (Ve – Vo)/(Vc –

Vo), where Vo was 7.96 ml, and Vc (geometric column volume) was 23.56 ml. With the calibration curve, the experimental molecular weights can be calculated.

2.5.9 TCA DOC precipitation

Trichloroacetic acid (TCA) precipitation utilizes the weak acid character of TCA that disrupts the hydrogen bond of the water molecules surrounding the protein, thus changing a protein to its insoluble state, which can be recovered by centrifugation and analyzed accordingly.

Analysis of proteins purified using the ÄKTA system could be improved regarding yield by adding Deoxycholate (DOC) to the regular trichloroacetic acid (TCA) precipitation. It precipitated proteins less concentrated than 1 µg/ml (Koontz 2014). To 1 ml of protein sample, 100 µl of 0.15 % (w/v) DOC was added and vortexed. The solution was incubated at room temperature for 10 min, complemented with 50 µl of 100 % (w/v) TCA, and vortexed. The protein precipitation was carried out while incubating on ice for 30 min and centrifuged at 10 000 * g at 4 °C for 10 min. The supernatant was carefully aspirated and discarded, while the white pellet was washed several times in 500 µl ice-cold acetone to remove leftover TCA. Each wash step followed centrifugation at 10 000 * g at 4 °C for 5 min, where the supernatant was discarded. Finally, the pellet is dried and resuspended in SDS buffer for analysis and processed accordingly before being analyzed by SDS-PAGE.

2.5.10 Anion-exchange chromatography (AIEX)

The ion exchange chromatography is a purification method that separates substances based on their charge. When a positively charged ion exchange resin with an affinity for molecules having a net negative surface charge is used, it is called anion-exchange chromatography (AIEX). The molecule of interest net charge in a specific buffer determines the resin choice. In order to separate the protein of interest from contaminating proteins, the column HiTrap® Q HP (1 ml, Cytiva) was used. All buffers were filter sterilized (0.22 µm, PES) and degassed before usage (30 min, sonification bath). The system was washed with H_2O and equilibrated with the start buffer (20 mM Tris-HCl, pH 8.0). The sample was dialyzed O/N against the start buffer, filtered, and centrifuged before being applied to the system. The flow rate was 0.5 ml/min, and the

sample elution was monitored spectroscopically at 280 nm. The elution followed a linear gradient with increasing concentrations of elution buffer (20 mM Tris-HCl, pH 8.0, 1 M NaCl), which reached 100 % at 20 CVs. To finalize the gradient, 3 ml was used to compensate for the system volume. During this elution, fractions were collected continuously, and appropriate elution fractions containing proteins were analyzed via SDS PAGE.

2.5.11 ALAS activity assay *in vitro*

Coupled enzyme activity assay

The enzymatic activity of ALAS can be determined by a coupled enzyme assay (Hunter & Ferreira 1995, Hunter & Ferreira 2005). Hereby, ALAS produces ALA, together with CoA-SH and CO_2, using the substrates glycine and succinyl-CoA. CoA-SH will then be used by the alpha-ketoglutarate dehydrogenase (α-KGD) to reduce nicotinamid-adenin-dinucleotid (NAD^+) to NADH. The reduction of NAD^+ to NADH is thought to be stoichiometric with ALA formation. This product formation can be detected spectroscopically at 340 nm (Figure 9). NADH can accumulate in levels far above the starting concentration of succinyl-CoA, thus allowing the determination of initial rates even when submicromolar concentrations of succinyl-CoA are used.

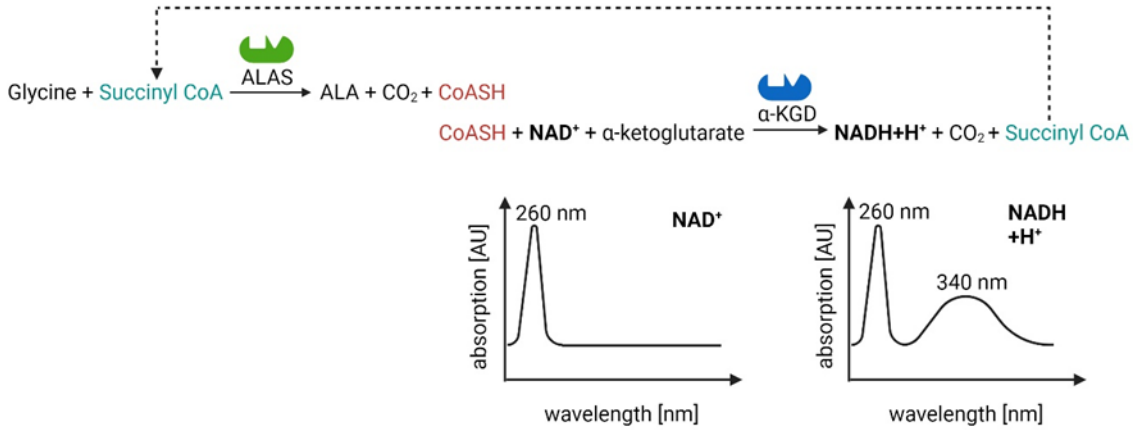


Figure 9: Coupled enzymatic reaction of ALAS for spectrometric detection of product formation. The enzymatic reaction of aminolevulinic acid synthase (ALAS) forms out of glycine and succinyl Coenzyme A (succinyl-CoA) the product 5-aminolevulinic acid (ALA), carbon dioxide and CoA-SH. The latter can be used by the enzyme α-ketoglutarate dehydrogenase (α-KGD) together with α-ketoglutarate to reduce nicotinamid-adenin-dinucleotid (NAD^+), while producing carbon dioxide and replenish the succinyl-CoA pool. This reduction is thought to be stoichiometric wit ALA formation. Replenish of the ALAS reaction enables working with submicromolar succinyl-CoA concentrations. The detection of product formation was done using spectroscopy measurements at 340 nm over time and the extinction factor NADH $\varepsilon_{340} = 6560\ M^{-1} * cm^{-1}$. Figure created with BioRender.com.

α-KGD (1 unit) will convert 1 µmol of NAD to NADH per min at pH 7.4 at 30 °C in the presence of saturating levels of coenzyme A. α-KGD from porcine heart is supplied as a 50 % glycerol solution consisting of ~9 mg per ml BSA, 30 % sucrose, 1.5 mM EDTA, 1.5 mM EGTA, 1.5 mM 2-mercaptoethanol, 0.3 % Trition X-100, 0.003 % sodium azide, 15 mM potassium phosphate pH 6.8 (Merck/Sigma Aldrich). TPP must be prepared freshly, and NAD aliquots must be stored at -80 °C and will only be stable for 6 months. PLP solubility is enhanced by adding powder in 20 mM HEPES buffer pH 7.5 and small amounts of 1 M NaOH until no precipitation is visible. It should be stored at -20 °C.

Reaction buffer:

20 mM HEPES (pH 7.5 NaOH)

3 mM $MgCl_2$

1 mM α-ketoglutaric acid sodium salt

100 µM NAD

0.25 mM thiamine pyrophosphate in H_2O

0.05 units α-KGD

2 µM ALAS

10 µM succinyl-CoA

10 mM glycine

Experiments were carried out at 30 °C in a total volume of 150 µl over a time period of 5 min. For general activity measurements of purified ALAS enzymes, all ingredients of the reaction buffer were mixed in a prewarmed quartz cuvette, except for glycine. This mixture was used as a blank, and the addition of glycine started the reaction. The produced amount of NADH can be quantified by the absorbance at 340 nm with the Lambert-Beer law using the extinction coefficient of NADH $\varepsilon_{340} = 6560\ M^{-1} * cm^{-1}$.

In order to determine kinetic parameters like K_m and v_{max} of RcA for both substrates, different concentrations for each substrate were tested. Thereby, one substrate concentration was varied, whereas the other substrate concentration remained constant. Tested succinyl-CoA concentrations: 0.05 µM, 0.1 µM, 0.221 µM, 1 µM, 5 µM, 10 µM and glycine concentrations 0.05 mM, 0.1 mM, 0.273 mM, 1 mM, 5 mM, 10 mM.

The observed rates were fitted to the Michaelis-Menten equation using nonlinear regression analysis software (Origin).

Colorimetric enzyme activity assay

The colorimetric detection of ALAS activity relies on the principle of ALA condensing to a pyrrole. The enzymatic activity of ALAS, through product formation from glycine and succinyl-CoA, can be quantified spectroscopically. The formed ALA is condensed with acetylacetone at pH 4.7 to yield 2-methyl-3-acetyl-5-propionic acid pyrrole (ALA-pyrrole). In the presence of modified Ehrlich reagent, ALA-pyrrole forms a colored complex which is measured at 554 nm in the spectrophotometer (Burnham 1970) (Figure 10).

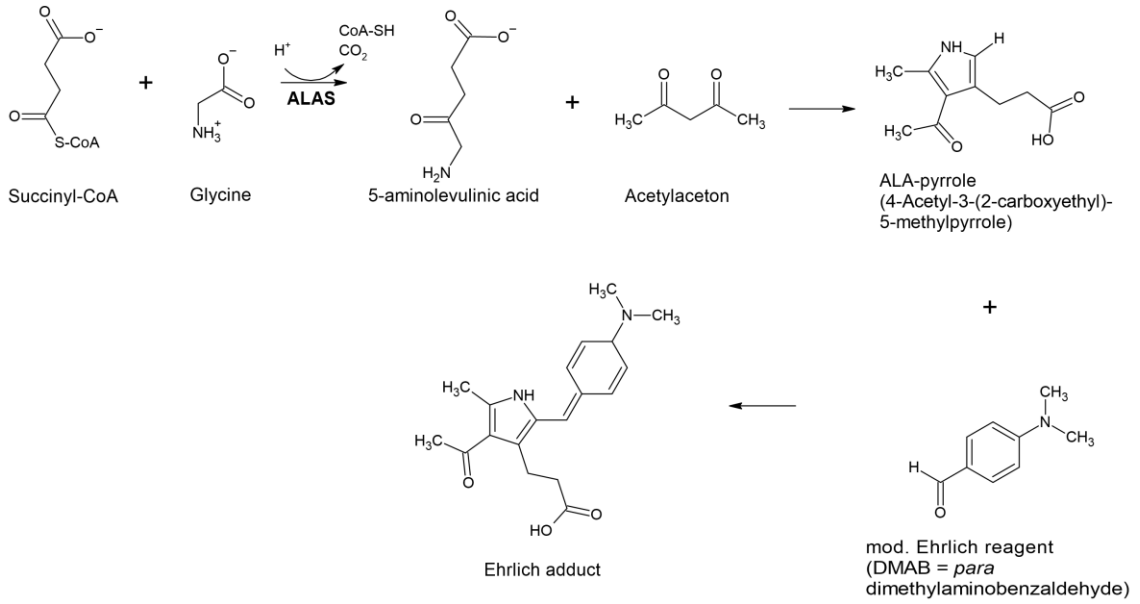


Figure 10: The colorimetric detection of ALAS product formation using Ehrlich reagent. The enzymatic production of 5-aminolevulinic acid (ALA) can be detected using pyrrole formation catalyzed by acetylacetone, which yields the ALA-pyrrole. The formation of a colored complex can be carried out by mixing this pyrrole with modified Ehrlich reagent and measuring the resulting Ehrlich adduct at 554 nm spectroscopically.

The enzyme was dialyzed in the following buffer (50 mM HEPES, 200 mM NaCl, 20 µM PLP, pH 7.5 NaOH). The modified Ehrlich reagent (for 50 ml= 30 ml glacial acetic acid dissolve 1 g p-Dimethyl-aminobenzaldehyde add 8 ml 70 % (v/v) perchloric acid and fill up to 50 ml with glacial acetic acid) had to be prepared freshly (Burnham 1970, Mauzerall & Granick 1956)). Experiments were carried out in a total volume of 1 ml. The reaction was started by adding glycine to the premixed ingredients of the reaction buffer. This mixture was incubated at 37 °C in a thermoblock for 30 min or 1 h. The reaction was terminated by adding 0.5 ml of 10 % (w/v) TCA and centrifugation of the

sample at 4 °C for 5 min at 17 800 * g at a benchtop centrifuge. The supernatant was transferred to test tubes containing 2 ml of 1 M acetate buffer (pH 4.7). To obtain a blank value, a sample consisting of 1 ml buffer was added to 2 ml of 1 M acetate buffer. 50 µl acetylacetone was added to the test tubes, and they were boiled in a water bath at 100 °C for 15 min. Upon cooling to room temperature, 3.5 ml of freshly prepared modified Ehrlich reagent was added and mixed. The absorption was measured after 20 min at 554 nm. Since glycine alone can interact with the Ehrlich reagent, this had to be subtracted from the determined value of the ALAS reaction. Therefore, a substrate control consisting of glycine in the buffer was used, that had the same concentration as in the reaction buffer. A standard curve with known ALA concentration was used to determine the amount of ALA produced by the ALAS enzyme with regard to the reaction time.

Reaction buffer:

50 mM HEPES, 200 mM NaCl, pH 7.5 (NaOH)

5 µM ALAS

0.5 mM PLP

5 mM glycine

100 µM succinyl-CoA

2.5.12 ALAS activity assay in cell lysate and porphyrin detection

Functional complementation of *E. coli* ST18, which is ALA auxotrophic, was used to determine ALAS activity *in vivo.* Therefore, plasmids containing either genes for *hemA* (GluTR), *alaS* (RcA or viral ALAS), or EV (empty vectors) were transformed in freshly prepared chemical competent cells of ST18 (see chapter 2.3.5, 2.3.6) and set up similar to the growth curve (2.3.7) except that it consisted of 50 ml medium. Plasmids pRcA and pMkA were used as positive controls for functional complementation. Plasmid pCYB1-ALAScryopt named vALAS was used for complementation for a viral freshwater ALAS. Empty vector controls were pCYB1 and pGEX-6P, respectively.

ALAS activity assay in cell lysate

In 50 ml medium, *E. coli* cultures consisting of *E. coli* ST18 (plus complementary plasmids or empty vectors) were inoculated and grown as stated before (see chapter 2.3.7) for 24 h covered in aluminum foil at 37 °C shaking. The cells were harvested by centrifugation of the culture for 20 min at 3,700 * g at 4 °C. The pellet was resuspended and washed twice with 50 mM Tris-HCl buffer pH 7.5. In order to lyse the cells, the pellets were resuspended in 500 µl lysis buffer (50 mM Tris-HCl buffer, pH 7.5) for each 100 mg of pellet wet weight and sonicated. This was carried out in an ice-water bath for a total of 1 min 30 sec sonification time with intervals of 30 sec sonification and 1 min cool down. Sonification pulse was 40 % intensity, 5 x cycle. The cell lysate was separated from debris by centrifugation for 30 min at 17,000 * g at 4 °C. The supernatant was taken off, and protein content was determined using Bradford reagent (Roti®Quant) in 96-well plate as recommended by the manufacturer (see chapter 2.5.4). Therefore, a calibration curve with known protein concentrations of BSA was used, and samples were measured in a plate reader (TECAN Infinite® F200 Pro) at 595 nm.

For determination of internal ALA formation, the protein content was set to 2 mg/ml, and substrates of ALAS (glycine, succinyl-CoA) and cofactor (PLP) were added to measure product formation (ALA) over time. In a 2 ml reaction tube, a 1 ml approach in 50 mM Tris-HCl buffer (pH 7.5) was carried out with 5 mM glycine, 100 µM succinyl-CoA, and 0.5 mM PLP. The reaction was started by the addition of the cell lysate [2 mg/ml] and incubated at 37 °C on a heating block for different time points (0 min, 15 min, 30 min, 60 min). The reaction was terminated by adding 0.5 ml 10 %(w/v) TCA. Following a centrifugation step at 4 °C for 5 min at 17,800 * g, the supernatant was mixed in a test tube with 2 ml of 1 M sodium acetate buffer (pH 4.7). Pyrrole formation was achieved by mixing the reaction mixture with 50 µl acetylacetone. The test tube was then capped and boiled in a water bath at 100 °C for 15 min. The cooling to room temperature followed the addition of 3.5 ml freshly prepared modified Ehrlich reagent (as described before, see chapter 2.5.11) to the reaction mix. Detection of pyrrole formation was measured after 5 min incubation at 554 nm. The extinction coefficient of ALA pyrrole 66,000 M^{-1} *cm^{-1} (Marver *et al.,* 1966) was used to determine ALA formation regarding the protein content used in mg of total protein content. Timepoint zero represents the

present ALA concentration in the cell lysate, but to calculate only the new ALA formation in the other timepoints, timepoint 0 was subtracted from it. This assay detects not only ALA but also amino ketones. Ehrlich reagent is not stable and should be prepared freshly. The data shown represents the mean values of all measurements and the respective standard deviation. Measurement was conducted with three independent biological replicates and two technical replicates. For statistical determination, values were tested for normal distribution using Shapiro-Wilk, and normally distributed data was analyzed using one-way ANOVA with Bonferroni correction and paired T-Test to determine significance levels of each timepoint against timepoint 0. Significance levels were assigned accordingly $p \leq 0.05$ equals *; $p \leq 0.01$ equals ** and $p \leq 0.001$ equals ***.

Porphyrin detection

Porphyrins are present as intermediates in the synthesis pathway of heme and other tetrapyrroles. They can be detected with spectroscopy and fluorescence measurements because of their highly conjugated π-electron system. For the detection, the debris-free lysed cells from the complementation experiment were used (see above). The UV-Vis and fluorescence measurements were conducted in quartz cuvettes using an Agilent 8453 series spectrophotometer and Jasco FP-8300 fluorometer with protein solutions set to 2 mg/ml concentration. The absorption curves were baseline subtracted using Origin software. The fluorescence measurements were done by excitation at 400 nm and emission reading after 415 nm to 750 nm. The settings were configured as follows: Excitation bandwidth and emission bandwidth were 5 nm while the response time was 50 msec at medium sensitivity.

2.5.13 ALA secretion into the medium of complemented *E. coli* strains

ALA is thought to be secreted through the transporter RhtA (Yang *et al.,* 2016) into the medium. The concentration of ALA in the medium of the *E. coli* strains was determined after conducting the growth curve (see chapter 2.3.7) using cell-free medium. Therefore, a colorimetric detection was done using modified Ehrlich reagent (Burnham 1970, Mauzerall & Granick 1956). At the end of the growth curve, a sample was taken from the culture supernatant after centrifugation for 5 min at 13,000 x g. From the now cell-free medium, 100 µl was used to determine ALA concentration using freshly prepared

modified Ehrlich reagent (as described before, see chapter 2.5.11). Therefore, a 100 µl sample of the cell-free medium was mixed with 0.5 ml 1 M sodium acetate buffer (pH 4.7) and 0.5 ml H_2O while 50 µl acetylacetone was added for pyrrole formation. The sample was then boiled for 10 min at 100 °C in a heating block. After cooling to room temperature, 3 ml modified Ehrlich reagent was added and absorption at 554 nm was measured instantly. The concentration of ALA was calculated by using a standard curve that was created with known ALA concentrations diluted in LB medium. The concentrations were then normalized to the cell culture $OD_{600\,nm}$ = 1.

The data shown represents the mean values of all measurements and the respective standard deviation. All measurements were performed in three biological replicates with three technical replicates. For statistical determination values were tested for normal distribution using Shapiro-Wilk, random distributed data was analyzed using the U-Test by Mann and Whitney, whereas normal distributed data was tested with unpaired T-Test. Significance levels were assigned accordingly $p \leq 0.05$ equals *; $p \leq 0.01$ equals ** and $p \leq 0.001$ equals ***.

2.5.14 Heme oxygenase assay

Utilizing the heme oxygenase assay enables the *in vitro* verification of the activity of the purified heme oxygenases (Wegele *et al*., 2004). The heme oxygenase assay can detect the enzymatic ability of an enzyme to form BV IX from heme by inhibiting nonspecific cleavage of heme through ROS formation due to the presence of certain substances in the enzyme assay mixture. Hemin was used as a substrate, and the heme oxygenase was assayed in equimolar amounts. The required electrons are supplied by a NADPH-regenerating system (NRS) to catalyze the heme degradation to BV IX. The enzyme glucose-6-phosphate-dehydrogenase (G-6-P-DH) oxidizes its substrate glucose-6-phosphate, thus reducing, with the gained electrons, $NADP^+$ to NADPH. The enzyme ferredoxin:$NADP^+$ oxidoreductase (FNR) catalyzes the reduction of oxidized ferredoxin by oxidizing NADPH. The reduced ferredoxin now serves as an electron donor for the heme oxygenase, which can catalyze the reduction of heme to BV IX (Figure 11). Crucial is the presence of 4,5-dihydroxy-1,3-benzenedisulfonate (Tiron), ascorbate, and catalase to verify enzymatic heme degradation and omit nonenzymatic heme

degradation. The employed ferredoxin was from the cyanophage P-SSM2, and FNR was PetH from *Synechococcus* sp. PCC 7002.

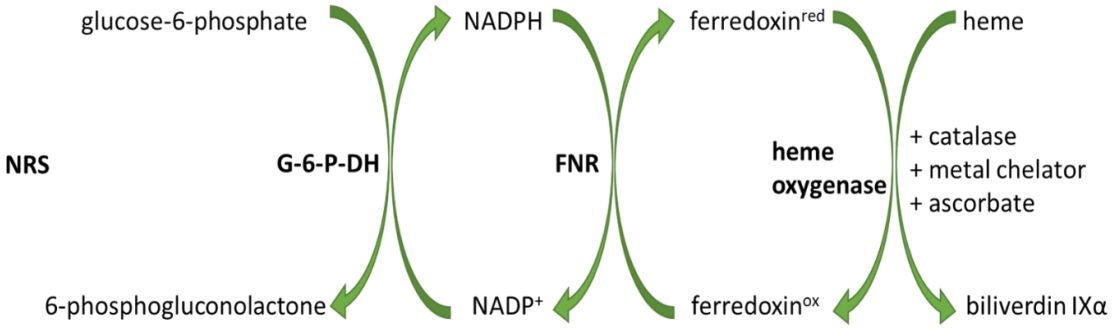


Figure 11: Schematic representation of the ferredoxin-dependent heme oxygenase assay. Heme oxygenase catalyzes heme reduction to biliverdin IXα (BV IXα), for which the ferredoxin serves as an electron donor. The oxidized ferredoxin is regenerated by NADPH and ferredoxin:NADP$^+$ oxidoreductase (FNR). NADP$^+$ is regenerated by the NADPH regenerating system (NRS) consisting of the enzyme glucose-6-phosphate-dehydrogenase (G-6-P-DH) and glucose-6-phosphate.

First, all components except for the NRS were mixed and transferred into a 3 mm quartz cuvette. The reaction was monitored at RT by UV-Vis spectroscopy, where spectra were recorded every 30 s when the reaction was started by the addition of NRS. The first spectrum was recorded before adding NRS. After the addition of NRS, the turnover was monitored for a total reaction time of 10 min. The reaction was stopped by adding 10 volumes of ice-cold 0.1 % (v/v) trifluoroacetic acid (TFA). Afterward, the samples were prepared for HPLC analyses (see chapter 2.5.16). All components for the assay and the heme oxygenase were present in KPi buffer (100 mM K_2HPO_4 pH 7.2).

Components of the standard approach	**Concentration of components**
Heme oxygenase and heme	20 µM each
BSA	0.15 mg/ml
$PetF_{P\text{-}SSM2}$ (ferredoxin)	4.6 µM
$PetH_{Syn}$ (FNR)	0.01 µM
Ascorbate	5 mM
Catalase	10 µM
Tiron	5 mM
NRS	7.1 µl
KPi buffer	fill up to 500 µl

NADPH regenerating system (NRS; 50 µl)	
NADP$^+$	7 mM
Glucose-6-phosphate	70 mM
Glucose-6-phosphate dehydrogenase	10 U/ml

2.5.15 Anaerobic FDBR assay

The anaerobic FDBR assay was used to detect the enzymatic activity of ferredoxin-dependent bilin reductases (FDBR) *in vitro* (Figure 12). The assay was performed as described previously with a few modifications (Ledermann *et al.*, 2016, Tu *et al.*, 2004). Its components serve the purpose of maintaining a continuous flow of electrons to enable the reducing activity of the reductase to be measured by UV-Vis spectroscopy. The required electrons are provided by the NADPH-regenerating system (NRS). Glucose-6-phosphate is oxidized by the enzyme glucose-6-phosphate dehydrogenase, whereby the obtained electrons enable the reduction of $NADP^+$ to NADPH. The electrons are then transferred from NADPH to $ferredoxin^{ox}$ by the enzyme ferredoxin:$NADP^+$ oxidoreductase (FNR). The reduced ferredoxin now acts as an electron donor for the bilin reductase, which catalyzes the reduction of BV IXα (Figure 12). Due to the different absorption maxima of the bilins, this enzymatic reduction can be detected spectroscopically. The assay was performed under anaerobic conditions as standard in order to stabilize and detect the substrate radicals formed during the reaction.

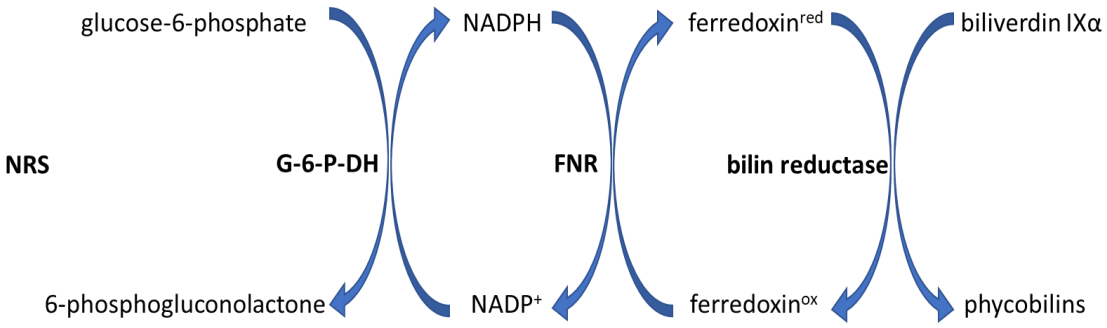


Figure 12: Schematic representation of the ferredoxin-dependent bilin reductase assay. Bilin reductases catalyze the reduction of BV IXα to phycobilins. The ferredoxin serves as an electron donor. The regeneration of the oxidized ferredoxin is done by NADPH and FNR. NADP+ is regenerated by the NADPH regenerating system (NRS) which consists of the enzyme glucose-6-phosphate-dehydrogenase (G-6-P-DH) and its substrate glucose-6-phosphate.

The components, except for the NRS and the FDBR:phycobilin complex, were gassed with nitrogen (N_2) in a quartz cuvette for approx. 15 min while stirring at 20 °C. This approach served as a reference for UV-Vis spectroscopy. Afterward, the FDBR:phycobilin complex was added, and the start of the reaction was triggered by the addition of the NRS. The reaction took place for 20 min at 20 °C and under continuous N_2 gassing. A spectrum in the range of 200 to 900 nm was measured every 30 seconds. The reaction was stopped by diluting the mixture 1:10 in ice-cold 0.1 % (v/v) TFA, which was then

prepared for HPLC analyses (see chapter 2.5.16). All components for the assay and the bilin reductase were present in TES-KCl buffer (25 mM TES-KOH pH 7.2, 100 mM KCl). Reaction products were identified by comparison of the retention time with known standards and whole spectrum analysis of the elution peaks.

Components of the standard approach	Concentration of components
BSA	10 µM
Glucose	100 mM
Glucose oxidase	50 U/min
Catalase	5 µM
$PetF_{P\text{-}SSM2}$ (ferredoxin)	1 µM
$PetH_{Syn}$ (FNR)	0.01 µM
FDBR:Phycobilin complex	10 µM
NRS	100 µl
TES buffer	fill up to 2 ml

NADPH regenerating system (NRS; 100 µl)	
$NADP^+$	8.2 mM
Glucose-6-phosphate	65 mM
Glucose-6-phosphate dehydrogenase	11 U/ml

2.5.16 High-performance liquid chromatography (HPLC)

HPLC (high-performance liquid chromatography) enables the separation of a mixture of substances according to their polarity. The stationary and mobile phases of HPLC have different polarities with which the substances interact in different ways. As a result, each substance of the analyte mixture exhibits a specific retention time, which allows for separation and characterization.

Sample preparation for HPLC

The bilins produced in the FDBR and heme oxygenase assay were separated from the non-polar components in the mixture by a solid phase extraction with C18 Sep-Pak® light cartridges from Waters. The columns were prepared with two runs, each with the following solvents:

Acetonitrile	3 ml
H_2O	3 ml
0.1 % (v/v) TFA	3 ml
10 % (v/v) MeOH in 0.1 % (v/v) TFA	3 ml

The prepared Sep-Pak® columns were then loaded with the reaction mixtures diluted before in 0.1 % (v/v) TFA. The columns were washed with 5 ml 0.1 % (v/v) TFA and eluted with 1 ml acetonitrile. After freezing the eluate for at least 1 h at -80 °C, the samples were transferred and freeze-dried overnight (Alpha 2-4 LSC plus, Christ). The dried samples were then dissolved in 10 - 20 µl dimethyl sulfoxide (DMSO) and filled to 300 µl with filter-sterilized HPLC solvent. After filtration with a 0.2 µm polytetrafluorethylene filter (Phenomenex), the samples were applied to the HPLC system.

HPLC solvent	
Acetone	50 % (v/v)
20 mM formic acid	50 % (v/v)

HPLC analysis of reaction products

The analysis of the bilin reaction products was carried out as described before (Ledermann *et al*., 2016, Tu *et al*., 2004). For the separation of the substance mixtures, an HPLC system from Agilent Technologies (1100 series) with a C18 reversed-phase column from Phenomenex (Luna 5µm C18 (2) 100A) was used. The column (storage solution = acetonitrile) was first washed with degassed distilled water (filtered and degassed) and then equilibrated with HPLC solvent (filtered and degassed). The samples were loaded manually via a Hamilton syringe through a 200 µl sample loop. The flow rate was constant at 0.6 ml/min over a 30 to 40 min measuring period. The separation of the linear tetrapyrroles was detected by a diode array detector at 380 nm, 560 nm, and 650 nm. For exact characterization, the samples were compared with retention times of known tetrapyrroles bought as standards or purified reaction products from characterized enzymes. Heme oxygenase product BV IXα was analyzed through the reaction product of BphO (Ho1 of *Pseudomonas aeruginosa*) activity. Meanwhile, the PcyA product PCB was analyzed using the reaction product of PcyA from *Anabaena* sp. PCC 7120.

2.5.17 Phytochrome assay

The FDBRs generally show low sequence similarity to each other. Therefore, PcyA or PcyX activity can be determined using a Phytochrome assay (Gambetta & Lagarias 2001, Terry & Lagarias 1991). Consequently, a phytochrome (photoreceptor) is required, which assembles autocatalytically with PCB, PΦB, and PEB by forming a thioether bond with its conserved cysteine residue. When bound to PCB or PΦB, photoconversion is possible due to the presence of the C15 C16 double bond at the C to D ring. This can be present in *E* or *Z* form when illuminated with red or far-red light. PEB has only a single bond and thus cannot photoconvert but forms a highly fluorescent phycobiliprotein named phytofluor (Figure 13)(Li *et al.,* 1995).

When PCB is bound to a phytochrome like Cph1 from *Synechocystis* PCC 6803 and illuminated by far-red light (730 nm), it produces the P_r form, while after red (636 nm) irradiation it produces a mixture of P_r and P_{fr}. When these forms are subtracted from each other (P_r-P_{fr}), they exhibit a specific delta absorbance maximum and minimum. The values can be used to determine whether PCB or PΦB is bound. For this, coexpression in *E. coli* BL21(DE3) was carried out using the plasmid the pET_His-tag Cph1 crystal and either pTD*ho1*_v*pcyA* or pTD_v*ho1*_v*pcyA*. The inoculation and growth conditions were as stated before. Shortly after reaching $OD_{600\ nm}$= 0.5, the culture was induced with either 0.1 mM or 1 mM IPTG and incubated at 17 °C O/N. The cells were harvested, the pellet resuspended in PBS buffer, sonicated, and cell debris was removed by centrifugation. The lysate containing Ho1 (CB_2), PcyA (CB_2), and Cph1 was further enriched by affinity chromatography, carried out using TALON column material (see 2.5.3). UV-VIS detection was carried out using a 1 mm quartz cuvette with PBS as blank. The lysate was incubated for 3 min in red illumination (P_{fr} form) and 3 min in far-red illumination (P_r form). Measurements were performed multiple times, resulting in a spectrum with reduced interference signals. Finally, the P_{fr} spectra get subtracted from the P_r spectra, resulting in the difference spectrum, which was smoothed using Origin 2021 software with filter FFT. The absorbance maximum for PCB is 650 nm, and the minimum is 706 nm, while for PΦB, the maximum is 670 nm and the minimum is 718 nm (Gambetta & Lagarias 2001, Terry & Lagarias 1991). The following filters were used: Interference filter red (636FS10-50), the light intensity with 12 micromoles $m^{-2}sec^{-1}$ /

*0.1 Watts m^{-2} / *10 Lux, and interference filter far-red (730FS10-50), the light intensity with 4 micromoles $m^{-2}sec^{-1}$ / *0.1 Watts m^{-2} / *10 Lux.

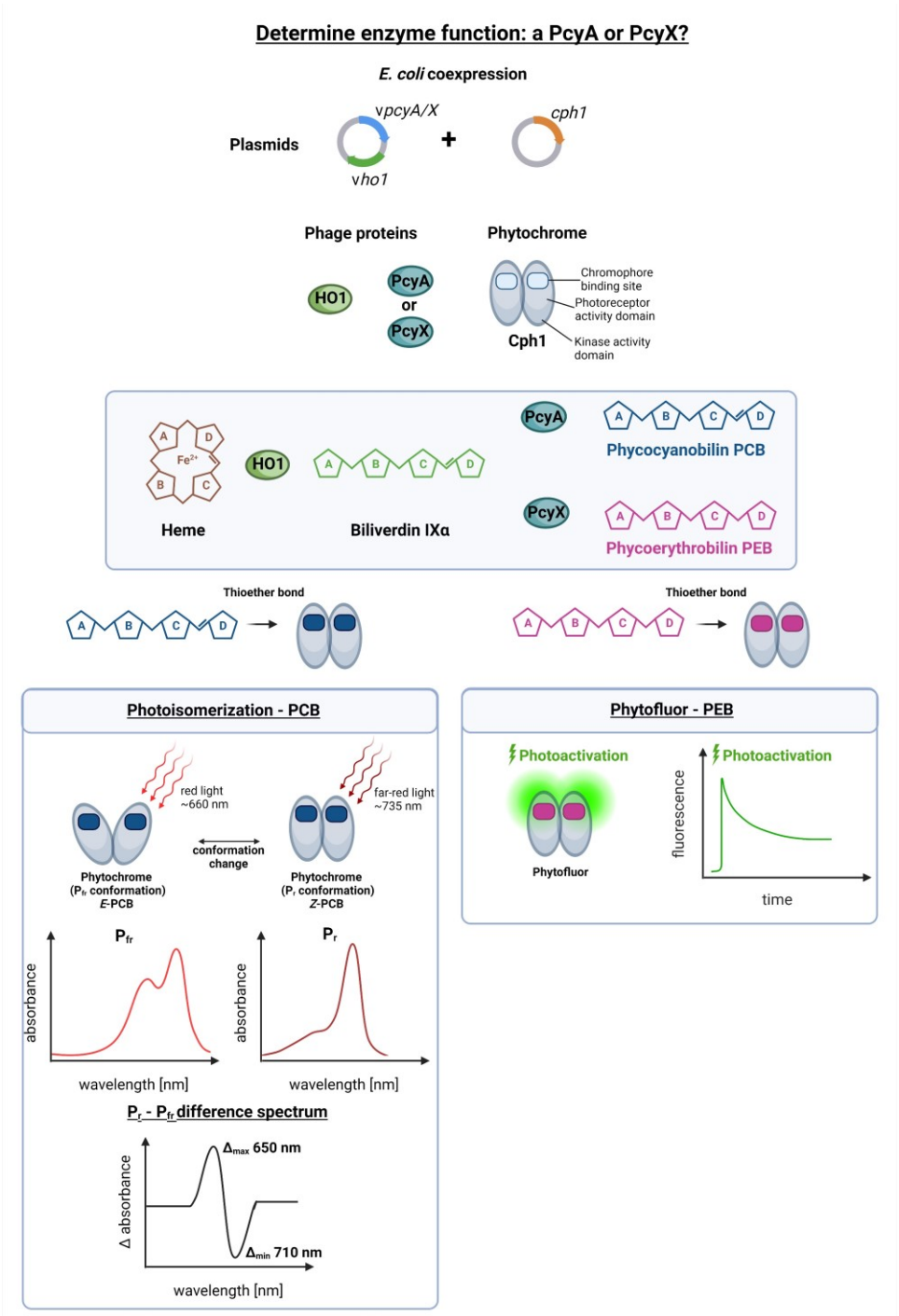


Figure 13: Determination of enzyme function regarding product formation through the presence of a phytochrome. The ability of phytochromes to bind specific linear tetrapyrroles was used to determine the enzymatic product formation of the viral enzyme PcyA/X belonging to the FDBRs. Heterologous coexpression in *E. coli* of the genes encoding for a heme oxygenase and the FDBR of unknown function, together with a gene encoding for a phytochrome enable the determination of the viral FDBR activity. The *E. coli* continuously produces heme which can be degraded using the introduced heme oxygenase to biliverdin IX that serves as substrate of the to be tested FDBR PcyA/X. The enzymatic activity of PcyA would result in the formation of a functional phytochrome, which can then be analyzed using red/far-red light illumination and gives a characteristic difference spectrum. Whereas if the tested FDBR is an PcyX this would result in the bilin PEB to be produced which can bind to the phytochrome but has fluorescence properties which can also be detected. Figure created with BioRender.com.

2.5.18 Thermal shift assay

The thermal shift assay is a method used to assess protein stability. It involves gradually heating a protein of interest while monitoring changes in fluorescence using a dye, such as SYPRO^{TM} orange. As the protein heats up, its hydrophobic regions are exposed, allowing the dye to bind and causing an increase in fluorescence. The protein melting point (T_M) is the temperature at which the protein is at the midpoint of a sigmoid curve. It can be calculated by analyzing the vertex of the first derivative of the best-fit curve. The temperature and fluorescence monitoring are done using a real-time PCR machine (BioRad CFX Connect^{TM} Real-Time System). The concentration of SYPRO^{TM} orange was 100x and 2 mg/ml protein in a 96-well plate. The thermal shift assay protocol consisted of the scan mode FRET starting with incubation at 10 °C for 15 min and then the first measurement. This is followed by a stepwise increase of the temperature for 0.5 °C until reaching 95 °C by a hold at each step for 10 sec and a measurement. Finally, samples were cooled down to 12 °C, measured after 30 sec, and held at this temperature.

2.6 *In silico* analysis

Phylogenetic trees

Sequences were aligned using Mafft and depicted in a phylogenetic tree using the Next Generation Phylogeny.fr web service (NGPhylogeny.fr, https://ngphylogeny.fr) by setting 'PhyML+SMS/OneClick' (Criscuolo & Gribaldo 2010, Guindon *et al.,* 2010, Katoh & Standley 2013, Lefort *et al.,* 2017, Lemoine *et al.,* 2018, Lemoine *et al.,* 2019). The output tree was then further adapted by Midpoint rooting and labeled using iTOL (interactive Tree of Life, version 6.8) via the webserver itol.embl.de (Letunic & Bork 2021). Further, bootstrap was assigned when values were between 0.75 and 1 and marked with a blue circle on the tree branch.

Multiple sequence alignment

Amino acid sequences of reference bacteria or phages were derived from NCBI. Freshwater and marine ALAS sequences were derived from the sources stated before (see 2.1.4). As a reference sequence, RcA or *Synechocystis* sp. PCC6803 Ho1 or PcyA was selected and are fixed at the top of the alignment. The alignment was created using Clustal Omega (Sievers *et al.,* 2011) and evaluated using ESPript v.3.0 (Robert & Gouet

2014). Important catalytic residues are marked with a black line at the top of the alignment, referring to crucial amino acid residues known from characterized proteins. For ALAS the crystallization publication of RcA was selected (Astner *et al*., 2005), for PcyA/X (Dammeyer *et al*., 2008a, Frankenberg & Lagarias 2003, Ledermann *et al*., 2016, Ledermann *et al*., 2018, Tu *et al*., 2007) and for Ho1 (Caignan *et al*., 2002, Lightning *et al*., 2001, Schuller *et al*., 1999, Schuller *et al*., 2001, Sugishima *et al*., 2004, Wegele *et al*., 2004). For ALAS alignment, residues marked in blue lines at the top are indicating crucial heme axial ligands of ALAS from *Caulobacter crescentus*, known to facilitate heme inhibition (Ikushiro *et al*., 2018).

Structural modeling

Prediction of the three-dimensional structure of the phage CB_2 proteins ALAScryopt, Ho1, and PcyA were generated using AlphaFold2 (Jumper *et al.,* 2021) with ColabFold (Mirdita *et al.,* 2022) by utilizing the AlphaFold2 advanced notebook. Overlays were done utilizing PyMOL v.2.0 (Schrödinger 2020). For overlay with ALAS, the crystal structure of ALAS from *Rhodobacter capsulatus*, RcA (RCSB PDB code: 2BWN, 2BWO, 2BWP) (Astner *et al*., 2005) was used. The overlay of vHo1 and vPcyA was done using *Synechocystis* sp. PCC 6803 Ho1 in complex with protoporphyrin IX containing iron (RCSB PDB code: 1WE1) (Sugishima *et al*., 2004) and PcyA from the same organism in complex with BV IXα (RCSB PDB code: 2D1E) (Hagiwara *et al.,* 2006).

3. Results

Many phage genomes carry auxiliary metabolic genes, often related to photosynthesis or other crucial steps in host metabolite synthesis. However, their role during infection of a bacterial host is not yet completely understood. Based on the finding of a conserved cluster of genes in metagenomic datasets, which seemed responsible for heme biosynthesis and linear tetrapyrrole formation, their function should be investigated.

Detection of novel three gene cassette in viral metagenomic datasets

A conserved gene cassette was detected in multiple viral metagenomic datasets from marine sampling sites. This gene cassette encodes for three enzymes responsible for the metabolism of tetrapyrroles, namely *ho1*, *pcyA/X,* and *alaS* (Figure 14). Previously identified in marine viral datasets were the *ho1* and *pcyX* genes that encode functional proteins involved in synthesizing linear tetrapyrroles from heme. The third gene, *alaS,* encodes for a putative aminolevulinic acid synthase (ALAS) that was never detected as a three-gene cassette in viral metagenomic data, nor has the viral enzyme been proven functional. Expanding the search to freshwater environments, the genes were also found in metagenomic datasets derived from freshwater sampling sites (Figure 14) (Chen *et al*., 2020). Here, the genes are organized differently, while *pcyA/X* and *ho1* are still present as a cassette, *alaS* is not located nearby but further up- and downstream of the phage contig (Figure 14). Regarding the organization of the found cassette, this occurrence was found to be habitat-specific as multiple datasets verified. Additionally, single ALAS sequences or broken cassettes were found in marine datasets (Figure 14) (Roitman 2022, Wegner *et al*., 2024).

Since the phages are not isolated and thus not culturable, the enzymes of interest were heterologously produced in *E. coli* to determine their function. Characterizing their activity may give physiological insights to further address their role and impact during phage infection.

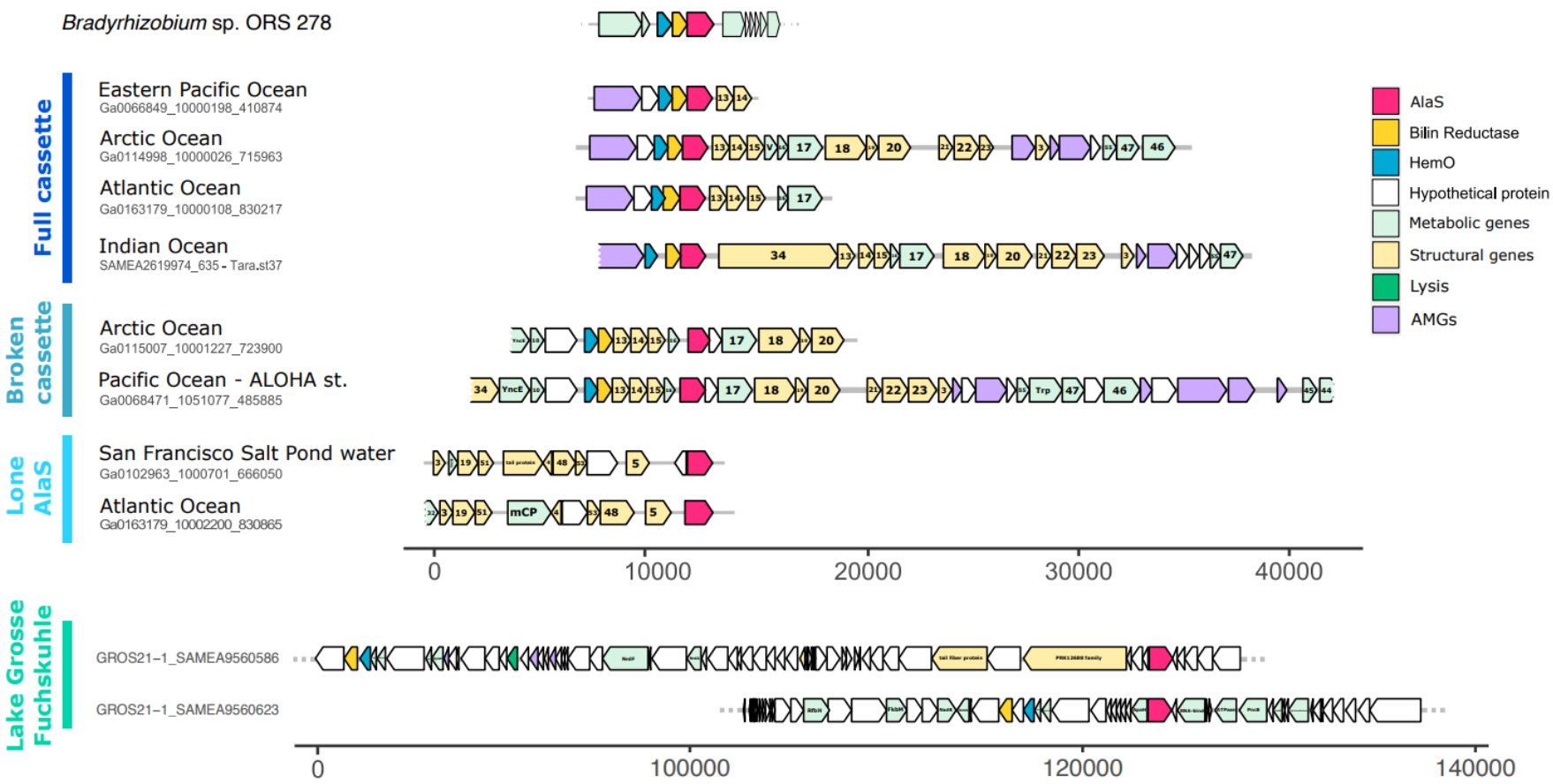


Figure 14: Genomic organization and representation of different constructed contigs containing *alaS, ho1* and an encoded FDBR of viral and bacterial origin. Displayed is schematic representation of characteristic genomic contigs from bacterial origin (‘*Bradyrhizobium* sp. ORS 278’), viral marine environment (labelled in different blue shades as ‘Full/Broken cassette’ and ‘Lone AlaS’) and viral freshwater environment (labelled as ‘Lake Grosse Fuchskuhle’ in turquoise). The contigs belonging to ‘Lone AlaS’ show the absence of the other genes encoding for an heme oxygenase (HemO) and an FDBR in its contigs. Respective genes are colored as described in the genome legend. The annotations of genes are depicted within the gene if available, while numbering of genes refer to ‘gene products (gp)’. The scale bar represents base pairs (bp) and dotted lines at contigs indicate that they are longer than displayed here. Sequences for the contigs can be found in Supplementary Data 1 of the publication in revision (Wegner *et al.,* 2024) and the picture was created by Sheila Roitman.

3.1 Characterization of viral heme oxygenase

Heme oxygenases are known to facilitate the first crucial step in linear tetrapyrrole biosynthesis by cleaving one methine bridge of heme, thus generating BV IX. Marine Ho1s from viral metagenomic data were characterized before and showed activity by forming BV IXα (Dammeyer *et al*., 2008a). In order to verify the activity of freshwater HOs found in viral metagenomic data from a peat bog named Crystal Bog (United States), one Ho1 from the phage CB_2 (Crystal_bog_pmoC-phage_2) was analyzed (Chen *et al*., 2020). This *ho1* gene occurs together with a *pcyA/X* and further downstream an *alaS* on the phage genome of CB_2.

Before characterizing the biochemical activity, conserved catalytic residues should be analyzed and determined for the viral Ho1 (vHo1). Essential residues are H17 (numbered as Ho1 from *Synechocystis* sp. PCC6803), the R/K10, and Y125 (Figure 15 A). While the H17 is crucial for the interaction with the iron in the heme, R/K10 and Y125 stabilize the bond to the heme propionate groups by forming a hydrogen bond (Y125) and electrostatic interactions (R/K10). As the histidine is conserved in all aligned HOs, there are some variances for the other residues. The HemO from *Pseudomonas aeruginosa* is known to cleave BV IX at the β and δ methine bridge due to the missing stabilization with the heme propionates (N11 and F125), leading to an in-plane rotated heme and the delta-meso carbon is now located within the HO fold (Caignan *et al*., 2002). The presence of the stabilizing amino acids in vHo1 thus suggests an HO activity capable of forming BV IXα. As sequence alignments are only indications of the activity, the structural fold is often crucial for enzyme activity, regarding the limitation of *in silico* analysis. Therefore, the vHo1 structure was predicted using AlphaFold and overlayed with a characterized Ho1 of known structure (Ho1 from *Synechocystis* sp. PCC6803). The overlay shows significant similarities between the vHo1 (red) and the cyanobacterial one (blue) (Figure 15 B). Only the C-terminus seems shorter compared to the cyanobacterial Ho1. Also, the coordination of the Fe-protoporphyrin IX reveals no difference from the cyanobacterial Ho1, except that R10 is K10 in the vHo1, this amino acid is also known from other Ho1s to be present at this position (Figure 15 A), e.g., heme oxygenase from *Neisseria meningitidis* 63006, forms BV IXα, and has the same amino acid at this position (Schuller *et al*., 2001).

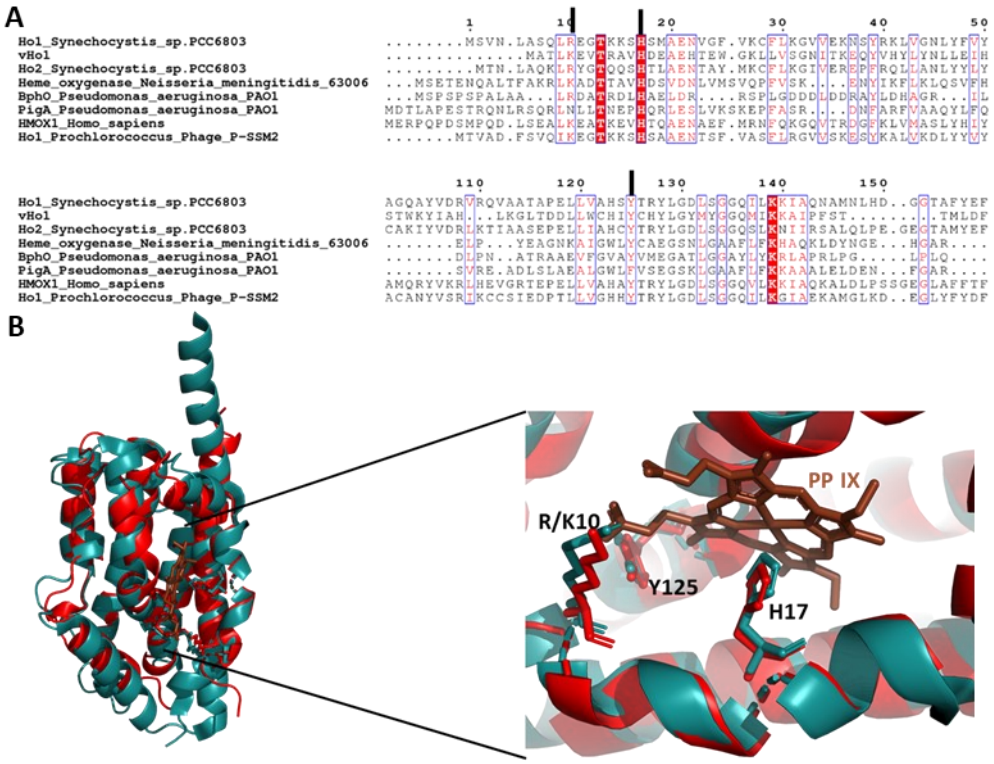


Figure 15: Sequence alignment and structural analysis of putative viral heme oxygenase. A Detailed view of reference sequences from the NCBI database which were aligned with sequence of viral heme oxygenase using Clustal Omega and ESPript 3. Highly conserved amino acids are highlighted in a red square with white lettering, similar amino acids are represented by a white box with red lettering, and non-similar amino acids are indicated by black lettering. Catalytic important amino acids, previously characterized (Caignan *et al.,* 2002, Lightning *et al.,* 2001, Schuller *et al.,* 1999, Schuller *et al.,* 2001, Sugishima *et al.,* 2004, Wegele *et al.,* 2004), are denoted by black lines at the top. Sequences can be found in the Appendix I + II. **B** Overlay model of monomeric Ho1 from *Synechocystis* sp. PCC6803 (blue, RCSB PDB code: 1WE1) (Sugishima *et al*., 2004) and the model of vHo1 predicted via AlphaFold (red), shows high structural similarity. The displayed overlay model was generated with PyMol, the magnified view shows binding of Fe-protoporphyrin IX (PP IX) with the important amino acids R10, H17, Y125 (sticks in blue/red) and PP IX in brown as chemical structure.

The predicted enzyme activity should now be demonstrated and compared with the experimental activity. Therefore, heterologous expression in *E. coli* of the codon adapted *ho1* was carried out. The protein vHo1 was tagged with an N-terminal GST-tag and purified via affinity chromatography (Figure 16). To minimize the possible interference on the following heme oxygenase assay, the N-terminal GST-tag was removed using PreScission Protease treatment and second affinity purification using glutathione agarose resin, which also removes the protease. The removed GST-tag (~26 kDa) could only be found in minor abundance in the purified vHo1 (~21 kDa) (Figure 17). The purified enzyme was dialyzed in the respective buffer to test the heme oxygenase activity.

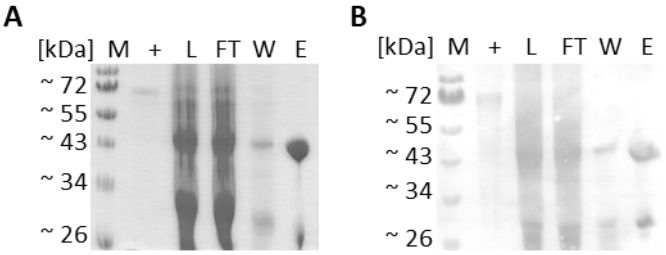


Figure 16: Purification of heterologous produced viral heme oxygenase of freshwater phage CB_2 via affinity chromatography. The N-terminal GST-tagged vHo1 from freshwater phage CB_2 was produced in *E. coli* BL21(DE3) in LB medium, the gene sequence of v*ho1* was codon adapted for *E. coli* codon usage. M= Size marker – Color Prestained Protein Standard, Broad Range (NEB); += positive control GST-ALAS from *Rhodobacter capsulatus;* L= Lysate; FT= Flow-through; W= Washing fractions; E= Elution fractions. **A** SDS-PAGE of the purification of GST-tagged vHo1 (MW= ~46 kDa) using glutathione agarose for affinity purification. **B** Corresponding western blot analysis of the SDS-PAGE shown in (**A**) to detect GST-tag. Respective antibodies used were goat anti-GST antibody (1:10 000 dilution) for detection of GST-tag, while for detection and visualization of goat IgG the antibody rabbit anti-goat IgG alkaline phosphatase (1:30 000 dilution) was used.

The heme oxygenase assay can detect the enzymatic ability of an enzyme to form BV IX from heme by inhibiting nonspecific cleavage of heme. Hereby, the absorbance changes over time are followed spectroscopically, and the enzyme is provided with electrons to reduce heme to BV IX. BphO from *P. aeruginosa* was used as a reference to determine the activity of the vHo1 enzyme (Figure 18 A).

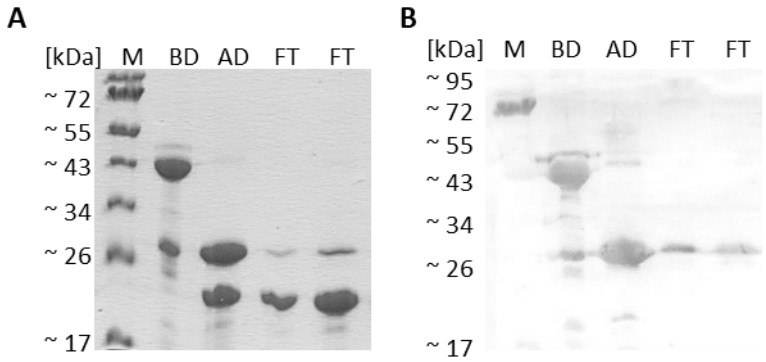


Figure 17: Second affinity purification of viral heme oxygenase of freshwater phage CB_2 after cleavage of N-terminal GST-tag. The fusion protein vHo1 was incubated over night with PreScission Protease with its cleavage site located between GST-tag and vHo1, following further affinity purification using glutathione agarose which also binds PreScission Protease which has an uncleavable GST-tag. M= Size marker - Color Prestained Protein Standard, Broad Range (NEB); BD= Before digest; AD= After digest; FT= Flow-through. **A** SDS-PAGE of the purification of cleaved vHo1 (MW= ~20 kDa) from the GST-tag (MW= ~26 kDa) using glutathione agarose for affinity purification. **B** Corresponding western blot analysis of the SDS-PAGE shown in (**A**) to detect residual GST-tag. Respective antibodies used were goat anti-GST antibody (1:10 000 dilution) for detection of GST-tag, while for detection and visualization of goat IgG the antibody rabbit anti-goat IgG alkaline phosphatase (1:30 000 dilution) was used.

Both enzymes were similarly purified, and the reaction showed characteristic similarities. At the start of the reaction, a dominant peak at 400 nm and a smaller signal at 500 nm is visible, of which the former can indicate the Soret band of the heme oxygenase since heme is not only the substrate of the enzyme but also a cofactor (Figure 18 A,B). These peaks dimmish over time, and at 350 and 640 nm, new peaks form, indicating heme degradation and BV IX formation. When loading the reaction product after stopping the assay onto the SepPak filter for further purification, these filter changed their color from white to green (Figure 18, upper pictures), indicating the presence of BV IX. HPLC analysis of the reaction products from the assays was carried out to verify the predicted Ho1 activity in producing BV IXα. The specific retention time can give insights into the BV IX isomer formed. As a reference for the reaction product of vHo1, purified BV IXα and the reaction product of BphO were also applied. At the retention time of 30 min, all three samples showed maximum absorbance at 650 nm (Figure 18 C) (Wegele *et al*., 2004). Only purified BV IXα had another peak at 21 min, most likely due to stock contamination. Therefore, the vHo1 can be characterized as heme degrading and BV IXα producing active enzyme.

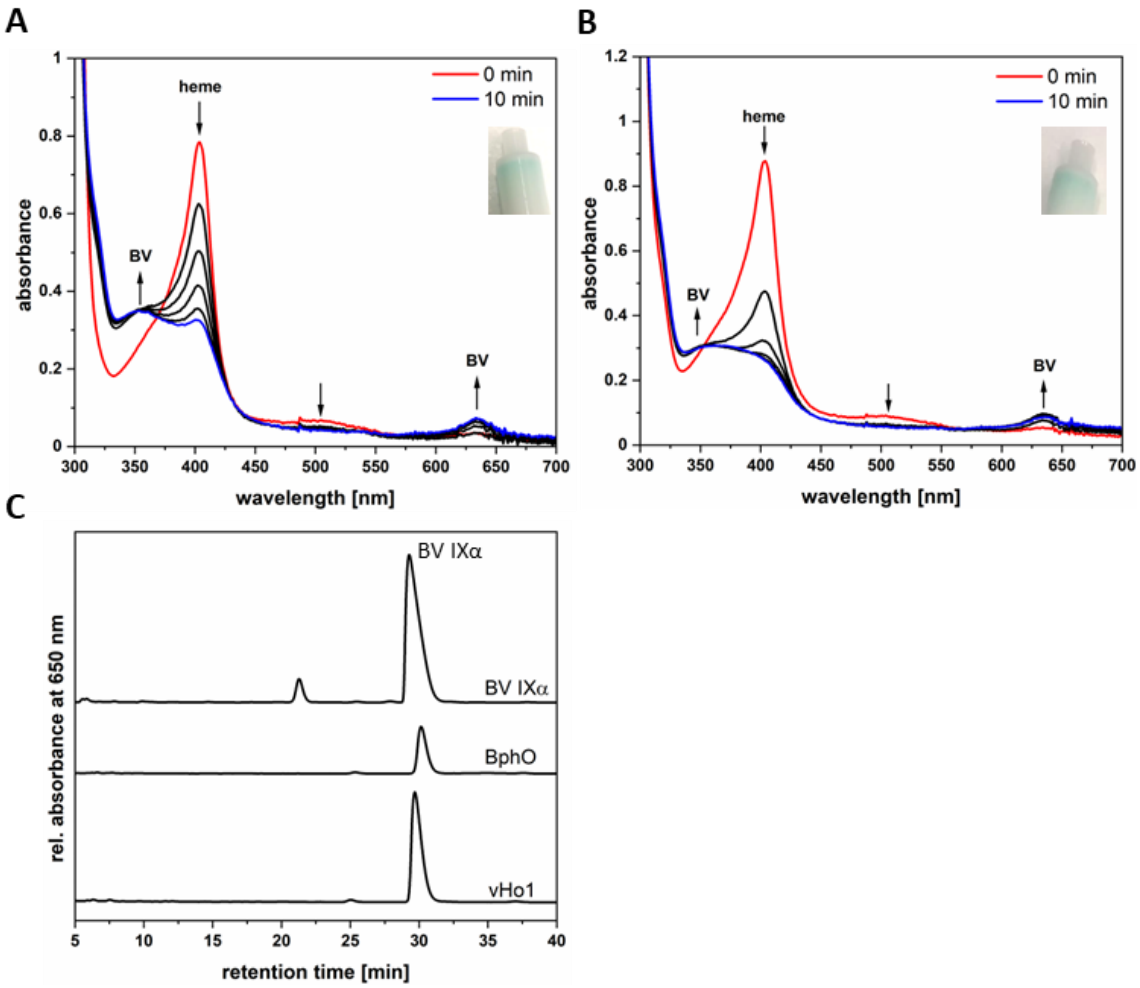


Figure 18: Heme oxygenase assay of BphO and vHo1 enzyme activity and HPLC analysis of the resulting products. **A + B** Time-resolved UV-Vis spectroscopy of an activity test employing HO and heme as the substrate. The total reaction time was 10 min. A single spectrum was recorded every 30 s and displayed every other minute. The arrows indicate the course of the absorbance during the reaction. Heme marks the initial spectrum after incubation with HO. Biliverdin IX (BV IX) marks absorbance maximum related to the formation of BV IXα. HO enzymes are incubated with heme before a coupled system producing electrons that are transferred via ferredoxins are added to start the reaction Arrows indicating product formation or radical formation ↑: absorption increase, ↓: absorption decrease. **A** Shows the enzymatic activity of BphO from *P. aeruginosa* which was affinity purified as vHo1 and used in the assay without GST-tag. BV IX formation can be observed as increase in absorbance at 640 nm and heme degrades at 400 nm over time. **B** vHo1 activity after purified using affinity chromatography with minor GST-tag contaminations. Measurements were conducted as for BphO. Formation of BV IX was measurable at 640 nm. **C** HPLC analysis of the reaction products from HOs, monitored at 650 nm. The analysis was carried out using Luna 5 µm reversed phase C18 column as stationary phase. The mobile phase consisted of 50 % (v/v) 20 mM formic acid and 50 % (v/v) acetone. Purified BV IXα as standard was also applied.

3.2 Characterization and determination of the viral ferredoxin-dependent bilin reductase

Similar to the characterization of vHo1, the viral FDBR next to it on the contig was characterized. Predicting the function of an FDBR is challenging because they share only ~10 % sequence similarity or less, making biochemical characterization essential (Ledermann *et al*., 2016). Despite this, FDBR enzymes share a similar 3D structure, and their key catalytic amino acids can help predict the reaction mechanism because they are specific for the respective FDBR and, when exchanged, can result in another product (Tu *et al*., 2007). Identifying the phycobilins formed during phage infection is crucial for understanding their role, as different phycobilins have varied functions in the cell, and not all are known to occur in phages. While the FDBRs PcyX and PebS form PEB, PcyA forms PCB. So far, the enzymes PcyX and PebS are known to be present only in phages, while PcyA is also present in cyanobacteria. However, the enzymes PebA and PebB can also produce PEB in cyanobacteria, which catalyze the same reaction as the phage PebS.

All FDBRs share a specific general fold, an α-/β-/α-sandwich fold, which can help characterize the protein family. An AlphaFold generated 3D-model of the viral FDBR confirmed the same fold. When overlayed with the characterized PcyA from *Synechocystis* sp. PCC6803 (Figure 19 B), the overlayed model shows distinct similarity regarding the general fold and the binding of BV IXα (green) (Figure 19 B, enhanced picture). A sequence alignment with the putative FDBR (named vPcyA/X) was analyzed to confirm the presence of the respective crucial catalytic residues. In general, the amino acid D105 (*Synechocystis* numbering) is responsible for the first deprotonation to start the reaction, which is present in all FDBRs analyzed (Figure 19 A) (Busch *et al*., 2011a, Busch *et al.,* 2011b, Tu *et al*., 2007). In order to distinguish between PebS and PcyA/X reaction, the amino acids E76, H88, D105, D206, and N219 are crucial. While PebS has a second Asp (D206) that is responsible for the second reduction from 15,16 DHBV to PEB, it is only catalytically important for PebS and PebB where an identical reduction is catalyzed (Busch *et al*., 2011a, Busch *et al*., 2011b). PcyA/X lack these residues but often carry an asparagine at the same position (data not shown). The present H88 in vPcyA/X is known to be only catalytically important for the PcyA/X reaction, where a proton channel is formed, and reprotonation of D105 occurs (Tu *et al*., 2007). Here, PebS has

V76, N88, and L219, whereas PcyA/X are quite conserved. Only the amino acid E76 is known to differ between PcyA and PcyX, whereas the latter has a D76 (Tu *et al*., 2007). Therefore, the vPcyA/X is most likely a vPcyA.

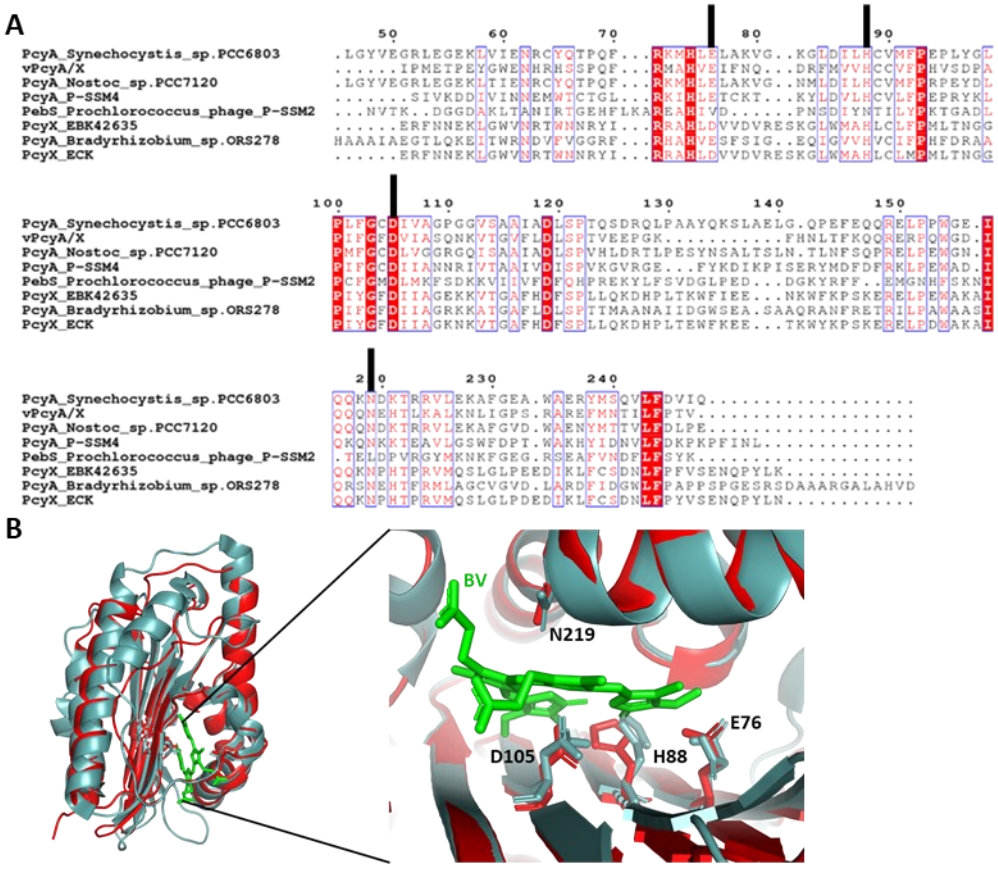


Figure 19: Sequence alignment and structural analysis of viral PcyA/X. A Detailed view of reference sequences from the NCBI database which were aligned with sequence of viral PcyA/X using Clustal Omega and ESPript 3. Highly conserved amino acids are highlighted in a red square with white lettering, similar amino acids are represented by a white box with red lettering, and non-similar amino acids are indicated by black lettering. Catalytic important amino acids, previously characterized in *Synechocystis* sp. PCC6803, are denoted by black lines at the top. Sequences can be found in the Appendix I + II. **B** Overlay model of monomeric PcyA from *Synechocystis* sp. PCC6803 (blue, RCSB PDB code: 2D1E) (Hagiwara *et al*., 2006) and the model of vPcyA predicted via AlphaFold (red), shows high structural similarity. The displayed overlay model was generated with PyMol, the magnified view shows binding of biliverdin IXα (BV IXα) with the important amino acids E76, H88, D105, and N219 (sticks in blue/red) (Frankenberg & Lagarias 2003, Hagiwara *et al*., 2005, Tu *et al*., 2007) and BV IXα in green as chemical structure.

In metagenomic data, heme oxygenases were always accompanied by PcyXs. Therefore, a phylogenetic tree of all found FDBRs in the metagenomic datasets and reference proteins should help determine their similarity (Figure 20). While all cyanobacterial enzymes required for PEB production (PebA, PebB) cluster together, the cyanophage PebS sequences are near to it. Also, cyanobacterial PcyA and cyanophage PcyAs are clustering near each other. Interestingly, the characterized PcyA from *Bradyrhizobium* sp. ORS278 (Jaubert *et al.,* 2007), a stem nodule-forming bacterium, is predicted to be on a genetic island and cluster with some freshwater FDBRs.

The metagenomic PcyXs are the most distant known FDBRs to the cyanobacterial FDBRs. They cluster together with most of the freshwater and marine metagenomic FDBR sequences. Sequence alignment of the metagenomic data indicated that only the vPcyA (in the alignment named Crystal Bog phage scaffold 2 70) and the two proteins below the *Bradyrhizobium* PcyA are PcyAs, whereas all others are PcyXs.

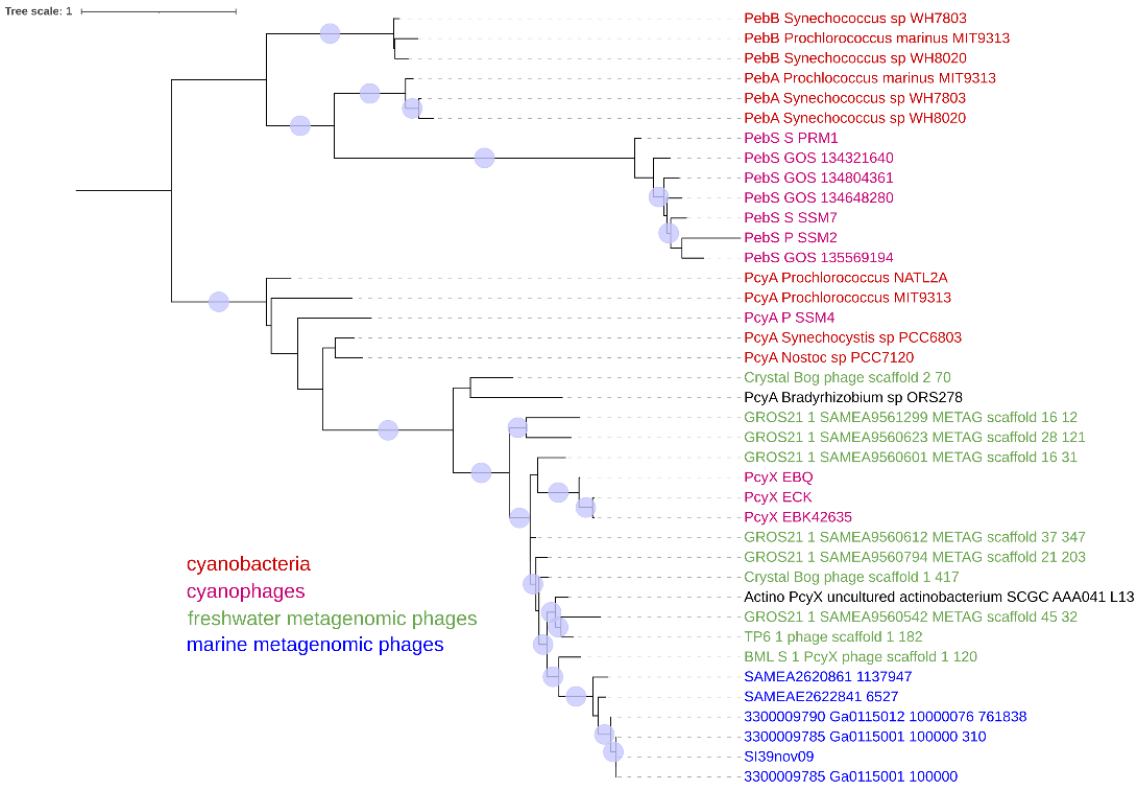


Figure 20: Phylogenetic tree of FDBRs from viral and cyanobacterial origin. The alignment of the sequences was done using Mafft and depicted in a phylogenetic tree using the Next Generation Phylogeny.fr web. The output tree was then further adapted by Midpoint rooting and labelling using iTOL (interactive Tree of Life, version 6.8) via the webserver. Further, bootstrap was assigned and when values were between 0.75 to 1 marked with a blue circle on the tree branch. The respective organisms could not be written in italic due to the restrictions of the program, the correct labelling is in the Appendix I + II. In general, there is a clustering recognizable that is traceable to the different reactions catalyzed by these enzymes. Sequences were obtained via NCBI and from sources stated before (see chapter 2.1.4) and colored respectively, they can be found in the Appendix I + II.

This novelty of a viral PcyA from metagenomic data should also be validated biochemically. Thus, the viral *pcyA* gene was codon adapted for *E. coli* codon usage and expressed heterologously in *E. coli.* The protein was N-terminal GST-tagged (MW= ~50 kDa) and purified via affinity purification (Figure 21). In order to determine the enzyme activity, an FDBR assay was carried out after removing the GST-tag using glutathione agarose affinity chromatography.

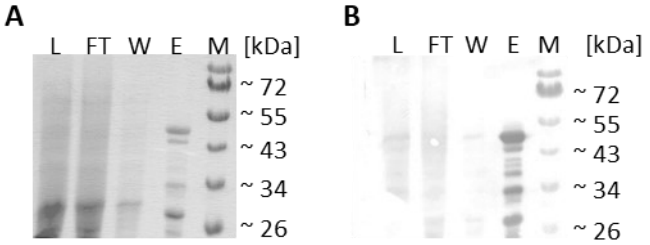


Figure 21: Purification of heterologous produced viral PcyA of freshwater phage CB_2 via affinity chromatography. The N-terminal GST-tagged vPcyA from freshwater phage CB_2 was produced in *E. coli* Rosetta pLysS in LB medium, the gene sequence of v*pcyA* was codon adapted for *E. coli* codon usage. M= Size marker - Color Prestained Protein Standard, Broad Range (NEB)*;* L= Lysate; FT= Flow-through; W= Washing fraction; E= Elution fraction. **A** SDS-PAGE of the purification of GST-tagged vPcyA (MW= ~50 kDa) using glutathione agarose for affinity purification. **B** Corresponding western blot analysis of the SDS-PAGE shown in (**A**) to detect GST-tag. Respective antibodies used were goat anti-GST antibody (1:10 000 dilution) for detection of GST-tag, while for detection and visualization of goat IgG the antibody rabbit anti-goat IgG alkaline phosphatase (1:30 000 dilution) was used.

In the second purification (Figure 22 A,B), a GST signal (~26 kDa) in the flow through fractions can be observed. Unfortunately, vPcyA is predicted with a size of (~25 kDa) and thus cannot be distinguished from the GST-tag in the SDS-PAGE. Nevertheless, the western blot (Figure 22 B) shows no strong signals for the GST-tag in the flow-through, thus, the PcyA is most likely present and contains only minor contaminations of the GST-tag. The protein was further concentrated and dialyzed in the respective buffer for the FDBR assay to verify the enzyme activity.

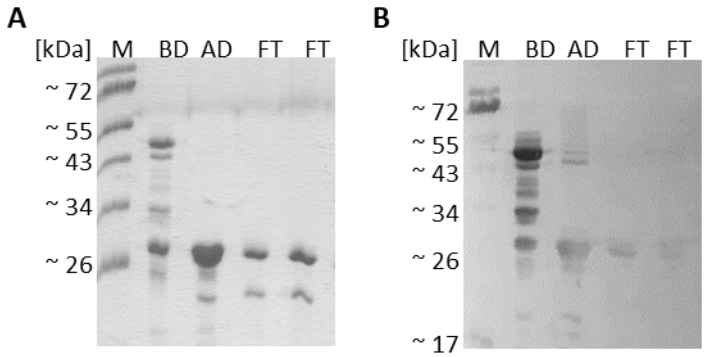


Figure 22: Second affinity purification of vPcyA of freshwater phage CB_2 after cleavage of N-terminal GST-tag. The fusion protein vPcyA was incubated over night with PreScission Protease with its cleavage site located between GST-tag and vPcyA. Further, affinity purification using glutathione agarose was carried out. This also removes the PreScission Protease from the FT as it has a non-cleavable GST-tag. M= Size marker - Color Prestained Protein Standard, Broad Range (NEB); BD= Before digest; AD= After digest; FT= Flow-through. **A** SDS-PAGE of the purification of cleaved vPcyA (MW= ~24 kDa) from the GST-tag (MW= ~26 kDa) using glutathione agarose for affinity purification. **B** Corresponding western blot analysis of the SDS-PAGE shown in (**A**) to detect residual GST-tag. Respective antibodies used were goat anti-GST antibody (1:10 000 dilution) for detection of GST-tag, while for detection and visualization of goat IgG the antibody rabbit anti-goat IgG alkaline phosphatase (1:30 000 dilution) was used.

The FDBR assay is carried out under anaerobic conditions and with BV IXα as substrate. A specific product formation regarding the enzyme activity can be observed spectroscopically over time. As a reference, a characterized PcyA from cyanobacterium *Anabaena* sp. PCC 7120 (*Anabaena*) was purified like vPcyA and analyzed with the FDBR assay (Figure 23 A). Here, the formation of a new signal at ~600 nm PCB can be observed while BV IXα is degraded over time (~680 nm). The formation of substrate radicals can also be observed at 420 nm, 470 nm, and 770 nm, which is characteristic for FDBR enzyme activity (Busch *et al*., 2011a, Busch *et al*., 2011b, Ledermann *et al*., 2016, Tu *et al*., 2004). Interestingly, the vPcyA enzyme does not show the same spectral characteristics as the cyanobacterial PcyA (Figure 23 A, B). While vPcyA also shows degradation of BV IXα at 680 nm, the formation of PCB at 600 nm is not prominent (Figure 23 B). However, similar to PcyA, the formation of radicals can be observed for vPcyA at 420 nm, 470 nm, and 770 nm. Here, the radical formation seems more dominant compared to vPcyA, which may indicate a slower or incomplete reaction.

In order to verify the activity of the vPcyA enzyme, the generated products in the FDBR assay were analyzed via HPLC analysis (Figure 23 C). As a reference for the specific retention times, the products generated through the PcyA reaction of *Anabaena* in the FDBR assay and purified PCB were investigated and compared with the vPcyA products. 3(*Z*)-PCB shows a retention time of 25.5 min, whereas its isomer 3(*E*)-PCB has a retention time of 14.9 min (Figure 23 C). For vPcyA, the significant retention times of its product were similar to 3(*Z*)-PCB and BV IXα, but a small peak was also detectable at 15 min, indicating the presence of 3(*E*)-PCB. The quite prominent peak of BV IXα at 30 min for vPcyA could indicate that the reaction was not completed.

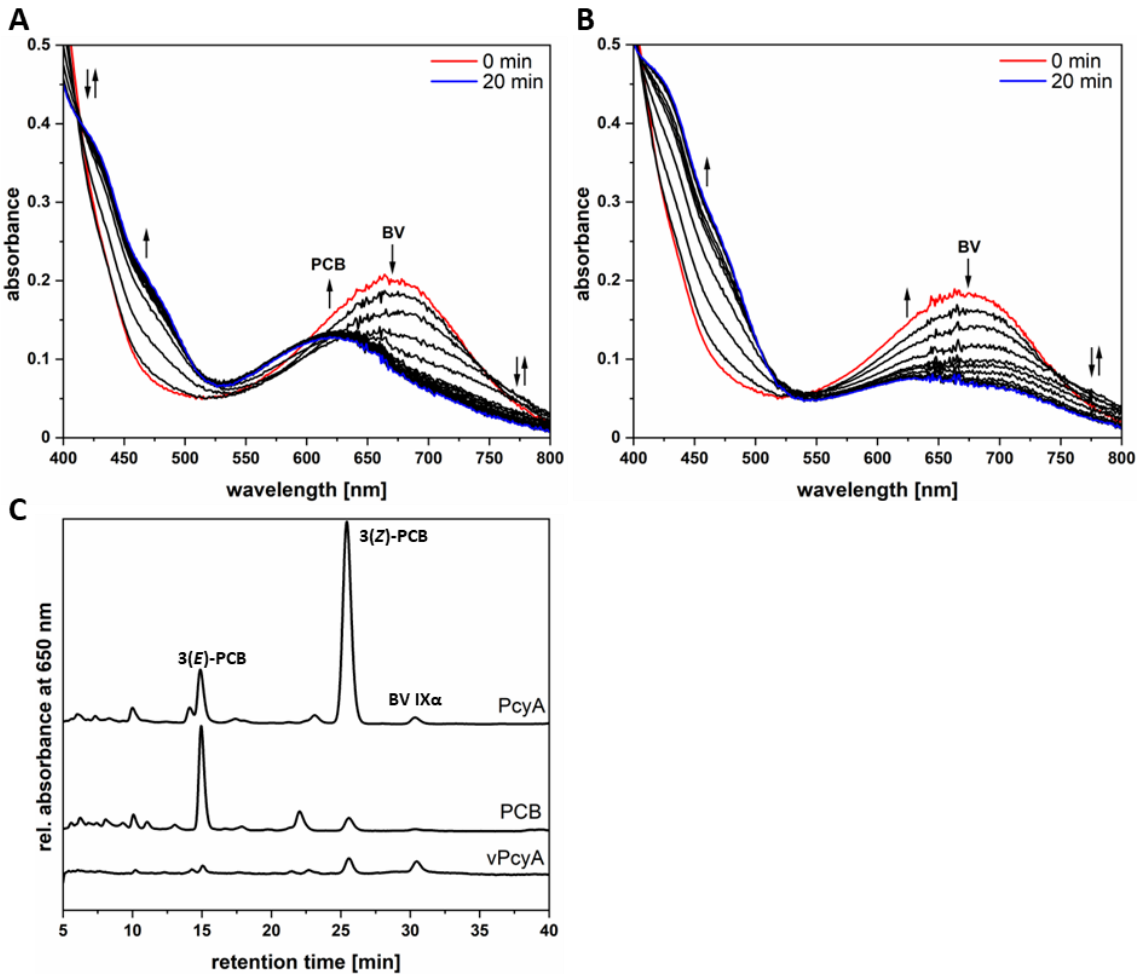


Figure 23: FDBR assay of PcyA and vPcyA enzyme activity and HPLC analysis of the resulting products. A + B Time-resolved UV-Vis spectroscopy of an anaerobic activity test employing PcyA and biliverdin IXα (BV IXα) as the substrate. The total reaction time was 20 min. A single spectrum was recorded every 30 s and displayed every minute. The arrows indicating the course of the absorbance during the reaction. BV IXα marks the initial spectrum after incubation of it with PcyA. PCB mark absorbance maximum related to the formation of PCB. PcyA enzymes are incubated with BV IXα anaerobically for 15 min before a coupled system producing electrons that are transferred via ferredoxins are added to start the reaction. Arrows indicating product formation or radical formation ↑: absorption increase, ↓: absorption decrease, ↓↑: radical formation. **A** Shows the enzymatic activity of PcyA from *Anabaena* sp. PCC 7120 which was affinity purified as vPcyA and used in the assay without GST-tag. PCB formation can be observed as increase in absorbance at 620 nm whereas radical formation at 420 nm, 470 nm and 775 nm and BV IXα degradation at 680 nm over time. **B** vPcyA affinity purified with minor GST-tag contaminations. Measurements were conducted as for PcyA from *Anabaena* sp. PCC 7120. No clear formation of PCB was measurable at 620 nm, but radical formation was detectable at 420 nm, 470 nm and 775 nm. **C** HPLC analysis of the reaction products from PcyAs, monitored at 650 nm. The analysis was carried out using Luna 5 µm reversed phase C18 column as stationary phase. The mobile phase consisted of 50 % (v/v) 20 mM formic acid and 50 % (v/v) acetone. Purified PCB as standard was also applied.

In order to confirm a functional enzyme, another approach was tested, which does not require the purification of a protein but a coexpression system. Therefore, an acceptor for the produced linear tetrapyrroles, namely the phytochrome Cph1 from *Synechocystis* sp. PCC 6803 is provided. Cph1 can bind PCB, PΦB, and PEB but not BV IXα and exhibits unique spectral properties upon binding. When PCB or PΦB is bound to Cph1, it can form a phytochrome that can photoconvert when illuminated with far-red

or red light and display a distinct absorbance. When PEB is bound to Cph1, it forms a phytofluor with strong fluorescence properties (Li *et al*., 1995). Since *E. coli* cannot use the produced BV IXα, it can be entirely reduced by the PcyA to PCB and bind to Cph1, making the tetrapyrrole more stable and forming a phytochrome. Thus, after the cells are lysed, the produced phycobilin can be determined based on its spectral properties.

The coexpression of a heme oxygenase and v*pcyA* yielded a turquoise-colored pellet of *E. coli* cells (Figure 24 A), which are typically pale brownish, indicating the enzymatic activity of the viral PcyA. The cells were lysed, and a difference spectrum could be determined after illumination with far-red (fr) and red (r) light (Figure 24 A). The difference spectra showed a maximum at 650 nm and a minimum at 706 nm, these specific values fit the literature values for Cph1 binding PCB (Gambetta & Lagarias 2001). Therefore, the viral PcyA is a real PcyA and produces PCB. Furthermore, to test the functionality of the enzymes of the gene cassette from this freshwater phage, the viral heme oxygenase of this cassette was also used to verify its coupled activity with the now-characterized PcyA. Therefore, the duet vector containing the genes of v*pcyA* and *ho1* was used, and the heme oxygenase gene was replaced with the viral v*ho1* gene from the cassette. After cell lysis and affinity chromatography using TALON column material to enrich the His_6-tagged Cph1 and the viral PcyA, difference spectra could be obtained after illumination with far-red and red light. Here, the maximum and minimum values of the difference spectrum are similar to the previous phytochrome assay and fit the literature-known values for Cph1 binding PCB (Gambetta & Lagarias 2001), thus indicating a functional contig of the phage CB_2.

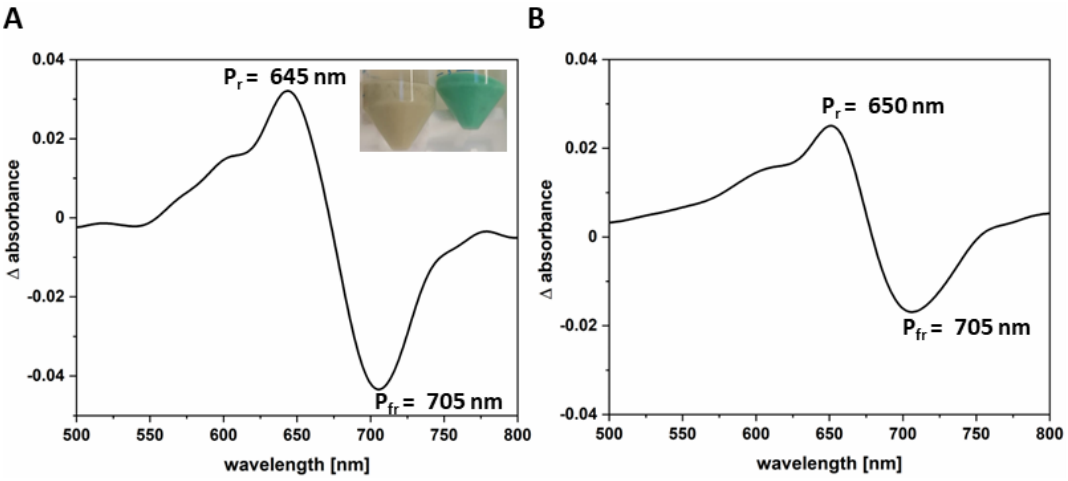


Figure 24: Coexpression and difference spectra of phytochrome *cph1* together with viral metagenomic heme oxygenase and FDBR gene. A Upper inset shows reference *E. coli* cell pellet (left) and pellets of *E. coli* cells coexpressing genes encoding for Cph1, Ho1 (P-SSM2) and vPcyA (CB_2) yielding a turquoise pellet (right). Of the latter a phytochrome difference spectrum was obtained following coexpression of genes encoding for Cph1, Ho1 and vPcyA (pTD*ho1*_v*pcyA* and pET his-tag cph1 crystal). *In vivo* chromophore assembly was detected after cell lysis through illumination of cell lysate with red light (636 nm – P_{fr} spectrum) and far-red light (730 nm – P_r spectrum). Difference spectra were calculated by the subtraction of the P_{fr} from the P_r spectrum. **B** Difference spectrum of phytochrome yielded after coexpression of genes encoding for Cph1, vHo1(CB_2) and vPcyA(CB_2) from Duet vector (pTDv*ho1*_v*pcyA*) and pET his-tag cph1 crystal. *In vivo* chromophore assembly was detected after cell lysis and affinity chromatography enrichment through TALON column material. Illumination of enriched protein fraction was carried out with red light (636 nm – P_{fr} spectrum) and far-red light (730 nm – P_r spectrum). Difference spectra were calculated by the subtraction of the P_{fr} from the P_r spectrum.

3.3 Characterization and determination of the viral aminolevulinic acid synthase

The novelty of the clusters found in viral metagenomic data relies on the presence of the gene *alaS*. The respective enzyme ALAS belongs to the C4 (Shemin) pathway and forms the first common tetrapyrrole precursor ALA by utilizing the substrates glycine, succinyl-CoA, and the cofactor PLP. This enzyme is only present in alphaproteobacteria, yeast, and higher non-plant eukaryotes.

In order to verify the activity of the found viral ALAS sequences in multiple metagenomic datasets of different origins, the sequences were first analyzed *in silico* following protein characterization *in vitro*. The viral ALAS sequences exhibited the highest sequence identity (~55 %) to ALAS sequences from the *Pelagibacteraceae*, a family in the class of alphaproteobacteria. In order to characterize the activity of the viral ALAS, a protein sequence alignment was carried out to determine the presence of catalytic import amino acids (Figure 25 A). Alphaproteobacterial ALAS sequences were used as a reference, especially the ALAS from *R. capsulatus* (RcA), whose crystal structure has been solved (numbering according to RcA). Some viral ALAS sequences showed changes in catalytic residues, but most were still highly similar to alphaproteobacterial ALASs. A crucial residue is K248, essential for forming Schiff's base linkage to the cofactor PLP. The presence of this residue was confirmed in all alphaproteobacterial and viral sequences. For the substrate binding of succinyl-CoA and glycine to ALAS and the PLP coordination, multiple residues are essential that are known from human ALAS1 and ALAS2, RcA and yeast ALAS (Astner *et al.*, 2005, Brown *et al.,* 2018). The variation of catalytic amino acids can provide insight into a possible relationship between the viruses, which is why some amino acids were analyzed in detail. Some viral ALAS showed changes in residues critical for PLP coordination, namely A115 and S277. For stabilizing succinyl-CoA phosphoadenosine, the amino acids S139, I/V149, K156, and I158 are essential, while S139 is variable in eukaryotes and other alphaproteobacteria. In the viral sequences, the occurrence of S139 is also variable, while I/V149 was either L in marine or M/I/L in freshwater viruses. K156 was conserved in alphaproteobacteria and marine viruses but variable in freshwater viruses. Highly conserved in all sequences was I158, and only in freshwater viruses also V occurred. The delineation of the succinyl-CoA

binding pocket is also facilitated by A143, which could be N in alphaproteobacteria or freshwater viruses and S in marine viruses.

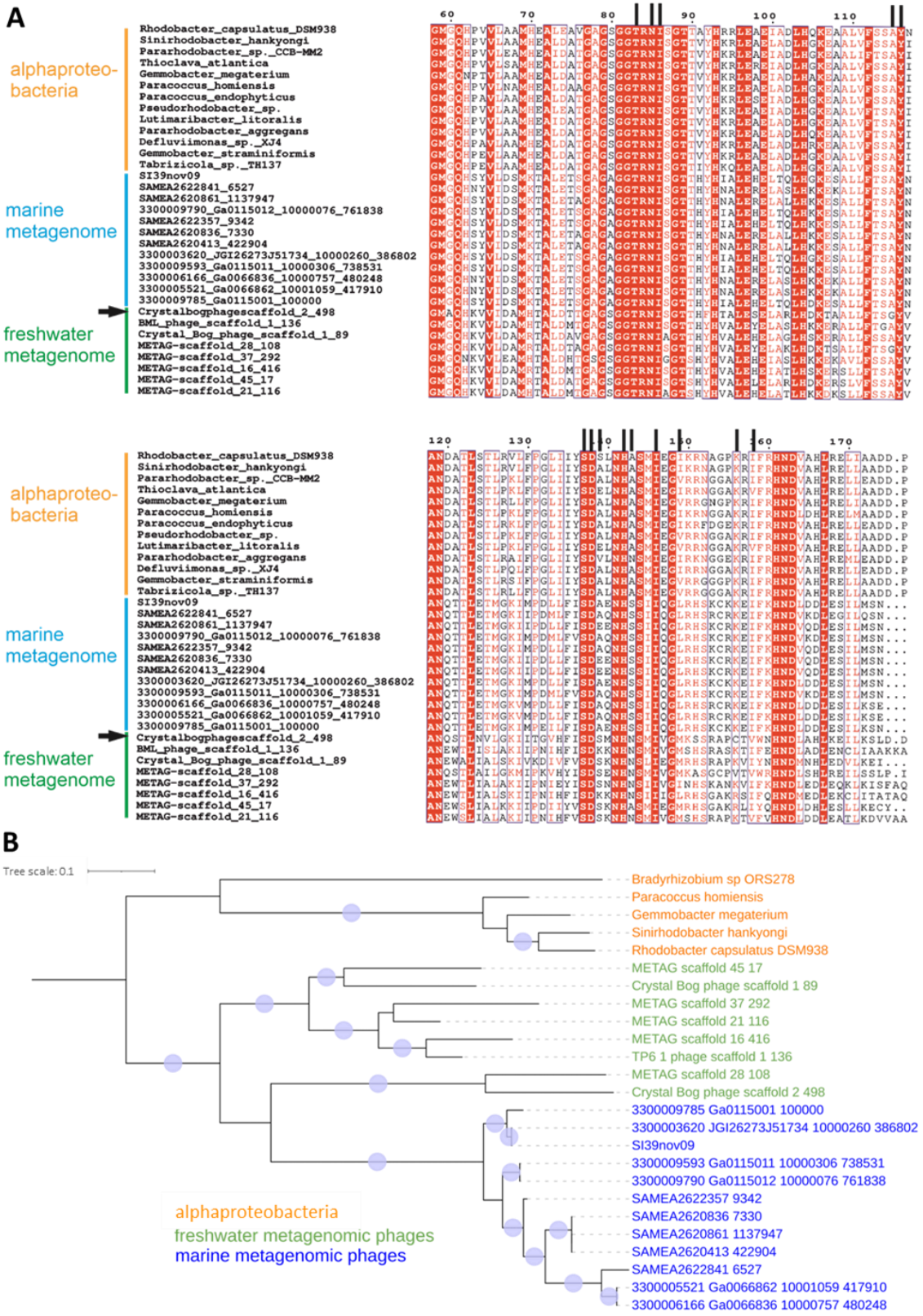


Figure 25: Amino acid sequence analysis and phylogenetic analysis of viral ALAS sequences. A Amino acid alignment of ALAS sequences from alphaproteobacteria and viral metagenomes revealed conserved catalytic residues. Reference sequences from the NCBI database (shown in yellow) were aligned with sequences from multiple sampling sites using Clustal Omega and ESPript 3. Highly conserved amino acid residues are highlighted in a red square with white lettering, similar amino acids are represented by a white box with red lettering, and non-similar amino acids are indicated by black lettering. Catalytic important amino acids, previously characterized in (Astner *et al.,* 2005), are denoted by black lines at the top. Analyzed sequences in this chapter is marked with an arrow. Sequences can be found in the Appendix I + II. The respective organisms could not be written in italic due to the restrictions of the program, the correct labelling is in the Appendix I + II. **B** Sequences were obtained via NCBI and from sources stated before (see chapter 2.1.4) and colored respectively. The alignment of the sequences was done using Mafft and depicted in a phylogenetic tree using the Next Generation Phylogeny.fr web. The output tree was then further adapted by Midpoint rooting and labelling using iTOL (interactive Tree of Life, version 6.8) via the webserver. Further, bootstrap was assigned and when values were between 0.75 to 1 indicated with a blue circle on the tree branch. In general, there is a clustering recognizable that is traceable to the different habitats these enzymes/sequences derived from. The respective organisms could not be written in italic due to the restrictions of the program, the correct labelling is in the Appendix I + II.

The sequence alignment demonstrated a viral amino acid distribution that seems specific to certain habitats, suggesting a potential phylogenetic relationship. The predicted phylogenetic tree (Figure 25 B) of alphaproteobacteria and viral ALAS sequences showed a habitat-specific clustering of the ALAS sequences. At the same time, the marine viral sequences are uniformly clustering together and indicate a stronger internal relationship compared to the freshwater viral sequences of ALAS. In general, viral ALAS sequences, regardless of their derived habitat, were quite distant related to the alphaproteobacterial ones.

Further *in silico* analysis was utilized to confirm the activity of the viral encoded ALAS enzymes regarding the structural similarity of ALAS enzymes. Therefore, AI-based structure prediction generated 3D models of viral ALAS proteins. There is no experimentally solved protein structure of a *Pelagibacteracea*-derived ALAS, which had the highest sequence similarity to the viral ALASs, so RcA was used for structural investigation, exhibiting a slightly reduced amino acid sequence similarity (~50 %). The previously analyzed contig of the freshwater phage CB_2 also consists of a further downstream located ALAS sequence, which was selected to determine the structural similarity compared to the alphaproteobacterial one. This viral ALAS candidate, ALAScry (Crystal_Bog_phage_scaffold_2_498), was modeled and overlaid with RcA (Figure 26 A, B, C). The amino acid sequence of ALAScry was obtained from a freshwater environment and curated bioinformatically by (Chen *et al*., 2020). The modeling was carried out using the AI tool AlphaFold to predict its structure and alignment to RcA (RCSB PDB code: 2BWN, 2BWO) using PyMol. The model reveals a highly similar structure of ALAScry, shown in red, and RcA in blue (Figure 26 A).

The structure of ALAScry is predicted to be homodimeric and reveals PLP-binding pockets situated at the interface of the dimer. The active site is located in the middle of each monomer (Figure 26 B) and consists of the glycine (green circle) and succinyl-CoA (purple circle) binding sites, while the cofactor PLP binding site is highlighted with yellow sphere. Thus, a homodimer was constructed similar to the known homodimeric form of RcA, where two active sites are present. Most catalytic amino acid residues mediating the enzymatic reaction of forming ALA were conserved in the ALAScry sequence. As K248 is essential for binding the cofactor PLP by forming Schiff's base linkage when no substrate is present (Figure 26 A, lower picture), the same amino acid in ALAScry was found to be in place to form the linkage.

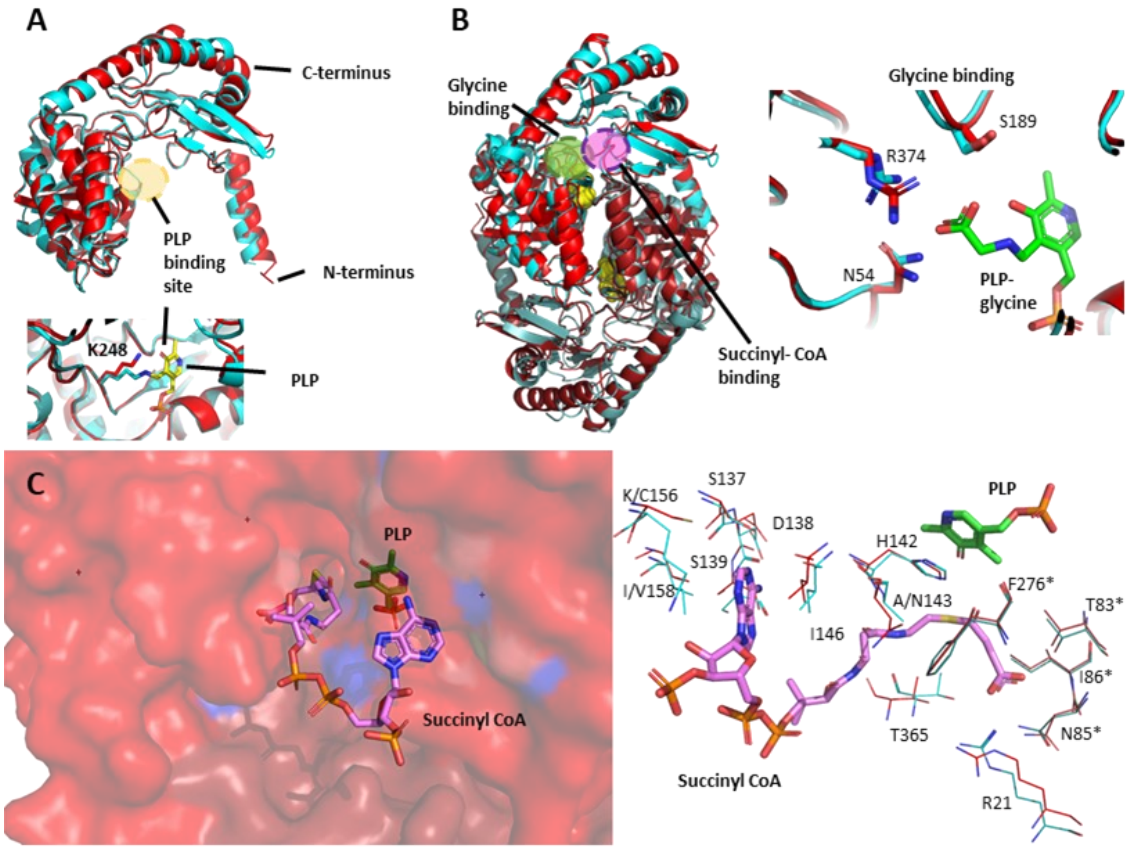


Figure 26: Structural modelling of ALAScry revealed high similarity to bacterial ALAS. A Overlay model of monomeric ALAS (orig. homodimer) from *Rhodobacter capsulatus* (RcA, blue, RCSB PDB: 2BWN) and the model of vALAS predicted via AlphaFold (red), shows high structural similarity. The displayed overlay model was generated with PyMol, the cofactor binding site of PLP is indicated with a yellow circle. A magnified view shows binding of PLP to its binding site with the important amino acid K248 (sticks in blue/red) and PLP in yellow as chemical structure forming a Schiff's base linkage. **B** Homodimeric view of vALAS with RcA overlay, indicated are glycine binding (green circle) and succinyl- CoA binding (purple circle) and an enhanced view of the three important amino acid residues of the glycine binding (N54, S189, R374) which is present as PLP-glycine intermediate (green chemical structure). **C** The substrate channel for succinyl-CoA binding is visualized (left picture), with the ribose moiety outside the enzyme (purple) and PLP (green) inside. The model (right picture) highlights all amino acids known to be important for succinyl-CoA binding/localization, with red sticks representing RcA amino acids and blue sticks representing vALAS amino acids. Some deviations in vALAS are observed at positions 143, 156, and 158, with the changed amino acids named following RcA. Asterisks indicate amino acids from the second monomer.

Furthermore, the ALAScry model was compared with the RcA crystal structures with the bound substrates to detect steric hindrances of amino acid residues crucial for binding substrate or catalysis (Figure 26 B,C). The essential amino acids for glycine binding N54, S189, and R374 were present in ALAScry and in the same position as the reference RcA. More amino acids are involved for the succinyl-CoA binding, and a substrate tunnel is formed, which can also be observed in ALAScry (Figure 26 C). Three amino acid residues that are part of succinyl-CoA binding differ from RcA in ALAScry, namely A/N143, K/C156, and I/V158, the first letter indicating the RcA amino acid residue. The change of A143 is also present in other alphaproteobacterial ALAS, and the exchange of I158 to V does not affect hydrophobicity at this position and should not interfere with the Van der Waals interaction. The exchange of lysine at position 156 to cysteine could lead to the steric loss of the hydrogen bond to the ribose. This could reduce the affinity for succinyl-CoA, which was only observed in some freshwater viral ALAS. ALAScry generally has the following amino acid changes compared to RcA A/G115, A/N139, and I/M149. A115 is critical in PLP coordination (Shoolingin-Jordan *et al.,* 2003), whereas all other named amino acid residues are essential for succinyl-CoA binding (Astner *et al*., 2005). Most ALAScry changes are likely negligible since they result in either another similar charged amino acid residue or occur in other bacterial, human, or yeast ALASs (Astner *et al*., 2005, Brown *et al*., 2018, Ikushiro *et al*., 2018) (Figure 25 A).

3.3.1 Freshwater viral ALAS

In order to characterize the enzymatic activity of viral ALAS, heterologous protein production was carried out in *E. coli*. Unfortunately, cloning the original and codon-adapted versions of the freshwater viral *alaS* gene (from crystal bog phage 2) into the pGEX-6P-3 vector, designed to produce a soluble N-terminal GST-tagged protein, resulted in the formation of inclusion bodies during protein production (data not shown). Also, further approaches of increasing solubility by changing conditions such as inducer concentration, temperature, etc. (see chapter 2.5.1) yielded no increase in protein solubility for ALAScry and the codon-adapted version ALAScryopt (data not shown). Therefore, the genes were cloned into the vector pColdTF, creating a fusion protein of ALAScry or ALAScryopt with a protein facilitating correct folding. This protein is an N-terminal trigger factor (TF), a prokaryotic ribosome-associated chaperone protein with an His_6-tag for affinity purification. The gene expression of the fusion protein is now cold-inducible, mediated through the *cspA* promotor, and regulated by the *lac* operator. Following the production, the purification via affinity chromatography was carried out using cobalt for the column matrix (TALON).

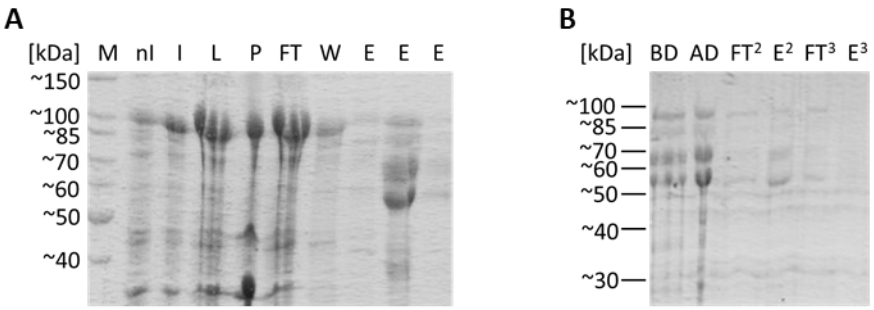


Figure 27: Affinity purification of heterologously produced viral ALAScry of freshwater phage CB_2 using affinity chromatography. The N-terminal fusion protein consists of trigger factor- His_6-tag-ALAScry and was produced in *E. coli* BL21(DE3) in LB medium, the gene sequence encoding ALAScry was codon adapted for *E. coli* codon usage. M= Size marker - Unstained Protein Standard, Broad Range (NEB); nI= Non-induced; I= Induced; L= Lysate; P= Pelleted cell debris; FT= Flow-through; W= Washing fraction; E= Elution fractions. **A** SDS-PAGE of the purification of fusion protein TF-ALAScry with His_6-tag (MW= ~97 kDa) using Cobalt column material for affinity purification. **B** Further purification analysis on SDS-PAGE of ALAScry (MW= ~46 kDa) after cleavage with PreScission protease thus separating ALAScry from Trigger factor with His_6-tag (MW= ~48 kDa) using cobalt affinity purification (FT^2, E^2). Further removal of PreScission protease from the FT^2, which should contain the cleaved ALAScry, was carried out using another affinity purification consisting of glutathione agarose (FT^3, E^3). FT^3 should then contain the purified enzyme ALAScry.

Protein production was observed at ~100 kDa (sample: induced I, lysate L), matching the predicted size of 97 kDa for the ALAScry-TF fusion protein (Figure 27 A). Elution fractions also showed additional lower-molecular-weight signals. For further purification, the cleaved ALAScry was separated with two further chromatography steps from other proteins, leaving ALAScry in the final flow-through (Figure 27 B, FT3).

As visible in the SDS gel (Figure 27 B FT^3), there are minor amounts of protein present in the third flow through at around ~100 kDa and ~50 kDa, which would indicate the presence of the full-length protein with TF (MW= 97 kDa) and cleaved TF (MW= 48 kDa) whereas for the ALAScry (MW= 46 kDa) no protein was detected. A similar result was obtained for the codon-adapted version ALAScryopt, indicating no solubility optimization using TF-fusion protein (data not shown). For enzyme characterization, elution fraction FT^3 was used in coupled enzyme activity measurement (see chapter 2.5.11). The reaction ALAS catalyzes (synthesizing ALA out of glycine and succinyl-CoA) is coupled to a second enzyme reaction, where the resulting CoA-SH from the ALAS reaction is used. The second enzyme reaction uses α-ketoglutarate dehydrogenase (α-KGD) to catalyze NAD reduction to NADH in the presence of its substrate α-ketoglutarate and CoA-SH.

For the establishment of the enzyme test, RcA, which was purified via glutathione agarose as the protein is GST-tagged, was used. The respective NADH production over time is displayed in the Appendix (Figure S 1). The partially purified ALAScry and ALAScryopt were analyzed for activity in the assay, and the amount of NADH over time was measured (Figure 28 A, B). The measured activity was lower than RcA, and because of the contamination in the sample, no significant presence of enzyme activity could be determined (Figure S 1). Thus, further purification methods were applied to achieve purer enzyme solutions and higher enzyme activity.

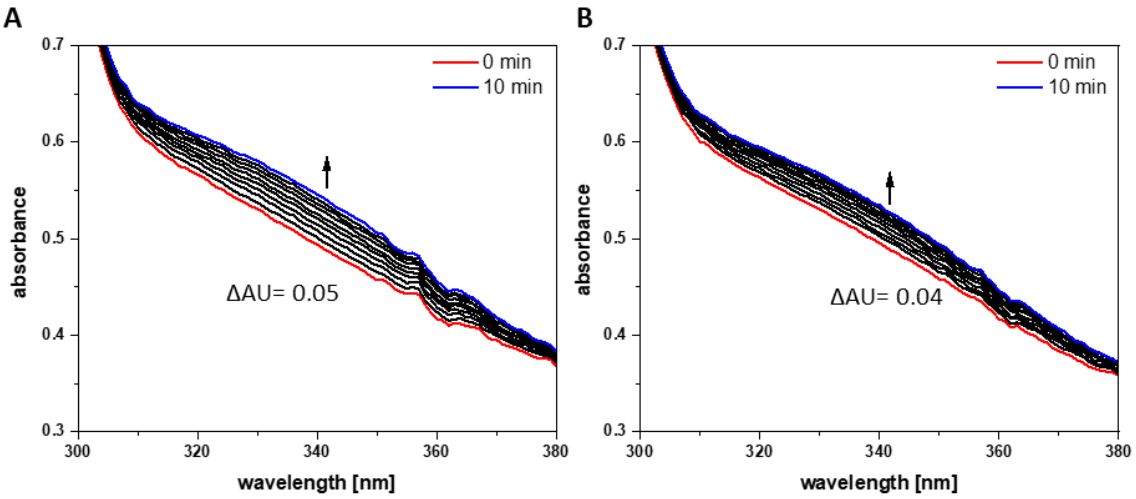


Figure 28: Coupled enzyme activity assay to measure spectroscopically the product formation of viral ALAS enzyme. A+B The coupled enzyme reaction consists of ALAS forming ALA and CoA-SH out of the substrates glycine and succinyl-CoA. The formation of CoA-SH can be used by the second enzyme reaction, which reduces NAD to NADH and replenish the first reaction with succinyl-CoA. The formation of NADH is stoichiometric with the formation of ALA and can be measured over time by spectroscopic measurements at 340 nm. The total reaction time was 10 min and the increase in absorbance over this time is displayed in absorbance units (AU). Single spectrum was recorded every 30 s and displayed every minute. The arrows indicating the course of the absorbance during the reaction. **A** Shows NADH formation of the partial purified and concentrated ALAScry enzyme. **B** Activity assay of the partial purified and concentrated ALAScryopt enzyme by NADH formation.

Further purification optimization was attempted using nickel column material, higher salt concentration, and imidazole in all buffers. The resulting SDS gel (Figure S 2) indicates some contaminating proteins. In order to further purify the ALAS enzyme, anion exchange chromatography (AIEX) followed by size exclusion chromatography (SEC) was carried out (Figure S 2, and data not shown). Unfortunately, both methods indicate that ALAScry and ALAScryopt form aggregates with the TF. Thus, another approach was tested in order to verify the enzyme activity.

Functional complementation proves the activity of vALAS by restoring growth deficit

The verification of enzymatic activity using purified enzymes was challenging to obtain. Therefore, functional complementation of an ALA auxotroph *E. coli* strain ST18 was chosen to prove the functionality of the viral enzyme. This strain has a deleted *hemA* gene that encodes for a GluTR (C5 pathway), which should be complemented with a plasmid containing a viral *alaS* gene (C4 pathway). In order to test whether ST18 can be used for functional complementation using a C4 pathway enzyme, an ST18 strain-producing RcA was used as a control experiment. Furthermore, another control was established to recover the ALA auxotrophic phenotype with another C5 pathway

enzyme, namely GluTR from *Methanopyrus kandleri*. Both *hemA* and *alaS* genes were controlled by a *tac* promotor and could complement the ALA-auxotrophic strain in the presence or absence of an inducer, while the strain ST18 with empty vector (EV) control showed no real growth (Figure 29 B). Thus, the ALA auxotrophic strain can be complemented with a C4 or C5 pathway enzyme, even without an inducer when a leaky promotor is present.

In order to verify the activity of viral sequences, a pre-test was conducted using vectors with different viral *alaS* genes for complementation, some with solubilization-enhancing tags and others without, to assess activity and avoid potential issues with protein folding. Only one vector showed functional complementation, referred to as vALAS from now on (ALAScryopt from phage CB_2 in pCYB1 vector without tag). Without inducing gene expression, GluTR- and RcA-producing strains could still reach optical densities like the strain ST18 (Figure 29 B,C). The ST18 strain producing vALAS was also able to complement the ALA-auxotrophic strain in contrast to the strain with EV control (Figure 29 A), demonstrating that vALAS encodes a functional enzyme that can synthesize ALA. As a control, the wild-type (WT) strain S17.1 and the Δ*hemA* deletion mutant strain ST18 showed no influence on growth during the presence or absence of IPTG (Figure 29 C). Curiously, the presence of IPTG did not led to enhanced growth in any of the strains. Instead, it appeared to induce cellular stress, possibly resulting in slower growth. This was verified for strains producing GluTR and RcA, whereas the vALAS-producing strain had only a minor growth deficit compared to the strain without IPTG induction (Figure 29 B). A growth deficit may happen because heme pathway enzymes are usually tightly controlled, and an overproduction of one enzyme can significantly disbalance the pathway. Interestingly, the *E. coli* strain ST18 producing the C5 pathway enzyme GluTR seems to be stronger restricted in its growth than the strain producing the C4 pathway enzyme RcA.

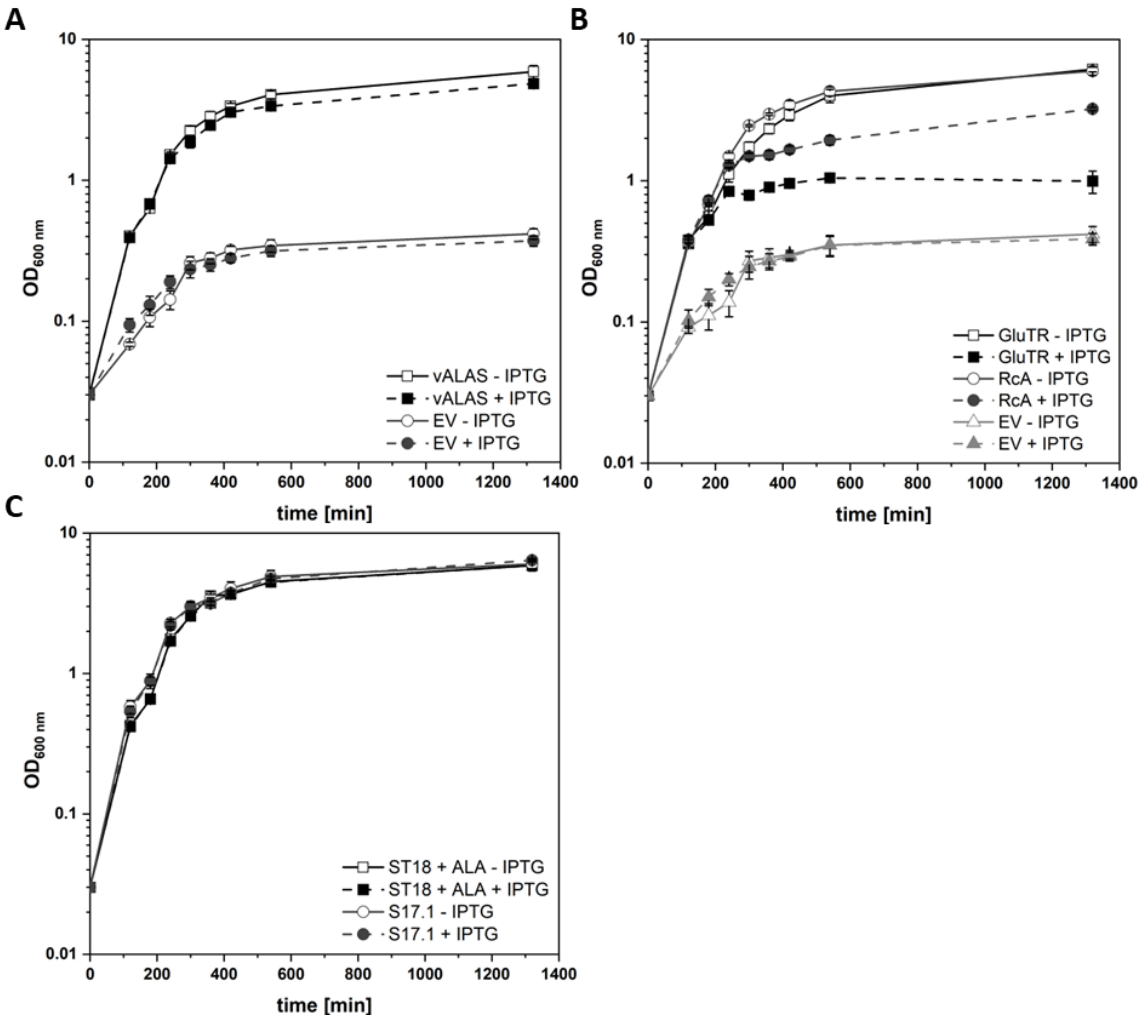


Figure 29: Growth curve of plasmid-based functional complementation of ALA-auxotrophic *E. coli* strain ST18. *E. coli* growth in the dark was measured in the presence (filled symbol) or absence (empty symbol) of gene expression inducer IPTG for *tac* promotor controlled C4 or C5 pathway genes in comparison to the strain ST18 with empty vector (EV) control. **A** vALAS producing strain complements the growth deficit of the auxotrophic *E. coli* strain regardless of induction of gene expression in contrast to the strain with EV control. **B** Strains producing GluTR and RcA showed complementation of growth deficit whereas the strain ST18 with EV control shows minimal growth. **C** Control strains ST18 supplemented with ALA and wild type S17.1 do not show an influence of the presence or absence of IPTG during growth.

In general, slow growth of the strain ST18 with EV control was detected and showed no further growth after reaching a certain threshold ($OD_{600\ nm} < 0.5$). One could speculate that the cell uses an internal ALA pool, which enables a limited ability to grow or maintain cellular functionality even without an enzyme producing ALA until this pool is depleted. Another test was carried out to verify that this limited growth does not indicate true complementation. Hereby, the cultures were plated out after finalizing the growth curve on a selective medium without ALA supplementation. There, it was apparent that the strain ST18 with EV control could not form colonies anymore as no growth was detectable (data not shown), whereas the plasmid-based complementation

using the C4 and C5 genes showed the formation of colonies (Figure 30), thus indicating proper recovery of the phenotype. The basal expression of the *tac* promotor was sufficient for complementing ALA auxotrophic growing *E. coli* strains when a C5 or C4 (bacterial or viral) gene was present (Lozano Terol *et al.,* 2021).

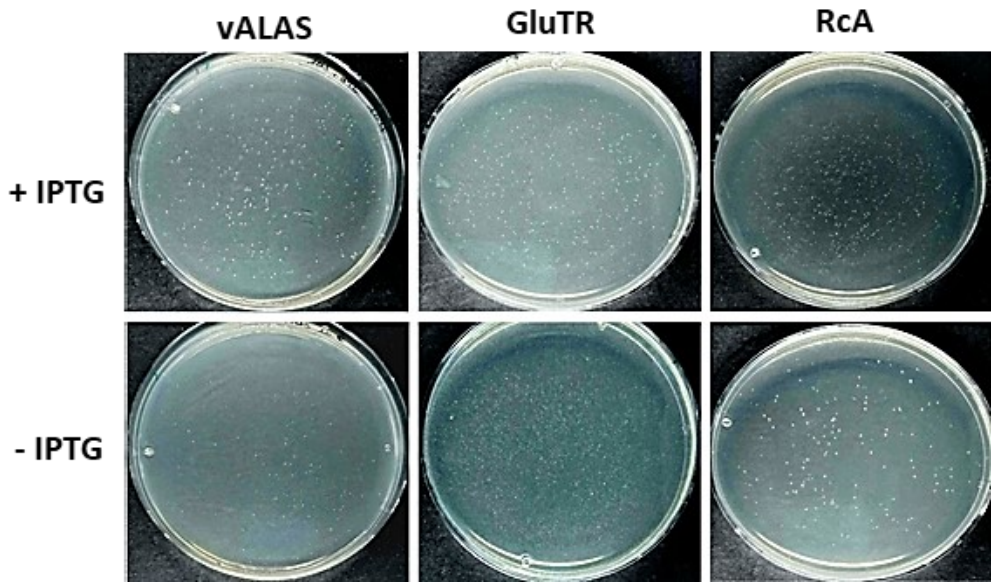
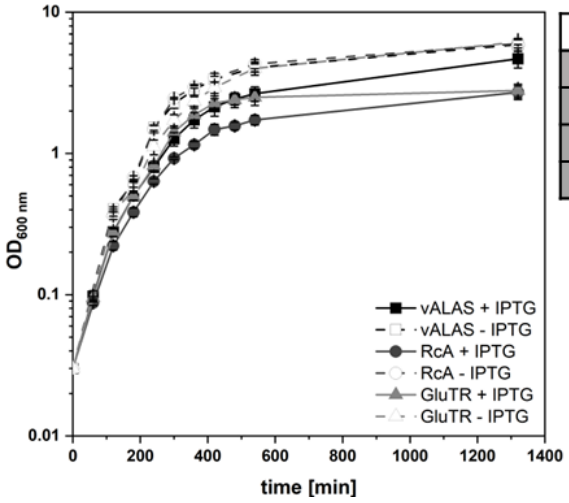


Figure 30: Growth on selective plates of functional complemented *E. coli* after growth in liquid culture. Plating of complemented *E. coli* strain ST18 grown for 24 h in the presence or absence of gene expression inducer, colonies indicating successful complementation with C4 or C5 pathway enzymes. All here tested strains displayed colony formation, even though some showed small colony morphology. For improving visibility, the pictures were adjusted regarding contrast and color saturation.

In order to determine the influence of IPTG concentration on growth and maybe facilitating cell stress because of heme overproduction, a lowered IPTG concentration was tested (Figure 31). This could not significantly improve the growth as, most likely, even a lowered IPTG concentration still yields a protein production of RcA or GluTR that is too high, resulting in negative feedback inhibiting both enzymes. Interestingly, the strain ST18 producing vALAS did not show any difference between the presence or absence of any tested IPTG concentration (Figure 29 A, Figure 31).



Growth rate [h^{-1}]			
	- IPTG	5 µM IPTG	500 µM IPTG
vALAS	0.56 ± 0.02	0.47 ± 0.08	0.52 ± 0.02
RcA	0.56 ± 0.03	0.27 ± 0.02	0.35 ± 0.03
GluTR	0.59 ± 0.07	0.27 ± 0.02	0.24 ± 0.01

Figure 31: Analysis of growth influence of gene inducer in plasmid-based functional complemented ALA-auxotrophic *E. coli* strain ST18. *E. coli* growth in the dark was measured in the reduced presence of 5 µM (filled symbol) or absence (empty symbol) of gene expression inducer IPTG for *tac* promotor controlled C4 or C5 pathway genes. The growth rate was determined in the exponential phase of growth in the presence of different IPTG concentrations and the absence of it for C4 and C5 pathway genes.

Complementation leads to the accumulation of porphyrins and ALA secretion into the medium

The complemented cultures were light-sensitive and had to be cultivated in the dark; otherwise, significant growth deficits were noticeable (data not shown). When cultivated in the dark, some cultures showed a color change, maybe due to the presence of pigments or pyrroles secreted into the medium or located in the cells. This finding would also explain the light-sensitivity of the cultures during the cultivation. Detecting the medium's color change and the culture's light sensitivity, the cell lysate was investigated for the accumulation of pigments, which can indicate tetrapyrrole pathway manipulation.

The vALAS expressing culture showed a strong red color of the medium and also the cells themselves (Figure 32 C). The absorption of the soluble lysate showed distinct signals at 400 nm, indicating a Soret band, and less pronounced signals at 500 nm, 538 nm, and 563 nm, displaying Q bands (Figure 32 A), suggesting the presence of porphyrins. Additionally, fluorescence spectroscopy showed a Stokes shift of the emission to 615 nm and a shoulder at 655 nm to 680 nm after excitation at 400 nm, confirming porphyrin accumulation (Figure 32 B). Interestingly, the other cell lysates do not show such a strong formation of colored compounds and only showed a broad shoulder with absorption maxima from 400 nm to 430 nm in spectroscopy measurements (Figure 32

A). The *E. coli* ST18 supplemented with ALA exhibited slight fluorescence in the cell lysate but with similar fluorescent maxima to vALAS (Figure 32 B).

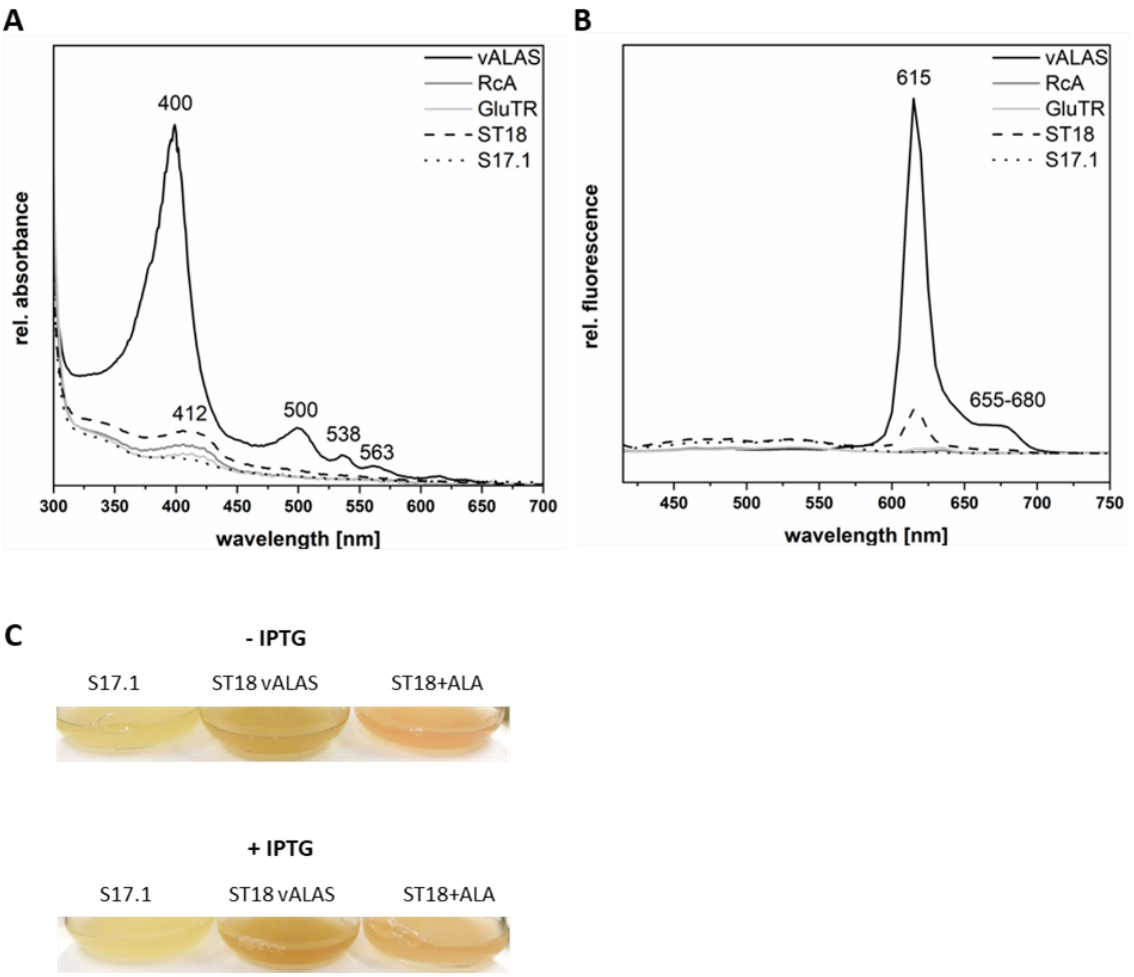


Figure 32: Viral ALAS leads to porphyrin accumulation in the cell lysate of complemented strains. A + B Cell lysate of in the dark cultivated *E. coli* S17.1 and ST18 with/without complementing C4 or C5 pathway enzymes was analyzed using absorption (**A**) and fluorescence **(B**) measurement. **A** Absorption scan detects Soret band at 400 nm and Q bands at 500 nm, 538 nm and 563 nm were observed for strains producing vALAS and a broad absorption peak from 400-430 nm for S17.1, ST18 supplemented with ALA and ST18 strains producing GluTR and RcA. **B** Fluorescence excitation at 400 nm show Stokes shift with emission at 615 nm and a shoulder at 655-680 nm, indicating the presence of porphyrins in the cell lysate of strains producing vALAS and ST18 supplemented with ALA. **C** Culture color in the presence and absence of IPTG when complemented with either vALAS or ALA from ST18. As a control S17.1 is displayed without complementation.

vALAS synthesizes ALA using the C4 substrates glycine and succinyl-CoA

To detect also the product of the C4 or C5 pathway enzymes, the formation of ALA was measured spectroscopically in the cell-free medium after growing the complemented strains for 24 h (Figure 33 A) using a colorimetric detection (Mauzerall & Granick 1956). The ALA secretion of the different complementation strains was compared to the ancestor strain of ST18, which has no deletion of the *hemA* gene, namely S17.1. Overall, a similar level of ALA concentration for all strains was detected when they were not induced with IPTG, even for S17.1 if it was grown supplemented with ALA (Figure 33 A).

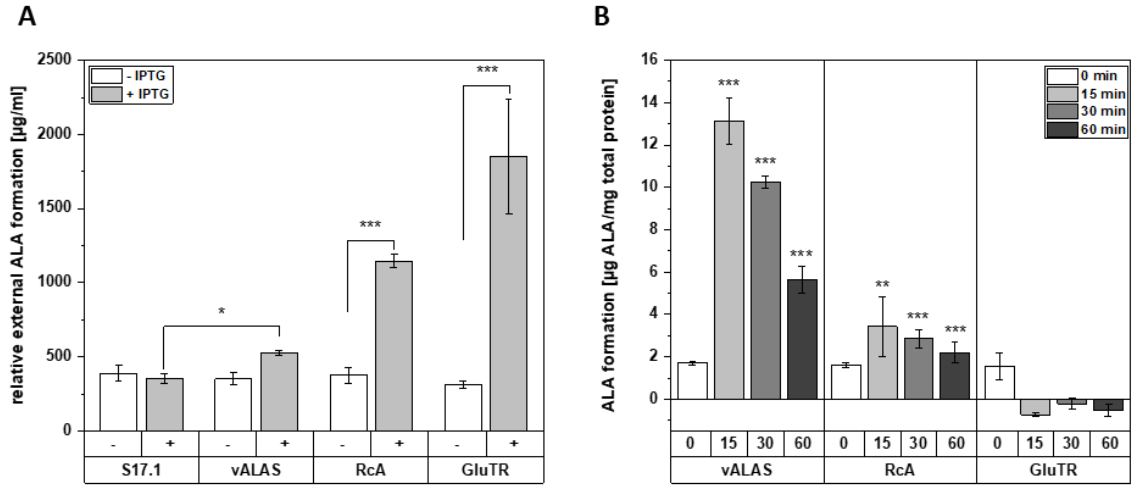


Figure 33: Viral ALAS leads to presence of ALA in the medium and shows distinct enzyme activity in the cell lysate of complemented *E. coli*. A External ALA was measured, in the cell free medium in which the *E. coli* strains (WT S17.1, the deletion mutant strain ST18 or its complemented strains of ST18) were cultivated before either in the presence or absence of gene inducer IPTG, to detect the presence of secreted product (ALA) using Ehrlich reagent. **B** Cell lysate with a adjusted protein concentration [2 mg/ml] of complemented *E. coli* ST18 strains, with either C4 or C5 genes induced with IPTG, was supplemented with specific substrate of C4 pathway enzyme (glycine and succinyl-CoA) to detect product formation over time using Ehrlich reagent. Time points demonstrate the respective amount of ALA that could be measured after subtracting the already present ALA in the cell (time point 0 min). Statistical test: Significance levels were assigned accordingly $p \leq 0.05$ = *; $p \leq 0.01$ = ** and $p \leq 0.001$ = ***.

In contrast, the IPTG addition and, thus, the expression of genes encoding for enzymes producing ALA led to significant differences in the secreted ALA concentration. While strains producing GluTR and RcA showed that significantly more ALA was present when IPTG was added to the culture, the strain producing vALAS showed a significant increase in the secreted ALA concentration compared to the WT S17.1 (Figure 33 A). Summarized, this would lead to the assumption that gene induction of C4 and C5 pathway enzymes promotes the formation of ALA or amino ketones in contrast to the non-induced cultures. Strains producing RcA and GluTR demonstrated even higher ALA formation than the strain producing vALAS and thus may indicate a more active enzyme.

Nevertheless, the presence of secreted ALA could be detected during the complementation of an ALA-auxotrophic strain, and for the strain producing vALAS, ALA levels were significantly elevated compared to the WT.

In order to prove the activity of the enzyme that was produced during the growth of the culture in the presence of IPTG, a C4 pathway-specific enzyme assay was developed. ALA formation was determined using Ehrlich reagent, and only the C4 substrates glycine and succinyl-CoA were used. Timepoint zero represents the already present ALA in the cell due to the complementation, whereas the later timepoints were subtracted from the first timepoint, thus showing only the newly formed ALA and, indirectly, the enzyme activity (Figure 33 B). Remarkably, the strain producing vALAS shows a significant formation of ALA over time, while the highest formation is at 15 min. The later time points show a reduced activity of vALAS, possibly due to the thermal instability of vALAS enzymes or because proteins of the downstream pathway are also present in the cell lysate, which may lead to the metabolization of ALA. In contrast, interference with the catalytic activity of ALAS by contaminants present in the cell lysate cannot be excluded. The strain-producing RcA seems to be active in this assay and shows a similar activity pattern as the strain-producing vALAS, with 15 min as the highest ALA production and decreasing ALA formation in later timepoints.

RcA is known to be feedback inhibited by heme, which is crucial for tight regulation of the biosynthesis pathway and can lead otherwise to the accumulation of photoreactive compounds that can form ROS or, in general, stress the cell. The assay's specificity is represented in the results of the strain that produces GluTR. Here, the presence of ALA at timepoint 0 min displays the concentration of present ALA, but there is no ALA formation at later timepoints, indicating that this assay only shows the activity of ALAS enzymes. ALA formation was also tested for the ancestor strain S17.1 cell lysate and the ALA auxotrophic strain ST18 supplemented with ALA (Figure S 3). There was no formation of ALA after timepoint 0 min observable, thus verifying the selectivity of the assay where a C5 pathway enzyme cannot interfere with the measurement of ALA formation. Thus vALAS is an active ALA forming C4 pathway enzyme.

3.3.2 Genomic context of freshwater phages carrying *alaS* genes

As this cluster of genes (v*ho1*, v*pcyA*, v*alaS*) found in marine and freshwater phages was verified to encode for active enzymes, at least for freshwater ALAS, the genomic context of the phages was analyzed further. The marine phages were only partially assembled to whole genomes and were omitted from further analyses as no activity of ALAS was verified so far. The freshwater phages from (Chen *et al*., 2020) were mostly fully assembled and predicted which host they infect. This was done regarding the co-occurrence of existing bacteria by time-series analyses and the similarity between the PmoC (part of a trimer of subunits building the enzyme methane monooxygenase, pMMO) sequences in phages and bacteria (Table 18), which was the focus of the publication. The phylogeny of these phages was carried out using a marker DNA polymerase and 12 concatenated protein subfamilies, indicating that all *pmoC*-phages are *Myoviridae* (Chen *et al*., 2020).

The proteins characterized in this work were from the phage CB_2, which is predicted to infect an alphaproteobacterial host, in this case, *Methylocystis* sp. (Chen *et al*., 2020). In addition to the proven activity of vHo1, vPcyA, and vALAS, which are essential for tetrapyrrole synthesis, the phage CB_2 has further AMGs that encode for enzymes important in methane metabolism (*cofF,* Coenzyme F420-3) and carbon fixation (RuBisCo, large subunit). Other phages like CB_1 and TP6_1 were also predicted to have the genes encoding for an ALAS, Ho1, and an FDBR. While CB_1 is also predicted to infect an alphaproteobacterium belonging to the *Methylocystis* sp., the phage TP6_1 is predicted to infect *Methyloparacoccus* sp., a gammaproteobacteria. As gammaproteobacteria do not possess the C4 pathway to produce ALA, it is quite interesting that a phage infecting them does have the C4 pathway enzyme ALAS encoded in its genome. The phage genomes were also analyzed regarding the presence of additional genes encoding for tRNAs, as these are important for protein translation during phage infection, while $tRNA^{Glu}$ is crucial for the C5 pathway. CB_1 and CB_2 harbored additional tRNAs, whereas CB_1 also has $tRNA^{Glu}$. Additionally, in the genome of CB_1 was the gene *ahbD* found that is used in heme synthesis for converting iron-coproporphyrin III into heme *b*, known as the last step of the siroheme-dependent pathway.

Table 18: Genomic content of analyzed freshwater phage genomes. Plus and minus signs indicate the presence or absence of the gene of interest in the respective genome, while a question mark indicates that the analysis has not yet been completed.

Phage	Predicted host	*alaS*	*ho1*	*pcyA/X*	Genes of tRNAs ($tRNA^{Glu}$)	*pmoC*
CB_1	Alphaproteobacteria – *Methylocystis* sp.	**+**	**+**	**+** (putative PcyX)	**+ (+)**	**+**
CB_2	Alphaproteobacteria – *Methylocystis* sp.	**+**	**+**	**+** (PcyA)	**+ (-)**	**+**
TP6_1	Gammaproteobacteria – *Methyloparacoccus* sp.	**+**	**+**	**+** (putative PcyX)	**?**	**+**

Subsequently, the predicted host families *Methylocystis* sp. and *Methyloparacoccus* sp. were analyzed regarding the presence of homologous proteins. Furthermore, the predicted host families were analyzed for potential acceptors of the products generated by the phage-encoded proteins verified in this study. This investigation could help to determine the benefit of the phage carrying these genes when infecting these hosts. Proteins like FDBRs were not found in any of the predicted hosts. In contrast, genes for ALA-forming enzymes were present, but only homologs that are typically required in native C4 or C5 pathways (Table 19). For some *Methylocystis* species, the presence of Ho1 could be verified. This led to an investigation of the presence of the proteins that could bind the respective products of phage and host Ho1, namely BV IXα and the phage-encoded FDBRs, which could yield PCB or another bilin as a product. Multiple enzymes could be a possible acceptor of these products, e.g., phytochromes or photoglobins. Photoglobins are fused to B_{12}-binding domains and may act regulatory in response to different light conditions in prokaryotes (Schneider *et al.,* 2022). The presence of phytochromes could not be verified for gammaproteobacteria but for some alphaproteobacteria. Here, some *Methylocystis* species carried genes that could encode for phytochromes but most likely belonged to bacteriophytochromes, thus only capable of binding BV IX and not PCB. Another possibility of product binding (PCB) would be photoglobins or GENOMES UNCOUPLED 4 (GUN4) (Wilde *et al.,* 2004). GUN4 is part of the magnesium chelatase and thus catalyzes a committing step in chlorophyll biosynthesis, thereby providing a photoprotective role by binding PCB. The

determination of these was not trivial and cannot definitely be determined as they are under the current research focus.

Table 19: Host usage prediction of viral genes encoding proteins. Plus and minus signs indicate the presence or absence of the gene of interest encoding for the respective protein in the analyzed genome, while a question mark indicates that the analysis has not yet been completed.

Predicted host	Infecting phages	ALA synthesis pathway present in host	Ho1 present in host	FDBR present in host	Phytochrome present in host (BV IX or PCB)	Photoglobin/ GUN4 homolog (PCB)
Alphaproteobacteria – *Methylocystis* sp.	CB_1, CB_2	C4 (ALAS)	+	-	+	?
Gammaproteobacteria –*Methyloparacoccus* sp.	TP6_1	C5 (GtrR + GsaM)	-	-	-	?

Interestingly, in some *Methylocystis* species, e.g., *M. rosea* SV97, there is a genetic cluster consisting of genes encoding for a magnesium chelatase, a photosynthesis transcriptional regulator PpsR, a phytochrome, a heme oxygenase, and a cobalamin B_{12} binding domain-containing protein. The latter is frequently wrongly annotated and most often a photoglobin. The transcriptional regulator, the phytochrome, and the heme oxygenase are orientated in one direction and most likely form an operon, regulating the genes together. This operon-like structure is similar to the lateral acquired genes of *Bradyrhizobium* sp. ORS 278 consists of a bacteriophytochrome, an anti-sigma factor, a Ho1, PcyA, and ALAS on a genetic island (Jaubert *et al*., 2007). Interestingly, this *Bradyrhizobium* has four ALAS genes that all vary a bit from each other regarding DNA and protein sequences, but all contain the crucial amino acids to be catalytically active (Figure S 8).

4. Discussion and outlook

The cycling of nutrients and carbon in the oceans is greatly influenced by marine bacteriophages. On their genomes are genes encoded that are thought to modulate and supplement the host metabolism, namely AMGs. In marine phages, these genes are often linked to photosynthesis, cycling of nutrients, carbon metabolism, and protein synthesis (Crummett *et al.,* 2016, Lindell *et al*., 2005, Puxty *et al.,* 2015, Shan *et al.,* 2008). The function and regulation of AMGs and their impact on the host are not yet sufficiently understood. Meanwhile, the current appearance of freshwater bacteriophage genomes grants insights into a new pool of possible AMGs.

In viral metagenomic data of marine origin, a three-gene cassette was found, encoding for proteins responsible for the first step in tetrapyrrole biosynthesis (*alaS*) and generating linear tetrapyrroles out of heme (*ho1, pcyA/X*). The novelty of this finding lies in the presence of *alaS* in this cassette, which was so far never shown to be present in such a viral cassette nor proven functional when originating from a viral context. Marine and artificial *alaS* genes (Table 20) were tested for solubility, but only ALASsyn showed slight solubility during purification (Figure S 4), where SEC and thermal shift assay revealed aggregation with the GST-tag (Figure S 5 and data not shown) and the enzyme showed no activity in the coupled enzyme activity test (data not shown) and non-significant activity in the colorimetric test (Figure S 6).

A study about bioinformatically curated large phage genomes from freshwater lakes with methanotrophic bacteria present also happened to find an *alaS* gene in bacteriophages (Chen *et al*., 2020). Here, the gene cassette was split, and *ho1* and *pcyA/X* were found further up-/downstream from *alaS* in the viral genome. Intriguingly, freshwater and marine viruses in the dataset of this study were shown to have these three genes. The genes were organized on the genome regarding the habitat from which the phages were derived (Figure 8, Figure 14). In marine viruses, the three gene cassette was conserved. Whereas, for freshwater phages, only *ho1* and *pcyA/X* were located next to each other, and *alaS* is localized further away, similar to the study of (Chen *et al*., 2020). This finding was further proven valid through curated phage genomes from freshwater lakes in Berlin, displaying the same organization as the freshwater phages

analyzed before, thus maybe indicating that marine and freshwater viral gene cassettes were derived from different ancestors. An earlier found viral gene cassette consisting only of *pcyX* and *ho1* from a marine source was already proven to encode functional enzymes (Ledermann *et al*., 2016). Therefore, the unrevealed activity of the viral *alaS*, which encodes for an ALAS, and the remaining genes of this cassette were investigated.

4.1 The activity and presence of ALAS in phages

The presented vALAS in this study from freshwater phage CB_2 was bioinformatically and biochemically analyzed. From the bioinformatic perspective, all crucial binding sites of the predicted model of vALAS using AlphaFold of the substrates (glycine and succinyl-CoA) and cofactor (PLP) were verified by comparing it to characterized enzymes from *Rhodobacteraceae*, human and yeast origin (Astner *et al*., 2005, Brown *et al*., 2018, Ikushiro *et al*., 2018, Shoolingin-Jordan *et al*., 2003). As this indicated an active viral enzyme, further biochemical analysis should be conducted using vALAS to determine enzyme kinetics and possible inhibition. The required purification of vALAS did not yield active and soluble enzyme to perform activity assays, as the formation of inclusion bodies was always observable. Different methods and codon adaptions were tested to increase solubility and yield but did not show improvement. Ultimately, vALAS activity was proven by complementing the lack of the $tRNA^{Glu}$-dependent C5 pathway (ALA-auxotrophic) in bacteria. vALAS is so far the first biochemically verified C4 pathway enzyme from phages to be active and able to catalyze the synthesis of ALA from glycine and succinyl CoA.

Regarding its origin, the gene encoding for vALAS was derived from a freshwater environment where the genome of the phage was bioinformatically curated by (Chen *et al*., 2020), and the host was predicted to be the alphaproteobacteria *Methylocystis* sp., based on *pmoC* (encodes for a subunit of methane monooxygenase) taxonomy. In their study, other phages were also shown to have the gene *alaS,* namely CB_1 and TP6_1. Both phages are in the same phylogenetic cluster, but TP6_1 has gammaproteobacteria as the host taxon, and the predicted host is *Methyloparacoccus* sp., while CB_1 is predicted to infect alphaproteobacteria *Methylocystis* sp. (Table 18). Bioinformatic predictions only suggest potential hosts, and since phages can have broad or narrow host ranges, they may infect hosts with C4 or C5 pathway enzymes. The occurrence of

alaS in different phages, which are predicted to infect diverse hosts of diverse phylogenetic groups, raises questions: 1) Why would an *alaS* gene be beneficial for a phage infecting a host that has the C5 pathway present and utilizing a different enzyme for ALA formation? 2) Why would a viral *alaS* gene be advantageous for a phage infecting a host with a genomic *alaS* gene in the C4 pathway? The answer to both questions may lie in the regulation of heme formation. The C4 and C5 pathway enzymes are known to be regulated tightly by heme (Burnham & Lascelles 1963, Chen *et al*., 1996, Ikushiro *et al*., 2018, Jones & Elliott 2010, Levicán *et al*., 2007). A lack of regulation could benefit phages infecting C4 and C5 organisms, allowing continuous heme production for optimal phage progeny production. Interestingly, the vALAS sequence was indicated bioinformatically to be not negatively inhibited by heme (Figure S 7), which was analyzed by verifying the enzyme activity.

vALAS seems to be not negative feedback inhibited by heme

Proving activity of vALAS was succeeded through plasmid-based functional complementation of an ALA-auxotrophic *E. coli* strain. Enzymatic activity for the marine viral ALAS could not be verified so far (Figure S 4 - Figure S 6). However, the freshwater vALAS producing strain could complement the growth deficit with or without an inducer for gene expression. The used control strains producing RcA and GluTR showed reduced growth when the inducer was present, which may be explained by the high enzyme activity of these enzymes (Figure 29)(Kaufholz *et al*., 2013, Moser *et al*., 1999). This could trigger a metabolic imbalance of the heme flux that stresses the cell and limits its growth, as known from the literature (Everse & Hsia 1997, Nakahigashi *et al*., 1991).

On the contrary, this would indicate that the vALAS is less active than RcA and GluTR. Also, negative feedback inhibition of the upregulated heme pathway by the end product (heme) could lead to this restricted growth in the control strains producing RcA and GluTR since an accumulation of heme is toxic for the cell (Kumar & Bandyopadhyay 2005). The vALAS may not affected since it seems to be missing the regulatory site for negative feedback inhibition of heme common in alphaproteobacteria (see blue bars, Figure S 7) (Ikushiro *et al*., 2018). These sites were identified in *Caulobacter crescentus* as an H/C motif, lacking either of these amino acid residues makes ALAS resistant to

heme-mediated inhibition (Ikushiro *et al*., 2018). It is worth noting that all alphaproteobacteria that were analyzed (see blue bars, Figure S 7) possess this H/C motif, while viral sequences either have only one residue, which is observed in marine phages or lack both residues, which was found among freshwater phages. This suggests that all analyzed viral ALASs are likely unaffected by heme-mediated feedback inhibition.

Considering the arising issue of heme toxicity of the host induced by phage infection, it is logic to expect the enzyme Ho1 and PcyA/X, encoded genes on the phage genome. During phage infection, these enzymes could detoxify elevated amounts of heme caused by increased ALA biosynthesis due to the presence of both pathways, the endogenous C4 or C5 pathway of the host and the viral ALAS. As the host C4 and C5 pathway enzymes are shown to be negatively inhibited by elevated heme levels, the presence of a heme-unregulated ALAS could benefit the virus. Especially if the ALA levels are expected not to increase significantly as vALAS seems to be a not highly active enzyme (Figure 29, Figure 33), and ALA overproduction was proven to cause cell damage and morphology changes in *E. coli* due to generating ROS (Zhu *et al.,* 2019). The formation of ALA is also regulated in alphaproteobacteria by iron availability and in photosynthetic active bacteria by oxygen tension, e.g., *R. sphaeroides* harboring the C4 pathway enzyme regulates oxygen tension by modulating trisulfide concentration (Hamza *et al.,* 1998, Lascelles 1960, Neuberger *et al.,* 1973, O'Brian 2015, Rodionov *et al.,* 2006, Sandy *et al.,* 1975, Zeilstra-Ryalls & Kaplan 1996). Therefore, a complete understanding of the regulation of vALAS is still pending.

The high amounts of vALAS triggered in this study's artificial system are not expected to occur in the native host during the virocell stage. In this stage, the active proliferation of the phage happens in the host cell, resulting in high requirements for energy as ATP and NADPH to meet the demand for viral protein and nucleic acid synthesis (Forterre 2013, Howard-Varona *et al*., 2020, Thompson *et al*., 2011). Here, a constant ALA formation is probably enhancing tetrapyrrole production. This could be beneficial for the phage as the cofactor heme is an integral part of the electron transport chain and other important redox reactions, while excessive heme levels would cause a disbalance of the highly responsive redox balance of the cell. In general, the effects of the vALAS produced in the virocell are only based on assumptions since it is unclear when the transcription of phage

alaS happens or how the regulation of vALAS occurs during infection. For cyanophages, different classes of transcripts (early/middle/late) during infection could be verified, thus regulating the expression of genes according to the current infection state of the host cell (Clokie *et al.,* 2006, Doron *et al.,* 2016). This scenario could also be possible for this phage and would be advantageous as the host DNA is often degraded during infection (Thompson *et al*., 2011). Host DNA degradation would not allow further ALA production by the native enzymes of the host, which ultimately leads to ALA depletion. The start of transcripts of vALAS in the middle or late part of infection would make vALAS favorable for the phage, as it will continuously produce ALA in the host without generating too much heme to stress the cell since the host GluTR or ALAS activity is successively replaced by the viral enzyme. This hypothesis remains to be verified.

Porphyrin accumulation, ROS detoxification, and low active vALAS

The photosensitivity of colored cultures and porphyrin formation detected in the vALAS cell lysate suggest successful complementation but also indicate an imbalance in the heme pathway, leading to porphyrin accumulation, which is not expected to occur during the phage infection in the host, as it is likely an *E. coli* effect. The red color in vALAS-producing *E. coli* and fluorescence measurements confirm the porphyrin accumulation, likely due to overproduction of the pathway's first enzyme (Bolt *et al.,* 1999). Unlike strains producing RcA and GluTR, which seem to be inhibited by heme and do not show porphyrin-derived fluorescence (Figure 32 B), the strain producing vALAS is less active and not subject to negative feedback inhibition, resulting in continued ALA formation and thus porphyrin accumulation. This is also in line with the detected external ALA formation. While strains producing GluTR and RcA show significantly high ALA formation upon induction of the respective gene expression, this is reduced for the strain producing vALAS, which still shows significantly more ALA formation than the WT (Figure 33 A). Here, the activity of the enzyme could also be the reason. Highly active enzymes like RcA and GluTR are forming high amounts of ALA or precursor of ALA in a short time span, thus triggering heme formation, which ultimately leads to the inhibition of the enzyme activity of ALAS or GluTR. However, this short activity of the fast-catalyzing enzymes could still yield high ALA concentration in the cell that will upregulate the exporter RhtA in *E. coli*, which exports ALA due to its broad specificity and therefore

prevents an excessive accumulation of intracellular heme (Kang *et al*., 2011a, Kang *et al*., 2011b).

Apart from extracellular ALA formation, the intracellular enzyme activity assay verified the activity of C4 pathway enzymes and indicated the influence of inhibition. When produced in the complementation strain and incubated with C4 pathway-specific substrates (glycine, succinyl-CoA, and cofactor PLP) to determine ALA formation, only the vALAS producing strain showed high ALA formation. Although RcA is a highly active C4 enzyme when purified, its strain was expected to produce more ALA than the vALAS producing strain, but this was not observed, likely due to feedback inhibition by heme in the cell lysate. This could not be tested for the strain producing GluTR as the assay is restricted to C4 pathway enzymes. However, literature does show an inhibitory effect of heme to GluTR or targeted proteolysis in the C5 pathway (Jones & Elliott 2010, Levicán *et al*., 2007, Rieble & Beale 1988, Wang *et al*., 1997).

In general, the effect of this assay, based on cell lysate, is also a decrease in ALA over an extended incubation time (>15 min). This indicates that the enzymes of the remaining heme pathway are present and are trying to cope with the enhanced ALA production by further metabolizing ALA. The observed reduced activity of vALAS could be derived from the artificial expression system of this study or displaying the actual enzyme activity. Similar reduced activity was shown for another phage enzyme encoding a transaldolase, which had one-third lower efficiency than the host enzyme (Thompson *et al*., 2011), making the reduced activity of vALAS plausible. This may be due to the presence of multiple copies of the phage genome that are synthesized during the replication in the host, potentially expressing *alaS* and thus compensating for a less active enzyme.

4.2 The presence of PcyA and Ho1 in phages

Light-harvesting and light-sensing

The characterization of vHo1 and vPcyA from the freshwater phage CB_2 proved the activity of the enzymes to produce BV IXα and PCB, respectively. The PCB produced by vPcyA was capable of forming a functional holophytochrome. So far, it is not clear why phages harbor these genes. As these genes encode important metabolic pathways in bacteria, they are AMGs. Regarding improving the phage infection, the state of the virocell, or the phage progeny numbers, AMGs comply quite well with the prerequisites of the phage host(-range). In this case, one could guess that a photosynthetic active bacterium utilizing phycobilins would be the most likely host since genes like *ho1* and *pcyA/X* are present on the host genome, as known from cyanophages encoding photosynthesis-related genes and infecting cyanobacteria capable of oxygenic photosynthesis (Dammeyer *et al*., 2008a, Fridman *et al*., 2017, Lindell *et al*., 2005, Shan *et al*., 2008). Nevertheless, neither the cassette from (Ledermann *et al*., 2016) consisting of *ho1* and *pcyX* nor the virus CB_2 (Chen *et al*., 2020), which has the here characterized genes v*ho1,* v*pcyA,* and v*alaS*, are predicted to infect oxygenic photosynthetic bacteria.

Since alphaproteobacteria are capable of performing anoxygenic photosynthesis using bacteriochlorophyll and not linear tetrapyrroles as light-harvesting pigments (Hohmann-Marriott & Blankenship 2011, Imhoff 2006), a direct benefit for photosynthesis cannot be determined regarding the AMG *pcyA/X*. Thus, another function of FDBRs could be of interest: the synthesis of chromophores for light-sensing phytochrome photoreceptors (Hughes & Lamparter 1999). Most bacteria use BV IXα as a chromophore in phytochromes and the alphaproteobacterium *Bradyrhizobium* sp. ORS278 harbors a genomic island with a gene for a bilin reductase (*pcyA*) and a bacteriophytochrome. The latter can bind PCB and measure light intensity, not light changes, which is untypical for bacteriophytochromes (Jaubert *et al*., 2007). The identified catalytic activity of the freshwater-derived vPcyA analyzed in this work generates a functional phytochrome chromophore. This aligns with the recently proven ancestor of PcyAs (prePcyAs) that evolved in non-photosynthetic proteobacteria, which are related to the predicted host of CB_2 *Methylocystis* sp. (Rockwell *et al.,* 2023). In this host, the presence of phytochromes was predicted by sequence similarity, but it is unclear which linear

tetrapyrrole it would bind (Table 19). Since a HO was also near the predicted phytochrome, a bacteriophytochrome binding BV IX would be the most logical explanation, as no further FDBRs could be detected in these strains. Further boosting this hypothesis is a relative of *Methylocystis* sp., namely *Methylobacterium radiotolerans,* another alphaproteobacterium present in the phyllosphere, which was proven to have HOs and photoreceptors present for far-red/red light-sensing by bacterial phytochromes with bound BV IX (Consiglieri *et al.,* 2020). If the host would have a phytochrome capable of binding BV IX, then phage-synthesized BV IXα produced by vHo1 would bind to it. A respective light trigger could then activate the phototransformation, which initiates the downstream signal process consistent with the phytochrome's function as a sensor. Alternatively, the phage FDBR enzyme would reduce the BV IXα instantly, and newly synthesized phytochromes would be without chromophore and thus without function, which could benefit the infection process of the phage by inhibiting crucial light-mediated regulations. Possible benefits for such a light sensor in the virocell would be the possibility of determining the optimal time to lyse the host cells if light conditions indicate that the environment is favorable for spreading phage progeny, e.g., during the day when other potential host cells are photosynthetically active. Or for timing of viral replication and modulating of host metabolism. Here, the virus could use the light sensor in a photosynthetic active host to ensure enhanced or optimized replication during periods of light when the host resources for replication are abundant, or the host's metabolic state is modulated in such a way that pathways tied to photosynthesis are inhibited or enhanced, as known from cyanophages. Cyanophages encode and express genes for the photosynthetic light reaction while inhibiting Calvin cycle genes, thus harvested light energy is not used for carbon fixation but creates reducing equivalents and energy used for phage replication (Thompson *et al*., 2011).

If a phytochrome is not the targeted binding partner of the BV IXα from the viral Ho1, then the viral FDBR present on the phage genome may be the key. It would generate PCB (verified for phage CB_2) or PEB (suspected for phage CB_1) from the virocell BV IXα pool. A possible binding partner for PCB could be GUN4, a conserved chlorophyll biosynthesis regulator that occurs often in oxygenic photosynthetic organisms, but

related genes were also found in non-photosynthetic active bacteria. It can regulate and stimulate magnesium chelatase activity in multiple organisms, e.g., plants, green algae, and cyanobacteria (Adhikari *et al.,* 2009, Adhikari *et al.,* 2011, Brzezowski *et al.,* 2014, Formighieri *et al.,* 2012, Hu *et al.,* 2021, Wilde *et al*., 2004). It is crucial because in the presence of light and oxygen, singlet oxygen generation can harm the enzymes of the chlorophyll synthesis pathway. The subunit of magnesium chelatase, the first committed enzyme of chlorophyll synthesis, is shown to be both stimulated and protected by heme-derived linear tetrapyrroles from self-sensitized photodamage and turnover via the formation of non-photosensitizing GUN4:bilin:porphyrin adducts. (Zhang *et al.,* 2021). The phototrophic alphaproteobacteria, which would be the host, lack GUN4 but possess BchJ, a homolog whose function has not been biochemically studied (Zhang *et al*., 2021). Further, the hypothesis of PCB as a protectant against oxidative stress is supported by studies of human alpha-2-macroglobulin. Here, PCB can bind the protein and prevent damage mediated by oxidative stress (Šunderić *et al.,* 2023).

The diverse possible binding partners make it reasonable for the FDBR catalyzing the product formation of PCB, namely PcyA, to occur in viruses. In contrast, PcyX produces PEB, which cannot work in such a system (Lamparter *et al.,* 2001, Li *et al*., 1995, Murphy & Lagarias 1997). Therefore, a role in iron supply during infection has been discussed (Ledermann *et al*., 2016, Ledermann *et al*., 2018).

HO as an iron scavenger and for heme detoxification

A heme oxygenase as an AMG could be favorable for the phage during infection as an iron scavenger, for which many (bacterial) HOs are known to function under iron-limiting conditions as iron is required for electron transfer reactions in energy metabolism (Richard *et al.,* 2019, Schmitt 1997). It was shown that the virus burst size was reduced under iron-limiting conditions and interfered with viral productivity (Slagter *et al.,* 2016). The identification of *ho1* and *pcyA/X* genes in this work strengthens this hypothesis as many of the metagenomic data are acquired from iron and nutrient-poor environments (De Baar *et al.,* 1995, Martin & Fitzwater 1988, Martin *et al.,* 1994) from the Tara Oceans expedition, lake Grosse Fuchskuhle (part of LIMNOS German lake dataset PRJEB47226)

and different peat bogs (Chen *et al*., 2020, Tara Oceans Consortium Coordinators & Tara Oceans Expedition Participants 2017, Wegner *et al*., 2024).

Heme oxygenases could also serve as a heme-degrading enzyme to provide a counter-regulating component to omit heme toxicity through the additional presence of the viral ALAS. In mammalian HOs, it was shown that the biliverdin-degrading enzyme, a biliverdin reductase, would bind to the mammalian HO and facilitate product release from the HO and accelerate catalytic turnover (Liu & Ortiz de Montellano 2000, Wang & Ortiz de Montellano 2003). The function of BV IX in non-photosynthetic bacteria remains elusive. Cyanobacteria, algae, and plants can use it for the synthesis of light-harvesting proteins, and the linear tetrapyrroles used for this are also known for their antioxidative effects (Anzaldi & Skaar 2010, Frankenberg-Dinkel 2004, Stocker 2004). This would make it reasonable for the phage to harbor this enzyme. While an ALAS upregulates the product entry into the tetrapyrrole pathway, a final enzyme like the Ho1 and the FDBR could decrease the accumulating toxic heme pool. The catalysis into BV IXα is then increased by viral PcyA/X, which promotes the product release of vHo1. Even though the vALAS is not as active as RcA, it would still maintain the metabolic flow into the tetrapyrrole pathway, producing a constant amount of tetrapyrroles. This would feed into the production of heme, chlorophyll, or different tetrapyrrole intermediates that would be favorable for the phage during infection since they have essential roles in, e.g., respiration, or if the host is photoactive, also in the photosystem.

A similar situation is already known for cyanophages that maintain the cell's photosynthetic capacity during the infection by expressing viral photosystem genes (Thompson *et al*., 2011). They also manipulated the photosynthetic electron transport chain in its cyanobacterial host (Fridman *et al*., 2017). This results in an enhanced cyclic electron flow around photosystem I, which could lead to the generation of additional ATP that is thought to facilitate phage replication. Some cyanophages harbor a Calvin cycle inhibitor, which was shown to shift the metabolism from carbon fixation to the pentose phosphate pathway, thus resulting in an increased supply of nucleotides for phage replication (Thompson *et al*., 2011). If the host is participating in anoxygenic photosynthesis, one could assume that viral ALAS is crucial since it facilitates the production of the cofactors for hemoproteins, which are common redox proteins in

photosynthesis and respiration. Also, bacteriochlorophyll ultimately results from the precursor ALA and is utilized as a main light-harvesting pigment. So does the anoxygenic photosynthetic *R. sphaeroides* synthesize product out of ALA, which is not only used for heme synthesis. ALA is also utilized as a precursor for bacteriochlorophyll and vitamin B_{12} (Cobalamin), which are essential cofactors for methionine and bacteriochlorophyll biosynthesis (Zappa *et al.,* 2010). Furthermore, the synthesized PCB of the phage CB_2 could also be an intermediate that can be further converted. There are some cyanobacteriochromes known that bind PCB and further isomerize it to PVB for increased green light-sensing (Rockwell *et al.,* 2012).

4.3 The biological relevance of the cassette

Host range and viral gene origin

When a phage containing vALAS infects a non-alphaproteobacterial host, it could gain three benefits: 1. reducing the genomic load of the phage DNA packaging by using one enzyme rather than two, as it is required for the C5 pathway, in a manner similar to PebS (where one unique phage enzyme comprises the function of two host enzymes, catalyzing the same reaction) 2. Maintaining the production of heme and tetrapyrroles to meet the need for tetrapyrrole cofactors such as cytochrome c oxidase, a respiratory heme-containing enzyme, and 3. Utilizing host $tRNA^{Glu}$ exclusively for protein synthesis, which may be useful during infection for phage progeny production (Figure 34). Because heme proteins can use iron as a terminal electron acceptor during respiration, uncoupling of heme and protein production is a known mechanism of organisms requiring vast numbers of heme proteins (Levicán *et al.,* 2005). Bioinformatic research further revealed that the vALAS did not have the previously identified heme binding sites for negative feedback inhibition (Ikushiro *et al*., 2018). This would allow the host to produce heme continuously and independently of the host-native ALA formation.

Both alphaproteobacteria and non-alphaproteobacteria are predicted to be infected by ALAS-bearing phages (Chen *et al*., 2020). However, because of their large genomes, the infecting phages are thought to have a broad host range, which raises the possibility that

some of the genes linked to photosynthesis may have come from oxygenic phototrophs, like cyanobacteria.

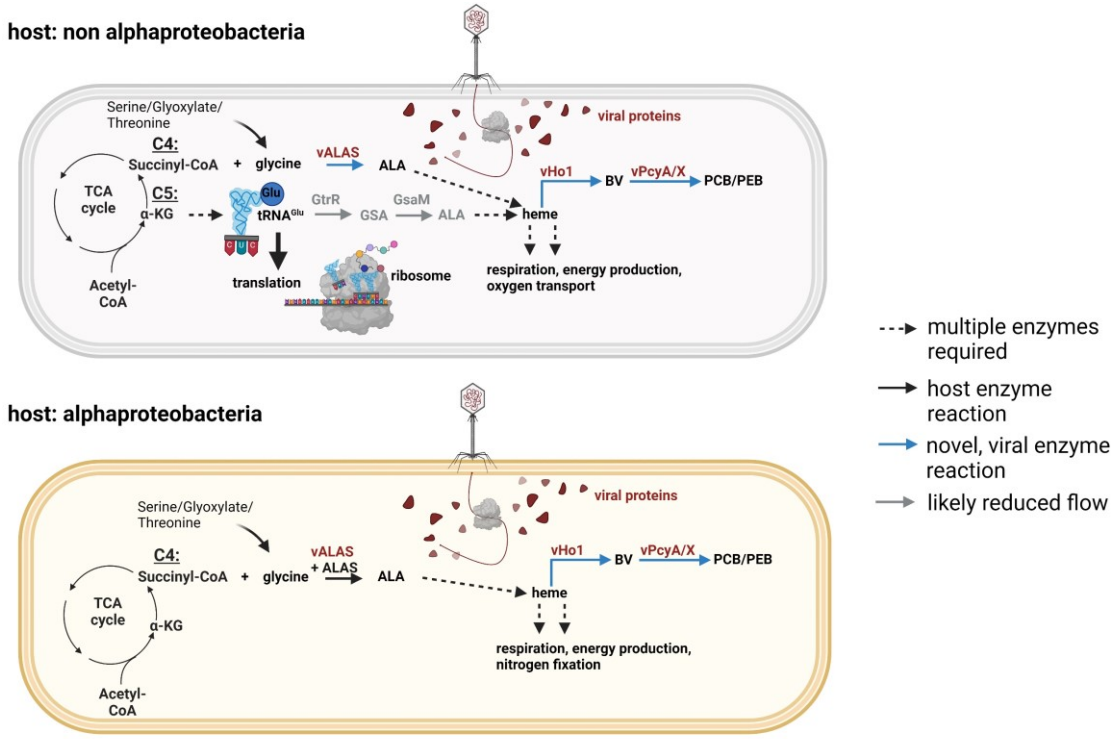


Figure 34: Proposed model of non-alphaproteobacterial and alphaproteobacterial cells during infection of phage carrying all three tetrapyrrole biosynthesis genes. Viral proteins are marked in red while different types of arrows indicate if multiple enzymes are required or display predicted changes in metabolic flow and reactions that occur. In non-alphaproteobacterial hosts, vALAS drives tetrapyrrole biosynthesis during infection, allowing the virocell to use glutamyl-tRNA for protein synthesis instead of using it for tetrapyrrole production. In the alphaproteobacterial hosts, vALAS likely complements the host's ALAS, ensuring sufficient tetrapyrrole production to support respiration and energy generation. In both host heme is also used by the phage proteins vHo1 and vPcyA/X to form phycobilins, like PCB or PEB regarding to the present FDBR, which are not known to be present in the host. See text for further details. Figure created with BioRender.com.

Another explanation for the origin of the genes could be *Bradyrhizobium* sp. ORS 278. This is the only alphaproteobacterium known that carries a genomic island including *ho1*, *pcyA*, and *alaS* and a phytochrome able to bind PCB (Jaubert *et al*., 2007). Considering the organism's ability to be photosynthetically active before and during the formation of nitrogen-fixing stem nodules in *Aeschynomene* (i.e., *A. indica, A. sensitiva*) that grows in floodplains or swamps, it is not unlikely that the viral cassette was obtained from this organism (Giraud *et al.,* 2000, Jaubert *et al.,* 2004, Jaubert *et al.,* 2009). *Bradyrhizobia* is thought to have evolved from a photosynthetic ancestor by acquisition of nodulation genes as the ability to stem nodulation of this legumes is hypothesized to be an evolutionary adaption to flooding (Giraud & Fleischman 2004, Ladha *et al.,* 1992, Loureiro *et al.,* 1995). Therefore, the origin of the cassette could be the *Bradyrhizobia*

ancestor or another unidentified organism that transferred it through horizontal gene transfer. However, it is striking that multiple hits were obtained during the search for the ALAS-encoding gene in this bacterium (Figure S 8). These hits correspond to four distinct ALAS sequences, whose genes were encoded within the genome of *Bradyrhizobium* sp. ORS278. Among the identified hits, it is significant that only the first hit is located on the genomic island, whereas the other hits for *alaS* sequences are not predicted to be linked with a genomic island. This finding raises questions as to why this organism possesses four *alaS* genes, whether all of them are functional, and where they derived from. Maybe one of them could be an ancestor of the viral ALAS sequences found in this study.

Influence of the viral gene cassette on different hosts

In order to structure the possible influences of the viral cassette regarding the host the phages infect, a schematic overview was created (Figure 34). If the reduced activity of vALAS shown in this study reflects the rate of ALA formation in the virocell, it could be beneficial for the infection process since the viral ALASs might be insensitive to heme feedback inhibition. The enzyme would be continuously active even when the host ALAS or GluTR is already downregulated, but not so active that heme would accumulate in toxic levels. If the enhanced ALA formation would still cause cellular stress, the Ho1 and PcyA/X could detoxify excessive heme levels as an instance of counter-regulation. If the phage is infecting a host that belongs to the bacteria with C5 pathway enzymes thus, acquiring a functioning C4 pathway enzyme could increase heme production. At the same time, the glutamate pool for the C5 pathway could be utilized independently of the infection process, thereby making heme biosynthesis independent from protein synthesis. The viral ALAS could also be differently regulated as the C5 pathway enzymes, thus pushing the metabolic flow toward more product formation.

During infection, the virocell is not thought to be stressed by the production of elevated porphyrin levels mediated by both the introduced vALAS and the native ALA-forming enzymes, as the vALAS activity seems to be low. If this would not be the case, the produced porphyrins could react with light or metal ions and generate ROS, which can cause oxidative damage to the cell (Duong *et al.,* 2014, Nakahigashi *et al*., 1991).

Excessive levels of ROS can be reduced by hemoproteins such as catalases and peroxidases. Given that ALA production also leads to the generation of heme, an increase in hemoproteins may assist in detoxifying the cell from detrimental ROS. ROS was also suggested to lead to capsid mispacking in the cyanophages P-SSP7 and P-HM2 (Laurenceau *et al.,* 2021). Interestingly, the presence of AMGs in the phage P-HM2 seemed to play a role in coping with ROS stress. While it contains many AMGs that control the host metabolism, a reduced occurrence of capsid mispackaging at higher irradiation was observed (Laurenceau *et al*., 2021). Additionally, it would be beneficial for the phage to have an additional *alaS* gene as backup, as host DNA is degraded during infection and is used to replicate the phage (Thompson *et al*., 2011, Wikner *et al.,* 1993). In general, viral *alaS* genes in phages could positively impact phage infection and replication in diverse ways.

Summarized, the function of an ALA-forming enzyme for the viral ALAS from a freshwater phage was proven. The comparison of marine and freshwater viruses revealed differences in the distribution of the gene cassette consistent with the habitat they were found in, thus indicating divergent evolutionary origin. The three genes *ho1, pcyA/X*, and *alaS* present in the viral genome are AMGs. While AMGs provide advantages to the phage during infection, their detailed mechanism remains elusive. The encoding enzymes may contribute to a continuous heme metabolic flow through a non-inhibited ALA production as an input and heme degradation to prevent heme toxicity as an output. Additionally, they may act as an iron scavenger and aid in detoxifying ROS that can arise during phage infection and through regulation of the heme pathway by the viral AMGs. This is the first time the AMGs encoding enzymes PcyA, Ho1, and ALAS from freshwater viruses could be proven as active enzymes biochemically. As the freshwater vALAS was found active using functional complementation, this was also tested with all marine ALAS constructs. No marine ALAS construct recovered the ALA-auxotrophic growth (data not shown), and no marine ALAS activity has been verified so far (Figure S 4 - Figure S 6). Next, marine ALAS could be expressed without a tag, as successfully carried out with freshwater vALAS, to see if it can recover the ALA-auxotrophic growth via functional complementation.

Further research could be focusing on the regulation of vALAS and its interaction with the heme pathway, which could provide insights into metabolic control mechanisms. Therefore, using the generated site-directed mutagenesis proteins impairing the active site of ALAS. Exploring the interplay between viral ALAS, heme production, and ROS detoxification could provide insights into the mechanisms phages employ to manipulate host metabolism and overcome stress response during infection. Also, the discovery of the evolutionary ancestor of this gene cassette would grant further insights into horizontal gene transfer and the host range of the phage.

5. Summary

Bacteriophages are key players in microbial ecosystems, often harboring auxiliary metabolic genes (AMGs) that influence host metabolism during infection. These AMGs are widely distributed in phage genomes and metagenomic sequences originating from phages, but their exact role remains enigmatic. This study explored AMGs from metagenomic datasets of diverse origins related to tetrapyrrole metabolism, a critical pathway for heme and pigment biosynthesis. Within this study, bioinformatic analysis of viral metagenomic data identified a three-gene cassette consisting of *ho1*, *pcyA/X,* and *alaS* genes. While viral *ho1* and *pcyX* have been found before in marine datasets and encode a heme oxygenase and a ferredoxin-dependent bilin reductase for linear tetrapyrrole biosynthesis, the identification of *alaS* is intriguing. *alaS* encodes a putative aminolevulinic acid synthase (ALAS), an enzyme that catalyzes the first committed step in tetrapyrrole biosynthesis in alphaproteobacteria to produce aminolevulinic acid (ALA). This so-called C4 pathway is in contrast to the C5 pathway of ALA formation, which is present in all other bacteria and archaea. This is the first report of a phage-encoded *alaS* in a cassette with linear tetrapyrrole synthesis-associated AMGs (*ho1, pcyA/X*), suggesting a potential role in increasing the metabolic flux through the tetrapyrrole pathway. The cassette was detected in various metagenomic datasets derived from marine and freshwater habitats, but in freshwater, *alaS* was not associated directly with this cassette and could be found further up-/downstream on the contig. Using a combination of *in silico* structure prediction and amino acid sequence alignment, the viral *alaS* gene seems to be a functional C4 pathway enzyme that uses glycine and succinyl-CoA for ALA synthesis. This was proven by functional complementation using plasmid-encoded viral ALAS (vALAS) that successfully recovered the growth of an ALA-auxotrophic *E. coli* C5 pathway mutant. Furthermore, the specific enzymatic formation of ALA by vALAS in complemented cell lysate was verified marking the first time a viral ALAS is enzymatically active. Further, the remaining freshwater-derived AMGs were analyzed for their activity. The vHo1 showcases heme conversion to biliverdin IXα, verifying its role as an active heme oxygenase. Surprisingly, vPcyA produces phycocyanobilin (PCB), the chromophore that forms a functional phytochrome, thus verifying it as an active PcyA enzyme. This study elucidates the distribution and roles of

these AMGs across various ecological niches, enhancing the understanding of phage-driven metabolic processes.

6. Zusammenfassung

Bakteriophagen sind wichtige Bestandteile mikrobieller Ökosysteme. Sie enthalten oft zusätzliche Stoffwechselgene (AMGs), die den Stoffwechsel des Wirts während der Infektion beeinflussen. Diese AMGs sind in Phagengenomen und metagenomischen Sequenzen, die von Phagen stammen, weit verbreitet, aber ihre genaue Rolle bleibt ungeklärt. In dieser Studie wurden AMGs aus metagenomischen Datensätzen unterschiedlicher Herkunft untersucht, die mit dem Tetrapyrrol-Stoffwechsel zusammenhängen, einem bedeutenden Stoffwechselweg für die Häm- und Pigmentbiosynthese. Im Rahmen dieser Studie identifizierte die bioinformatische Analyse viraler Metagenomdaten eine Drei-Gen-Kassette, bestehend aus *ho1*-, *pcyA/X*- und *alaS*-Genen. Während die viralen *ho1*- und *pcyX*-Gene bereits davor in marinen Datensätzen gefunden wurden und für eine Häm-Oxygenase (Ho1) und eine Ferredoxin-abhängige Bilin-Reduktase (FDBR) für die lineare Tetrapyrrol-Biosynthese kodieren, ist die Identifizierung von *alaS* interessant. *alaS* kodiert für eine mutmaßliche Aminolävulinsäure-Synthase (ALAS), ein Enzym, das den ersten Schritt der Tetrapyrrol-Biosynthese in Alpha-Proteobakterien katalysiert, um Aminolävulinsäure (ALA) zu produzieren. Dieser sogenannte C4-Weg steht im Gegensatz zum C5-Weg der ALA-Bildung, der in allen anderen Bakterien und Archaeen vorhanden ist. Dies ist der erste Nachweis eines von einem Phagen kodierten *alaS* in einer Kassette mit Tetrapyrrolbiosynthese assoziierten AMGs (*ho1*, *pcyA/X*), was auf eine mögliche Rolle bei der Erhöhung des Stoffwechselflusses durch den Tetrapyrrolweg schließen lässt. Die Kassette wurde in verschiedenen metagenomischen Datensätzen aus Meeres- und Süßwasserhabitaten entdeckt, aber in Süßwasser war *alaS* nicht direkt mit dieser Kassette assoziiert und konnte weiter entfernt auf dem Kontig gefunden werden. Durch eine Kombination aus *In-silico*-Strukturvorhersage und Aminosäuresequenz-Alignment erweist sich das virale *alaS*-Gen als funktionelles C4-Enzym, das Glycin und Succinyl-CoA für die ALA-Synthese verwendet. Dies wurde durch eine funktionelle Komplementierung mit plasmidkodiertem viralen ALAS (vALAS) bewiesen, die erfolgreich das Wachstum einer ALA-auxotrophen *E. coli* C5-Mutante wiederherstellte. Darüber hinaus wurde die spezifische enzymatische Bildung von ALA durch vALAS in komplementiertem Zelllysat nachgewiesen, womit zum ersten Mal gezeigt wurde, dass ein virales ALAS enzymatisch

aktiv ist. Außerdem wurden die übrigen aus Süßwasser stammenden AMGs auf ihre Aktivität hin untersucht. Die virale Ho1 zeigt eine Umwandlung von Häm in Biliverdin IXα, was seine Rolle als aktive Häm-Oxygenase bestätigt. Überraschenderweise produziert das virale PcyA Phycocyanobilin (PCB), den Chromophor für die Bildung eines funktionsfähigen Phytochroms, und bestätigt damit seine Rolle als aktives PcyA-Enzym. Diese Studie gibt Aufschluss über die Verteilung und die Rolle dieser AMGs in verschiedenen ökologischen Nischen und verbessert das Verständnis der von Phagen gesteuerten Stoffwechselprozesse.

References

Adhikari ND, Orler R, Chory J, Froehlich JE, Larkin RM. 2009. Porphyrins promote the association of GENOMES UNCOUPLED 4 and a Mg-chelatase subunit with chloroplast membranes. J Biol Chem **284**(37):24783-24796.

Adhikari ND, Froehlich JE, Strand DD, Buck SM, Kramer DM, Larkin RM. 2011. GUN4-porphyrin complexes bind the ChlH/GUN5 subunit of Mg-chelatase and promote chlorophyll biosynthesis in *Arabidopsis*. Plant Cell **23**(4):1449-1467.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol **215**:403-410.

Andel F, Hasson K, Gai F, Anfinrud PA, Mathies RA. 1997. Femtosecond time-resolved spectroscopy of the primary photochemistry of phytochrome. Biospectros **3**(6):421-433.

Anzaldi LL, Skaar EP. 2010. Overcoming the heme paradox: heme toxicity and tolerance in bacterial pathogens. Infect Immun **78**(12):4977-4989.

Astner I, Schulze JO, van den Heuvel J, Jahn D, Schubert W-D, Heinz DW. 2005. Crystal structure of 5-aminolevulinate synthase, the first enzyme of heme biosynthesis, and its link to XLSA in humans. EMBO J **24**:3166–3177.

Banerjee R, Ragsdale SW. 2003. The many faces of vitamin B12: catalysis by cobalamin-dependent enzymes. Annu Rev Biochem **72**:209-247.

Battersby AR. 2000. Tetrapyrroles: the pigments of life. Nat Prod Rep **17**:507-526.

Beale S, Castelfranco P. 1973. ^{14}C incorporation from exogenous compounds into delta-aminolevulinic acid by greening cucumber cotyledons. Biochem Biophys Res Commun **52**:143-149.

Beale S. 1990. Biosynthesis of the tetrapyrrole pigment precursor, aminolevulinic acid, from glutamate. Plant Physiol **93**:1273-1279.

Beale SI, Cornejo J. 1991. Biosynthesis of phycobilins. 3 (Z)-phycoerythrobilin and 3 (Z)-phycocyanobilin are intermediates in the formation of 3 (E)-phycocyanobilin from biliverdin IX alpha. J Biol Chem **266**(33):22333-22340.

Beator J, Kloppstech K. 1993. The circadian oscillator coordinates the synthesis of apoproteins and their pigments during chloroplast development. Plant Physiol **103**:191-196.

Béjà O, Fridman S, Glaser F. 2012. Viral clones from the GOS expedition with an unusual photosystem-I gene cassette organization. ISME J **6**(8):1617-1620.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The protein data bank. Nucleic Acids Res **28**(1):235-242.

Besemer J, Borodovsky M. 1999. Heuristic approach to deriving models for gene finding. Nucleic Acids Res **27**:3911-3920.

Bianchetti CM, Yi L, Ragsdale SW, Phillips GN, Jr. 2007. Comparison of apo- and heme-bound crystal structures of a truncated human heme oxygenase-2. J Biol Chem **282**(52):37624-37631.

Blumenstein A, Vienken K, Tasler R, Purschwitz J, Veith D, Frankenberg-Dinkel N, Fischer R. 2005. The *Aspergillus nidulans* phytochrome FphA represses sexual development in red light. Curr Biol **15**(20):1833-1838.

Bolt EL, Kryszak L, Zeilstra-Ryalls J, Shoolingin-Jordan PM, Warren MJ. 1999. Characterization of the *Rhodobacter sphaeroides* 5-aminolaevulinic acid

synthase isoenzymes, HemA and HemT, isolated from recombinant *Escherichia coli*. Eur J Biochem **265**:290-299.

Bonnett R, McDonagh AF. 1973. The meso-reactivity of porphyrins and related compounds. Part VI. Oxidative cleavage of the haem system. The four isomeric biliverdins of the IX series. J Chem Soc Perkin Trans:881-888.

Bottomley SS, May BK, Cox TC, Cotter PD, Bishop DF. 1995. Molecular defects of erythroid 5-aminolevulinate synthase in X-linked sideroblastic anemia. J Bioenerg Biomembr **27**:161-1688.

Brown BL, Kardon JR, Sauer RT, Baker TA. 2018. Structure of the mitochondrial aminolevulinic acid synthase, a key heme biosynthetic enzyme. Structure **26**:580-589.

Bryant DA, Glazer AN, Eiserling FA. 1976. Characterization and structural properties of the major biliproteins of *Anabaena* sp.. Arch Microbiol **110**:61-75.

Bryant DA, Hunter CN, Warren MJ. 2020. Biosynthesis of the modified tetrapyyroles - the pigments of life. J Biol Chem **295**(20):6888-6925.

Brzezowski P, Schlicke H, Richter A, Dent RM, Niyogi KK, Grimm B. 2014. The GUN 4 protein plays a regulatory role in tetrapyrrole biosynthesis and chloroplast-to-nucleus signalling in *Chlamydomonas reinhardtii*. Plant Cell **79**(2):285-298.

Buchfink B, Reuter K, Drost HG. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods **18**(4):366–368.

Buchholz HH, Michelsen M, Parsons RJ, Bates NR, Temperton B. 2021. Draft genome sequences of pelagimyophage Mosig EXVC030M and pelagipodophage Lederberg EXVC029P, isolated from Devil's Hole, Bermuda. Microbiol Resour Announc **10**(7):e01325-20.

Burnham B, Lascelles J. 1963. Control of porphyrin biosynthesis through a negative-feedback mechanism. Studies with preparations of δ-aminolaevulate synthetase and δ-aminolaevulate dehydratase from *Rhodopseudomonas spheroides*. Biochem J **87**(3):462.

Burnham BF. 1970. Aminolevulinic acid synthase (*Rhodopseudomonas spheroides*), p 195-200, Methods in Enzymology, vol 17. Academic Press, Elsevier, Amsterdam, Netherlands.

Busch AWU, Reijerse EJ, Lubitz W, Frankenberg-Dinkel N, Hofmann E. 2011a. Structural and mechanistic insight into the ferredoxin-mediated two-electron reduction of bilins. Biochem J **439**(2):257-264.

Busch AWU, Reijerse EJ, Lubitz W, Hofmann E, Frankenberg-Dinkel N. 2011b. Radical mechanism of cyanophage phycoerythrobilin synthase (PebS). Biochem J **433**(3):469-476.

Butler WL, Norris K, Siegelman H, Hendricks S. 1959. Detection, assay, and preliminary purification of the pigment controlling photoresponsive development of plants. Proc Natl Acad Sci USA **45**(12):1703-1708.

Caignan GA, Deshmukh R, Wilks A, Zeng Y, Huang HW, Moënne-Loccoz P, Bunce RA, Eastman MA, Rivera M. 2002. Oxidation of heme to β- and δ-biliverdin by *Pseudomonas aeruginosa* heme oxygenase as a consequence of an unusual seating of the heme. J Am Chem Soc **124**(50):14879-14892.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics **10**(421).

Camargo AP, Roux S, Schulz F, Babinski M, Xu Y, Hu B, Chain PSG, Nayfach S, Kyrpides NC. 2023. Identification of mobile genetic elements with geNomad. Nat Biotechnol:1303-1312.

Cartwright GE, Deiss A. 1975. Sideroblasts, siderocytes, and sideroblastic anemia. N Engl J Med **292**(4):185-193.

Celis AI, DuBois JL. 2019. Making and breaking heme. Curr Opin Struct Biol **59**:19-28.

Chen LX, Meheust R, Crits-Christoph A, McMahon KD, Nelson TC, Slater GF, Warren LA, Banfield JF. 2020. Large freshwater phages with the potential to augment aerobic methane oxidation. Nat Microbiol **5**(12):1504-1515.

Chen W, Wright L, Lee S, Cosloy S, Russell C. 1996. Expression of glutamyl-tRNA reductase in *Escherichia coli*. Biochim Biophys Acta Gene Struct Expression **1309**(1-2):109-121.

Chim N, Iniguez A, Nguyen TQ, Goulding CW. 2010. Unusual diheme conformation of the heme-degrading protein from *Mycobacterium tuberculosis*. J Mol Biol **395**(3):595-608.

Chu GC, Sönnichsen FD, Ikeda-Saito M, Tomita T, Yoshida T. 1999. The heme complex of HmuO, a bacterial heme degradation enzyme from *Corynebacterium diphtheriae*: structure of the catalytic site. J Biol Chem **274**(35):24490-24496.

Clokie MRJ, Shan J, Bailey S, Jia Y, Krisch HM, West S, Mann NH. 2006. Transcription of a 'photosynthetic' T4-type phage during infection of a marine cyanobacterium. Environ Microbiol **8**:827-835.

Consiglieri E, Xu QZ, Zhao KH, Gärtner W, Losi A. 2020. The first molecular characterisation of blue- and red-light photoreceptors from *Methylobacterium radiotolerans*. Phys Chem Chem Phys **22**:12434-12446.

Cornejo J, Beale S. 1988. Algal heme oxygenase from *Cyanidium caldarium*. Partial purification and fractionation into three required protein components. J Biol Chem **263**(24):11915-11921.

Criscuolo A, Gribaldo S. 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol Biol **10**(210).

Crummett LT, Puxty RJ, Weihe C, Marston MF, Martiny JBH. 2016. The genomic content and context of auxiliary metabolic genes in marine cyanomyoviruses. Virology **499**:219-229.

Culley AI, Lang AS, Suttle CA. 2003. High diversity of unknown picorna-like viruses in the sea. Nature **424**:1054-1057.

Dailey HA, Dailey TA, Gerdes S, Jahn D, Jahn M, O'Brian MR, Warren MJ. 2017. Prokaryotic heme biosynthesis: multiple pathways to a common essential product. Microbiol Mol Biol Rev **81**(1):e00048-16.

Dammeyer T, Bagby SC, Sullivan MB, Chisholm SW, Frankenberg-Dinkel N. 2008a. Efficient phage-mediated pigment biosynthesis in oceanic cyanobacteria. Curr Biol **18**(6):442-448.

Dammeyer T, Hofmann E, Frankenberg-Dinkel N. 2008b. Phycoerythrobilin synthase (PebS) of a marine virus - crystal structure of the biliverdin complex and the substrate-free form. J Biol Chem **283**(41):27547-27554.

Davis SJ, Vener AV, Vierstra RD. 1999. Bacteriophytochromes: phytochrome-like photoreceptors from nonphotosynthetic eubacteria. Science **286**(5449):2517-2520.

De Baar HJ, de Jong JT, Bakker DC, Löscher BM, Veth C, Bathmann U, Smetacek V. 1995. Importance of iron for plankton blooms and carbon dioxide drawdown in the Southern Ocean. Nature **373**(6513):412-415.

De Lorenzo V, Cases I, Herrero M, Timmis KN. 1993. Early and late responses of TOL promoters to pathway inducers: identification of postexponential promoters in *Pseudomonas putida* with *lacZ-tet* bicistronic reporters. J Bacteriol **175**(21):6902-6907.

Diekert G, Weber B, Thauer RK. 1980. Nickel dependence of factor F_{430} content in *Methanobacterium thermoautotrophicum*. Arch Microbiol **127**:273-278.

Donegan RK, Moore CM, Hanna DA, Reddi AR. 2019. Handling heme: the mechanisms underlying the movement of heme within and between cells. Free Radic Biol Med **133**:88-100.

Doron S, Fedida A, Hernandez-Prieto MA, Sabehi G, Karunker I, Stazic D, Feingersch R, Steglich C, Futschik M, Lindell D, Sorek R. 2016. Transcriptome dynamics of a broad host-range cyanophage and its hosts. ISME J **10**(6):1437-55.

Duong HD, Lee JW, Rhee JI. 2014. Generation of reactive oxygen species from 5-aminolevulinic acid and glutamate in cooperation with excited CdSe/ZnS QDs. Proc SPIE Biosensing and Nanomedicine VII **9166**:153-170.

Everse J, Hsia N. 1997. The toxicities of native and modified hemoglobins. Free Radic Biol Med **22**(6):1075-1099.

Ferreira G, Dailey H. 1993. Expression of mammalian 5-aminolevulinate synthase in *Escherichia coli*. Overproduction, purification, and characterization. J Biol Chem **268**(1):584-590.

Filée J, Tétart F, Suttle CA, Krisch H. 2005. Marine T4-type bacteriophages, a ubiquitous component of the dark matter of the biosphere. Proc Natl Acad Sci USA **102**(35):12471-12476.

Formighieri C, Ceol M, Bonente G, Rochaix J-D, Bassi R. 2012. Retrograde signaling and photoprotection in a *gun4* mutant of *Chlamydomonas reinhardtii*. Mol Plant **5**(6):1242-1262.

Forterre P. 2013. The virocell concept and environmental microbiology. ISME J **7**:233-236.

Foyer CH, Lelandais M, Kunert KJ. 1994. Photooxidative stress in plants. Physiol Plant **92**:696-717.

Frankenberg-Dinkel N. 2004. Bacterial heme oxygenases. Antioxid Redox Signal **6**(5):825-834.

Frankenberg N, Mukougawa K, Kohchi T, Lagarias JC. 2001. Functional genomic analysis of the HY2 family of ferredoxin-dependent bilin reductases from oxygenic photosynthetic organisms. Plant Cell **13**(4):965-978.

Frankenberg N, Lagarias JC. 2003. Phycocyanobilin:ferredoxin oxidoreductase of *Anabaena* sp. PCC 7120: biochemical and spectroscopic characterization. J Biol Chem **278**:9219-9226.

Fridman S, Flores-Uribe J, Larom S, Alalouf O, Liran O, Yacoby I, Salama F, Bailleul B, Rappaport F, Ziv T, Sharon I, Cornejo-Castillo FM, Philosof A, Dupont CL, Sanchez P, Acinas SG, Rohwer FL, Lindell D, Beja O. 2017. A myovirus encoding both photosystem I and II proteins enhances cyclic electron flow in infected *Prochlorococcus* cells. Nat Microbiol **2**(10):1350-1357.

Froehlich AC, Noh B, Vierstra RD, Loros J, Dunlap JC. 2005. Genetic and molecular analysis of phytochromes from the filamentous fungus *Neurospora crassa*. Eukaryot Cell **4**(12):2140-2152.

Frydman RB, Feinstein G. 1974. Studies on porphobilinogen deaminase and uroporphyrinogen III cosynthase from human erythrocytes. Biochim Biophys Acta Enzymol **350**(2):358-373.

Gambetta GA, Lagarias JC. 2001. Genetic engineering of phytochrome biosynthesis in bacteria. Proc Natl Acad Sci USA **98**:10566-10571.

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, 3rd, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods **6**(5):343-345.

Gibson K, Neuberger A, Scott J. 1955. The purification and properties of δ-aminolaevulic acid dehydrase. Biochem J **61**(4):618-629.

Gibson K, Laver W, Neuberger A. 1958. Initial stages in the biosynthesis of porphyrins. 2. The formation of delta-aminolaevulic acid from glycine and succinyl-coenzyme A by particles from chicken erythrocytes. Biochem J **70**(1):71-81.

Gill SC, von Hippel PH. 1989. Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem **182**:319-326.

Giraud E, Hannibal L, Fardoux J, Verméglio A, Dreyfus B. 2000. Effect of *Bradyrhizobium* photosynthesis on stem nodulation of *Aeschynomene sensitiva*. Proc Natl Acad Sci USA **97**(26):14795-14800.

Giraud E, Fleischman D. 2004. Nitrogen-fixing symbiosis between photosynthetic bacteria and legumes. Photosynth Res **82**:115-130.

Glazer AN, Hixson CS. 1977. Subunit structure and chromophore composition of rhodophytan phycoerythrins. *Porphyridium cruentum* B-phycoerythrin and b-phycoerythrin. J Biol Chem **252**(1):32-42.

Glazer AN. 1988. Phycobilisomes, p 304-312, Methods in Enzymology, vol 167. Academic Press, Elsevier, Amsterdam, Netherlands.

Grant SGN, Jessee J, Bloom FR, Hanahan D. 1990. Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. Proc Natl Acad Sci USA **87**:4645-4649.

Grombein S, Rüdiger W, Zimmermann H. 1975. The structures of the phytochrome chromophore in both photoreversible forms. Biol Chem **356**(2):1709-1714.

Grote A, Hiller K, Scheer M, Munch R, Nortemann B, Hempel DC, Jahn D. 2005. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic Acids Res **33**:526-531.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol **59**(3):307-321.

Hagiwara Y, Sugishima M, Takahashi Y, Fukuyama K. 2006. Crystal structure of phycocyanobilin: ferredoxin oxidoreductase in complex with biliverdin IXα, a key enzyme in the biosynthesis of phycocyanobilin. Proc Natl Acad Sci USA **103**(1):27-32.

Hamza I, Chauhan S, Hassett R, O'Brian MR. 1998. The bacterial Irr protein is required for coordination of heme biosynthesis with iron availability. J Biol Chem **273**:21669-21674.

Hamza I, Dailey HA. 2012. One ring to rule them all: trafficking of heme and heme synthesis intermediates in the metazoans. Biochim Biophys Acta Mol Cell Res **1823**(9):1617-1632.

Hännig AL. 2011. Aminolevulinic acid synthase of *Rhodobacter capsulatus*. Doctorate, Technical University of Braunschweig.

Hirotsu S, Chu GC, Unno M, Lee D-S, Yoshida T, Park SY, Shiro Y, Ikeda-Saito M. 2004. The crystal structures of the ferric and ferrous forms of the heme complex of HmuO, a heme oxygenase of *Corynebacterium diphtheriae*. J Biol Chem **279**(12):11937-11947.

Hohmann-Marriott MF, Blankenship RE. 2011. Evolution of photosynthesis. Annu Rev Plant Biol **62**:515-548.

Hörtensteiner S, Kräutler B. 2011. Chlorophyll breakdown in higher plants. Biochim Biophys Acta Bioenerg **1807**(8):977-988.

Howard-Varona C, Lindback MM, Bastien GE, Solonenko N, Zayed AA, Jang H, Andreopoulos B, Brewer HM, Glavina del Rio T, Adkins JN, Paul S, Sullivan MB, Duhaime MB. 2020. Phage-specific metabolic reprogramming of virocells. ISME J **14**(4):881-895.

Hu JH, Chang JW, Xu T, Wang J, Wang X, Lin R, Duanmu D, Liu L. 2021. Structural basis of bilin binding by the chlorophyll biosynthesis regulator GUN4. Protein Sci **30**:2083-2091.

Hughes J, Lamparter T. 1997. A prokaryotic phytochrome. Nature **386**(6626).

Hughes J, Lamparter T. 1999. Prokaryotes and phytochrome. The connection to chromophores and signaling. Plant Physiol **121**:1059-1068.

Hunter GA, Ferreira GC. 1995. A continuous spectrophotometric assay for 5-aminolevulinate synthase that utilizes substrate cycling. Anal Biochem **226**:221-224.

Hunter GA, Ferreira GC. 1999. Pre-steady-state reaction of 5-aminolevulinate synthase: evidence for a rate-determining product release. J Biol Chem **274**(18):12222-12228.

Hunter GA, Ferreira GC. 2005. 5-aminolevulinate synthase: catalysis of the first step of heme biosynthesis. Cell Mol Biol **55**(1):102-110.

Hunter GA, Zhang J, Ferreira GC. 2007. Transient kinetic studies support refinements to the chemical and kinetic mechanisms of aminolevulinate synthase. J Biol Chem **282**:23025-23035.

Hwang HW, Lee JR, Chou KY, Suen CS, Hwang MJ, Chen C, Shieh RC, Chau LY. 2009. Oligomerization is crucial for the stability and function of heme oxygenase-1 in the endoplasmic reticulum. J Biol Chem **284**(34):22672-22679.

Ikeuchi M, Ishizuka T. 2008. Cyanobacteriochromes: a new superfamily of tetrapyrrole-binding photoreceptors in cyanobacteria. Photochem Photobiol Sci **7**(10):1159-1167.

Ikushiro H, Nagami A, Takai T, Sawai T, Shimeno Y, Hori H, Miyahara I, Kamiya N, Yano T. 2018. Heme-dependent inactivation of 5-aminolevulinate synthase from *Caulobacter crescentus*. Scientific Rep **8**(142288).

Ilag LL, Jahn D. 1992. Activity and spectroscopic properties of the *Escherichia coli* glutamate 1-semialdehyde aminotransferase and the putative active site mutant K265R. Biochemistry **31**:7143-7151.

Imhoff JF. 2006. The phototrophic alpha-proteobacteria, p 41-64, The Prokaryotes, vol 5. Springer New York, New York, United States of America.

Ishizuka T, Narikawa R, Kohchi T, Katayama M, Ikeuchi M. 2007. Cyanobacteriochrome TePixJ of *Thermosynechococcus elongatus* harbors phycoviolobilin as a chromophore. Plant Cell Physiol **48**(9):1385-1390.

Jacobson TB, Callaghan MM, Amador-Noguez D. 2021. Hostile takeover: how viruses reprogram prokaryotic metabolism. Annu Rev Microbiol **75**:515-539.

Jaubert M, Zappa S, Fardoux J, Adriano J-M, Hannibal L, Elsen S, Lavergne J, Verméglio A, Giraud E, Pignol D. 2004. Light and redox control of photosynthesis gene expression in *Bradyrhizobium*: dual roles of two PpsR. J Biol Chem **279**(43):44407-44416.

Jaubert M, Lavergne J, Fardoux J, Hannibal L, Vuillet L, Adriano JM, Bouyer P, Pignol D, Giraud E, Verméglio A. 2007. A singular bacteriophytochrome acquired by lateral gene transfer. J Biol Chem **282**:7320-7328.

Jaubert M, Hannibal L, Fardoux J, Giraud E, Verméglio A. 2009. Identification of novel genes putatively involved in the photosystem synthesis of *Bradyrhizobium* sp. ORS 278. Photosynth Res **100**:97-105.

Jiang Z, Swem LR, Rushing BG, Devanathan S, Tollin G, Bauer CE. 1999. Bacterial photoreceptor with similarity to photoactive yellow protein and plant phytochromes. Science **285**(5426):406-409.

Johnson BH, Hecht MH. 1994. Recombinant proteins can be isolated from *E. coli* cells by repeated cycles of freezing and thawing. Bio/Technol **12**:1357-1360.

Jones AM, Elliott T. 2010. A purified mutant HemA protein from *Salmonella enterica* serovar *typhimurium* lacks bound heme and is defective for heme-mediated regulation *in vivo*. FEMS Microbiol Lett **307**:41-47.

Jordan PM, Laghai-Newton A. 1986. Purification of 5-aminolevulinate synthase. Methods Enzymol **123**:435-443.

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. Nature **596**(7873):583-589.

Kang Z, Wang Y, Gu P, Wang Q, Qi Q. 2011a. Engineering *Escherichia coli* for efficient production of 5-aminolevulinic acid from glucose. Metab Eng **13**:492-498.

Kang Z, Wang Y, Wang Q, Qi Q. 2011b. Metabolic engineering to improve 5-aminolevulinic acid production. Bioeng Bugs **2**(6):342-345.

Kannangara CG, Gough SP. 1978. Biosynthesis of delta-aminolevulinate in greening barley leaves: glutamate 1-semialdehyde aminotransferase. Carlsb Res Commun **43**:185-194.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol **30**:772-780.

Kaufholz AL, Hunter GA, Ferreira GC, Lendrihas T, Hering V, Layer G, Jahn M, Jahn D. 2013. Aminolaevulinic acid synthase of *Rhodobacter capsulatus*: high-resolution kinetic investigation of the structural basis for substrate binding and catalysis. Biochem J **451**:205-216.

Kehoe DM, Grossman AR. 1996. Similarity of a chromatic adaptation sensor to phytochrome and ethylene receptors. Science **273**(5280):1409-1412.

Kehoe DM. 2010. Chromatic adaptation and the evolution of light color sensing in cyanobacteria. Proc Natl Acad Sci USA **107**:9029-9030.

Kieft K, Zhou Z, Anantharaman K. 2020. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. Microbiome J **8**(90).

Kohchi T, Mukougawa K, Frankenberg N, Masuda M, Yokota A, Lagarias JC. 2001. The *Arabidopsis HY2* gene encodes phytochromobilin synthase, a ferredoxin-dependent biliverdin reductase. Plant Cell **13**(2):425-436.

Koontz L. 2014. TCA precipitation. Methods Enzymol **541**:3-10.

Kořený L, Oborník M, Horáková E, Waller RF, Lukeš J. 2022. The convoluted history of haem biosynthesis. Biol Rev **97**:141-162.

Kumar S, Bandyopadhyay U. 2005. Free heme toxicity and its detoxification systems in human. Toxicol Lett **157**(3):175-188.

Kwon SJ, De Boer AL, Petri R, Schmidt-Dannert C. 2003. High-level production of porphyrins in metabolically engineered *Escherichia coli*: systematic extension of a pathway assembled from overexpressed genes involved in heme biosynthesis. Appl Environ Microbiol **69**:4875-4883.

Ladha J, Pareek R, Becker M. 1992. Stem-nodulating legume-*Rhizobium* symbiosis and its agronomic use in lowland rice. Adv Soil Sci **20**:147-192.

Lagarias JC, Rapoport H. 1980. Chromopeptides from phytochrome. The structure and linkage of the P_r form of the phytochrome chromophore. J Am Chem Soc **102**(14):4821-4828.

Lamparter T, Esteban B, Hughes J. 2001. Phytochrome Cph1 from the cyanobacterium *Synechocystis* PCC6803 purification, assembly, and quaternary structure. Eur J Biochem **268**:4720-4730.

Lascelles J. 1960. The synthesis of enzymes concerned in bacteriochlorophyll formation in growing cultures of *Rhodopseudomonas spheroides*. J Gen Microbiol **23**:487-498.

Laurenceau R, Raho N, Forget M, Arellano AA, Chisholm SW. 2021. Frequency of mispackaging of *Prochlorococcus* DNA by cyanophage. ISME J **15**:129-140.

Ledermann B, Béjà O, Frankenberg-Dinkel N. 2016. New biosynthetic pathway for pink pigments from uncultured oceanic viruses. Environ Microbiol **18**:4337-4347.

Ledermann B, Schwan M, Sommerkamp JA, Hofmann E, Béjà O, Frankenberg-Dinkel N. 2018. Evolution and molecular mechanism of four-electron reducing ferredoxin-dependent bilin reductases from oceanic phages. FEBS J **285**:339-356.

Lefort V, Longueville JE, Gascuel O. 2017. SMS: smart model selection in PhyML. Mol Biol Evol **34**:2422-2424.

Lemoine F, Domelevo Entfellner JB, Wilkinson E, Correia D, Dávila Felipe M, De Oliveira T, Gascuel O. 2018. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. Nature **556**(7702):452-456.

Lemoine F, Correia D, Lefort V, Doppelt-Azeroual O, Mareuil F, Cohen-Boulakia S, Gascuel O. 2019. NGPhylogeny.fr: new generation phylogenetic services for non-specialists. Nucleic Acids Res **47**:260-265.

Letunic I, Bork P. 2021. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res **49**:293-296.

Levicán G, Katz A, Valenzuela P, Söll D, Orellana O. 2005. A $tRNA_{Glu}$ that uncouples protein and tetrapyrrole biosynthesis. FEBS Lett **579**:6383-6387.

Levicán G, Katz A, De Armas M, Núñez H, Orellana O. 2007. Regulation of a glutamyl-tRNA synthetase by the heme status. Proc Natl Acad Sci USA **104**(9):3135-3140.

Levin EY, Coleman D. 1967. The enzymatic conversion of porphobilinogen to uroporphyrinogen catalyzed by extracts of hematopoietic mouse spleen. J Biol Chem **242**(18):4248-4253.

Li L, Murphy JT, Lagarias JC. 1995. Continuous fluorescence assay of phytochrome assembly *in vitro*. Biochemistry **34**(24):7923-7930.

Lightning LK, Huang HW, Moënne-Loccoz P, Loehr TM, Schuller DJ, Poulos TL, Ortiz De Montellano PR. 2001. Disruption of an active site hydrogen bond converts human heme oxygenase-1 into a peroxidase. J Biol Chem **276**:10612-10619.

Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW. 2005. Photosynthesis genes in marine viruses yield proteins during host infection. Nature **438**(7064):86-89.

Liu Y, Ortiz de Montellano PR. 2000. Reaction intermediates and single turnover rate constants for the oxidation of heme by human heme oxygenase-1. J Biol Chem **275**(8):5297-5307.

Loureiro MdF, James E, Sprent J, Franco A. 1995. Stem and root nodules on the tropical wetland legume *Aeschynomene fluminensis*. New Phytol **130**(4):531-544.

Lozano Terol G, Gallego-Jara J, Martínez RAS, Vivancos AM, Díaz MC, Puente TdD. 2021. Impact of the expression system on recombinant protein production in *Escherichia coli* BL21. Front Microbiol **12**(682001).

Luo XQ, Wang P, Li JL, Ahmad M, Duan L, Yin LZ, Deng QQ, Fang BZ, Li SH, Li WJ. 2022. Viral community-wide auxiliary metabolic genes differ by lifestyles, habitats, and hosts. Microbiome J **10**(190).

Lyles KV, Eichenbaum Z. 2018. From host heme to iron: the expanding spectrum of heme degrading enzymes used by pathogenic bacteria. Front Cell Infect Microbiol **8**(198).

Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, Madhusoodanan N, Kolesnikov A, Lopez R. 2022. Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Res **50**:276-279.

Mandalari C, Losi A, Gaertner W. 2013. Distance-tree analysis, distribution and co-presence of bilin-and flavin-binding prokaryotic photoreceptors for visible light. Photochem Photobiol Sci **12**(7):1144-1157.

Maniatis T, Fritsch EF, Sambrook J. 1987. Molecular cloning: a laboratory manual, vol 1. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, United States of America.

Mann NH. 2003. Phages of the marine cyanobacterial picophytoplankton. FEMS Microbiol Rev **27**(1):17-34.

Martin JH, Fitzwater SE. 1988. Iron deficiency limits phytoplankton growth in the north-east Pacific subarctic. Nature **331**(6154):341-343.

Martin JH, Coale K, Johnson K, Fitzwater S, Gordon R, Tanner S, Hunter C, Elrod V, Nowicki J, Coley T. 1994. Testing the iron hypothesis in ecosystems of the equatorial Pacific Ocean. Nature **371**(6493):123-129.

Marver HS, Tschudy DP, Perlroth MG, Collins A, Hunter G. 1966. The determination of aminoketones in biological fluids. Anal Biochem **14**:53-60.

Mauzerall D, Granick S. 1956. The occurrence and determination of δ-aminolevulinic acid and porphobilinogen in urine. J Biol Chem **219**(1):435-446.

May BK, Dogra SC, Sadlon TJ, Bhasker CR, Cox TC, Bottomley SS. 1995. Molecular regulation of heme biosynthesis in higher vertebrates. Prog Nucleic Acid Res Mol Biol **51**:1-51.

McCormac AC, Fischer A, Kumar AM, Söll D, Terry MJ. 2001. Regulation of *HEMA1* expression by phytochrome and a plastid signal during de-etiolation in *Arabidopsis thaliana*. Plant J **25**:549-561.

Millard AD, Zwirglmaier K, Downey MJ, Mann NH, Scanlan DJ. 2009. Comparative genomics of marine cyanomyoviruses reveals the widespread occurrence of *Synechococcus* host genes localized to a hyperplastic region: implications for mechanisms of cyanophage evolution. Environ Microbiol **11**:2370-2387.

Mirdita M, Schutze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. 2022. ColabFold: making protein folding accessible to all. Nat Methods **19**(6):679-682.

Moraes CC, Kalil SJ. 2009. Strategy for a protein purification design using C-phycocyanin extract. Bioresour technol **100**(21):5312-5317.

Moser J, Lorenz S, Hubschwerlen C, Rompf A, Jahn D. 1999. *Methanopyrus kandleri* glutamyl-tRNA reductase. J Biol Chem **274**:30679-30685.

Mukherjee S, Stamatis D, Li CT, Ovchinnikova G, Bertsch J, Sundaramurthi JC, Kandimalla M, Nicolopoulos PA, Favognano A, Chen IMA, Kyrpides NC, Reddy TBK. 2023. Twenty-five years of Genomes OnLine Database (GOLD): data updates and new features in v.9. Nucleic Acids Res **51**:957-963.

Muramoto T, Kohchi T, Yokota A, Hwang I, Goodman HM. 1999. The *Arabidopsis* photomorphogenic mutant *hy1* is deficient in phytochrome chromophore biosynthesis as a result of a mutation in a plastid heme oxygenase. Plant Cell **11**(3):335-347.

Murphy JT, Lagarias JC. 1997. The phytofluors: a new class of fluorescent protein probes. Curr Biol **7**:870-876.

Murphy MJ, Siegel LM. 1973. Siroheme and sirohydrochlorin. J Biol Chem **248**:6911-6919.

Murphy MJ, Siegel LM, Kamin H, DerVartanian DV, Lee JP, LeGall J, Peck HD. 1973. An iron tetrahydroporphyrin prosthetic group common to both assimilatory and dissimilatory sulfite reductases. Biochem Biophys Res Commun **54**:82-88.

Murphy MJ, Siegel LM, Tove SR, Kamin H. 1974. Siroheme: a new prosthetic group participating in six electron reduction reactions catalyzed by both sulfite and nitrite reductases. Proc Natl Acad Sci USA **71**:612-616.

Nagano S. 2016. From photon to signal in phytochromes: similarities and differences between prokaryotic and plant phytochromes. J Plant Res **129**:123-135.

Nakahigashi K, Nishimura K, Miyamoto K, Inokuchi H. 1991. Photosensitivity of a protoporphyrin-accumulating, light-sensitive mutant (*visA*) of *Escherichia coli* K-12. Proc Natl Acad Sci USA **88**:10520-10524.

Nambu S, Matsui T, Goulding CW, Takahashi S, Ikeda-Saito M. 2013. A new way to degrade heme: the *Mycobacterium tuberculosis* enzyme MhuD catalyzes heme degradation without generating CO. J Biol Chem **288**:10101-10109.

Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Krypides NC. 2021. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol **39**:578-585.

Neidle EL, Kaplan S. 1993. Expression of the *Rhodobacter sphaeroides hemA* and *hemT* genes, encoding two 5-aminolevulinic acid synthase isozymes. J Bacteriol **175**:2292-2303.

Neuberger A, Sandy JD, Tait GH. 1973. Control of 5-aminolaevulinate synthetase activity in *Rhodopseudomonas spheroides* the purification and properties of an endogenous activator of the enzyme. Biochem J **136**(3):491-499.

Nies M, Wehrmeyer W. 1980. Isolation and biliprotein characterization of phycobilisomes from the thermophilic cyanobacterium *Mastigocladus laminosus* Cohn. Planta **150**:330-337.

Novy R. 2001. Overcoming the codon bias of *E. coli* for enhanced protein expression. inNovations **12**:1-3.

O'Brian MR. 2015. Perception and homeostatic control of iron in the rhizobia and related bacteria. Annu Rev Microbiol **69**(1):229-245.

Oberortner E, Cheng JF, Hillson NJ, Deutsch S. 2017. Streamlining the design-to-build transition with build-optimization software tools. ACS Synth Biol **6**:485-496.

Paiva-Silva GO, Cruz-Oliveira C, Nakayasu ES, Maya-Monteiro CM, Dunkov BC, Masuda H, Almeida IC, Oliveira PL. 2006. A heme-degradation pathway in a blood-sucking insect. Proc Natl Acad Sci USA **103**(21):8030-8035.

Petříček M, Petříčková K, Havlíček L, Felsberg J. 2006. Occurrence of two 5-aminolevulinate biosynthetic pathways in *Streptomyces nodosus* subsp. *asukaensis* is linked with the production of asukamycin. J Bacteriol **188**:5113-5123.

Pham VH, Kim J. 2012. Cultivation of unculturable soil bacteria. Trends Biotechnol **30**(9):475-484.

Philosof A, Yutin N, Flores-Uribe J, Sharon I, Koonin EV, Béjà O. 2017. Novel abundant oceanic viruses of uncultured marine group II Euryarchaeota. Curr Biol **27**:1362-1368.

Puxty RJ, Millard AD, Evans DJ, Scanlan DJ. 2015. Shedding new light on viral photosynthesis. Photosynth Res **126**:71-97.

Rastrojo A, Alcamí A. 2017. Aquatic viral metagenomics: lights and shadows. Virus Res **239**:87-96.

Ratliff M, Zhu W, Deshmukh R, Wilks A, Stojiljkovic I. 2001. Homologues of neisserial heme oxygenase in gram-negative bacteria: degradation of heme by the product of the *pigA* gene of *Pseudomonas aeruginosa*. J Bacteriol **183**:6394-6403.

Rhie GE, Beale SI. 1992. Biosynthesis of phycobilins. Ferredoxin-supported NADPH-independent heme oxygenase and phycobilin-forming activities from *Cyanidium caldarium*. J Biol Chem **267**(23):16088-16093.

Richard KL, Kelley BR, Johnson JG. 2019. Heme uptake and utilization by gram-negative bacterial pathogens. Front Cell Infect Microbiol **9**(81).

Rieble S, Beale SI. 1988. Transformation of glutamate to δ-aminolevulinic acid by soluble extracts of *Synechocystis* sp. PCC 6803 and other oxygenic prokaryotes. J Biol Chem **263**:8864-8871.

Rihtman B, Bowman-Grahl S, Millard A, Corrigan RM, Clokie MRJ, Scanlan DJ. 2019. Cyanophage MazG is a pyrophosphohydrolase but unable to hydrolyse magic spot nucleotides. Environ Microbiol Rep **11**:448-455.

Rippka R, Deruelles J, Waterbury NB. 1979. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. J Gen Microbiol **111**:1-61.

Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res **42**:320–324.

Rockwell NC, Martin SS, Feoktistova K, Lagarias JC. 2011. Diverse two-cysteine photocycles in phytochromes and cyanobacteriochromes. Proc Natl Acad Sci USA **108**(29):11854-11859.

Rockwell NC, Martin SS, Gulevich AG, Lagarias JC. 2012. Phycoviolobilin formation and spectral tuning in the DXCF cyanobacteriochrome subfamily. Biochemistry **51**:1449-1463.

Rockwell NC, Duanmu D, Martin SS, Bachy C, Price DC, Bhattacharya D, Worden AZ, Lagarias JC. 2014. Eukaryotic algal phytochromes span the visible spectrum. Proc Natl Acad Sci USA **111**(10):3871-3876.

Rockwell NC, Martin SS, Lagarias JC. 2023. Elucidating the origins of phycocyanobilin biosynthesis and phycobiliproteins. Proc Natl Acad Sci USA **120**(17):e2300770120.

Rodionov DA, Gelfand MS, Todd JD, Curson ARJ, Johnston AWB. 2006. Computational reconstruction of iron- and manganese-responsive transcriptional networks in α-proteobacteria. PLoS Comput Biol **2**:1568-1585.

Rodriguez-Brito B, Li L, Wegley L, Furlan M, Angly F, Breitbart M, Buchanan J, Desnues C, Dinsdale E, Edwards R. 2010. Viral and microbial community dynamics in four aquatic environments. ISME J **4**(6):739-751.

Roitman S, Hornung E, Flores-Uribe J, Sharon I, Feussner I, Béjà O. 2018. Cyanophage-encoded lipid desaturases: oceanic distribution, diversity and function. ISME J **12**:343-355.

Roitman S. 2022. Expanding the scope of the virocell: viral-host interplay in the marine environment. Doctorate, Technion-Israel Institute of Technology.

Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, Poulos BT, Solonenko N, Lara E, Poulain J, Pesant S, Kandels-Lewis S, Dimier C, Picheral M, Searson S, Cruaud C, Alberti A, Duarte CM, Gasol JM, Vaque D, Tara Oceans C, Bork P, Acinas SG, Wincker P, Sullivan MB. 2016. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. Nature **537**(7622):689-693.

Sandy JD, Davies RC, Neuberger A. 1975. Control of 5-aminolaevulinate synthetase activity in *Rhodopseudomonas spheroides* a role for trisulphides. Biochem J **150**:245-257.

Schacter BA, Nelson EB, Marver HS, Masters BSS. 1972. Immunochemical evidence for an association of heme oxygenase with the microsomal electron transport system. J Biol Chem **247**(11):3601-3607.

Schmitt MP. 1997. Utilization of host iron sources by *Corynebacterium diphtheriae*: identification of a gene whose product is homologous to eukaryotic heme oxygenases and is required for acquisition of iron from heme and hemoglobin. J Bacteriol **179**:838-845.

Schneider T, Tan Y, Li H, Fisher JS, Zhang D. 2022. Photoglobin, a distinct family of non-heme binding globins, defines a potential photosensor in prokaryotic signal transduction systems. Comput Struct Biotechnol J **20**:261-273.

Scholnick PL, Hammaker LE, Marver HS. 1972. Soluble δ-aminolevulinic acid synthetase of rat liver. J Biol Chem **247**(13):4132-4137.

Schön A, Krupp G, Gought S, Berry-Lowet S, Kannangarat CG, Soll D. 1986. The RNA required in the first step of chlorophyll biosynthesis is a chloroplast glutamate tRNA. Nature **322**:281-284.

Schrödinger LLC. 2020. The PyMOL Molecular Graphics System, Version 2.0.

Schuller DJ, Wilks A, Ortiz de Montellano PR, Poulos TL. 1999. Crystal structure of human heme oxygenase-1. Nat Struct Biol **6**(9):860-867.

Schuller DJ, Zhu W, Stojiljkovic I, Wilks A, Poulos TL. 2001. Crystal structure of heme oxygenase from the gram-negative pathogen *Neisseria meningitidis* and a comparison with mammalian heme oxygenase-1. Biochemistry **40**(38):11552-11558.

Schulze JO, Schubert WD, Moser J, Jahn D, Heinz DW. 2006. Evolutionary relationship between initial enzymes of tetrapyrrole biosynthesis. J Mol Biol **358**(5):1212-1220.

Sehnal D, Bittrich S, Deshpande M, Svobodová R, Berka K, Bazgier V, Velankar S, Burley SK, Koča J, Rose AS. 2021. Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. Nucleic Acids Res **49**(1):431-437.

Shan J, Jia Y, Clokie MR, Mann NH. 2008. Infection by the 'photosynthetic' phage S-PM2 induces increased synthesis of phycoerythrin in *Synechococcus* sp. WH7803. FEMS Microbiol Lett **283**(2):154-161.

Sharon I, Alperovitch A, Rohwer F, Haynes M, Glaser F, Atamna-Ismaeel N, Pinter RY, Partensky F, Koonin EV, Wolf YI, Nelson N, Béjà O. 2009. Photosystem I gene cassettes are present in marine virus genomes. Nature **461**:258-262.

Sharrock RA. 2008. The phytochrome red/far-red photoreceptor superfamily. Genome Biol **9**(8):e230.

Shemin D, Rittenberg D. 1945. The utilization of glycine for the synthesis of a porphyrin. J Biol Chem **159**(2):567-568.

Shoolingin-Jordan PM, Al-Daihan S, Alexeev D, Baxter RL, Bottomley SS, Kahari ID, Roy I, Sarwar M, Sawyer L, Wang SF. 2003. 5-Aminolevulinic acid synthase: mechanism, mutations and medicine. Biochim Biophys Acta **1647**:361–366.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol **7**:539.

Simon R, Preifer U, Phuler A. 1983. A broad host range mobilization system for *in vivo* genetic enginering: transposon mutagenesis in gram negative bacteria. Nat Biotechnol **1**:784-791.

Skaar EP, Gaspar AH, Schneewind O. 2004. IsdG and IsdI, heme-degrading enzymes in the cytoplasm of *Staphylococcus aureus*. J Biol Chem **279**(1):436-443.

Slagter HA, Gerringa LJA, Brussaard CD. 2016. Phytoplankton virus production negatively affected by iron limitation. Front Mar Sci **3**(156).

Sommerkamp JA, Frankenberg-Dinkel N, Hofmann E. 2019. Crystal structure of the first eukaryotic bilin reductase GtPEBB reveals a flipped binding mode of dihydrobiliverdin. J Biol Chem **294**(388):13889-13901.

Sonani RR, Singh NK, Kumar J, Thakar D, Madamwar D. 2014. Concurrent purification and antioxidant activity of phycobiliproteins from *Lyngbya* sp. A09DM: an antioxidant and anti-aging potential of phycoerythrin in *Caenorhabditis elegans*. Process Biochem **49**(10):1757-1766.

Steward GF, Culley AI, Mueller JA, Wood-Charlson EM, Belcaid M, Poisson G. 2013. Are we missing half of the viruses in the ocean? ISME J **7**(3):672-679.

Stocker R. 2004. Antioxidant activities of bile pigments. Antioxid Redox Signal **6**(5):841-849.

Studier FW, Moffatt BA. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J Mol Biol **189**:113-130.

Studier FW. 2005. Protein production by auto-induction in high-density shaking cultures. Protein Expr Purif **41**(1):207-234.

Studier FW. 2014. Stable expression clones and auto-induction for protein production in *E. coli*, p 17-32, Structural Genomics: general applications, vol 1091. Humana Press, Totowa, United States of America.

Sugishima M, Omata Y, Kakuta Y, Sakamoto H, Noguchi M, Fukuyama K. 2000. Crystal structure of rat heme oxygenase-1 in complex with heme. FEBS Lett **471**(1):61-66.

Sugishima M, Sakamoto H, Noguchi M, Fukuyama K. 2003. Crystal structures of ferrous and CO-, CN-, and NO-bound forms of rat heme oxygenase-1 (HO-1) in complex with heme: structural implications for discrimination between CO and O_2 in HO-1. Biochemistry **42**:9898-9905.

Sugishima M, Migita CT, Zhang X, Yoshida T, Fukuyama K. 2004. Crystal structure of heme oxygenase-1 from cyanobacterium *Synechocystis* sp. PCC 6803 in complex with heme. Eur J Biochem **271**:4517-4525.

Sugishima M, Hagiwara Y, Zhang X, Yoshida T, Migita CT, Fukuyama K. 2005. Crystal structure of dimeric heme oxygenase-2 from *Synechocystis* sp. PCC 6803 in complex with heme. Biochemistry **44**:4257-4266.

Sugishima M, Wada K, Fukuyama K, Yamamoto K. 2020. Crystal structure of phytochromobilin synthase in complex with biliverdin IXα, a key enzyme in the biosynthesis of phytochrome. J Biol Chem **295**(3):771-782.

Šunderić M, Gligorijević N, Milčić M, Minić S, Nedić O, Nikolić M. 2023. Phycocyanobilin is a new binding partner of human alpha-2-macroglobulin that protects the protein against oxidative stress. J Biomol Struct Dyn:1-11.

Sutherland GR, Baker E, Callen DF, Hyland VJ, Mayt BK, Bawdent MJ, Healyt HM, Borthwickt A. 1988. 5-Aminolevulinate synthase is at 3p21 and thus not the primary defect in X-linked sideroblastic anemia. Am J Hum Genet **43**(3):331-335.

Suttle CA. 2007. Marine viruses - Major players in the global ecosystem. Nat Rev Microbiol **5**:801-812.

Székely AJ, Breitbart M. 2016. Single-stranded DNA phages: from early molecular biology tools to recent revolutions in environmental microbiology. FEMS Microbiol Lett **363**(6):fnw027.

Tara Oceans Consortium Coordinators, Tara Oceans Expedition Participants. 2017. Registry of all samples from the Tara Oceans Expedition (2009-2013). PANGAEA.

Terauchi K, Montgomery BL, Grossman AR, Lagarias JC, Kehoe DM. 2004. RcaE is a complementary chromatic adaptation photoreceptor required for green and red light responsiveness. Mol Microbiol **51**(2):567-577.

Terry MJ, Lagarias JC. 1991. Holophytochrome assembly: coupled assay for phytochromobilin synthase *in organello*. J Biol Chem **266**:22215-22221.

Thauer RK, Bonacker LG. 1994. Biosynthesis of coenzyme F_{430}, a nickel porphinoid involved in methanogenesis, p 210-227, The Biosynthesis of the Tetrapyrrole

Pigments Ciba Foundation Symposium 180. John Wiley & Sons Ltd, Chichester, United Kingdom.

Thoma S, Schobert M. 2009. An improved *Escherichia coli* donor strain for diparental mating. FEMS Microbiol Lett **294**(2):127-132.

Thompson LR, Zeng Q, Kelly L, Huang KH, Singer AU, Stubbe J, Chisholm SW. 2011. Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. Proc Natl Acad Sci USA **108**:757-764.

Tsai SF, Bishop D, Desnick R. 1987. Purification and properties of uroporphyrinogen III synthase from human erythrocytes. J Biol Chem **262**(3):1268-1273.

Tu SL, Gunn A, Toney MD, Britt RD, Lagarias JC. 2004. Biliverdin reduction by cyanobacterial phycocyanobilin:ferredoxin oxidoreductase (PcyA) proceeds via linear tetrapyrrole radical intermediates. J Am Chem Soc **126**:8682-8693.

Tu SL, Rockwell NC, Lagarias JC, Fisher AJ. 2007. Insight into the radical mechanism of phycocyanobilin-ferredoxin oxidoreductase (PcyA) revealed by X-ray crystallography and biochemical measurements. Biochemistry **46**:1484-1494.

Ulijasz AT, Cornilescu G, von Stetten D, Kaminski S, Mroginski MA, Zhang J, Bhaya D, Hildebrandt P, Vierstra RD. 2008. Characterization of two thermostable cyanobacterial phytochromes reveals global movements in the chromophore-binding domain during photoconversion. J Biol Chem **283**(30):21251-21266.

Ulijasz AT, Cornilescu G, von Stetten D, Cornilescu C, Escobar FV, Zhang J, Stankey RJ, Rivera M, Hildebrandt P, Vierstra RD. 2009. Cyanochromes are blue/green light photoreversible photoreceptors defined by a stable double cysteine linkage to a phycoviolobilin-type chromophore. J Biol Chem **284**(43):29757-29772.

Wang J, Ortiz de Montellano PR. 2003. The binding sites on human heme oxygenase-1 for cytochrome P450 reductase and biliverdin reductase. J Biol Chem **278**:20069-20076.

Wang LY, Brown L, Elliott M, Elliott T. 1997. Regulation of heme biosynthesis in *Salmonella typhimurium:* activity of glutamyl-tRNA reductase (HemA) is greatly elevated during heme limitation by a mechanism which increases abundance of the protein. J Bacteriol **179**(9):2907-2914.

Wang W, Malcolm BA. 1999. Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange™ site-directed mutagenesis. Biotechniques **26**(4):680-682.

Wegele R, Tasler R, Zeng Y, Rivera M, Frankeinberg-Dinkel N. 2004. The heme oxygenase(s)-phytochrome system of *Pseudomonas aeruginosa*. J Biol Chem **279**:45791-45802.

Wegner H, Roitman S, Kupczok A, Braun V, Woodhouse JN, Grossart HP, Zehner S, Béjà O, Frankenberg-Dinkel N. 2024. Identification of Shemin pathway genes for tetrapyrrole biosynthesis in bacteriophage sequences from aquatic environments. Nat Commun(accepted).

Weinstein JD, Beale SI. 1983. Separate physiological roles and subcellular compartments for two tetrapyrrole biosynthetic pathways in *Euglena gracilis*. J Biol Chem **258**(11):6799-6807.

Whatley SD, Ducamp S, Gouya L, Grandchamp B, Beaumont C, Badminton MN, Elder GH, Holme SA, Anstey AV, Parker M, Corrigall AV, Meissner PN, Hift RJ, Marsden JT, Ma Y, Mieli-Vergani G, Deybach JC, Puy H. 2008. C-terminal

deletions in the *ALAS2* gene lead to gain of function and cause X-linked dominant protoporphyria without anemia or iron overload. Am J Hum Genet **83**:408-414.

Wikner J, Vallino JJ, Steward GF, Smith DC, Azam F. 1993. Nucleic acids from the host bacterium as a major source of nucleotides for three marine bacteriophages. FEMS Microbiol Ecol **12**:237-248.

Wilde A, Mikolajczyk S, Alawady A, Lokstein H, Grimm B. 2004. The *gun4* gene is essential for cyanobacterial porphyrin metabolism. FEBS Lett **571**(1-3):119-123.

Wilks A, Schmitt MP. 1998. Expression and characterization of a heme oxygenase (HmuO) from *Corynebacterium diphtheriae*: iron acquisition requires oxidative cleavage of the heme macrocycle. J Biol Chem **273**(2):837-841.

Wilks A. 2002. Heme oxygenase: evolution, structure, and mechanism. Antioxid Redox Signal **4**(4):603-614.

Wilks A, Heinzl G. 2014. Heme oxygenation and the widening paradigm of heme degradation. Arch Biochem Biophys **544**:87-95.

Wu R, Smith CA, Buchko GW, Blaby IK, Paez-Espino D, Kyrpides NC, Yoshikuni Y, McDermott JE, Hofmockel KS, Cort JR, Jansson JK. 2022. Structural characterization of a soil viral auxiliary metabolic gene product - a functional chitosanase. Nat Commun **13**:5485.

Yang P, Liu W, Cheng X, Wang J, Wang Q, Qi Q. 2016. A new strategy for production of 5-aminolevulinic acid in recombinant *Corynebacterium glutamicum* with high yield. Appl Environ Microbiol **82**:2709-2717.

Yoshihara S, Katayama M, Geng X, Ikeuchi M. 2004. Cyanobacterial phytochrome-like PixJ1 holoprotein shows novel reversible photoconversion between blue-and green-absorbing forms. Plant Cell Physiol **45**(12):1729-1737.

Zappa S, Li K, Bauer CE. 2010. The tetrapyrrole biosynthetic pathway and its regulation in *Rhodobacter capsulatus*. Adv Exp Med Biol **675**:229-250.

Zeilstra-Ryalls JH, Kaplan S. 1996. Control of *hemA* expression in *Rhodobacter sphaeroides* 2.4.1: regulation through alterations in the cellular redox state. J Bacteriol **178**(4):985-993.

Zhang J, Ferreira GC. 2002. Transient state kinetic investigation of 5-aminolevulinate synthase reaction mechanism. J Biol Chem **277**(47):44660-44669.

Zhang W, Willows RD, Deng R, Li Z, Li M, Wang Y, Guo Y, Shi W, Fan Q, Martin SS, Rockwell NC, Lagarias JC, Duanmu D. 2021. Bilin-dependent regulation of chlorophyll biosynthesis by GUN4. Proc Natl Acad Sci USA **118**(20):e2104443118.

Zhu C, Chen J, Wang Y, Wang L, Guo X, Chen N, Zheng P, Sun J, Ma Y. 2019. Enhancing 5-aminolevulinic acid tolerance and production by engineering the antioxidant defense system of *Escherichia coli*. Biotechnol Bioeng **116**:2018-2028.

Zhu W, Hunt DJ, Richardson AR, Stojiljkovic I. 2000. Use of heme compounds as iron sources by pathogenic *Neisseriae* requires the product of the *hemO* gene. J Bacteriol **182**(2):439-447.

Zhu W, Lomsadze A, Borodovsky M. 2010. *Ab initio* gene identification in metagenomic sequences. Nucleic Acids Res **38**(12):e132.

Appendix

I. Viral amino acid sequences derived from metagenomic data

Viral freshwater sequences

Viral ALAS sequences:

>Crystalbogphagescaffold_2_498

MLKSKTQDAIAQSLQKLKSDGNYRVFTDILRERGHFPKAMFYGKYGIKHITNWCANDYLGMAQHKVVIDAMHTALDTTGAGSGGTRNI
SGTTHFHTALESELARLHDKTAALLFTSGYVANQSTLNVLGKIITGVHFISDSMNHNSMIVGMKSSRAPCTVWNHNDLAHLKEILKSL
DDDAQPIIAMEGVYSMDGDKGLISDVCDIARQYGAMVYVDEVHAVGLYGVRGAGIAEEQHCVDGVDIIQGTLGKAFGIQGGYIAGGRD
LIDMIRSYANGFIFSTSMSPVLCAGALASVRYVQDHPEARSRMFEVVDKVKEQLAQENFNVMPGNTHIVPLIIGNAVRCKQVSDWLLD
EHNIYIQPIGYPTVPIGTERLRITPTPNHTDAMICDLRKSLRNSIDKYI

>TP6_1_phage_scaffold_1_136

MIDKKPFANLLQTLKDNGKYRVFNDILRENGKFPNAIWYGPYNIKNIVNWCSNDYLGMGQHKVVLDAMHTALDMTGAGSGGTRNISGT
THYHVALEYELASLHKKERALLFTGAYVANEWTLISLAKIIPNIHFISDKKNHNSMIVGMSHSRASKTIFEHNDLADLENCLIAAKKA
DHIPCIVFESVYSMDGDVSLIKEICDLADKYDAITYIDEVHAVGLYGDHGAGKLEELNLQGRVDIVNGTLGKAFGVQGGYIAADAVII
DAIRSIAAGFIFTTSMSPVTCAGALAAVKYLKEHNEIREKHQDRARKLKHRLAKAGLPVMPSTTHIVPVLVGDAKRCKAMSDRLLEEF
NIYVQSINAPTVPEGTERLRFAPTPYHTDGMIEDLVNALKIVFEE

>Crystal_Bog_phage_scaffold_1_89

MINKSLFEQTITDLKFAGKYRVFNDILRECGDYPHALWYGPYCIKKIVNWCSNDYLGMGQHKVVLDAMRTALDAVGAGSGGTRNIAGT
THYHVALEHELAGLHSKERALLFSSAYVANEWALIALSKIVKDIVFVSDSKNHNSLIEGMRHSRAPKVIYNHNDMNHLEDVLKEIKGT
PCIVFESVYSMDGYVSNLREICDLADKYGAITYLDEVHAVGLYGETGGGKSEKYGLQKRIDIINGTLGKAYGVGGGYIAADNVVIDSI
RSLANGFIFTTSISPVLCAGAMASVKYLKEHNELRIKHQERASTLKNKLRDAGITLLKNETHIVPVMVNDAKKCKEISDRLLNDYDIY
SQCINYPTVPVGSERLRFAPTPLHTDAMIEDLVNALVEIIKP

>METAG-scaffold_28_108

MIKAKTQDIIAQSLEKLKSEGNYRVFTDILRQRGDFPQAIWYGKYNIKQITNWCSNDYLGMGQHKVVLDAMRTALDTAGAGSGGTRNI
SGTTHYHVALEYELAKLHDKTSALLFTSGYVANQSTLAILGKMIPKVHYISDSENHNSMIVGMKASGCPVTVWRHNDLSHLREILSSL
PIDTQPIIALEGVYSMDGDKGLISDVCEIARLNAAMVYVDEVHAVGLYGPRGAGIAEEQKCVDGVDIIQGTLAKAFGVQGGYIAAGRD
LIDMVRSYAQGFIFSTSMSPVLCAGALASVKYVQDRHDLRQSIFKKAQETREHLLSNGIEVLVNDGHIVPIMIKDAKKCKMISDWLLE
ERAIYLQPINYPTVPWGTERLRATPTPNHTEADIDYLTRSLKEILKI

>METAG-scaffold_37_292

MINKEPFERLIAELKESGKYRVFNDIVRETGKFPNAIWYGPYNIKNITNWCSNDYLGMGQNKVVLEAMHTALDHTGSGSGGTRNIGGT
SHYHVALEHEIASLHKKEKAVLFSSAYVANEWTLIALAKIIPNIEYISDSNNHNSIIVGINHSKANKVIFKHNDLEDLEQKLKISFAQ
GKTPCVVFESVYSMDGDVSPMAEMCKLANKYKAITYIDEVHAVGLYGKTGGGKIEELGLEDKIDIINGTLGKAFGVQGGYIACDKIVA
DAIRSVAAGFIFTTSMSPVTCSGALAAIKWLKDHNEVRDKHQERATLLKTKLKAAGIPVMECSTTHIVPVLVGDAKRAKAMSDALLND
YSIYVQAINYPTVDVGTERLRFAPTPFHDDGMIEDLVNALVQVFSTY

>METAG-scaffold_16_416

MIDKAPFNDLIRTLKDNGKYRVFNDILRENGRFPNAIWYGPYAIKNIVNWCSNDYLGMGQHKVVLDAMRTALDMTGAGSGGTRNISGT
THYHVALEHEIATLHSKERSLLFTSAYVANEWTLIALSKIIPNIHFISDKKNHNSIIIGLRHSGAKRSIFQHNDMEDLEKCLITATAQ
HQTPCIVFESVYSMDGDVSMIKEICELADKYDAITYIDEVHAVGLYGKHGAGKLEELDLQGRVDIVNGTLGKAFGVQGGYIAADAVVI
DAIRSVAAGFIFTTSMSPVTCAGALAAIKYLKEHGELRDKHQERARKLKHCLEKAGLPVMPSTTHIVPILVGDAKKCKEMSDRLLNEF
DLYSQAINFPTVPEGTERLRFAPTPFHDDGMIEDLVNALLIVFDY

>METAG-scaffold_45_17

MLNTKPFEAVIQSLKDQGKYRVFNDIMRERGDFPYAMWYGPYGVKKIVNWCSNDYLGMGQHKVVIDAMRTALDATGAGSGGTRNISGT
SHYHVALELELARLHDKERSLVFSSAYVANEWSLIALKKIIPDIIYVSDSKNHASMIEGMRHSGAKRLIYQHNDLDHLESLLKECYVG
TPCIVFESVYSMDGSVSKMQEICDLADKYEAITYLDEVHAVGLYGEHGGGYSELLGLQDRIDIINGTLGKAFGVTGGYIASDDVVIDA
IRSIASGFIFTTSLSPVICSGALAAVKYLKDHNELRVQHQDRSRKLKFDLLNAGLPVMMNDTHIVPVLVGDATVCKYISSVLIEEKDI
YCQPINYPTVPIGTERLRFAPTPLHTDGMIEELVSALTDIMLTKKL

>METAG-scaffold_21_116

MINKQPFENLIQELKDSGKYRVFNDILRENGKFPAAIWYGPYNIKNIVNWCSNDYLGMGQHKVVLDAMHTALDMTGAGSGGTRNIAGT
SHYHVALEHELATLHKKDKSLLFSSAYVANEWSLIALAKIIPNIHFVSDSKNHNSMIVGMSHSRAPKTVFVHNDLDDLEATLKDVVAA
QQTPCVVFESVYSMDGDVSLIREICDLADKYGAITYLDEVHAVGLYGEHGAGKLEELGLQDRVDIVNGTLGKAYGVQGGYIAADAVIV

DAIRSIAAGFIFTTSMSPVTCSGALAAVKYLKANSELRVQHQDRARKLKYRLAKAGIPVMESTTHIVPVLVGDAKRAKAMSDMLLNEY
NIYVQPINYPTVPVGTERLRFAPTPLHDDGMIEDLIQALTEVFAKV

Viral heme oxygenase sequences:

>CB2_vHo1

MATLKEVTRAVHDEAEHTEWGKLLVSGNITKEQYVHYLYNLLEIHQAIESRGVITMPEILRVKAIISDIEATDNTISPVTLDSTWKYI
AHLKGLTDDLLWCHIYCHYLGYMYGGQMIKKAIPFSTTMLDFDNRAECIAYIRQHLENANHNEAIAAFKWATFMFNDLWNHYKDGK

Viral PcyA/X sequences:

>BML_S_1_PcyX_phage_scaffold_1_120

MTTVWVKLLDLADQIENILSENSTEIYEKGMDRFNQPGWVNRVYANDNYRRAHIDIVDARESKGLWMMHCCIFPQVNNTSPIFGYDVI
AGKEKITGFFHDFSPVCCEDHPMSKHFKETTDTLTWKKERQLPDWALQIFSKGMIAAGNIKAVDEKDFRNMDIMIENLKYYIENVGKY
KSNFDFTKTHNKYSHYQKQNPHTPKAMAALGLSKEDVEIFIKDCLFPEINR

>Crystal_Bog_phage_scaffold_2_70

MTDYIELLDKTARQLNHIISAQPDVIPMETPEYGWENHRHSSPQFRMAHVEIFNQDRFMVVHCCVFPHVSDPAPIFGFDVIASQNKVT
GVFLDLSPTVEEPGKFHNLTFKQQRERPQWGDIFSQNWIACRPDKTEMMSIVAESQRLLKHYLTNIINKRTAPVDAIREGQNRYCKQQ
QQNEHTLKALKNLIGPSRAREFMNTILFPTV

>BML_phage_scaffold_1_182

MSEVWNTLIDIQHLLEDHFKRTGQEIQEPGMERFNRPGWVNRVWTSNLYRRAHIDVIDARESKGLWMMHCCIFPHIHNPAPIYGFDVI
AGRNKITGCFHDFSPAGDKDHPLIQWFSEQVKNLEWKRSRDLPEWAERIFTPYIITASNVKDSYELDQIFDIANRSVAHYLDCVAETN
NTVTDTTEAQNYYAYNQKLNPHNPRVMTNLGLDEEEVRIFVQDCLFPDLR

>Crystal_Bog_phage_scaffold_1_417

MSNVWPRLIEIQDLFIDKLNKTGLEIFEPGMDRFNQPGWVNRVWGSRNYRRAHVDVVDARETKGLWMMHCCIFPHVDNNAPIFGYDVV
AGKNKITGCFHDFSITDDPSHPLIEWFGNEVSKLTWRKKRDLPDWAQRIFSEHMIAAGNVSKEDELDQIFAMIEKNLSHYLECVGETS
GTTFNNAEEQNYYCENQKQNPHTPKVMSSLGLDEEDVRVFIQDCLFPEIV

>GROS21-1_SAMEA9560612_METAG-scaffold_37_347

MSKVWDTLINIQHLLEKEFDRTGTEQFEPGMDRFNQPGWINRVWSSPSYRRAHVDVVDARETKGLWMMHCCIFPHLHNPAPIFGFDVV
AGKNKITGCFIDYSPTEDKFHHMLDYFGEEVSRYEWIKKRELPDWAKRIFSQHMVAAGNVSDESELAQISSLASILINHYLETVGETN
NRVLDTSSYQNYYCDNQKQNPHTPRVMVSLGLSEDDVQHFIQECLFPEIR

>GROS21-1_SAMEA9560623_METAG-scaffold_28_121

MSHIWNKMQDCADVMRDKMLRMGEIDQDPILEKYNWENHVYRSSYFRRGHVEVVDQRENYGLFILHATVFPRVDSDAPIWGFDAVCGK
NKITGAFHDFSLVDTEDNFMYQWFKDTTKDITWKKERELPDWAKAIFSPSMVAVGNVQDELEVDQFIKLGLDTLDYYISNANLCKIVG
KDYTEQQNRYCYYQKQNPHVVRSMVAMGYEQHIIENFVEEVLFPEIK

>GROS21-1_SAMEA9560794_METAG-scaffold_21_203

MQTSNVWDTLINIQELLETNFGRTGTEIFEPGMDRFNQPGWVNRVWTSDLYRRAHVDVVDARDTKGLWMMHCCIFPHLHNDAPIFGFD
VVAGKSKITGCFYDFSPAGDSEHPLCEWFADETSQLEWNKKRKLPEWAERIFSESMVAAGNIQKEEELEQIFQMAKRGVEHYLQAVGE
TNHTASSTKDAQNYYAQNQKQNPHTPKVMVSLGLTEEDVTVFIQECLFPEIK

>GROS21-1_SAMEA9561299_METAG-scaffold_16_12

MSLIFEKIAPLADLIADKFEKISTRVTEPGIEHFDWYNKVYTNYKFRRAHIEVLDRTEKNNMWIMHATVFPHLNDPSPIFGFDIVCTG
KKVSGIFHDFSITTNDEHYMQKGFEGSVRDLSWKKERELPEWGKKIFSKSMIAAGGISSEQELEQIINVCLRNLDYYLANVGTVITLE
PISAVKEKQNYYCKCQKENPYPVRMLINFGLTKEQAENFVNNHLFPEVS

>GROS21-1_SAMEA9560601_METAG-scaffold_16_31

MNDFSNTLLSIQHSLEESFNRTGTEIQEEGMERFNQANWINRVWSSKSYRRAHLNVVDCRESSGLWMMHCCIFPHLHNTAPIYGFDIV
ALRNKITGCFHDFSPTADKQHPMINWFSTESKKLNWNRTRELPDWAKLIFSNDVIAASNVQEEDELNQIVELVKRTTEYYLNNVRQSN
DAIDETIIFQNYYSYNQKQNPHLYTPRIMESLGLDHEGVTIFVEKCLFPEIFNLLH

>GROS21-1_SAMEA9560542_METAG-scaffold_45_32

MSEIWNRLIDIENHYINRFTETGKEILTDDISALGWSSRSWSSDTYRLANIVTADVRETKGMWMMHCCVFPHMENTAPIFGFDVVAGK
NKITGCFHDFSSTGFASHPLIEWFGHEVSQLEWRKTRELPEWAQRIFSPHIIAAANVNEGDELEQIISMSTRNLDHYIDTVGETDGDC
TDNTKGQNYYCENQKLNPHNPRVMTNLGLSEEQIRFFIEQCMFPEA

Viral marine sequences

Viral ALAS sequences:

>SI39nov09

MEHLDKFKQVITELKDDGRYRVFNDILRTRGSYPNAIWYSKYSIKKIVNWCSNDYLGMGQHSYVLDSMKTALETSGAGAGGTRNISGT
THYHIALEHELTQLHGKESALLFTSAYNANQTTLETMGKIMPDLLFISDAENHSSIIQGLRHSKCKKEIFKHNDLDDLESILMSNPGP
KCVVFESVYSMDGDIAPVKEIADLCKKYKAISYIDEVHGVGLYGPNGAGICERDKVDVDIINGTLAKAYGVQGGYICGKREFIDAIRS
MASAFIFTTSLSPVLCAGALTSIKYVKDHPELREKLQERAQKTKEELTRQGIEVLQNDSHIVPVIIGDAKKCKAVSDELLYKDGIYVQ
PINYPTVAVGTERLRFTPTPFHTDTMIFDMVVKVKSAMRRCGRT

>SAMEA2622841_6527

MKELEKFTQVIEDYKSDGRYRTFNDIIRIKGKYPHAIWYSKYSIKNIVNWCSNDYLGMGQHNYVIDSMKTALETSGAGAGGTRNISGT
THYHNALERELASLHKKEKALLFTSAYNANQTTLETMGKIIPDMLFISDAQNHSSIIQGLRHSRCRKEIFKHNDVKDLEGILQSNPGP
KCVVFESVYSMDGDIAPVKEIVDLCNKYNAISYIDEVHAVGLYGEEGAGICERDNVEVDIINGTLAKAFGVQGGYIAGKRDFIDAIRS
MASAFIFTTSVSPVICAGALTSVKYVRDHPELREKIHERANKTKEELTRQGIEVMKNDSHIVPVIIGEAKRCKAISDELLYKEGIYVQ
PINWPTVAVGTERLRFTPTPFHTDKLIFDMVVKVKSAIKRCGKGLKYDR

>SAMEA2620861_1137947

MEHLDKFQQVIKDYKDDGRYRTFNDIIRKRGDYPNAIWYSKYSIKNIVNWCSNDYLGMGQHSYVIDSMKTALETAGAGAGGTRNISGT
THYHNALERELALLHKKESALLFTSAYNANQTTLETMGKIIPDLLFISDEENHSSIIQGLRHSKCRKEIFKHNDVQDLESILMSNPGP
KCVVFESVYSMDGDIAPVKEIIEVSKKYNAITYIDEVHAVGLYGETGAGICERDKVEVDIINGTLAKAYGVQGGYITGKREFIDAIRS
MASAFIFTTSLSPVICAGALTSIKYVKDHPELREKIQERARKTKEEIERQGIEVLKNDSHIVPVIIGDPIKCKAVSDELLYKEGIYVQ
PINWPTVKRGTERLRFTPTPFHTDAHIFDMVVKLKSALKRCGKKK

>3300009790_Ga0115012_10000076_761838

MEHLDKFKAVIKDLKEDGRYRVFNDILRTRGDYPNAIWYSKYSIKKIVNWCSNDYLGMGQHSYVLDSMKTALETSGAGAGGTRNISGT
THYHIALEHELTQLHKKESALLFTSAYNANQTTLETMGKIMPDMLFVSDAQNHSSIIQGLRHSKCRKEIFKHNDVKDLESILMSNPGP
KCVVFESVYSMDGDIAPVKEIVEVCKKYNAITFIDEVHAVGLYGPTGAGICERDGIEVDIINGTLAKAYGVQGGYIAGKREFIDAIRS
MASAFIFTTSLSPVLCAGALTSIKYVKDHPELRMKLQERALKTKEELERAGIEVLKNDSHIVPVIIGDAKKCKAVSDELLYKNGIYVQ
PINYPTVAVGTERLRFTPTPFHTDAMIFDMVVKVKSAMRRCGNKK

>SAMEA2622357_9342

MEHLDKFKAVIKDLKDDGRYRVFNDILRTRGDYPNAIWYSKYSIKKIVNWCSNDYLGMGQHSYVLDSMKTALETSGAGAGGTRNISGT
THYHIALEHELTQLHKKESALLFTSAYNANQTTLETMGKIMPDLLFISDAQNHSSIIQGLRHSKCRKEIFKHNDVQDLESILMSNPGP
KCVVFESVYSMDGDIAPVKEIVEVCKKYNAITFIDEVHAVGLYGKTGAGICERDNVDVDIINGTLAKAFGVQGGYIAGKREFIDAIRS
MASAFIFTTSLSPTICAGALTSIKYVKDHPELREKLQERAKKTKEEIERQGIEVLKNDSHIVPVIIGDPKKCKAVSDELLYKNGIYVQ
PINWPTVARGTERLRFTPTPFHTDAHIFDMVVKLKSALKRCGKKK

>SAMEA2620836_7330

MEHLDKFQQVIKDYKDDGRYRTFNDIIRKRGDYPNAIWYSKYSIKNIVNWCSNDYLGMGQHSYVIDSMKTALETAGAGAGGTRNISGT
THYHNALERELALLHKKESALLFTSAYNANQTTLETMGKIIPDLLFISDEENHSSIIQGLRHSKCRKEIFKHNDVQDLESILMSNPGP
KCVVFESVYSMDGDIAPVKEIIEVSKKYNAITYIDEVHAVGLYGETGAGICERDKVEVDIINGTLAKAYGVQGGYITGKREFIDAIRS
MASAFIFTTSLSPVICAGALTSIKYVKDHPELREKIQERARKTKEEIERQGIEVLKNDSHIVPVIIGDPIKCKAVSDELLYKEGIYVQ
PINWPTVKRGTERLRFTPTPFHTDAHIFDMVVKLKSALKRCGKKK

>SAMEA2620413_422904

MEHLDKFQQVIKDYKDDGRYRTFNDIIRKRGDYPNAIWYSKYSIKNIVNWCSNDYLGMGQHSYVIDSMKTALETAGAGAGGTRNISGT
THYHNALERELALLHKKESALLFTSAYNANQTTLETMGKIIPDLLFISDEENHSSIIQGLRHSKCRKEIFKHNDVQDLESILMSNPGP
KCVVFESVYSMDGDIAPVKEIIEVSKKYNAITYIDEVHAVGLYGETGAGICERDKVEVDIINGTLAKAYGVQGGYITGKREFIDAIRS
MASAFIFTTSLSPVICAGALTSIKYVKDHPELREKIQERARKTKEEIERQGIEVLKNDSHIVPVIIGDPIKCKAVSDELLYKEGIYVQ
PINWPTVKRGTERLRFTPTPFHTDAHIFDMVVKLKSALKRCGKKK

>3300003620_JGI26273J51734_10000260_386802

MEHLDKFKQVITELKDDGRYRVFNDILRTRGSYPNAIWYSKYSIKKIVNWCSNDYLGMGQHSYVLDSMKTALETSGAGAGGTRNISGT
THYHIALEHELTQLHGKESALLFTSAYNANQTTLETMGKIMPDLLFISDAENHSSIIQGLRHSKCKKEIFKHNDLDDLESILMSNPGP
KCVVFESVYSMDGDIAPVKEIADLCKKYKAISYIDEVHGVGLYGPNGAGICERDKVDVDIINGTLAKAYGVQGGYICGKREFIDAIRS
MASAFIFTTSLSPVLCAGALTSIKYVKDHPELREKLQERAQKTKEELTRQGIEVLQNDSHIVPVIIGDAKKCKAVSDELLYKDGIYVQ
PINYPTVAVGTERLRFTPTPFHTDTMIFDMVVKVKSAMRRCGRTK

>3300009593_Ga0115011_10000306_738531

MEHLDKFKAVIKDLKEDGRYRVFNDILRTRGDYPNAIWYSKYSIKKIVNWCSNDYLGMGQHSYVLDSMKTALETSGAGAGGTRNISGT
THYHIALEHELTQLHKKESALLFTSAYNANQTTLETMGKIMPDMLFVSDAQNHSSIIQGLRHSKCRKEIFKHNDVKDLESILMSNPGP
KCVVFESVYSMDGDIAPVKEIVEVCKKYNAITFIDEVHAVGLYGPTGAGICERDGIEVDIINGTLAKAYGVQGGYIAGKREFIDAIRS

MASAFIFTTSLSPVLCAGALTSIKYVKDHPELRMKLQERALKTKEELERAGIEVLKNDSHIVPVIIGDAKKCKAVSDELLYKNGIYVQ
PINYPTVAVGTERLRFTPTPFHTDAMIFDMVVKVKSAMRRCGNKK

>3300006166_Ga0066836_10000757_480248

MEHLDKFKQVIDDYKSDGRYRTFNDIIRTRGKYPHAIWYSKYSIKNIVNWCSNDYLGMGQHNYVIDSMKTALETSGAGAGGTRNISGT
THYHNALERELASLHKKEKALLFTSAYNANQTTLETMGKVMPDLLFISDAQNHSSIIQGLRHSRCRKEIFKHNDLDDLESILKSEPGP
KCVVFESVYSMDGDIAPVKEIADLCKKYNAISYIDEVHAVGLYGKEGAGICERDNVEVDIINGTLAKAFGVQGGYIAGKREFIDTIRS
MASAFIFTTSVSPVICAGALTSVKYVRDHPELRDKIHERANKTKEELERQGIEVMKNDSHIVPVIIGEAKRCKAVSDELLYKEGIYVQ
PINWPTVAVGTERLRFTPTPFHTDNLIFDMVVKVKAAIKRCGKKLNYD

>3300005521_Ga0066862_10001059_417910

MEHLDKFKQVIDDYKSDGRYRTFNDIIRTRGKYPHAIWYSKYSIKNIVNWCSNDYLGMGQHNYVIDSMKTALETSGAGAGGTRNISGT
THYHNALERELASLHKKEKALLFTSAYNANQTTLETMGKVMPDLLFISDAQNHSSIIQGLRHSRCRKEIFKHNDLDDLESILKSEPGP
KCVVFESVYSMDGDIAPVKEIADLCKKYNAISYIDEVHAVGLYGKEGAGICERDNVEVDIINGTLAKAFGVQGGYIAGKREFIDTIRS
MASAFIFTTSVSPVICAGALTSVKYVRDHPELRDKIHERANKTKEELERQGIEVMKNDSHIVPVIIGEAKRCKAVSDELLYKEGIYVQ
PINWPTVAVGTERLRFTPTPFHTDNLIFDMVVKVKAAIKRCGKKLNYD

>3300009785_Ga0115001_100000

MEHLDKFKKIITELKADGRYRVFNDILRTRGSYPNAIWYSKYSIKKIVNWCSNDYLGMGQHSYVIDSMKTALETSGAGAGGTRNISGT
THYHIALEHELTQLHDKESALLFTSAYNANQTTLETMGKIMPEILFISDAENHSSIIQGLRHSKCKKEIFKHNDLDDLESILKSNPGP
KCVVFESVYSMDGDIAPVKEIADLCKKYKAISYIDEVHGVGLYGPKGAGICERDNVDVDIINGTLAKAYGVQGGYICGKREFIDAIRS
MASAFIFTTSLSPVLCAGALTSIKYVKDHPELREQLQERAQKTKEELTRQGIEVLQNDSHIVPVIIGDAKKCKAVSDELLYKDGIYVQ
PINYPTVAVGTERLRFTPTPFHTDMMIFDMVVKVKSAMRRCGKTK

Viral PcyA/X sequences:

>SAMEA2620861_1137947

MLSMKQDNVLVLQLTYLKKCCNLLRITKRGLVMAQRSRIWEMLEDATKNLMSKFDDEGKEIVEESMAKFNRPEDGWINRVWETPEARR
CHIDVVDARESKKLYMFHCVVIPHFHTPAPIWGLDVIAGPNKVTGFFHDWSPLSGKREMDHPMVEWFCEESKTYEPSKVRELPDWALQ
IFSPGMIAAGNINTEKELTNALSLACTDLGPYFTLLRRYKQDFNLNSNIKSEKEVKEAQNRYAKFQRENPHTPRTMKALGLPEKDIEE
FCTDALFPYAE

>SAMEAE2622841_6527

MKQDNVLDLQQIYLKIFTNLLNKTRRGLVMAQYKRSRIWNMLEETTTLIMAVFEREGKEIFEPTMEKFNRPKDGWINRVWETPEARRC
HLDVVDARATKGLYMFHCCVFPRLNSPAPIYGLDVIAGAKKVTGFFHDFSPLAKRDHSMVDWFVKESSNYKPSKVRELPEWALKIFSP
GMVAASNITTEKELNHALSLSQCNLGAYFTLLRREKNNKTDITEQEIKDAQNRYAKHQRENPHTPRVMKSLGLPEDDVKEFCTNALFP
YVE

>SI39nov09

MKRGLVMERRSRIWEMLEQHTHSIIANFEREGEEIFEPAMKKFNRPEEGWVNRVWKTPEARRCHLDVVDARDEKGLFMFHCCVFPNLT
SEAPIFGLDVIAGAKKVTGFFHDFSPLAKRDHSMVDWFVKESKNYTPSKARPLPDWAMKIFSPGMIAAGNINTEKELTQALSMAQSNL
QVYFTLLRRNKEVGDLQEIKDAQNRYAKHQRENPHTPRVMLSLGLPEDDVKEFCTDALFPYVE

>Ga0115001_100000_310

MKRGLVMERRSRIWEMLEQHTQSIIANFEREGEEIFEPAMKKFNRPEEGWVNRVWKTPEARRCHLDVVDARDEKGLFMFHCCVFPNLT
SEAPIFGLDVIAGAKKVTGFFHDFSPLAKRDHSMVDWFVKEASNYTPSKARPLPDWAMKIFSPGMIAAGNINTEKELTQALSMAQANL
QVYFTLLRRNNEVGDLQEIKDAQNRYAKHQRENPHTPRVMLSLGLPEDDVKEFCTDALFPYVE

>Ga0115001_100000

MKRGLVMERRSRIWEMLEQHTQSIIANFEREGEEIFEPAMKKFNRPEEGWVNRVWKTPEARRCHLDVVDARDEKGLFMFHCCVFPNLT
SEAPIFGLDVIAGAKKVTGFFHDFSPLAKRDHSMVDWFVKEASNYTPSKARPLPDWAMKIFSPGMIAAGNINTEKELTQALSMAQANL
QVYFTLLRRNNEVGDLQEIKDAQNRYAKHQRENPHTPRVMLSLGLPEDDVKEFCTDALFPYVE

>Ga0115012_10000076 761838

MKRGLAMERRSRIWEMLEQHTHSIIANFEREGEEIFEPTMKKFNRPEQGWVNRVWATPEARRCHLDVVDARDEKGLYMFHCCVFPKLT
STAPIYGLDVIAGAKKVTGFFHDFSPLAKKDHSMVDWFVKEASLYTPSKPRPLPDWAMKIFSPGMIAAGNINTEKELTQALSMAQSNL
SVYFTLLRREKEQGDIQEIKDAQNRYAKHQRENPHTPRVMLSLGLPEDDVKEFCSDALFPYVE

II. Further sequences used for phylogenetic analysis:

ALAS sequences:

>*Rhodobacter_capsulatus*_DSM938

MDYNLALDKAIQKLHDEGRYRTFIDIEREKGAFPKAQWNRPDGGKQDITVWCGNDYLGMGQHPVVLAAMHEALEAVGAGSGGTRNISG
TTAYHRRLEAEIADLHQKEAALVFSSAYNANDATLSTLRVLFPGLIIYSDSLNHASMIEGIKRNAGPKRIFRHNDVAHLRELIAADDP
AAPKLIAFESVYSMDGDFGPIKEICDIAEEFGALTYIDEVHAVGMYGPRGAGVAERDGLMHRIDIFNGTLAKAYGVFGGYIAASARMV
DAVRSYAPGFIFSTSLPPAIAAGAQASIAFLKTAEGQKLRDAQQMHAKVLKMRLKALGMPIIDHGSHIVPVVIGDPVHTKAVSDMLLS
DYGVYVQPINFPTVPRGTERLRFTPSPVHDLKQIDGLVHAMDLLWARCALNRAEASA

>*Sinirhodobacter_hankyongi*

MDYDLALDQAIQKLHDEGRYRVFIDIEREKGAFPKAVWTRPDGSKKEITIWCGNDYLGMGQHPVVLAAMHEALDATGAGSGGTRNISG
TTVYHKRLEAELADLHGKEAALVFSSAYIANDATLSTLRVLFPGLIIYSDALNHASMIEGIKRNAGPKRIFRHNDVAHLRELLAADDP
AAPKLIAFESVYSMDGDFGKITEICDLAKEFNALTYLDEVHAVGMYGPRGAGIAEKLGEMGRIDIFNGTLGKAFGVFGGYVAGTAKMM
DALRSYAPGFIFTTSLPPVIAAGAAASIAFLKTAAGQELRDKQQLHAKILKMRLKALGMPIIDHGSHIVPVVIGDPKHTKALSDMLLE
EFGVYVQPINFPTVPRGTERLRFTPSPVHDLKQIDHLVRGMDTLWSRCALNRAEASA

>*Pararhodobacter*_sp._CCB-MM2

MTYDTALDAALQRLHDEGRYRTFIDIERKKGHFPHAVWTRKDGSTKDITVWCGNDYLGMGQHPVVLEAMHEAIDATGAGSGGTRNISG
TTVYHKRLEAEIADLHGKEAALIFTSAYIANDATLSTLPKLFPGLIIYSDALNHASMIEGVRRNGGAKRIFRHNDVAHLRELLEADDP
AAPKLIAFESIYSMDGDFGPIEAICDLADEFNALTYIDEVHAVGMYGPRGAGVAERDGLMHRIDIINATMAKAFGVMGGYIAASAKMV
DAVRSYAPGFIFTTSIPPAVAAGCTASIAFLKTAAGQELRDKQQLHARILKMRLKGLGMPIIDHGSHIVPVHVGNPVHCKMLSDMLLE
DHGVYVQPINFPTVPRGTERLRFTPSPVHDPKEIDALVKAMDKLWSHCALNRAEMSA

>*Thioclava_atlantica*

MDYDRALDQAIGKLHEEGRYRVFIDIEREKGNFPHAVWTKPDGTRKDITIWCGNDYLGMGQHPVVLSAMHEALDATGAGSGGTRNISG
TTVYHKRLEREIADLHGKEAALVFSSAYIANDATLSTLPKLFPGLIIYSDELNHASMIEGIRRNGGAKRVFRHNDVAHLRELLAADDP
AAPKLIAFESVYSMDGDFGPISEICDLAEEFGALTYLDEVHAVGMYGPRGAGIAEQLGEMGRIDIFNGTLGKAFGVFGGYIAASARMV
DAVRSYAPGFIFTTSLPPAVAAGAAASIAHLKTAAGQELRDAQQLHAKVLKMRLKAMGMPIIDHGSHIVPVIIGDPVHTKALSDMLLD
EFGVYVQPINFPTVPRGTERLRFTPSPVHDLGMIDGLVKAMDKLWSRCELNRAEASA

>*Paracoccus_endophyticus*

MNYDAALDQAIDRLHQEGRYRTFIDIERRKGAFPQAIWTRPDGSETEITVWCGNDYLGMGQHPAVLAAMHEALDATGAGSGGTRNISG
TTVYHKRLESEIADLHGKEAALVFSSAYIANDATLSTLRKLFPGLIIYSDALNHASMIEGIKRFDGEKRIFRHNDVAHLRELLEADDP
ALPKLIAFESIYSMDGDFGPIKAICDLAQEFHALTYLDEVHAVGMYGPRGGGVAERDGLMGRIDIFNGTLGKAFGVFGGYIAASARMV
DAIRSYAPGFIFTTSIPPAVAAGAAESIRLLKTAEGQALRDAQQLHARILKMRLRGLGLPIDDRGSHIVPVHVGNPVHCKMLSDMLLA
DFGIYVQPINFPTVPRGTERLRFTPSPVHDSKQIDHLVRAMDALWSRCALNRAHAVA

>*Pseudorhodobacter*_sp.

MNFDTALDAALTRLHDEGRYRTFIEIEREKGQFPHAVWTRPDGSKKDITVWCGNDYLGMGQHPVVLAAMHEALEATGAGSGGTRNISG
TTVYHKRLEAEIADLHGKEAALVFTSAYIANDATLSTLPKLFPGLIIYSDALNHASMIEGVRRNGGAKRIFRHNDVAHLRELIAADDP
AAPKLIAFESIYSMDGDFGPIKDICDLAEEVGALTYIDEVHAVGMYGPRGAGVAERDGQMARIDIINATLAKAYGVFGGYIAASAKMV
DAVRSYAPGFIFTTSLPPAVAAGAAASIAFLKTAEGQKLRDSQQLNARILKMRLRGLGLPIIDHGSHIVPVHVGNPIHCKMLSDMLLE
DHGIYAQPINFPTVPRGTERLRFTPSPVHDAGQIDGLIRAMDSLWQHCALNRAEISA

>*Lutimaribacter_litoralis*

MDYSAQLDQALNALHSEGRYRTFIDIERKKGQFPHAVWQRPDGTEKDITVWCGNDYLGMGQHPVVLAAMHEAIDATGAGSGGTRNISG
TTVYHKRLEAELADLHGKEAALVFTSAYIANDATLSTLPKLFPGLIIYSDELNHASMIEGVRRNGGAKRVFRHNDVAHLRELLAADDP
AAPKLIAFESVYSMDGDFGPIEAICDLADEFNALTYIDEVHAVGMYGPRGGGVAERDRLMHRLDIINGTLAKAYGVMGGYIAASAKMC
DAVRSYAPGFIFSTSLAPTLAAGATASVAFLKTAEGQKLRDKHQTQAKILKTRLKGLGLPIIDHGSHIVPVIVGDPVHTKMLSDMLLD
GYGAYVQPINFPTVPRGTERLRFTPSPVHGAAEMDHLVRALDGLWAQCKLNRAEMSA

>*Pararhodobacter_aggregans*

MTYDSALDTALQRLHDEGRYRTFIDIERKNGQFPHAVWSKEDGSEKDITVWCGNDYLGMGQHPVVLAAMHEALDATGAGSGGTRNISG
TTVYHKRLEAELADLHQKEAALVFTSAYIANDATLSTLRAIFPGLIIYSDVLNHNSMIEGIRKNGGEKRIFRHNDVAHLREMLEADDP
AAPKLIAFESIYSMDGDFGPIEAFCDLADEFNALTYIDEVHAVGMYGPRGGGVAERDGLMDRIDIINGTMAKAIGVFGGYIAASAKMV
DAIRSYAPGFIFTTSLPPTVAAGCAASIAFLKTAEGQALRDKQQLHAQILKMRLKGLGLPIIDHGSHIVPVHVGHPVHCKLISDMLLD
RYGVYVQPINFPTVPRGTERLRFTPSPVHDPREIDRVVTAMDQLWAHCALNRAELSA

>*Defluviimonas*_sp._XJ4

MDYDKALDGAIQRLHDEGRYRTFIDIERRKGAFPHAVWRRPDGTESEITVWCGNDYLGMGQHPAVLAAMHEALDATGAGSGGTRNISG
TTVYHKRLEAELADLHRKEAALVFSSAYIANDATLSTLPQLFPGLIIYSDALNHASMIEGIRRNGGAKRIFRHNDVAHLRELLEADDP
AAPKLIAFESIYSMDGDFGPIAAICDLARDFGALTYLDEVHAVGMYGPRGGGVAERDGLLDRIDILNGTLGKAFGVFGGYIAASAKMV

DAIRSYAPGFIFTTSLPPAVAAGAAASIAFLKTDEGQKLRDAQQLHARILKMRLKGLGMPIIDHGSHIVPVHVGNPIHCKMLSDMLLA
EFGVYVQPINFPTVPRGTERLRFTPSPVHDLGQIDHLVRAMDSLWSHCALNRAELSA

>*Gemmobacter_straminiformis*

MDYEAALDASLQRLHDEGRYRTFIDIERVNGHFPKAIWRRPDGTSQDITVWCGNDYLGMGQHPEVLAAMHEALDATGAGSGGTRNISG
TTVYHKRLEAELADLHGKDAALVFSSAYIANDATLSTLRSIFPGLIIYSDALNHASMIEGVRRGGGPKRIFRHNDVAHLRELLAADDP
AAPKLIAFESIYSMDGDFGPIKDICDVAAEFNALTYLDEVHAVGMYGPHGAGVAERDGEMHRVDIINATLGKAYGVFGGYIAASARMV
DAVRSYAPGFIFTTSLPPVVAAGAAAAVRVLKTAQHLRDAQQLHARVLKMRLKALGLPIIDHGSHIVPVHVGDPVHCKMLSDMLLDRY
GIYVQPINFPTVPRGTERLRFTPSPVHDLKQIERLVTAMDALWRHCALNRAEASA

>*Tabrizicola*_sp._TH137

MNYEAALDASLQRLHDEGRYRTFIDIVRKQGHFPQAIWRRPDGTERDITVWCGNDYLGMGQHPEVLAAMHEALDATGAGSGGTRNISG
TTIYHKQLEAELADLHGKEAALVFSSAYIANDATLSTLRLLFPGLIIYSDALNHASMIEGVRRGGGAKRIFRHNDVAHLRELLAADDP
AAPKLIAFESIYSMDGDFGPIKDICDLAEEFNALTYIDEVHAVGMYGPRGAGVAERDGQMHRIDIINATLAKAYGVFGGYIAASAKMV
DAIRSYAPGFIFTTSLPPAVAAGAAASVRILKTAQHLRDAQQTHARILKMRLKGLGLPIIDHGSHIVPVHVGDPVHCKMLSDMLLENY
GIYVQPINFPTVPRGTERLRFTPSPVHDARQIDHLVKAMDALWRHCALNRAEVSA

>*Gemmobacter_megaterium*

MTYDAQLDAALQRLHDEGRYRTFIDIERRKGHFPHAVWRKPDGSEQDITVWCGNDYLGMGQNPTVLAAMHEALDATGAGSGGTRNISG
TTVYHKRLEAEIADLHAKEAALVFSSAYIANDATLSTLRLLFPGLIIYSDALNHASMIEGVRRGNGPKRIFRHNDVAHLRELIAADDP
AAPKLIAFESIYSMDGDFGPIREICDLAEEVGALTYLDEVHAVGMYGPRGAGVAERDGQMHRIDIINATLGKAFGVFGGYIAASARMV
DAIRSYAPGFIFTTSLPPVVAAGAAASIRLLKTEDGQKLRDAQQLHARILKMRLKGMGLPIIDHGSHIVPVHVGNPVHCKMLSDMLLE
DFGIYVQPINFPTVPRGTERLRFTPSPVHDSGQIDRLVHAMDSLWAQCALNRAELSA

>*Paracoccus_homiensis*

MNYDAALDKAIGRLHEEGRYRTFIDIERAKGRFPQAIWTRPDGESQEITVWCGNDYLGMGQHPVVLNAMHEALDAAGAGSGGTRNISG
TTVYHKELEAELADLHGKEAALVFSSAYIANDATLSTLPKLFPGLIIYSDALNHASMIEGIKRNGGAKRIFRHNDLDHLRELMAADDP
AAPKLIAFESIYSMDGDFGPIEAICDLADEFGALTYLDEVHAVGMYGPRGGGVAERDGLLDRIDIFNGTLGKAFGVFGGYIAATAKMC
DAIRSYAPGFIFTTSLPPTVAAGAAASIRFLKSAEGQHLRDAQQLHARILKMRLKSLGMPIIDHGSHIVPVHVGHPVHCKQLSDMLLR
DYGIYVQPINFPTVPRGTERLRFTPSPVHDSRQIDHLVRAMDALWSHCALNRAELSA

>WP_011924405.1 5-aminolevulinate synthase [*Bradyrhizobium* sp. ORS 278] alaS on genomic island

MDYEAMFRGSLAELRGEGAYRVFADLERMVGAFPRAIYRRDGVAKVVTIWCSNDYLGMGENPQVLGAMHRALDRSGAGAGGTRNISGT
SHAHVMLEAELAALHGKPAALLFNSGYASNWATLGTLGARLPNCVMFSDELNHASIIEGIRHSRCEVRIFKHNDVADLRAKLAAVEAR
VPKIVVFESVYSMDGDIAPISDICDVADAHGAMTYLDEVHAVGLYGHEGGGISELHGVRDRLTMIEGTLAKAFGVVGGYIAGSRDLCD
FIRSFASGFIFSTALPPSVAAGALESIKLLRRDGRRRARHQNRVSLLRKKLDQAGIAHLENPSHIVPIMVGEPRLCKQMSDRLLDLHD
IYVQPINYPTVPRGTERLRLTPTPQHTDEDIAKLVDALIEVRSSLVPTSQQGRVSRNGRPLTRTRSAHRTGSEHRKPSAETVEATV

>WP_011924757.1 5-aminolevulinate synthase [Bradyrhizobium sp. ORS 278]

MNYEDYFRRQLDGLRREGRYRVFADLERKAGAFPRATYRGPHGEREVSVWCSNDYLGMGQHPKVLEVMHEALDRCGAGAGGTRNISGT
NHYHILLEEELADLHGKQAALLFNSGYVSNWTSLSTLAGRMPGCVILSDELNHASMIEGIRHSRCETRIWRHNDVADLRDKLSDLDPK
VPKLIAFESVYSMDGDISPISDICDVADEYNAMTYLDEVHAVGLYGPRGGGISERDGVADRLTVIEGTLAKAFGVIGGYITASNTVCD
FVRSFASGFIFSSSLPPAVAAGALTSIRHLKASNHERARHQDRVARLRRRLDEAGVSHLDNPSHIVPVMVRDPVLCKRISDILLDRYG
VYVQPINYPTVPRGTERLRITPSPYHSDEDIEHLVSALSEIWAEMGLAQAA

>WP_011926517.1 5-aminolevulinate synthase [Bradyrhizobium sp. ORS 278]

MKNYDRYFDHALDQLHQERRYRVFADLERRAGRFPHAIWHSPEGPRDVVIWCSNDYLGMGQHPKVIGAMVDVAARMGTGAGGTRNIAG
TNHPLVELERELADLHGKEAALVFTSGYVSNETGISTIAKLLSNCLLLSDALNHNSMIEGVRKAGTEKQVWRHNDLAHLEELLQAADP
DRPKLIVCESLYSMDGDVAPLHRICDLADRYGALTYVDEVHAVGMYGPRGGGIAERDGAMARIDILEGTLAKAFGCLGGYIAANANLI
DAVRSYAPGFIFTTALPPPVCAAATAAIRHLKTSQWERDRHQDRAARLKAVLNASGLPVMPSDTHIVPVHVGDPEKCKSASDLLLAEH
GIYIQPINYPTVPRGLERLRITPSPYHDDGLIDALAEALVDVWDRLGLPRAAQAMAAE

>WP_012030145.1 5-aminolevulinate synthase [Bradyrhizobium sp. ORS 278]

MDYSQYFSAALGRLHDERRYRVFADLERIAGRFPHATWHSPNGPRNVVIWCSNDYLGMGQHPKVVGAMVETATRVGTGAGGTRNIAGT
HHPLVQLEAELADLHGKEAALLFTSGYVSNQTGIATLAKLIPDCLILSDALNHNSMIEGVRQSGCERQIWRHNDLAHLEQLLIAAGPS
RPKLIACESLYSMDGDVAPLAAICDLAERYNAMTYVDEVHAVGMYGARGGGIAERDGVMHRIDILEGTLAKAFGCLGGYIAAKGEIID
AVRSYAPGFIFTTALPPPICSAATAAIKHLKSSSWERERHQDRAARVKAILNAAGLPVMDNDTHIVPLFVGDPEKCKQASDLLLEEHG
IYIQPINYPTVAKGTERLRITPSPYHDDGLIDRLAEALLQVWERLGLALHKKPLAAE

FDBR sequences:

>PebA_*Synechococcus*_sp.WH8020

MFDSFLNELHSDITKRGGSPLPLPEGLEECRSSKSSSVIQSWLWDVPGFRRWRVTRLDAGDSLQVFNSVAYPDYNYDHPLMGVDLLWF
GARQKLVAVLDFQPLVQDKDYLDRYFSGLKELNQRFPDLNGEETMRSFDPNQYFSSWLLFCRGGAEQADLSLPKAFSAFLKAYWDLHD
NAKSIPSTIPPEEVKNLQDKYDIYSAERDPAHGLFTSHFGKDWSNRFLHEFLFPASSSHK

>PebA_*Prochlococcus_marinus*_MIT9313

MFDPLLEKLHSSIRIQGGETAAVPDGLRECRNEKKNSWIRSWLWQVPGFRRWRVSRLDAGESLQVLNSVAYPNYNIDQPLMGLDLLWF
GKRQKLVAILDFQPLIQDQSYLERHFQGLKTLQNRFPELSGEETMRLFDPNQYFSPWLLFCRGGAEKATNSLPEAFNAFLHCYWELHQ
QNSDKASLIPAAEVKQLQIAYDIYSAERDPAHGLFTSHFGKAWSDRFLHEFLFPASTKADSSPPADADDDLPR

>PebA_*Synechococcus*_sp.WH7803

MFDSFLNKLHQGIHQRGGQPAAVPEGLEHCHSSKSGGSIDSWLWSVPGFRRWRVTRLDAGESLQVLNSVAYPDHSLDHPLMGVDLLWF
GARQKLVAVLDFQPLIQDQDYLDRHFQGLKALHERFPELNGEETMRSFDPNQYFSPWLLFCRGGAQEAEESLPAAFDAFLTAYWAMHD
QALDVEAAGTSASTLSINDVERLQEAYDVYSAERDPAHGLFTSHFGKDWSDRFLHEFLFPASQST

>PebB_*Synechococcus*_sp.WH8020

MTNQRFKSTDPVNIEGWSWQPFLEDAIKRLEGLNVEPYPVPDRFLQREDQTGSKSKSIPVTTATWACKTEKFRQVRAACVSAGSAASV
LNFVINPKSTYDLPFFGGDLVTLPAGHLLALDLQPAIKTDEVHTTHVWDRLIPIFERWRDQLPYGGPIPEEAQPFFSPGFLWTRLPLG
EEGDELIQSIVRPAFNDYLDLYLELAASAERVTDERSEVLLQGQRKYTDYRAEKDPARGMLTRFHGSEWTEAYIHTVLFDL

>PebB_*Prochlorococcus_marinus*_MIT9313

MKPSRLSSLDPVKMPEWRWAPFLSHAINAFIPLKPEPYPVAPEFLQREGKTGSKSQPIRVTTCTWACRTKKFRQVRAACVEAGRSASV
LNFVINPYHTFDLPFFGADLVTLPSGHLLALDLQPAITSDERHTKQVWERLMPIFEHWRVHLPEGGPIPEEAKPYFSPGFLWTRLPLS
IEGNQLIDEVIMPAFKDYLNLYLDLVEMAEEVSPQRAFKLLEGQKRYLSYRAKKDPARAMLARFHGHQWTESYIHNVLFDL

>PebB_*Synechococcus*_sp.WH7803

MSSLVRTSSLDPVQIKGWRWQPFLDKAISALETFRLSPYPIDETFLQREGSTGSKSKPVPVTTATWACATDKLRQVRCACVEAGTAAS
VLNFVINPSCRFDLPFFGADLVTLPNGHLLALDLQPVDKADPEHTGPVWERLMPVFERWQAELPDGGPIPEEAQPYFSPAFLWTRIPL
GEEGDALIERVIRPAFIEYLQLYLKLVAEAQPVNDDRAASLLAGQKRYTAYRAEKDPARGMLTRFYGSEWTESYIHDVLFDLEKAAST
SLS

>PebS_S-SSM7

MKLKLMNLWKNYKSLLYEVFPELYHHSTWAEWEGKGTSLTAKLYGTDKDWYINKSREVEIWNEKSCIYNTIIYPRTGENLPCFGMDLM
GFFEKKVIIVFDFQHPIENCPFSVQGLPKAEQDYRFFEMGNHFSDNIYVRYCTFAEVDEHLDMFKKYLTVYRDMLESKKPSQNLMHKT
YHDFDKYMRKLDPVGGYLSGKFGKEKSESLVNDFLFTYG

>PebS_S-PRM1

MDTSNDIWKNYKKVLWETFPDLENICDWADWEGKNLNLSAKLYNSPYILKAREVEIWNEKTCIYNTIIYPKTGSDLPCFGMDLMMFFP
KKVVLTFDFQHPRENYLFSVPNLPKCEGGIRFFEPGNHFSENLYIRKCTSEQVDDYLNDFKTYLQIFAGMLNSKKPTGLDISKYTDFD
QYMTKLDPVAGYLSSNFGKEQSEKLVKEFLFTY

>PebS_GOS_134648280

MNIWSNYKRVLHETFPLHGAAGSVWAQWEGKGTSLLARTYTNQYFIKAREVEIWNEKSCIYNNIIYPKTGSNLPCFGMDLMGFFDKKV
IIVFDFQHPTENYLFSVEGLPKGKGDYRFFEPGNHFSENIYIQYCTMNEVDDYLDMFKKYLLKYKEMVDEAKPTGMDTSVYKDFDTYM
TKLDPVKGYLSGKFGSEKSESLVNDFLFCYK

>PebS_GOS_134321640

LMMKPNKHSLNSMTKLMQTSNSLWNQYKTAIWETFPDLELDCEWADWRSLQFSSSNVSNLSAKIYVNKHILKSREVEIWDNKSCIYNN
IIYPRTGSNLPCFGMDLMGFFDKKVIIVFDFQHPVENYLFSHPDLPKADGSFRFFEPGNHFSENVYVAKCAMSEINEHLDIFKKYLTT
YKDMLECEKPNGNDFSTYCDFDSYMKKLDPVSGYLSNKFGKEKADSLINDFLFCYG

>PebS_GOS_135569194

MQIKKKMNLWKEYKDALHESIDLHNGVGHVWAQWESKGTTLLAKTYSNQHLIKSREVEIWSDTSCIYNNILYPKTGSNLPCFGMDLMA
FNENRVIVVFDFQHPVENYLFSIEGLPKAEKDYRFFEMGNHFSENIFVRYCKMEEVNAYLSTFKEYLTNFKFMLELERPTGTDTSEYK
DFDAYMTRLDPVGGYLKGKFGKEKAESLVNDFLFQYG

>PebS_GOS_134804361

MQTLKWMQRKMQCIVRIKMNLWVDYKKILHETFPLHNGVGSVWAEWERKGTWLTAKTYTAPYIIKSREVEIFNEKSCIYNNIIYPKTG
SNLPCFGMDLMGFNENRVIIVFDFQHPVENYLFSVEGLPKQEGNIRFFETGNHFSENIYVVKCTFKEVEDHLDMFKTYLTKYRDMLEL
EKPKGTDTSLYKDFDAYMTKLDPVSGYLTGKFGKEKAENLVNDFLFTYG

>PebS_*Prochlorococcus*_phage_P-SSM2

MTKNPRNNKPKKILDSSYKSKTIWQNYIDALFETFPQLEISEVWAKWDGGNVTKDGGDAKLTANIRTGEHFLKAREAHIVDPNSDIYN
TILYPKTGADLPCFGMDLMKFSDKKVIIVFDFQHPREKYLFSVDGLPEDDGKYRFFEMGNHFSKNIFVRYCKPDEVDQYLDTFKLYLT
KYKEMIDNNKPVGEDTTVYSDFDTYMTELDPVRGYMKNKFGEGRSEAFVNDFLFSYK

>Actino_PcyX(uncultured_actinobacterium_SCGC_AAA041-L13)

MNSVWDSLIQLETFFETAFNASGTLINDPSMDRFNQPGWVNLVWTGQNYRRAHIDVVDARHSKGLWMMHCCIFPHTHNPAPIFGFDVI
AGKSKITGCFYDYSPAGDVEHPMLDWFSSEAAKLQWNKTRKLPEWAERIFSSSMIAASNVSKPEEVEQILSIAKKGIDQYLVAVGETN
KTAISTAHEQNFYCENQKLNPHTPKVMTSLGLSEADVTAFIQECLFPEIK

>PcyX_EBK42635

MIWERLIKWKDETIEVLNNNLVEYNEPGMERFNNEKLGWVNRTWNNRYIRRAHLDVVDVRESKGLWMAHLCLFPMLTNGGPIYGFDII
AGEKKVTGAFHDFSPLLQKDHPLTKWFIEENKWFKPSKERELPEWAKAIFSGGMIAAGNVREEDELNKICTMAVSNLNNYIDKIRNHE
GEAEMADVIKAQNYYSEHQQKNPHTPRVMQSLGLPEEDIKLFCSDNLFPFVSENQPYLK

>PcyX_ECK

MIWERLIKWKDETIDVLNKHLVEYNEPGMERFNNEKLGWVNRTWNNRYIRRAHLDVVDVRESKGLWMAHLCLMPMLTNGGPIYGFDII
AGKNKVTGAFHDFSPLLQKDHPLTEWFKEETKWYKPSKERELPDWAKAIFSGGMIAAGNVQEEKELNQICTMAVSNLENYIDKIRNHE
GEAEMIDVIKAQNYYSEHQQKNPHTPRVMQSLGLPDEDIKLFCSDNLFPYVSENQPYLN

>PcyX_EBQ

MIWERLIKWQEETIELLNKELVEYKEPGMERFNNDEFGWVNRTWKNKYIRRAHVDVVDVRDTKGLWMAHVCLFPELTNGGPIYGFDII
AGKNKVTGAFHDFSPLLQKHHPLTEWFLEETSWYKPSKERELPDWAKAIFSGGMVAAGNVTEEKELNQICTMATSNLANYIDKIRTHD
GESNREDVIKAQNYYCEHQQMNPHTPRVMQSLGLPDGDIKLFCQDNLFPKI

>PcyA_*Bradyrhizobium* sp. ORS278

MSDGDDGDDLICDLQHAAEDFAADLRAVPALERVRVPDFHAAAIAEGTLQKEITWRNDVFVGGRFRHAHVESFSIGEQIGVVHVCIFP
HFDRAAPIFGFDIIAGRKKATGAFLDLSPTTMAANAIIDGWSEASAAQRANFRETRILPAWAASIFSRSALAIRPASRHEVASVVALG
RSALAYYLDAHLATAAEAEMQVAQRKYIEAQRSNEHTFRMLAGCVGVDLARDFIDGWLFPAPPSPGESRSDAAARGALAHVD

>PcyA_*Nostoc*_sp. PCC7120 (*Anabaena* sp. PCC 7120)

MSLTSIPSLREQQHPLIRQLADCIEEVWHQHLDLSPYHLPAELGYVEGRLEGEKLTIENRCYQTPQFRKMHLELAKVGNMLDILHCVM
FPRPEYDLPMFGCDLVGGRGQISAAIADLSPVHLDRTLPESYNSALTSLNTLNFSQPRELPEWGNIFSDFCIFVRPSSPEEEAMFLGR
VREFLQVHCQGAIAASPVSAEQKQQILAGQHNYCSKQQQNDKTRRVLEKAFGVDWAENYMTTVLFDLPE

>PcyA_P-SSM4

MIVDDVAQTIRKITSGLPDVKHLPEDPYRSIVKDDIVINNEMWTCTGLRKIHLETCKTKYLDVLHCVLFPEPRYKLPIFGCDIIANNR
IVTAAIVDISPVKGVRGEFYKDIKPISERYMDFDFRKLPEWADIFSPHCKFMRLHKQTEQIMYVQLLEEYLQVYVNAVSKAEKCMDID
ATYDRYQDQVYYCQQQKQNKKTEAVLGSWFDPTWAKHYIDNVLFDKPKPFINL

>PcyA_*Prochlorococcus*_MIT9313

LNLLSKSLTKTKLIDPLILTLLQNIKVQRSKLNDLNCIEVDPKLSNIISNEEGKELYIENEFYKAKGFRKLHIEVAEFSKSLKILHCV
FFPDPKYDIPIFGMDLVKVNELVSAAIVDLSPSSKNQNLKYDHLLSHIDKSVFKSKREIPIWGNIFSKNVFFASLKNESEKNAFCKIV
DNYLSVLIQLSQSTSPDSDYEIIEERINYQKNYCVQQMKNEKTSLVLLKYFDKVWVDEYIKKVLFDF

>PcyA_*Synechocystis*_sp.PCC6803

MAVTDLSLTNSSLMPTLNPMIQQLALAIAASWQSLPLKPYQLPEDLGYVEGRLEGEKLVIENRCYQTPQFRKMHLELAKVGKGLDILH
CVMFPEPLYGLPLFGCDIVAGPGGVSAAIADLSPTQSDRQLPAAYQKSLAELGQPEFEQQRELPPWGEIFSEYCLFIRPSNVTEEERF
VQRVVDFLQIHCHQSIVAEPLSEAQTLEHRQGQIHYCQQQQKNDKTRRVLEKAFGEAWAERYMSQVLFDVIQ

>PcyA_*Prochlorococcus*_NATL2A

MESVFCIQSELNPLLLEAFELIKNRINSLNDLEPLYIDPMLSRIYRKQNEEKLFIINEFYQAKGFRKIHLEVAKLGKSLEILHCVFFP
DPCYDLPIFGADLVVNSNNISAGIVDLSPVGKHLPDCLISQMRSLKVSKFTEPGKLPEWGFIFSPYVAFIRPVDFLEEKSFLELIDQY
LSLLLSLLVKVKRDEINSLDAMCRLNYQKRYCLNQKRNDKTRGILTKFFGSSWADEYINKILFEC

Heme oxygenase sequences:

>Ho1_*Synechocystis*_sp.PCC6803

MSVNLASQLREGTKKSHSMAENVGFVKCFLKGVVEKNSYRKLVGNLYFVYSAMEEEMAKFKDHPILSHIYFPELNRKQSLEQDLQFYY
GSNWRQEVKISAAGQAYVDRVRQVAATAPELLVAHSYTRYLGDLSGGQILKKIAQNAMNLHDGGTAFYEFADIDDEKAFKNTYRQAMN
DLPIDQATAERIVDEANDAFAMNMKMFNELEGNLIKAIGIMVFNSLTRRRSQGSTEVGLATSEG

>Ho2_*Synechocystis*_sp.PCC6803

MTNLAQKLRYGTQQSHTLAENTAYMKCFLKGIVEREPFRQLLANLYYLYSALEAALRQHRDNEIISAIYFPELNRTDKLAEDLTYYYG
PNWQQIIQPTPCAKIYVDRLKTIAASEPELLIAHCYTRYLGDLSGGQSLKNIIRSALQLPEGEGTAMYEFDSLPTPGDRRQFKEIYRD
VLNSLPLDEATINRIVEEANYAFSLNREVMHDLEDLIKAAIGEHTFDLLTRQDRPGSTEARSTAGHPITLMVGE

>Heme_oxygenase_*Neisseria_meningitidis*_63006

MSETENQALTFAKRLKADTTAVHDSVDNLVMSVQPFVSKENYIKFLKLQSVFHKAVDHIYKDAELNKAIPELEYMARYDAVTQDLKDL
GEEPYKFDKELPYEAGNKAIGWLYCAEGSNLGAAFLFKHAQKLDYNGEHGARHLAPHPDGRGKHWRAFVEHLNALNLTPEAEAEAIQG
AREAFAFYKVVLRETFGLAADAEAPEGMMPHRH

>BphO_*Pseudomonas_aeruginosa*_PAO1

MSPSPSPALAALRDATRDLHAELDRRSPLGDDDLDDRAYLDHAGRILGWLEPLERALRDNRSGWPAALRADARLVKSTWLESDLLAGG
MSRAQVEALPRCADLPNATRAAEVFGVAYVMEGATLGGAYLYKRLAPRLPGLPLQWLQGYGQATGVRWQEFLEQLARQIDSPEAIGLA
QDAAQATFLSFRRWVLDEA

>PigA_*Pseudomonas_aeruginosa*_PAO1

MDTLAPESTRQNLRSQRLNLLTNEPHQRLESLVKSKEPFASRDNFARFVAAQYLFQHDLEPLYRNEALARLFPGLASRARDDAARADL
ADLGHPVPEGDQSVREADLSLAEALGWLFVSEGSKLGAAFLFKKAAALELDENFGARHLAEPEGGRAQGWKSFVAILDGIELNEEEER
LAAKGASDAFNRFGDLLERTFA

>HMOX1_*Homo_sapiens*

MERPQPDSMPQDLSEALKEATKEVHTQAENAEFMRNFQKGQVTRDGFKLVMASLYHIYVALEEEIERNKESPVFAPVYFPEELHRKAA
LEQDLAFWYGPRWQEVIPYTPAMQRYVKRLHEVGRTEPELLVAHAYTRYLGDLSGGQVLKKIAQKALDLPSSGEGLAFFTFPNIASAT
KFKQLYRSRMNSLEMTPAVRQRVIEEAKTAFLLNIQLFEELQELLTHDTKDQSPSRAPGLRQRASNKVQDSAPVETPRGKPPLNTRSQ
APPLRWVLTLSFLVATVAVGLYAM

>Ho1_*Prochlorococcus*_Phage_P-SSM2

MTVADFSVQIKEGTKKSHSAAENTSFVASFLRGVVSKESYKALVKDLYYVYRTLEEEFEKHKDHPVVGKLYLPELNRVNALERDLRFY
YGPIWRSLIMPSEACANYVSRIKCCSIEDPTLLVGHHYTRYLGDLSGGQILKGIAEKAMGLKDEGLYFYDFDKIENAKKYKDGYRAIL
NGLDVDQHQVDAIIVEANYAFRLNMYMFDTLEGNWFQSLIQMIFGFIKSILKKRKSE

III. Viral nucleotide sequences used for cloning

>3300009785_Ga0115001_100000(ALASsyn)

ATGGAACATTTAGATAAATTTAAAAAAATAATAACAGAATTAAAAGCTGACGGTAGATACCGTGTATTCAATGATATCTTACGAACCA
GAGGTAGTTATCCTAACGCAATCTGGTATTCAAAATACTCAATCAAAAAAATAGTCAACTGGTGTTCAAATGATTATCTAGGCATGGG
CCAACATTCTTATGTTATAGATAGCATGAAAACAGCACTGGAGACGAGCGGAGCGGGTGCTGGAGGGACGAGAAACATCTCCGGTACA
ACTCACTATCACATTGCATTAGAACACGAATTAACTCAATTACACGACAAAGAAAGTGCGTTATTATTCACTTCAGCATACAACGCTA
ATCAAACAACTTTAGAAACTATGGGCAAGATTATGCCTGAAATTTTATTTATTTCAGACGCCGAAAATCATTCTTCTATCATACAAGG
ATTAAGACATAGTAAATGTAAAAAAGAAATATTTAAACATAATGATTTAGACGATTTAGAAAGTATCTTAAAGTCTAACCCTGGTCCT
AAATGTGTAGTATTTGAAAGTGTCTATAGTATGGACGGAGATATTGCACCTGTAAAAGAGATTGCTGATTTATGTAAAAAGTATAAAG
CTATATCGTACATTGATGAAGTACACGGAGTTGGTTTATACGGACCAAAAGGTGCTGGTATTTGTGAACGAGATAATGTTGATGTTGA
TATTATTAATGGAACATTAGCAAAGGCCTATGGTGTACAGGGAGGTTACATTTGTGGAAAGAGAGAGTTTATAGACGCAATTAGAAGT
ATGGCAAGTGCGTTTATATTTACAACTAGTTTATCGCCTGTACTATGTGCCGGTGCTTTAACAAGTATTAAGTATGTTAAAGACCATC
CAGAATTAAGAGAACAACTACAAGAACGAGCTCAAAAAACAAAAGAAGAATTAACAAGACAAGGTATAGAAGTATTACAAAACGATAG
TCATATCGTACCAGTAATTATAGGTGACGCTAAAAAATGTAAAGCAGTTTCAGATGAATTACTTTACAAAGATGGTATCTATGTACAA
CCTATTAACTATCCTACGGTTGCTGTTGGTACTGAAAGACTAAGATTTACACCTACACCATTTCATACAGATATGATGATATTTGATA
TGGTAGTTAAAGTTAAATCTGCTATGCGAAGATGTGGTAAAACTAAATGA

>ALASopt

ATGGAACACCTGGACAAATTCAAAAAAATCATCACCGAACTGAAAGCTGACGGTCGTTACCGTGTTTTCAACGACATCCTGCGTACCC
GTGGTTCTTACCCGAACGCTATCTGGTACTCTAAATACTCTATCAAAAAAATCGTTAACTGGTGCTCTAACGACTACCTGGGTATGGG
TCAGCACTCTTACGTTATCGACTCTATGAAAACCGCTCTGGAAACCTCTGGTGCTGGTGCTGGTGGTACCCGTAACATCTCTGGTACC
ACCCACTACCACATCGCTCTGGAACACGAACTGACCCAGCTGCACGACAAAGAATCTGCTCTGCTGTTCACCTCTGCTTACAACGCTA
ACCAGACCACCCTGGAAACCATGGGTAAAATCATGCCGGAAATCCTGTTCATCTCTGACGCTGAAAACCACTCTTCTATCATCCAGGG
TCTGCGTCACTCTAAATGCAAAAAAGAAATCTTCAAACACAACGACCTGGACGACCTGGAATCTATCCTGAAATCTAACCCGGGTCCG
AAATGCGTTGTTTTCGAATCTGTTTACTCTATGGACGGTGACATCGCTCCGGTTAAAGAAATCGCTGACCTGTGCAAAAAATACAAAG
CTATCTCTTACATCGACGAAGTTCACGGTGTTGGTCTGTACGGTCCGAAAGGTGCTGGTATCTGCGAACGTGACAACGTTGACGTTGA
CATCATCAACGGTACCCTGGCTAAAGCTTACGGTGTTCAGGGTGGTTACATCTGCGGTAAACGTGAATTCATCGACGCTATCCGTTCT
ATGGCTTCTGCTTTCATCTTCACCACCTCTCTGTCTCCGGTTCTGTGCGCTGGTGCTCTGACCTCTATCAAATACGTTAAAGACCACC
CGGAACTGCGTGAACAGCTGCAGGAACGTGCTCAGAAAACCAAAGAAGAACTGACCCGTCAGGGTATCGAAGTTCTGCAGAACGACTC
TCACATCGTTCCGGTTATCATCGGTGACGCTAAAAAATGCAAAGCTGTTTCTGACGAACTGCTGTACAAAGACGGTATCTACGTTCAG
CCGATCAACTACCCGACCGTTGCTGTTGGTACCGAACGTCTGCGTTTCACCCCGACCCCGTTCCACACCGACATGATGATCTTCGACA
TGGTTGTTAAAGTTAAATCTGCTATGCGTCGTTGCGGTAAAACCAAATAA

>ALASorg

ATGGAACATTTAGATAAATTTAAAAAAATAATAACAGAATTAAAAGCTGACGGTAGATACCGTGTATTCAATGATATCTTACGAACCA
GAGGTAGTTATCCTAACGCAATCTGGTATTCAAAATACTCAATCAAAAAAATAGTCAACTGGTGTTCAAATGATTATCTAGGCATGGG
CCAACATTCTTATGTTATAGATAGCATGAAAACAGCACTGGAGACGAGCGGAGCGGGTGCTGGAGGGACGAGAAACATCTCCGGTACA
ACTCACTATCACATTGCATTAGAACACGAATTAACTCAATTACACGACAAAGAAAGTGCGTTATTATTCACTTCAGCATACAACGCTA
ATCAAACAACTTTAGAAACTATGGGCAAGATTATGCCTGAAATTTTATTTATTTCAGACGCCGAAAATCATTCTTCTATCATACAAGG
ATTAAGACATAGTAAATGTAAAAAAGAAATATTTAAACATAATGATTTAGACGATTTAGAAAGTATCTTAAAGTCTAACCCTGGTCCT
AAATGTGTAGTATTTGAAAGTGTCTATAGTATGGACGGAGATATTGCACCTGTAAAAGAGATTGCTGATTTATGTAAAAAGTATAAAG
CTATATCGTACATTGATGAAGTACACGGAGTTGGTTTATACGGACCAAAAGGTGCTGGTATTTGTGAACGAGATAATGTTGATGTTGA
TATTATTAATGGAACATTAGCAAAGGCCTATGGTGTACAGGGAGGTTACATTTGTGGAAAGAGAGAGTTTATAGACGCAATTAGAAGT
ATGGCAAGTGCGTTTATATTTACAACTAGTTTATCGCCTGTACTATGTGCCGGTGCTTTAACAAGTATTAAGTATGTTAAAGACCATC
CAGAATTAAGAGAACAACTACAAGAACGAGCTCAAAAAACAAAAGAAGAATTAACAAGACAAGGTATAGAAGTATTACAAAACGATAG
TCATATCGTACCAGTAATTATAGGTGACGCTAAAAAATGTAAAGCAGTTTCAGATGAATTACTTTACAAAGATGGTATCTATGTACAA

CCTATTAACTATCCTACGGTTGCTGTTGGTACTGAAAGACTAAGATTTACACCTACACCATTTCATACAGATATGATGATATTTGATA
TGGTAGTTAAAGTTAAATCTGCTATGCGAAGATGTGGTAAAACTAAATGA

>ALAScry (crystal bog phage scaffold 2 498)

ATGTTAAAATCAAAAACACAAGATGCTATTGCACAAAGTCTACAGAAATTAAAATCTGATGGCAATTATCGAGTTTTTACAGATATTC
TGAGAGAGCGCGGTCATTTTCCCAAAGCCATGTTCTATGGCAAATATGGTATTAAACATATTACAAATTGGTGTGCAAATGATTATCT
CGGTATGGCGCAACATAAAGTCGTTATTGATGCTATGCATACTGCGCTCGACACAACTGGTGCTGGTAGTGGTGGTACTCGTAATATA
TCCGGTACTACTCATTTTCATACAGCCTTAGAAAGTGAACTGGCTAGACTACACGATAAAACTGCTGCACTGTTATTCACCAGTGGGT
ATGTTGCTAATCAATCCACACTTAATGTATTAGGTAAAATAATCACAGGAGTTCATTTTATATCCGATAGTATGAATCATAATAGCAT
GATTGTTGGTATGAAAAGCAGTCGAGCACCCTGTACTGTTTGGAATCATAATGATCTAGCTCATCTTAAGGAAATACTGAAAAGTCTA
GACGATGATGCACAACCTATTATTGCCATGGAAGGTGTGTATAGCATGGATGGCGACAAAGGTCTTATATCTGATGTATGTGATATTG
CTCGTCAGTACGGTGCTATGGTTTATGTAGATGAAGTTCACGCTGTAGGTCTTTATGGTGTTCGGGGTGCTGGTATTGCAGAAGAACA
ACATTGTGTAGACGGAGTTGATATCATACAGGGAACATTAGGCAAAGCATTTGGTATACAGGGAGGATATATTGCAGGCGGACGTGAT
CTTATTGATATGATACGCAGCTATGCTAACGGATTTATATTCAGTACCAGTATGAGTCCTGTACTATGTGCAGGTGCACTAGCCAGTG
TGCGTTATGTTCAAGATCATCCAGAAGCAAGAAGTCGCATGTTTGAGGTAGTTGATAAAGTCAAAGAACAATTGGCACAAGAAAACTT
CAATGTCATGCCAGGTAATACACATATTGTACCCCTAATAATAGGCAATGCTGTGCGCTGTAAACAAGTGTCTGATTGGTTATTAGAT
GAACATAATATCTATATACAACCAATTGGATATCCTACAGTCCCTATTGGAACAGAAAGATTACGAATTACCCCTACTCCAAATCATA
CAGATGCAATGATTTGCGATTTACGAAAATCTTTAAGAAATTCTATTGACAAATATATTTAA

>ALAScryopt (JGI adapted)

ATGCTTAAATCTAAAACCCAGGACGCCATTGCGCAGTCTCTGCAGAAACTGAAATCTGATGGTAATTATCGAGTCTTTACCGATATAC
TGAGAGAGCGTGGACATTTTCCCAAGGCAATGTTCTATGGTAAATATGGGATTAAGCATATAACCAATTGGTGTGCTAACGACTATCT
GGGAATGGCCCAGCATAAAGTCGTGATCGATGCTATGCACACGGCGCTGGATACCACTGGCGCAGGCTCTGGGGGCACTCGAAACATC
TCCGGGACCACTCATTTCCATACAGCACTGGAGAGTGAGCTGGCTCGTCTGCATGATAAGACGGCTGCGTTATTGTTTACATCAGGTT
ATGTTGCCAACCAATCAACACTCAATGTGCTGGGTAAAATTATTACAGGGGTGCATTTTATCAGCGACTCAATGAATCACAACAGCAT
GATTGTCGGCATGAAATCATCCAGAGCTCCCTGCACGGTATGGAATCATAATGATCTGGCTCATCTGAAAGAAATATTAAAAAGTCTG
GATGATGATGCGCAACCGATCATAGCTATGGAAGGGGTCTATTCTATGGACGGGGATAAAGGGCTGATCTCCGATGTCTGCGATATTG
CCCGTCAATACGGTGCGATGGTGTATGTGGACGAAGTCCACGCCGTGGGACTGTATGGCGTTCGCGGCGCGGGGATTGCCGAAGAACA
GCACTGCGTTGATGGTGTGGACATTATACAGGGCACCTTAGGCAAGGCGTTCGGTATCCAGGGAGGTTATATTGCCGGAGGCCGGGAT
CTTATTGATATGATTCGCTCTTACGCGAATGGTTTCATTTTTTCAACAAGTATGTCCCCAGTACTGTGTGCTGGGGCCCTGGCATCGG
TTCGTTATGTACAAGACCATCCGGAGGCAAGAAGCCGGATGTTTGAGGTCGTAGATAAAGTAAAAGAACAGTTGGCGCAAGAAAATTT
TAATGTCATGCCTGGAAATACCCACATTGTGCCCTTAATAATTGGTAATGCGGTACGTTGCAAACAAGTATCTGACTGGTTACTTGAT
GAACATAACATTTATATCCAGCCAATTGGTTATCCGACCGTACCTATAGGTACAGAACGCCTTCGTATTACCCCGACACCTAATCATA
CCGACGCTATGATCTGCGATCTGCGCAAAAGCCTTCGCAATTCCATTGATAAATATATATAA

>SI39nov09 (SI)

ATGGAACATTTAGACAAGTTTAAACAAGTAATAACAGAATTAAAAGATGACGGTAGATACCGTGTATTCAATGATATCCTACGAACCA
GAGGTAGTTATCCTAACGCAATCTGGTATTCAAAATACTCAATCAAAAAAATAGTTAACTGGTGTTCAAATGATTATCTAGGTATGGG
ACAGCATTCCTATGTGTTAGATAGTATGAAAACAGCACTGGAGACGAGCGGAGCGGGCGCTGGAGGGACGAGAAACATCTCCGGCACA
ACTCACTATCACATTGCTTTAGAACACGAATTAACTCAATTACACGGCAAAGAAAGTGCGTTATTATTCACTTCAGCATACAATGCTA
ATCAAACAACTTTAGAAACAATGGGCAAGATTATGCCTGACTTGTTGTTTATTTCTGACGCCGAAAATCACTCTTCTATCATACAAGG
ATTAAGACATAGTAAATGTAAAAAAGAAATATTTAAACATAATGATTTAGATGATTTAGAAAGTATCTTAATGTCTAACCCTGGTCCT
AAATGTGTAGTATTTGAAAGTGTGTATAGTATGGACGGAGATATTGCACCTGTAAAAGAGATTGCTGATTTATGTAAAAAGTATAAAG
CTATATCTTATATTGATGAAGTACACGGAGTTGGTTTATACGGACCTAACGGTGCTGGTATTTGTGAACGAGATAAAGTTGATGTTGA
TATTATCAATGGAACATTAGCTAAGGCCTACGGTGTACAGGGAGGTTACATTTGTGGAAAGAGAGAGTTTATAGACGCAATTAGAAGT
ATGGCAAGTGCGTTTATATTTACAACTAGTTTATCGCCTGTACTATGTGCTGGTGCATTAACAAGTATTAAGTATGTTAAAGACCATC
CTGAATTAAGAGAAAAACTACAAGAACGAGCACAAAAAACAAAAGAAGAATTAACAAGACAAGGTATAGAAGTATTACAAAACGATAG
TCATATTGTACCAGTAATTATAGGTGACGCTAAAAAATGTAAAGCAGTTTCAGATGAATTACTTTATAAAGATGGTATCTATGTACAA
CCTATTAATTATCCTACGGTTGCTGTTGGTACTGAAAGATTAAGATTTACACCTACACCATTTCATACAGATACAATGATATTTGATA
TGGTAGTTAAAGTTAAATCTGCTATGCGAAGATGTGGTAGAACTAAATGA

>SI39nov09 (SIopt)

ATGGAGCATCTGGATAAATTTAAACAAGTGATAACTGAGTTGAAAGATGATGGTCGTTATCGTGTTTTTAATGATATTTTGCGTACCA
GAGGAAGCTACCCGAATGCCATCTGGTACAGTAAATATTCTATTAAAAAAATTGTTAATTGGTGTAGCAATGATTATCTGGGTATGGG
CCAGCACTCATACGTGCTCGATAGCATGAAAACTGCTTTGGAAACCTCTGGCGCCGGAGCCGGTGGTACCAGAAACATTTCAGGCACC
ACCCATTACCACATTGCACTTGAACATGAGTTGACTCAGCTTCATGGTAAAGAAAGCGCGTTACTGTTTACCTCGGCATACAATGCTA
ATCAGACTACGCTGGAGACCATGGGCAAAATTATGCCCGATCTTCTGTTTATCTCCGATGCGGAAAATCATAGCTCTATAATTCAAGG
TCTGCGACACAGTAAATGTAAAAAGGAAATCTTTAAGCACAATGACTTGGATGACCTGGAGAGCATACTGATGAGTAATCCAGGTCCT
AAATGCGTGGTTTTCGAAAGCGTTTACTCAATGGACGGTGACATCGCGCCAGTTAAAGAAATAGCGGATCTGTGTAAAAAATATAAAG
CTATTAGCTACATTGATGAAGTGCATGGAGTTGGTCTATATGGTCCGAACGGTGCAGGAATCTGTGAACGTGACAAAGTAGACGTTGA
TATCATTAACGGTACCTTAGCGAAGGCATATGGCGTGCAAGGTGGTTATATATGCGGGAAACGCGAATTTATCGATGCGATTAGGTCA
ATGGCCAGCGCATTTATTTTTACGACTTCCTTATCTCCGGTACTGTGTGCGGGCGCGTTAACTAGCATTAAGTACGTTAAAGATCACC
CAGAACTGCGCGAAAAATTACAGGAGCGTGCACAGAAGACCAAAGAAGAATTAACGCGTCAAGGTATAGAGGTACTTCAGAACGATTC
TCATATTGTGCCGGTCATAATCGGTGATGCTAAAAAGTGCAAAGCCGTTAGCGATGAACTGTTATACAAGGATGGGATATATGTTCAG
CCGATAAACTATCCGACAGTAGCTGTTGGCACGGAGAGACTGCGTTTTACTCCAACGCCATTTCATACCGATACAATGATTTTTGACA
TGGTCGTAAAAGTTAAAAGTGCGATGCGACGTTGTGGGCGTACCAAGTAA

>3300009790_Ga0115012_10000076 761838 (33)

ATGGAACATTTAGATAAATTTAAGGCTGTAATAAAAGATTTAAAAGAAGATGGTAGATACCGTGTCTTTAATGATATTCTACGAACTA
GAGGCGATTATCCTAACGCAATCTGGTATTCAAAATACTCAATCAAAAAAATAGTTAACTGGTGTTCAAATGATTATCTAGGTATGGG
ACAACACTCTTATGTGCTAGATAGTATGAAAACAGCACTGGAGACGAGCGGAGCGGGCGCTGGAGGGACTAGAAACATATCCGGCACT
ACTCACTATCATATTGCTTTAGAACACGAATTAACTCAATTACACAAGAAAGAAAGTGCGTTATTATTCACTTCAGCATACAATGCTA

ATCAAACAACTTTAGAAACAATGGGCAAGATTATGCCTGATATGTTGTTTGTATCAGACGCACAAAATCACTCTTCAATCATACAAGG
TTTAAGGCATAGTAAATGTAGAAAAGAAATATTTAAACATAATGATGTAAAAGATTTAGAAAGTATATTAATGTCTAACCCAGGACCT
AAATGTGTTGTATTTGAAAGTGTATATTCTATGGACGGTGATATTGCACCTGTAAAAGAAATTGTTGAAGTATGTAAAAAATATAATG
CAATAACTTTTATAGATGAAGTACACGCTGTTGGTCTTTACGGTCCGACAGGTGCTGGTATTTGTGAGAGAGACGGTATTGAAGTAGA
TATTATTAACGGAACATTAGCCAAGGCCTACGGTGTACAAGGTGGATACATTGCAGGAAAGAGAGAGTTTATAGACGCAATAAGAAGT
ATGGCAAGTGCGTTTATATTTACAACTAGTTTATCTCCAGTATTATGCGCTGGTGCATTGACTAGTATAAAATATGTCAAAGACCATC
CTGAATTAAGAATGAAATTACAGGAGAGAGCATTAAAAACAAAAGAAGAACTTGAAAGAGCAGGTATAGAAGTATTAAAAAACGATAG
TCATATCGTACCTGTAATTATAGGTGACGCTAAAAAATGTAAAGCAGTTTCAGATGAATTACTTTACAAAAATGGTATCTATGTTCAA
CCAATTAATTATCCTACGGTTGCTGTTGGTACTGAAAGATTAAGATTTACTCCAACACCATTTCATACAGACGCAATGATATTTGATA
TGGTAGTTAAAGTAAAATCTGCTATGAGAAGATGTGGTAATAAGAAATGA

>ALASart

ATGGAACACCTCGACAAATTCAAAAAAATCATCACCGAGCTCAAAGCGGACGGACGTTACCGCGTCTTCAACGACATCCTCCGCACGC
GCGGCTCGTACCCCAACGCGATCTGGTACTCGAAGTACTCGATCAAGAAGATCGTCAACTGGTGCTCGAACGACTATCTGGGCATGGG
CCAGCACTCGTATGTCATCGACTCGATGAAGACCGCGCTGGAGACCTCGGGGGCCGGTGCCGGCGGCACCCGCAACATCTCGGGCACC
ACGCACTATCACATCGCCCTGGAACACGAACTGACGCAGCTGCACGATAAGGAATCCGCGCTTCTTTTCACGTCCGCCTATAATGCCA
ATCAGACGACGCTCGAGACGATGGGCAAGATCATGCCCGAGATCCTGTTCATCTCCGACGCCGAGAACCACTCGTCGATCATCCAGGG
GCTGCGCCACagcAAGTGCAAGAAGGAGATCTTCAAGCACAAtGACCTGGACGACCTGGAgTCgATCCTGAAgTCgAAtCCGGGcCCG
AAgTGCGTCGTCTTCGAATCgGTCTATTCgATGGAtGGcGACATCGCcCCGGTcAAgGAAATCGCcGACCTGTGCAAgAAgTAtAAgG
CcATCTCgTAtATCGACGAAGTcCAtGGcGTcGGcCTGTAtGGcCCcAAAGGcGCGGGcATCTGCGAACGcGACAACGTgGACGTgGA
CATCATCAACGGTACgCTGGCGAAAGCcTACGGcGTcCAGGGcGGcTAtATCTGCGGcAAACGgGAgTTCATCGAtGCGATCCGgTCg
ATGGCcTCcGCGTTCATCTTCACCACCTCgCTGTCgCCGGTTCTGTGCGCGGGcGCGCTGACCTCgATCAAgTAtGTTAAgGACCACC
CGGAgCTGCGcGAgCAGCTGCAGGAgCGcGCGCAaAAgACCAAgGAgGAgCTcACCCGgCAGGGgATCGAgGTTCTcCAGAACGACag
cCACATCGTTCCGGTTATCATCGGcGACGCGAAgAAgTGtAAgGCGGTgTCgGACGAACTtCTGTACAAgGACGGcATCTACGTTCAG
CCGATCAACTACCCGACgGTgGCGGTgGGcACCGAACGgCTGCGTTTCACCCCcACCCCGTTCCAtACCGAtATGATGATCTTCGACA
TGGTTGTTAAgGTTAAgTCgGCcATGCGcCGcTGtGGgAAgACCAAgTgA

>CB2_ho1_JGIadapted

ATGGCTACACTGAAAGAGGTGACTCGTGCAGTCCATGATGAAGCAGAGCACACCGAATGGGGAAAATTGCTGGTCTCTGGAAACATTA
CCAAGGAGCAATATGTACACTATTTGTACAATTTATTAGAGATTCACCAGGCGATAGAATCACGTGGTGTGATAACAATGCCTGAAAT
CCTGAGAGTAAAGGCCATCATTTCGGATATAGAGGCGACCGATAATACCATCTCACCTGTTACACTGGATAGCACTTGGAAATATATA
GCTCATCTGAAAGGCCTTACTGATGATCTGTTGTGGTGCCATATTTATTGTCATTATTTAGGTTATATGTACGGTGGGCAAATGATAA
AAAAGGCGATCCCATTTAGCACGACCATGCTTGACTTTGATAATCGGGCAGAATGCATCGCATACATCCGTCAACATCTGGAGAACGC
CAATCATAATGAGGCTATTGCAGCTTTCAAATGGGCAACTTTTATGTTTAATGATTTGTGGAATCACTATAAAGATGGAAAGTGA

>CB2_ho1_original (was not used in this thesis)

ATGGCAACATTAAAGGAAGTCACTCGTGCCGTCCACGACGAAGCCGAACACACTGAATGGGGTAAGTTACTTGTGAGCGGTAATATCA
CAAAAGAACAATATGTACATTATCTATATAATCTGTTGGAAATTCATCAGGCAATTGAATCACGAGGCGTGATAACCATGCCAGAAAT
ATTGAGAGTTAAGGCAATTATATCAGACATAGAAGCTACTGATAATACGATTAGTCCCGTTACATTAGACAGCACTTGGAAATATATT
GCACATCTTAAAGGTTTGACAGATGATTTACTTTGGTGTCACATTTATTGTCATTATTTGGGGTACATGTATGGCGGACAAATGATCA
AGAAGGCTATACCATTTAGCACAACTATGTTAGATTTTGACAATAGAGCCGAATGTATTGCATATATAAGACAACATCTAGAGAATGC
CAATCACAATGAAGCTATAGCAGCATTTAAATGGGCAACCTTTATGTTTAATGATCTTTGGAATCATTATAAAGATGGAAAGTAA

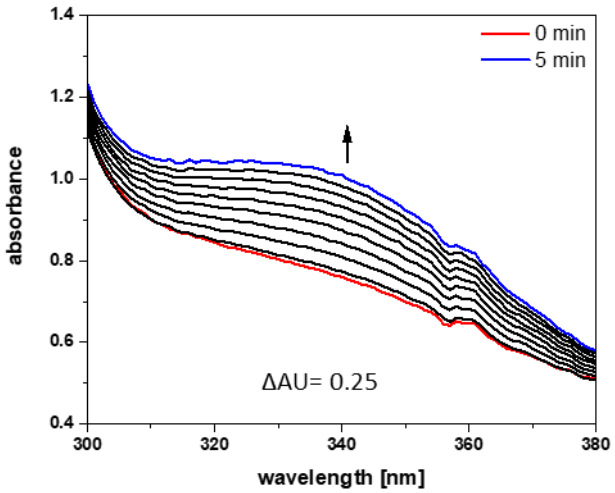
>CB2_pcyA_JGIadapted

GTGACCGATTACATAGAACTACTCGATAAAACCGCAAGACAGTTAAACCACATCATTTCCGCACAACCCGATGTTATTCCAATGGAAA
CCCCGGAATATGGGTGGGAAAATCATCGCCACTCTTCTCCTCAGTTTCGCATGGCCCATGTTGAAATCTTTAACCAAGATCGTTTTAT
GGTTGTGCATTGCTGTGTTTTTCCACATGTCTCCGATCCAGCCCCCATCTTCGGATTTGACGTCATCGCATCGCAGAACAAGGTCACC
GGTGTATTTTTGGATCTCTCTCCCACCGTTGAAGAACCAGGCAAATTCCACAATCTGACCTTTAAACAGCAACGTGAACGCCCTCAGT
GGGGGGATATTTTTTCTCAGAACTGGATTGCGTGTCGGCCGGACAAAACGGAAATGATGAGCATTGTTGCCGAGTCTCAGCGCCTCCT
GAAACATTACTTGACAAATATAATAAATAAACGAACAGCTCCGGTGGACGCGATCAGAGAAGGACAAAACCGTTACTGTAAACAGCAG
CAGCAGAACGAACATACACTCAAAGCTCTTAAAAATCTGATTGGGCCATCGCGTGCAAGGGAATTTATGAACACGATTCTGTTTCCAA
CCGTATAA

>CB2_pcyA_original (was not used in this thesis)

GTGACTGACTATATAGAACTGTTGGACAAAACTGCACGTCAGTTGAATCATATTATATCAGCTCAGCCAGATGTAATACCAATGGAAA
CTCCGGAATACGGTTGGGAAAACCATAGGCATTCGAGTCCACAGTTCCGTATGGCTCACGTGGAAATATTCAATCAGGATCGCTTCAT
GGTGGTACATTGTTGTGTATTTCCACATGTGAGCGATCCTGCTCCAATTTTTGGATTTGACGTTATTGCAAGCCAAAATAAAGTAACA
GGAGTATTTTTAGATCTGTCCCCCACAGTTGAAGAACCTGGAAAATTTCACAATTTAACATTTAAACAGCAGCGCGAACGTCCTCAAT
GGGGTGATATATTCTCACAAAATTGGATAGCATGTCGTCCTGATAAAACTGAAATGATGTCGATTGTTGCAGAAAGTCAACGATTATT
AAAACATTATCTAACCAATATTATTAATAAAAGAACTGCGCCGGTAGATGCAATACGAGAAGGTCAAAATCGCTATTGTAAACAACAG
CAACAAAATGAGCATACACTAAAAGCTCTTAAAAATCTAATTGGTCCTAGCCGTGCTAGAGAATTTATGAACACAATATTGTTCCCAA
CTGTTTAG

IV. Further experimental data



k_M value	literature (RcA + murine)	experimental data
Glycine	0.273 mM – 51 mM	0.194 mM
Succinyl CoA	0.221 µM – 55 µM	0.565 µM
v_{max} value	literature (RcA)	experimental data
	3.74 µM/min	7.961 µM/min

	product formation [µM NADH/min]	enzyme activity [nmol/min/mg]
ALAScry	0.83	9.34
ALAScryopt	0.63	7.09
RcA	7.62	85.83

Figure S 1: Coupled enzyme activity assay to measure spectroscopically the product formation of ALAS from *R. capsulatus* (RcA). The coupled enzyme reaction consists of ALAS forming ALA and CoA-SH out of the substrates glycine and succinyl-CoA. The formation of CoA-SH can be used by the second enzyme reaction, which reduces NAD to NADH and replenish the first reaction with succinyl-CoA. The formation of NADH is stoichiometric with the formation of ALA and can be measured over time by spectroscopic measurements at 340 nm. The total reaction time was 5 min and the increase in absorbance over this time is displayed in absorbance units (AU). Single spectrum was recorded and displayed every 30 s. The arrows indicating the course of the absorbance during the reaction. Diagram shows NADH formation of the purified and concentrated RcA enzyme without GST-tag. Below are the enzymatic values k_M and v_{max} of RcA compared with literature values (Ferreira & Dailey 1993, Hunter & Ferreira 1995, Jordan & Laghai-Newton 1986). Further, production formation and enzyme activity of RcA and viral ALAScry and ALAScryopt were calculated.

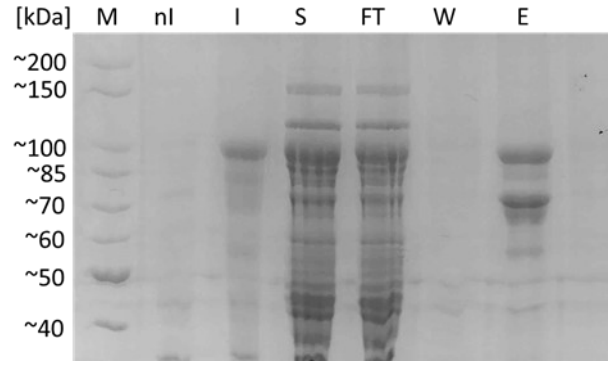


Figure S 2: Improved affinity purification of heterologously produced viral ALAScry of freshwater phage CB_2. The N-terminal fusion protein consists of trigger factor- His_6-tag-ALAScry and was produced in *E. coli* BL21(DE3) in LB medium, the gene sequence encoding ALAScry was codon adapted for *E. coli* codon usage. SDS-PAGE of the purification of fusion protein TF-ALAScry with His_6-tag (MW= ~97 kDa) using Nickel column material for affinity purification with increased salt concentration and imidazole present in all buffers. M= Size marker - Unstained Protein Standard, Broad Range (NEB)*;* nI= Non-induced; I= Induced; S= Lysate; FT= Flow-through; W= Washing fraction; E= Elution fraction.

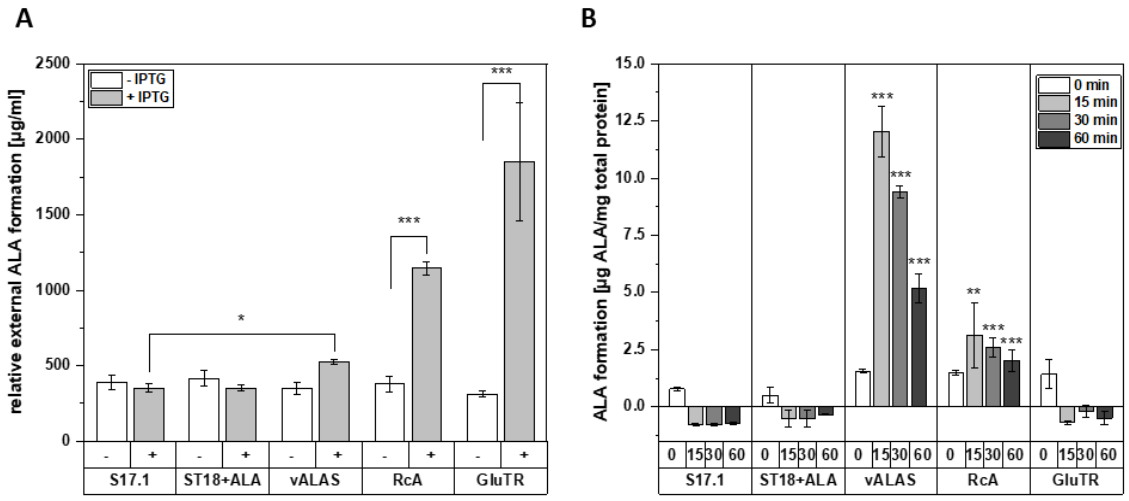


Figure S 3: Viral ALAS leads to enhanced ALA presence in the medium and shows distinct enzyme activity in the cell lysate. A External ALA concentration was measured, in the cell free medium in which the WT (S17.1), deletion mutant strain ST18 or its complemented *E. coli* strains of ST18were cultivated before, to detect the presence of secreted product (ALA) using Ehrlich reagent. **B** Cell lysate [2 mg/ml] of WT (S17.1), deletion mutant strain ST18 or its complemented *E. coli* strains of ST18 were grown in the presence of IPTG and supplemented with specific substrate of C4 pathway enzyme (glycine and succinyl-CoA) to detect product formation over time using Ehrlich reagent. Time points demonstrate the respective amount of ALA that could be measured after subtracting the already present ALA in the cell (time point 0 min). Statistical test: Significance levels were assigned accordingly $p \leq 0.05$ = *; $p \leq 0.01$ = ** and $p \leq 0.001$ = ***.

Table 20: Overview of marine and artificial generated ALAS genes in different vectors. The different marine *alaS* genes displayed in the table were tested for solubility by altering transcription and translation using, e.g., different expression temperatures, inducer concentration, expression duration, addition of rare tRNAs, codon adaption of the native gene to *E. coli* or to RcA. But only one protein showed slight solubility, ALASsyn with an N-terminal GST-tag (MW= 70 kDa). This should be purified for characterization using affinity chromatography.

origin	Synthetic gene	Construct	Utilized tag
Marine	ALASsyn (codon adapted - EMBOSS backtransseq)	pGEX_ALASsyn	N-term. GST
	ALASopt (codon adapted- jCAT)	pGEX_ALASopt	N-term. GST
		pET28_ALASopt	N-term. His_6
	ALASorg	pGEX_ALASorg	N-term. GST
		pET28_ALASorg	N-term. His_6
	ALAS_Sl	pGEX_ALAS_Sl	N-term. GST
	ALAS_Slopt (codon adapted – JGI)	pGEX_ALAS_Sl	N-term. GST
	ALAS_33	pGEX_ALAS_33	N-term. GST
Artificial	ALASart (codon adapted to RcA)	pGEX_ALASart	N-term. GST
		pColdTF_ALASart	N-term. TF +His_6

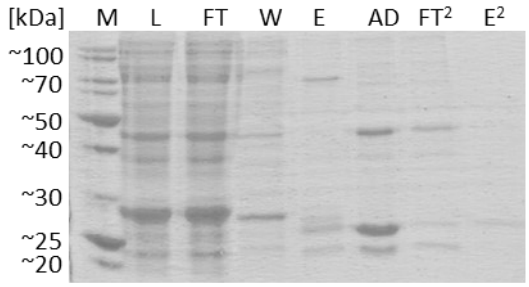


Figure S 4: Purification of heterologously produced viral ALASsyn of marine phage via affinity chromatography. The N-terminal GST-tagged ALASsyn was produced in *E. coli* Codon Plus RIL in LB medium with betaine and sorbitol, the gene sequence of ALASsyn was codon adapted for *E. coli* codon usage. M= Size marker - Unstained Protein Standard, Broad Range (NEB); L= Lysate; FT= Flow-through; W= Washing fraction; E= Elution fraction, AD= After digest. SDS-PAGE of the purification of GST-tagged ALASsyn (MW= ~70 kDa) using glutathione agarose for affinity purification, following cleavage with PreScission Protease and second affinity purification (FT^2, E^2). The cleaved fusion protein ALASsyn (MW= 42 kDa) was further purified from the GST-tag (MW= 26 kDa) and the PreScission protease using glutathione agarose for affinity purification.

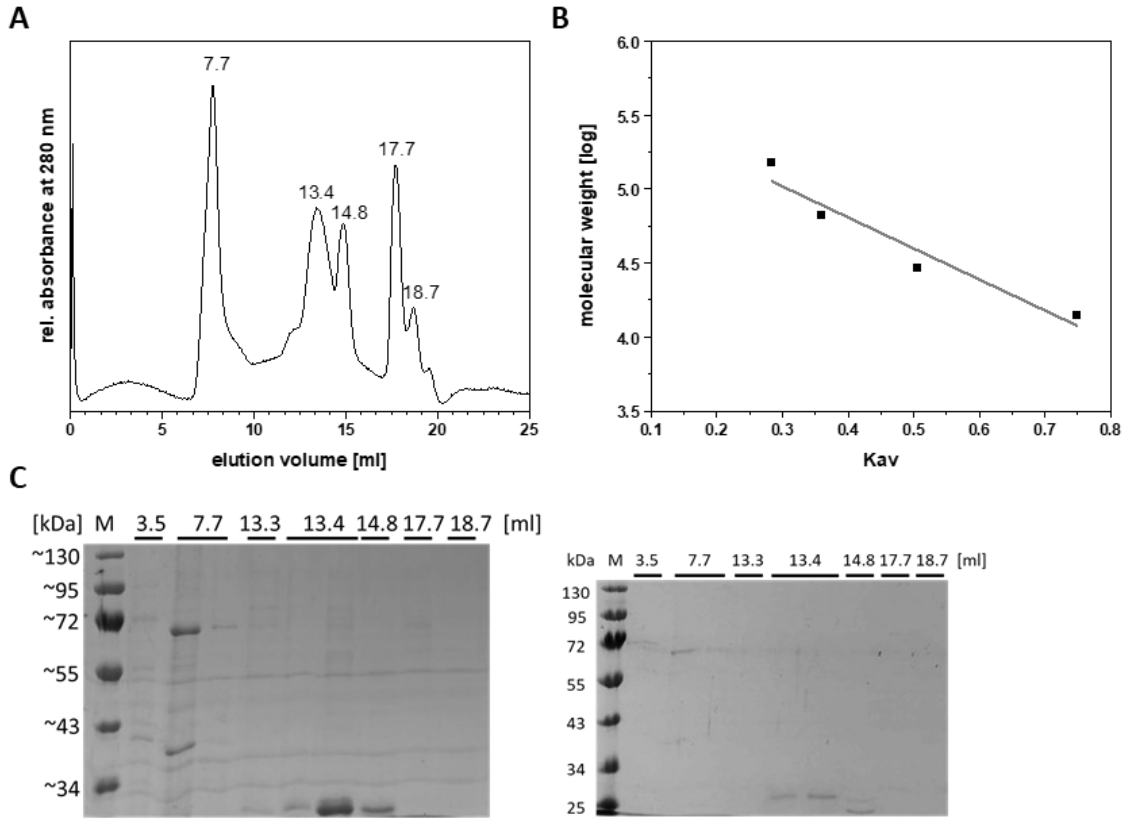


Figure S 5: Determination of oligomerization state of ALASsyn using size exclusion chromatography (SEC). A SEC performed with purified ALASsyn protein. The experiment was carried out using FPLC system Äkta Pure with Superdex 200 10/300 GL column, equilibrated with HEPES buffer (50 mM HEPES, 200 mM NaCl, 10 mM DTT, 20µM PLP, pH 7.0 (NaOH) and purified ALASsyn after affinity purification. Eluting proteins were detected using absorbance measurement at 280 nm. **B** Calibration curve of Superdex 200 10/300 GL column. The standards used for calibration were: alcohol dehydrogenase (150 kDa), albumin (66.5 kDa), carbonic anhydrase (29 kDa) and lysozyme (14 kDa). The void volume (Vo) was determined using Blue dextran (2 000 kDa). For size determination the calibration curve was blotted using linear regression in Origin where the gel-phase distribution coefficient (Kav) was plotted against the logarithm of the molecular weight. Kav = (Ve – Vo)/(Vc – Vo) where Vo was 7.96 ml and Vc (geometric column volume) was 23.56 ml. With the calibration curve the experimental molecular weights could be calculated. **C** SDS-PAGE of collected fractions after SEC, which were concentrated using TCA-DOC precipitation and the corresponding western blot using GST antibodies. Respective antibodies used were goat anti-GST antibody (1:10 000 dilution) for detection of GST-tag, while for detection and visualization of goat IgG the rabbit anti-goat IgG alkaline phosphatase (1:30 000 dilution) was used. M= Color Prestained Protein Standard, Broad Range (NEB).

Enzyme/Control	OD_{554} -blank	OD_{554} - blank – substrate control
$ALAS_{RcA}$	0.361	0.284
$ALAS_{syn1}$	0.099	0.022
$ALAS_{syn2}$	0.058	-0.018
ALA	3.5	--
Background activity	0.029	--

Figure S 6: Colorimetric ALAS enzyme activity assay for measuring viral ALAS activity by spectroscopic determination of product formation. Purified enzyme is dialyzed in HEPES buffer (50 mM HEPES, 200 mM NaCl, 20 µM PLP, pH 7.5 (NaOH). The reaction is carried out adding substrates glycine and succinyl-CoA and the cofactor PLP while incubating at 37 °C for 30 min (RcA) or 1 h (ALASsyn). After termination of the reaction the formed ALA can be detected spectroscopically by inducing pyrrole formation that can be detected using Ehrlich reagent and measured at 554 nm after 20 min. Abbreviation: Blank consists of buffer only; BG= background activity, consists of purified enzyme without substrates, PC= Positive control, consist of ALA only; SC= Substrate control, consists of 5 mM glycine and 100 µM succinyl-CoA used in the assay; RcA= *R. capsulatus* ALAS purified enzyme; ALAS= ALASsyn purified enzyme aggregate. Absorbance measurements at 554 nm are displayed below and calculation of subtraction of blank measurements and substrate control measurements for ALAS activity assay indicate no significant activity of ALASsyn.

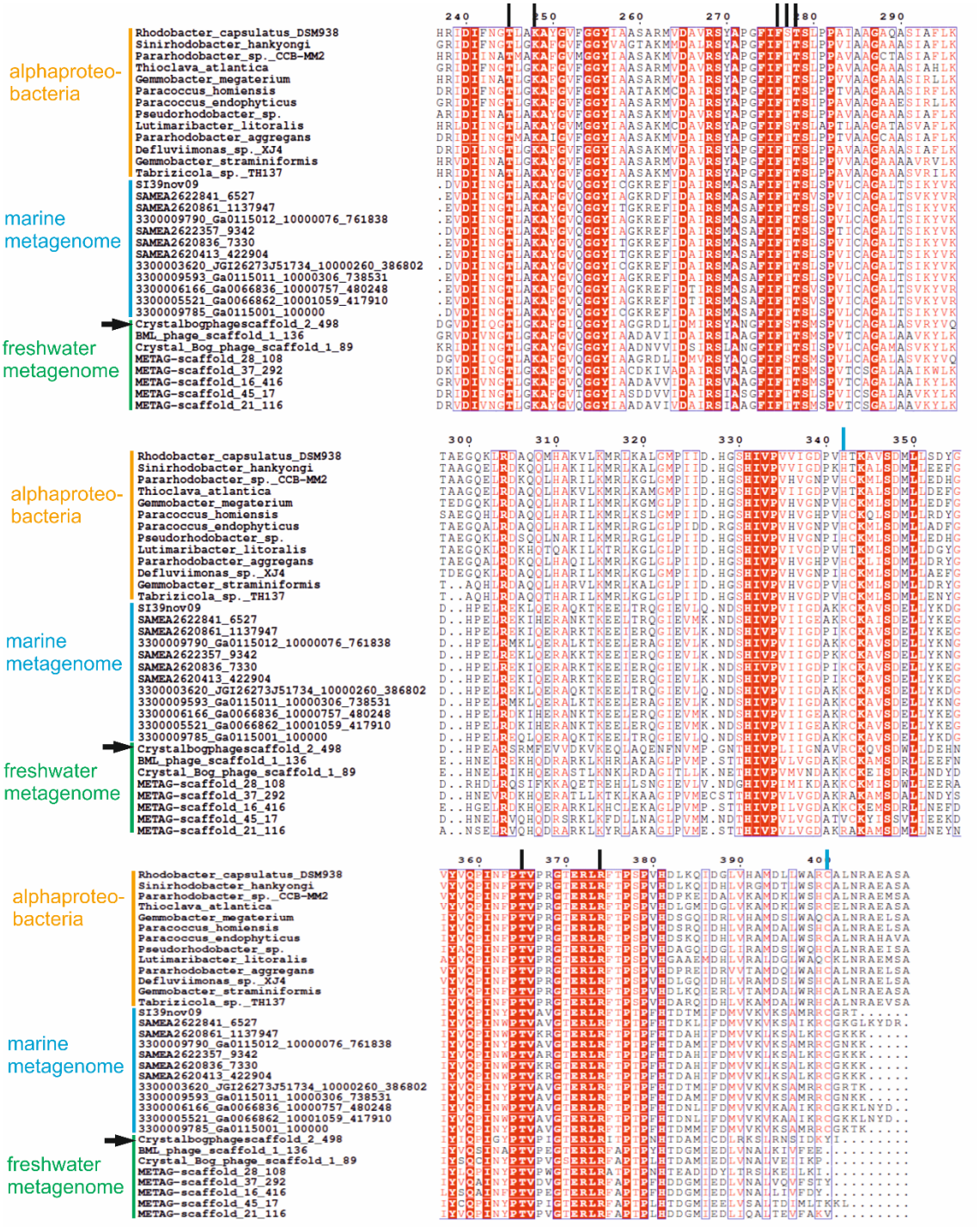


Figure S 7: Amino acid alignment of ALAS sequences from alphaproteobacteria and viral metagenomes revealed conserved catalytic residues. Reference sequences from alphaproteobacteria (shown in yellow) from the NCBI database were aligned with sequences from multiple sampling sites of marine (blue) and freshwater (green) viruses using Clustal Omega and ESPript 3. In this study analyzed ALAS from freshwater metagenome is indicated with a black arrow on the left side of the sequence. Highly conserved amino acids are highlighted in a red square with white lettering, similar amino acids are represented by a white box with red lettering, and non-similar amino acids are indicated by black lettering. Catalytic important amino acids, previously characterized in (Astner *et al.*, 2005), are denoted by black lines at the top. Heme axial ligands of ALAS, known to facilitate heme inhibition, are marked in blue lines at the top (Ikushiro *et al.*, 2018). The respective organisms could not be written in italic due to the restrictions of the program, the correct labelling is in the Appendix I + II.

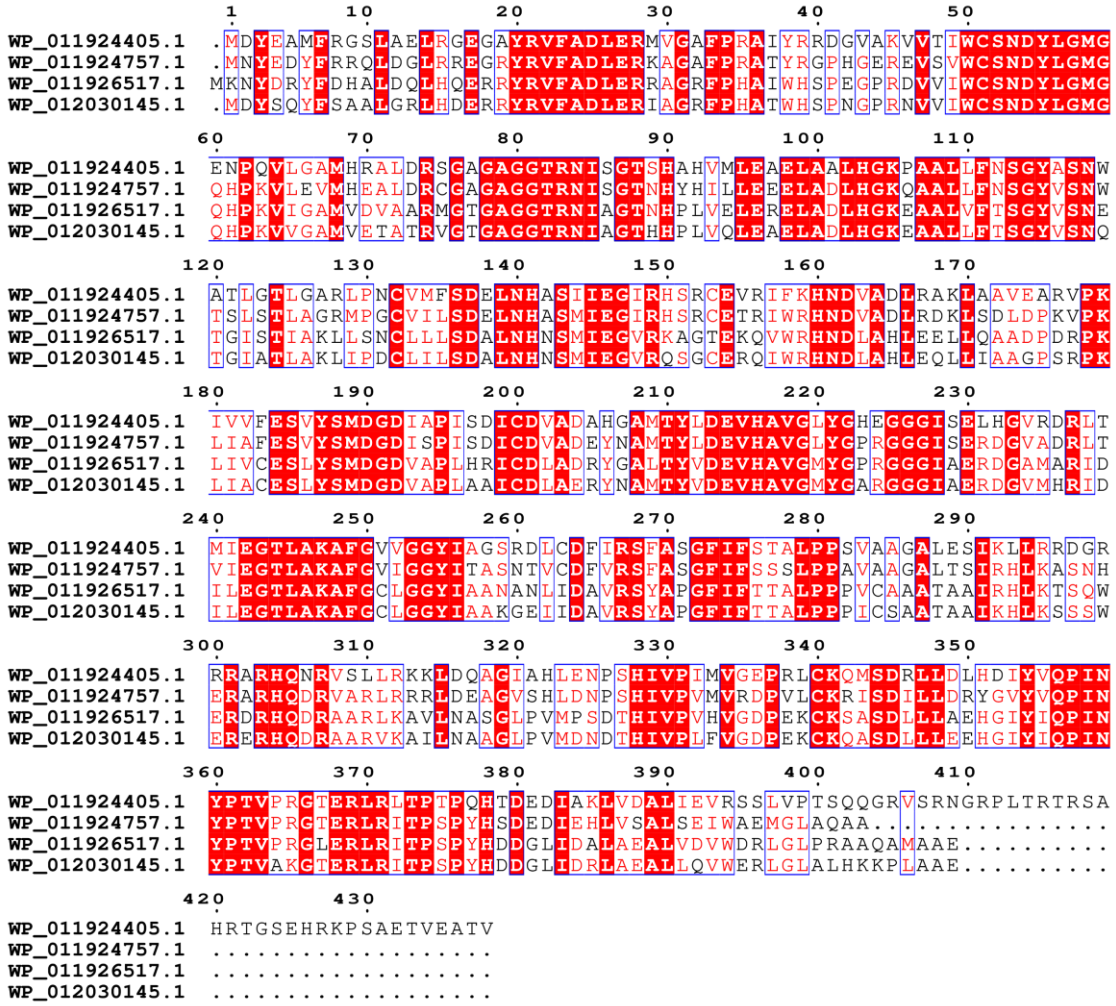


Figure S 8: Amino acid alignment of predicted ALAS proteins present as four genes on the genome of *Bradyrhizobium* sp. ORS 278. On the genomic island located is *alaS* encoding WP_011924405.1 (NCBI Reference Sequence) from 1354962-1356140 bp , while all other three genes are not part of a genomic island (as indicated by IslandViewer 4; https://www.pathogenomics.sfu.ca/islandviewer/). WP_011924757.1 from 1767763-1768974 bp, WP_011926517.1 from 3785771-3787003 bp, and WP_012030145.1 from 6932074-6933303 bp. Amino acid sequences where derived from the NCBI database and can be found in Appendix II. They were aligned using Clustal Omega and ESPript 3. Highly conserved amino acids are highlighted in a red square with white lettering, similar amino acids are represented by a white box with red lettering, and non-similar amino acids are indicated by black lettering.

Curriculum Vitae

Persönliche Daten

Helen Wegner

Berufserfahrung

02/2024	08/2024	**Teamleitung mikrobiologisches Labor** Nestlé Health Science, Osthofen
03/2020	12/2023	**Wissenschaftliche Mitarbeiterin (RPTU, Kaiserslautern)** Doktorandin in der Abteilung Mikrobiologie
08/2017	02/2018	**Wissenschaftliche Hilfskraft (TUK, Kaiserslautern)** Abteilung Phytopathologie

Studium und schulische Laufbahn

10/2017	02/2020	**Master Studium (TUK, Kaiserslautern)** Biology - Microbial and Plant Biotechnology
10/2013	07/2017	**Bachelor Studium (TU KL, Kaiserslautern)** Biowissenschaften
06/2010	06/2013	**Allgemeines Abitur (Marie Baum Schule, Heidelberg)** Biotechnologisches Gymnasium

Konferenzteilnahmen (Auswahl)

Tetrapyrrole Discussion Group 2023	Poster Präsentation - Auszeichnung mit Poster Preis Canterbury, Vereinigtes Königreich
Fachbereichsseminar 2022	Wissenschaftlicher Vortrag Thallichtenberg, Deutschland
ProSyn Fest 2022	Poster Präsentation Cordoba, Spanien
Geomicrobiology Workshop 2019	Betreuung und organisatorische Unterstützung Kaiserslautern, Deutschland

Publikationen

Publikation 2024	Helen Wegner *et al.* 2024. *Nature Communications* accepted for publication
Publikation 2021	Natascha Tomazic, Kristina E. Overkamp, Helen Wegner, *et al.* 2021. *Biochimica et Biophysica Acta* (BBA) - Bioenergetics, Volume 1862, Issue 12.

Eidesstattliche Erklärung

Erklärungen lt. § 6 (4) der Promotionsordnung des Fachbereichs Biologie vom 28.11.2018:

Ich erkläre wahrheitsgemäß, dass ich die eingereichte Dissertation selbst angefertigt und alle für die Arbeit benutzten Hilfsmittel und Hilfestellungen in der Arbeit angegeben habe. Die Dissertation oder Teile hiervon habe ich bisher nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht. Ebenso habe ich die gleiche oder eine andere Abhandlung nicht bei einem anderen Fachbereich oder einer anderen Universität als Dissertation eingereicht.

_____

Ort, Datum, Unterschrift

Darlegung des Eigenanteil

Helen Wegner

Functional characterization of viral auxiliary metabolic genes involved in tetrapyrrole biosynthesis

Die Konzeption, die Planung, das Versuchsdesign und die Methodenentwicklung der vorliegenden Arbeit wurde in Teilen von Prof. Dr. Nicole Frankenberg-Dinkel und PD Dr. Susanne Zehner unterstützt. Die Literaturrecherche, die Interpretation und die Diskussion der Ergebnisse wurden, sofern im Folgenden nicht anders dargelegt, nicht von Dritten unterstützt. Teile der bioinformatischen Datenanalyse von Metagenomdaten wurden durchgeführt von Dr. Anne Kupczok und Dr. Sheila Roitman. Die Abbildung der Contigs wurde von Dr. Sheila Roitman erstellt. Die Datenerhebung wurde in Teilen durch Dr. Jason Woodhouse, Prof. Dr. Hans-Peter Grossart und Prof. Dr. Oded Béjà ermöglicht. Dr. Jason Woodhouse und Prof. Dr. Hans-Peter Grossart erstellten das LIMNOS Dataset durch Probennahme, -aufbereitung, Analyse und Auswertung. Drei Plasmide wurden von Vanessa Braun kloniert, alle weiteren verwendeten Plasmide die nicht selbst kloniert worden sind entsprechend gekennzeichnet.

Teile dieser Arbeit liegen zur Veröffentlichung vor: “Identification of Shemin pathway genes for tetrapyrrole biosynthesis in bacteriophage sequences from aquatic environments” Helen Wegner, Sheila Roitman, Anne Kupczok, Vanessa Braun, Jason Woodhouse, Hans-Peter Grossart, Susanne Zehner, Oded Béjà, Nicole Frankenberg-Dinkel. Nature Communications, akzeptiert für Veröffentlichung.

Die ersten Entwürfe für diese Veröffentlichung wurden von der Autorin dieser Dissertation angefertigt, die Fertigstellung des Manuskripts bis zur Einreichung wurde von Prof. Dr. Nicole Frankenberg-Dinkel mehrheitlich unterstützt.

Die vorliegende Einschätzung über die erbrachte Leistung von Dritten wurde mit den genannten Personen einvernehmlich abgestimmt.

_____ _____

Datum, Unterschrift Doktorandin Datum, Unterschrift Betreuerin

Darlegung aller benutzten Hilfsmittel und Hilfestellungen

Helen Wegner

Functional characterization of viral auxiliary metabolic genes involved in tetrapyrrole biosynthesis

Die (statistische) Analyse und Visualisierung der meisten Daten erfolgte mit Hilfe von Microsoft Excel (Microsoft, Version 2021), Origin (OriginLab, Version 2021) und Corel DRAW Graphics Suite (Corel Corporation, Version 2023). Einige Abbildungen wurden mit BioRender (BioRender.com) hergestellt und entsprechend kennzeichnet, die Veröffentlichungsrechte dafür liegen vor. *In silico* Analysen wurden mit Clustal Omega (Sievers *et al*. 2011), ESPript v.3.0 (Robert and Gouet 2014), Mafft, AlphaFold2 (Jumper *et al*., 2021) mit ColabFold (Mirdita *et al*. 2022), PyMOL v.2.0 (Schrödinger and DeLano 2020), Next Generation Phylogeny.fr über den Webserver (NGPhylogeny.fr, https://ngphylogeny.fr), Mol* über den Webserver (RCSB.org) und iTOL (interactive Tree of Life, version 6.8) über den Webserver itol.embl.de (Letunic & Bork, 2021) durchgeführt. Chemische Darstellungen wurden mit Hilfe von ChemDraw (PerkinElmer, Version 5) und ChemSketch (Advanced Chemistry Development, Version 2023 1.2) erstellt.

Diese Arbeit wurde von Steffen Heck (Doktorand, RPTU, Mikrobiologie) Korrektur gelesen.

_____

Datum, Unterschrift Doktorandin

Danksagung

An dieser Stelle möchte ich mich bei all denjenigen bedanken, die mich während der Anfertigung dieser Doktorarbeit unterstützt und motiviert haben.

Mein herzlicher Dank gilt meiner Doktormutter Prof. Dr. Nicole Frankenberg-Dinkel für die Möglichkeit, an einem so faszinierenden Projekt arbeiten zu dürfen. Speziell bedanken möchte ich mich für die hilfreichen Anregungen und die konstruktive Kritik bei der Erstellung dieser Arbeit.

Bei Prof. Dr. Stefanie Müller-Schüssele bedanke ich mich für die freundliche Übernahme des Korreferats.

Außerdem danke ich Prof. Dr. Michael Schroda für die Bereitschaft der Übernahme des Vorsitzes der Prüfungskommission.

Bedanken möchte ich mich bei der Arbeitsgruppe Mikrobiologie der RPTU. Die früheren und momentanen Arbeitskollegen haben eine spannende Zeit im Labor und außerhalb ermöglicht. Besonderen Dank möchte ich PD Dr. Susanne Zehner, Dr. Eugenio Pérez Patallo und Dr. Michelle Gehringer aussprechen, für die umfassende Unterstützung und fruchtvollen Diskussionen. Mein größter Dank geht an Steffen Heck, ohne dich wäre die Zeit nicht halb so schön gewesen.

Meinem Vater möchte ich danken für seinen emotionalen Rückhalt, besonders sein Interesse an meiner Arbeit hat mich beflügelt und weiter motiviert.